Summary

This chapter began with a discussion of a claim made in a grocery store tabloid that poor diet causes marital problems. Actually, there was no specific evidence that diet causes the problems—only that diet and marital difficulties are associated. However, the *Enquirer* failed to specify the exact strength of the association. The rest of the chapter was designed to help you be more specific than the *Enquirer* by learning to specify associations with precise mathematical indexes known as *correlation coefficients.*

First, we presented pictures of the association between two variables; these pictures are called *scatter diagrams.* Second, we presented a method for finding a linear equation to describe the relationship between two variables. This regression method uses the data in raw units. The results of regression analysis are two constants: A *slope* describes the degree of relatedness between the variables, and an *intercept* gives the value of the *Y* variable when the *X* variable is 0. When both of the variables are in standardized or *Z* units, the intercept is always 0 and drops out of the equation. In this unique situation, we solve for only one constant, which is *r,* or the *correlation coefficient.*

When using correlational methods, we must take many things into consideration. For example, correlation does not mean the same thing as causation. In the case of the *National Enquirer* article, the observed correlation between diet and problems in marriage may mean that diet causes the personal difficulties. However, it may also mean that marriage problems cause poor eating habits or that some *third variable* causes both diet habits and marital problems. In addition to the difficulties associated with causation, we must always consider the strength of the correlational relationship. The *coefficient of determination* describes the percentage of variation in one variable that is known on the basis of its association with another variable. The *coefficient of alienation* is an index of what is not known from information about the other variable.

A *regression line* is the best-fitting straight line through a set of points in a scatter diagram. The regression line is described by a mathematical index known as the regression equation. The *regression coefficient* is the ratio of covariance to variance and is also known as the slope of the regression line. The regression coefficient describes how much change is expected in the *Y* variable each time the *X* variable increases by one unit. Other concepts discussed were the *intercept,* the *residual* (the difference between the predicted value given by a regression equation and the observed value), and the *standard error of estimate* (the standard deviation of the residuals obtained from the regression equation).

The field of *multivariate analysis* involves a complicated but important set of methods for studying the relationships among many variables. *Multiple regression* is a multivariate method for studying the relationship between one criterion variable and two or more predictor variables. A similar method known as *discriminant analysis* is used to study the relationship between a categorical criterion and two or more predictors. *Factor analysis* is another multivariate method for reducing a large set of variables down to a smaller set of composite variables.

Correlational methods are the most commonly used statistical techniques in the testing field. The concepts presented in this overview will be referred to throughout the rest of this book.

Scanned with CamScanner

APPENDIX 3.1: Calculation of a Regression Equation and a Correlation Coefficient

In this appendix, we consider the relationship between team performance and payroll for teams in baseball's National League. The Data are from the 2016 season and available on the Internet at www.espn.com. The 2016 season was of particular interest to baseball fans because the World Series pitted the Chicago Cubs with a payroll of more than $154 million against the Cleveland Indians with a payroll of a mere $74 million. The Cubs won the Series, raising the question of whether there is a relationship between expenditure and performance of professional baseball teams.

In this example, payroll for National League teams is measured as mean player salary (expressed in millions of dollars) whereas performance is measured by the number of games won[1]. The data are shown in Table 3.5 and summarized in Figure 3.13. Each dot in the figure represents one team. In 2016, there was a positive relationship between payroll and performance. In other words, teams with higher median salaries had better performance. As Figure 3.13 indicates, each increase in expenditure

