Correlation reveals relationship but not necessarily causation.

CHAPTER 13

Correlational Research

INSTRUCTIONAL OBJECTIVES

After studying this chapter, the student will be able to:

1. Describe the nature of correlational research.
2. Describe the ways correlational research is used.
3. Describe the design of correlational research.
4. Discuss the limitations of correlational research.
5. Distinguish between correlational and ex post facto research.
6. List different types of correlational coefficients and state the conditions for their appropriate use.
7. Interpret correlation coefficients in terms of sign, magnitude, statistical significance, and practical significance.
8. Test a hypothesis about a correlation coefficient, r.
9. Define how large a random sample is needed to reject a null hypothesis for a given population correlation.
10. Define predictor and criterion.
11. Define multiple regression and explain when it is used.
12. Define discriminant analysis and explain its purpose.
13. Define factor analysis and explain its purpose.
14. Distinguish exploratory factor analysis and confirmatory factor analysis.
15. Define partial correlation and explain its purpose.
16. Identify studies in which *canonical correlation*, *path analysis*, or *structural equation modeling* would be appropriate.

Correlational research is nonexperimental research that is similar to ex post facto research in that they both employ data derived from preexisting variables. There is no manipulation of the variables in either type of research. They differ in that in ex post facto research, selected variables are used to make comparisons between two or more existing groups, whereas correlational research assesses the relationships among two or more variables in a single group. Ex post facto research investigates possible cause-and-effect relationships;

correlational research typically does not. An advantage of correlational research is that it provides information about the strength of relationships between variables. An ex post facto researcher might define those who make more than $200,000 per year as high earners and those who make less than $40,000 per year as low earners and then compare the mean percent of income paid in taxes for each group. The researcher's data show that the average percent of income paid in taxes by the low earners, at 19 percent, is greater than the 10 percent paid by the high earners. The conclusion is that low-income earners pay a higher percent of their income in taxes than do high-income earners. A correlational researcher would record the income and the percent of income paid in taxes for all people in the study. This researcher might report a correlation coefficient of −.6, indicating a strong negative correlation between the two variables.

Correlational research produces indexes that show both the direction and the strength of relationships among variables, taking into account the entire range of these variables. This index is called a **correlation coefficient**. Recall from Chapter 6 that in interpreting a **coefficient of correlation**, one looks at both its sign and its size. The sign (+ or −) of the coefficient indicates the direction of the relationship. If the coefficient has a positive sign, this means that as one variable increases, the other also increases. For example, the correlation between height and weight is positive because tall people tend to be heavier and short people lighter. A negative coefficient indicates that as one variable increases, the other decreases. The correlation between outdoor air temperature during the winter months and heating bills is negative; as temperature decreases, heating bills rise.

The size of the correlation coefficient indicates the strength of the relationship between the variables. The coefficient can range in value from +1.00 (indicating a perfect positive relationship) through 0 (indicating no relationship) to −1.00 (indicating a perfect negative relationship). A perfect positive relationship means that for every *z*-score unit increase in one variable there is an identical *z*-score unit increase in the other. A perfect negative relationship indicates that for every unit increase in one variable there is an identical unit decrease in the other. Few variables ever show perfect correlation, especially in relating human characteristics.

THINK ABOUT IT 13.1

Interpret each of the following:

1. The correlation between time spent watching television and time spent reading is −.44.
2. The correlation between socioeconomic status and number of museums visited is .21.
3. The correlation between days absent from school and kindergarten reading scores is −.58.

Answers

1. The more time spent watching TV, the less time spent reading; there is a negative relationship.
2. The higher the socioeconomic status, the more museums visited; there is a positive relationship.
3. The more days absent, the lower the reading scores; there is a negative relationship.

USES OF CORRELATIONAL RESEARCH

Correlational research is useful in a wide variety of studies. The most useful applications of correlation are (1) assessing relationships, (2) assessing consistency, and (3) prediction.

ASSESSING RELATIONSHIPS

Correlational research methods are used to assess relationships and patterns of relationship among variables in a single group of subjects. For instance, correlational research is used to answer questions such as the following: Is there a relationship between math aptitude and achievement in computer science? What is the direction and strength of this relationship, if any? You would most likely predict that a positive relationship would be found between scores on a math aptitude test and grades in computer science. A correlational study would determine the extent of any relationship between these variables.

The following are additional examples of questions that could be investigated in a correlational study: What is the relationship between self-esteem and academic achievement? Is there a relationship between musical aptitude and mathematics achievement among 6-year-olds? and What is the relationship between watching media violence and aggression in children?

In some correlational studies, the researcher may be able to state a hypothesis about the expected relationship. For example, from phenomenological theory you might hypothesize that there is a positive relationship between first-grade children's perceptions of themselves and their achievement in reading. In other instances, the researcher may lack the information necessary to state a hypothesis.

Recall from Chapter 9 that the correlation between test scores and selected external variables is a widely used source of evidence in validity studies.

