Third Edition

USING AND INTERPRETING STATISTICS

A Practical Text for the Behavioral, Social, and Health Sciences

ERIC W. CORTY

The Pearson Correlation Coefficient

13

LEARNING OBJECTIVES

- Differentiate difference tests from relationship tests.
- Define and describe a relationship.
- Compute a Pearson *r*.
- Interpret the results of a Pearson *r*.
- Take into account the effects of a confounding variable.

13.1 Introduction to the Pearson Correlation Coefficient

13.2 Calculating the Pearson Correlation Coefficient

13.3 Interpreting the Pearson Correlation Coefficient

13.4 Calculating a Partial Correlation

CHAPTER OVERVIEW

Previous chapters on hypothesis tests focused on two types of tests—*t* tests and ANOVAs. With these tests, cases are assigned to or classified in groups on the basis of an explanatory variable, and then the outcomes of the groups are compared. These tests are called "difference" tests because they determine whether there is a *difference* in the mean of the dependent variable between the groups. Classic experiments in which an experimental group is compared to a control group to see which has a better mean outcome are examples of difference tests.

In this chapter, we're going to learn about another type of test, what is called a "relationship" test. With relationship tests, there is one group of cases. Each case in the group is measured on two variables to determine if a *relationship*, or an association, exists between the two variables. For example, we could measure a group of college students to determine if there is a relationship between how extroverted they are and how often they date.

13.1 Introduction to the Pearson Correlation Coefficient

There are multiple statistical tests that measure relationships by calculating correlation coefficients. In subsequent chapters, we'll cover the Spearman rank-order correlation coefficient and the chi-square test of independence. But, the focus of this chapter is the head of the relationship test household, the most commonly used relationship test, the Pearson correlation coefficient. Usually, it is referred to as the Pearson *r*, or simply *r*.

Here is what we'll cover in the first part of the chapter as we introduce r:

- Defining Pearson r and "relationship"
- Exploring what a relationship says about cause and effect
- Seeing how to visualize a relationship
- Learning the difference between weak and strong relationships
- Seeing how z scores define the Pearson r
- Learning how the Pearson r quantifies relationship strength
- Exploring the two directions a relationship can take
- Learning about conditions that affect Pearson r

Defining Pearson *r* and Relationship

Measures of association are statistics that quantify the degree of relationship between two variables. If two variables are related, they are said to be *correlated.* This means that the variables vary together systematically, that a change in one variable is associated with a change in the other. For example, alcohol consumption and blood alcohol level vary together systematically—the more alcohol a person consumes, the higher the level of alcohol in his or her blood. Depending on sex and weight as shown in Table 13.1, the relationship between the number of drinks consumed and blood alcohol level is well established.

The Pearson r is a specific measure of association. A **Pearson correlation coefficient** quantifies the degree of linear relationship between two interval and/or ratio variables. Because the Pearson r uses interval and/or ratio variables, distances between points can be measured, z scores calculated, and graphs made. The graphs, called scatterplots, can be examined for the degree to which the relationship between the two variables takes the form of a straight line because the Pearson r measures the degree of linear relationship. (We'll discuss nonlinear relationships, and why Pearson r is not appropriate to use in those cases, later in this chapter.)

TABLE 13.1 The Relationship Between Number of Drinks Consumed and Estimated Blood Alcohol Percentage

	Body Weight in Pounds					
	Women			Men		
Drinks	120	140	160	140	160	180
0	.00	.00	.00	.00	.00	.00
1	.04	.03	.03	.03	.02	.02
2	.08	.07	.06	.05	.05	.04
3	.11	.10	.09	.08	.07	.06
4	.15	.13	.11	.11	.09	.08
5	.19	.16	.14	.13	.12	.11
6	.23	.19	.17	.16	.14	.13

Blood alcohol level varies systematically with the number of drinks consumed, based on a person's sex and weight.
Source: Pennsylvania Liquor Control Board.

A Common Question

Q If the Pearson r only measures association between two interval and/or ratio variables, what is used for ordinal and/or nominal variables?

A If one variable is ordinal and the other is an ordinal, interval, or ratio variable, then association can be measured with a test called the Spearman rank-order correlation coefficient, or Spearman r for short. Association between two nominal variables is measured with the chi-square test of independence. Both of these tests are covered in Chapter 15.

Correlation, Causation, and Association

If two variables correlate, the relationship *may* be a cause-and-effect relationship, but it doesn't have to be. Statisticians love to say, "Correlation is not causation." That's because correlation between two variables does not guarantee a cause-and-effect relationship between them. A correlation between two variables only guarantees there is an *association* between the two variables, not that one causes the other.

To understand this, look at **Figure 13.1**. This type of graph, called a scatterplot, displays the relationship between two variables. For a random sample of 24 states, this graph shows the association between the number of community hospitals in a state (on the X-axis) and the number of deaths per year in the state (on the Y-axis). Each point represents a state. For example, Delaware, the state at the bottom left of

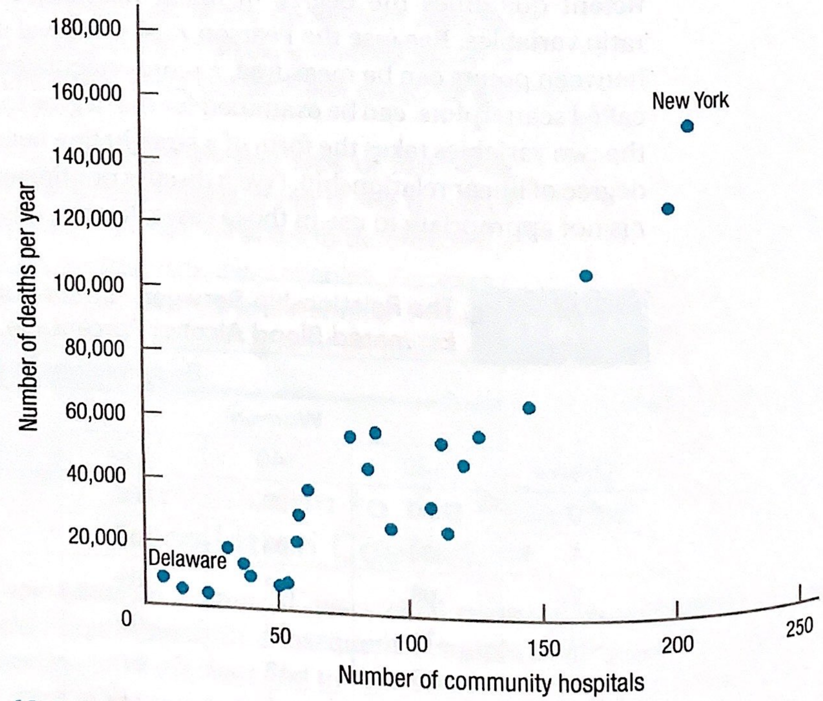


Figure 13.1 Relationship at a State Level Between Number of Community Hospitals and Number of Deaths This scatterplot illustrates the relationship between the number of community hospitals per state and the number of deaths per year in that state. States with more hospitals have more deaths. (Credit: Thanks to Jillian Mrozowski, who obtained these data.)

the graph, has 6 community hospitals and about 7,000 deaths per year; New York, the state at the top right, has 206 community hospitals and about 153,000 deaths.

The scatterplot shows clear evidence of an association between the two variables. The two variables vary together systematically: states like Delaware, with few hospitals have few deaths, and states like New York, with a lot of hospitals, experience a lot of deaths.

The relationship can be read in either direction. So far, this relationship has been viewed as states with more hospitals have more deaths. But, the scatterplot can be viewed just as legitimately as showing that states with more deaths have more hospitals.

If two variables are correlated, it simply means that they vary together systematically. A cause-and-effect relationship may exist between them, but there does not have to be. The number of deaths in a state and the number of community hospitals are correlated, but it doesn't seem plausible that there is a cause-and-effect relationship between them. If such a relationship did exist, then a state could reduce the number of people who die each year by closing down hospitals. That doesn't seem to be a course of action likely to work.

The number of hospitals per state and the number of deaths vary together systematically, but they don't cause each other. Their correlation most likely results because each is associated with a third variable, population. States with more people, like New York, have more of everything than states with fewer people, like Delaware. New York, with a population of almost 20 million, has more pencils, cars, barbers, and murders than Delaware, with under a million residents. New York also has more community hospitals and more deaths, simply because more people live there.

Correlation just means two variables systematically vary together. If the goal of science is to understand cause and effect, it may seem that drawing a conclusion that two variables are associated represents a second-place finish. Not so. Correlation does not *have to* mean causation, but it *may* mean causation. If there is a cause-and-effect relationship between two variables, then they are correlated. An association exists between how much alcohol a person consumes and what his or her blood alcohol level is because consuming alcohol causes the alcohol concentration in the blood to rise.

Correlation may not prove cause and effect, but it can suggest cause. For example, finding that children who watch more violent TV tend to exhibit more aggressive behaviors doesn't prove that TV is the culprit, but it certainly raises such a question and leads to more research. Thus, correlations are not proof of cause and effect, but they often serve as a jumping off point for using experimental techniques to explore cause and effect.

Studies in which the variables in a relationship test are not manipulated by the researcher do not address cause and effect. In a relationship test, it is rare that the explanatory variable is an independent variable and the outcome variable is called a dependent variable. Most commonly, they are just called *X* and *Y* and that's what we'll do here. However, if the researcher believes there is an order to the relationship, then the explanatory variable is called the *predictor variable* and the outcome variable is called the *outcome variable*. The language is meant to be straightforward—the variable that comes first, that is thought of as leading to or influencing the other variable, is the **predictor variable**. The one that comes second, and is influenced by the first variable, is the **outcome variable**.

A Common Question

Q Can a correlation give cause-and-effect information?

A There is a difference between correlation, referring to the statistical test, and a correlational study. In the latter, the explanatory variable is not controlled by the experimenter, so a cause-and-effect conclusion cannot be reached. But, it would be possible to design a study where the experimenter manipulates the explanatory variable and where the results would be analyzed with a Pearson *r*. In such a case, a correlation coefficient would give cause-and-effect information.

Visualizing Relationships

The degree to which two variables vary together determines whether the relationship is strong or weak. The strength of the relationship can be visualized in a scatterplot. Figure 13.2 shows the relationship between randomly generated numbers. If *X* values are randomly selected and each *X* is paired with a randomly selected *Y*, then the two do not vary together systematically. Figure 13.2 shows a rectangular scatterplot where there is *no* relationship, what is called a zero correlation, between two variables. Cases with low values on *X* could have low, medium, or high values on *Y*. And, the same is true for cases with medium or high values on *X*.

Though the rectangular scatterplot shows a zero correlation clearly, that is not how statisticians illustrate a zero correlation. Statisticians use a circular scatterplot to demonstrate no relationship between two variables. If both variables are normally distributed, which is one of the assumptions for the Pearson *r*, then the scatterplot has a circular shape when no relationship exists. Figure 13.3 uses the traditional circular shape to illustrate the lack of relationship between two normally distributed variables.

What does a scatterplot look like when there *is* a relationship between two variables and they vary systematically? Figure 13.4 illustrates the relationship between the temperatures of objects measured in both degrees Fahrenheit and Celsius. This figure is an example of the strongest possible linear relationship between two variables, a **perfect relationship.** In a perfect relationship, all the data points fall along a straight line. Figure 13.4 shows that knowing an object's temperature in Fahrenheit tells one exactly what it is in Celsius.

Strength of Relationships

A strong relationship is one in which cases' scores on variable *X* are closely allied with their scores on variable *Y*. When looking at a scatterplot, strength is shown by how well the points fall along a straight line. The more the points fall in a straight line, the stronger the association. Figure 13.5 gives four scatterplots. As the points form less of a line and more of a blob, as they move from a line to an oval or a circle, the relationship grows weaker.

Two situations exist in which the points in a scatterplot form a straight line, but the relationship is not a perfect one. These occur when the line is horizontal or vertical. When the line is horizontal, for example, every value of *X* is paired with the same

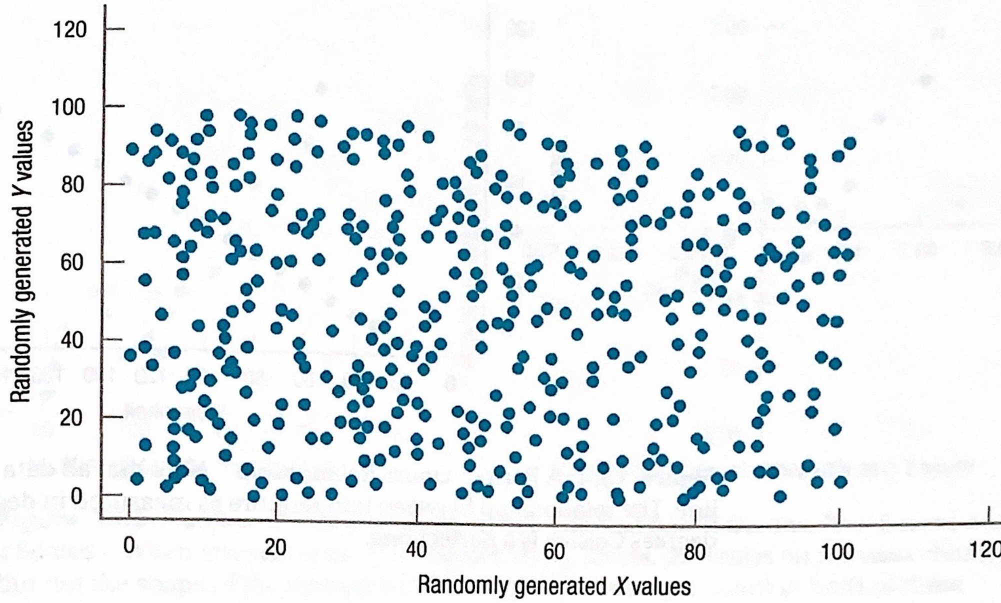


Figure 13.2 Lack of Relationship Between Randomly Generated Values of *X* and Randomly Generated Values of *Y* When the two variables don't vary together systematically, there is no relationship between *X* and *Y*. This is one example of what no relationship looks like.

