**CHAPTER 13**

**SCENES** FROM THE FIELD

It has been thoroughly demonstrated that regular physical activity is associated with better physical well-being and overall mental health. Since both of these outcomes are often identified as problematic to achieve in older adults, studying nursing interventions that can improve an older adult’s health and well-being is important. Khazaee-pool et al. (2015) conducted a randomized controlled trial to determine if a program of regular exercise could improve happiness in older adults.

These nurse researchers recruited volunteers from a public park. Each person was then checked against inclusion and exclusion criteria, and a final sample of 120 adults was identified and randomly assigned to either the experimental group, who participated in an 8-week physical exercise program, or a control group, who received no specific physical activity recommendations.

Variables that were collected included demographic data and responses on the Oxford Happiness Inventory. The latter instrument takes a broad look at personal happiness, and its results have been determined to be a reliable and valid reflection of satisfaction with life, positive mood, mental health, self-esteem, and efficiency. The measure itself is an ordinal, Likert-type scale.

Some interesting data were revealed by the researchers’ descriptive analysis. For example, a significant inverse relationship between age and level of happiness was identified before the experiment began, meaning older subjects were less happy. This lent credence to the need for interventions for this age group. Females were happier than males, but overall those persons with lower incomes reported being less happy. There was a significant inverse relationship between being dependent on others and lower happiness—a relationship that was found to reverse itself after the exercise program ended. This was a particularly important finding, in that older people who are active feel less dependent on others for their day-to-day care.

Almost all of the scale scores improved after members of the experimental group engaged in the 8-week program of physical activity. Self-esteem, life satisfaction, and global happiness showed the greatest improvements, but statistically significant differences were found in the results of all six scales that the researchers administered.

The results of this study were reported in a way that made their interpretation relatively easy. Two tables were used to display demographic characteristics of the subjects and the results of statistical tests. The authors reported test statistics as well as *p* values. They noted which findings were clinically significant and focused the discussion section on the nursing implications of the findings. Enough data were provided for the reader to get a general sense of the magnitude of differences and to determine which findings were large enough to be clinically important.

It has been clear for some time that participating in a regular program of physical activity has wide-ranging implications for health as we age. This study indicates that mental health and happiness can be improved through physical activity as well. These findings—the result of a single randomized trial—represent evidence that can support nursing practices in the care of older adults.

Khazaee-pool, M., Sadeghi, R., Majilessi, F., & Rahimi, F. (2015). Effects of physical exercise programme on happiness among older people. *Journal of Psychiatric & Mental Health Nursing, 22*(1), 47–57.

**Introduction**

If descriptive analysis answers the question “What is going on?”, then inferential analysis answers the question “Are you sure?” The word *inferential* means that the reader can infer something about a population’s response to an intervention based on the responses of a carefully selected sample. Inference requires the calculation of numerical values to enhance the researcher’s and reader’s confidence that the intervention resulted in the outcome and to rule out the possibility that something else did. In quantitative inferential analysis, that “something else” is error in all its forms—sampling error, measurement error, standard error, even random error—and inferential analysis allows the researcher to quantify its effect.

When reading quantitative analysis, it is important to focus on both the probability of error and the certainty of the estimates. When quantitative analysis is used as evidence for nursing practice, the nurse should also consider the size of the effect and determine whether it attains both statistical and clinical significance. When creating quantitative analysis, the focus is on choosing the correct test and making appropriate decisions about its application. All of these factors are critical for ensuring that the relationship between intervention and outcome is one that is defined with certainty so the results can be expected to be replicated in different settings among different people.

**Some General Rules of Quantitative Analysis**

Quantitative analysis is a highly systematic and organized process. Data for quantitative analysis are represented numerically, so the reliability of data collection, accuracy of data entry, and appropriateness of analytic processes are critical for drawing the correct conclusions. Because these types of analyses can be complex, a plan for analysis and reporting is determined when the research methods and procedures are designed. A wide variety of statistical tests are available, each of which has specific requirements for data preparation and analysis. There are, however, some general guidelines for conducting all types of quantitative analyses:

* *Select tests* a priori. The specific statistical tests that will be used for quantitative analysis are selected before the experiment begins. The selection of specific tests is based on the research question to be answered, the level of measurement, the number of groups, and the nature of the data. Selecting the tests before the data are generated reduces the chance that a researcher might use a test that is less conservative or more favorable by eliminating this source of researcher bias.
* *Run all the tests identified*. The researcher must run all tests that were identified *a priori*. Looking at the data and then deciding which tests to run can create bias. Although the specific version of a test may be dictated by the nature of the data (for example, using a test for non-normal data), the researcher should not pick and choose tests after reviewing the data.
* *Report all the tests that were run*. The researcher must report the results from each test that was run. Selectively reporting or retaining data to support a personal viewpoint is a form of researcher bias and is unethical.