TABLE 3.5 Games Won and Median Salaries for Teams in Baseball's National

Team	Average Salary (X)	Games Won (Y)	X^2	Y^2	XY
Dodgers	$8.85	91	$78.35	8,281	805.49
Giants	$6.89	87	$47.47	7,569	599.44
Cubs	$6.18	103	$38.23	10,609	636.85
Cardinals	$5.72	86	$32.74	7,396	492.10
Nationals	$5.67	95	$32.10	9,025	538.28
Mets	$5.36	87	$28.68	7,569	465.93
Rockies	$4.51	75	$20.30	5,625	337.94
Pirates	$4.15	78	$17.23	6,084	323.79
Padres	$4.06	68	$16.46	4,624	275.88
Reds	$3.60	68	$12.95	4,624	244.68
Diamondbacks	$3.57	69	$12.75	4,761	246.37
Phillies	$3.36	71	$11.28	5,041	238.50
Marlins	$3.09	79	$9.56	6,241	244.31
Brewers	$2.77	73	$7.68	5,329	202.31
Braves	$2.76	68	$7.62	4,624	187.70
SUM (Σ)	$70.53	1,198	373.42	97,402	5,839.57

[1]In this example we use total payroll divided by 25, which is the number of players allowed on the roster. Some of the clubs pay more than 25 players. This may explain why estimates of average salary differ in different data sets.

Scanned with CamScanner

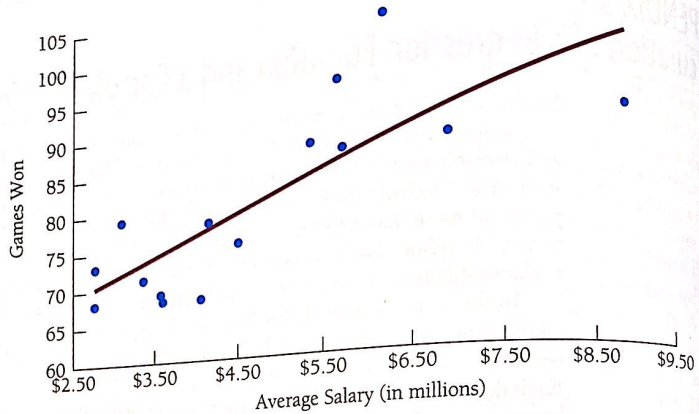

FIGURE 3.13
Average Salary in millions of dollars versus games won by National League baseball teams in 2016.

is associated with an increase in performance. The regression coefficient (4.9 suggests that for each million dollar increase in mean salary, the team's performan increases by an average of 4.93 games per season. We also did the same exerc for total payroll. In 2016 the Los Angeles Dodgers had the highest payroll in t National League, at $221 million. In contrast, the Atlanta Brave had a total payr of $69 million. The correlation between total payroll and games won was 0.77 a the regression equation was:

$$Y = 4.93X + 56.70$$

This tells us that an owner must spend about $5,070,994 (or about $202,840 each of 25 players) to win one game. Overall, the relationship is significant, and best explanation is that there is an association between payroll and performance.

Calculation of a Regression Equation (Data from Table 3.5)

Formulas:

$$b = \frac{N(\Sigma XY) - (\Sigma Y)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}$$

$$a = \overline{Y} - b\overline{X}$$

STEPS

1. Find N by counting the number of pairs of observations. $N = 15$.
2. Find ΣX by summing the X scores.

$$\$8.85 + \$6.89 + \$6.18 + \ldots + \$2.76 = \$70.53$$

3. Find ΣY by summing the Y scores.

$$91 + 87 + 103 + \ldots + 1198$$

4. Find ΣX^2. Square each X score and then sum them.

$$(8.85)^2 + (6.89)^2 + (6.18)^2 + \ldots (2.76)^2 = 373.42$$

Scanned with CamScanner

Summary

Measurement error is common in all fields of science. Psychological and educational specialists, however, have devoted a great deal of time and study to measurement error and its effects. Tests that are relatively free of measurement error are considered to be reliable, and tests that contain relatively large measurement error are considered unreliable. In the early part of the 20th century, Charles Spearman worked out the basics of contemporary theories and methods of reliability. Test score and reliability theories have gone through continual refinements.

When we evaluate *reliability,* we must first specify the source of measurement error we are trying to evaluate. If we are concerned about errors that result from tests being given at different times, then we might consider the *test–retest method* in which test scores obtained at two different points in time are correlated. On other occasions, we may be concerned about errors that arise because we have selected a small sample of items to represent a larger conceptual domain. To evaluate this type of measurement error, we could use a method that assesses the internal consistency of the test such as the *split-half method.* The KR_{20} method and *alpha coefficient* are other methods for estimating the internal consistency of a test.

