Understanding Correlation: Factors That Affect the Size of r

Author(s): Laura D. Goodwin and Nancy L. Leech

Source: *The Journal of Experimental Education*, Spring, 2006, Vol. 74, No. 3 (Spring, 2006), pp. 251-266

Published by: Taylor & Francis, Ltd.

Stable URL: https://www.jstor.org/stable/20157427

The Journal of Experimental Education, *2006*, *74*(3), 251–266

Understanding Correlation: Factors That Affect the Size of *r*

LAURA D. GOODWIN
NANCY L. LEECH
University of Colorado at Denver and Health Sciences Center

ABSTRACT. The authors describe and illustrate 6 factors that affect the size of a Pearson correlation: (a) the amount of variability in the data, (b) differences in the shapes of the 2 distributions, (c) lack of linearity, (d) the presence of 1 or more "outliers," (e) characteristics of the sample, and (f) measurement error. Also discussed are ways to determine whether these factors are likely affecting the correlation, as well as ways to estimate the size of the influence or reduce the influence of each.

Key words: correlation, errors, interpretation, Pearson product–moment correlation

CORRELATION IS A COMMONLY USED STATISTIC in research and measurement studies, including studies conducted to obtain validity and reliability evidence. Understanding the meaning of a simple correlation is key to understanding more complex statistical techniques for which the simple correlation is the foundation. In basic statistics courses, students typically learn about the conceptual meaning of "relationship" between two variables (including size and direction), how to calculate and interpret a sample correlation, how to construct scattergrams or scatterplots to graphically display the relationship, and how to conduct an inferential test for the significance of the correlation and interpret the results. What is often missing in class discussions and activities, however, is a focus on factors that can affect the size of the statistic based on the characteristics of the correlation or the particular dataset used for the calculation of the correlation. Without a solid understanding of these factors, students and researchers

Address correspondence to: Laura D. Goodwin, University of Colorado at Denver and Health Sciences Center, Downtown Denver Campus, 1380 Lawrence Street, Suite 1400, Campus Box 137, PO Box 173364, Denver, CO 80217. E-mail: Laura.Goodwin@cudenver.edu

can find it difficult to "diagnose" a low correlation—or just to fully interpret results of simple or multivariate statistical analyses in the correlation "family."

The purpose of this article is to describe and illustrate six factors that affect the size of correlations, including (a) the amount of variability in either variable, *X* or *Y*; (b) differences in the shapes of the two distributions, *X* or *Y*; (c) lack of linearity in the relationship between *X* and *Y*; (d) the presence of one or more "outliers" in the dataset; (e) characteristics of the sample used for the calculation of the correlation; and (f) measurement error. Where possible, we illustrate the effects of these characteristics on the size of a correlation with a hypothetical data example.

Although correlation is a fairly basic topic in statistics courses, the effects on correlations of the six factors or characteristics discussed in this article are rarely presented succinctly in one place. As will be shown next, several of these characteristics are not typically discussed in basic statistics textbooks at all.

A correlation describes the relationship between two variables. Although there are a number of different correlation statistics (Glass & Hopkins, 1996), the one that is used most often is the Pearson product–moment correlation coefficient, hereafter referred to as *correlation* in this article. This statistic describes the size and direction of the linear relationship between two continuous variables (generically represented by *X* and *Y*), and ranges in value from −1.0 (perfect negative relationship) to +1.0 (perfect positive relationship); if no relationship exists between the two variables, the value of the correlation is zero. The symbol r_{xy} (or r) is used to represent the correlation calculated with a set of sample data. The correlation requires that both variables (*X* and *Y*) are measured on interval or ratio scales of measurement. A formula for the correlation is:

$$r_{xy} = \frac{s_{xy}}{s_x s_y},$$

where s_{xy} is the covariance between the two variables and s_x and s_y are the standard deviations for *X* and *Y*, respectively. Another commonly presented formula in textbooks uses z scores:

$$r_{xy} = \frac{\Sigma z_x z_y}{n},$$

where z_x and z_y are the z scores for each individual on the *X* and *Y* variables, respectively. The squared correlation (r^2) is a very useful statistic. It indicates the proportion of *shared variance*, or the proportion of the total variance in *Y* that is predictable from a regression equation.

Understanding the meaning of correlation is very important for measurement practitioners and researchers. Simple correlations are used in many measurement studies, such as studies aimed at obtaining validity and reliability evidence. Furthermore, many multivariate statistical procedures—such as multiple regression, factor analysis, path analysis, and structural equation modeling—build on, or are extensions of, simple correlations. Rodgers and Nicewander (1988) outlined 13 ways of interpreting a correlation. These included the interpretation of a correlation as the standardized slope of the regression line, as the proportion of variability accounted for, and as a function of test statistics. Rovine and von Eye (1997) added a 14th way: as the proportion of matches. In general, it is helpful to understand that correlations can be used and interpreted in multiple ways.