**Types of Quantitative Analysis**

Quantitative analysis refers to the analysis of numbers. Quantifying the values of variables involves counting and measuring them; these counts and measures result in numbers that can be mathematically manipulated to reveal information. The researcher can think of quantitative analysis as an interpreter that takes raw data and turns it into something understandable.

Many types of quantitative analyses are possible. Those available to the researcher can be categorized in several ways: by the goals of the analysis, by the assumptions made about the data, and by the number of variables involved.

**Goals of the Analysis**

Quantitative analyses are useful for many research goals. In particular, research questions that focus on evaluating differences between groups (for example, between an experimental group and a control group) are amenable to quantitative analysis. This type of research leads to some of the strongest studies when it comes to presenting evidence for practice. Quantitative tests are appropriate to assess the nature and direction of relationships between subjects or variables, including the capacity to predict an outcome given a set of characteristics or events. Researchers also use quantitative methods to sort data; for example, a clinician may identify characteristics that enable him or her to differentiate people at risk for falls from those who are not at risk. Quantitative analyses aid in data reduction by grouping variables into overall classifications, which then helps researchers determine, as an example, which clusters of symptoms may predict complications.

Quantitative analyses are also classified as descriptive or inferential based on the aims of the study. Descriptive studies are concerned with accurately describing the characteristics of a sample or population. Inferential analyses are used to determine if results found in a sample can be applied to a population—a condition necessary for confidently generalizing research as a basis for evidence for nursing practice ([**Schmidt & Brown, 2012**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_bib8)).

**Assumptions of the Data**

Quantitative analyses are generally grouped into two major categories based on assumptions about the data: parametric and nonparametric. The key differentiating factor for these categories is the assumptions made about the distribution, or shape, of the data. **Parametric tests** are based on the assumption that the data fall into a specified distribution—usually the normal (bell-shaped) distribution. This assumption holds only when interval- or ratio-level measures are collected or when samples are large enough to achieve normality. In reported healthcare research, parametric tests are the most common, even when doubt exists about whether the basic assumptions have been met. Other tests, however, are specific to data that are not normally distributed. If a normal distribution cannot be assumed, then such **nonparametric tests** are needed. This group of tests is “distribution free,” meaning the tests do not rely on a specific distribution to generate accurate results. Nonparametric tests are becoming more common in research, particularly in health care, where many variables are not normally distributed.

**Parametric tests:** Statistical tests that are appropriate for data that are normally distributed (that is, fall in a bell curve).

**Nonparametric tests:** Statistical tests that make no assumptions about the distribution of the data.

Parametric tests are usually desirable because they are sensitive to relatively small differences, they are commonly available in most software packages, and they are readily recognizable by the reader. Nevertheless, they are sometimes applied erroneously to data sets for which the distribution of the data has not been shown to be normal; in such a case, they may result in misleading conclusions. Small deviations from normality may be acceptable with these tests because most parametric tests are **robust tests**, or capable of yielding reliable results even when their underlying assumptions have been violated. Nevertheless, some data are so non-normal as to require a test that makes no such assumptions.

**Robust tests:** Statistical tests that are able to yield reliable results even if their underlying assumptions are violated.

Compared to parametric tests, nonparametric tests are not as commonly applied, are less recognizable, and are not always available in analytic packages. These tests are also relatively insensitive and require large samples to run effectively. When possible, researchers should strive to collect data in a form that can be expected to be normally distributed and to create sample sizes that allow them to use parametric tests. However, researchers should use the appropriate category of test for the data, meaning they should specifically evaluate the distribution of the results prior to making the final decision on which tests to use. Most parametric tests have a nonparametric counterpart, enabling researchers to apply the correct class of test to the data.