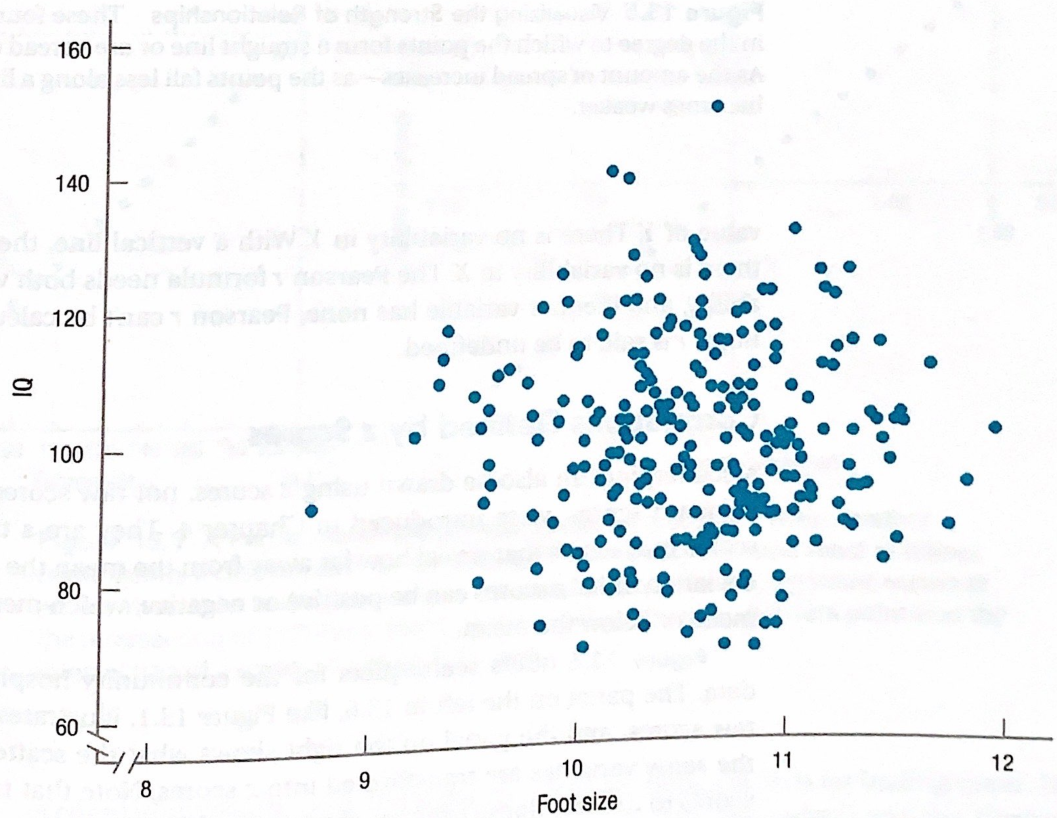


Figure 13.3 Zero Correlation Between Two Normally Distributed Variables If there is no correlation between two normally distributed variables, the scatterplot looks more like a circle than a rectangle. This shows the hypothetical lack of a relationship between foot size and intelligence.

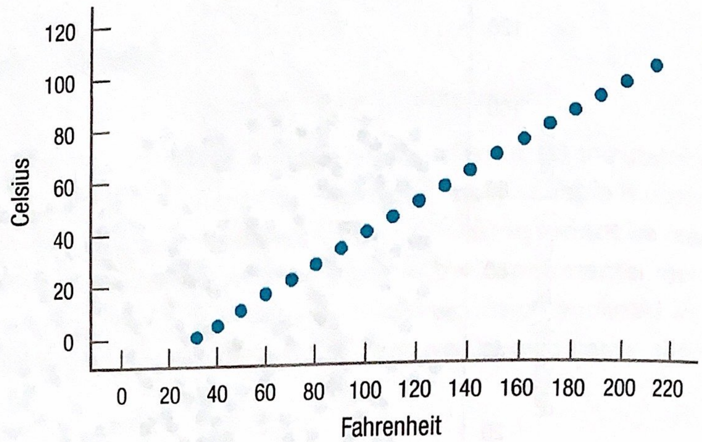


Figure 13.4 A Perfect Linear Relationship Note that all data points fall on a straight line. The relationship between temperature as measured in degrees Fahrenheit and degrees Celsius is a perfect one.

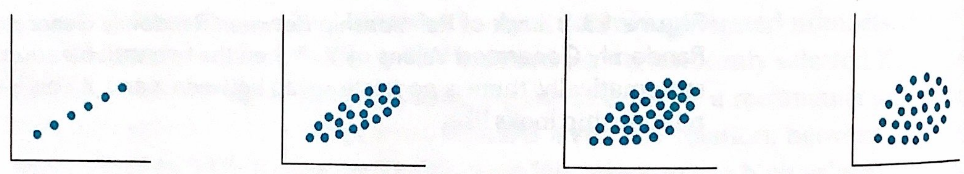


Figure 13.5 Visualizing the Strength of Relationships These four scatterplots differ in the degree to which the points form a straight line or are spread out to form an oval. As the amount of spread increases—as the points fall less along a line—the relationship becomes weaker.

value of *Y*. There is no variability in *Y*. With a vertical line, the opposite is true and there is no variability in *X*. The Pearson *r* formula needs both variables to have variability, and if either variable has none, Pearson *r* can't be calculated. In these situations, *r* is said to be undefined.

Correlations Defined by *z* Scores

A scatterplot can also be drawn using *z* scores, not raw scores. *z* scores, also called standard scores, were introduced in Chapter 4. They are a transformation of raw scores into scores that reveal how far away from the mean the scores are in standard deviation units. *z* scores can be positive or negative, which means they fall above the mean or below the mean.

Figure 13.6 offers scatterplots for the community hospital/number of deaths data. The panel on the left in 13.6, like Figure 13.1, illustrates the relationship with raw scores, and the panel on the right shows what the scatterplot looks like when the same variables are transformed into *z* scores. Note that transforming from raw scores to *z* scores doesn't change the relationship. The pattern of the dots, the shape, is exactly the same, whether raw scores or *z* scores are being plotted.

Figure 13.7 shows the scatterplot for the relationship between temperature as measured in Fahrenheit and Celsius both with raw scores (the panel on the left) and with *z* scores (the panel on the right). These *z* scores give a new perspective on what

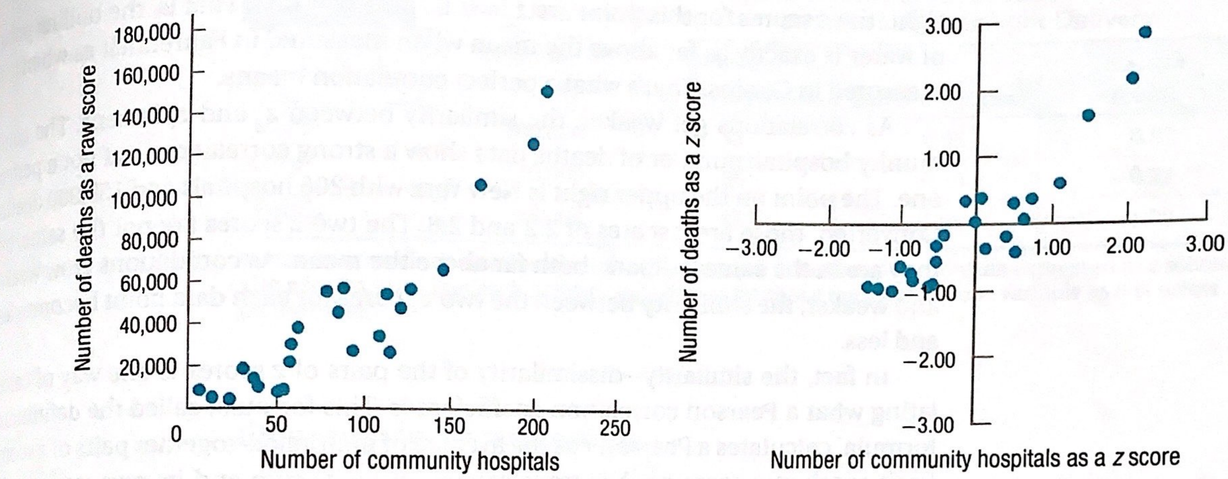


Figure 13.6 Number of Community Hospitals and Number of Deaths: Raw Scores and *z* Scores When raw scores are transformed into *z* scores, the scales on the axes change, but not the shape of the scatterplot. The dots form the same pattern in both of these scatterplots.

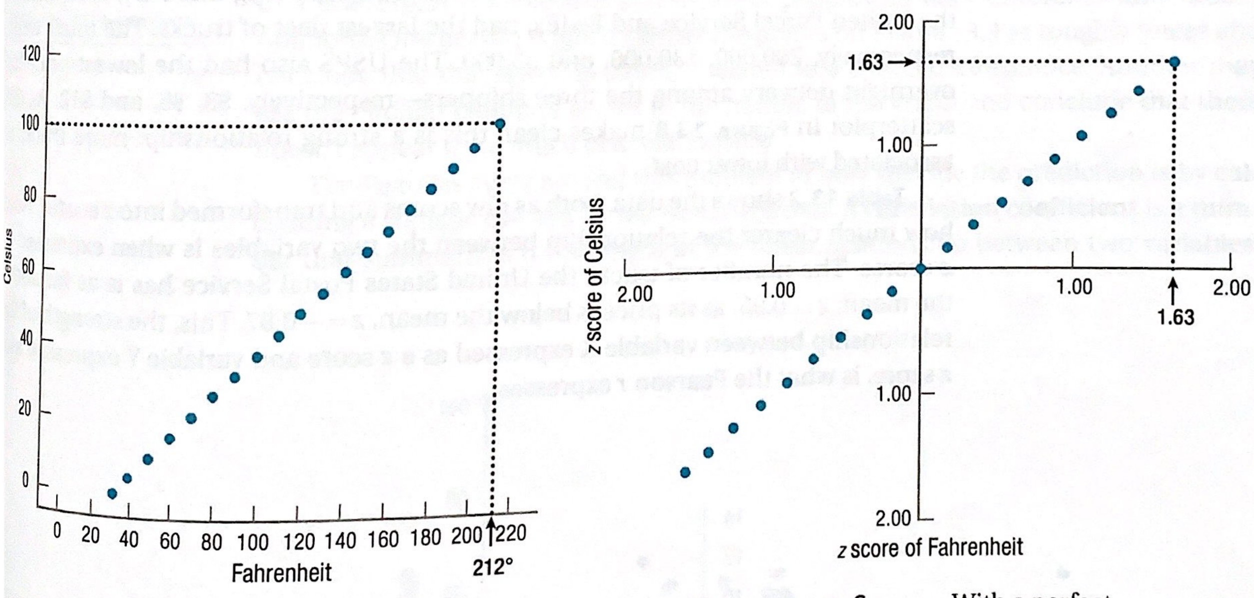


Figure 13.7 A Perfect Relationship: Raw Scores vs. *z* Scores With a perfect relationship, a case's *z* score on the *X* variable has exactly the same value as it does on the *Y* variable. In the panel on the left, the data point for boiling water occurs at the intersection of 212°F and 100°C. In the panel on the right, the data point is at the intersection of *z* scores of 1.63 on both axes.

a correlation means. Look at the dot on the upper right, which is for boiling water. In the panel on the left, the *X* value for this point is 212° (Fahrenheit) and the *Y* value is 100° (Celsius). To a person who doesn't know much about Fahrenheit and Celsius, those scores don't sound similar as they are 112 points apart. In the panel on the right, the z scores for this point are $z_X = 1.63$ and $z_Y = 1.63$. That is, the boiling point of water is exactly as far above the mean when measured in Fahrenheit as when it is measured in Celsius. That's what a perfect correlation means.

As correlations get weaker, the similarity between z_X and z_Y lessens. The community hospital/number of deaths data show a strong correlation, but not a perfect one. The point on the upper right is New York with 206 hospitals and 153,000 deaths. Converted, those are z scores of 2.2 and 2.8. The two z scores are not the same, but they are in the same ballpark, both far above the mean. As correlations grow weaker and weaker, the similarity between the two z scores for each data point becomes less and less.

In fact, the similarity–dissimilarity of the pairs of z scores is one way of calculating what a Pearson correlation coefficient is. This formula, called the definitional formula, calculates a Pearson r as the average of multiplied-together pairs of z scores. Each case's raw score on X is transformed into a z score and its raw score on Y is transformed into a z score. The two z scores for each case are multiplied together and the mean of these products is calculated. The mean of all such multiplied-together scores is a Pearson correlation coefficient. Thus, the stronger the correlation, the more similar—on average—the z scores are.

Here is a quick example to illustrate the benefit of the z score method. Many years ago, the United States Postal Service ran an ad showing that they, compared to the United Parcel Service and FedEx, had the largest fleet of trucks. The totals were, respectively, 200,000, 130,000, and 35,000. The USPS also had the lowest price for overnight delivery among the three shippers—respectively, \$3, \$6, and \$12. As the scatterplot in Figure 13.8 makes clear, this is a strong relationship: more trucks is associated with lower cost.

Table 13.2 shows the data, both as raw scores and transformed into z scores. Note how much clearer the relationship between the two variables is when expressed in z scores. The number of trucks the United States Postal Service has is as far above the mean, $z = 0.95$, as its price is below the mean, $z = -0.87$. This, the strength of the relationship between variable X expressed as a z score and variable Y expressed as a z score, is what the Pearson r expresses.

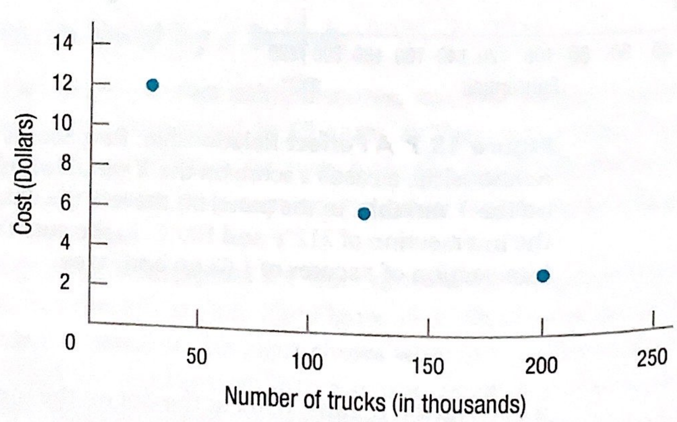


Figure 13.8 Relationship Between Fleet Size and Cost of Overnight Letter This scatterplot shows a strong relationship between how many trucks a delivery service has and the amount it charges for overnight delivery.

TABLE 13.2 Size of Delivery Service Fleet and Cost of Overnight Delivery

	Number of Trucks	Overnight Delivery Cost	z Score (trucks)	z Score (cost)
USPS	200,000	$3	0.95	−0.87
UPS	130,000	$6	0.10	−0.22
FedEx	35,000	$12	−1.05	1.09

Notice how much clearer the relationship is between fleet size and cost when expressed as *z* scores. The *z* scores show that each delivery service is as far above the mean on one variable as it is below on the other.

Quantifying Relationships

A correlation coefficient is a number that summarizes the strength of the relationship between two variables.

Scatterplots are great for visualizing relationships, but their interpretation is subjective. For example, can the owner of a baseball team buy his or her way to the World Series? **Figure 13.9** is a scatterplot showing the relationship between the payrolls of Major League Baseball teams and their winning percentages. How strong is the association? Just by looking at the graph, it is difficult to tell how strong the relationship between these two variables is. Deciding how strong a relationship is on the basis of a scatterplot is subjective and different people can have different—and valid—opinions. One person might interpret the points on Figure 13.9 as roughly linear and conclude that the payroll is strongly related to a team's performance. Another may see the circular clump of scores in the middle of the graph and conclude that there doesn't appear to be much of a relationship.

The Pearson *r* gets around this problem of how specific the prediction is by calculating a statistic called a *correlation coefficient*. A **correlation coefficient** is a number that summarizes the strength of the linear relationship between two variables.