Although virtually all basic statistics textbooks cover the topic of correlation, very few describe and illustrate all six of the characteristics discussed here. We recently examined a sample of 30 statistics textbooks published in the past 10 years.[1] All of the books we reviewed were written for basic and intermediate-level statistics courses. Of the six characteristics that affect the value of r, lack of linearity was covered most often (in 22, or 73%, of the textbooks). Next in order of frequency of coverage was the lack of variability in X or Y (in 17, or 57%, of the textbooks), followed by the presence of outliers (in 11, or 37%, of the textbooks). The effect of measurement error on r was covered in only 4 (13%) of the books, and the effect that dissimilar shapes of distributions for X and Y have on the maximum size of r was covered in just 2 (7%) of the books reviewed. Characteristics of samples often overlap with other factors that affect the size of r, such as variability or presence of outliers. Thus, it was difficult to "rate" the textbooks on this dimension; however, very few of the books included descriptions or examples that went beyond the influence of variability on r. In terms of the other five dimensions, we found only 1 textbook that directly addressed all of them.

The Hypothetical Dataset

For purposes of illustrating the effects that various characteristics have on the size of r, we constructed a hypothetical dataset (Table 1). We purposefully kept the data simple so that interested readers could easily replicate the analyses. The hypothetical dataset consists of scores on five variables X, Y_1, Y_2, Y_3, and Y_4 for 30 subjects. Each variable has six levels (or "scores"), and the numbers assigned to the levels of the variables range from 1 to 6. Used as a class example, these variables might be questions about "interest in the field of statistics," "comfort with math," "anxiety about statistics," and so on. Responses could be solicited with a Likert-type scale. It is important to note that both variables in a correlation must be measured on an interval or ratio measurement scale to use the Pearson correlation statistic; we assume, therefore, that these are interval-level data.[2]

Table 2 presents descriptive statistics for each variable. These include mea-

[1]A list of the textbooks reviewed is available from the first author.

[2]To keep this example simple, our hypothetical data consist of one-item measures. In "real life," measuring abstract constructs such as "anxiety" or "interest" with just one-item measures is, of course, inappropriate; the resulting sets of scores would most likely be quite unreliable.

TABLE 1. Hypothetical Dataset for Correlation Illustrations

Participant	X	Y_1	Y_2	Y_3	Y_4
1	1	2	2	2	1
2	2	2	2	1	2
3	2	1	1	2	2
4	2	2	2	1	3
5	2	3	3	2	3
6	3	2	2	2	4
7	3	3	3	1	4
8	3	3	3	1	5
9	3	3	3	1	5
10	3	3	3	1	5
11	3	3	3	1	5
12	3	3	3	1	6
13	3	4	4	1	6
14	3	4	4	1	6
15	3	4	4	1	6
16	4	3	3	1	4
17	4	3	3	1	4
18	4	3	3	1	5
19	4	4	4	1	5
20	4	4	4	2	5
21	4	4	4	3	5
22	4	4	4	3	6
23	4	4	4	4	6
24	4	4	4	4	6
25	4	4	4	3	6
26	5	5		3	3
27	5	5		4	3
28	5	6		5	2
29	5	5		5	2
30	6	5		6	1

TABLE 2. Descriptive Statistics for the Five Hypothetical Variables (N = 30)

Statistic	X	Y_1	Y_2	Y_3	Y_4
M	3.50	3.50	3.16	2.17	4.20
Mdn	3.50	3.50	3.00	1.50	5.00
Mode(s)	3.00	3.00	3.00	1.00	5.00
	4.00	4.00	4.00	6.00	
SD	1.11	1.11	.85	1.49	1.63
Skewness	.00	.00	–.77	1.11	–.55
Kurtosis	.06	.06	.06	.20	–.92

sures of central tendency (mean, median, mode), standard deviations, and values for the skewness and kurtosis of each distribution. Note that the distributions of both X and Y_1 were intentionally constructed to have identical values of these statistics and to be symmetrical in shape. The correlation between these two variables serves as the "original" correlation in this article—allowing for subsequent comparisons by varying the values of the Y variable (i.e., Y_2, Y_3, and Y_4). The correlation between X and Y_1 is .83; a scattergram illustrating this relationship is shown in Figure 1.