The standard of reliability for a test depends on the situation in which the test will be used. In some research settings, bringing a test up to an exceptionally high level of reliability may not be worth the extra time and money. On the other hand, strict standards for reliability are required for a test used to make decisions that will affect people's lives. When a test has unacceptably low reliability, the test constructor might wish to boost the reliability by increasing the test length or by using factor analysis to divide the test into homogeneous subgroups of items. In research settings, we can sometimes deal with the problem of low reliability by estimating what the correlation between tests would have been if there had been no measurement error. This procedure is called *correction for attenuation.*

Evaluating the reliability of behavioral observations is also an important challenge. The percentage of items on which observers agree is not the best index of reliability for these studies because it does not take into consideration how much agreement is to be expected by chance alone. Correlation-like indexes such as kappa or phi are better suited to estimate reliability in these behavioral studies.

Reliability is one of the basic foundations of behavioral research. If a test is not reliable, then one cannot demonstrate that it has any meaning. In the next chapter, we focus on how the meaning of tests is defined.

Scanned with CamScanner

Are Scores Really Getting Worse?

In September 2011, newspapers across the nation reported that declining SAT scores mean that the U.S. educational system is losing ground. According to the College Board, SAT reading scores have dropped 3 points, while math scores have dropped 1 point and writing scores 2 points.

What do we make of these declines? Education journalists are quick to cite declining investments in public education. Further, they attribute some of the decline to more poorly prepared teachers, to student apathy, and to a general decline in the academic preparation of American students.

There are other explanations for the decline and, before we even go there, it is worth considering whether there has been, indeed, a measureable decline. First, we must consider the reliability of the SAT components. The Educational Testing Service ETS reports that the reliability of the SAT is quite high. For critical reading, mathematics, and writing, it suggests that reliabilities tend to be about 0.90. However, there may be some problems with the reliability of the scoring of essays. For example, different readers who evaluate the same essay do not always agree. For a group of 2006 college-bound seniors, the interrater reliability of the evaluation of essays ranged from 0.77 to 0.81. When students were asked to write more than one essay, the reliability of the estimates was only 0.67 (Ridgers et al., 2011).

Using the formula for the standard error of measurement, we can calculate the 95% confidence interval around the mean of 489 for the writing component of the SAT. For simplicity, we will use the SD of 100, which is the standard for the SAT.

The standard error of measurement is:

$$S_m = 100\sqrt{1 - .67} = 57.45$$

Using the standard error of measurement, we can say that we are 95% confident that a person's true score falls between two values.

What to Do about Low Reliability

Often, test constructors want their tests to be used in applied settings, but analysis reveals inadequate test reliability. Fortunately, psychometric theory offers some options. Two common approaches are to increase the length of the test and to throw out items that run down the reliability. Another procedure is to estimate what the true correlation would have been if the test did not have measurement error.

Increase the Number of Items

According to the domain sampling model, each item in a test is an independent sample of the trait or ability being measured. The larger the sample, the more likely that the test will represent the true characteristic. In the domain sampling model, the reliability of a test increases as the number of items increases.

Scanned with CamScanner

Many thousands of college-bound students take the SAT, so there is a lot of precision for estimating mean scores. However, we can use the formula to estimate the usual range to be expected around different scores. For example, we can estimate the 68% confidence interval, which would be equivalent to one standard deviation above or below the mean score of 489 for the writing component of the SAT. This would result in a range from 432 to about 546. The observed difference of 2 points is well within the range defined by one adjusted standard deviation from the mean.