**Number of Variables in the Analysis**

Quantitative analyses can be classified in terms of the number of variables that are to be considered. In practice, such tests are usually classified by both the number and the type of variables involved.

**Univariate analysis** involves a single variable. Such analyses are the primary focus of descriptive and summary statistics. The term univariate analysis may also be applied when the study involves a single dependent variable or when only one group is included. For example, differentiating whether blood pressure is affected more by exercise in the morning or the evening is a univariate analysis; that is, even though two groups (morning and evening) are included in the study, the analysis focuses on a single dependent variable—blood pressure.

**Univariate analysis:** Analysis of a single variable in descriptive statistics or a single dependent variable in inferential analysis.

**Bivariate analysis** is the analysis of the relationship between two variables. The most common form of bivariate analysis is correlation. Bivariate analysis is also used to determine if a single variable can predict a specified outcome. For example, determining if blood pressure is associated with sodium intake is a bivariate analysis. In this case, two variables—blood pressure and sodium—are analyzed to determine any relationship between them.

**Bivariate analysis:** Analysis of two variables at a time, as in correlation studies.

**Multivariate analysis** is the simultaneous analysis of multiple variables. This endeavor may address the effects of multiple predictors on a single outcome, the differences between groups on several effects, or the relationships between multiple factors on multiple outcomes. For example, determining if blood pressure is different in the morning or evening, and is associated with sodium intake, weight, and stress level, is an example of a multivariate analysis.

**Multivariate analysis:** The simultaneous analysis of multiple variables.

These analyses become more sophisticated and complex as more variables are added to either side of the equation. On the one hand, a research study may require a simple univariate analysis to achieve a descriptive goal. On the other hand, an experiment may require the complexities of a full factorial multivariate analysis of variance—a calculation of the effect of multiple factors on multiple outcomes, taking into account both their main effects and the effects that occur when factors interact. Quantitative analyses can accommodate the complexity of human responses to interventions and illness by reflecting the multivariate nature of health and illness.

**An Overview of Quantitative Analysis**

**Inferential analysis** is undertaken to determine if a specific result can be expected to occur in a larger population, given that it was observed in a sample. In statistics, a sample (part of the population) is used to represent a target population (all of the population); the question then becomes whether the same results found in the sample would be found in the larger target population. Quantitative research as evidence for practice is useful only when it can be generalized to larger groups of patients than those who were directly studied in the experiment. Inferential analysis allows the nurse to recommend that an intervention be used and to do so with an identified level of confidence that it is evidence based.

**Inferential analysis:** Statistical tests to determine if results found in a sample are representative of a larger population.

Inferential analysis is fundamentally an analysis of differences that occur between samples and populations, between groups, or over time because something changed. In experimental research, the change is an intervention. In case-control studies, the “change” is a risk factor; in causal-comparative studies, it is a characteristic or an event. Inference is used to determine if an outcome was affected by the change.

It is not enough, however, to see a difference between two samples and assume that the difference is the same as would be expected in a larger population. Samples, by their very nature, are different than the populations from which they were drawn. Samples are made up of individuals and, particularly in small samples, we cannot be sure that the sample exactly matches the population’s characteristics. These differences—the ones that are due to the sampling process—are quantified as **standard error**. One might view standard error as the differences between samples and populations that are expected simply due to the nature of sampling.

**Standard error:** The error that arises from the sampling procedure; it is directly affected by variability and indirectly affected by sample size.

**Statistical Significance**

Historically, researchers were interested in finding an objective way to decide if differences between treatment groups and control groups were important. Comparing the differences to standard error made sense; it told the researcher if the change was real. For many decades, researchers have relied on the statistic that is yielded by this comparison—the probability of standard error—to determine if changes are different from random errors ([**Scott & Mazhindu, 2014**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_bib9)). This comparison of observed differences to standard error forms the basis for most inferential tests.

