Selecting the Right Statistical Test

16

LEARNING OBJECTIVES

- Define the different tasks of statistics.
- Select the correct statistical test.

CHAPTER OVERVIEW

Most people have a toolbox with a few tools in it. There's probably a hammer, a screwdriver, and a pair of pliers. Most people know that if a screw has to be tightened, a pair of pliers won't be much use, so they reach for the screwdriver. Statistical tests are tools—pliers, screwdrivers, and hammers—designed for different research jobs. This book has filled your statistical toolbox with a set of tools for the most commonly encountered statistical jobs. Now it is time to learn how to pick the right statistical tool for the particular research job at hand.

16.1 Review of Statistical Tasks

The tool that the statistician selects depends on the task, and there are three different types of tasks:

1. Summarizing a set of data
2. Deciding if a group of cases differs from another group of cases on some underlying parameter
3. Determining if two variables vary together systematically

The first type of task involves what are called *descriptive statistics*, statistics used to describe a set of observations. Reporting that the average length of stay in a psychiatric hospital is two weeks is an example of a descriptive statistic.

The other two tasks for statisticians involve *difference tests* and *relationship tests*. Difference tests look for differences in some underlying parameter among groups of cases. For example, reporting that it takes longer, on average, to treat someone for depression than for anxiety would be the result of a difference test. Relationship tests see if variables correlate. Finding that the longer a person has suffered from depression tends to increase the difficulty in treating it would be the result of a relationship test.

Corty_3e_Ch16_617-642_Final HR.indd 617 21/01/16 7:41 pm

Chapter 1 covered three different types of studies—experimental designs, quasi-experimental designs, and correlational designs. Table 16.1 summarizes their similarities and differences. Knowing what type of study was conducted is important because different types of studies call for different statistics:

- *Descriptive studies* call for the use of descriptive statistics.
- *Experimental designs* and *quasi-experimental designs* divide participants into groups, so they are analyzed with difference tests.
- *Correlational designs,* which measure two variables in a single sample of participants, call for relationship tests.

TABLE 16.1 Comparing Correlational, Experimental, and Quasi-Experimental Studies

	Correlational	Experimental	Quasi-Experimental
Is the explanatory variable manipulated/ controlled?	No	Yes	No
Cases are . . .	Measured for naturally occurring variability on both variables	Assigned to groups by an experimenter using random assignment	Classified in groups on the basis of naturally occurring status
The study can draw a firm conclusion about cause and effect.	No	Yes	No
Do you need to worry whether confounding variables exist?	Yes	No	Yes
Question being asked by the study	Is there a relationship between the two variables?	Do the different populations possess different amounts of the dependent variable?	Do the different populations possess different amounts of the dependent variable?
Advantage of study	A correlational design can study conditions that can't be manipulated.	An experimental study can draw conclusions about cause and effect.	A quasi-experimental design can study conditions that can't be manipulated.

The flowchart in Figure 16.1 shows how to determine whether a study is experimental, quasi-experimental, or correlational, as well as whether it is descriptive.

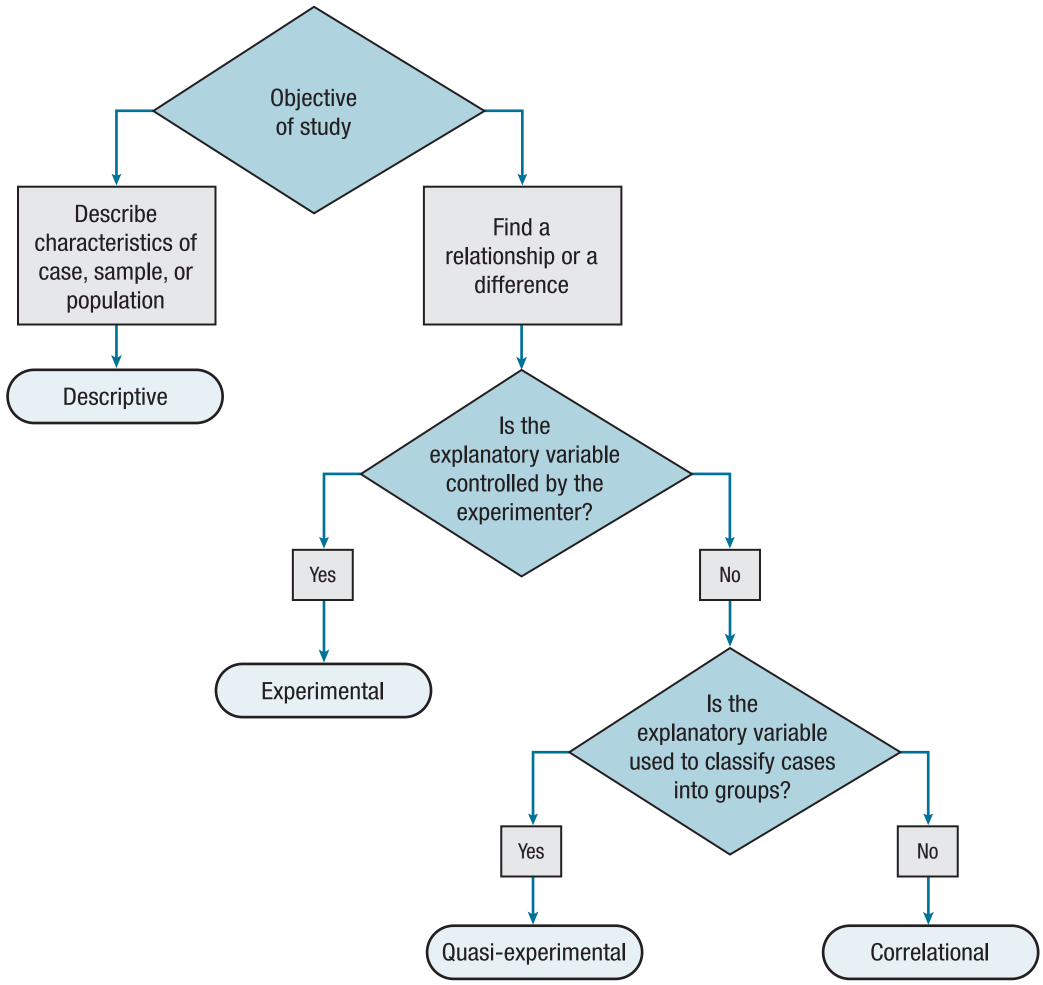


Figure 16.1 How to Choose: Type of Study This flowchart can be used to determine whether a study is descriptive, correlational, experimental, or quasi-experimental. The type of study depends on whether the explanatory variable is manipulated by the experimenter and, if the explanatory variable is not manipulated, whether cases are classified into groups.

The flowchart in Figure 16.1 may involve making a decision about the explanatory variable, often called the independent variable, in a study. Studies often have two variables, an explanatory variable and an outcome variable, commonly referred to, respectively, as the independent variable and the dependent variable. It is important to be able to tell which is which. The explanatory variable is the one used to assign or classify cases into groups or to predict the outcome, and the outcome variable is the one where the outcome of the study is measured.

16.2 Descriptive Statistics

The most basic task in statistics is taking a set of numbers and summarizing them in some way. Statistical techniques that do this are called **descriptive statistics** because these techniques are used to describe a set of cases.