ASSESSING CONSISTENCY

In Chapter 9, we noted that the reliability (consistency) of a test can be assessed through correlating test–retest, equivalent-forms, or split-half scores. Correlation can be used to measure consistency (or lack thereof) in a wide variety of cases. For example, how consistent are the independently assigned merit ratings given by the principal and the assistant principal to teachers in a school? How much agreement is there among Olympic judges rating the performance of a group of gymnasts? When a researcher asks a group of teachers to rank the severity of disruption created by each item on a list of behavior disorders, to what extent do their rankings agree?

PREDICTION

If you find that two variables are correlated, then you can use one variable to predict the other. The higher the correlation, the more accurate the prediction. Prediction studies are frequently used in education. For example, correlational research has shown that high school grades and scholastic aptitude measures are related to college grade point average (GPA). If a student scores high on aptitude tests and has high grades in high school, he or she is more likely to make high grades in college than is a student who scores low on the two predictor variables. Researchers can predict with a certain degree of accuracy a student's probable freshman GPA based on high school grades and aptitude test scores. This prediction will not hold for every case because other factors, such as motivation, initiative, or study habits, are not considered. However, in general, the prediction is good enough to be useful to college admissions officers.

THINK ABOUT IT 13.2

Criticize the conclusions reached in the following examples:

1. The correlation between two variables in an investigation turned out to be negative. The researcher reported that there was no relationship between the variables.
2. A scatterplot showed that the points were all close to a straight line. The researcher concluded that this indicated a positive correlation between the variables.

Answers

1. There is a relationship: The negative correlation means that high scores on one variable are associated with low scores on the other. As one variable increases, the other decreases.
2. It could also indicate a negative correlation. It depends on the direction of the straight line.

DESIGN OF CORRELATIONAL STUDIES

The basic design for correlational research is straightforward. First, the researcher specifies the problem by asking a question about the relationship between the variables of interest. The variables selected for investigation are generally based on a theory, previous research, or the researcher's observations. Because of the potential for spurious results, we do not recommend the "shotgun" approach in which one correlates a number of variables just to see what might show up. The population of interest is also identified at this time. In simple correlational studies, the researcher focuses on gathering data on two (or more) measures from a single group of subjects. For example, you might correlate vocabulary and reading comprehension scores for a group of middle school students. Occasionally, correlational studies investigate relationships between scores on one measure for logically paired groups such as twins, siblings, or husbands and wives. For instance, a researcher might want to study the correlation between the SAT scores of identical twins.

The following is an example of a typical correlational research question: What is the relationship between quantitative ability and achievement in science among high school students? The researcher determines how the constructs, ability and achievement, will be quantified. He or she may already be aware of well-accepted operational definitions of the constructs, may seek definitions in sources such as those described in Chapter 4, or may develop his or her own operational definitions and then assess their reliability and validity. In the example, the researcher may decide that quantitative ability will be defined as scores on the School and College Ability Test, Series III (SCAT III), and science achievement will be defined as scores on the science sections of the Sequential Tests of Educational Progress (STEP III).

You learned in Chapters 8 and 9 that it is important to select or develop measures that are appropriate indicators of the constructs to be investigated, and that it is especially important that these instruments have satisfactory reliability and are valid for measuring the constructs under consideration. In correlation research, the size of a coefficient of correlation is influenced by the adequacy of the measuring instruments for their intended purpose. Instruments that are too easy or too difficult for the participants in a study would not discriminate among them and would result in a smaller correlation coefficient than instruments with appropriate difficulty levels. Studies using instruments with low reliability and questionable validity are unlikely to produce useful results.

Following the selection or development of instruments, the researcher specifies his or her population of interest and draws a random sample from that population. Finally, the researcher collects the quantitative data on the two or more variables for each of the students in the sample and then calculates the coefficient(s) of correlation between the paired scores. Before calculating the coefficient, the researcher should look at a scatterplot or a graph of the relationship between the variables.

CORRELATION COEFFICIENTS

There are many different kinds of correlation coefficients. The researcher chooses the appropriate statistical procedure primarily on the basis of (1) the scale of measurement of the measures used and (2) the number of variables.

PEARSON PRODUCT MOMENT COEFFICIENT OF CORRELATION

In Chapter 6, we introduced you to the Pearson product moment correlation coefficient, symbolized *r*, which is the most widely used descriptive statistic of correlation. Recall that the Pearson coefficient is appropriate for use when the variables to be correlated are normally distributed and measured on an interval or ratio scale. We briefly mention some of the other indexes of correlation without going into their computation. Interested students should consult statistics books for the computational procedures.