Amount of Variability in *X* or *Y*

It is well known among statisticians that, other things being equal, the value of r will be greater if there is more variability among the observations than if there is less variability. However, many researchers are unaware of this fact (Glass &

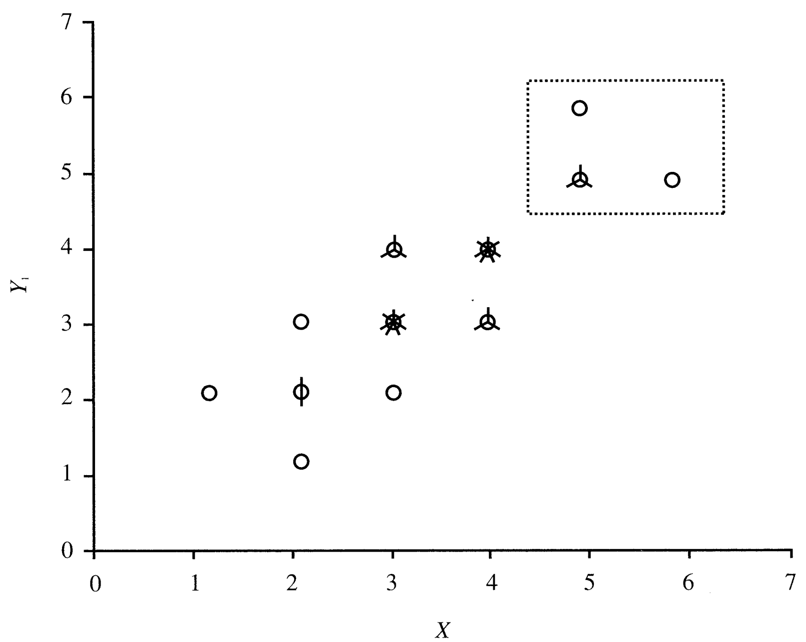

FIGURE 1. The full scattergram represents the relationship between X and Y_1, whereas the scattergram without the boxed values represents the relationship between X and Y_2. A circle represents one case. Circles with spikes represent multiple cases. Each spike represents one case.

Hopkins, 1996), and it is common for students in basic statistics courses (and even some students in intermediate- and advanced-level courses) to have difficulty comprehending this concept. Examples of this characteristic of r can be found quite easily, and it is often termed *range restriction*, *restriction of range*, or *truncated range* by authors of statistics and measurement textbooks (e.g., Abrami, Cholmsky, & Gordon, 2001; Aron & Aron, 2003; Crocker & Algina, 1986; Glenberg, 1996; Harris, 1998; Hopkins, 1998; Spatz, 2001; Vaughan, 1998). In predictive validity studies, the phenomenon occurs when a test is used for selection purposes; subsequently, the scores obtained with the test are correlated with an outcome variable that is only available for those individuals who were selected for the educational program or job. For example, the correlation between SAT scores and undergraduate grade point average (GPA) at some selective universities is only about .20 (Vaughan). This does not necessarily mean that there is little relationship between SAT scores and college achievement, however. The range of SAT scores is small at selective colleges and universities that use SAT scores as a criterion for admission. Furthermore, GPAs can be restricted in elite colleges. Other things being equal, the correlation between SAT scores and GPAs would be greater if there were a greater range of scores on the SAT and a greater range of GPAs. Other examples of range restriction can be attributed to the sampling methods used. For example, if individuals are chosen to participate in a study based on a narrow range of scores on a variable, correlations between that variable and any other variables will be low. The "ultimate" situation in which low variability influences a correlation occurs when there is no variability on either X or Y. In that case, the correlation between the variable with no variability and any other variable is not even defined (Hays, 1994).

To illustrate the relationship between variability and the size of a correlation, we reduced the amount of variability in both X and Y by removing the five highest scoring cases from the Y variable. This can be seen in the second distribution of Y (Y_2) in Table 1, in which all remaining participants' scores range from 1 to 4 (rather than 1–6 in the original Y_1 distribution; by removing 5 cases from Y, those same cases are also removed from X when the correlation is calculated). With those 5 cases removed, the value of the correlation shrinks from .83 to .71. Although the nature of the relationships among the remaining 25 cases is essentially the same as it was when the original correlation was calculated, the size of the correlation now is smaller due to the shrinkage in the variability in the data. Therefore, it appears that the relationship is smaller, too. This can also be seen in Figure 1, in which the removed cases are surrounded by a dotted box. Without those cases in the scattergram, the relationship is seen as less strong (more of a "circle" in shape).

In trying to determine why a correlation might be lower than it was expected to be (or, perhaps, lower than other researchers have reported), examining the amount of variability in the data can be very helpful. This can be done visually (by looking at a scatterplot), as well as by calculating the variances or standard deviations. In terms of restriction of range, there are procedures available for the estimation of the correlation for the entire group from the correlation obtained with the selected group (Glass & Hopkins, 1996; Gulliksen, 1950; Nunnally & Bernstein, 1994; Thorndike, 1982). However, the equation used to estimate the unrestricted correlation requires knowledge of the standard deviations for X and Y for the entire group and also requires several assumptions that are rarely tenable in practical situations (Crocker & Algina, 1986). Furthermore, the obtained estimates are often imprecise unless the sample size, N, is very large (Gullickson & Hopkins, 1976; Linn, 1983). As Glass and Hopkins noted, the main value of using the equation is not to actually estimate the size of the unrestricted correlation but, rather, to illuminate "the consequence of restricted or exaggerated variability on the value of r so that it can be interpreted properly" (p. 122).

