PSYCHOMETRIKA—VOL. 51, NO. 4, 559–562
DECEMBER 1986

A FAMILY OF CHANCE-CORRECTED ASSOCIATION COEFFICIENTS FOR METRIC SCALES

FRITS E. ZEGERS

UNIVERSITY OF GRONINGEN

A chance-corrected version of the family of association coefficients for metric scales proposed by Zegers and ten Berge is presented. It is shown that a matrix with chance-corrected coefficients between a number of variables is Gramian. The members of the chance-corrected family are shown to be partially ordered.

Key words: association coefficient, metric scale, product-moment correlation.

A family of association coefficients for metric scales was derived by Zegers and ten Berge (1985), with coefficients of the form

$$g(X, Y) = \frac{2 \sum U_{ix} U_{iy}}{\sum U_{ix}^2 + \sum U_{iy}^2}, \tag{1}$$

where, as in the remainder of the present paper, summation runs from $i = 1$ to N, and the symbol U denotes a "uniforming" transformation of the original variables. The form of the transformation U depends on the scale type of the variables. For each of the four metric scale types the general formula (1) results in a specific association coefficient.

In the present paper a chance-corrected version of the g-coefficient defined by (1) will be proposed. The matrix with g-coefficients between a number of variables is Gramian (positive semi-definite; Zegers & ten Berge, 1985). It will be shown that the chance-corrected g-coefficient has the very same property. It will also be shown that as a result of the optimality properties of the uniforming transformations the four members of the chance corrected g-coefficient family are partially ordered.

The Chance-Corrected g-Coefficient

A well-known approach to chance-correction of an association or agreement coefficient A is to define

$$A' = \frac{A_0 - A_c}{A_m - A_c}, \tag{2}$$

where A' is the chance-corrected coefficient, A_0 is the observed coefficient, A_c is the coefficient under chance and A_m is the maximum value of the coefficient regardless of the marginals. Cohen's kappa (Cohen, 1960) is an example of a coefficient of this type.

Given the sample distributions of X and Y, and therefore of U_x and U_y, the value of $g(X, Y)$ under chance, $g_c(X, Y)$, is by definition the expectation of $g(X, Y)$ over all permutations of the scores on X. The denominator of (1) is invariant under permutations of this kind. The expectation of the numerator of (1) over permutations of the X-scores is given

The author is obliged to Jos ten Berge for helpful comments.

Requests for reprints should be sent to Frits E. Zegers, University of Groningen, Department of Psychology, Grote Markt 31/32, 9712 HV Groningen, THE NETHERLANDS.

0033-3123/86/1200-7052$00.75/0

by

$$\underset{p}{E}\,(2\sum U_{ix}U_{iy}) = 2N^{-1}\sum U_{ix}\sum U_{iy} \tag{3}$$

(Kendall & Stuart, 1961, p. 474). This yields

$$g_c(X, Y) = \frac{2N^{-1}\sum U_x \sum U_y}{\sum U_x^2 + \sum U_y^2}, \tag{4}$$

where the index i has been dropped for convenience. Correcting the g-coefficient according to (2), with one as the maximum value of g and (4) as the value under chance, we obtain

$$g'(X, Y) = \frac{2(\sum U_x U_y - N^{-1}\sum U_x \sum U_y)}{\sum U_x^2 + \sum U_y^2 - 2N^{-1}\sum U_x \sum U_y}. \tag{5}$$

It can easily be verified that for the difference scale and the interval scale the corrected g-coefficients are identical to the corresponding g-coefficients; that is, for the additivity coefficient (a_{xy}) and the product-moment correlation (r_{xy}), respectively.