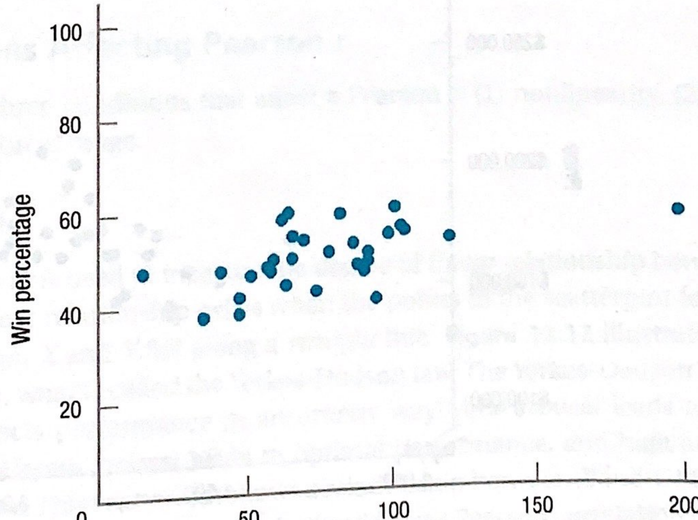


Figure 13.9 Relationship Between Team Payroll and Team Success in Major League Baseball This scatterplot shows the relationship between the payrolls of Major League Baseball teams and their winning percentages.

For the Pearson *r*, the correlation coefficient is abbreviated as *r* (think of *r* as being short for "relationship"). For a Pearson correlation, *r* is a value that ranges from −1.00 to +1.00. An *r* value of zero is less than a weak relationship; it means that there is no relationship between the two variables. Figure 13.2, where the dots form a rectangle, and Figure 13.3, where the dots form a circle, have Pearson *r* values of zero.

As the *r* value moves further away from zero, in either a negative or positive direction, it represents a stronger relationship between *X* and *Y*. For example, an *r* of −.60 represents a stronger relationship than an *r* of .30. Pearson *r* values of −1.00 and 1.00, though they differ in sign, both represent perfect relationships.

Direction of the Relationship

Though Pearson *r* values of −1.00 and 1.00 both represent perfect relationships, the two indicate different types of relationships because their signs differ. The sign of a Pearson *r*, either positive or negative, gives information about the *direction* of the relationship. There are two options for direction: positive or negative.

- **Positive relationships.** Positive *r*'s are found for what are called direct relationships. **Direct relationships** have scatterplots where the points tend to fall along a line moving from the bottom on the left, up and to the right. This means that cases with low scores on the *X* variable tend to have low scores on the *Y* variable and those with high scores on the *X* variable will tend to have high scores on the *Y* variable. For example, within a community there's a positive relationship between the size of a house and how much it costs. In general, smaller houses cost less money and bigger houses cost more money. This positive relationship is illustrated in Figure 13.10.

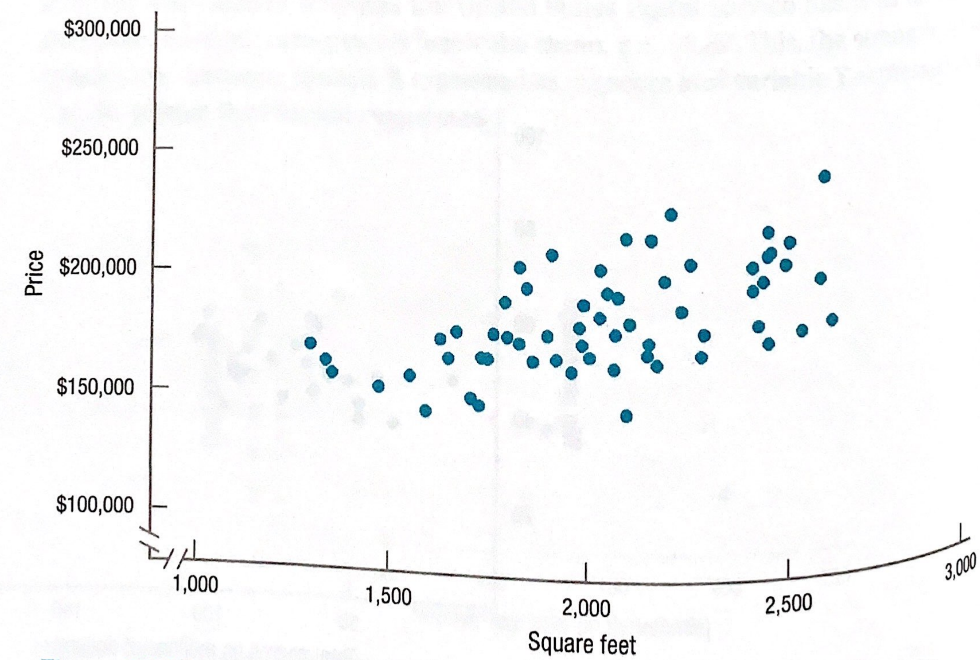


Figure 13.10 Positive Relationship Between the Size of a House and Its Cost In a positive relationship, the data points in a scatterplot fall along a diagonal line that moves up and to the right. In this scatterplot, as the size of the house increases, the price generally does as well.

- **Negative relationships.** Negative *r*'s are also called inverse relationships. **Inverse relationships** have scatterplots where the points fall along a line moving from the top left, down and to the right. This means that cases with low scores on *X* tend to have high scores on *Y* and cases with high scores on the *X* variable will tend to have low scores on the *Y* variable. For example, **Figure 13.11** shows a relationship between a car's horsepower and its fuel economy. The relationship is a negative one—in general, lower horsepower means better fuel economy and higher horsepower means worse fuel economy.

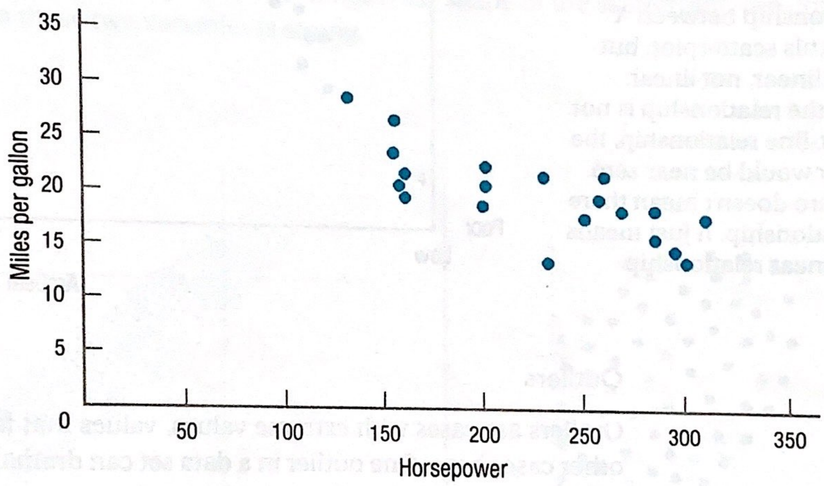


Figure 13.11 Negative Relationship Between Horsepower and Fuel Economy In a negative relationship, points form a line moving downward and to the right. This means that as one variable increases, the other decreases. Here, as the horsepower of cars goes up, miles per gallon go down. (Credit: Thanks to Griffon Olon who collected these data.)

Conditions Affecting Pearson *r*

There are three conditions that affect a Pearson *r*: (1) nonlinearity, (2) outliers, and (3) restriction of range.

Nonlinearity

The Pearson *r* is used to measure the degree of *linear* relationship between two variables. A linear relationship exists when the points in the scatterplot for the relationship between *X* and *Y* fall along a *straight* line. **Figure 13.12** illustrates a nonlinear association, what is called the Yerkes-Dodson law. The Yerkes-Dodson law states that arousal affects performance in an orderly way—low arousal leads to poor performance, moderate arousal leads to optimal performance, and high arousal impairs performance. This scatterplot shows a relationship between *X* and *Y*, but the points in the scatterplot don't fall on a straight line. Using a Pearson correlation to quantify this relationship would provide a value close to zero, meaning there is little *linear* relationship between the two variables. The Pearson *r* only measures how much *linear* relationship exists between two variables. If the relationship appears nonlinear, do not use a Pearson *r* to measure it.

The Pearson r only measures how much linear relationship exists between two variables.

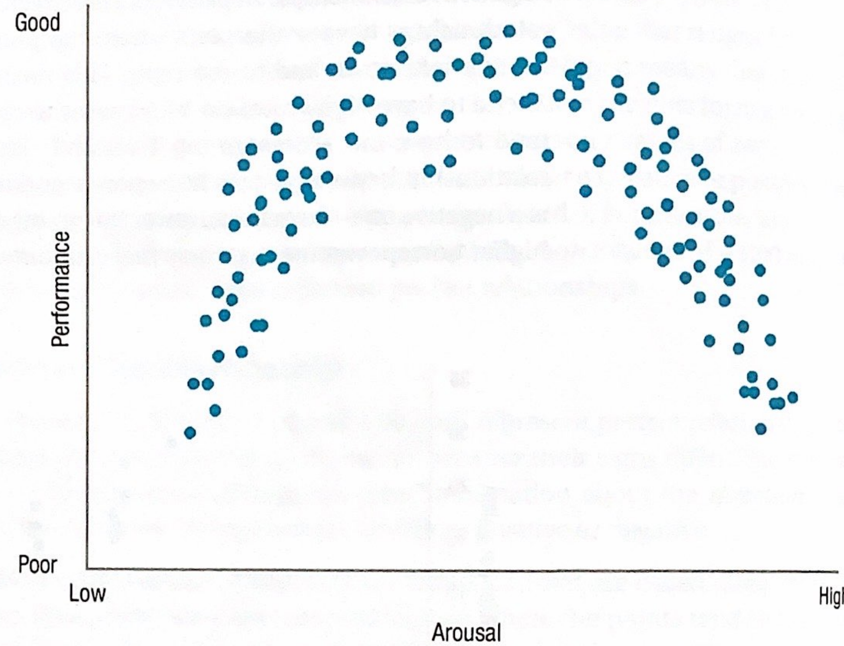


Figure 13.12 Example of a Curvilinear Relationship There is a relationship between X and Y in this scatterplot, but it is curvilinear, not linear. Because the relationship is not a straight-line relationship, the Pearson r would be near zero. An r of zero doesn't mean there is no relationship. It just means that no linear relationship exists.

Outliers

Outliers are cases with extreme values, values that fall far away from the values that other cases have. One outlier in a data set can dramatically change the value of a correlation. The left panel in **Figure 13.13** shows a scatterplot for a small data set. The points form a rough circle and the correlation between X and Y is zero.

The right panel in Figure 13.13 adds one data point, an outlier with extreme values on both X and Y. This data point has an X value and a Y value that are dramatically higher than any other case. As a result of adding this one case, the correlation between X and Y changes from $r = .00$ to $r = .63$.

Outliers inflate the strength of a relationship between two variables. Because outliers may exist, it is always a good idea to create and inspect a scatterplot before calculating a correlation coefficient.

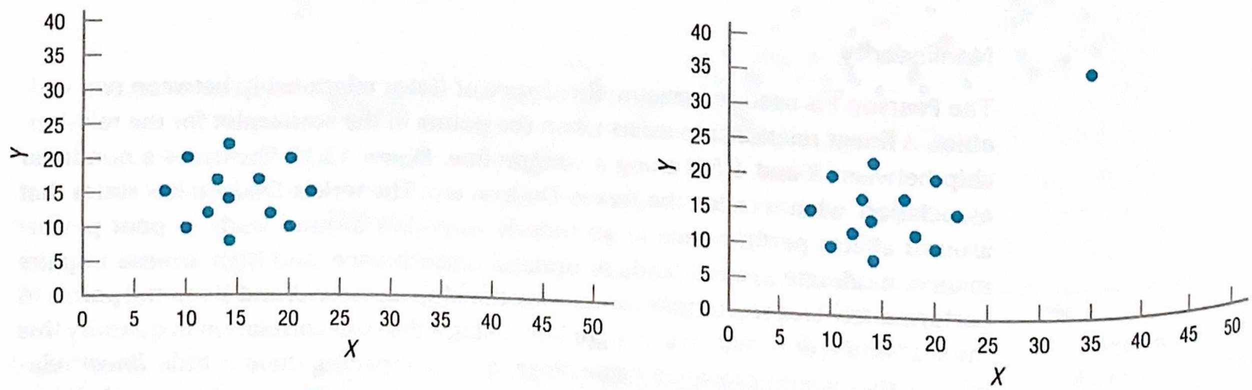


Figure 13.13 Effect of an Outlier on a Pearson Correlation The correlation in the left scatterplot is zero. In the right scatterplot, the correlation is .63 thanks to the addition of one case. The additional case is an outlier, with extreme values on X and Y.

Restriction of Range

While outliers inflate the value of a correlation, restriction of range tends to deflate the value of a correlation. This means that a restricted range could lead a researcher to conclude that there is less of a relationship between two variables than actually exists.

What is restriction of range? Let's consider an unrestricted range first. Look at the left panel of **Figure 13.14**, a scatterplot showing the hypothetical relationship between IQ and GPA in a sample of high school students. There is a full (unrestricted) range of IQ scores, ranging from 70 to 130. And, there is a full (unrestricted) range of GPA, from 0 to 4. Judging by the noncircular shape of the scatterplot, the correlation between these two variables is strong.

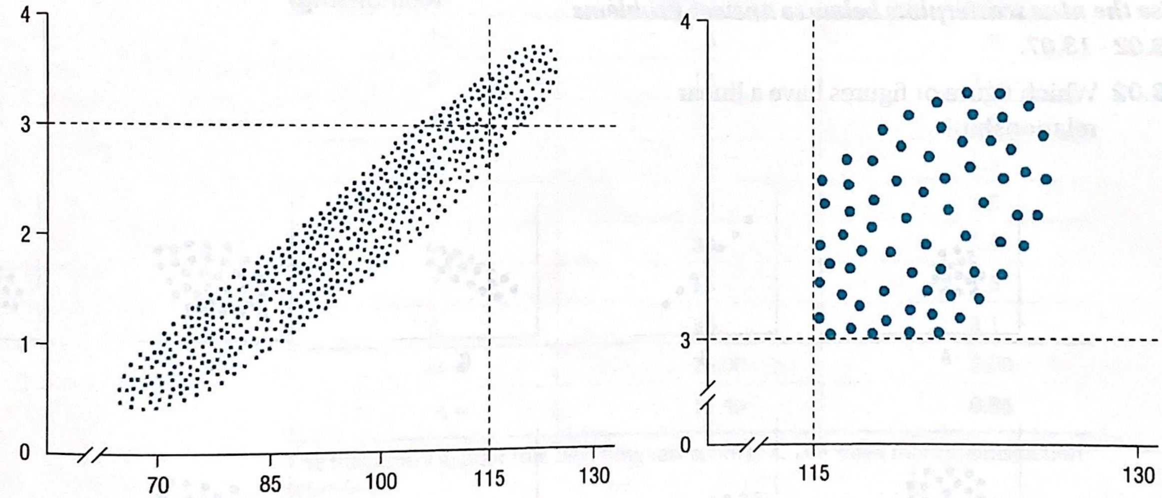


Figure 13.14 Effect of Restriction of Range on Correlation Coefficients The left scatterplot shows the strong relationship between IQ and GPA for a large sample of high school students. The right scatterplot depicts the relationship when the analysis is restricted to people with IQs of 115 or higher *and* GPAs of 3 or higher. Note how the shape of this sample with restricted ranges on both variables is more circular, which means the relationship between IQ and GPA will be much weaker. A restricted range deflates the value of a correlation.