The Shapes of the Distributions of X and Y

The correlation can achieve its maximum value of 1.0 (positive or negative) only if the shapes of the distributions of X and Y are the same (Glass & Hopkins, 1996; Hays, 1994; Nunnally & Bernstein, 1994). Carroll (1961) showed that the maximum value of r, when the distributions of X and Y do not have the same shape, depends on the extent of dissimilarity (or lack of similarity in skewness and kurtosis): the more dissimilar the shapes, the lower the maximum value of the correlation. Nunnally and Bernstein also noted that the effect on the size of r depends on how different the shapes of the distributions are, as well as how high the correlation would be if the distributions had identical shapes. In terms of the latter, the effect is greater if the correlation between the same-shaped distributions is greater (other things being equal). For example, if the correlation were .90 between same-shaped distributions, changes in the shape of one of the distributions could reduce the size of the correlation to .80 or .70. On the other hand, if the correlation were .30 between same-shaped distributions, even dramatic changes in the shape of one of the distributions will have relatively little effect on the size of r (assuming that N is fairly large—approximately 30 or more subjects—so that there is some stability in the data). Nunnally and Bernstein also discussed the situation wherein one variable is dichotomous and the other is normally distributed. They showed that the maximum value of the correlation is about .80, which can occur only if the p value (difficulty index) of the dichotomous variable is .50; as the p value deviates from .50 (in either direction), the ceiling on the correlation becomes lower than .80.

To illustrate this characteristic of the correlation, we retained the original X variable's distribution (which is symmetrical) but altered the distribution of the Y variable. As compared with the original distribution for Y (Y_1), the distribution of Y_3 is skewed positively. (See the value of the skewness statistic in Table 2.) The correlation between X and Y_3 is now .68, which is illustrated in Figure 2. Note that Y_3 has a differently shaped distribution and greater variance than Y_1. Yet, Y_3 has a lower correlation with X than Y_1 has with X.

If different-shaped distributions attenuate r, one or both variables can be transformed so that the distributions become more similar in shape. However, nonlinear transformations of X or Y will have only a small effect on the size of the correlation unless the transformations markedly change the shapes of the distributions (Glass & Hopkins, 1996).

Lack of Linearity

The correlation measures the extent and direction of the *linear* relationship between X and Y. If the actual relationship between X and Y is not linear—rather, if it is a *curvilinear* or *nonlinear* relationship—the value of r will be very low and might even be zero. Although the relationships between most variables examined in educational and behavioral research studies are linear, there are interesting examples of nonlinear relationships among adults between age and psychomotor

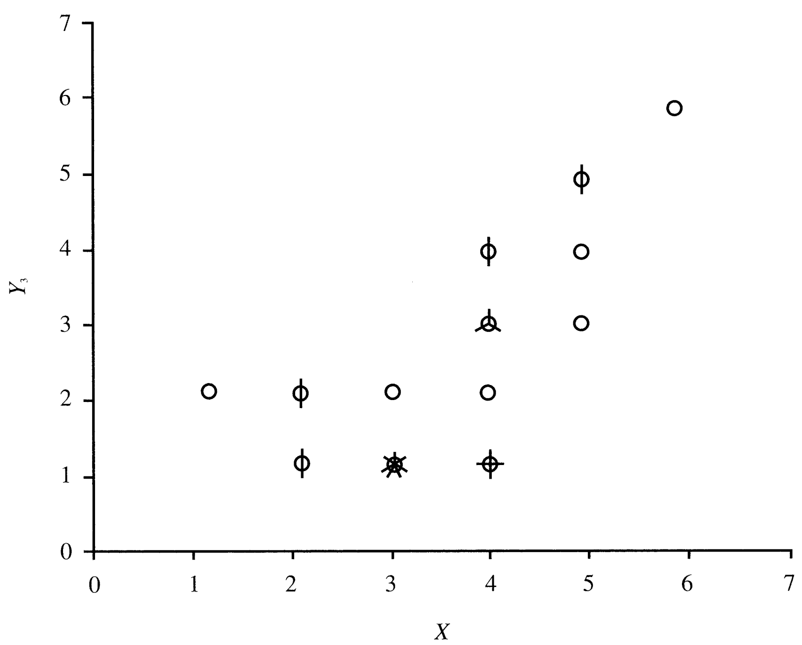

FIGURE 2. Scattergram of X and Y_3.

skills that require coordination (Glass & Hopkins, 1996). Also, some researchers studying the relationships between anxiety and test performance have reported curvilinear relationships (Hopkins, 1998). Abrami et al. (2001) described the anxiety–test performance relationship: "One of the most famous examples of a curvilinear relationship in the social sciences is the inverted U-shaped relationship between personal anxiety and test performance. It is now well known that 'moderate' levels of anxiety optimize test performance" (p. 434).