For the absolute scale the corrected identity coefficient is

$$\begin{aligned} e'_{xy} &= \frac{2(\sum X_i Y_i - N^{-1}\sum X_i \sum Y_i)}{\sum X_i^2 + \sum Y_i^2 - 2N^{-1}\sum X_i \sum Y_i} \\ &= \frac{2S_{xy}}{S_x^2 + S_y^2 + (M_x - M_y)^2}, \end{aligned} \tag{6}$$

where S_{xy} is the sample covariance of X and Y, S_x^2 and S_y^2 are the sample variances, and M_x and M_y are the sample means of X and Y, respectively. It may be noted that e'_{xy} is identical to the coefficient of equality proposed by Jobson (1976).

For ratio scales the corrected proportionality coefficient is

$$\begin{aligned} p'_{xy} &= \frac{\sum X_i Y_i - N^{-1}\sum X_i \sum Y_i}{(\sum X_i^2 \sum Y_i^2)^{1/2} - N^{-1}\sum X_i \sum Y_i} \\ &= \frac{S_{xy}}{(N^{-1}\sum X_i^2)^{1/2}(N^{-1}\sum Y_i^2)^{1/2} - M_x M_y}. \end{aligned} \tag{7}$$

In order to prove that the matrix with g'-coefficients between a number of variables is Gramian, a lemma on Gramian matrices will be proved first.

A Lemma on Gramian Matrices

Lemma. If Z is a Gramian matrix of order k, with elements z_{ij} in the range $(-1, 1)$, $i, j = 1, 2, \ldots, k$, then also the $k \times k$ matrix Z^* with elements $z^*_{ij} = (1 - z_{ij})^{-1}$, $i, j = 1, 2, \ldots, k$, is Gramian.

Proof. By applying the binomial expansion to the elements of Z^* (Abramowitz & Stegun, 1972, Equation 3.6.10), it is clear that Z^* is an infinite sum of $k \times k$ matrices $Z^{(p)}$ with elements z_{ij}^p, $p = 0, 1, \ldots$ Obviously, $Z^{(0)} = 11'$ is Gramian. The matrix $Z^{(1)} = Z$ is Gramian by definition, and for $p > 1$, $Z^{(p)}$, being the Hadamard (element-wise) product of Gramian matrices, is Gramian (Schur, 1911). Therefore, Z^*, being an infinite sum of Gramian matrices, is Gramian.

A Matrix With g'-Coefficients is Gramian

In order to prove that the $k \times k$ matrix G with g'-coefficients between k variables is Gramian it will be shown that G can be expressed as two times the Hadamard product of three Gramian matrices G_1, G_2 and G_3.

Define G_1 as the matrix with h, j-th element $(\sum U_h^2 + \sum U_j^2)^{-1}$, G_2 as the matrix with h, j-th element $S_{U_h U_j}$ and G_3 as the matrix with h, j-th element $(1 - t_{jh})^{-1}$, with

$$t_{jh} = \frac{2N^{-1} \sum U_h \sum U_j}{\sum U_h^2 + \sum U_j^2}, \tag{8}$$

then it can be verified that $G = 2G_1 * G_2 * G_3$, where $*$ denotes the Hadamard product. Matrix G_1 is Gramian (Zegers & ten Berge, 1985), and G_2, being a symmetric covariance matrix, is Gramian as well. It can be verified that the $k \times k$ matrix with elements t_{jh} is Gramian and that $t_{jh}^2 < 1$. Therefore, G_3 is a matrix of the type Z^* defined in the Lemma presented above, which shows that G_3 is Gramian. This concludes the proof of the Gramian property of G.

The matrix G is Gramian with unit diagonal entries. This means that the g'-coefficient is an E-correlation coefficient in the terminology of Vegelius (1978) and Janson & Vegelius (1978).

Optimality Properties of the Uniforming Transformations

The uniforming transformations U defined by Zegers & ten Berge (1985) were chosen for reasons of convenience. It can be shown, however, that for each of the four metric scales the uniforming transformation maximizes the absolute value of the g'-coefficient within the class of permissible transformations.

