Measures of Association
How to Choose?

HARRY KHAMIS, PhD

Researchers in sonography, as well as other areas, often wish to measure the strength of relationship or association between two variables. For example, one may wish to determine if, on the average, total cholesterol level increases as age increases for adult American men. However, there are a very large number of measures or coefficients (i.e., a number that indicates the strength of the relationship between two variables) from which to choose. It is not infrequent to find a researcher selecting an incorrect coefficient to measure a given association, thereby possibly rendering a false or misleading conclusion. The choice of the proper measure of association is based on, among other things, the characteristics of each of the two variables involved. This article enumerates every case that can be encountered by the researcher and provides an appropriate measure of association that can be used.

Key words: coefficient, relationship, ordinal, nominal, continuous

Very often, as a part of the general analyses of a set of data, a researcher wishes to determine the strength of the relationship between two variables using a single *coefficient* or *measure of association*—namely, a number (often between –1 and +1 or between 0 and 1) that is used as a measure of how strongly the two variables are related. In such instances, it is important that the appropriate measure is used to assess the strength of the relationship. If an inappropriate measure is used, then the resulting value is meaningless, and misleading results may be concluded. For example, the commonly used Pearson correlation coefficient is not necessarily the correct measure of association in every instance. To decide on the appropriate measure of

From Wright State University, Dayton, Ohio.

Correspondence: Harry Khamis, PhD, Wright State University, 130 Mathematical and Microbiological Sciences Bldg, 3640 Col. Glenn Hwy, Dayton, OH 45435. E-mail: harry.khamis@wright.edu.

DOI: 10.1177/8756479308317006

association, one must identify the *level of measurement* (defined and discussed below) of each variable being studied. Based on this information, an appropriate measure of association can be identified as outlined below. Indeed, for any given situation, there may be several different measures of association that are valid. Rather than providing an exhaustive list of all possible such measures, this article provides recommendations of just one or two measures for each practical situation encountered by the researcher. Ample references are provided, directing the researcher to sources of further reading. Several real-data examples are provided, illustrating the measures discussed.

In the next section, levels of measurement are defined and discussed. In subsequent sections, the process of selecting an appropriate measure of association for any given situation is outlined, and a summary and conclusion are provided.

Levels of Measurement

The *level of measurement* (or *measurement scale*) of a variable is its designation as continuous or discrete.

A *continuous*-valued variable has values that, at least theoretically, come from a continuum of the real number line. For such variables, there are, theoretically at least, no gaps in the possible values of the variable. Examples of continuous-valued variables are gestational age, blood pressure, body mass index, left ventricular ejection fraction (calculated as the percentage of blood expelled in a cycle), size of rotator cuff calcification (in cm), and coracoacromial distance (in cm). Consider gestational age: it is possible that there is a fetus with a gestational age of 60.32 days, even though the recorded value is 60 days. The reason that the more precise measurement, 60.32 days, is not recorded is due perhaps to the lack of accuracy of the measuring instrument.

A *discrete*-valued variable has values that are discrete or "separated." For such variables, the values do not come from a continuum of the real number line; rather, there are gaps between the values of the variable. Usually, such variables have only a few possible values. Examples of discrete-valued variables are gender (with values male, female), severity of disease (with values mild, moderate, severe), type of operation (with values standard, modified, laparoscopic), New York borough (with values Manhattan, Queens, etc.). Because of the way these variables are defined, it is not possible to observe a value between two given values. For example, each patient is either male or female with no other possible designation.

In the case of discrete variables, there are two subcategories of measurement scale, ordinal and nominal. An ordinal variable is a discrete variable having an order associated with its levels. For example, severity of disease is an ordinal variable because the "moderate" level represents a somewhat more severe disease state than the "mild" level, and the "severe" level corresponds to a more severe condition than the "moderate" level. If the levels of a discrete variable do not have any order associated with them, then the variable is called a discrete *nominal* variable. Type of operation is a nominal variable because the three different types of operation do not have any order associated with them. You might argue that indeed there is an order associated with these three levels—for instance, one type of operation may be more expensive than another type of operation. However, this is a different variable from the one defined earlier. If the variable of interest is cost of operation, with levels inexpensive, moderate, and expensive, then indeed this would be an ordinal variable. However, type of operation is a nominal variable.

Measures of Association—How to Choose

Suppose you wish to study the relationship between two variables by using a single measure or coefficient. There are many considerations that go into selecting an optimum measure of association; these are briefly discussed in the Summary section. For simplification, we base our selection of an appropriate measure exclusively on what type of variable we are measuring (i.e., its level of measurement)—namely, whether it is (1) continuous, (2) discrete ordinal, or (3) discrete nominal.