The best way to detect a curvilinear relationship between two variables is to examine the scattergram. If a curvilinear relationship exists between X and Y, the Pearson correlation should not be used; it will seriously underestimate the strength of the relationship. Instead, the correlation ratio, or eta (η), is defined by Vogt (1999) as "a correlation coefficient that does not assume that the relationship between two variables is linear" (p. 99). This statistic allows one to calculate the size of the relationship between any two variables and would be a method of choice. It is a "universal measure of relationship" (Nunnally & Bernstein, 1994, p. 137) because it measures the relationship between two variables regardless of the form of the relationship (linear or nonlinear). Also, it can be used with nominal or continuous (interval- or ratio-level) data. Like r^2, η^2 indicates the proportion of "shared variance" between the two variables.

Another approach to take if the relationship is nonlinear is to transform one or both variables—by raising variables to powers, expressing variables as logarithms, or taking square roots of variables (Abrami et al., 2001). Creating a scattergram and computing r again after the transformations may show a linear relationship that was not apparent prior to transformation. However, the linear relationship is now with the transformed variable, not the original variable. This can make interpretation more complex.

To illustrate a nonlinear relationship, we altered the original Y variable again. The values of the new Y variable (Y_4) are included in Table 1, and the scattergram showing the relationship between X and Y_4 is shown in Figure 3. The calculated value of the Pearson correlation between X and Y_4 is zero. If we did not look at the scattergram in this case, we would erroneously conclude that there is no relationship between the two variables when, in fact, there is a very strong and likely meaningful relationship. The value of η is .91, and η^2 is .83.

Presence of One or More Outliers

An *outlier* can be defined as a score or case that is so low or so high that it stands apart from the rest of the data. Reasons for outliers include data collection errors, data entry errors, or just the fact that a valid (although unusual) value occurred (Brase & Brase, 1999); inadvertent inclusion of an observation from a different population (Glenberg, 1996); or a subject not understanding the instructions or wording of items on a questionnaire (Cohen, 2001). As with many of the

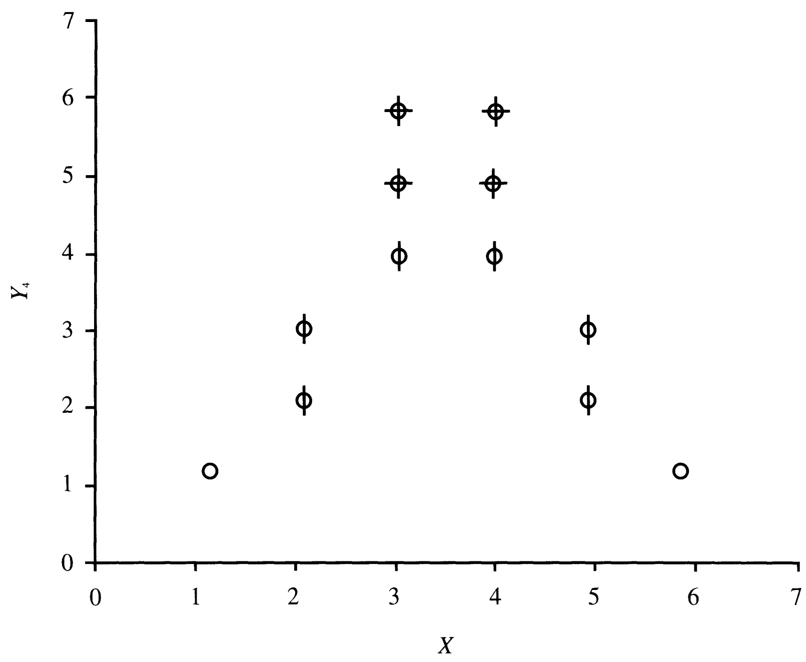

FIGURE 3. Scattergram of X and Y_4.

characteristics of a dataset that can affect the size of r, an outlier's effect will be greater in a small dataset than in a larger one. The presence of an outlier in a dataset can result in an increase or decrease in the size of the correlation, depending on the location of the outlier (Glass & Hopkins, 1996; Lockhart, 1998).

To illustrate the effect of an outlier on the correlation between X and Y, we added a 31st case to the original X and Y distributions (i.e., the X and Y_1 distributions in Table 1). We assigned a value of 9 on X and 10 on Y for this additional case. The new scattergram is shown in Figure 4, and the value of r is .91. (Recall that the original correlation was .83; adding the outlier case "stretched" the scattergram and increased the calculated value of r.)