The flowchart in Figure 16.2 walks through how to choose the correct descriptive statistic. Choosing the correct descriptive statistic depends on three things:

1. At what level the variable is measured: nominal, ordinal, or interval/ratio
2. What is being described: an individual case or a sample of cases
3. What is to be described: central tendency or variability

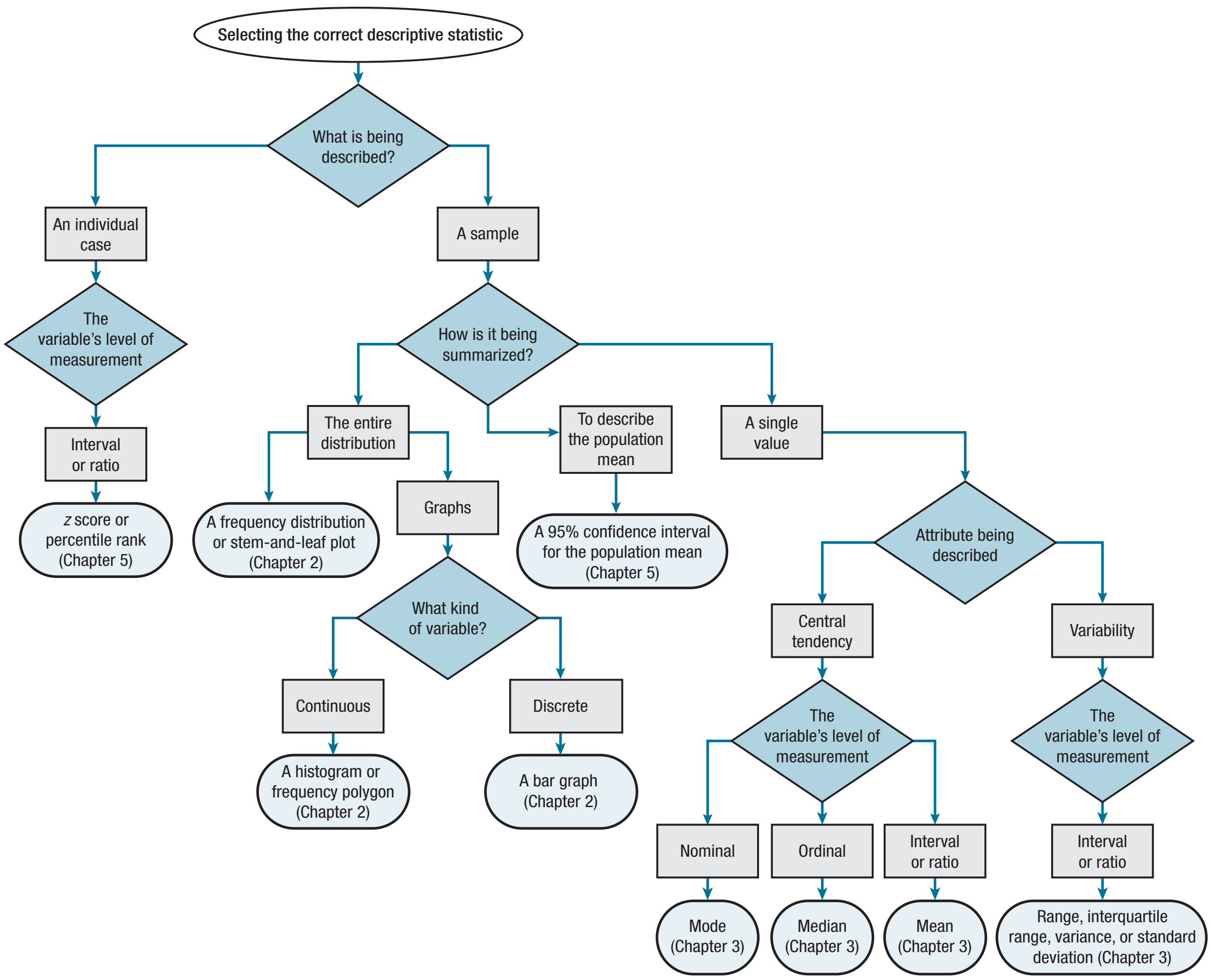


Figure 16.2 How to Choose: Selecting a Descriptive Statistic This flowchart is used to select the correct descriptive statistic, depending on (1) whether one is trying to describe an individual case, a sample, or a population; (2) what attribute one wants to describe (central tendency or variability); and (3) the variable's level of measurement (nominal, ordinal, interval, or ratio).

Knowing the level of measurement for the variable—nominal, ordinal, interval, or ratio—is important in statistics (see Chapter 1). Table 16.2 shows the distinctions among the four levels of measurement, and the flowchart in Figure 16.3 leads one through the steps for determining the level of measurement of a variable.

TABLE 16.2 Level of Measurement: Information Contained in Numbers

	Same/Different	Direction of Difference (more/less)	Amount of Distance (equality of units)	Proportion (absolute zero point)
Nominal	✓			
Ordinal	✓	✓		
Interval	✓	✓	✓	
Ratio	✓	✓	✓	✓

As levels of measurement become more complex, numbers contain more information.

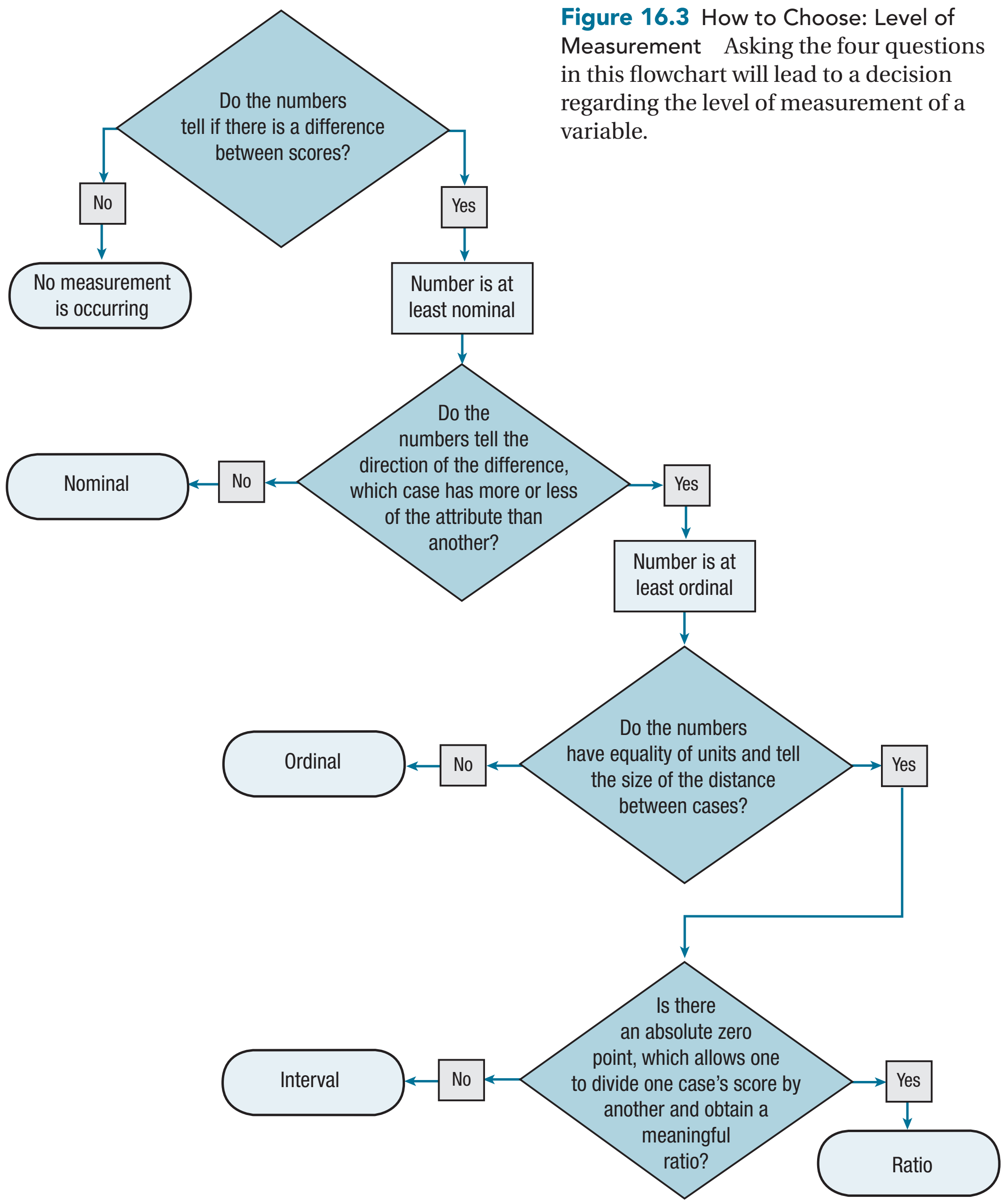


Figure 16.3 How to Choose: Level of Measurement Asking the four questions in this flowchart will lead to a decision regarding the level of measurement of a variable.

Describing Individual Cases

Sometimes the objective is to describe a single case. If the characteristic being measured for an individual case is measured at the interval or ratio level, it can be left as a raw score or transformed. Transformed scores have an advantage because they put scores into context. This book covered two types of transformed scores in Chapter 4:

- *z* scores reveal (1) whether a score falls above or below the mean, and (2) how far away from the mean it falls.
- Percentile ranks tell the percentage of scores in the distribution that fall at or below a score.

For example, a person with an IQ score of 115 could have his or her score transformed into a *z* score of 1.00 or a percentile rank of 84.13, both of which provide more information than the raw score.