On the basis of this information, we choose an appropriate measure of association with which to analyze the relationship between the two variables. From a practical point of view, the six possible combinations of variables encountered by researchers are as follows:

1. Continuous-continuous
2. Continuous-ordinal
3. Continuous-nominal
4. Ordinal-ordinal
5. Ordinal-nominal
6. Nominal-nominal

For each of these combinations of variables, one or more measures of association that accurately assess the strength of the relationship between the two variables are discussed below. The following is not an exhaustive list of all possible measures of association but rather the most commonly used and practically useful measures. Most of these measures are available on most statistical software packages. The mathematical formulas and more extensive details for these measures can be found in many statistics texts; three basic references for such measures are Goodman and Kruskal,[1] Liebetrau,[2] and Khamis.[3]

1. CONTINUOUS-CONTINUOUS

Consider two continuous variables. In most instances, the *Pearson correlation coefficient* (also called the *Pearson product-moment correlation coefficient*), *r*, is appropriate for measuring the strength of the linear relationship between them. The value of *r* lies between –1 and +1. Values close to –1 indicate a strong negative linear relationship (as the value of one variable increases, the value of the other variable decreases; e.g., consider age and physical stamina for adult Americans—as people get older, they have less physical stamina in general). Values close to +1 indicate a strong positive linear relationship (as the value of one variable increases, so does the value of the other variable; e.g., consider years of experience and annual income among professional sonographers). Values of *r* close to zero indicate no linear relationship between the two variables. A value of *r* close to zero can occur (1) if the two variables are independent (i.e., knowledge of the value of one variable in no way improves the ability to predict the value of the other variable), (2) the two variables are highly variable, or (3) the two variables have a nonlinear relationship.

There is no universal rule for interpreting a given value of *r*, but informal guidelines have been given. For most studies involving medical, biomedical, biological, health care, sociological, educational, and psychological data, the following guidelines are appropriate:[4]

r	Interpretation of Linear Relationship
0.8	Strong positive
0.5	Moderate positive
0.2	Weak positive
0.0	No relationship
–0.2	Weak negative
–0.5	Moderate negative
–0.8	Strong negative

The above guidelines are generally in agreement with Cohen's[5] recommended guidelines:

$|r| < 0.3 \rightarrow$ Weak relationship
$0.3 \le |r| \le 0.5 \rightarrow$ Moderate relationship
$|r| > 0.5 \rightarrow$ Strong relationship

Suppose one or both of the continuous variables of interest have extreme values (sometimes called outliers). Annual income may be such a variable. For example, although the vast majority of workers within a given organization have annual incomes relatively close to the average annual income, there are a few individuals with positions at the highest levels of the administrative hierarchy (CEO, president, vice presidents, etc.) whose incomes are extremely high. In this case, a more appropriate measure of a linear relationship is the *Spearman rank correlation coefficient*. This is especially true if a test of the statistical significance of the strength of the relationship is desired. Its values also range from –1 to +1, and it is interpreted in the same way as *r*.

For further reading on these coefficients, refer to almost any standard statistics text—for example, Sokal and Rohlf[6] or Zar.[7]

Example 1. In a study to determine whether left atrial size, pressure, and ejection fraction are useful in diagnosing patients with left ventricle diastolic dysfunction through noninvasive means,[8] the Pearson correlation coefficient between (1) the left atrial pressure evaluated through pulmonary wedge pressure and (2) the E/A wave velocity ratio is $r = 0.77$. This can be characterized as a "strong" positive linear relationship between the two variables.

2. CONTINUOUS-ORDINAL

If one variable is continuous and the other is ordinal, then an appropriate measure of association is *Kendall's coefficient of rank correlation tau-sub-b,* τ_b. If the two variables are denoted by X (continuous) and Y (ordinal), then consider the levels of Y to be numerically coded according to the order of the levels (e.g., assign 1, 2, 3, . . . to the levels). Then Kendall's τ_b uses the numerical values of X and the coded numerical values of Y to render a number (coefficient) between –1 and +1 that measures the strength of relationship between X and Y. For further reading, see Liebetrau.[2]

If the ordinal variable, Y, has a large number of levels (say, five or six or more), then one may use the Spearman rank correlation coefficient to measure the strength of association between X and Y. In doing so, one must be careful in numerically coding the levels of Y in a practically meaningful way, keeping in mind that a metric is being imposed by the coding scheme. See Chatfield[9(p45)] and Luce and Narens[10] for further discussion. A typical example of this treatment is if Y represents degree of agreement, with the following levels: 1 = *very strongly agree*, 2 = *strongly agree*, 3 = *agree*, 4 = *neutral*, 5 = *disagree*, 6 = *strongly disagree*, and 7 = *very strongly disagree*. Another example involves income ranges where each level is coded with the midpoint between the lowest and highest income of the range (e.g., a range of $50,000–$75,000 would be coded 62,500).