As is the case with nonlinear relationships, one simple way to detect the presence of one or more outliers is to examine the scattergram; statistical outlier analysis (e.g., Tukey, 1977) can also be useful above and beyond analysis by visual inspection. If an outlier is present, the researcher should first check for data collection or data entry errors. If there were no errors of this type and there is no obvious explanation for the outlier—the outlier cannot be explained by a third variable affecting the person's score—the outlier should not be removed. If there

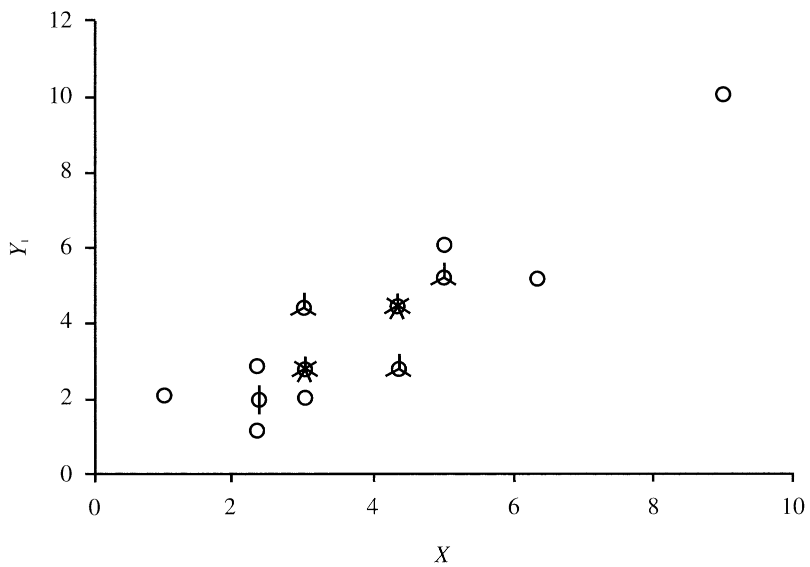

FIGURE 4. Scattergram of X and revised Y_1 to include an outlier.

is a good reason for a participant responding or behaving differently than the rest of the participants, the researcher can consider eliminating that case from the analysis; however, the case should not be removed only because it does not fit with the researcher's hypotheses (Field, 2000). Sometimes the researcher has to live with an outlier (because he or she cannot find an explanation for the odd response or behavior). Also, as Cohen (2001) noted, the outlier might represent an unlikely event that is not likely to happen again—hence, the importance of replication of the study.

Characteristics of the Sample

Unique characteristics of a sample can affect the size of r. Sometimes these characteristics overlap with one or more of the other situations already discussed—truncated range, presence of outliers, skewed data. Sprinthall (2003), for example, described a study reported by Fancher (1985) in which the correlations between IQ scores and grades decreased as the age of the participants increased: .60 for elementary school students, .50 for secondary school students, .40 for college students, and only .30 for graduate students: " . . . the lower IQs are consistently and systematically weeded out as the students progress toward more intellectually demanding experiences" (Sprinthall, p. 286). Of course, the amount of variability in the IQ scores likely decreased, too, across the various sample groups.

In addition to examples in which the unique characteristics of the sample coincide with one of the other characteristics that affect *r*, there are situations when the correlation is different for one group versus another group because of the nature of the participants studied. For example, we expect that the relationship between shoe size and spelling ability would be fairly large and positive when calculated on a sample of children aged 4 through 10 years but would be negligible when calculated on a sample of college freshmen. Combining different subgroups into one group prior to calculating a correlation also can produce some interesting results. Glenberg (1996) showed how the correlation between the age of widowers and their desire to remarry was positive for two separate subgroups: one group of fairly young widowers and one group of fairly old widowers. When the two subgroups were combined into a total group, however, the relationship between the two variables actually became negative.

Finally, sample selection can affect the strength of relationships calculated with correlations. "Sometimes misguided researchers select only the extreme cases in their samples and attempt to look at the relationship between the two variables" (Runyon, Haber, & Coleman, 1994, p. 136). The example presented by Runyon et al. dealt with the relationship between scores on a depression test and performance on a short-term memory task. A researcher might administer a depression measure to a group of patients and then select only those patients who scored in the top 25% and the bottom 25%. The calculated correlation between the two variables would be artificially enhanced by the inclusion of only the two extreme groups, providing an erroneous impression of the true relationship between depression and short-term memory.

Measurement Error

As noted earlier, the effects of measurement error on correlations are rarely presented in basic statistics textbooks; only 4 (13%) of the 30 books we reviewed included this topic as part of the discussion of correlation. Measurement error, which decreases the reliability of the measures of the variables, can be attributed to a variety of sources: intraindividual factors (fatigue, anxiety, guessing, etc.), administrative factors, scoring errors, environmental factors, ambiguity of questions, too few questions, and so on. In classical test theory, *reliability* is defined as the ratio of true-score variance to observed-score variance (Hopkins, 1998; Thompson, 2003). Other things being equal, the correlation between two variables will be lower when there is a large amount of measurement error—or low measurement reliability—than when there is a relatively small amount of measurement error. This makes sense, given the fact that reliability is the correlation of a test with itself (Lockhart, 1998); if a test does not correlate with itself, it cannot correlate with another variable. Consequently, the reported value of r may "substantially underestimate the true correlation between the underlying variables that these imperfect measures are meant to reveal" (Aron & Aron, 1994, p. 90). Thus, the reliability of a measure places an upper bound on how high the correlation can be between the measured variable and any other variable; the *reliability index*, which is the square root of the reliability coefficient, indicates the maximum size of the correlation (Hopkins).

