implications of using data. Digital tools provide an easy, quick form of data collection. However, use of the digital tools may be limited because of (a) limitations involving the use of listservs and obtaining e-mail addresses, (b) limitations of the technology itself, (c) lack of a population list, and (d) the questionable representativeness of the sample data (Mertler, 2001).

How to Decide What Types to Choose

Confronted by these many options for collecting quantitative data, which one or ones will you use? To select your data sources, ask yourself the following questions:

- *What am I trying to learn about participants from my research questions and hypotheses?* If you are trying to learn about individual behaviors of parents at a student–parent conference meeting, you could use a behavioral checklist and record observations. If you are trying to measure the attitudes of teachers toward a bond issue, attitudinal questions or an attitudinal instrument will be required.
- *What information can you realistically collect?* Some types of data may not be collectible in a study because individuals are unwilling to supply them. For example, precise data on the frequency of substance abuse in middle schools may be difficult to collect; identifying the number of student suspensions for substance abuse is much more realistic.
- *How do the advantages of the data collection compare with its disadvantages?* In our discussion of each data source, we have talked about the ideal situations for data collection. Given the ease or difficulty of collecting data, each type needs to be assessed.

How would you now advise that Maria collect her data? Assume that she now seeks to answer the general quantitative research question "Why do students carry weapons in high school?" and the following subquestions:

a. "How frequently do students feel weapons are carried into high school?"
b. "What general attitudes do high school students hold toward the possession of weapons in the schools?"
c. "Does participation in extracurricular activities at school influence attitudes of students toward possession of weapons?"
d. "Are student suspensions for possession of weapons on the increase in high schools?"

Before looking at the answers provided, list the type of information that Maria might collect for subquestions a through d.

To answer these subquestions, Maria first needs to locate or develop a questionnaire to send out to a sample of high school students in the school district. Her data collection will consist mainly of attitudinal data. This questionnaire will measure student *attitudes* toward frequency of weapon possession (question a), assess student *attitudes* toward possession of weapons (question b), and gather *factual data* about the students (question c), such as age, level of education, race, gender, and extent of participation in extracurricular activities. To answer question d, she will contact the school officials of several high schools and ask if she can obtain reports on student suspensions—school documents that report quantitative data. In summary, she will collect both attitudinal and factual information.

MyLab Education Self-Check 5.4

MyLab Education Application Exercise 5.3: Collecting Quantitative Data: Collecting Data

- Look as well at how the researchers have interpreted (discussed if the instrument measured what it is intended to measure) the scores in light of their intended use
- Evaluate whether the authors provide good evidence that links their interpretation to their use

What types of evidence should researchers seek to establish validity? Impara (2010) provided a useful summary of AERA et al.'s *Standards* (1999). He directed readers to examine closely Chapter 1 from the *Standards* on "validity," and then presented an extended list of examples of evidence to document validity. Only a few of the examples are mentioned here.

The *Standards* mention five categories of evidence as shown in Table 5.3: evidence based on test content, response processes, internal structure, relations to other variables, and the consequences of testing. In the discussion to follow, the word "testing" will be equivalent to "instrument."

TABLE 5.3

Sources of Validity Evidence and Examples

Validity Evidence	Types of Tests or Instruments to Which Validity Evidence Is Applicable	Type of Evidence Sought	Examples of Evidence
Evidence based on test content	Achievement tests, credentialing tests, and employment tests	Evidence of an analysis of the test's content (e.g., themes, wording, and format) and the construct it is intended to measure	• Examine logical or empirical evidence (e.g., syllabi, textbooks, and teachers' lesson plans) • Have experts in the area judge
Evidence based on response processes	Tests that assess cognitive processes, rate behaviors, and require observations	Evidence of the fit between the construct and how individuals taking the test actually performed	• Interviews with individuals taking tests to report what they experienced/were thinking • Interviews or other data with observers to determine if all are responding to the same stimulus in the same way
Evidence based on internal structure	Applicable to all tests	Evidence of the relationship among test items, test parts, and the dimensions of the test	• Statistical analysis to determine if factor structure (scales) relates to theory, correlation of items
Evidence based on relations to other variables	Applicable to all tests	Evidence of the relationship of test scores to variables external to the test	• Correlations of scores with tests measuring the same or different constructs (convergent/discriminant validity) • Correlations with scores and some external criterion (e.g., performance assessment—test-criterion validity) • Correlations of tests scores and their prediction of a criterion based on cumulative databases (called meta-analysis—validity generalization)
Evidence based on the consequences of testing	Applicable to all tests	Evidence of the intended and unintended consequences of the test	• Benefits of the test for positive treatments for therapy, for placement of workers in suitable jobs, for prevention of unqualified individuals from entering a profession, for improvement of classroom instructional practices, and so forth

Source: Adapted from Impara (2010) and AERA et al. (1999).