The performance of the Spearman rank correlation coefficient is comparable to that of Kendall's τ_b, with the former being somewhat better for large sample sizes (see Zar[7(p392)]).

Example 2. A medical researcher who was a client at the Wright State University Statistical Consulting Center obtained data on 52 subjects. One of the questions posed in the survey was as follows: "Have you had a lot of energy in the last four weeks?" The possible responses were 1 = *none of the time*, 2 = *a little of the time*, 3 = *some of the time*, 4 = *a good bit of the time*, 5 = *most of the time*, and 6 = *all of the time*. This researcher wished to assess the strength of relationship between this ordinal variable and the age of the subject. For these data, Kendall's τ_b is 0.08. There appears to be no strong relationship between these two variables. The Spearman rank correlation coefficient for these data, with the coding given above, is 0.10, very close to the value of Kendall's τ_b, rendering the same conclusion. This illustrates the closeness of Kendall's τ_b with the Spearman rank correlation coefficient mentioned above.

3. CONTINUOUS-NOMINAL

If one variable is continuous and the other is nominal with just two categories, then use the *point-biserial correlation coefficient*. This coefficient ranges between –1 and +1. Values close to ±1 indicate a strong positive/negative relationship, and values close to zero indicate no relationship between the two variables. If the nominal variable has more than two levels, then one can calculate the point-biserial correlation between the continuous variable and all possible pairs of levels of the nominal variable; this would result in $\frac{k\,(k-1)}{2}$ such coefficients, where k represents the number of levels of the nominal variable. For further reading, see Tate.[11,12]

The calculation of the point-biserial correlation coefficient is accomplished by coding the two levels of the binary variable "0" and "1" and obtaining the Pearson correlation coefficient between the continuous variable and this coded binary variable.

Example 3. The following data show the absolute value of the difference in labor time from six hours (median labor time) for expectant mothers and whether they receive analgesia (Y, N); see Kotz and Johnson.[13(p279)]

Y	14.8	12.4	10.1	7.1	6.1	4.6	3.2	3.0	2.4	2.3	2.1
	0.8	0.1									
N	13.8	5.8	4.3	3.5	3.3	2.8	2.8	2.5	1.7	1.7	1.5
	1.3	1.3	1.2	1.2	1.1	0.7	0.6	0.5	0.2	0.2	

The point-biserial correlation coefficient for these data is 0.36. There is a "moderately" strong association between extreme labor time (either very short or very long) and the use of analgesia in expectant mothers. Specifically, receiving analgesia is associated with more extreme labor time.

4. ORDINAL-ORDINAL

If both variables are ordinal, then an appropriate measure of association is Kendall's τ_b. If both ordinal variables have a large number of levels, then an appropriate numerical coding scheme can be used and the Spearman rank correlation coefficient calculated. See the discussion of the continuous-ordinal case above.

Example 4. The following data come from a study of economic voting behavior by Kuklinski and West:[14]

	Expected Financial Well-Being		
	Better	Same	Worse
Present Financial Well-Being			
Better	70	85	15
Same	10	134	41
Worse	27	60	100

The value of Kendall's τ_b for these data is 0.39. This represents a "moderately" strong association between these two variables: better present financial well-being tends to correspond to better expected financial well-being.

It is interesting to note that, if the levels of each variable are coded 1, 2, and 3 for worse, same, and better, respectively, then the Spearman rank correlation coefficient is 0.42. This value is very close to the value for τ_b, once again illustrating the closeness of these two measures.

5. ORDINAL-NOMINAL

If one variable is nominal and the other is ordinal, then the *rank-biserial correlation coefficient* can be used to measure association. The same discussion as for the continuous-nominal case (see above) applies here. See Cureton.[15]

Example 5. In the study discussed in example 3, the rank-biserial correlation coefficient is 0.43. This value is somewhat larger than the point-biserial correlation coefficient of 0.36. Because there may be one or more outliers in the data set (e.g., 13.8 in the "N" group appears to be an outlier), the rank-biserial correlation coefficient is more appropriate. The general conclusion is the same as that of example 3: there is a moderately strong association between extreme labor time and use of analgesia.