Describing a Sample of Cases

When a set of scores is a sample from a population, it can be summarized (1) by presenting the entire distribution or (2) with a single value that describes some aspect of the sample. As was covered in Chapter 2, the entire distribution can be summarized in a table (i.e., a frequency distribution or stem-and-leaf plot) or as a graph. Graphs have a visual advantage over tables. Which graph can be used (histogram, frequency polygon, or bar graph) depends on whether the variable is continuous or discrete. Figure 16.4 is a flowchart for determining if a variable is continuous or discrete.

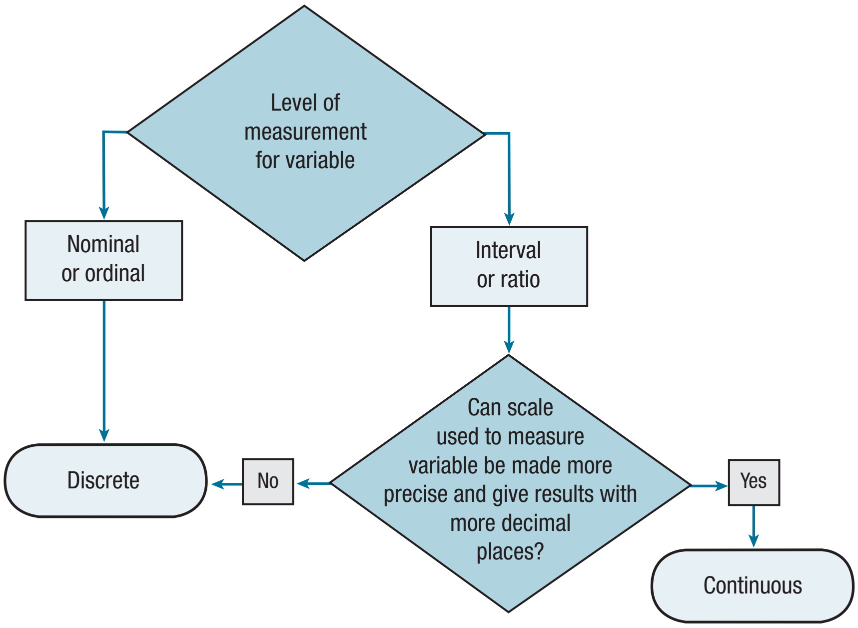


Figure 16.4 How to Choose: Continuous Numbers vs. Discrete Numbers To determine if the variable is continuous or discrete, decide whether the variable is (1) nominal or ordinal, or (2) interval or ratio. Then follow the flowchart.

When using a descriptive statistic to describe a sample of cases, a researcher can describe the central tendency or the variability in a set of cases (see Chapter 3). Usually, both are described. For example, a researcher might report both the average age of the participants *and* the range of ages.

A measure of central tendency for a variable is chosen depending on its level of measurement (see Table 16.3). But other factors—such as the shape of the distribution (see Chapter 3)—also need to be taken into account when selecting a measure of central tendency.

- To describe central tendency in a sample for a variable measured at the interval or ratio level, the go-to option is the mean (M).
- Central tendency for an ordinal-level variable is best measured with the median (Mdn).
- Central tendency for a nominal variable can only be measured with a mode.

TABLE 16.3 How to Choose: Which Measure of Central Tendency for Which Level of Measurement

	Measure of Central Tendency		
Level of Measurement	Mode	Median	Mean
Nominal	✓		
Ordinal	✓	✓	
Interval or ratio	✓	✓	✓

Not all measures of central tendency can be used with all levels of measurement. When more than one measure of central tendency may be used, choose the one that utilizes more of the information available in the numbers. If considering a mean, be sure to check the shape of the data set for skewness and modality.

All of the descriptive statistics for variability (see Chapter 3) are used for interval-level or ratio-level variables. These statistics include the range, interquartile range (IQR), variance (s^2), and standard deviation (s). The standard deviation is the most commonly used measure of variability. Different abbreviations are used for variance and standard deviation depending on whether one is referring to a population value or sample value. A lowercase Greek sigma, σ, is used for population standard deviations and s for sample standard deviations.

Just like seeing a mountain from several sides gives a better idea of what it looks like, using multiple perspectives—multiple descriptive statistics—gives a better sense of a set of data. Here are multiple ways a teacher might report how well her class had done on a test. Each descriptive statistic by itself answers a specific question. Taken together, they provide a detailed and comprehensive view of class performance:

- The teacher might use a frequency distribution, grouping scores into A's, B's, and so on to provide an overview of the class performance.
- A single score like the mean would summarize the class' performance, giving some sense of how easy (or difficult) the test was and how well (or how poorly) the class as a whole had done.

- The standard deviation would indicate how much variability existed in the scores, how tightly they are packed around the mean.
- The range would tell how badly the worst student had performed and how well the best student had performed. It would also reveal how much distance separated the best and the worst students.
- If a student came in to talk about his or her performance, the teacher might transform that student's score into a percentile rank in order to compare that student's performance to the rest of the class.

Sometimes the results from a sample are used to describe a population. The population standard deviation, σ, is an example of this. σ is a point estimate, a single value. Chapter 5 covered confidence intervals, which give a range within which it is likely that the population value falls. If one has a sample measured on an interval- or ratio-level variable and wants to estimate what the average value is in the larger population, a confidence interval for the mean is the way to go.

Practice Problems 16.1

16.01 Read each scenario and decide whether the study is descriptive, experimental, quasi-experimental, or correlational.

a. An ethologist attaches motion sensors to house cats and to wild cats so that he can measure the amount of time that each type of cat spends resting/sleeping each day. He's curious if wild cats spend more, less, or the same amount of time sleeping/resting as do house cats.

b. A physical fitness instructor is curious about how much time people spend at the gym. At his gym, people have to swipe in and swipe out. He collects data for a week, finding that the amount of time ranges from less than a minute to more than 9 hours, with a median time of 77 minutes.

c. An industrial/organizational psychologist obtained a random sample of employees at a large company. Among the questions she asked them was how good they thought the company was. She found that as the ages of the workers increased, so did their ratings of the "goodness" of the company.

16.02 Read the scenarios below and for each determine the level of measurement for the variable.

a. People who were not Chinese and who could not read Chinese viewed samples of Chinese writing written by Chinese people of different ages. They were asked to judge whether the passage was written by a pre-schooler (1), an elementary school–age child (2), a middle schooler (3), a high schooler (4), a young adult (5), a middle-aged person (6), or a senior citizen (7). The numbers in parentheses are the scores assigned to each answer.

b. Researchers wanted to find out if people could judge what sex a person was from a distance. They had participants watch videos of people walking toward them from 200 yards away and classify the subjects as –1 (male), 0 (can't tell), or 1 (female).

c. Some personality psychologists wanted to measure "manliness." They showed women pictures of men and had them rate the images as –1 (negative levels of manliness), 0 (neutral level of manliness), 1 (a small degree of manliness), 2 (more than a little, but less than average level of manliness), 3 (average level of manliness), 4 (somewhat above average in manliness), 5 (far above average in manliness), and 6 (off the top of the chart in terms of manliness).

16.03 Read each scenario and then select the appropriate descriptive statistic.

a. John attends a large university and is one of 450 people in introductory psychology. The last exam posed 73 questions, and John answered 62 correctly. John goes to see the professor because he would like to know how his performance compares to others in the same class.
b. The physical fitness instructor in Practice Problem 16.01(b) above wants to display the whole range of amounts of time people spend in the gym, as well as how "popular" the different amounts of time are.
c. The researchers in Practice Problem 16.02(b) above want to report what the average perceived sex is.
d. A hospital clothing manufacturer wants to re-size its one-size-fits-all hospital gowns now that Americans have grown larger. The manufacturer measures the weights of a sample of hospital patients from all 50 states and wishes to know the average weight of all Americans.

16.3 Hypothesis Tests I: Difference Tests

Let's move from selecting the correct descriptive statistic to selecting the right difference test. Difference tests are the most popular type of hypothesis test used in psychology. *t* tests and ANOVAs are examples of difference tests. **Difference tests** are used when cases are sorted into groups that are defined on the basis of an explanatory variable. They are used to answer the question of whether groups differ on some outcome variable. For example, whether people with depression or people with anxiety are more impaired by their disorder would be answered by a difference test.