COEFFICIENT OF DETERMINATION

Unsophisticated consumers of research often assume that a correlation indicates percentage of relationship, for example, that an *r* of .60 means the two variables are 60 percent related. In fact, *r* is the mean *z*-score product for the two variables, not a percentage. The absolute size of the correlation coefficient (how far it is from zero) indicates the strength of the relationship. Thus, a correlation of −.4 indicates a stronger relationship than a +.2 because it is further from zero. The sign has nothing to do with the strength of the relationship. Another way to see how closely two variables are related is to square the correlation coefficient. When you square the Pearson *r*, you get an index called the **coefficient of determination**, r^2, which tells you how much of the variance of *Y* is in common with the variance of *X*. A correlation of +.60 or −.60 means that the two variables have ($.60^2$) or 36 percent of their variance in common with each other. If the two variables were caffeine and reaction time, then the amount of caffeine one has consumed would be associated with 36 percent of the variance in one's reaction time. That leaves 64 percent of the variance in reaction time associated with factors other than variation in caffeine intake. The notion of common variance is illustrated in

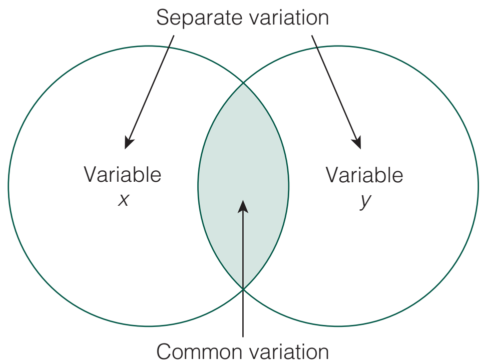

Figure 13.1 The Common Variance between Two Variables

Figure 13.1, in which the total amount of variation in each variable is represented by a circle. The overlap of the circles represents the common variance.

An increase in the *r* results in an accelerating increase in r^2. A correlation of .20 yields a coefficient of determination of .04. An *r* of .4 yields an r^2 of .16. An *r* of .8 yields an r^2 of .64, and so on. The coefficient of determination is a useful index for evaluating the meaning of size of a correlation. It also reminds one that positive and negative correlations of the same magnitude, for example, $r = .5$ and $r = -.5$, are equally useful for prediction and other uses because both have the same coefficient of determination, $r^2 = .25$. The coefficient of determination ranges from 0 to +1.00. If it is 1.00 ($r = +1.00$), you can predict individuals' scores on one variable perfectly from their scores on the other variable.

The Pearson *r* and r^2 are only appropriate where the relationship between *X* and *Y* is *linear*. Linear means that a straight line is a good fit for showing the tilt of the cloud of data in a scattergram. In Chapter 6, several scattergrams are shown. All of them are linear except Figure 6.13. Look at those figures and note the difference between this figure and the rest.

Fortunately, most correlations found in the behavioral sciences are linear. However, before you proceed to calculate and interpret a Pearson *r* for your data, have your computer print out a scatter diagram for your data. If the relationship is not linear, the Pearson *r* is not appropriate for assessing the relationship between variables.

SPEARMAN RHO COEFFICIENT OF CORRELATION

Spearman rho (ρ), an ordinal coefficient of correlation, is used when the data are ranks. For example, assume the principal and assistant principal have independently ranked the 15 teachers in their school from first, most effective, to fifteenth, least effective, and you want to assess how much their ranks agree. You can calculate the Spearman's rho by putting the paired ranks into the Pearson *r* formula or by using a formula developed specifically for rho that is simpler that the Pearson *r* if you are calculating "by hand."

Spearman rho is interpreted in the same way as the Pearson *r*. Like the Pearson product moment coefficient of correlation, it ranges from −1.00 to +1.00. When each individual has the same rank on both variables, the rho correlation will be +1.00, and when their ranks on one variable are exactly the opposite of their ranks on the other variable, rho will be −1.00. If there is no relationship between the rankings, the rank correlation coefficient will be 0.

THE PHI COEFFICIENT

The **phi (φ) coefficient** is used when both variables are genuine dichotomies scored 1 or 0. For example, phi would be used to describe the relationship between the gender of high school students and whether they are counseled to take college preparatory courses or not. Gender is dichotomized as male = 0, female = 1. Being counseled to take college preparatory courses is scored 1, and not being so counseled is scored 0. It is possible to enter the pairs of dichotomous scores (1's and 0's) into a program that computes Pearson *r*'s and arrive at the phi coefficient.

If you find the phi coefficient in school A is −.15, it indicates that there is a slight tendency to counsel more boys than girls to take college preparatory courses. If in school B the phi coefficient is −.51, it indicates a strong tendency in the same direction. As with the other correlations, the phi coefficient indicates both direction and strength of relationships.

A variety of correlation coefficients are available for use with ordinal and nominal data. These include coefficients for data that are more than just pairs; for example, assessing the agreement of three or more judges ranking the performance of the same subjects. We highly recommend Siegel and Castellan's *Nonparametric Statistics* (1988). We consider it a remarkably well organized and easy to understand text.