6. NOMINAL-NOMINAL

Consider two discrete nominal variables whose association is of interest. Suppose that both variables have just two levels; the resulting data display is in the form of a two-by-two *contingency table*. One common way of measuring association in such a table is to use the *phi coefficient*, φ. Values of φ lie between 0 and 1. Values of φ close to 0 indicate very little association, and values close to 1 indicate nearly perfect predictability. Fleiss[16] provides, as a rule of thumb, that any value of φ less than 0.30 or 0.35 may be taken to indicate no more than trivial association. See Fleiss[16] for further discussion.

For two nominal variables in which at least one of the variables has more than two levels, a useful measure of association is *Goodman and Kruskal's lambda*, λ. The value λ is the relative decrease in the probability of error in guessing the level of one of the variables as between the level of the other variable known and unknown. The value λ lies between 0 and 1; values close to 1 correspond to a strong association. For further details, see Goodman and Kruskal.[1]

Example 6. The following contingency table represents a cross-classification of hair color and eye color in males (see Kendall[17(p300)]). Is there an association between hair and eye color in males?

	Hair Color			
	H_1	H_2	H_3	H_4
Eye color				
E_1	1768	807	189	47
E_2	946	1387	746	53
E_3	115	438	288	16

The value of λ for these data is 0.21. The reduction in the probability of error in predicting the level of one factor is 0.21 by knowing the level of the other factor, compared with not knowing the level of the other factor. That is, you can eliminate about 20% of your errors in predicting the level of one of the factors if you know the level of the other factor.

Reliability of the Estimated Coefficient

Because a given coefficient of correlation or association is calculated from a sample, it is measured with a certain margin of error. Generally, the smaller the sample size, the larger the margin of error. Your statistical software may provide a standard error (SE) along with the estimate of the coefficient. Very often, the margin of error can be calculated as approximately twice the standard error. Then, it can be concluded with approximately 95% confidence that the interval,

estimated coefficient ± 2*SE,

contains the "true" or "population" coefficient value. See Goodman and Kruskal[1] for further details.

By the true or population value, we mean the following: consider a population of subjects that is under study, such as all 50- to 60-year-old Caucasian Americans. In practice, one obtains a random sample from this population (e.g., randomly select 100 subjects from the population), obtains values of two variables of interest (e.g., age and cholesterol level), and then calculates the coefficient of interest (e.g., the Pearson correlation coefficient). This value is the *estimated coefficient*. If the coefficient value had been calculated from the entire population instead of only 100 randomly selected subjects, then the true or population value of the coefficient would have been obtained.

Example 7. For the data in example 2, where the sample size is n = 52, the SE of the estimated Kendall's τ_b is 0.0991. Then, with 95% confidence, the interval 0.08 ± 0.1982 or [–0.12, 0.28] contains the population coefficient value. Because this interval contains zero, we cannot be highly confident that this coefficient is not zero. In fact, the data do not adequately support the claim that the coefficient differs from zero. Consequently, we cannot conclude that age is related to "recent energy level" for the population from which these 52 individuals were selected.

Example 8. For the data of example 4, the SE of Kendall's τ_b coefficient is 0.04. The 95% confidence interval is approximately 0.39 ± 0.08 or [0.31, 0.47]. There is strong statistical evidence that this coefficient is not zero because zero is not contained in the interval. Note that the SE is much smaller here than for the data in example 7; that is because the sample size is much larger: n = 542.

Summary

In this article, appropriate measures of association have been provided based on the types (or levels of measurement) of the variables involved. For any given situation, there may be a wide variety of measures of association to choose from. The recommendations given in this article can be summarized as follows:

	Variable *X*		
Variable *Y*	Nominal	Ordinal	Continuous
Nominal	φ or λ	Rank biserial	Point biserial
Ordinal	Rank biserial	τ_b or Spearman	τ_b or Spearman
Continuous	Point biserial	τ_b or Spearman	Pearson or Spearman

φ = phi coefficient, λ = Goodman and Kruskal's lambda, τ_b = Kendall's τ_b.

For any pair of variables, say *X* and *Y*, whose association is to be studied, identify the type of variable for each (namely, nominal, ordinal, or continuous) and choose the measure of association by referring to the table above. For example, if one variable is discrete ordinal and the other is continuous (it does not matter which variable is called *X* and which is called *Y*), then an appropriate measure for assessing the strength of association is Kendall's τ_b or the Spearman correlation coefficient—notice the intersection of the row corresponding to "ordinal" and the column corresponding to "continuous" or, equivalently, the row corresponding to "continuous" and the column corresponding to "ordinal." For more details about the selection of the measure of association for this case, see the discussion under "continuous-ordinal" above.