Suppose a researcher decided to restrict her examination of the relationship between IQ and GPA to students likely to be accepted at Ivy League colleges. Thus, she restricted her sample to students with IQs above 115 *and* GPAs above 3.00. The two lines in the top panel are the cut-off values for her sample and the few cases that fall in the upper right quadrant are the new sample.

This subsample has a restricted range on *both* the *X* variable and the *Y* variable. The new sample is shown in a scatterplot all by itself, in the right panel of Figure 13.14. What is the shape of this scatterplot? The points fall roughly in a circle, meaning that little correlation will be found between *X* and *Y*. Restricting the range of one or both variables in a correlation deflates its value.

Practice Problems 13.1

Apply Your Knowledge

13.01 Make a scatterplot for these data:

X	Y	X	Y
100	110	80	95
90	85	100	95
85	95	110	115
90	95	85	80
80	85	90	105
110	125		

Use the nine scatterplots below to answer Problems 13.02–13.07.

13.02 Which figure or figures have a linear relationship?

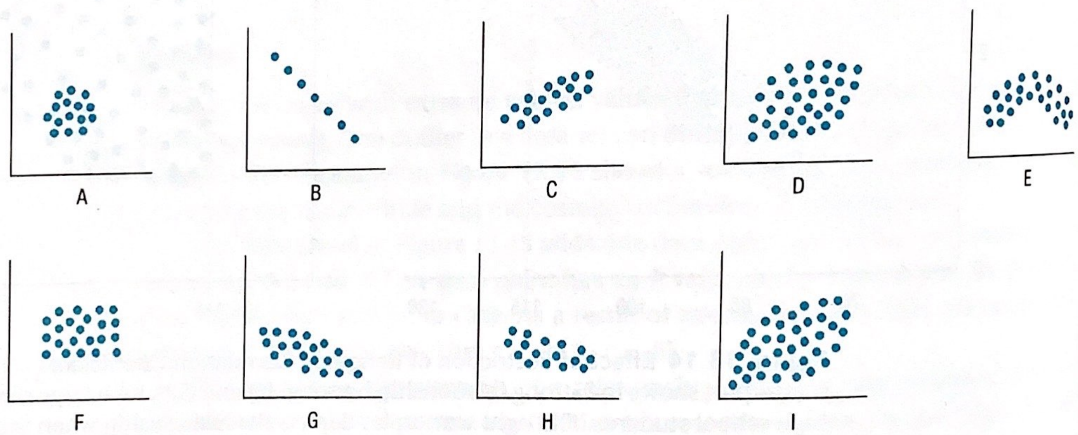
13.03 Which figure or figures represent no relationship?

13.04 Which figure or figures have a weak linear relationship?

13.05 Which figure or figures have a strong, but not perfect, linear relationship?

13.06 Which figure or figures have a perfect linear relationship?

13.07 Which figure or figures have a negative linear relationship?



13.2 Calculating the Pearson Correlation Coefficient

To learn how to calculate a Pearson r, imagine some data collected by a marriage therapist. Dr. Paik was interested in seeing whether there was a relationship between gender role flexibility and marital satisfaction. Gender role flexibility refers to the ability to express both male and female traits. Dr. Paik wanted to find out if women's marital satisfaction correlated with how gender role flexible their husbands were.

To measure gender role flexibility, he used the Role Flexibility Test (RFT). The RFT is scored on an interval level and scores range from 0 to 40. Higher scores mean more role flexibility. To measure marital satisfaction, he asked the women to grade their husbands A–F on a number of dimensions. He averaged these grades together, and then he calculated a marital "GPA." Just like an academic GPA, a marital GPA ranges from 0 to 4, with 0 = F and 4 = A.

Dr. Paik obtained a random sample of eight heterosexual married couples from his city and measured two characteristics for each couple (see the data in Table 13.3). *X* was the husband's level of gender role flexibility, and *Y* was the wife's rating of marital satisfaction. Dr. Paik's question was simple. He did not specify a direction, direct or inverse, for the relationship, he just wanted to know: Are the two variables related?

TABLE 13.3 Data for Dr. Paik's Study of the Relationship Between Gender Role Flexibility and Marital Satisfaction

Case	X Gender Role Flexibility	Y Marital Satisfaction
1.	8	.8
2.	15	2.0
3.	22	1.5
4.	31	2.3
5.	35	1.5
6.	38	3.3
7.	15	1.5
8.	36	3.1
$M =$	25.00	2.00
$s =$	11.49	0.86

The husband's gender role flexibility test score is *X*. The wife's marital satisfaction score is *Y*.

Step 1 Pick a Test

The first step in hypothesis testing involves picking the correct statistical test. Here, one group of cases ($N = 8$) is used to examine the linear relationship between two interval/ratio level variables. This calls for a Pearson *r*.

Step 2 Check the Assumptions

The Pearson *r* has four assumptions that should be met in order to proceed with the test. Three of the assumptions are familiar (random sample, independence of observations, and normality). One assumption, linearity, is new.

- *Random sample.* This assumption says that the sample in which the correlation is being calculated is a random sample from the population to which the results will be generalized. As with *t* tests and ANOVAs, the random sample assumption is robust to violation. With the marital satisfaction study, the random sample assumption is not violated. This is one of those rare times there is a random sample. The researcher can generalize the results of the study to the whole city. However, the sample size is small, so replication should be on Dr. Paik's mind already.

"Tom and Harry despise crabby infants" is the mnemonic for the six steps of hypothesis testing: (1) Pick a **test**, (2) check the **assumptions**, (3) list the **hypotheses**, (4) set the **decision rule**, (5) **calculate** the test statistic, and (6) **interpret** the results.

- *Independence of observations.* This assumption says each case in the sample is independent of every other case, that the cases don't influence each other. This assumption is not robust, so a researcher shouldn't use the Pearson *r* if it is violated. Even though the cases are couples, the independence of observations assumption is not violated for the marital satisfaction study because each couple was measured individually and was only in the sample once.
- *Normality.* This assumption says that each of the variables in a Pearson *r*, both *X* and *Y*, is normally distributed in the population. This assumption is robust to violation, especially if the sample size is large. The normality assumption is not violated for the marital satisfaction data. It seems reasonable to assume, in the larger population of the city, that both gender role flexibility and marital satisfaction are normally distributed.
- *Linearity.* This assumption is a new one. It says that the relationship between the two variables in the population is a linear one. The linearity assumption is tested by making a scatterplot for the data. If the dots in the scatterplot fall on a curve, as they do in Figure 13.12, then this assumption is violated. If the dots form an irregular shape, like that seen in Figure 13.13, or fall in a straight line, then the linearity assumption is not violated. This assumption is not robust, so if it is violated, the test should not proceed. The scatterplot for Dr. Paik's data appears in **Figure 13.15**, which shows the points falling roughly along a straight line that moves up and to the right. The shape is not curvilinear, so the linearity assumption is not violated for the marital satisfaction data.

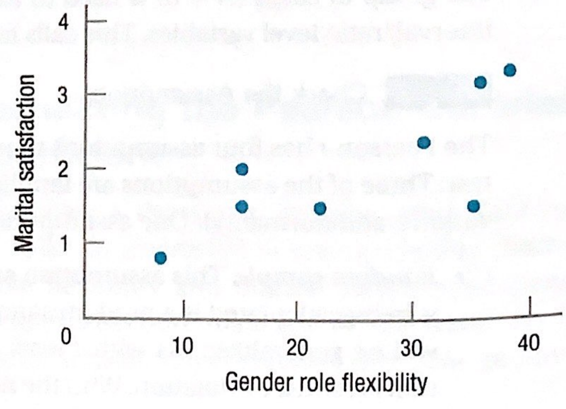


Figure 13.15 Relationship Between Male Gender Role Flexibility and Female Marital Satisfaction The data points in this scatterplot for the most part fall along a straight line, so the linearity assumption is not violated.

With no assumptions violated, Dr. Paik can proceed to the next step.

Step 3 List the Hypotheses

The null and alternative hypotheses are statements about populations, not samples. For the population value of a correlation, the Greek letter rho, ρ, will be used as the abbreviation. Though ρ looks like the letter "p," it is the Greek letter "r" and is pronounced "row." ρ represents the *population* value of a correlation, while r is the value of a correlation when it is calculated for the data in a *sample*. The correlation coefficient Dr. Paik is calculating for the sample of eight cases in the marital satisfaction study is an r. If the correlation coefficient were calculated for all the married couples in the city, it would be a ρ, a population value.

For a nondirectional or two-tailed test, the null hypothesis states that there is no relationship, in the population, between X and Y. This is expressed by saying that ρ equals zero:

$$H_0: \rho = 0$$

The alternative hypothesis says there is *some* relationship between X and Y in the population. It might be a strong relationship (near 1.00), a very weak relationship (say, .01), or anywhere between those extremes. It might be a positive relationship or it might be a negative relationship. It just is something other than a zero-relationship. So, the alternative hypothesis is

$$H_1: \rho \neq 0$$

If a researcher has a directional hypothesis, then he or she would specify what the expected direction of the relationship is before any data are collected. This direction is then stated as the alternative hypothesis. For example, if the researcher expected a direct relationship between X and Y, that is, a correlation coefficient with a positive sign, then the alternative hypothesis would be $\rho > 0$; for an expected inverse relation, the alternative hypothesis would read $\rho < 0$. The null hypothesis is then set so that the two hypotheses together are all-inclusive and mutually exclusive. If the alternative hypothesis is $\rho > 0$, then a null hypothesis of $\rho \leq 0$ would satisfy those criteria.

Step 4 Set the Decision Rule

The sampling distribution of r is based on the population value of the correlation being zero. If the null hypothesis is nondirectional, then the sampling distribution of r would look like Figure 13.16. Note the following points about the sampling distribution:

- All the values of r fall from -1.00 to 1.00.
- The distribution is centered on zero, and $r = .00$ has the highest frequency.
- The distribution is symmetric. One is just as likely to draw a random sample that has a negative r as a positive r.
- As one moves away from $r = .00$, the frequencies become smaller.

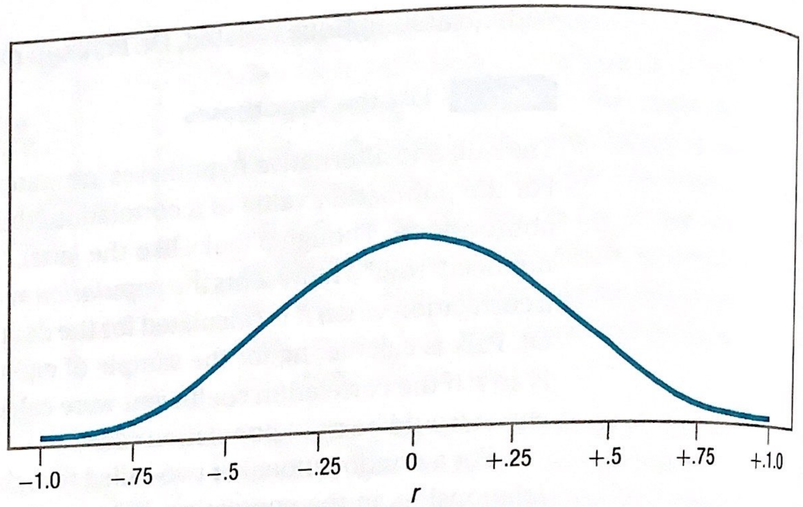


Figure 13.16 Example of a Sampling Distribution of Pearson *r* Centered at zero, *r* is symmetrical and ranges from −1.00 to 1.00. Note that the frequencies decrease as *r* moves away from zero.

As with the sampling distribution for *t* and *F*, this sampling distribution can be divided into common and rare zones:

- The common zone is the middle section of the sampling distribution, the section centered around zero where it would be common to find the *r* value for a sample when the null hypothesis is true.
- The rare zone is the extreme sections of the sampling distribution. If the null hypothesis is true, it is rare to find a sample *r* value that falls in this section.

The decision rule for the Pearson *r* is similar to the other tests:

- If the value of *r* calculated for the sample falls in the rare zone, reject the null hypothesis.
- If the value of *r* calculated for the sample falls in the common zone, fail to reject the null hypothesis.

The boundary between the common zone and the rare zone is called the critical value of *r*, abbreviated r_{cv}. Most commonly, the rare zone is set to be 5% of the sampling distribution. Phrased in the language of statistics, this means alpha (α) is set at .05, giving a 5% chance of a Type I error.

Besides deciding what alpha level to use, a researcher has to decide whether to do a one-tailed or two-tailed test. With a one-tailed test, the hypotheses are directional, so the entire rare zone is placed in one tail of the sampling distribution. A directional alternative hypothesis makes a statement about the direction of ρ, such as $\rho > 0$ or $\rho < 0$. Directional hypotheses need to be stated before data are collected.

With a two-tailed test, the hypotheses are nondirectional, so the rare zone is split evenly between the two tails. Two-tailed tests are more common. For example, the hypotheses for the marital satisfaction study are nondirectional, with the alternative hypothesis simply stating $\rho \neq 0$. The hypotheses are nondirectional because Dr. Paik hasn't predicted whether higher gender role flexibility scores or lower gender role flexibility scores are associated with more marital satisfaction.

Two-tailed tests with $\alpha = .05$ are the "default" option for the Pearson correlation coefficient. This is what Dr. Paik plans to use. When there is a two-tailed test and $\alpha = .05$, the middle 95% of the sampling distribution is the common zone and the rare zone is made up of the 2.5% on the far left of the sampling distribution *and* the 2.5% on the far right. The sampling distribution for the marital satisfaction study, with the rare and common zones marked, is shown in **Figure 13.17**.

Critical values of r are listed in Appendix Table 6, a portion of which is in **Table 13.4**. Finding the critical value of r follows the same process as finding a critical value of t:

- First, select a column to use based on whether it is a one-tailed test or a two-tailed test *and* what the level of alpha is.

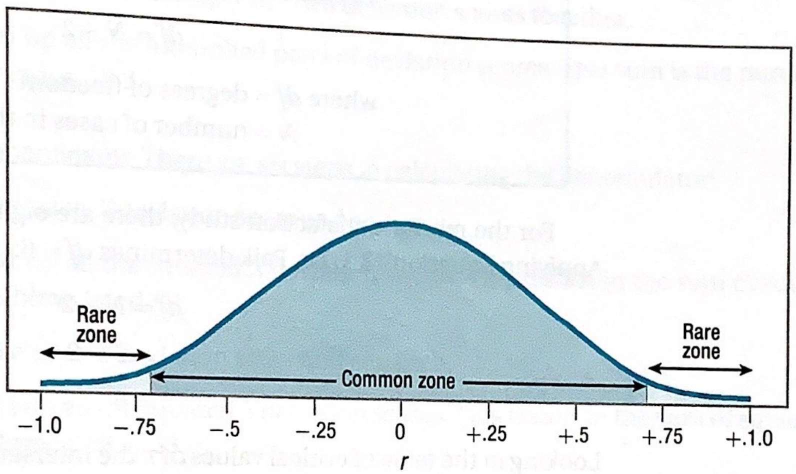


Figure 13.17 Common Zone and Rare Zones for the Sampling Distribution of r for Dr. Paik's Marital Satisfaction Study The critical value of r is $\pm.707$. If the observed value of r falls in the common zone of the sampling distribution, one will fail to reject the null hypothesis. If r falls on either of the lines or in either rare zone, reject the null hypothesis.

