Behavior Research Methods, Instruments, & Computers
1996, 28 (1), 128-130

Program Abstracts/Algorithms

A Turbo Pascal program for the computation of scale-dependent association coefficients

STEF DECOENE
Katholieke Universiteit, Leuven, Belgium

A Turbo Pascal program is described that computes scale-dependent association coefficients belonging to Zeger's general family of association coefficients. Members of this family can be constructed that are adapted to the scale type of the variables, in that they are insensitive to admissible transformations of the variables but sensitive to inadmissible transformations. The program can compute (chance-corrected) association coefficients between variables of eight different scale types and between variables of mixed scale type.

In the literature, a large number of association coefficients have been described, all with their particular advantages and disadvantages (see, e.g., Suen & Ary, 1989; Vegelius, 1978). However, it would be quite useful if coefficients, differing in appropriateness with respect to scale types, would be otherwise equivalent. In this way, both the comparison of coefficients included in the family and the development of new members would be made easier. One such a set of association coefficients has been described by Janson and Vegelius (1982; Vegelius, 1978), given a systematic treatment in Zegers (1986a, 1986b, 1991; Zegers & Ten Berge, 1985, 1986), and extended by Stine (1989).

The general format for coefficients belonging to this family is

$$1-\frac{2\sum_{i=1}^{N}(X_i-Y_i)^2}{\sum_{i=1}^{N}X_i^2+\sum_{i=1}^{N}Y_i^2}, \tag{1}$$

which is equivalent to the computationally simpler

$$\frac{2\sum_{i=1}^{N}X_iY_i}{\sum_{i=1}^{N}X_i^2+\sum_{i=1}^{N}Y_i^2}. \tag{2}$$

In Formula 1, the identity of the variables X and Y is expressed as their mean squared difference, which is 0 if the variables are identical and becomes larger the more the variables differ. By subtracting the sum of squared differences from +1, the coefficient is limited to +1, in the case of identity. By dividing the mean squared difference between X and Y by the denominator given in Formula 1, the coefficient cannot attain values less than -1 in the case of reversed identity. Because Formula 1 satisfies the requirements of a scalar product between normalized vectors in a Euclidean space (Vegelius, 1978), members belonging to this family are called *E-coefficients*.

The family of association coefficients based on Formula 1 has several properties that make members quite interesting as measures of correlation and interobserver agreement. Association coefficients derived from Formula 1 are insensitive to the order of associated variables (*symmetry*) and have a fixed minimum and maximum (i.e., they are restricted to a *bounded* interval; cf. above). If the association between variables X and Y is perfect and the association between Y and Z is perfect, then the association between X and Z will be perfect (*transitivity*). Moreover, association of a variable with itself is perfect. If two variables are statistically independent, then the association is zero. Finally, there is an additional characteristic that is of special practical importance. Given a set of variables between which association coefficients are computed, it can be shown that the corresponding association coefficient matrix is positive semidefinite (*Gramian*). Although a discussion of this characteristic is not appropriate here (see Zegers, 1986b, for discussion and proof that coefficients based on Formula 1 are Gramian), it is important to note its consequences. If a coefficient matrix is Gramian it can be used in various multivariate techniques, such as principal component analysis. Moreover, the association matrix can be converted into a Euclidean distance matrix by computing for each association value

$$d_{XY} = \sqrt{1-g_{XY}}, \tag{3}$$

with d_{XY} the distance between variables X and Y, and g_{XY} the association between X and Y. In this way, a Gramian coefficient matrix can be used in metric multidimensional scaling techniques.

Scale-Dependent Modifications of the General Association Coefficient

Formula 1 is a coefficient of absolute identity and should therefore be used when the variables are absolute scales. However, by defining X and Y in Formula 1 as uniforming transforms of one another, the formula can be modified to express relational agreement and scale-dependent correlation in general. The term *relational agreement* was coined by Stine (1989) to express the identity between values of two variables up to admissible transformations, as

The author is a postdoctoral fellow of the Beglian National Fund of Scientific Research. Help from National Fund of Scientific Research Grant S 2/5–ID E 120 is gratefully acknowledged. Correspondence should be addressed to S. Decoene, Centrum voor Mathematische Psychologie en Psychologische Methodologie, Tiensestraat 102, 3000 Leuven, Belgium (e-mail: stef.decoene@psy.kuleuven.ac.be).

defined in measurement theory (e.g., Krantz, Luce, Suppes, & Tversky, 1971; Roberts, 1979). That is, these coefficients measure the degree to which judges attribute the same empirical structure to the task at hand or the degree to which two variables with the same kind of structure correlate. In particular, E-coefficients can be constructed that are (1) insensitive to admissible transformations of the variables but (2) sensitive to inadmissible transformations (Stine, 1989; Zegers, 1986b; Zegers & Ten Berge, 1985).