TABLE 5.4

Types of Scales Used in Quantitative Research

Type of Scale	Examples of Questions Using the Scale
Nominal scale (uses categories)	How much education have you completed? _____ No college _____ Some college _____ Bachelor's degree _____ Graduate or professional work What is your class rank? _____ Freshman _____ Sophomore _____ Junior _____ Senior
Ordinal scale (uses categories that imply or express rank order)	Has your adviser helped you select courses? _____ Not at all _____ To a small extent _____ To some extent _____ To a great extent _____ To a very great extent Rank your preference for type of graduate-level instruction from 1 to 4. _____ Activity-based learning _____ Lecture _____ Small-group learning _____ Discussion
Quasi-interval or interval/ratio scale (uses continuous equal intervals)	School is a place where I am thought of as a person who matters. _____ Strongly agree _____ Agree _____ Undecided _____ Disagree _____ Strongly disagree Colleges and universities should conduct research to solve economic problems of cities. _____ Strongly agree _____ Agree _____ Undecided _____ Disagree _____ Strongly disagree

in data analysis programs) also have two types: interval/quasi-interval and ratio scales. These types of scales are shown in Table 5.4.

Scales of measurement are response options to questions that measure (or observe) variables in categorical or continuous units. It is important to understand scales of measurement to assess the quality of an instrument and to determine the appropriate statistics to use in data analysis.

Nominal Scales Researchers use **nominal (or categorical) scales** to provide response options where participants check one or more categories that describe their traits, attributes, or characteristics. These scales do not have any order. An example of a nominal scale would be gender, divided into the two categories of male and female (either one could be listed first as a response option). Another form of a nominal scale would be a checklist of "yes" or "no" responses. A semantic differential scale, popular in psychological research, is another type of nominal scale. This scale consists of bipolar adjectives that the participant uses to check his or her position. For example, in a psychological study of talented teenagers, researchers were interested in studying the teenagers' emotional responses to their everyday activities (Csikszentmihalyi, Rathunde, Whalen, & Wong, 1993). The researchers used a semantic differential scale for teenagers

HOW DO YOU ANALYZE THE DATA?

After you prepare and organize the data, you are ready to analyze it. You analyze the data to address each one of your research questions or hypotheses. Questions or hypotheses in quantitative research require that you do the following:

- Describe trends in the data for a single variable or question on your instrument (e.g., "What is the self-esteem of middle school students?"). To answer this question, we need **descriptive statistics** that indicate general tendencies in the data (mean, median, and mode); the spread of scores (variance, standard deviation, and range); or a comparison of how one score relates to all others (z scores or percentile rank). We might seek to describe any of our variables: independent, dependent, control, or mediating.
- Compare two or more groups on the independent variable in terms of the dependent variable (e.g., "How do boys and girls compare in their self-esteem?"). To answer this question, we need **inferential statistics** in which we analyze data from a sample to draw conclusions about an unknown population. We assess whether the differences of groups (their means) or the relationship among variables is much greater or less than what we would expect for the total population if we could study the entire population.
- Relate two or more variables (e.g., "Does self-esteem relate to an optimistic attitude?"). To answer this question, we also use inferential statistics. Alternatively, we could test a hypothesis about the relationship of variables (e.g., "Self-esteem predicts an optimistic attitude among middle school children") using inferential statistics.
- Examine the relationship among variables through more advanced statistical procedures, such as regression analysis, meta-analysis, factors analysis, discriminant function analysis, path analysis, structural equation modeling or hierarchical linear modeling. These advanced statistical procedures are beyond the scope of this book but are introduced in Chapter 11.

Thus, we describe results to a single variable or question, or infer results from a sample to a population. In all quantitative research questions or hypotheses, we study individuals sampled from a population. However, in descriptive questions, we study only a single variable at a time; in inferential analysis, we analyze multiple variables at the same time. In addition, from comparing groups or relating variables, we can make predictions about the variables. We can test hypotheses that make predictions comparing groups or relating variables.

Conduct Descriptive Analysis

How do we analyze the data to describe trends? We use **statistics**, the calculations of values based on numbers. Many helpful books provide details about different statistics, their computation, and their assumptions (e.g., Field, 2013; Gravetter & Wallnau, 2013; Triola, 2018). We focus here on the statistics typically used in educational research.