TABLE 13.4 Section of Appendix Table 6, Table of Critical Values of r

df	$\alpha = .10$, two-tailed *or* $\alpha = .05$, one-tailed	$\alpha = .05$, two-tailed *or* $\alpha = .025$, one-tailed
1	.988	.997
2	.900	.950
3	.805	.878
4	.729	.811
5	.669	.754
6	.621	.707
7	.582	.666
8	.549	.632
9	.521	.602
10	.497	.576

The critical value of r, r_{cv}, is found at the intersection of the column for the selected alpha level and number of tails and the row for the degrees of freedom.

- Then, find the row that contains r_{cv} by selecting the row with the correct degrees of freedom. (If there is no row for the degrees of freedom one needs, follow *The Price Is Right* rule and use the row for the degrees of freedom that is closest without going over.)

Equation 13.1 shows how to calculate degrees of freedom for a Pearson correlation coefficient.

Equation 13.1 Formula for Calculating Degrees of Freedom (*df*) for a Pearson Correlation Coefficient

$$df = N - 2$$

where df = degrees of freedom
N = number of cases in the sample

For the marital satisfaction study, there are eight cases in the sample, so $N = 8$. Applying Equation 13.1, Dr. Paik determines $df = 6$:

$$\begin{aligned} df &= N - 2 \\ &= 8 - 2 \\ &= 6 \end{aligned}$$

Looking in the table of critical values of *r*, the intersection of the row where $df = 6$ and the column where $\alpha = .05$, two-tailed, we determine that the critical value of *r* is .707. There is no sign attached to this critical value. Dr. Paik has opted for a two-tailed test and the sampling distribution is symmetric, so $r_{cv} = \pm.707$. The decision rule can now be written for the marital satisfaction study:

- If $r \leq -.707$ *or* if $r \geq .707$, reject H_0.
- If $-.707 < r < .707$, fail to reject H_0.

Step 5 Calculate the Test Statistic

Equation 13.2 is the computational formula for calculating the Pearson *r*. Computational formulas are mathematically easier than definitional formulas. Equation 13.2 may look complex, but it isn't difficult when broken down into chunks.

Equation 13.2 Formula for Calculating *r*, the Pearson Correlation Coefficient

$$r = \frac{\Sigma[(X - M_X)(Y - M_Y)]}{\sqrt{SS_X SS_Y}}$$

where r = Pearson correlation coefficient
X = a case's score on variable X
M_X = mean score on variable X
Y = a case's score on variable Y
M_Y = mean score on variable Y
SS_X = sum of the squared deviation scores for variable X
SS_Y = sum of the squared deviation scores for variable Y

The easiest way to complete Equation 13.2 is in three pieces: one for the numerator, one for the denominator, and one to finish the division.

The Numerator. There are four steps in calculating the numerator:

1. Take each X score and subtract the mean of the X scores from it. The resulting values are the X deviation scores.
2. Take each Y score and subtract the mean of the Y scores from it. The resulting values are the Y deviation scores.
3. For each case, multiply the two deviation scores together.
4. Add up all the multiplied pairs of deviation scores. This sum is the numerator in Equation 13.2.

The Denominator. There are six steps in calculating the denominator:

1. Take each X deviation score and square it.
2. Add up all the squared X deviation scores. This results in the sum of squares for X, abbreviated SS_X.
3. Take each Y deviation score and square it.
4. Add up all the squared Y deviation scores. This results in the sum of squares for Y, abbreviated as SS_Y.
5. Multiply SS_X from Step 2 by SS_Y from Step 4.
6. The square root of the product found in Step 5 is the denominator in Equation 13.2.

Final Calculations. The last step in calculating Pearson r is to take the numerator and divide it by the denominator, yielding the r value.

Doing the Calculations. The marital satisfaction data are re-presented in Table 13.5. The first two columns contain the X value and the Y value for each case (the same

TABLE 13.5 Calculating Pearson r for Dr. Paik's Marital Satisfaction Study

	1	2	3	4	5	6	7
	X	Y	$X - M_X$	$Y - M_Y$	$(X - M_X)(Y - M_Y)$	$(X - M_X)^2$	$(Y - M_Y)^2$
	8	.8	−17.00	−1.20	20.40	289.00	1.44
	15	2.0	−10.00	0.00	0.00	100.00	0.00
	22	1.5	−3.00	−0.50	1.50	9.00	0.25
	31	2.3	6.00	0.30	1.80	36.00	0.09
	35	1.5	10.00	−0.50	−5.00	100.00	0.25
	38	3.3	13.00	1.30	16.90	169.00	1.69
	15	1.5	−10.00	−0.50	5.00	100.00	0.25
	36	3.1	11.00	1.10	12.10	121.00	1.21
M	25.00	2.00			$\Sigma = 52.70$	$\Sigma = SS_X = 924.00$	$\Sigma = SS_Y = 5.18$

The original data can be found in column 1 (gender role flexibility) and column 2 (marital GPA). Columns 3–7, as outlined in the text, are steps in the calculation of the Pearson r.

information as in Table 13.3). But, Table 13.5 contains five new columns (numbered as columns 3–7), each of which will be explained below as it is used.

To find Pearson r, calculate the numerator first:

1. Column 3, labeled "$X - M_X$," is used to calculate deviation scores for the X variable. It is completed by subtracting the mean for the X scores, which is 25.00 here, from each X score. The first case has an X score of 8.00, so the deviation score is $8.00 - 25.00 = -17.00$. To complete column 3, repeat this step for each of the X scores.
2. Column 4, labeled "$Y - M_Y$," calculates deviation scores for the Y variable. The same action is taken here for each Y value that was done for each X value in column 3. That is, the mean for Y is subtracted from each Y score. For example, the first case has a Y score of 0.8 and M_Y is 2.00, so the deviation score is $0.8 - 2.00 = -1.20$. To complete column 4, repeat this step for each of the Y scores.
3. In column 5, labeled "$(X - M_X)(Y - M_Y)$," the two deviation scores for each case are multiplied together. For the first case, which is in the first row, $-17.00 \times -1.20 = 20.40$. For each row, column 3 is multiplied by column 4.
4. Finally, add up all the deviation scores that were multiplied together in column 5. At the bottom of column 5, it says $\Sigma = 52.70$. That sum of the multiplied-together deviation scores is the numerator Dr. Paik will use to calculate r.

The next step to find Pearson r is calculating the denominator:

1. In column 6, labeled "$(X - M_X)^2$," square the X deviation scores from column 3. The X deviation score for the first case was -17.00, so $-17.00^2 = 289.00$. Repeat this step for each of the scores in column 3.
2. Once all the X deviation scores have been squared, add them all up at the bottom of column 6. $SS_X = 924.00$.
3. In column 7, labeled "$(Y - M_Y)^2$," square the Y deviation scores from column 4. For example, for the first case, $-1.20^2 = 1.44$. Repeat this step for each of the scores in column 4.
4. Next, add up the squared Y deviation scores and write the value at the bottom of column 7. $SS_Y = 5.18$.
5. Then, multiply together the two sums of squared deviation scores, 924.00 (Step 2) and 5.18 (Step 4): $924.00 \times 5.18 = 4{,}786.32$.
6. Next, to find the denominator, take the square root of the product that was calculated in Step 5: $\sqrt{4{,}786.32} = 69.1832 = 69.18$.

Now that the numerator of Equation 13.2 (52.70) and the denominator of Equation 13.2 (69.18) are known, Dr. Paik can calculate r:

$$
\begin{aligned}
r &= \frac{\Sigma[(X - M_X)(Y - M_Y)]}{\sqrt{SS_X SS_Y}} \\
&= \frac{52.70}{69.18} \\
&= .7618 \\
&= .76
\end{aligned}
$$

Dr. Paik now knows that $r = .76$. His next task is to interpret this correlation. We'll turn to that after one more experience going through the first five steps of hypothesis testing for a Pearson r.

A Common Question

Q Why is r reported as .76, not 0.76?

A APA format says not to use a zero before the decimal point for correlations.

Worked Example 13.1 Imagine that a developmental psychologist, Dr. Solomon, wanted to determine if there were a relationship between the age at which children started to walk and their intelligence at age 16. She went to a pediatrician's office and randomly selected 10 charts of 16-year-old girls. In the charts, she found the age (in months) at which each girl started walking and then she gave each girl a standard IQ test.

In this instance, there is a clear order to the variables—age of walking comes first and IQ is measured later. This is an example of a relationship test with a predictor variable (age) and an outcome variable (IQ). When the variables can be classified as predictor and outcome, the convention is to make the predictor variable X and the outcome variable Y. That is how the data are shown in **Table 13.6**, with age of walking as X and IQ as Y. Remember, correlations measure association. The fact that one variable is used to predict the other should not be taken to mean that the one causes the other.

TABLE 13.6 Data for Dr. Solomon's Study Exploring the Relationship Between Age of First Walking and IQ at Age 16

	Age of Walking (in months)	Intelligence at Age 16
1.	9	115
2.	15	100
3.	11	90
4.	14	95
5.	13	115
6.	10	100
7.	12	110
8.	18	90
9.	17	105
10.	16	90
M	13.50	101.00
s	3.03	9.94

A Pearson r will be used to see if age of first walking, the X variable, predicts IQ, the Y variable.

Step 1 Pick a Test. There is one group of cases in which two interval/ratio level variables were measured in order to see if a relationship exists between them. This calls for a Pearson *r*.

Step 2 Check the Assumptions.

- *Random sample.* The random sample assumption is not violated. The sample is from one pediatric practice, though, so the results shouldn't be generalized beyond that practice.
- *Independence of observations.* Each case is in the sample only once. There's no reason to think that the cases influence each other in terms of age of first walking or IQ. The independence of observations assumption is not violated.
- *Normality.* Researchers consider IQ to be normally distributed. It seems reasonable to consider age at first walking normally distributed as well. This assumption is not violated.
- *Linearity.* There is no obvious curvilinear relationship in the scatterplot (**Figure 13.18**), so this assumption is not violated.

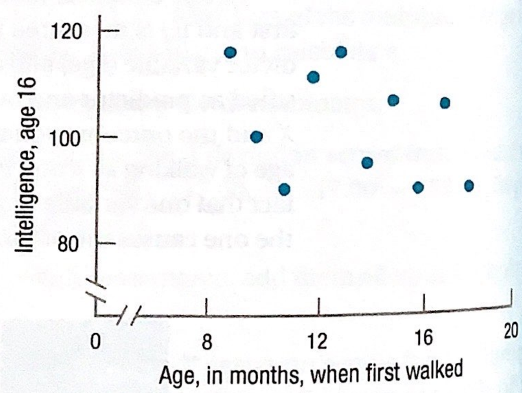


Figure 13.18 Dr. Solomon's Data Showing the Relationship Between Age of First Walking and Intelligence at Age 16 There is no obvious curvilinear relationship in this scatterplot, so the linearity assumption is not violated.

With no assumptions violated, Dr. Solomon can proceed with the Pearson *r*.

Step 3 List the Hypotheses. Dr. Solomon has not made a prediction about the direction of the relationship, so her hypotheses are nondirectional or two-tailed:

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

Step 4 Set the Decision Rule. With $N = 10$, the degrees of freedom are calculated with Equation 13.1:

$$\begin{aligned} df &= N - 2 \\ &= 10 - 2 \\ &= 8 \end{aligned}$$

Dr. Solomon is examining whether there is a relationship between age of first walking and IQ, either positive or negative, so the test is two-tailed. She's willing to have a 5% chance of a Type I error, so $\alpha = .05$. Next, she consults the table of critical values of r, Appendix Table 6. The intersection of the row with 8 degrees of freedom and the column for $\alpha = .05$, two-tailed gives $r_{cv} = \pm.632$. The decision rule is:

- If $r \leq -.632$ *or* if $r \geq .632$, reject H_0.
- If $-.632 < r < .632$, fail to reject H_0.

Figure 13.19 shows how Dr. Solomon sketched the decision rule as a sampling distribution of r, marking the rare and common zones.

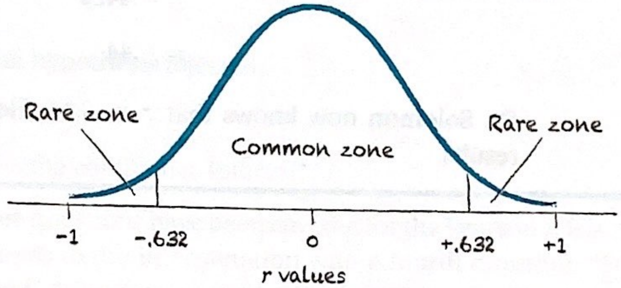


Figure 13.19 Decision Rule for Dr. Solomon's IQ/Walking Study This sketch of a sampling distribution of r uses the critical value of r, $\pm.632$, to mark the rare and common zones.

Step 5 **Calculate the Test Statistic.** **Table 13.7** shows the deviation scores for age of first walking (X) and IQ (Y), the deviation scores that were multiplied together, and the squared deviation scores.

- For the numerator in Equation 13.2, add up the deviation scores that were multiplied together. The bottom of column 5 gives the result: -120.00.
- The sums of the squared deviation scores are 82.50 (SS_X in column 6) and 890.00 (SS_Y in column 7):
- Multiply these together: $82.50 \times 890.00 = 73{,}425.00$.
- Take the square root of this product to find the denominator: $\sqrt{73{,}425.00} = 270.97$.

TABLE 13.7 Calculating Pearson r for Dr. Solomon's Age of Walking/IQ Study

	1	2	3	4	5	6	7
	X	Y	$X - M_X$	$Y - M_Y$	$(X - M_X)(Y - M_Y)$	$(X - M_X)^2$	$(Y - M_Y)^2$
	9	115	−4.50	14.00	−63.00	20.25	196.00
	15	100	1.50	−1.00	−1.50	2.25	1.00
	11	90	−2.50	−11.00	27.50	6.25	121.00
	14	95	0.50	−6.00	−3.00	0.25	36.00
	13	115	−0.50	14.00	−7.00	0.25	196.00
	10	100	−3.50	−1.00	3.50	12.25	1.00
	12	110	−1.50	9.00	−13.50	2.25	81.00
	18	90	4.50	−11.00	−49.50	20.25	121.00
	17	105	3.50	4.00	14.00	12.25	16.00
	16	90	2.50	−11.00	−27.50	6.25	121.00
M	13.50	101.00			$\Sigma = -120.00$	$\Sigma = SS_X = 82.50$	$\Sigma = SS_Y = 890.00$

The calculations in this table lead through the steps necessary to obtain the numerator and denominator for calculating a Pearson correlation coefficient.

Now that the numerator (−120.00) and the denominator (270.97) are known, the calculation of Pearson r is straightforward:

$$r = \frac{\Sigma[(X - M_X)(Y - M_Y)]}{\sqrt{SS_X SS_Y}}$$

$$= \frac{-120.00}{270.97}$$

$$= -.4429$$

$$= -.44$$

Dr. Solomon now knows that $r = -.44$. Her next step will be to interpret the results.