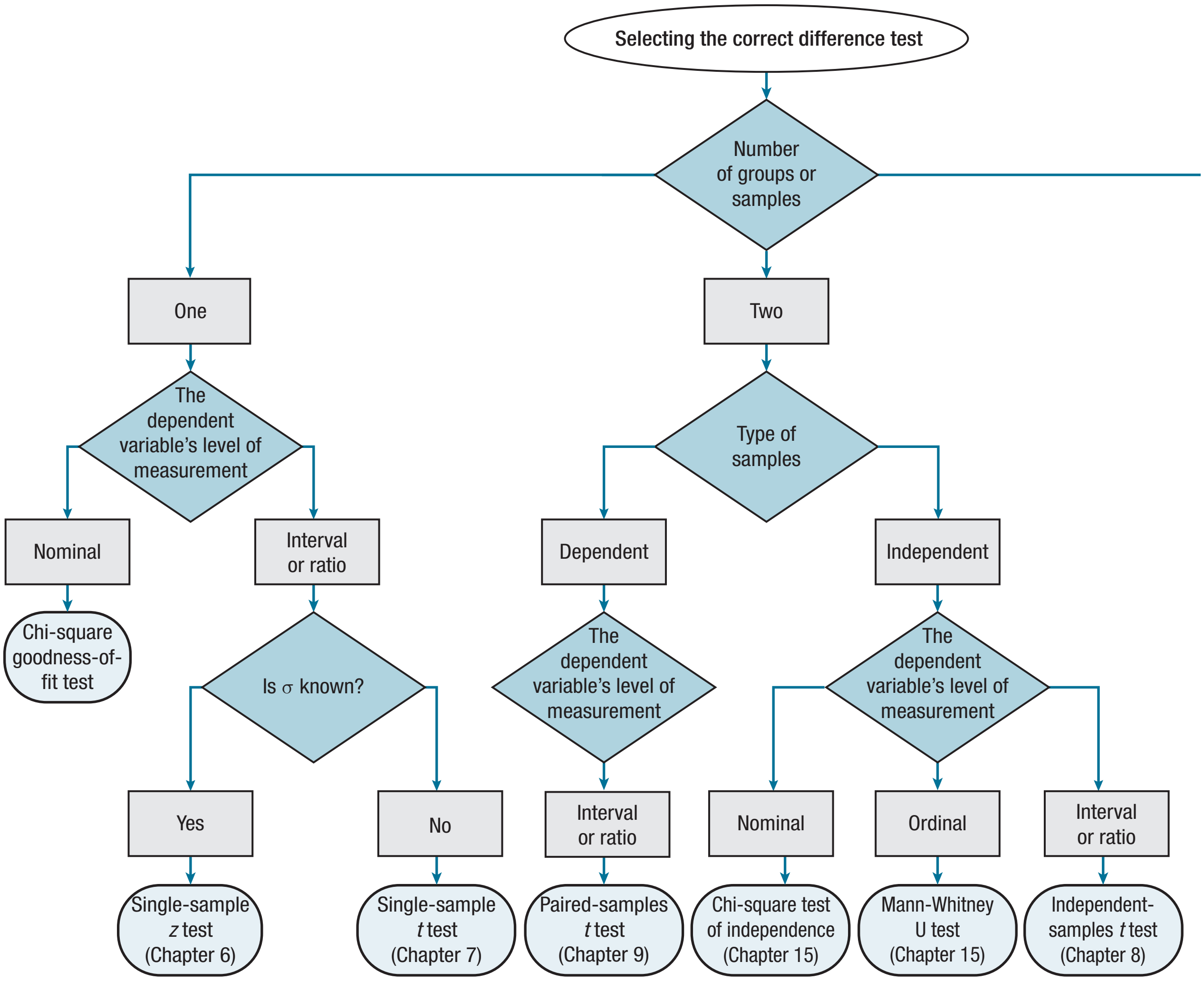
Difference tests are the go-to tests for experimental and quasi-experimental designs. In experimental designs, the researcher controls the assignment of participants to groups on the basis of the independent variable. With quasi-experimental designs, participants are classified as being in groups on the basis of their naturally occurring status on the grouping variable. In either event, a difference test is used to compare the groups in terms of the dependent variable.

The flowchart in Figure 16.5 shows how to choose the correct difference test to answer a research question. Picking the right test depends on (1) the number of groups of cases, (2) the number of explanatory variables, (3) the type of samples—independent or dependent, and (4) the level of measurement for the dependent variable.

Whether a study is experimental or quasi-experimental, the first decision point involves how many groups of cases there are: one, two, or three or more groups. If just one group exists, then the sample is being compared to a specified value or a population value. The specific test to be used in making that comparison is determined by the dependent variable's level of measurement.

If there is more than one group, it is important to determine whether the comparison is between just two groups or among three or more groups. This determination leads to the next decision point. When the comparison is between two groups, the next decision point is whether the groups are independent samples or dependent samples (see Chapters 8 and 9 on *t* tests).

Table 16.4 presents guidelines for determining the type of samples—independent or dependent. Once the type of samples has been determined for a two-sample test, the specific difference test to use depends on the level of measurement for the dependent variable.


Selecting the correct difference test
Number of groups or samples
One
Two
The dependent variable's level of measurement
Nominal
Interval or ratio
Chi-square goodness-of-fit test
Is σ known?
Yes
No
Single-sample z test (Chapter 6)
Single-sample t test (Chapter 7)
Type of samples
Dependent
Independent
The dependent variable's level of measurement
Interval or ratio
Paired-samples t test (Chapter 9)
The dependent variable's level of measurement
Nominal
Ordinal
Interval or ratio
Chi-square test of independence (Chapter 15)
Mann-Whitney U test (Chapter 15)
Independent-samples t test (Chapter 8)

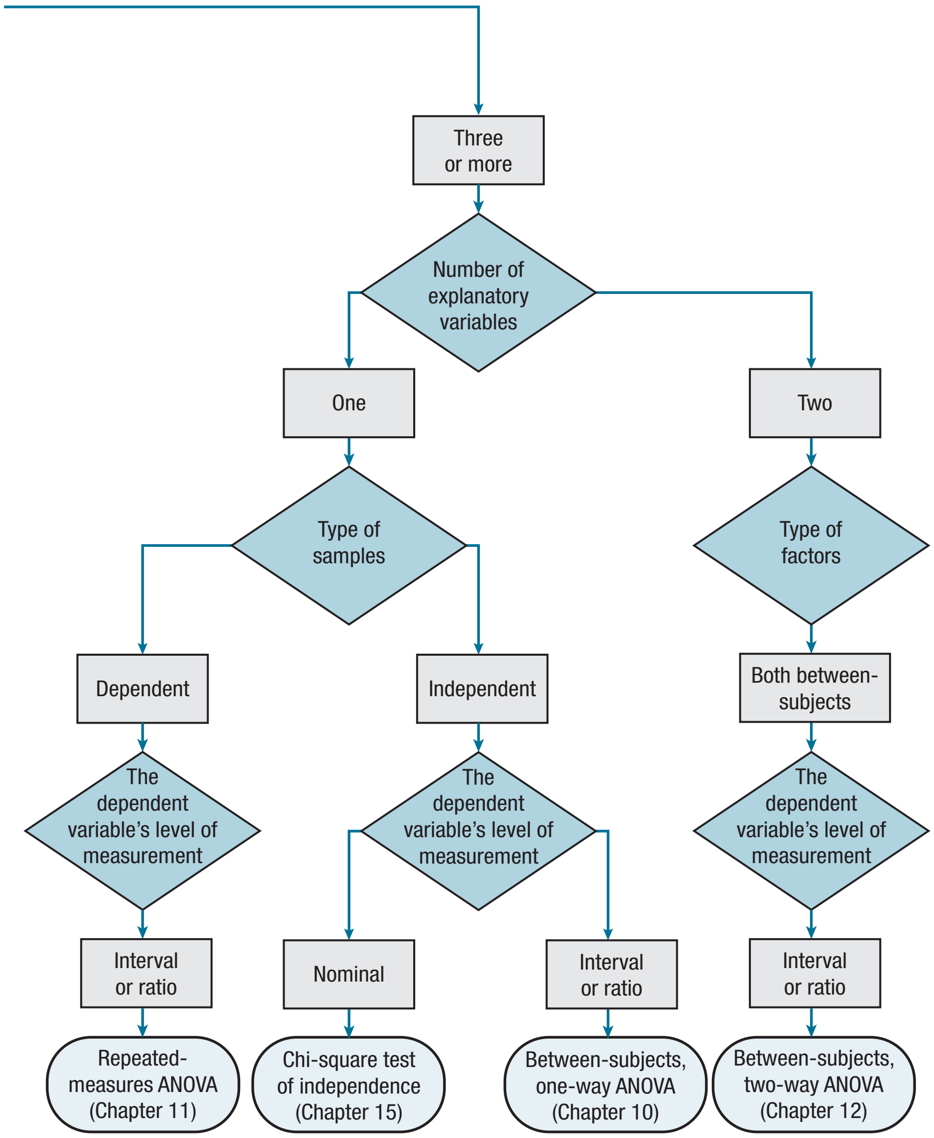


Figure 16.5 How to Choose: Selecting the Correct Difference Test Choosing a difference test depends on the number of samples, the type of samples, the dependent variable's level of measurement, and the number of explanatory variables.