The modification of Formula 1 to take the scale type of the variables into account consists of specifying uniforming transforms that "change the two variables so that they approximate one another as much as possible without altering what they indicate about the empirical events that they describe" (Stine, 1989, p. 344). Zegers and Ten Berge derived uniforming transforms for the absolute, ratio, additive, and interval scales (Zegers & Ten Berge, 1985; see also Zegers, 1986a), for nominal scales (Zegers & Ten Berge, 1986), and for cases in which interval, ordinal, and nominal scales are mixed (Zegers & Ten Berge, 1986). Stine (1989) provides additional uniforming transforms for log-ratio, log-interval, and ordinal scales.

With respect to the case of mixed scale types, the formula can deal with combinations of interval, ordinal, and nominal scales. Because it is an E-coefficient, it can be used in multivariate and multidimensional analysis techniques. Consequently, the use of the scale-dependent E-coefficients allows one to compute an association matrix based on variables of differing scale type and, at the same time, allows one to use analysis techniques commonly thought to be restricted to interval scales.

Scale-Dependent Coefficients Corrected for Chance

Besides the basic scale-dependent association coefficients, Zegers and Ten Berge (1985, 1986; see also Zegers, 1986a) provided a way to correct them for chance association. To do this, they used the method proposed by Cohen (1960), in which an observed association value was corrected by comparing it with the maximal possible value of association and the association to be expected by chance, combined into Formula 4:

$$g'' = \frac{g_o - g_c}{g_m - g_c}, \tag{4}$$

with g_o the observed association, g_m the maximal possible value of association, and g_c the association to be expected by chance.

Following this logic, it can be shown (Zegers & Ten Berge, 1985; see also Zegers, 1986a) that the general chance-corrected coefficient in Formula 1 is

$$c^{g_{XY}} = \frac{2\left(\sum_{i=1}^{N} X_i Y_i - N^{-1}\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i\right)}{\sum_{i=1}^{N} X_i^2 + \sum_{i=1}^{N} Y_i^2 - 2N^{-1}\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i} . \tag{5}$$

SDAC: A Turbo Pascal Program to Compute Scale-Dependent Euclidean Association Coefficients

SDAC computes association coefficients for variables of eight different scale types and for variables of mixed scale type (interval, ordinal, or nominal). The scale types included are the absolute, ratio, log-ratio, additive, interval, log-interval, ordinal, and nominal scales. The formulas are taken from Stine (1989) and Zegers (1986a). Computation of log-interval scaled variables should be done with care. Unusual values can occur, and this could possibly be due to the fact that the dispersion measure, as defined in Stine (1989) and computed in SDAC with the function GEOMEANDEV, is not robust enough. Although SDAC ignores this problem, it may be advisable to transform log-interval scales into interval scales (both contain equivalent information in the measurement theoretical sense). For nominal scales, the program function is based on a formula that does not presuppose that the variables have an equal number of categories. Chance-corrected versions of these association coefficients can also be computed.

The program was written in Turbo Pascal Version 6, but it does not use any special features not already present in previous versions. There is one version that requires a mathematical coprocessor, and one version that does not. An ordinary PC/AT 286 with 640K suffices.

The data should be written into an ASCII file and cannot be larger than 100 × 100. The balance between number of judges and number of variables can be changed in the code (to, e.g., a 150 × 50 matrix), as long as the data matrix does not contain more than 10,000 data points. The program assumes that the columns are the variables/judges, the rows are the stimuli/items, and the cells are real-valued. In addition, one should also write a data definition file that indicates the scale type of each variable on the first row and the number of categories on the second row, in case nominal scales are included.