CONSIDERATIONS FOR INTERPRETING A CORRELATION COEFFICIENT

The coefficient of correlation may be simple to calculate, but it can be tricky to interpret. It is probably one of the most misinterpreted and/or overinterpreted statistics available to researchers. Various considerations need to be taken into account when evaluating the practical utility of a correlation. The importance of the numerical value of a particular correlation may be evaluated in four ways: (1) considering the nature of its population and the shape of its distribution, (2) its relation to other correlations of the same or similar variables, (3) according to its absolute size and its predictive validity, or (4) in terms of its statistical significance.

THE NATURE OF THE POPULATION AND THE SHAPE OF ITS DISTRIBUTION

The value of an observed correlation is influenced by the characteristics of the population in which it is observed. For example, a mathematics aptitude test that has a high correlation with subsequent math grades in a regular class where students range widely on both variables would have a low correlation in a gifted class. This is because the math aptitude scores in the gifted class are range restricted (less spread out) compared to those in a regular class. Range restrictions of either the predictor or the criterion scores reduce the strength of the observed correlation. Before proceeding to interpret your correlation results, produce a scattergram to determine if you have range restriction problem.

Also, if your population differs from the population in which a correlation was reported, that correlation only provides an estimate of correlation in your population of interest. The more your population differs from the original population, the less useful the estimate becomes.

In planning a correlational study, if you think variables such as home language or gender will influence your correlation of interest, you can draw random samples of equal numbers from each subgroup to assess the influence of these variables.

COMPARISON TO OTHER CORRELATIONS

A useful correlation is one that is higher (in either direction) than other correlations of the same or similar variables. For example, an *r* of .75 would be considered low for the relationship between the results of two equivalent forms of an achievement test because equivalent forms of most achievement tests correlate with each other by more than .90. A correlation of .80 between a measure of academic aptitude and GPA of middle school students would be considered high because the correlation for other measures of academic aptitude and GPA for this population is typically approximately .70.

As we have previously stated, a measure that can be used with high school seniors that correlates .60 with their subsequent college freshman GPA would be excellent because currently available measures correlate between .40 and .45 with college GPAs.

PRACTICAL UTILITY

Always consider the practical significance of the correlation coefficient. Although a correlation coefficient may be statistically significant, it may have little practical utility. With a sample of 1000, a very small coefficient such as .08 would be statistically significant at the .01 level. But of what practical importance would this correlation be? Information on *X* only accounts for less than 1 percent ($.08^2 = .0064$ or 00.64 percent) of the variance in *Y* (r^2). In this case, it would hardly be worth the bother of collecting scores on a predictor variable, *X*, to predict another variable, *Y*. You want to avoid the **significance fallacy**—the assumption that a statistically significant correlation also has practical significance. Statistical significance alone is not sufficient. How worthwhile a correlation may be is partly a function of its predictive utility in relation to the cost of obtaining predictor data. A **predictor** with a high correlation that is difficult and expensive to obtain may be of less practical value than a cheap and easy predictor with a lower correlation. Also, note that a correlation coefficient only describes the degree of relationship between given operational definitions of predictor and predicted variables in a particular research situation for a given sample of subjects. It can easily change in value if the same variables are measured and correlated using different operational definitions and/or a different sample.

Failure to find a statistically significant relationship between two variables in one study does not necessarily mean there is *no* relationship between the variables. It only means that in that particular study, sufficient evidence for a relationship was not found. Recall from Chapter 6 that other factors, such as reliability of the measures used and range of possible values on the measures, influence the size of a correlation coefficient.

STATISTICAL SIGNIFICANCE

In evaluating the size of a correlation, it is important to consider the size of the sample on which the correlation is based. Without knowing the sample size, you do not know if the correlation could easily have occurred merely as a result of chance or is likely to be an indication of a genuine relationship. If there were fewer than 20 cases in the sample (which we would not recommend), then a "high" *r* of .50 could easily occur by chance. You should be very careful in attaching too much importance to large correlations when small sample sizes are involved; an *r* found in a small sample does not necessarily mean that a correlation exists in the population.

To avoid the error of inferring a relationship in the population that does not really exist, the researcher should state the null hypothesis that the population correlation equals 0 (H_0: $\rho_{xy} = 0$) and then determine whether the obtained sample correlation departs sufficiently from 0 to justify the rejection of the null hypothesis. In Chapter 7, we showed you how to use Table A.3 in the Appendix, which lists critical values of *r* for different numbers of degrees of freedom (*df*). By comparing the obtained *r* with the critical values of *r* listed in the table, you can determine the statistical significance of a product moment correlation. For example, assume a correlational study involving the paired math and spelling test scores of 92 students yields a correlation of .45.

Recall that for the Pearson *r* the degrees of freedom are the number of paired scores minus 2 ($n - 2$). In Table A.3, we find that for 90 degrees of freedom, $r = .2050$ or greater is statistically significant at the .05 level of significance, .2673 or greater is statistically significant at the .01 level, and .3375 or greater is statistically significant at the .001 level (all two tailed). Therefore, the hypothesis that the population correlation is zero can be rejected at the .01 level and even at the .001 level; therefore, you conclude that there is a positive relationship between math and spelling scores.