TABLE 16.4 How to Choose: Guidelines for Determining If Samples Are Independent Samples or Dependent Samples

- Is each sample a random sample from its respective population?
  - If yes, independent samples.
  - If no, then there is no information about the type of samples.
- Is the size of each sample different? (That is, does $n_1 \neq n_2$?)
  - If yes, independent samples.
  - If no, then there is no information about the type of samples.
- Do the samples consist of the same cases measured at more than one point in time or in more than one condition?
  - If yes, dependent (paired) samples.
  - If no, then there is no information about the type of samples.
- Does the selection of cases for one sample determine the selection of cases for the other sample? Are cases paired together in some way?
  - If yes, dependent (paired) samples.
  - If no, independent samples.

When there are three or more samples, the flowchart branches out depending on (1) the number of explanatory variables, (2) whether the samples are independent or dependent, and (3) the dependent variable's level of measurement.

To see how the flowchart works, imagine a clinical psychologist who wanted to see how the loss of a parent before age 6 affected a child's adult level of psychological functioning. She obtained a large and representative sample of adults and classified them as belonging to one of four groups: (1) neither parent had died before the child, who is now an adult, was 6 years old; (2) the mother died before the child was 6 years old; (3) the father died before the child was 6 years old; or (4) both parents died before the child was 6 years old. To measure adult psychological functioning, she used an interval-level measure of psychological resilience. What statistical test should she use to find out if the mean level of psychological resilience differs among these four groups? The difference test flowchart, Figure 16.5, will help determine the answer.

- The first decision point involves how many groups of cases exist. There was one sample of cases, which was divided into four groups, so proceed down the flowchart path on the far right side for three or more groups.
- The next decision point involves how many explanatory variables exist. In this situation, there is one grouping variable, early parental death, that has four levels (none died, mother died, father died, both died).
- The next choice point involves the type of samples, dependent or independent. In this example, each case is classified as belonging to only one group and cases are not paired with each other. Who is classified as being in one group has no impact on who is classified as being in one of the other groups. The samples are independent.
- The final question to be addressed involves the level of measurement of the dependent variable. The dependent variable, psychological resilience, is measured at the interval level, which leads to the selection of a between-subjects, one-way ANOVA as the appropriate statistical test.

A Common Question

Q What should be done if there is no test listed in the flowchart? For example, what happens if the dependent variable for the early parental loss study was ordinal, not interval?

A The flowcharts here only choose among the statistical tests taught in this book. To find the correct test for a different situation, see the SPSS section at the end of this chapter, or consult a more advanced text.

Worked Example 16.1 For practice choosing the correct difference test, imagine an economist who wanted to investigate whether parents invest equally in their children's education depending on the children's birth order. From around the United States, he obtained a sample of families with two (and only two) children in which both children had attended college. His dependent variable was how much money the parents contributed to each child's college education. Is there a difference in mean parental funding for first-born vs. second-born children?

- There are two samples, a sample of first-born children and a sample of second-born children.
- The children in the two samples are paired together—they come from the same family—so the samples are dependent samples.
- The dependent variable, dollars spent, is a ratio-level variable.

Following the difference test flowchart, Figure 16.5, leads to selecting the paired-samples *t* test as the correct test to analyze the results.

Practice Problems 16.2

Select the correct statistical test.

16.04 A public health researcher wanted to know if hand washing or hand sanitizer was more effective in removing germs. She randomly assigned visitors at a hospital to wash their hands with soap and water or to use hand sanitizer. She then cultured the visitors' hands to see how many bacterial colonies there were.

16.05 People who wanted to lose weight were randomly assigned to drink 1 cup, 2 cups, or 3 cups of green tea each day. After three months, each person was weighed to determine how many pounds he or she had lost.

16.06 People from countries with spicy cuisines were matched in terms of age and sex with people from countries with bland cuisines. All participants completed the Willingness to Engage in Risky Behavior Scale. Are people who eat spicy food more likely to engage in risky behavior?

16.4 Hypothesis Tests II: Relationship Tests

Having covered the selection of descriptive statistics and difference tests, it is time to turn to relationship tests. **Relationship tests** are meant to examine whether a correlation exists between two variables. Relationship tests are used when there is one sample of cases and each case is measured on two variables in order to see if the variables vary together systematically. For example, one might measure both attractiveness and number of dates in college students to see if the two are associated. The flowchart for selecting the correct relationship test is shown in Figure 16.6.

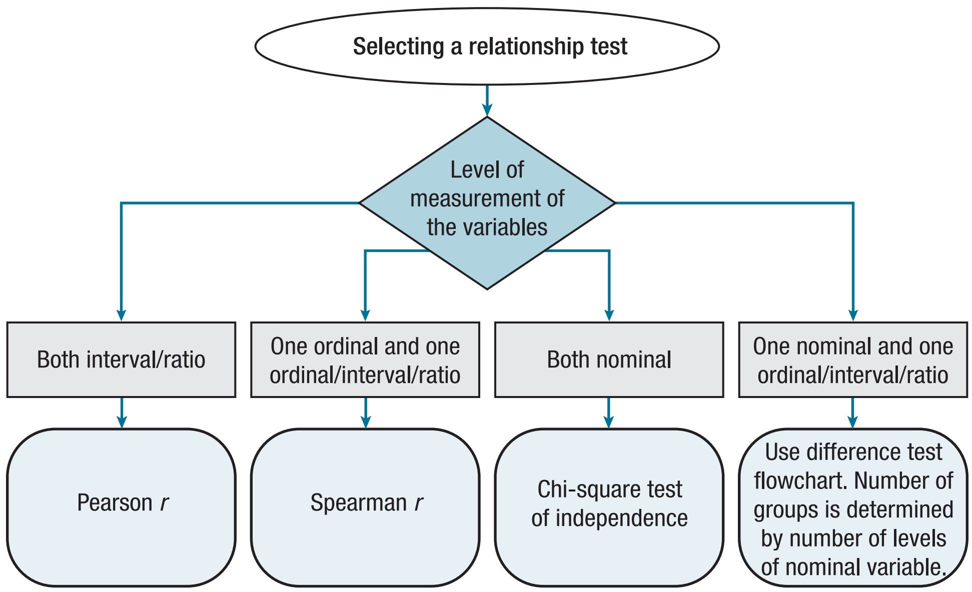


Figure 16.6 How to Choose: Selecting the Correct Relationship Test The selection of the correct relationship test depends on the level of measurement of the variables. If one variable is nominal and the other is not, then a difference test is used.

The decision about which relationship test to use in a given situation is determined by the level of measurement for each of the two variables:

- If both variables are measured at the interval or ratio level, then the Pearson *r* is used (see Chapter 13).
- If one variable is ordinal and the other is ordinal or interval or ratio, use a Spearman *r* (see Chapter 15).
- If both variables are nominal-level, then the chi-square test of independence is used (see Chapter 15).
- When one of the variables is nominal and the other is not, the relationship question can be conceptualized as a difference question. To do so, treat the nominal variable as the explanatory variable and use it to classify cases into groups. Then, the difference test flowchart (Figure 16.5) can be used to determine the appropriate test.

For an example of how to select the correct relationship test, imagine an animal behaviorist who first trained rats to run a maze to find food in a goal box and then measured how many days it took for the behavior to cease (for the rats to stop running to the goal box) once food was no longer being placed in the goal box. She wondered if smarter rats, those who learned the maze more quickly, were also quicker in learning a reward was no longer present. Is there a relationship between the times it takes to learn these two tasks?

This calls for a relationship test. To select the appropriate test, use the flowchart in Figure 16.6. To use that flowchart, the level of measurement must be known for each of the two variables—number of trials to learn the maze and number of days to extinction. Using the flowchart in Figure 16.2, both variables are classified as being at the ratio level. Turning back to the relationship test flowchart, Figure 16.6, when both variables are measured at the interval or ratio level, the Pearson *r* is the test to use.