Practice Problems 13.2

Apply Your Knowledge

13.08 Read each scenario and decide if the data can be analyzed with a Pearson r.

a. A researcher from a facial tissue manufacturer rates the severity of people's colds on a 15-point interval scale and measures how many tissues they use in a 24-hour period. She wants to determine if there is a relationship between the severity of a cold and tissue use.

b. A high school counselor obtains each student's class rank and IQ score. She wants to know if a relationship exists between class rank and IQ.

c. People who use Apple computers and those who use Windows-based computers are measured on an interval level of creativity. Is there a relationship between the type of computer a person uses and his or her level of creativity?

13.09 A dietician goes to a mall on a Sunday afternoon and finds people shopping by themselves who are willing to complete questionnaires about weekly food consumption and weekly exercise. She wants to see if any relationship exists between caloric consumption and caloric expenditure. She is planning to use a Pearson r. (a) Check as many assumptions as possible and (b) decide if she can proceed.

13.10 A cosmetician develops a theory that the longer a man's hair is, the more tattoos and piercings he is likely to have. She obtains a sample of 37 men, measures how long their hair is in millimeters, and counts how many tattoos and piercings each has.

a. List her hypotheses.

b. With $\alpha = .05$, list her decision rules.

13.11 Given the following values for X and Y, complete the table:

	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7
	X	Y	$X - M_X$	$Y - M_Y$	$(X - M_X)(Y - M_Y)$	$(X - M_X)^2$	$(Y - M_Y)^2$
	10	20					
	14	24					
	9	37					
M	11.00	27.00			$\Sigma =$	$\Sigma = SS_X =$	$\Sigma = SS_Y =$

13.12 Given $\Sigma[(X - M_X)(Y - M_Y)] = 80.00$, $SS_X = 200.00$, and $SS_Y = 90.00$, what is r?

13.3 Interpreting the Pearson Correlation Coefficient

Step 6 Interpret the Results

To interpret a Pearson correlation coefficient, there are three questions to be addressed. The answers to these three questions provide the raw material from which the researcher selects the most salient pieces to build a four-point interpretation. The three questions are the same ones used in the interpretation of an independent-samples *t* test:

- Was the null hypothesis rejected?
- How big is the effect?
- How wide is the confidence interval?

After those three questions have been covered for the Pearson *r*, it is time to add more nuance and depth to the interpretation with a fourth question, "Did this test have adequate power?" As we learned in Chapter 6, power is the likelihood of being able to reject the null hypothesis when it should be rejected. Having adequate power is usually raised as a concern when the null hypothesis is not rejected. So, we will save our exploration of power until the next worked example, when Dr. Solomon interprets the Pearson *r* for the relationship between age of walking and intelligence. Be sure to read the worked example—it covers new ground.

Until then, we will follow Dr. Paik as he interprets the results from his marital satisfaction study. In that study, he obtained a random sample of eight couples from the city where he lives. For each couple, he measured the husband's degree of gender role flexibility and the wife's level of marital satisfaction in order to see if the two variables were related.

Was the Null Hypothesis Rejected?

To determine if the null hypothesis is rejected, Dr. Paik will use the decision rule from Step 4. For the marital satisfaction study, the observed value of *r* was calculated, in Step 5, to be .76. In Step 4, the critical value of *r* was found to be ±.707. Dr. Paik has to decide:

- Is $.76 \leq -.707$ *or* is $.76 \geq .707$?
- Is $-.707 < .76 < .707$?

The second part of the first statement, $.76 \geq .707$, is true, so the null hypothesis is rejected. **Figure 13.20** shows how the results fall in the rare zone. The null hypothesis is rejected and the results are called statistically significant. Dr. Paik can say that this Pearson *r* of .76 is statistically different from zero. In APA format, he would write

$$r(6) = .76, p < .05$$

APA format for a Pearson *r* contains five pieces of information: (1) what test is being done, (2) how many cases there are, (3) the observed value of the test statistic, (4) what alpha level was selected, and (5) whether the null hypothesis was rejected.

1. The *r* tells that the test was a Pearson *r*.

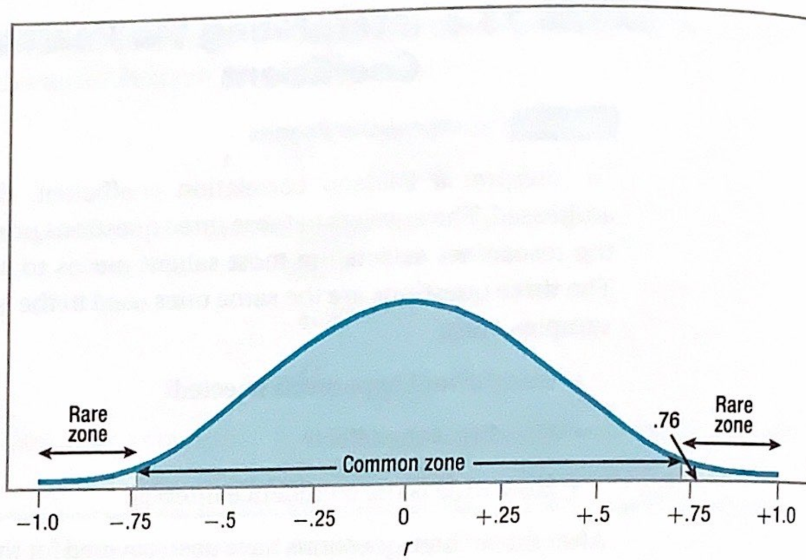


Figure 13.20 Implementing the Decision Rule for Dr. Paik's Marital Satisfaction Study The observed value of *r*, .76, falls in the rare zone of the sampling distribution. It is an unusual result if the null hypothesis is true, so the null hypothesis is rejected.

2. The 6 in the parentheses is the degrees of freedom for the test. Degrees of freedom for a Pearson correlation coefficient are calculated by subtracting 2 from the sample size. Thus, 2 can be added to the degrees of freedom to determine the sample size, which was 8.
3. The .76 is the value of the test statistic found in the sample.
4. The .05 shows what level alpha was set and that a 5% chance of making a Type I error exists.
5. Finally, $p < .05$ reveals that the results landed in the rare zone, and the null hypothesis is rejected.

What conclusion can Dr. Paik draw so far? He has rejected the null hypothesis, which means he is forced to accept the alternative hypothesis, that there is a relationship between these two variables. Being "forced" to accept the alternative hypothesis is not a hardship. Remember, it is almost always the alternative hypothesis that the researcher believes to be true. Dr. Paik should bear in mind that there is a 5% chance that his decision to reject the null hypothesis is wrong. He should state that the alternative hypothesis is *probably* true.

If the alternative hypothesis is true, it doesn't tell much besides the fact that the population value of the correlation (ρ) between *X* and *Y* is not zero. Phrased without the negative, Dr. Paik can conclude that some relationship exists between the two variables in the population, but he doesn't know how much. At this point, how strong ρ is remains unknown, but it is possible to comment on its direction, whether the relationship is direct or inverse:

- If the null hypothesis is rejected and the sign of the *r* is positive, then a researcher would conclude that the association between *X* and *Y* is a positive one. This means that, in general, cases with high scores on *X* have high scores on *Y*, and cases with low scores on *X* have low scores on *Y*. This is a direct relationship.

- If the null hypothesis is rejected and the sign of the r is negative, then a researcher would conclude that the association between X and Y is a negative one. This means that, in general, cases with high scores on X will have low scores on Y, and cases with low scores on X will have high scores on Y. This is an inverse relationship.
- If the null hypothesis is not rejected, then r is not statistically different from zero. This means sufficient evidence does not exist to conclude that there is a relationship between X and Y. One can't say that the relationship isn't a zero relationship, and one can't assert that it is a zero relationship. The researcher is left in limbo.

At this point, Dr. Paik's interpretation would read something like this:

> The relationship between a husband's gender role flexibility and a wife's level of marital satisfaction was examined in a random sample of married couples in a city. The relationship was positive and statistically significant [$r(6) = .76$, $p < .05$]. For couples in this city, there is a direct relationship between these two variables: higher levels of female marital satisfaction are associated with higher levels of male gender role flexibility.

How Big Is the Effect?

It is a good idea to quantify the size of the effect whether the observed value of r was statistically significant or not:

- If the Pearson r was statistically significant, an effect probably exists in the population. The question is how large is the effect? How strong is the relationship between the two variables? Calculating an effect size will answer this question.
- If the Pearson r was not statistically significant, then not enough evidence is available to conclude that an effect exists. However, a Type II error might have been made, and there was a failure to find an effect when one exists. When the null hypothesis was not rejected, if the effect size is meaningful, then the researcher should raise a concern about Type II error.

The effect size used for r has a formal name, the **coefficient of determination,** but it is commonly called r^2. This is the same effect size that was calculated for the between-subjects, one-way ANOVA.

r^2 tells the percentage of variability in one variable that is accounted for by the other variable. The amount of variability that can be explained ranges from 0% to 100%. The closer the size of the effect is to 100%, the stronger it is. The closer the size of the effect is to 0%, the weaker it is.

r^2 can be thought of as indicating how much overlap occurs between what the two variables measure. If $r^2 = 100\%$, then the two variables overlap 100% and measure exactly the same thing (though they are on different scales). This occurs if $r = 1.00$ or $r = -1.00$. For example, the correlation between temperatures measured in Fahrenheit and Celsius is 1.00 and $r^2 = 100\%$. Fahrenheit and Celsius measure the same thing. Saying something's temperature is 212° in Fahrenheit is exactly the same as saying it is 100° Celsius.

An r^2 of 0% would occur if there were no relationship between X and Y, if r were zero. This means that no overlap, none at all, exists between the two variables. An r^2

near zero means the effect is very weak; near 100% means the effect is very strong. Very high r^2 values are rare. Much more common are low to mid-level values. Cohen (1988) provides standards for judging effect sizes for r^2:

- Small $\approx$1%
- Medium $\approx$9%
- Large $\approx$25%

The formula for calculating r^2 is shown in Equation 13.3. To calculate r^2, take the Pearson r, square it, and multiply that number by 100 to turn it into a percentage.

Equation 13.3 Formula for Calculating r^2, the Percentage of Variability in One Variable That Is Accounted for by the Other Variable

$$r^2 = (r)^2 \times 100$$

where r^2 = percentage of variability in one variable that is accounted for by the other variable

r = Pearson r

For the marital satisfaction study, $r = .76$. Here are the calculations for r^2:

$$\begin{aligned} r^2 &= (r)^2 \times 100 \\ &= .76^2 \times 100 \\ &= .5776 \times 100 \\ &= 57.76\% \end{aligned}$$

The two variables in this correlational study, husbands' gender role flexibility and wives' marital satisfaction, explain almost 58% of the variability in each other. This is a large effect, indicating that the two variables are strongly correlated with each other. Remember, correlation is not causation, so a researcher needs to be careful about phrasing the results. The order in which the variables are listed has implications for the conclusion about the order of the relationship. There's a difference between "The husbands' gender role flexibility explains 58% of the variability in the wives' marital satisfaction" and "The wives' marital satisfaction explains 58% of the variability in the husbands' gender role flexibility." The variable that is listed first tends to be perceived as the influential one. Unless one wants to suggest an order to the relationship, it is better to say something like the following: "The two variables, husbands' gender role flexibility and wives' marital satisfaction, explain 58% of the variability in each other."

How Wide Is the Confidence Interval?

Most of the interpreting done so far has been based on r, the correlation coefficient calculated for the data in the *sample*. The task of inferential statistics is to use the sample to draw a conclusion about the larger *population* of cases. A confidence interval allows a researcher to use a sample statistic to estimate the range within which the population parameter probably falls. A confidence interval for Pearson r makes a statement about ρ, the population correlation coefficient, based on r, the sample correlation coefficient.

Calculating the 95% confidence interval for ρ, 95% CIρ, is a three-step procedure:

1. First, transform the observed r value into a z score, z_r. This is called a Fisher's r to z transformation. Appendix Table 7 makes the transformation easy.
2. Then, use z_r along with Equations 13.4 and 13.5 to calculate the confidence interval.
3. This confidence interval is in z_r format, not r format. So, the final step is to use Appendix Table 8 to transform the confidence interval back into r value format.

Step 1 Transform r to z_r

Appendix Table 7, shown in **Table 13.8**, is used to transform r to z_r. The transformation is necessary because the sampling distribution of r becomes less normally distributed as ρ deviates more from zero.

TABLE 13.8 Appendix Table 7: Transformation Table for Fisher's r to z

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
.1	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19
.2	0.20	0.21	0.22	0.23	0.24	0.26	0.27	0.28	0.29	0.30
.3	0.31	0.32	0.33	0.34	0.35	0.37	0.38	0.39	0.40	0.41
.4	0.42	0.44	0.45	0.46	0.47	0.48	0.50	0.51	0.52	0.54
.5	0.55	0.56	0.58	0.59	0.60	0.62	0.63	0.65	0.66	0.68
.6	0.69	0.71	0.73	0.74	0.76	0.78	0.79	0.81	0.83	0.85
.7	0.87	0.89	0.91	0.93	0.95	0.97	1.00	1.02	1.05	1.07
.8	1.10	1.13	1.16	1.19	1.22	1.26	1.29	1.33	1.38	1.42
.9	1.47	1.53	1.59	1.66	1.74	1.83	1.95	2.09	2.30	2.65

In this table, the rows represent the first digit of an r value and the columns represent the second digit of an r value. For example, $r = .76$ is broken down into the row for .7 and the column for .06. The z_r value is found at the intersection of the row and column. An r value of .76 becomes a z_r of 1.00. (Be sure to maintain the sign associated with the r. If r had been −.76, z_r would have been −1.00.)

The rows in Appendix Table 7 represent the first digit of a two-digit r value and the columns represent the second digit of the two-digit r value. For example, an r of .76 is broken down into .7 for the row and .06 for the column. The numbers at the intersections of the row and column are the r value transformed into a z value. Table 13.7 demonstrates this for the Pearson r for the marital satisfaction study where $r = .76$. The intersection of the row for r's that start with .7 and the column for r's that end in .06 gives $z_r = 1.00$. The original r, .76, was positive, so the z_r value is also positive. The r of .76 is transformed into a z_r of 1.00.

Step 2 Calculate the 95% Confidence Interval for the z Value

The formula for calculating the 95% confidence interval around the z_r value is given in Equation 13.4.

Equation 13.4 95% Confidence Interval for z_r

$$95\%\text{CI}_{z_r} = z_r \pm 1.96 s_r$$

where $95\%\text{CI}_{z_r}$ = 95% confidence interval for ρ, expressed in z_r units

z_r = Pearson r transformed into z format, using Appendix Table 7

s_r = standard error of r (Equation 13.5)

This formula says that 1.96 standard errors of r, abbreviated s_r, are added to and subtracted from z_r. So before calculating 95%CIρ, s_r needs to be calculated using Equation 13.5.

Equation 13.5 Standard Error of r

$$s_r = \frac{1}{\sqrt{N-3}}$$

where s_r = standard error of r

N = number of pairs of cases used in calculating the Pearson r

For Dr. Paik's marital satisfaction study, where $N = 8$, s_r is calculated like this:

$$\begin{aligned} s_r &= \frac{1}{\sqrt{N-3}} \\ &= \frac{1}{\sqrt{8-3}} \\ &= \frac{1}{\sqrt{5}} \\ &= \frac{1}{2.2361} \\ &= 0.4472 \\ &= 0.45 \end{aligned}$$

A Common Question

Q Are a 1 and a 3 always used in Equation 13.4 to calculate s_r?