DETERMINING SAMPLE SIZE

The Pearson product moment correlation is a form of effect size. Therefore, Table A.3 in the Appendix can be used to determine the needed sample size for a predetermined level of significance and predetermined tolerable probability of Type I error.

For example, a researcher developed a measure of how much a person is willing to sacrifice to achieve success and found it had very satisfactory reliability when administered to high school seniors. He thinks it may be a useful predictor of success in college. Since previous research has shown that the predictor variables high school GPA, ACT test scores, and CEEB test scores of high school seniors all correlate around .40 with the criterion variable college freshmen GPAs, the researcher decides that if his scores correlate .40 or higher with college GPAs it is worth further investigation. If it is less than .40, it is not worth pursuing. He sets his desired level of significance at the two-tailed .01 level. You see in Table A.3 that if the true population correlation is .3932 or greater with 40 degrees of freedom, then $40 + 2$ subjects randomly selected from that population are needed to reject the null hypotheses that the population correlation is zero.

Recall from Chapter 7 that the larger the sample, the more likely the sample statistics are to approximate the population parameters. Note that this is true only when generalizing results from a random sample back to the population from which it was drawn. If the researcher drew the sample from high school seniors in Peoria, Illinois, he could only directly apply results to Peoria, Illinois, seniors. The usefulness of the result for predicting scores for a different population depends on how similar that population is to the Peoria senior population.

Before disseminating the results of this study, the researcher should calculate the scores on his sacrifice-for-success test with high school GPA and ACT and CEEB scores. If any or all of these scores correlate highly with the sacrifice-for-success measure, it is largely repeating information already known. Therefore, it is not adding enough to the prediction of college GPA to be worthwhile. If the correlations are low, the sacrifice-for-success scores would be useful for increasing the predictive validity of the combined weighted scores currently in use.

THINK ABOUT IT 13.3

If this population Pearson product moment correlation is .40 or greater, how large should a randomly selected sample be to reject the null hypothesis that the population correlation is zero with a two-tailed test at the .001 level?

Answer

$70 + 2 = 72$

CORRELATION AND CAUSATION

In evaluating a correlational study, one of the most frequent errors is to interpret a correlation as indicating a **cause-and-effect relationship**. Correlation is a necessary but never a sufficient condition for causation. For example, if a significant positive correlation is found between the number of hours of television watched per week and above average body weight among middle school pupils, that does not prove that excessive television watching causes obesity. Recall from Chapter 12 that when the independent variable is not under the investigator's control, alternate explanations must be considered. In this example, reverse causality is plausible. Perhaps the more overweight a child is, the more he or she is inclined to choose television watching instead of physical activities, games, and interacting with peers. The common-cause explanation is also plausible. Perhaps differences in family recreational patterns and lifestyle account for both differences in weight and time spent watching television.

Consider another example. Assume a researcher finds a relationship between measures of self-esteem and academic achievement (grades) for a sample of students. Table 13.1 summarizes the possibilities for interpreting this observed relationship. Any number of factors could act together to lead to both self-esteem and academic achievement: previous academic experiences, parents' education, peer relationships, motivation, and so on.

Table 13.1 Possible Interpretations of a Relationship between Self-esteem and Academic Achievement

Self-esteem	→	Achievement (self-esteem causes achievement)
Achievement	→	Self-esteem (achievement causes self-esteem)
Intelligence	→ →	Self-esteem (a third factor causes both) Achievement
Home environment	→ →	Self-esteem (a third factor causes both) Achievement

Table 13.2 Possible Interpretations of a Relationship between Children's Watching Television Violence and Aggression

Watching TV violence	→	Aggression in children
Aggressive children	→	Choose to watch violent TV shows
Other factors such as home environment	→ →	Both children's watching TV violence and aggression

Let us consider the example of the relationship between the amount of violence children watch on television and their aggression. Most research has shown a relationship between these two variables, which many people assume is causal. However, Table 13.2 shows other explanations for this relationship.

We must stress, however, that correlation can bring evidence to bear for cause and effect. The Surgeon General's warning about the dangers of cigarette smoking is, in part, based on studies that found positive correlations between the number of cigarettes smoked per day and incidence of lung cancer and other maladies. Here, reverse causality (cancer leads to cigarette smoking) is not a credible explanation. Various common-cause hypotheses (e.g., people who live in areas with high air pollution smoke more and have higher cancer rates) have been shown not to be the case.

Although correlational research does not permit one to infer causality, it may generate causal hypotheses that can be investigated through experimental research methods. For example, finding the correlation between smoking and lung cancer led to animal experiments that allowed scientists to infer a causal link between smoking and lung cancer. Because the results of correlational studies on humans agree with the results of experimental studies on animals, the Surgeon General's warning is considered well founded.