Sampling is another explanation for the changes in scores between 2010 and 2011. We would expect the average performance on the test to be the same if the population of test takers was also the same. Increasing the number of less-prepared test takers is another alternative explanation for the decline in the average score. In 2011, more people than ever before took the SAT. Historically, the SAT had been taken by students most directed toward obtaining an education at a 4-year college. When the population of test takers expands, it is most likely that the population becomes more inclusive of students who have not had exposure to college preparation, including coaching for standardized tests. In 2011, not all test takers had completed the college preparative curriculum, which includes 4 years of English, 3 years of math, 3 years of natural science, and 3 years of social science and history. Those who did not complete the college prep core scored an average of 143 points lower than those who had finished the traditional preparation. This might be regarded as a positive sign because it represents increasing access to college and greater interest in more people obtaining a college education.

A medical example will clarify why longer tests are more reliable. Suppose that you go to the doctor with indigestion. You want the doctor to make a reliable judgment about what is causing it. How comfortable would you feel if the doctor asked only one question before making a diagnosis? You would probably feel more comfortable if the doctor asked many questions. In general, people feel that the more information a doctor obtains by asking questions and performing tests, the more reliable the diagnosis will be. This same principle applies to psychological tests.

A decision to increase the number of items in a test might engender a long and costly process. With new items added, the test must be reevaluated; it may turn out to fall below an acceptable level of reliability. In addition, adding new items can be costly and can make a test so long that few people would be able to sit through it. Fortunately, by applying the Spearman-Brown prophecy formula, one can estimate how many items will have to be added in order to bring a test to an acceptable level of reliability.

The prophecy formula for estimating how long a test must be to achieve a desired level of reliability is another case of the general Spearman-Brown method

Scanned with CamScanner

for estimating reliability. Algebraic manipulations of the general formula allow one to solve it for the length needed for any desired level of reliability:

$$N = \frac{r_d(1 - r_o)}{r_o(1 - r_d)}$$

where

N = the number of tests of the current version's length that would be needed

r_d = the desired level of reliability

r_o = the observed level of reliability based on the current version of the test

Consider the example of the 20-item CES-D test that had a reliability for medical students of .87. We would like to raise the reliability to .95. Putting these numbers into the prophecy formula, we get

$$N = \frac{.95(1 - .87)}{.87(1 - .95)} = \frac{.124}{.044} = 2.82$$

These calculations tell us that we would need 2.82 tests of the same length as the current 20-item test to bring the reliability up to the desired level. To find the number of items required, we must multiply the number of items on the current test by N from the preceding formula. In the example, this would give $20 \times 2.82 = 56.4$. So the test would have to be expanded from 20 to approximately 56 items to achieve the desired reliability, assuming that the added items come from the same pool as the original items and that they have the same psychometric properties.

The decision to expand a test from 20 to 56 items must depend on economic and practical considerations. The test developer must first ask whether the increase in reliability is worth the extra time, effort, and expense required to achieve this goal. If the test is to be used for personnel decisions, then it may be dangerous to ignore any enhancement of the test's reliability. On the other hand, if the test is to be used only to see if two variables are associated, the expense of extending it may not be worth the effort and cost.

When the prophecy formula is used, certain assumptions are made that may or may not be valid. One of these assumptions is that the probability of error in items added to the test is the same as the probability of error for the original items in the test. However, adding many items may bring about new sources of error, such as the fatigue associated with taking an extremely long test.

As an example of a situation in which increasing the reliability of a test may not be worthwhile, consider a 40-item test with a reliability of .50. We would like to bring the reliability up to .90. Using the prophecy formula, we get

$$N = \frac{.90(1 - .50)}{.50(1 - .90)} = \frac{.90(.50)}{.50(.10)} = \frac{.45}{.05} = 9$$

These figures tell us that the test would have to be nine times its present length to have a projected reliability of .90. This is calculated as $9 \times 40 = 360$ items long. Creating a test that long would be prohibitively expensive, and validating it would require a considerable investment of time for both test constructors and test takers. Also, new sources of error might arise that were not present in the shorter measure.

Scanned with CamScanner

For example, many errors may occur on the longer test simply because people get tired and bored during the long process of answering 360 questions. There is no way to take these factors into account by using the prophecy formula.