Let's complete one more relationship test example, one in which one variable is nominal and the other isn't. Imagine a researcher who has a random sample of male college students and a random sample of female college students, and who wonders if there is a relationship between sex (male or female) and GPA. To determine what statistical test should be used, go to the relationship test flowchart, Figure 16.6. Notice that one variable is nominal (sex) and the other variable is interval (GPA). This combination of variables prompts us to exit the relationship test flowchart and move to the difference test flowchart. The relationship question is now phrased, "Is there a difference between male and female students in terms of mean GPA?"

In the difference test flowchart, the nominal-level variable, sex, is treated as the explanatory variable, and the interval/ratio variable, GPA, as the dependent variable. Here are the answers to the choice points in Figure 16.5 for the sex/GPA test decision:

- The explanatory variable has two levels, male and female, so two groups are being compared.
- Each sample was a random sample from its population, so the two samples are independent.
- The dependent variable, GPA, is measured at the interval level.

These choice points lead to the selection of an independent-samples *t* test as the appropriate test to analyze the data to determine if a relationship exists between sex and mean GPA. A difference test can answer a relationship question.

Worked Example 16.2

A criminologist developed a theory that elementary school teachers can quickly tell which kids are going to be a problem in their classroom. After only two weeks of school, he had asked teachers to make a judgment for each child: whether he or she "would be trouble in the classroom." A dozen years later, as the same kids are ready to graduate from high school, he tracked down all their records and determined which ones had been arrested, spent time in jail, dropped out of school, and so on. Anyone for whom one of these events had occurred was classified as "having gotten into trouble in life." What test should the criminologist use to see if an association exists between a prediction made by a first-grade teacher, after only two weeks of observation, and getting into serious trouble over the next 12 years?

This means asking if there is a relationship between the two variables, so the place to start is with the flowchart in Figure 16.6. The first question concerns the level of measurement of the variables. Both variables are nominal. Thus, the appropriate statistical test is the chi-square test of independence.

Practice Problems 16.3

Select the correct statistical test.

16.07 A hair salon owner watches people walk by his shop, notes whether they are male or female, and classifies them as having long hair or short hair. Does a relationship exist between sex and hair length?

16.08 Is there a relationship, for adults, between how many text messages they send per week and how many minutes they talk on their phones?

16.09 An education researcher wants to see if class rank in high school is associated with class rank in college.

Application Demonstration

To see how selecting the correct statistical test works in real life, here are three classic studies in psychology: one that uses descriptive statistics, one that uses difference tests, and one that uses relationship tests.

Obedience to Authority

Perhaps the most famous study in psychology was reported by Stanley Milgram in 1963. Called a "behavioral study in obedience," Milgram investigated whether normal people would behave in inhumane ways just because they were following orders.

In the study, 40 males, from age 20 to 50, served as participants and were called "teachers." The teachers were asked to test another participant, called the "learner," to see how well he had memorized a list of words. For every word the learner got wrong, the teacher was asked to punish him with a shock. And, after delivering a shock, the teacher was to adjust the shock generator to deliver a higher level of shock as the next punishment. The shock generator had 30 levels, ranging from 15 to 450 volts, and the levels were labeled from "Slight Shock" to "Danger: Severe Shock." At 300 volts, the learner stopped responding—presumably, he was now unconscious or dead from the shocks—and the teacher was instructed to treat no response as a wrong answer, to administer a shock, and to advance to the next question. (By the way, the learner was a confederate of the experimenter, offered incorrect answers according to a script, and received no shocks.)

This is all that there was to Milgram's experiment. It included no experimental group vs. control group aspect. All he included was an experimental group. Milgram's question was simple: How many participants would continue to give shocks all the way up to 450 volts? And his answer—a descriptive statistic that 26 of 40 normal men, 65%, were willing to shock a man to death in a psychology experiment simply because a researcher in a lab coat asked them to—shocked a nation. Sometimes the simplest statistic is the most powerful.

Television and Aggression

Leonard Eron was one of the first to study the relationship between children watching violence on TV and behaving aggressively in real life. In 1972, along with some colleagues, he presented data on children followed for 10 years, from 9 to 19 years old.

When the children were 9 years old, the researchers measured the number of violent shows the children liked to watch. At the same time, Eron asked the children to answer questions about their peers. From this, he was able to develop a measure for each child as to how aggressive he or she was, as rated by peers. Similar measures were obtained when the kids were 19 years old.

Eron used Pearson correlation coefficients to analyze the results. Here are the highlights of his findings:

- There was a positive, statistically significant correlation of .21 between violent TV watching at age 9 and rating of aggressiveness by peers at age 9.
- The correlation between watching violent TV at age 9 and rating of aggressiveness by peers 10 years later was also positive and statistically significant. But, the relationship was a stronger one: $r = .31$.
- The correlation between violent TV watching at age 19 and aggression at age 19 was not statistically different from zero.

Taken together, these correlations suggest that early viewing of violent TV is more strongly related to aggression 10 years later than it is to aggression as a child. These relationship tests suggest that there is a critical period in childhood during which images from television may have an impact on personality. Eron's research has led to thousands of other studies about the effects of violence on television. Simple correlations can make powerful points.

Language and Memory

Elizabeth Loftus, a cognitive psychologist, is one of the people most responsible for eyewitness testimony no longer being held in high esteem. One of her early studies, published with John Palmer in 1974, used difference tests to make this point.

In that study, participants were shown movies of car crashes and asked to estimate how fast the cars were traveling when the crash occurred. Different participants were asked slightly different questions, ranging from how fast the cars were going when they *contacted* each other to how fast they were going when they *smashed into* each other. There were five different options: smashed, collided, bumped, hit, and contacted. Loftus's research question was whether the different words would elicit different estimates of speed.

And, that was what she found. Asked how fast cars were traveling when they "contacted" each other, the mean response was 32 mph. Each word that suggested more speed was associated with an increase of perceived speed by about 2 mph, until cars that "smashed into" each other were going almost 41 mph. The analysis of variance that Loftus used to analyze these results showed that the results were statistically significant. Here's a *difference* test that made a difference—it showed that perceptions can be manipulated by a person asking about them.

DIY

Many people have pet theories about the world. A friend of mine in graduate school, many years ago, believed that no matter what she bought at the grocery store, the cost always turned out to be about $15 per bag. I believe that the day after a presidential election, many more supporters of the loser have removed bumper stickers from their cars than have supporters of the winner.

Do you have a pet theory? If not, develop one. What statistical test would you use to answer it?

SUMMARY

Define the different tasks of statistics.

- Statistical tests are tools, appropriate for specific tasks. Tasks include summarizing a set of data with descriptive statistics, using difference tests to compare groups in experimental and quasi-experimental designs, and using relationship tests to see if two variables vary together systematically (relationship tests).

Select the correct statistical test.

- Descriptive statistics summarize a set of numbers to describe an individual case or a sample of cases in terms of central tendency and/or variability. Difference tests are used when cases are sorted into groups on the basis of an explanatory variable. Relationship tests are used when there is one sample of cases and each case is measured on two variables to see if the variables covary.

KEY TERMS

descriptive statistics – statistics used to describe a set of observations.

difference tests – statistical tests that look for differences among groups of cases.

relationship tests – statistical tests that determine if two variables in a group of cases covary.

CHAPTER EXERCISES

Review Your Knowledge

16.01 Descriptive statistics ____ a set of observations.

16.02 To determine if one group differs from another, the type of test to use is a ____ test.

16.03 Relationship tests are used to see if ____ vary together systematically.

16.04 If cases are divided into groups, the researcher is probably conducting a ____ test.

16.05 The ____ is thought of as the causal agent.

16.06 The two characteristics of a data set most commonly described are ____ and ____.

16.07 Scores are transformed in order to put them into ____.

16.08 In order to choose the correct ____ to display a variable's frequency, it matters if the variable is discrete or continuous.

16.09 The ____ of the distribution of the data, as well as the variable's ____, determines what measure of central tendency should be chosen.

16.10 This book covered no measures of variability for ____ or ____ variables.

16.11 Level of measurement *is / is not* important in deciding what difference test to use.

16.12 Two variables are measured on each case in one sample of cases. This sounds like a ____ test is being planned.

Apply Your Knowledge

Determining the type of study

For Exercises 16.13–16.20, determine if the study is descriptive, experimental, quasi-experimental, or correlational.

16.13 A restaurant chain surveys its customers to learn demographic characteristics such as age in years, marital status, and annual income.