PARTIAL CORRELATION

The correlation techniques discussed so far are appropriate for examining the relationship between two variables. In most situations, however, a researcher must deal with more than two variables, and we need procedures that examine the relationship among several variables. **Partial correlation** is a technique used to determine what correlation remains between two variables when the effect of another variable is eliminated. We know that correlation between two variables may occur because both of them are correlated with a third variable. Partial correlation controls for this third variable. For example, assume you are interested in the correlation between vocabulary and problem-solving skills. Both these variables are related to a third variable, chronological age. For example, 12-year-old children have more developed vocabularies than 8-year-old children, and they also have more highly developed problem-solving skills. Scores on vocabulary and problem solving will correlate with each other because both are correlated with chronological age. Partial correlation would be used with such data to obtain a measure of correlation with the effect of age removed. The remaining correlation between two variables when their correlation with a third variable is removed is called a *first-order partial correlation*. Partial correlation may be used to remove the effect of more than one variable. However, because of the difficulty of interpretation, partial correlation involving the elimination of more than one variable is not often used.

MULTIPLE REGRESSION

Multiple regression is a correlational procedure that examines the relationships among several variables. Specifically, this technique enables researchers to find the best possible weighting of two or more independent variables to yield a maximum correlation with a single dependent variable. For example, colleges use data submitted by prospective freshmen to predict first-semester GPA. The predictor values may be scores on the SAT subtests [SAT verbal (SATV) and SAT math (SATM)], along with students' relative high school rank (RHSR). Relative high school rank, found by dividing a student's rank in the high school graduating class by the size of the class, adjusts for the variation in size of graduating classes. Table 13.3 shows the simple correlations between each of the predictors and the criterion.

You can see in Table 13.3 that none of the variables has a very high correlation with freshman GPA; the best single predictor is relative high school rank. However, we can use all three variables in a multiple regression analysis to determine the correlation of the best possible weighted combination of the three predictor variables with GPA. Computer programs produce a **prediction equation** with the **predictor variables** weighted in the appropriate way to yield the highest correlation with GPA and hence the best prediction. The university can use the equation with similar groups of prospective students whose SAT scores and relative high school ranks are known to predict their as yet unknown GPA at the university.

The regression equation would look as follows: $Y' = a + b_1X_1 + b_2X_2 + b_3X_3$, where Y is the predicted score (GPA); a, b_1, b_2, and b_3 are constants provided by the regression analysis; and X_1, X_{12}, and X_3 are the independent variables (RHSR, SATV, and SATM, respectively). Assume a student has the following scores: SATV = 510, SATM = 540, and RHSR = 21. The student's predicted GPA at the university would be 2.97:

$$Y' = 2.0813 + (-.0131)21 + .0014(540) + .0008(510) = 2.97$$

The regression analysis also yields R, the **coefficient of multiple correlation**, which indicates the relationship between the predictor variables in combination and the criterion. If we square R to get the coefficient of determination, we know the amount of variability in the criterion that is due to differences in scores on the predictor variables. For example, if R is .50, then 25 percent of the variability in GPA can be accounted for by the combined SATV, SATM, and RHSR scores.

Table 13.3 Correlations of Each Predictor with the Criterion

	SATV	SATM	RHSR[a]
GPA	0.31	0.39	−0.42

[a]The negative correlation between relative high school rank and GPA is due to the way that rank in class is measured. The highest achiever in the class has a rank of 1 (the lowest number); the lowest achiever in the class has a rank equal to the size of the class (a high number). The students with the lowest *size* rank are predicted to have the highest GPA; hence, the correlation coefficient is negative.

In the development of a multiple regression equation, the variables should be measured on an interval scale. It is possible, however, to put categorical variables such as gender, social class, marital status, political preference, and the like into a prediction equation if they are recoded as binary variables. For instance, if the variable is gender, 1's can be assigned to females and 0's to males. Such recoded variables are referred to in multiple regression as **dummy variables**.

Because the computations are very complex, multiple regression is done by computer. Computer programs are available that provide not only the multiple correlation coefficient (R) and the prediction equation but also the proportion of variance in the criterion accounted for by the combination of predictors (R^2) and tests of statistical significance for the R and for the contribution of each predictor. For further discussion of multiple regression, see Cohen, Cohen, West, and Aiken (2003).

THINK ABOUT IT 13.4

Match the procedure listed in the left column with the definition in the right column:

1. Spearman rho	a. Shows sign and magnitude of correlation between two nominal variables
2. Pearson *r*	b. Shows sign and magnitude of correlation between two ordinal variables
3. Multiple regression	c. Shows sign and magnitude of correlation between two interval variables
4. Phi coefficient	d. Uses a number of independent variables to predict a single dependent variable
5. Eta correlation	e. Used when the relationship between two variables is curvilinear

Answers

1\. b; 2. c; 3. d; 4. a; 5. e

FACTOR ANALYSIS

Factor analysis, or **exploratory factor analysis**, is a family of techniques used to detect patterns in a set of interval-level variables (Spicer, 2005). Factor analysis begins with a table of pairwise correlations (Pearson *r*'s) among all the variables of interest; this table is called a **correlation matrix**. The purpose of the analysis is to try to reduce the set of measured variables to a smaller set of underlying factors that account for the pattern of relationships. The search follows the law of parsimony, which means that the data should be accounted for with the smallest number of factors. This reduction of the number of variables serves to make the data more manageable and interpretable.