Factor and Item Analysis

The reliability of a test depends on the extent to which all of the items measure one common characteristic. Although psychologists set out to design test items that are consistent, often some items do not measure the given construct. Leaving these items in the test reduces its reliability. To ensure that the items measure the same thing, two approaches are suggested. One is to perform factor analysis (see Chapter 3 and Brown, 2015). Tests are most reliable if they are *unidimensional.* This means that one factor should account for considerably more of the variance than any other factor. Items that do not load on this factor might be best omitted.

Another approach is to examine the correlation between each item and the total score for the test. This form of item analysis (see Chapter 6) is often called **discriminability analysis.** When the correlation between the performance on a single item and the total test score is low, the item is probably measuring something different from the other items on the test. It also might mean that the item is so easy or so hard that people do not differ in their response to it. In either case, the low correlation indicates that the item drags down the estimate of reliability and should be excluded.

Correction for Attenuation

Low reliability is a real problem in psychological research and practice because it reduces the chances of finding significant correlations between measures. If a test is unreliable, information obtained with it is of little or no value. Thus, we say that potential correlations are attenuated, or diminished, by measurement error.

4.3 FOCUSED EXAMPLE

Measurement Error in the Era on Big Data

For any measured variable, the difference between the true score and the observed score results from measurement error. This error is common in all biological, social, and behavioral sciences, but it can be controlled through systematic measure development. Small controlled studies often devote great attention to measurement precision. In contrast, users of big data sets may be unaware of how study variables were measured and low reliability of some measures can be expected. This can be a serious problem because it reduces the chances of finding significant relationships among measures. Thus, estimates of correlations and effect sizes are attenuated by measurement error. Large sample size may help reduce this bias, but if the measures are of very low reliability, the analysis will be focused on random variation. For example, suppose two variables in a big data set have reliability coefficients of .050 and 0.30. Even if these two variables were perfectly correlated in the real world ($r = 1.0$), the maximum expected observed correlation would be only 0.38. A modest correlation would be unlikely to be detected, even in a very large data set. Having more observations does not necessarily resolve the issue of measurement error.

Scanned with CamScanner

Multivariate Analysis (Optional)

Multivariate analysis considers the relationship among combinations of three or more variables. For example, the prediction of success in the first year of college from the linear combination of SAT verbal and quantitative scores is a problem for multivariate analysis. However, because the field of multivariate analysis requires an understanding of linear and matrix algebra, a detailed discussion of it lies beyond the scope of this book.

On the other hand, you should have at least a general idea of what the different common testing methods entail. This section will familiarize you with some of the multivariate analysis terminology. It will also help you identify the situations in which some of the different multivariate methods are used. Several references are available in case you would like to learn more about the technical details (Brown, 2015; Gravetter & Wallnau, 2016; Vogt & Johnson, 2015).

General Approach

The correlational techniques presented to this point describe the relationship between only two variables such as stress and illness. To understand more fully the causes of illness, we need to consider many potential factors besides stress. Multivariate analysis allows us to study the relationship between many predictors and one outcome, as well as the relationship among the predictors.

Multivariate methods differ in the number and kind of predictor variables they use. All of these methods transform groups of variables into linear combinations. A *linear combination* of variables is a weighted composite of the original variables. The weighting system combines the variables in order to achieve some goal. Multivariate techniques differ according to the goal they are trying to achieve.

A linear combination of variables looks like this:

$$Y' = a + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k$$

where Y' is the predicted value of Y, a is a constant, X_1 to X_k are variables and there are k such variables, and the b's are regression coefficients. If you feel anxious about such a complex-looking equation, there is no need to panic. Actually, this equation describes something similar to what was presented in the section on regression. The difference is that instead of relating Y to X, we are now dealing with a linear combination of X's. The whole right side of the equation creates a new composite variable by transforming a set of predictor variables.