The certainty with which a researcher can say “these effects are not due to standard error” is subject to rules of probability. The calculations produce a *p* value, or the probability the results were due to standard error. If the *p* value is very small, then the probability that the results were due to error is very small, and the researcher can be very confident that the effects of the intervention are real. The largest a *p* value can be, and still be considered significant, is 0.05, or 5%. If the *p* value is very large (greater than 0.05 or 5%), then the probability that the results were due to error is very large, and the researcher cannot conclude that the intervention had an effect greater than would be expected from random variations. When the *p* value is very small, indicating that the probability the results were due to chance is also very small, then the test is said to have *statistical significance*. It is the comparison of differences to standard error and the calculation of the probability of error that give inferential analysis its strength. Nevertheless, statistical significance is just one of the important measures that determine whether research is truly applicable to practice. Many readers rely solely on the *p* value to determine whether results are useful as evidence for practice; this value is often misinterpreted as scientific evidence of an effect, when it is actually a measure of error. Significance testing does not eliminate the need for careful thought and judgment about the relevance of the findings for clinical practice.

Inferential analysis yields more than a *p* value, and it is these other statistics—the test statistic, confidence intervals, and difference scores—that enable the researcher to quantify whether the difference is important (clinical significance). Statistical significance tells us the findings are *real*; clinical significance tells us if the results are *important for practice*. This type of analysis forms the basis of some of the strongest scientific evidence for effective nursing practices.

**Clinical Significance**

In evidence-based practice, the concern for statistical significance has been augmented with a broader focus on clinical significance. Clinical significance is generally expected to reflect the extent to which an intervention can make a real difference in patients’ lives. Clinicians are more frequently interested in these kinds of findings than in whether the result was due to chance. However, while well-established means are available to assess statistical significance, no single measure can identify a result’s clinical significance. Nevertheless, several statistics can inform the clinical evaluation of the importance of a finding, including confidence intervals, minimum important difference (MID), and effect size.

Estimates of population values can be expressed as **point estimates** or interval estimates. A point estimate represents a single number. In other words, a researcher measures a value in a sample, calculates a statistic, and then concludes that the population value must be exactly that number. In reality, samples rarely produce statistics that exactly mimic the population value (called a parameter). In this respect, a point estimate is less accurate than an estimate that includes a range of numbers. The likelihood that a single point will match the population value is quite small; the likelihood that the actual value could be captured in a range of numbers is considerably better. This range of numbers represents a **confidence interval** and is used to estimate population parameters.

**Point estimate:** A statistic derived from a sample that is used to represent a population parameter.

**Confidence interval:** A range of values that includes, with a specified level of confidence, the actual population parameter.

A confidence interval enables the researcher to estimate a specific value, but it also provides an idea of the range within which the value might occur by chance. It is defined as the range of values that the researcher believes, with a specified level of confidence, contains the actual population value. Although it sounds counterintuitive, confidence intervals are more accurate in representing population parameters than are point estimates because of the increased likelihood that an interval will contain the actual value.

Confidence intervals are helpful in quantitative analysis in several ways. The calculation of a confidence interval takes into account the effects of standard error, so the probability that the interval actually contains the population value can be determined. Confidence intervals allow the measurement of sampling error on an outcome and express it in numeric terms. They provide more information than *p* values do; they enable the evaluation of magnitude of effect so the nurse can determine whether results are large enough to be of clinical importance. Confidence intervals are constructed using several pieces of information about the data, combined with a decision made by the researcher regarding how much confidence is necessary. Although such intervals may be constructed around any statistic, the mean is typically used for this purpose. Confidence intervals for mean estimates and for estimates of the mean differences between groups are popular ways to measure experimental differences, and they represent a critical aspect of application of research as evidence for practice. Their construction requires that the researcher calculate the mean value and then determine the range around the mean that is due to standard error. This range is affected by the level of confidence in the results that the researcher must have (usually 95% or 99%), the amount of variability among subjects, and the size of the sample.

[**FIGURE 13.1**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_fig1) indicates how a confidence interval of the mean is constructed. At the most fundamental level, confidence intervals show the range of possible differences in effect caused by the intervention. They help the nurse determine whether the observed differences suggest true benefits or just minor changes. The confidence interval is particularly helpful in determining how close the treatment came to *no difference*. For example, a confidence interval for a weight-loss strategy might be from 0.25 to 2.0, meaning that the strategy could result in a loss of as much as 2 pounds or as little as 0.25 pound. Such an interval is helpful in judging evidence about which treatment might be statistically sound but has little clinically meaningful effect.