A Yes. They are constants.

The standard error of r from Equation 13.5 can now be used in Equation 13.4 to calculate the 95% confidence interval for ρ in z_r units:

$$\begin{aligned} 95\%\text{CI}_{z_r} &= z_r \pm 1.96 s_r \\ &= 1.00 \pm (1.96 \times 0.45) \\ &= 1.00 \pm 0.8820 \\ &= \text{from } 0.1180 \text{ to } 1.8820 \\ &= \text{from } 0.12 \text{ to } 1.88 \end{aligned}$$

The 95% confidence interval for ρ, expressed in z_r units, ranges from 0.12 to 1.88. All that is left is to transform the confidence interval back into r units.

Step 3 Transform the Confidence Interval from z_r Units to r Values

Transforming from z_r units to r units uses Appendix Table 8, a reversal of Appendix Table 7. Appendix Table 8 is shown in Table 13.9.

TABLE 13.9 Appendix Table 8: Transformation Table for Fisher z to r

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.1	.10	.11	.12	.13	.14	.15	.16	.17	.18	.19
0.2	.20	.21	.22	.23	.24	.24	.25	.26	.27	.28
0.3	.29	.30	.31	.32	.33	.34	.35	.35	.36	.37
0.4	.38	.39	.40	.41	.41	.42	.43	.44	.45	.45
0.5	.46	.47	.48	.49	.49	.50	.51	.52	.52	.53
0.6	.54	.54	.55	.56	.56	.57	.58	.58	.59	.60
0.7	.60	.61	.62	.62	.63	.64	.64	.65	.65	.66
0.8	.66	.67	.68	.68	.69	.69	.70	.70	.71	.71
0.9	.72	.72	.73	.73	.74	.74	.74	.75	.75	.76
1.0	.76	.77	.77	.77	.78	.78	.79	.79	.79	.80
1.1	.80	.80	.81	.81	.81	.82	.82	.82	.83	.83
1.2	.83	.84	.84	.84	.85	.85	.85	.85	.86	.86
1.3	.86	.86	.87	.87	.87	.87	.88	.88	.88	.88
1.4	.89	.89	.89	.89	.89	.90	.90	.90	.90	.90
1.5	.91	.91	.91	.91	.91	.91	.92	.92	.92	.92
1.6	.92	.92	.92	.93	.93	.93	.93	.93	.93	.93
1.7	.94	.94	.94	.94	.94	.94	.94	.94	.94	.95
1.8	.95	.95	.95	.95	.95	.95	.95	.95	.95	.96
1.9	.96	.96	.96	.96	.96	.96	.96	.96	.96	.96
2.0	.96	.96	.97	.97	.97	.97	.97	.97	.97	.97
2.1	.97	.97	.97	.97	.97	.97	.97	.97	.97	.98
2.2	.98	.98	.98	.98	.98	.98	.98	.98	.98	.98
2.3	.98	.98	.98	.98	.98	.98	.98	.98	.98	.98
2.4	.98	.98	.98	.98	.98	.99	.99	.99	.99	.99
2.5	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
2.6	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
2.7	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
2.8	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
2.9	.99	.99	.99	.99	.99	.99	.99	.99	.99	.99
3.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The intersection of the row for the first two digits of z (1.8) with the column for the third digit of z (.08) gives the r value into which the z_r value is transformed. $z_r = 1.88$ is transformed into $r = .95$. Be sure to maintain the sign associated with the z value.

Appendix Table 8 is set up the same way as Appendix Table 7. Each row covers the first decimal place of a z_r value and each column is the second decimal place of a z_r value. At the intersection of the row and column, the z_r value is transformed back into an r value. Of course, be sure to maintain the sign. If the z_r was a negative number, then the r value is a negative value as well.

The first z_r value for the marital satisfaction study, the one at the lower end of the confidence interval, was 0.12. Looking at the intersection of the 0.1 row with the .02 column in the z to r table, Dr. Paik finds $r = .12$. The second z_r value for the marital satisfaction study was 1.88. The intersection of the 1.8 row with the .08 column gives $r = .95$. The 95% confidence interval for ρ, the population value of the relationship between male gender role flexibility and female marital satisfaction, ranges from .12 to .95 when expressed in Pearson correlation coefficient units.

The observed r in the sample was .76, which is a strong and positive correlation. But, the confidence interval indicates uncertainty about the strength of the relationship between gender role flexibility and marital satisfaction in the population. ρ could be (1) an almost perfect correlation coefficient of .95, or (2) a correlation of .12 that is much closer to zero, or (3) a correlation anywhere between those two extremes. And, there's a fourth option, a 5% chance that this confidence interval doesn't capture ρ. A very wide confidence interval, like this one ranging from .12 to .95, tells a researcher that there is a lot of uncertainty about how strong the correlation is in the underlying population.

To interpret a confidence interval for ρ, pay attention to three factors:

1. Whether the confidence interval captures zero. Capturing zero means it is possible that there is no relationship in the population between the two variables. In other words, it is possible that $\rho = 0$, just like the null hypothesis said. The confidence interval should capture zero when the null hypothesis is not rejected.
2. How close the confidence interval comes to zero. The closer an end of the confidence interval comes to zero, the weaker the relationship may be in the population.
3. How wide the confidence interval is. The wider the confidence interval is, the less sure a researcher is of the population value of the correlation coefficient.

For Dr. Paik's marital satisfaction study, the results were statistically significant and the null hypothesis was rejected. As a result, the confidence interval for ρ should not capture zero and that's exactly what happened. There's little reason, beyond the possibility of a Type I error, to think $\rho = 0$. The whole confidence interval, from .12 to .95, is on the positive side of zero. This leaves the researcher confident that the relationship between gender role flexibility and marital satisfaction is a direct one.

The low end of the confidence interval is .12. Using Equation 13.4 to calculate r^2 for this correlation coefficient, Dr. Paik finds $r^2 = 1.44\%$. This is a small effect, which tells Dr. Paik that the size of the effect in the overall population might also be small. On the other hand, if $\rho = .95$, the other end of the confidence interval, then $r^2 = 90.25\%$, which represents a huge effect.

Dr. Paik is bothered by the width of the confidence interval. It ranges from near 0 (.12) to near 1 (.95). That's a wide range. This lack of precision in the confidence interval makes it unclear how much association exists between role flexibility and marital satisfaction in the population. The relationship in the population from which this sample came could be near zero or near perfect.

The reason the confidence interval is so wide is that the sample size was small. If *N* had been 80 (not 8), then s_r would have been 0.11 (not 0.45), and the 95% confidence interval would have been a much more precise .65 to .84, with the low end much further away from zero. Such a confidence interval would inspire more confidence that, as a result of this study, something was known about the relationship between these two variables in the larger population.

Putting It All Together

Before writing the interpretation, Dr. Paik reviews the scatterplot made in Step 2 when evaluating the assumptions. Seen in Figure 13.14, it is a graphic display of the relationship between gender role flexibility (X) and marital satisfaction (Y). Scatterplots help one think about the strength and direction of relationships. The scatterplot here shows a direct and strong relationship—higher levels of gender role flexibility are associated with more marital satisfaction. Below is Dr. Paik's four-point interpretation in which he (1) tells what the study was about, (2) indicates its main results, (3) explains what they mean, and (4) makes suggestions for future research:

> In this study, the relationship between a husband's flexibility in his gender role and a wife's degree of marital satisfaction was assessed in a random sample of eight married couples from one city. There was a statistically significant, direct relationship [$r(6) = .76$, $p < .05$] between the two variables: couples with a high score on one variable tended to have a high score on the other. Unfortunately, the sample size in the present study was small, making it impossible to say how strong the relationship between these two variables is in the larger population. The study should be replicated with a larger sample size in order to better determine the strength of the relationship.
>
> In addition, this is a correlational study, so it does not address whether (1) more gender flexible husbands lead to more satisfied wives, or (2) wives who are more satisfied give their husbands the leeway to be more gender flexible, or (3) that a third variable—such as age, education level, or socioeconomic status—could influence both marital satisfaction and gender role flexibility. Future research should attempt to determine the order of the relationship.

Worked Example 13.2 For practice interpreting a Pearson correlation coefficient, let's return to Dr. Solomon's study that investigated the relationship between the age at which children begin walking and their later IQ. In that study, ten 16-year-old girls were randomly selected from a pediatrician's practice. They were given IQ tests and their charts were examined to determine the age at which they started to walk. Dr. Solomon used a Pearson correlation coefficient to examine the relationship between the two variables and found $r = -.44$.

Was the null hypothesis rejected? The null hypothesis said that there was no relationship between the two variables. With $df = 8$, $\alpha = .05$, and a two-tailed test, the critical value of r was $\pm.632$. Here is the decision rule:

- If $-.44 \leq -.632$ *or* if $-.44 \geq .632$, reject H_0.
- If $-.632 < -.44 < .632$, fail to reject H_0.

The second statement is true, so Dr. Solomon has failed to reject the null hypothesis. As shown in Figure 13.21, the observed r of $-.44$ falls between $-.632$ and .632, which is in the common zone of the sampling distribution of r. That means $-.44$ is a result that will commonly occur if the null hypothesis is true. There is no reason to reject the null hypothesis.

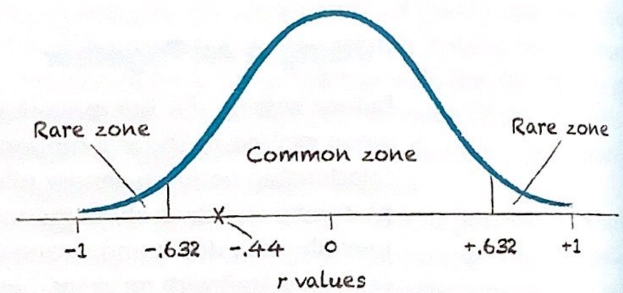


Figure 13.21 Implementing the Decision Rule for Dr. Solomon's Age of Walking/IQ Study The observed r of $-.44$ falls in the common zone of the sampling distribution of r. It is a common result when the null hypothesis is true, so the null hypothesis is not rejected.

With the observed r in the common zone, the null hypothesis is not rejected, and the results are called "not statistically significant." Dr. Solomon has to state that the observed r of $-.44$ is not statistically different from zero. When she calculates the confidence interval for ρ, she should find that it captures zero.

When a researcher fails to reject the null hypothesis, he or she can't speculate about the direction of the relationship because there isn't enough evidence to say one exists.

The null hypothesis wasn't rejected, so Dr. Solomon can't draw a conclusion about the direction of the relationship. The observed r, $-.44$, may be negative, but she can't say that there is a negative relationship between age at first walking and IQ in the population. That is, she can't conclude that people who start walking at a younger age have higher IQs at age 16. When a researcher fails to reject the null hypothesis, he or she can't speculate about the direction of the relationship because there isn't enough evidence to say a relationship exists.

In APA format, Dr. Solomon reports the results as

$$r(8) = -.44, p > .05$$

- The "r" tells what test was done, a Pearson r.
- The "8" is the degrees of freedom. Adding 2 to it gives the number of participants: 10.
- "$-.44$" is the observed value of the test statistic, Pearson r.
- ".05" indicates the alpha level.
- "$p > .05$" reveals that results like this occur more than 5% of the time when the null hypothesis is true. That is, the null hypothesis was not rejected.

How big is the effect? Though Dr. Solomon failed to reject the null hypothesis, she should still calculate an effect size to help her consider the possibility of Type II error.

Using Equation 13.3, she calculates the effect size, r^2:

$$\begin{aligned} r^2 &= (r)^2 \times 100 \\ &= -.44^2 \times 100 \\ &= 0.1936 \times 100 \\ &= 19.36\% \end{aligned}$$

An r^2 of 19.36% is more than a medium effect, according to Cohen (1988). As a result, Dr. Solomon is worried that a Type II error may have been made and that she missed finding a relationship between age of walking and IQ. She is going to recommend replicating the study with a larger sample size in order to have a better chance of finding the effect if it exists.

How wide is the confidence interval? Calculating the confidence interval for ρ is a three-step process.

Step 1 Convert r to z_r. To do this, Dr. Solomon uses Appendix Table 7. Given $r = -.44$, she finds the z_r transformation for this at the intersection of the row for .4 and the column for .04. The r was negative, so the z_r should be, too: $z_r = -0.47$.

Step 2 Calculate the Confidence Interval in z_r Units. The first thing Dr. Solomon does in this step is use Equation 13.5 to calculate s_r. The only variable to plug into the equation is the sample size, which is 10:

$$\begin{aligned} s_r &= \frac{1}{\sqrt{N-3}} \\ &= \frac{1}{\sqrt{10-3}} \\ &= \frac{1}{\sqrt{7}} \\ &= \frac{1}{2.6458} \\ &= 0.3780 \\ &= 0.38 \end{aligned}$$

Now that she knows $s_r = 0.38$ and $z_r = -0.47$, she can use Equation 13.4 to calculate the confidence interval in z_r units:

$$\begin{aligned} 95\%\text{CI}_{z_r} &= z_r \pm 1.96 s_r \\ &= -0.47 \pm (1.96 \times 0.38) \\ &= -0.47 \pm 0.7448 \\ &= \text{from } -1.2148 \text{ to } 0.2748 \\ &= [-1.21, 0.27] \end{aligned}$$

Step 3 Convert the Confidence Interval Back into r Units. The confidence interval, in z_r units, ranges from −1.21 to 0.27. In order to interpret the confidence interval, Dr. Solomon needs to convert it back to Pearson correlation coefficient units using Appendix Table 8. She will maintain the sign of the values, positive or negative, in the conversion process.

- At the intersection of the row for 1.2 and the column for .01, she transforms a z_r of −1.21 into $r = -.84$.
- At the intersection of the row for 0.2 and the column for .07, she transforms a z_r of .27 into $r = .26$.

- Her 95% confidence interval for the population value of the correlation between age at first walking and IQ at age 16 is [−.84, .26].

There are several things Dr. Solomon can note about this confidence interval:

- The confidence interval captures zero. This was expected because the null hypothesis was not rejected. The confidence interval capturing zero means it is possible, in the population, that there is a zero relationship between age at first walking and IQ.
- The confidence interval doesn't just come close to zero, it includes zero. This means that the relationship might not just be small: it might be nonexistent.
- The confidence interval is very wide. It goes all the way from far below zero, −.84, to a modest distance above zero, .26. With such a wide range, the population value is uncertain. Is the population value positive, zero, or negative? Is it a weak relationship or a strong one? With a range from −.84 to .26, one just can't tell. The study should be replicated with a larger sample size in order to obtain a more precise estimate of ρ.

Step 4 Was There Adequate Power? How Likely Was a Type II Error? Dr. Solomon is already concerned that a Type II error might have been made. That is, she's concerned the study might've been underpowered. Remember, power is the probability that a null hypothesis that should be rejected is rejected. Statisticians like power to be at least .80 to be considered adequate. That is, if the null hypothesis should be rejected, they want an 80% chance of being able to do so.

Power is the flipside to Type II error, it is heads to error's tails. If the null hypothesis should be rejected, then there are only two options: it is rejected or it is not rejected. If there is an 80% chance of rejecting it, then there's a 20% chance of failing to reject it. Beta, β, is the probability of Type II error. As discussed in Chapter 6, power + $\beta = 1.00$, so if power = .80, then $\beta = .20$.