An Example Using Multiple Regression

Suppose we want to predict success in law school from three variables: undergraduate GPA, rating by former professors, and age. This type of multivariate analysis is called **multiple regression,** and the goal of the analysis is to find the linear combination of the three variables that provides the best prediction of law school success. We find the correlation between the criterion (law school GPA) and some composite of the predictors (undergraduate GPA plus professor rating plus age). The combination of the three predictors, however, is not just the sum of the three scores. Instead,

Scanned with CamScanner

we program the computer to find a specific way of adding the predictors that will make the correlation between the composite and the criterion as high as possible. A weighted composite might look something like this:

$$\begin{aligned}\text{law school GPA} = {} & .80\ (Z \text{ scores of undergraduate GPA}) \\ & +.54\ (Z \text{ scores of professor ratings}) \\ & +.03\ (Z \text{ scores of age})\end{aligned}$$

This example suggests that undergraduate GPA is given more weight in the prediction of law school GPA than are the other variables. The undergraduate GPA is multiplied by .80, whereas the other variables are multiplied by much smaller coefficients. Age is multiplied by only .03, which is almost no contribution. This is because .03 times any Z score for age will give a number that is nearly 0; in effect, we would be adding 0 to the composite.

The reason for using Z scores for the three predictors is that the coefficients in the linear composite are greatly affected by the range of values taken on by the variables. GPA is measured on a scale from 0 to 4.0, whereas the range in age might be 21 to 70. To compare the coefficients to one another, we need to transform all the variables into similar units. This is accomplished by using Z scores (see Chapter 2). When the variables are expressed in Z units, the coefficients, or weights for the variables, are known as *standardized regression coefficients* (sometimes called *B*'s or betas). There are also some cases in which we would want to use the variables' original units. For example, we sometimes want to find an equation we can use to estimate someone's predicted level of success on the basis of personal characteristics, and we do not want to bother changing these characteristics into Z units. When we do this, the weights in the model are called *raw regression coefficients* (sometimes called *b*'s).

Before moving on, we should caution you about interpreting regression coefficients. Besides reflecting the relationship between a particular variable and the criterion, the coefficients are affected by the relationship among the predictor variables. Be careful when the predictor variables are highly correlated with one another. Two predictor variables that are highly correlated with the criterion will not both have large regression coefficients if they are highly correlated with each other as well. For example, suppose that undergraduate GPA and the professors' rating are both highly correlated with law school GPA. However, these two predictors also are highly correlated with each other. In effect, the two measures seem to be of the same thing (which would not be surprising, because the professors assigned the grades). As such, professors' rating may get a lower regression coefficient because some of its predictive power is already taken into consideration through its association with undergraduate GPA. We can only interpret regression coefficients confidently when the predictor variables do not overlap and are uncorrected. They may do so when the predictors are uncorrected.

Discriminant Analysis

Multiple regression is appropriate when the criterion variable is continuous (not nominal). However, there are many cases in testing where the criterion is a set of categories. For example, we often want to know the linear combination of variables that differentiates passing from failing. When the task is to find the linear combination

Scanned with CamScanner

of variables that provides a maximum discrimination between categories, the appropriate multivariate method is **discriminant analysis.** An example of discriminant analysis involves attempts to determine whether a set of measures predicts success or failure on a particular performance evaluation.

Sometimes we want to determine the categorization in more than two categories. To accomplish this, we use multiple discriminant analysis.

Discriminant analysis has many advantages in the field of test construction. One approach to test construction is to identify two groups of people who represent two distinct categories of some trait. For example, say that two groups of children are classified as "language disabled" and "normal." After a variety of items are presented, discriminant analysis is used to find the linear combination of items that best accounts for differences between the two groups. With this information, researchers could develop new tests to help diagnose language impairment. This information might also provide insight into the nature of the problem and eventually lead to better treatments.

Factor Analysis

Discriminant analysis and multiple regression analysis find linear combinations of variables that maximize the prediction of some criterion. Factor analysis is used to study the interrelationships among a set of variables without reference to a criterion. You might think of factor analysis as a data-reduction technique. When we have responses to a large number of items or a large number of tests, we often want to reduce all this information to more manageable chunks. In Figure 3.1, we presented a two-dimensional scatter diagram. The task in correlation is to find the best-fitting line through the space created by these two dimensions. As we add more variables in multivariate analysis, we increase the number of dimensions. For example, a three-dimensional plot is shown in Figure 3.12. You can use your imagination to visualize what a larger set of dimensions would look like. Some people claim they can visualize more than three dimensions, while others feel they cannot. In any case, consider that points are plotted in the domain created by a given dimension.