16.14 A developmental psychologist has a teacher rate the boys in his class in terms of how helpful they are. He then classifies the boys as to whether they were ever Boy Scouts or not, and compares the helpfulness of the two groups.

16.15 A clinical psychologist obtained a large sample of adults. Each adult kept track of all the food and beverages he or she consumed for a week. From this, the psychologist calculated the average grams of caffeine consumed per day. The psychologist believed that caffeine, a stimulant, might be used to self-medicate for symptoms of ADHD. So, she also had all participants complete an ADHD symptom inventory. She wanted to see if people with more ADHD symptoms consumed more caffeine.

16.16 An economist wondered if job satisfaction was associated with salary. She obtained a sample of first-year teachers from all around the country and found each one's salary, adjusted for the local cost of living. She also had the teachers complete a job satisfaction survey.

16.17 A cognitive psychologist wondered if the type of computer one used affected how one thought. She put together a sample of kids, randomly divided them into two groups, and gave Apple Macs to one group and Windows PCs to the other group. A year later, she administered the Smith & Jones Measure of Creative and Divergent Thinking to each child.

16.18 A real estate company wondered why some of its agents were more successful than others. The company classified its agents as high-volume producers, medium-volume producers, and low-volume producers. All agents took a personality test and the company compared the three groups on the personality traits measured.

16.19 A milk-processing plant had strict standards regarding how much milk went into each container. A quart container, for example, was supposed to contain no less than 31.5 ounces and no more than 32.5 ounces. Every day, quality control randomly sampled 100 quart containers, precisely measured their contents, and prepared a graph showing the distribution of amount per container.

16.20 A researcher on superstition decided to investigate if walking under a ladder really brought about bad luck. He took 100 people to his laboratory, one at a time. As each person entered, he flipped a coin. If it turned up heads, he had the person walk under a ladder; tails, the person walked next to the ladder. A week later, all the participants returned and reported if anything unlucky had happened to them during the week.

Determining a variable's level of measurement

For Exercises 16.21–16.26, determine (a) whether the variable is measured at the nominal, ordinal, interval, or ratio level, and (b) whether it is continuous or discrete.

16.21 Ounces of milk, as measured in Exercise 16.19

16.22 Whether anything unlucky or not happened, as measured in Exercise16.20

16.23 Marital status

16.24 Level of anxiety before a test, as measured 0–4 on a scale marked 0 = none, 1 = slight, 2 = moderate, 3 = considerable, and 4 = extreme

16.25 Heights of randomly selected pieces of land as measured by meters above or below sea level

16.26 Number of light bulbs in a house

Selecting the right descriptive statistic

For Exercises 16.27–16.30, select the appropriate statistical technique from among the options in Figure 16.2.

16.27 A measure of variability in IQ scores for students in law school

16.28 The most common marital status among 50-year-old men living in Rhode Island

16.29 A graph showing the number of high-, medium-, and low-volume producers from the real estate company in Exercise 16.18

16.30 The average salary for a sample of first-year teachers in the United States

Selecting the correct difference test

For Exercises 16.31–16.38, select the appropriate statistical technique from among those in Figure 16.5.

16.31 10% of the U.S. population is left-handed. Several recent presidents (Barack Obama, Bill Clinton, George H. W. Bush) have been left-handed. This caused a cognitive psychologist to wonder if presidents were representative of the U.S. population in terms of handedness. What test should she use?

16.32 A demographer was curious if an age difference existed between men and women who were getting married for the first time. From wedding announcements in local papers throughout the United States, he randomly sampled 250 couples for whom it was the first marriage for both the bride and the groom. From the announcements, he obtained their ages. What statistical test should he use to compare the ages of husbands and wives?

16.33 In some educational programs, there are a lot of options. One nursing program, for example, allows its students to decide between (a) traditional, semester-long courses; (b) online courses; (c) weekend-only courses; or (d) courses that meet for 8 hours a day, 6 days in a row. The dean of the school wanted to find out if the type of course content had any impact on how much was learned. At the end of its programs, she administered a 50-question, multiple-choice test about basic nursing facts. What test should she use to see if the groups differ on the mean percentage of questions answered correctly?

16.34 Students often gain a few pounds when they move away to college and start eating institutional food. A nutritionist weighed first-year students at the start of the semester, right before Thanksgiving, on their return to school in January, and at the end of the second semester. What statistical test should he use to see if their mean weight changes over time?

16.35 An ecologist obtained a random sample of 85 small rural towns and a random sample of 72 small suburban towns. For each town, he calculated pounds of trash per resident per year that went to the landfill. What statistical test should he use to see if rural and suburban residents differ in the average amount of trash they generate every year?

16.36 A dentist wanted to examine if milk consumption as a child had an impact on the need for dentures by age 50. He classified his patients as having consumed a lot of milk as a child (yes vs. no) and as to whether they wore dentures by age 50 (yes vs. no). What statistical test should he use to determine this?

16.37 The same dentist wanted to find out if his practice was unusual in terms of how many cavities his patients had. He learned that the average American has 11.23 decayed, filled, or missing permanent teeth. In his dental practice, the mean was 8.76, with a standard deviation of 3.42. What statistical test should he use?

16.38 A textbook publisher calculated the percentage of deadlines met by each of its authors. It then classified the authors as having made money for the company, broken even, or lost money for the company. What statistical test should the company use to see if the three types of authors differ in the degree to which they met their deadlines?

Selecting the right relationship test

For Exercises 16.39–16.42, select the appropriate statistical technique from among those in Figure 16.6.

16.39 The dentist from Exercises 16.36 and 16.37 has a new question to contemplate. He wondered if people get most of their cavities when they are young or if they occur consistently over time. To investigate this, he decided to examine the relationship between the ages of his patients and how many of their permanent teeth were decayed, filled, or missing. What test should he use?

16.40 An anti-doping agency decided to investigate if an athlete's eliteness was related to the likelihood of doping. It unexpectedly drug-tested the top 10 finishers in all the track and field events at a large meet. From this, it calculated the percentage of first-place finishers who tested positive, the percentage of second-place finishers who tested positive, and so on. The agency then wanted to examine the relationship between the place in which an athlete finished and the likelihood of testing positive. What test should the agency use?

16.41 A member of a bicycle club wondered if a relationship existed between how expensive a person's bicycle was and how fast that person rode. He found riders with bikes in two categories. One group rode bikes that were classified by a cycling magazine as "decent bikes for not too much money." The others rode bikes classified as "amazing, but even Bill Gates would think twice about spending this much for a bike." The club member then asked each person to ride solo on a track and time his or her fastest lap. What test should the club member use to see if there's a relationship between the cost of the bike and the speed of the rider?

16.42 Some people like cats and some like dogs. Cats are generally thought of as more independent and aloof. A personality psychologist wondered if a relationship existed between the personality trait of introversion/extroversion and the type of pet a person preferred. He obtained a sample of cat owners and a sample of dog owners. Each pet owner took a personality test and was classified as an introvert or an extrovert. What statistical test should be used to see if the two variables are related?

Expand Your Knowledge

In this series of questions, there are no headings that tell which flowchart to use. Plus, some gaps occur in the flowchart. For example, if the scenario has two dependent samples with an ordinal-level dependent variable, there is no such test in the flowchart. Tests exist for such situations; they just weren't covered in this book. So, if a scenario in this series calls for a "missing" test, just write, "No such test in book" as your answer.

16.43 A neonatologist is curious to find out what the long-term effect of premature birth is on mental development. He goes to a large school district, assembles a sample of sixth graders, and determines each student's IQ score. From each student's parents, he finds out if the student was full-term or premature at birth. What statistical test should he complete to figure out if there is a relationship between prematurity and IQ?

16.44 A summer camp director wanted to find out if campers' degree of homesickness changed over time. She took a sample of them at her camp and at three points in time—the beginning, middle, and end of camp—administered the interval-level Homesickness Inventory. What statistical test should she choose to analyze her data?

16.45 A high school principal claimed that academic talent was evenly distributed among the social strata at his high school. To test this, he classified the students as jocks, cheerleaders, techies, stoners, band geeks, and so on and compared class rank among the different groups. What statistical test should he use?

16.46 A health-care researcher obtained a sample of respondents from throughout the United States and asked each one to indicate how worried he or she is about the avian flu on an interval-level scale. Before reporting the results for how worried Americans are about avian flu, the researcher wants to make sure the sample may be representative of the United States in terms of geographic distribution, based on the U.S. Census Bureau's division of the United States into four regions (Northeast, Midwest, South, and West). What statistical test should he use to answer this question?