There are two types of situations in which factor analysis is typically used. In the first, a researcher is interested in reducing a set of variables to a smaller set. For example, assume a technology company uses 10 different tests to select computer programmers. Factor analysis could be used to identify perhaps four underlying dimensions measured by those 10 tests so that tests of the four dimensions could be used just as effectively in the selection process as the 10 original tests.

The second type of situation is when researchers use factor analysis to determine the characteristics or underlying structure of a measuring instrument such as a measure of intelligence, personality, or attitudes. Assume a researcher has developed a new scale to measure self-esteem and thinks it is unidimensional (measuring one single dimension). If this is true, factor analysis should yield only one factor. In other cases, a researcher may be interested in investigating the nature of the underlying factors in an existing scale. In Chapter 9, we discussed the use of factor analysis in establishing the construct validity of tests.

Let us illustrate factor analysis with a simple example. Imagine that you have scores on six different subscales of an aptitude measure for 300 subjects. The correlations among all the pairs of scores are shown in Table 13.4. Each subject is shown both on the horizontal rows and vertical columns. The Pearson *r* for each pair of variables is shown where the columns and rows intersect.

The question is: Is there a simpler structure underlying these 15 correlations? Table 13.4 shows that all of the subscales are positively correlated, and we assume all of the correlations are statistically significant. The first two subscales (vocabulary and analogies) form a separate subgroup because they correlate .50 with each other but do not correlate with the other subscales. The next two subscales (arithmetic and numerical reasoning) correlate .55 and thus form another subgroup, and likewise the last two subscales (picture completion and block design) have a correlation of .52 with each other but negligible correlations with other subscales. In other words, the pattern of correlations among these variables seems to reflect three underlying factors, which we label verbal, numerical, and spatial.

Table 13.4 Correlations among Subscale Scores

	1	2	3	4	5	6
1. Vocabulary	—	.50	.15	.12	.12	.15
2. Analogies	—	—	.12	.15	.10	.18
3. Arithmetic	—	—	—	.55	.15	.12
4. Numerical reasoning	—	—	—	—	.20	.22
5. Picture completion	—	—	—	—	—	.52
6. Block design	—	—	—	—	—	—

In this simple example, we are dealing with only six variables. In most cases, there would be a greater number, and it would not be so easy to discover the factors by inspection. Thus, researchers turn to factor analysis. The mathematics of factor analysis is beyond the scope of this book. However, basically it involves searching for the clusters of variables that are all correlated with each other. The first cluster identified is called the first factor, and it represents the variables that are most intercorrelated with each other. Then, other factors are identified that account for decreasing amounts of the variance. The factor is represented as a score, which is generated for each subject in the sample. Next, a correlation coefficient is computed between subjects' factor score and their score on the particular variable entered into the factor analysis. This correlation between a variable and a factor is called the **factor loading**. The higher its loading, the more a variable contributes to and defines a particular factor. A factor loading is interpreted like a correlation coefficient: The larger it is (either positive or negative), the stronger the relationship of the variable to the factor. The result of the factor analysis is a factor matrix, which shows the number of important underlying factors and the weight (loading) of each original variable on the resulting factors. The square of the factor loading is the proportion of common variance between the test and the factor. Table 13.5 shows what the hypothetical factor matrix resulting from a factor analysis of the intelligence test in the previous example might look like.

The first two tests load strongly on factor 1; we might call this underlying factor "verbal ability." The next two tests load strongly on factor 2, which we might label "numerical ability." The last two tests load strongly on factor 3, which we might label "spatial ability." Our simple example thus suggests that there were three factors underlying performance on the intelligence test. This procedure did not involve a hypothesis to be tested but, rather, something to be explored. How does one decide on the "correct" number of factors? The first criterion is that all the factors should be interpretable; an uninterpretable factor serves no practical or theoretical function (Spicer, 2005). Second, the factors should account for a satisfactory amount of shared variance in the data. What is "satisfactory" is defined by the researcher. Some writers suggest that the analysis keep extracting factors as long as a factor accounts for at least another 10 percent of the variance. "There is general agreement that overfactoring is preferable to underfactoring" (Spicer, 2005, p. 195).

Table 13.5 Hypothetical Factor Matrix from an Analysis of Scores on Six Subscales[a]

Test Subscale	Factor 1	Factor 2	Factor 3
Vocabulary	**.91**	.40	.30
Analogies	**.87**	.30	.20
Arithmetic	.25	**.90**	.15
Numerical reasoning	.22	**.80**	.10
Picture completion	.15	.10	**.85**
Block design	.09	.05	**.75**

[a]The variables loading most strongly on each factor are set in boldface.