Several cautionary notes need to be made as follows.

- Establishment of a strong relationship between two variables does not necessarily imply a *cause-effect relationship*. Although a strong association is *necessary* for establishment of a cause-effect relationship, it is not *sufficient*.
- Establishment of a strong relationship between two variables does not necessarily imply *agreement* between the two variables, nor does it necessarily imply high *reliability*. The notions of agreement and reliability are quite different from association. For further reading about agreement between two variables, see Bland and Altman[18] and Cohen[19]; for reliability, see Fleiss.[20]
- When assessing the strength of association between two variables, it is important to adjust for the effects of important confounders. This can be done with the use of *partial* correlation coefficients. See Fisher and van Belle[21] for further discussion.
- If one of the two variables under study is antecedent to the other, or there is an a priori cause-effect relationship, or there is a dependent-independent variable structure, then the variables are handled *asymmetrically*. In this case, prediction is often the research goal, and special measures of association are recommended. See Goodman and Kruskal[1] for further details.
- When nominal or ordinal variables are involved, then there may be several *ties* in the data. That is, several subjects may have the same *X* and *Y* values. For instance, in example 6, many subjects (1768) had hair color H_1 and eye color E_1. There are several different procedures for handling ties, so it is possible for a given coefficient to give rise to slightly different values for the same data set. Most statistical software packages will have a default procedure for handling ties.

Conclusion

It is important that the measure or coefficient used to assess the strength of association between two variables is appropriate for the data involved. A coefficient that is designed to measure the strength of association between two continuous variables, such as the Pearson correlation coefficient, should not be used to assess the strength of relationship between two ordinal variables or between an ordinal variable and a nominal variable, for example. This article provides a simple way of choosing an appropriate measure of association for a given pair of variables by classifying the variables according to their levels of measurement. By using these recommendations, it is hoped that research in sonography, as well as in other areas of medical research, will lead to more accurate and more reliable results.

References

1. Goodman LA, Kruskal WH: *Measures of Association for Cross Classifications*. New York, Springer-Verlag, 1979.
2. Liebetrau AM: *Measures of Association*. Beverly Hills, CA, Sage, 1983.
3. Khamis HJ: Measures of association, in Armitage P, Colton T (eds): *Encyclopedia of Biostatistics*. 2nd ed. New York, John Wiley, 2004.
4. Newton RR, Rudestam KE: *Your Statistical Consultant*. Thousand Oaks, CA, Sage, 1999.
5. Cohen J: *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ, Lawrence Erlbaum, 1988.
6. Sokal RR, Rohlf FJ: *Biometry*. New York, W. H. Freeman, 1995.
7. Zar JH: *Biostatistical Analysis*. Upper Saddle River, NJ, Prentice Hall, 1996.
8. Valentine AL, Pope J, Read T: Evaluation of left atrial size and pressure as it relates to left ventricle diastolic measurements in patients with left ventricle hypertrophy. *J Diagn Med Sonography* 2003;19:73–79.
9. Chatfield C: *Problem Solving: A Statistician's Guide*. 2nd ed. New York, Chapman & Hall, 1995.
10. Luce RD, Narens L: Measurement scales on the continuum. *Science* 1987;236:1527–1532.

11. Tate RF: Correlation between a discrete and a continuous variable, point-biserial correlation. *Ann Math Stat* 1954;25:603–607.
12. Tate RF: The theory of correlation between two continuous variables when one is dichotomized. *Biometrika* 1955;42:205–216.
13. Kotz S, Johnson NL: *Encyclopedia of Statistical Sciences*. Vol. 1. New York, John Wiley, 1982.
14. Kuklinski JH, West DM: Economic expectations and voting behavior in United States House and Senate elections. *Am Polit Sci Rev* 1981;75:436–447.
15. Cureton EE: Rank-biserial correlation. *Psychometrika* 1956;21:287–290.
16. Fleiss JL: *Statistical Methods for Rates and Proportions*. 2nd ed. New York, John Wiley, 1981.
17. Kendall MG: *The Advanced Theory of Statistics*. London, Charles Griffin, 1948.
18. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.
19. Cohen J: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
20. Fleiss JL: *The Design and Analysis of Clinical Experiments*. New York, John Wiley, 1986.
21. Fisher LD, van Belle G: *Biostatistics, a Methodology for the Health Sciences*. New York, John Wiley, 1993.