When the program is started, one is asked to give the name of the data file and its structure (number of rows and colums). Next, the program will ask you to define the data set as either uniform or mixed. A uniform matrix is defined as one in which all variables are of the same scale type and which does not contain any nominally scaled variables. A mixed matrix is defined as one in which different scale types, including nominal ones, are present.

If one works with a uniform matrix, there are three possibilities in computing association coefficients: (1) selection of two variables, (2) computation of the association matrix, and (3) computation of a split-half association coefficient. In each case, you are asked to select an association coefficient (chance-corrected or not). In the second case, one can ask either to write the association values to an output file or to transform the association values to distance measures (in case you would like to use the output in, for example, multidimensional scaling).

In the third case, the built-in RANDOMIZE procedure of Turbo Pascal is used to select a random half of the variables. Then, variables in each of the two groups are aggregated, and the two aggregates are defined as the new variables. The aggregation is done in one of three possible ways. If the scale type of the variables as defined in the data definition file is ratio, additive, or power, the geometric mean is used to average over variables—for ratio scales, the geometric mean is invariant in the measurement theoretic sense (Aczel & Roberts, 1989; Narens & Luce, 1993). If the variables are interval, log-interval, or absolute, the arithmetic mean is used. If the variables are ordinal, the median is used as the merging function. The split-half option cannot be used if nominal variables are involved.

If the data matrix is of a mixed type, only the first two options are available. If two variables are of the same scale type, the appropriate chance-corrected association coefficient is computed. If they are of a different scale type, the following procedure is used: (1) If the two variables involved are both nominal, the appropriate nominal scale association coefficient is used. (2) If the two variables involved are nominal and ordinal, or nominal and interval or higher, the mixed scale association coefficient is used. (3) If the two variables are ordinal and interval or higher, then the interval-scaled variable is rank ordered, and the ordinal scale association coefficient is computed. (4) If one variable is additive or power and the other is ratio, then the additive/power variable is transformed to a ratio scale (by taking the logarithm), and the ratio scale association coefficient is computed. (5) If one variable is log-interval and the other is interval, the log-interval variable is transformed to an interval scale (by exponentiation), and the interval scale association coefficient is computed. (6) For all other combinations, the interval scale association coefficient is used.

Availability

The computer program, including the executable and source file, can be obtained by writing to Stef Decoene, Centrum voor Mathematische Psychologie, Tiensestraat 102, 3000 Leuven, Belgium, and including $25 for copying and shipping.

REFERENCES

Aczel, J., & Roberts, F. (1989). On the possible merging functions. *Mathematical Social Sciences*, **17**, 205-243.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement*, **20**, 37-46.

Janson, S., & Vegelius, J. (1982). Correlation coefficients for more than one scale type. *Multivariate Behavioral Research*, **17**, 255-269.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement. Vol. 1. Additive and polynomial representation.* New York: Academic Press.

Narens, L., & Luce, R. D. (1993). Further comments on the "nonrevolution" arising from axiomatic measurement theory. *Psychological Science*, **4**, 127-130.

Roberts, F. (1979). *Measurement theory*. Reading, MA: Addison-Wesley.

Stine, W. W. (1989). Interobserver relational agreement. *Psychological Bulletin*, **106**, 341-347.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data*. Hillsdale, NJ: Erlbaum.

Vegelius, J. (1978). On the utility of the E-correlation coefficient concept in psychological research. *Educational & Psychological Measurement*, **38**, 605-611.

Zegers, F. E. (1986a). A family of chance-corrected association coefficients for metric scales. *Psychometrika*, **51**, 559-569.

Zegers, F. E. (1986b). *A general family of association coefficients*. Groningen: Boomker.

Zegers, F. E. (1991). Coefficients for interrater agreement. *Applied Psychological Measurement*, **15**, 321-333.

Zegers, F. E., & Ten Berge, J. M. F. (1985). A family of association coefficients for metric scales. *Psychometrika*, **50**, 17-24.

Zegers, F. E., & Ten Berge, J. M. F. (1986). Correlation coefficients for more than one scale type: An alternative to the Janson and Vegelius approach. *Psychometrika*, **51**, 549-557

(Manuscript received September 6, 1994;
revision accepted for publication January 9, 1995.)