The reduction in the size of a correlation due to measurement error is called *attenuation*, and there is a *correction for attenuation* that allows one to estimate what the correlation between the two variables would be if all measurement error were removed from both measures (Hopkins, 1998; Muchinsky, 1996; Nunnally & Bernstein, 1994). To use this equation, the researcher needs to know the reliability of each measure:

$$r^* = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}},$$

where r^* is the estimated correlation; r_{xy} is the calculated correlation between the two variables; and r_{xx} and r_{yy} are the reliability coefficients of the measures of X and Y, respectively. This equation really results in an *estimate* rather than a *correction*—that is, the estimate of the correlation between two variables if both measures were perfectly reliable. Nunnally and Bernstein advised caution in the use of the formula, especially because it can be used to fool one into believing that a higher correlation has been found than what actually occurred—and, with very small samples, the corrected correlation can even surpass 1.0! They also noted some appropriate uses:

> However, there are some appropriate uses of the correction for attenuation given good reliability estimates. One such use is in personality research to estimate the correlation between two traits from imperfect indicators of these traits. Determining the correlation between traits is typically essential in this area of research, but if the relevant measures are only modestly reliable, the observed correlation will underestimate the correlations among the traits. (p. 257)

As an example of the use of the correction for attenuation formula, assume that the correlation between measures of two traits is .65; the reliabilities of the measures are .70 and .80. The estimated correlation between two perfectly reliable measures here is .86.

Summary and Conclusion

In this article, we have discussed and illustrated six factors that affect the size of a correlation coefficient. When confronted with a very low or zero correlation, researchers should ask questions about these factors. Is there a lack of variability in the data? Do the marginal distributions have dissimilar shapes? Is there a nonlinear or curvilinear relationship between the two variables? Are there one or more outliers in the dataset? Are there other unique characteristics of the sample that might be responsible for an unusually low value of r? Is the measurement reliability for either variable (or both) low? To detect these possible problems or explanations for low correlations, various strategies can be used. Examining scatterplots will help identify a lack of variability in the data, a lack of linearity in the relationship, or the presence of outliers. Various descriptive statistics—standard deviations or variances and skewness statistics—can help identify a lack of variability or dissimilar distribution shapes, respectively. Complete descriptions of samples and sampling methods may help identify special characteristics of the sample that might be responsible for unusually low (or high) values of r. And, finally, the reliability coefficients will reveal the possibility of an attenuated correlation due to measurement error. It is also possible, of course, that none of the six factors are responsible for the low value of r and that, instead, there just is no relationship between the two variables. Furthermore, it is also possible that, although one or more of the six characteristics is affecting the size of r, the problem is exacerbated by a small sample size; if the sample were larger, the extent of the adverse effect would be reduced. Correlations calculated on data collected from a small sample (say, 30 or fewer subjects) can be affected substantially by any changes in scores, including the addition of an outlier or transformations of the variables. The effects on correlations of dissimilar distribution shapes are greater for small samples than large ones, as well.

For most of the factors described in this article, there are strategies available to either reduce the effect of the phenomenon or estimate the "true" strength of the relationship between the two variables. Data transformations can be useful when dissimilar distribution shapes or a lack of linearity occur. For nonlinear relationships, calculating a different relationship statistic—η—is also appropriate. If the "culprit" is the presence of one or more outliers (which can result in a spuriously high or low correlation), reasons for the presence of the outliers should be explored; if a reasonable explanation can be found, the researcher may be able to justify removing the odd case(s) and then recalculating the correlation. When restriction of range or measurement error exists, equations are available to estimate what the correlations would be if the data were unrestricted or error-free. (The equation used to estimate an unrestricted correlation requires information and assumptions that are rarely tenable, however.)

There are several other interesting and important correlation-related topics that are beyond the scope of this article but deserve brief mention. Knowing that correlations are not sensitive to some factors or conditions is as important for students and researchers as knowing about the factors that do affect the size of r. One common source of confusion that belongs in this arena pertains to the fact that correlations are "mean-free." Using interrater reliability as an example can result in a powerful illustration of this fact. If two raters tend to agree on the relative placement of participants' scores but differ dramatically in the levels of the scores assigned, the correlation will be very high and positive but the two raters' means will differ greatly. Similarly, linear transformations of scores (such as converting raw scores to z scores) will not change the correlation between those data and another variable. A second common misconception is that sample size (N) has a direct relationship to the size of r—a misconception that often results in erroneous interpretations of reliability and validity coefficients (Goodwin & Goodwin, 1999). Although small samples can result in unstable or inaccurate results (Hinkle, Wiersma, & Jurs, 2003), the size of N itself has no direct bearing on the size of the calculated value of r.