16.47 A medical school dean wondered whether students who did better in her medical school became better physicians. To test this, she gathered a sample of medical students in their final year and determined their GPAs. She also administered to each student the Differential Diagnosis Test (DDT). The DDT presents students with 50 diagnostic dilemmas. Each dilemma is considered equivalent, and scores on the DDT can range from 0 (no questions answered correctly) to 50 (all answered correctly). What statistical test should the dean use to answer her question?

16.48 Graduates of a particular college were classified as having (a) liberal arts degrees or (b) professional degrees. They were also classified as (1) planning to pursue graduate education or (2) not going on for graduate education. When the graduates returned for their 25th reunion, they completed an interval-level, life satisfaction scale. Is life satisfaction influenced by type of degree and/or graduate study? What statistical test will answer this question?

16.49 A sample of students at a college were asked to classify themselves as (1) Asian; (2) Black; (3) Hispanic; (4) Native American; (5) White; or (6) other. Each person was also asked whether his or her family would consider it acceptable (yes or no) if he or she married someone from a different racial category. What statistical test should be used to see if the degree of acceptability differs among racial/ethnic categories?

16.50 The very last question in this book is a challenging one. It is rare that a researcher does a study that involves just one statistical analysis. Usually, there are multiple dependent variables and a branching series of questions and analyses. Here is an example. An infectious disease specialist believes she has developed a vaccine for the common cold. In order to test its efficacy, she designs a study. As a first step, potential participants call in and provide demographic information (age, sex, race) and complete an interval-level measure of general physical health. After this, they learn more about the study, and about 40% of those who originally called decide that they wish to take part in the study.

The actual study is a double-blind study in which the volunteers will be inoculated either with the active vaccine or a placebo. At the end of the first week, each participant completes a side effect checklist, which counts how many side effects each person has experienced.

For the next 12 months, the research team will keep track of each participant and whether he or she contracts a cold. If a participant contracts a cold, the research team will note how many days have elapsed from the vaccine until the cold. In addition, the researchers will assess the severity of the cold in terms of two dimensions: (1) measuring how many days the cold lasts, and (2) having the cold sufferers report how congested they felt on an ordinal scale.

For each scenario below, decide which statistical test or tests should be used to answer the research question. Be aware—many of these questions involve multiple dependent variables and multiple analyses.

a. The researcher wants to know if those who volunteered for the study differ from those who chose not to volunteer.

b. The researcher wants to know if her sample of volunteers is demographically similar to the U.S. population.

c. Did the active vaccine have more side effects than the placebo?

d. The primary measure of the vaccine's effectiveness is whether or not people in the experimental group, compared to the control group, did or did not contract a cold during the year after the vaccine. How should she answer this question?

e. Secondary measures of the vaccine's effectiveness were, for those who contracted colds, how long the cold lasted and how sick they were. What tests should the researcher do to answer these questions?

f. For the control group participants, the researcher was interested in the relationship between their general physical health and contracting a cold. How can this question be answered?

g. For the control group participants who contracted a cold, the researcher was interested in the relationship between general physical health and the severity of the cold. How should she answer this question?

SPSS

SPSS, like most serious software programs, has a variety of help options. One help option is Statistics Coach, which leads users through a series of questions to the selection of a statistical test. Let's see how it works, using as an example the study in which the effect of early parental loss is measured years later in adults. There were four independent samples (no loss, mother died, father died, both died), and the dependent variable was measured at the interval level.

Figure 16.7 shows the help options available within SPSS. Statistics Coach is on the fifth line down. Clicking on it opens up the screen shown in Figure 16.8, where SPSS asks, "What do you want to do?" There are seven possible answers to the question. Option 1 and option 2 involve descriptive statistics, option 4 applies to difference tests, and option 5 relationship tests. The other three options lead to more advanced statistics.

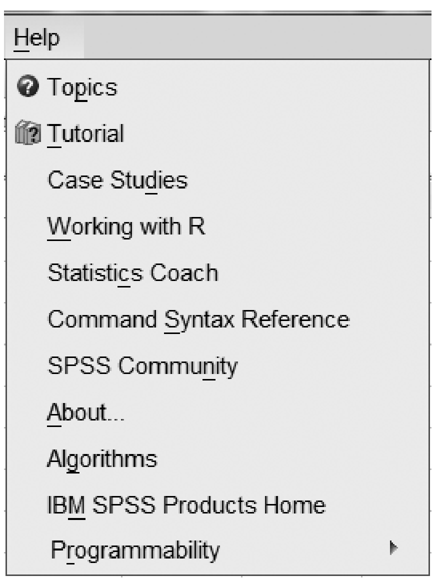


Figure 16.7 The SPSS Help Menu Statistics Coach leads an SPSS user through a series of questions to select an appropriate statistical test. (Source: SPSS)

Figure 16.8 Initiating the Statistics Coach The first step is to decide what one wants to do statistically. (Source: SPSS)

Clicking on the option for difference tests leads to the query, seen in Figure 16.9, about the kind of data. SPSS calls interval and ratio data "scale" data. Here, it is checking how the cases are divided into groups, on the basis of the level of measurement for the variable. Click on the option for "scale numeric data divided into groups."

Corty_3e_Ch16_617-642_Final HR.indd 639 21/01/16 7:41 pm

The next question, seen in Figure 16.10, concerns how many groups there are. This study involves three or more groups.

Statistics Coach > Statistics Coach

Compare groups for significant differences

What kind of data do you want to compare?

Data in categories (nominal, ordinal)

Scale numeric data divided into groups

Figure 16.9 Determining Level of Measurement In this step, SPSS asks one to choose whether one's data are nominal/ordinal or interval/ratio. (Source: SPSS)

Statistics Coach > Statistics Coach > Compare groups for significant differences

Scale numeric data divided into groups

How many groups or variables do you want to compare?

One group or variable compared to a known value

Two groups or variables

Three or more groups

Figure 16.10 Deciding How Many Groups One Has In this step, one decides whether to do a single-sample test, a two-sample test, or a multiple-sample test. (Source: SPSS)

Clicking on three or more groups brings up the next screen, Figure 16.11, which inquires how many "grouping" or "factor" variables exist. These are what the text calls independent variables, and there is only one of them.

Statistics Coach > Statistics Coach > Compare groups for significant differences > Scale numeric data divided into groups

Three or more groups

How many grouping (factor) variables do you have?

One (e.g, revenue for three groups defined by region)

Two or more (e.g., revenue for groups defined by division within each region)

Figure 16.11 Verifying the Number of Independent Variables In this step, the decision is whether to do a one-way test or a multiway test. (Source: SPSS)

Clicking on one grouping variable brings up the next-to-last screen, Figure 16.12, which asks whether the data are normally distributed (i.e., a parametric test), not normally distributed (i.e., a nonparametric test), or if one wants to check.

Statistics Coach > Statistics Coach > Compare groups for significant differences > Scale numeric data divided into groups > Three or more groups

One (e.g, revenue for three groups defined by region)

Which test do you want?

Test that assumes data are normally distributed within groups

Test that does not assume data are normally distributed

Check normality of data

Figure 16.12 Parametric or Nonparametric? In this step, one decides whether a parametric or nonparametric test should be done. (Source: SPSS)

Corty_3e_Ch16_617-642_Final HR.indd 640 21/01/16 7:41 pm

Selecting normally distributed leads to a decision about what test should be done, a one-way ANOVA. As can be seen in Figure 16.13, SPSS also provides information about where the commands are located.

Statistics Coach > Statistics Coach > Compare groups for significant differences > Scale numeric data divided into groups > Three or more groups > One (e.g. revenue for three groups defined by region)

Test that assumes data are normally distributed within groups

To Obtain a One-Way Analysis of Variance

This feature requires the Statistics Base option.

1. From the menus choose:

   Analyze > Compare Means > One-Way ANOVA...

2. Select one or more scale, numeric test variables for the Dependent List.
3. Select a numeric grouping variable (a variable that divides cases into three or more groups) for the Factor.

The factor (grouping) variable must be numeric. If your grouping variable is a string (alphanumeric) variable, use Automatic Recode on the Transform menu to convert the string values to integers.

Related information:
One-Way ANOVA

Figure 16.13 Selecting the Test Not only does SPSS select a test, but it also provides guidance about where to find the commands and what commands to select. (Source: SPSS)

Corty_3e_Ch16_617-642_Final HR.indd 642 21/01/16 7:41 pm