An additional advantage of the confidence interval is that it is reported on the same relative scale as the outcome itself. In other words, a *p* value of 0.02 tells us nothing about how effective a treatment was; its scale matches that of probability, ranging from 0 to 1.0. The confidence interval, in contrast, is reported on the original scale of the outcome measure, so it is more meaningful ([**Pandis, 2013**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_bib7)). For example, knowing that “the weight lost as a result of the treatment was between 3.5 and 6 pounds” gives the nurse valuable information for counseling patients and judging the success of an intervention. [**Table 13.1**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_tbl1) explains the interpretation of confidence intervals.

A measure of clinical significance that is being more commonly reported in study write-ups is the MID. This clinically derived number represents the smallest difference in scores between the experimental and control groups that would be of interest as evidence for practice. For example, a weight change of 4 ounces may be of no importance in a study of adult males, but it would most certainly be clinically notable in a sample of preterm infants. There are three commonly accepted ways to find the MID:

**FIGURE 13.1** What Makes Up a Confidence Interval?

**Table 13.1** Confidence Intervals Interpreted: An Example

| **Statistic** | **Confidence Interval** | **What It Means** |
| --- | --- | --- |
| Mean number of distressful symptoms reported by patients with terminal cancer | 3.5 to 4.1 symptoms | Patients with terminal cancer have, on average, between 3.5 and 4.1 distressful symptoms, inclusive. |
| Proportion of patients with terminal cancer who report distressful symptoms | 79.2% to 88.7% | Between 79.2% and 88.7% of patients with terminal cancer report distressful symptoms, inclusive. |
| Mean difference between number of symptoms experienced by men and women as distressful | −0.24 to 1.4 symptoms | The average difference between the number of symptoms reported as distressful by men and women could be nothing (zero, which appears in this interval). There are no statistical differences between these groups. |
| Mean difference between number of symptoms experienced by members of a treatment group and members of a control group | −1.2 to -0.3 symptoms | A nursing intervention had the effect of reducing the number of symptoms perceived as distressful. The decrease could be as little as 0.3 of a symptom or as much as 1.2 symptoms. There is a significant effect, but it is very small. |
| Mean difference between number of symptoms experienced by those with early-stage disease and those with late-stage disease | 2.1 to 4.3 symptoms | People late in the terminal stages of cancer experience more distressful symptoms than those in early stages of the same disease. They could experience as few as 2.1 symptoms or as many as 4.3 symptoms, which demonstrates a significant effect of considerable magnitude. This is a clinically meaningful finding. |

1. Compare the change in the outcome to some other measure of chance (anchor based).
2. Compare the change to a sampling distribution to determine its probability.
3. Consult an expert panel ([**Fethney, 2010**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_bib1)).

The MID is determined prior to statistical analysis. The confidence interval is reviewed; if it contains the MID value, then the results are clinically significant. If the confidence interval does not include the MID, however, then the results are not considered clinically significant, whether they are statistically significant or not.

To see how the MID is used in practice, suppose a researcher testing the effects of massage therapy on hypertension consults the literature and an expert panel and determines that a change of 5 mm Hg in blood pressure would be considered clinically significant. If the confidence interval for the change were 1 to 3 mm Hg, the interval does not include 5 mm Hg, so the result is not clinically significant. If the confidence interval were 2 to 10 mm Hg, however, then the researcher could conclude that the findings were clinically significant.

Clinical importance can also be represented by **effect size**, meaning the size of the differences between experimental and control groups. Effect size can provide the nurse with a yardstick by indicating how much of an effect can be expected and, therefore, it provides more information for evidence-based practice than statistical significance alone.

**Effect size:** The size of the differences between experimental and control groups compared to variability; an indication of the clinical importance of a finding.

Effect size is calculated in many different ways, but all have in common a formula that takes into account the size of differences and the effects of variability. Interpreting effect size is relatively easy: Larger numbers represent a stronger effect, and smaller numbers represent a weak one. Effect size can also be discerned from a confidence interval. If the confidence interval for a mean difference includes zero or nears zero on one end, for example, then the difference could be nothing. In such a case, clinical significance becomes a critical consideration.