This preceding discussion illustrates exploratory factor analysis (EFA) because a researcher does not test any formal hypotheses about the number of underlying factors. The number of factors is determined empirically rather than being specified a priori. It is distinguished from the more advanced technique called confirmatory factor analysis (CFA), which we describe briefly next.

CONFIRMATORY FACTOR ANALYSIS

Confirmatory factor analysis, like exploratory factor analysis, "is used to examine the relationships between a set of measured variables and a smaller set of factors that might account for them" (Spicer, 2005, p. 199). Confirmatory factor analysis, however, assumes relatively precise advance knowledge and allows a researcher to specify a priori what these relationships might look like and then to test the accuracy of these formal hypotheses.

The first step in CFA is to specify a model made up of a number of hypotheses about the number of underlying factors, whether or not they are correlated, and which variables are expected to load on which factors. The output of CFA allows the researcher to evaluate the factor model overall and at the level of individual variable–factor relationships. The researcher can use CFA and compare different models or factor solutions that might be proposed.

Researchers often use both EFA and CFA in the construction and evaluation of measuring instruments. They begin with EFA and then move to CFA at later stages in the research. CFA is beyond the scope of an introductory text. Interested readers may refer to Pedhazur (2006), Loehlin (2004), or Thompson (2004).

OTHER COMPLEX CORRELATIONAL PROCEDURES

Several more complex techniques are available to investigate correlation of more than two variables. These analyses require more sophistication with statistics than is usually needed in a beginning research course. We briefly describe these techniques and refer interested students to other texts.

Canonical correlation is a generalization of multiple regression that adds more than one dependent variable (criterion) to the prediction equation. For more information on canonical correlation, see Thompson (1984).

Discriminant analysis is a statistical procedure related to multiple regression, but it differs in that the criterion is a categorical variable rather than a continuous one. A good source for this procedure is Huberty (1994).

Structural equation modeling (SEM) is a popular technique used in the analysis of causality. SEM combines confirmatory factor analysis and path analysis to test both a measurement model and a structural model. We refer the reader to Bentler and Chou (1988), Pedhazur (2006), and Loehlin (2004) for further discussion of SEM.

Pedhazur (2006) defines **path analysis** as "a method for studying direct and indirect effects of variables hypothesized as causes of variables treated as effects."

Pedhazur further states, "Path analysis is intended *not* to discover causes but to shed light on the tenability of the causal models a researcher formulates based on knowledge and theoretical considerations" (p. 769).

THINK ABOUT IT 13.5

Match the procedure in the left column with the definition in the right column:

1. Factor analysis	a. Uses multiple independent variables to predict more than one dependent variable
2. Discriminant analysis	b. Reduces a matrix of correlations among variables to a few underlying constructs
3. Canonical correlation	c. Uses a number of variables to predict membership in categories
4. Partial correlation	d. Uses theory to specify and test models of causation among variables
5. Path analysis	e. Determines the relationship between two variables when the effect of another variable is eliminated

Answers

1. b; 2. c; 3. a; 4. e; 5. d

SUMMARY

Correlational research is nonexperimental research that studies the direction and strength of relationships among variables. It gathers data on two or more quantitative variables from the same group of subjects (or from two logically related groups) and then determines the correlation among the variables.

Correlational procedures are widely used in educational and psychological research. They enable researchers to better understand certain phenomena and to make predictions. Correlational designs are often valuable for generating hypotheses that can be further investigated in experimental or ex post facto research. Correlations must be interpreted appropriately. When assessing a correlation coefficient, one must take into account the population from which the sample was drawn, the shape of the distribution, the sample size, and its statistical and practical significance. The most serious error is to interpret correlation per se as an indicator of causation.

A number of different types of correlation coefficients are used with variables that are measured on different types of scales. Multiple regression is used to find the relationship between two or more independent variables and a dependent variable. It yields a prediction equation that the researcher can use later to predict the dependent variable for a new group of subjects, when the researcher has information only on the independent variables.

Some more sophisticated correlational procedures include partial correlation, discriminant analysis, factor analysis, canonical correlation, path analysis, and structural equation modeling. It is important to know the type of research situation in which each of these techniques would be useful.

Table A.3 Critical Values of the Pearson Product Moment Correlation Coefficient

	Level of Significance for a Directional (One-Tailed) Test				
	.05	.025	.01	.005	.0005
	Level of Significance for a Nondirectional (Two-Tailed) Test				
df = *N* − 2	.10	.05	.02	.01	.001
1	.9877	.9969	.9995	.9999	1.0000
2	.9000	.9500	.9800	.9900	.9990
3	.8054	.8783	.9343	.9587	.9912
4	.7293	.8114	.8822	.9172	.9741
5	.6694	.7545	.8329	.8745	.9507
6	.6215	.7067	.7887	.8343	.9249
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8471
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3233	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3578	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

Source: From *Statistical Tables for Biological, Agricultural, and Medical Research,* edited by R. A. Fisher and F. Yates. Pearson Education Limited. Copyright © 1938 by Oliver and Boyd. Reproduced with permission of the publisher.