Appendix Table 9 makes it easy to find power for a Pearson r. Power depends on the number of cases and the strength of the correlation. The rows in the table are for different sample sizes and the columns offer different correlations. Dr. Solomon had 10 cases and found $r = -.44$. For both rows and columns, follow the *Price Is Right* rule and select the option that is closest without going over. The sign of the correlation does not matter. So, Dr. Solomon looks at the intersection of the row with 10 cases and the column for $r = .40$, where she finds power = .20. As she would like power to be .80 and it is only .20, this study is quite underpowered.

With power of .20, $\beta = .80$. If there really is a relationship between age of walking and intelligence and Dr. Solomon tries to find it with only 10 cases, then there's an 80% chance the results will not find evidence of a relationship. Those are not very good odds.

Appendix Table 9 can also be used to see how many cases Dr. Solomon would need to have an 80% chance of finding a relationship if the relationship had the strength of $r = .40$. She would simply go down the column until she reaches power of .80 and then look to the left to see how many cases would be needed. Dr. Solomon should have used 48 cases.

Putting it all together. Finally, Dr. Solomon is ready to interpret results:

This study investigated if a relationship existed between the age at which children started walking and their IQs at age 16. Using a random sample of 10 girls from a pediatrician's practice, there wasn't enough evidence to conclude that the age at which a child begins to walk provides any information about intelligence [$r(8) = -.44, p > .05$]. However, the sample size was small, so a relationship may exist that the researcher failed to find. Future research should replicate the study with a sample size of at least 48 participants to increase the chances of finding a relationship if one does exist. The sample should also include boys and be drawn from multiple sites.

A Common Question

Q A lot of interpretations in this book suggest replication with a larger sample size in order to get a better idea if an effect exists and/or how strong it is. Is this really a common recommendation in studies?

A It's not as common as in this book. In order to make the math easy in the examples, small sample sizes are used. Most real studies use larger sample sizes, making it easier to reject the null hypothesis when it is false and yielding narrower confidence intervals.

Practice Problems 13.3

Apply Your Knowledge

13.13 If $N = 22$, $\alpha = .05$, $r_{cv} = .444$, and $r = .63$, (a) state whether the null hypothesis was rejected, (b) indicate whether the results are statistically significant, and (c) report the results in APA format.

13.14 (a) If $r = -.34$, what is r^2? (b) Is this a small, medium, or large effect?

13.15 Given $N = 12$ and $r = .40$, calculate the 95% confidence interval for ρ.

13.16 If power = .72, what is β?

13.17 Given $N = 33$ and $r = .37$, what is power?

13.18 If $r(19) = .36$, $p > .05$, how likely is Type II error?

13.19 If a researcher thinks $r = .30$, how many cases would he need to have in order to have power of .80? Of .95?

13.20 A clinical psychologist randomly sampled 402 parents-to-be in the United States and administered to each couple an interval-level measure of marital harmony. She tracked down all of the children 18 years later and to each child administered an interval-level measure of mental health. Both scales were scored so that higher scores indicated more marital harmony or more mental health. She found $r(400) = .38$, $p < .05$. Given 95% CI [.29, .36] and $r^2 = 14.44\%$, write a four-point interpretation.

13.4 Calculating a Partial Correlation

We started this chapter with a scatterplot showing that there was an association between the number of community hospitals per state and the number of deaths per state. That correlation did not mean that having hospitals caused the deaths. Rather, it seemed more plausible that a confounding variable, the population of a state, caused both the number of hospitals and the number of deaths. Now, we are going to learn about a technique called *partial correlation* that can be used to quantify objectively if and how a confounding variable exerts influence.

A confounding variable is a third variable, let's call it Z, that is not measured and/or not controlled, that correlates with both X and Y, and that potentially explains why X and Y are correlated. Here is an example. Suppose we give \$10 each to ten 21-year-old college students on a Friday night and ask them to rendezvous with us at midnight. When they return, we find out how much money they spent and measure how well they can walk a straight line. We find a relationship—the more money a student has spent, the worse he or she is at walking the line. Clearly, a third variable, the amount of alcohol a student has purchased, would explain both why his or her funds are depleted and why his or her performance is impaired. If we took into account each student's alcohol consumption, the correlation between money spent and performance would be explained.

A **partial correlation** mathematically removes the influence of a third variable on a correlation. It is a useful technique that moves a correlational study a little bit in the direction of an experimental study. Partial correlations do not allow cause-and-effect conclusions to be drawn, but they can make potential cause-and-effect conclusions less or more plausible. In the student study above, if the results turned out as described, it would be less plausible to conclude that spending money is the cause of impaired performance.

How can a partial correlation make a cause-and-effect conclusion more plausible? Suppose a researcher finds a relationship between the number of years a person has smoked cigarettes and the degree of impairment of lung function. This researcher claims that the impairment is caused by smoking. Another researcher points out that people who have been smoking longer are usually older, and posits that it is age that causes the impaired lung function. If the effect of age is removed and the correlation between years of smoking and impaired lung function remains strong, then the notion that it is the number of years of smoking that leads to impaired lung function remains plausible.

The abbreviation for a partial correlation is r_{XY-Z}, the correlation between X and Y minus the influence of Z. The formula for a partial correlation, shown in Equation 13.6, makes use of three pieces of information: the correlation between X and Y, the correlation between X and Z, and the correlation between Z and Y.

Equation 13.6 Partial Correlation of X with Y, Controlling for Z

$$r_{XY-Z} = \frac{r_{XY} - (r_{XZ} \times r_{YZ})}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

where r_{XY-Z} = the partial correlation of X and Y controlling for the influence of Z

r_{XY} = the correlation between X and Y

r_{XZ} = the correlation between X and Z

r_{YZ} = the correlation between Y and Z

Let's use a partial correlation to control for the effect of population on the correlation between the number of community hospitals per state and the number of deaths per state. The scatterplot in Figure 13.1 at the start of this chapter shows a strong and direct relationship. Not surprisingly, the correlation is strong and significant: $r(23) = .89$, $p < .05$. But, the correlation between the number of hospitals (X) and the state population (Z) was .92, and for the number of deaths (Y) and the population, it was .98.

Using $r_{XY} = .89$, $r_{XZ} = .92$, and $r_{YZ} = .98$, we can complete Equation 13.6:

$$\begin{aligned}
r_{XY-Z} &= \frac{r_{XY} - (r_{XZ} \times r_{YZ})}{\sqrt{(1 - r^2_{XZ})(1 - r^2_{YZ})}} \\
&= \frac{.89 - (.92 \times .98)}{\sqrt{(1 - .92^2)(1 - .98^2)}} \\
&= \frac{.89 - .9016}{\sqrt{1 - .9604)(1 - .8464)}} \\
&= \frac{-.0116}{\sqrt{.0396 \times .1536}} \\
&= \frac{-.0116}{\sqrt{.0061}} \\
&= \frac{-.0116}{.0781} \\
&= -.1485 \\
&= -.15
\end{aligned}$$

The correlation of the number of hospitals with the number of deaths was .89. But, when population was controlled for, the correlation fell to −.15. It went from statistically significant to failing to reject the null hypothesis. This means that there is not enough evidence to suggest the number of community hospitals has any causal impact on the number of deaths in a given state—the apparent association between the two variables can be explained by the states' populations.

A Common Question

Q How many degrees of freedom does a partial correlation have?

A *df* for a partial correlation is $N - 3$. For the community hospital study, there were 25 states, so $df = 25 - 3 = 22$. The critical value of r is $\pm .396$; thus, the observed r of $-.15$ falls in the common zone. The results would be reported as $r(22) = -.15$, $p > .05$.

Worked Example 13.3 A researcher gathered a random sample of 160 students at her college and measured their physical and mental health on two dimensions: the distance that they are able to run in 30 minutes and happiness, as indicated on a survey. She theorized, from a biopsychosocial perspective, that if there is a mind–body connection, then a relationship should exist between physical health and psychological health.

She found in her sample that the correlation between the two variables was .37 [$r(158) = .37, p < .05$], suggesting that a correlation does indeed exist between physical health and mental health.

She tells a colleague, an obesity researcher, about these findings, and the colleague suggests that the observed relationship could be explained by the body mass index (BMI)—people with high BMIs are in poorer mental and physical health. Luckily, the biopsychosocial researcher had recorded the heights and weights of her subjects, so she calculates BMIs and finds that the correlation between physical health and BMI is .42; for BMI and emotional health, it is .18. Using these values, she calculates a partial correlation:

$$r_{XY-Z} = \frac{r_{XY} - (r_{XZ} \times r_{YZ})}{\sqrt{(1 - r^2_{XZ})(1 - r^2_{YZ})}}$$

$$= \frac{.37 - (.42 \times .18)}{\sqrt{(1 - .42^2)(1 - .18^2)}}$$

$$= \frac{.37 - .0756}{\sqrt{(1 - .1764)(1 - .0324)}}$$

$$= \frac{.2944}{\sqrt{.8236 \times .9676}}$$

$$= \frac{.2944}{\sqrt{.7969}}$$

$$= \frac{.2944}{.8927}$$

$$= .3298$$

$$= .33$$

The partial correlation of .33 is reduced slightly from .37, but it is still statistically significant: $r(157) = .33, p < .05$. The relationship between physical and emotional health is not explained by BMI.

Practice Problems 13.4

13.21 Using $N = 54$, $r_{XY} = .53$, $r_{XZ} = .46$, and $r_{YZ} = .25$, calculate r_{XY-Z}.

13.22 Using $N = 27$, $r_{XY} = .35$, $r_{XZ} = .18$, and $r_{YZ} = .17$, calculate r_{XY-Z}.

Application Demonstration

To see how the Pearson r is used in real research, let's examine a study that involves both psychology and political science. A researcher wanted to know if living in a nation with better government services was associated with greater happiness. He used the average happiness ratings reported by citizens of 130 nations (Ott, 2011). Happiness ratings by individuals could range from 0 ("Right now I am living the worst possible life") to 10 ("Right now I am living the best possible life").

The individual ratings were averaged together to yield a score for a country. The country with the greatest average happiness was Denmark ($M = 8.00$) and the lowest average happiness was Togo ($M = 3.24$); the United States was near the top ($M = 7.26$).

The researcher also developed a measure of government services for each of the 130 nations. Nations that provided better public services had better regulations, experienced less corruption, etc., received higher scores on government services.

The scatterplot showing the relationship between the two variables can be seen in Figure 13.22. The r value for these data is .75, and the results are statistically significant [$r(128) = .75$, $p < .05$]. There is a positive relationship between the two variables in these countries—better government services are associated with happier citizens. Squaring the correlation gives the effect size, $r^2 = 56.25\%$. According to Cohen (1988), this is a large effect.

The 95% confidence interval for ρ ranges from .66 to .82. Note that (1) the confidence interval doesn't capture zero, (2) the bottom end (.66) is far from zero, and (3) the confidence interval is quite narrow. One can conclude that there is

Figure 13.22 Relationship Between the Quality of a Nation's Governments and the Happiness of the People This scatterplot suggests a positive relationship between the quality of government and the happiness of the people. Data from Ott, 2011.

probably a strong, positive correlation between these two variables in the larger population of countries.

The researcher thought of government services as predicting, or leading to, happiness. But, do better government services *cause* more happiness? This is a correlational study where nothing is manipulated. All that one can be sure of is that there is an *association* between better government services (X) and more happiness (Y). Correlation is not causation, so we shouldn't draw conclusions about cause and effect.

When two variables are correlated, three possible explanations exist for the correlation:

1. X causes Y.
2. Y causes X.
3. Some other variable, Z, causes X and Y.

For the happiness/government services data, all three explanations are plausible. First, it is possible that X causes Y, that better government services lead to more happiness. This seems sensible—if government provides better services to its citizens, then their lives should be better and they should be happier.

The second explanation, Y causes X, also seems plausible. Happier citizens could cause better government. It seems possible that happier people will make their government function better because they are more likely to vote, more willing to serve on committees, less likely to disobey laws, and so on.

The third explanation—that Z causes both X and Y—is plausible if one can think of a third variable, Z, that has an impact on both X and Y. Wealth is a third variable that springs to mind. Nations with more wealth might have better government services because they can afford to spend more money on providing them. And, nations with more wealth might have happier citizens because, well, because money does help buy happiness. So, perhaps wealth causes *both* better government services and more happiness.

Untangling cause and effect in a correlational study is a challenge for any researcher. Looking at the scatterplot (Figure 13.22) and the .75 r value, there is clearly an association between the two variables. Just as clearly, the relationship is a direct one. But, how can a researcher put the results in context and explain what they mean? Here's an interpretation that addresses, even embraces, the uncertainty.

A researcher examined the relationship between the quality of government services and happiness in 130 nations around the world. He found that there was a strong, positive association between these two variables [$r(128) = .75, p < .05$]. This means that governments with better services had happier citizens. One explanation for the observed relationship is that countries with better government services provide an environment that leads to more happiness among their citizens. Improving government then should raise up citizens.

But, this study is correlational and other explanations are possible. Perhaps the relationship goes in the other direction. That is, happy citizens lead to better government because happy citizens are more likely to be involved civically and to be concerned for, and supportive of, the welfare of others.

probably a strong, positive correlation between these two variables in the larger population of countries.

The researcher thought of government services as predicting, or leading to, happiness. But, do better government services *cause* more happiness? This is a correlational study where nothing is manipulated. All that one can be sure of is that there is an *association* between better government services (X) and more happiness (Y). Correlation is not causation, so we shouldn't draw conclusions about cause and effect.

When two variables are correlated, three possible explanations exist for the correlation:

1. X causes Y.
2. Y causes X.
3. Some other variable, Z, causes X and Y.

For the happiness/government services data, all three explanations are plausible. First, it is possible that X causes Y, that better government services lead to more happiness. This seems sensible—if government provides better services to its citizens, then their lives should be better and they should be happier.

The second explanation, Y causes X, also seems plausible. Happier citizens could cause better government. It seems possible that happier people will make their government function better because they are more likely to vote, more willing to serve on committees, less likely to disobey laws, and so on.

The third explanation—that Z causes both X and Y—is plausible if one can think of a third variable, Z, that has an impact on both X and Y. Wealth is a third variable that springs to mind. Nations with more wealth might have better government services because they can afford to spend more money on providing them. And, nations with more wealth might have happier citizens because, well, because money does help buy happiness. So, perhaps wealth causes *both* better government services and more happiness.

Untangling cause and effect in a correlational study is a challenge for any researcher. Looking at the scatterplot (Figure 13.22) and the .75 r value, there is clearly an association between the two variables. Just as clearly, the relationship is a direct one. But, how can a researcher put the results in context and explain what they mean? Here's an interpretation that addresses, even embraces, the uncertainty.

A researcher examined the relationship between the quality of government services and happiness in 130 nations around the world. He found that there was a strong, positive association between these two variables [$r(128) = .75$, $p < .05$]. This means that governments with better services had happier citizens. One explanation for the observed relationship is that countries with better government services provide an environment that leads to more happiness among their citizens. Improving government then should raise up citizens.

But, this study is correlational and other explanations are possible. Perhaps the relationship goes in the other direction. That is, happy citizens lead to better government because happy citizens are more likely to be involved civically and to be concerned for, and supportive of, the welfare of others.