Another misconception students and researchers sometimes develop is that a correlation can be interpreted as a proportion or a percentage; therefore, understanding the difference between r and r^2 is a very useful way to prevent this misconception. The limitations of statistical significance tests—particularly in terms of the ease with which a correlation can be found to be statistically significant when the sample size is very large—is another important aspect of the study of correlation; distinguishing between statistical and practical significance can be crucial. Finally, no discussion of correlation is complete without emphasizing that the correlations found in correlational research studies cannot be interpreted as causal relationships between two variables. However, as one of the reviewers of this article pointed out, in an experimental study where random assignment is used, a correlation (point biserial) can be computed. In that case, a causal inference can be drawn for the relationship between the grouping variable and the outcome variable.

Given that correlations are so widely used in research in education and the behavioral sciences—as well as in measurement research aimed at estimating validity and reliability—it is critical that students have knowledge of the important (and sometimes subtle) factors that can affect the size of r. Knowledge of the role these factors play is also very helpful when students or researchers find unexpectedly low correlations in their research. An unexpectedly low correlation might be "explained" by one or more of the factors that affect the size of r; knowing this, a researcher would be encouraged to continue with his or her line of research rather than abandon it under the mistaken impression that there is no relationship between the variables of interest. It is also important to note that some factors—such as outliers and sample characteristics—can result in spuriously high correlations. In all cases, researchers should be advised to carefully consider possible contributing factors when interpreting correlational results.

REFERENCES

Abrami, P. C., Cholmsky, P., & Gordon, R. (2001). *Statistical analysis for the social sciences: An interactive approach.* Needham Heights, MA: Allyn & Bacon.

Aron, A., & Aron, E. N. (1994). *Statistics for psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Aron, A., & Aron, E. N. (2003). *Statistics for psychology* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
Brase, C. H., & Brase, C. P. (1999). *Understanding statistics: Concepts and methods* (6th ed.). Boston: Houghton Mifflin.
Carroll, J. B. (1961). The nature of the data, or how to choose a correlation coefficient. *Psychometrika, 26,* 247–272.
Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York: Wiley.
Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Fort Worth, TX: Harcourt Brace Jovanovich.
Fancher, R. E. (1985). *The intelligence men.* New York: Norton.
Field, A. (2000). *Discovering statistics using SPSS for Windows.* London: Sage.
Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
Glenberg, A. M. (1996). *Learning from data: An introduction to statistical reasoning.* Mahwah, NJ: Erlbaum.
Goodwin, L. D., & Goodwin, W. L. (1999). Measurement myths and misconceptions. *School Psychology Quarterly, 14,* 408–427.
Gullickson, A. R., & Hopkins, K. D. (1976). Interval estimation of correlation coefficients corrected for restriction of range. *Educational and Psychological Measurement, 36,* 9–25.
Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.
Harris, M. B. (1998). *Basic statistics for behavioral science research* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
Hays, W. L. (1994). *Statistics* (5th ed.). Fort Worth, TX: Harcourt Brace College Publishers.
Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston: Houghton Mifflin.
Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Needham Heights, MA: Allyn & Bacon.
Linn, R. L. (1983). Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement, 20,* 1–16.
Lockhart, R. S. (1998). *Introduction to statistics and data analysis for the behavioral sciences.* New York: W. H. Freeman.
Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56,* 63–75.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
Rodgers, J. L., & Nicewander, W. L. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42,* 59–66.
Rovine, M. J., & von Eye, A. (1997). A 14th way to look at a correlation coefficient: Correlation as the proportion of matches. *The American Statistician, 51,* 42–46.
Runyon, R. P., Haber, A., & Coleman, K. A. (1994). *Behavioral statistics: The core.* New York: McGraw-Hill.
Spatz, C. (2001). *Basic statistics: Tales of distributions* (7th ed.). Belmont, CA: Wadsworth/Thomson Learning.
Sprinthall, R. C. (2003). *Basic statistical analysis* (7th ed.). Boston: Allyn & Bacon.
Thompson, B. (Ed.). (2003). *Score reliability: Contemporary thinking on reliability issues.* Thousand Oaks, CA: Sage.
Thorndike, R. L. (1982). *Applied psychometrics.* Boston: Houghton Mifflin.
Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.
Vaughan, E. D. (1998). *Statistics: Tools for understanding data in the behavioral sciences.* Upper Saddle River, NJ: Prentice-Hall.
Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (2nd ed.). Thousand Oaks, CA: Sage.