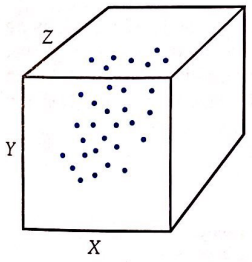

FIGURE 3.12
A three-dimensional scatter plot might be represented by this box. In addition to plotting points on the *X* and *Y* axes, we must locate them in relation to a third *Z* axis.

In factor analysis, we first create a matrix that shows the correlation between every variable and every other variable. Then we find the linear combinations, or *principal components,* of the variables that describe as many of the interrelationships among the variables as possible. We can find as many principal components as there are variables. However, each principal component is extracted according to mathematical rules that make it independent of or uncorrected with the other principal components. The first component will be the most successful in describing the variation among the variables, with each succeeding component somewhat less successful. Thus, we often decide to examine only a few components that account for larger proportions of the variation. Technically, principal components analysis and true factor analysis differ in how the correlation matrix is created. Even so, principal components are often called *factors.*

Once the linear combinations or principal components have been found, we can find the correlation between the original items and the factors. These correlations are called *factor loadings.* The expression "item 7 loaded highly on factor I" means

Scanned with CamScanner

The Factors of Trust

Rotter (1967) described a scale for the measurement of interpersonal trust. *Trust* was defined as "an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon" (p. 651). However, after the publication of the original trust article, several authors reported that trust seems to be composed of several independent factors (Chun & Campbell, 1974; Kaplan, 1973; Wright & Tedeschi, 1975). In each case, the items were given to a large group of people, and the results were subjected to factor analysis. This procedure reduces the many items down to a smaller number of *factors*, or linear combinations of the original items. Then *item loadings*, or the correlations of the original items with the factors, are studied in order to name the factors. The table that follows shows the loadings of the items on three of the factors (Kaplan, 1973).

Once they have obtained the factor loadings, researchers must attempt to name the factors by examining which items load highly on them. In this case, an item was used to help interpret a factor if its item loading on the factor was greater than .35 or less than −.35. Three factors of trust were found.

Factor I: Institutional trust. This represented trust toward major social agents in society. It included items regarding the competence of politicians, such as "This country has a dark future unless we can attract better people into politics" (−.67). Many of the items conveyed the idea of misrepresentation of public events by either the government or the mass media. For example, some items with high loadings were "Most people would be horrified if they knew how much news the public hears and sees is distorted" (−.69) and "Even though we have reports in newspapers, radio, and TV, it is hard to get objective accounts of public events" (−.67).

Factor II: Sincerity. Items loading highly on sincerity tended to focus on the perceived sincerity of others. These items included "Most idealists are sincere and usually practice what they preach" (.62) and "Most people answer public opinion polls honestly" (.58). Nearly all the items with high loadings on the second factor began with the word "most." Because of this loose wording, it would be possible for people to agree with the items because they believe in the sincerity of most people in a given group but still feel little trust for the group because of a few "rotten eggs." Thus, a woman could believe most car repairers are sincere but still service her car herself because she fears being overcharged.

Factor III: Caution. This contained items that expressed fear that some people will take advantage of others, such as "In dealing with strangers, one is better off being cautious until they have provided evidence that they are trustworthy" (.74) and "In these competitive times you have to be alert or someone is likely to take advantage of you" (.53). Note that caution appears to be independent of perceived sincerity.

The data imply that generalized trust may be composed of several dimensions. It also implies that focusing on specific components of trust rather than the generalized case will likely help researchers the most in using this trust scale.

Focused Example adapted from Rotter (1967); table taken from Kaplan (1973).

Scanned with CamScanner