Both statistical significance and clinical significance are needed to ensure that a change in practice is warranted based on the study results. A change in practice should not be recommended if the researcher cannot demonstrate that the results are real—in other words, that the results are statistically significant. Change also is not warranted if the results are not important enough to warrant the effort involved in changing a practice. Effect size is needed to draw this conclusion. No single statistic can be used to establish the usefulness of a study; all of the results taken together provide the nurse with evidence that can be applied confidently.

**Designs That Lend Themselves to Quantitative Analysis**

Some research questions naturally lend themselves to quantitative analysis. Experimental, quasi-experimental, causal-comparative, and case-control designs are all particularly well suited to inferential analysis, for example. A research question must lend itself to collection of numerical data to be appropriate for quantitative testing, and a relationship, difference, or effect must be the focus of the study. Qualitative studies are not appropriate candidates for statistical testing because—unlike quantitative research—they do not involve analysis of numbers.

Predictive and correlation studies also yield statistics, but they do not necessarily focus on differences in groups. A correlation coefficient yields a *p* value that represents the probability that the relationship observed is not due to standard error. That relationship, however, may be a weak one, so statistical significance is less important for a correlation study than interpretation of the size and direction of the coefficient. Predictive studies involving regression also yield quantitative output. The regression itself is compared to the mean to determine the value of the prediction. Thus a *p* value in a regression simply indicates that the values are related in a linear way and that the regression line is a better predictor than the mean.

The primary consideration in any quantitative analysis is appropriate interpretation. That is, the researcher’s interpretation of the results should not exceed what the data will support. Appropriately reporting statistical tests and associated effect sizes, along with confidence intervals, is a complete and accurate way to draw conclusions about evidence for nursing practice. All three of these results should be incorporated into the quantitative research report.

**Selecting the Appropriate Quantitative Test**

Conducting an appropriate quantitative analysis depends on the ability of the researcher to select the appropriate statistical test. This is often the most daunting part of the analysis process because many tests are available to the nurse researcher. Each has specific requirements and yields particular information, which may then be appropriate in specific circumstances. The researcher must make decisions about the appropriateness of a statistical test based on the following factors:

* The requirements of the research question
* The number of groups to be tested
* The level of measurement of the independent and dependent variables
* The statistical and mathematical assumptions of the test

Each of these elements is considered in determining which group of tests to select and, from that group, which particular version will best answer the question without producing misleading results. Questions about differences between two groups, for example, will be answered with different tests than would be used with three or four groups. Likewise, interval data, which are assumed to be normally distributed, are tested differently than nominal data, which must be represented as proportions or rates. All statistical tests have assumptions; the data must meet the assumptions for the results to be interpreted correctly.

The tests most commonly used in intervention research are tests of means and proportions. When dependent variables are measured as interval numbers (for example, heart rate, length of stay, and cost), then the mean value can be calculated and compared. When dependent variables represent nominal or ordinal data, then frequencies, rates, or proportions are tested. In each case, tests may be conducted to determine if differences exist between two, three, or more groups.

**Tests of Differences Between Two Group Means**

Frequently, the research question of interest is whether an intervention, risk factor, or condition made a difference in a specific outcome between two groups. The typical experiment, in which the outcomes for a treatment group are compared to those for a control or comparison group, falls into this category. If the outcome can be expressed as a mean, then a *z* or *t* test is appropriate to use for these differences. These statistical tests indicate whether the differences in mean values between two groups are statistically significant and clinically important. [**FIGURE 13.2**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_fig2) depicts the decision process that results in a *z* or *t* test for a quantitative analysis.

The two tests essentially accomplish the same end: They generate a statistic that reflects the differences between the groups compared to standard error—a *p* value that quantifies the probability that standard error is responsible for the outcome and a confidence interval for mean differences that enables the quantification of effect size. The *z* test is appropriate for large samples or when testing an entire population. The *t* test is best used with smaller samples, generally including fewer than 30 subjects ([**Hoffman, 2015**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_bib4)). [**FIGURE 13.3**](https://jigsaw.vitalsource.com/books/9781284138887/epub/EPUB/xhtml/26_Chapter13.xhtml#ch13_fig3) shows how the *t* test is calculated and the calculations that are represented by each part of the formula.

**FIGURE 13.2** A Decision Tool for a *z* or *t* Test

**FIGURE 13.3**