Choosing a Descriptive Statistics Test

Descriptive statistics will help you summarize the overall trends or tendencies in your data, provide an understanding of how varied your scores might be, and provide insight into where one score stands in comparison with others. These three ideas are the central tendency, variability, and relative standing. Figure 6.3 portrays the statistical procedures that you can use to provide this information.

A measure of relative standing is the percentile rank. A **percentile rank** of a particular score is the percentage of participants in the distribution with scores at or below a particular score. You use it to determine where in a distribution of scores an individual's score lies in comparison with other scores. In Table 6.3, we see that an individual with a score of 94 is at the 80th percentile, with 20% of the participants having scores above this individual and 80% of the participants having scores at or below this individual.

Another measure of relative standing is the **standard score**, a calculated score that enables a researcher to compare scores from different scales. It involves the transformation of a raw score into a score with relative meaning. A **z score** is a popular form of the standard score, and it has a mean of 0 and a standard deviation of 1. This yields a z score, or a standard score that has the advantage of enabling you to compare scores from one instrument to scores from another instrument. Using standardized scores is also central to calculating many statistics. The procedure is to take a score, subtract it from the mean, and divide it by the standard deviation. In Table 6.3, we see that a person with a score of 60 has a z score of -1.57, or a score that is more than one and a half standard deviations below the average (or mean).

Conduct Inferential Analysis

Descriptive statistics help you analyze descriptive questions. However, when you compare groups or relate two or more variables, inferential analysis comes into play. The basic idea is to look at scores from a sample and use the results to draw inferences or make predictions about the population. Often, you cannot study the entire population because of size and cost, so we instead examine a sample that has been carefully chosen from the population.

When you study this sample and obtain scores, several approaches exist for determining if the sample scores you receive are a good estimate of the population scores (e.g., see Vogt & Johnson, 2016). Ask yourself the following:

1. Is the sample score (e.g., the mean difference between two groups) probably a *wrong* estimate of the population mean? The procedure you use to examine this question is hypothesis testing. **Hypothesis testing** is a procedure for making decisions about results by comparing an observed value of a sample with a population value to determine if no difference or relationship exists between the values. This is the traditional way to test whether the sample mean is a good estimate of the population mean. It provides a yes–no answer: Either the sample mean is a good estimate, or it is not. Because we can never absolutely prove that the sample is a good estimate, we try to establish whether it is a *wrong* estimate.

2. How confident are you that your sample score is right? This is the confidence interval approach. A **confidence interval** or **interval estimate** is the range of upper and lower statistical values that is consistent with observed data and is likely to contain the actual population mean. In this approach, you determine an interval or range in which your population score would likely fall. In this sense, confidence intervals give us more flexibility than the yes–no options of hypothesis testing.

3. Does the sample score or differences between two groups make practical sense? This is the effect size approach. **Effect size** is a means for identifying the practical strength of the conclusions about group differences or about the relationship among variables in a quantitative study. Effect sizes tell us how different the sample values are and allow us to make a judgment as to whether this is meaningful based on our knowledge of measures, the participants, and the data collection effort.

The reason we have more than one approach is that some researchers have felt that a yes–no hypothesis testing answer to our quantitative questions and hypotheses leads to

misinterpretations and errors (Finch, Cumming, & T
and effect sizes provide a more practical reading
it is useful to report all three estimates of your po
dence interval, and the effect size (Wilkinson & Ta

Hypothesis Testing

There are five steps in hypothesis testing: (a) ide
(b) set the level of significance, or alpha level, (c
statistic; and (e) make a decision about rejecting

1. *Identify your null and alternative hypothe*
about the population and is typically stated using
relationship" or "no association"). The alternative
ence (or relationship or association), and the dir
or negative (alternative directional hypotheses)
nondirectional hypotheses).

Returning to the data for high school stude
alternative hypothesis as follows:

Null Hypothesis:

There is no difference between smokers

Alternative Hypothesis (nondirectio

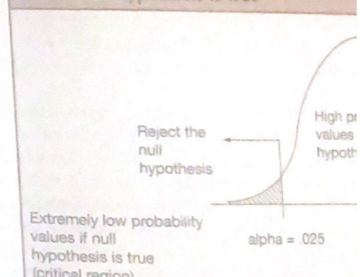
There is a difference between smokers
written another way:

Smokers are more depressed than nons

2. *Set the level of significance, or alpha l*
were to collect a large number of sample me
difference"), the theoretical distribution wo
curve, as illustrated in Figure 6.6. In this f

FIGURE 6.6

Normal Curve Distribution of Mean Differen
If the Null Hypothesis Is True



Source: Adapted from Gravetter, F. J., & Wallnau, L
Belmont, CA: Thomson Learning. Reprinted with p
of Cengage Learning; permission conveyed throu