To verify the optimality properties of the uniforming transformations, consider the general linear transformations $U_x = b_x X + a_x$ and $U_y = b_y Y + a_y$. Using these transformations in (5), we get, after some algebra

$$g'(X, Y) = \frac{2b_x b_y S_{xy}}{b_x^2 S_x^2 + b_y^2 S_y^2 + \{b_x(M_x + a_x) - b_y(M_y + a_y)\}^2}. \tag{9}$$

The constants a_x and a_y only appear in the last (squared) term of the denominator of (9). Clearly, for every choice of b_x and b_y the absolute value of $g'(X, Y)$ is maximized by taking $a_x = -M_x$ and $a_y = -M_y$. Therefore, if the scale type permits an additive transformation, centering the variables around zero is optimal.

Furthermore, rearranging the terms in (9) yields

$$g'(X, Y) = \frac{2S_{xy}}{\{(b_x b_y^{-1})^{1/2} T_x - (b_x^{-1} b_y)^{1/2} T_y\}^2 + 2T_x T_y - 2(M_x + a_x)(M_y + a_y)}, \tag{10}$$

with $T_x^2 = S_x^2 + (M_x + a_x)^2$ and T_y^2 defined in a parallel fashion. The constants b_x and b_y only appear in the first (squared) term of the denominator of (10). For every choice of a_x and a_y the absolute value of $g'(X, Y)$ is maximized by taking $(b_x b_y^{-1})^{1/2} T_x = (b_x^{-1} b_y)^{1/2} T_y$, which can be simplified to $b_x T_x = b_y T_y$, and hence $b_x = cT_x^{-1}$ and $b_y = cT_y^{-1}$ for arbitrary $c > 0$. If the scale type permits rescaling of the variables the uniforming transformation defined by Zegers and ten Berge (1985) implies taking $b_x = T_x^{-1}$ and $b_y = T_y^{-1}$, which is an optimal choice.

As a result of the optimizing properties of the uniforming transformations the four g'-coefficients are partially ordered. Both the additivity coefficient $a'_{xy} = a_{xy}$ and the corrected proportionality coefficient p'_{xy} exceed or equal (in absolute value) the corrected identity coefficient e'_{xy}. The product-moment correlation exceeds or equals (in absolute value) the other three members of the g'-coefficient family.

Discussion

Zegers and ten Berge (1985) advocated using the identity coefficient e_{xy} as a measure of profile similarity when both levels and dispersions are to be respected. Obviously, the corrected identity coefficient e'_{xy} is another candidate for this situation. The two coef-

ficients, however, take account of level in different ways. The e-coefficient is affected by any change in level of X and/or Y, whereas the e'-coefficient is only affected by differential changes of the X and Y levels. It can be concluded that e'_{xy} should be chosen as a coefficient of profile similarity when dispersions and differences in level are to be respected, whereas e_{xy} should be chosen when dispersions are to be respected as well as levels in an absolute sense. Similar considerations play a role in choosing a coefficient to express the degree of interrater agreement.

Another difference between e_{xy} and e'_{xy} is important in the context of profile similarity. Like the H-coefficient of profile similarity (Sjöberg & Holley, 1967), which is identical to p_{xy}, the identity coefficient e_{xy} is insensitive to changing the polarity (sign) of one or more variables in the profiles. It can readily be verified that e'_{xy} is sensitive to such changes in polarity.

Tucker's congruence coefficient (Tucker, 1951) is often used as a measure of factorial invariance. The congruence coefficient has been criticized because it cannot be zero when both factors have only positive loadings (Pinneau & Newhouse, 1964, p. 275). The corrected proportionality coefficient p'_{xy} does not suffer from this shortcoming. However, p'_{xy} has two drawbacks as a measure of factorial invariance. Firstly, the size of p'_{xy} is affected by changing the sign of one of the columns of factor loadings and secondly p'_{xy} is sensitive to changing the sign of one or more of the variables.

References

Abramowitz, M., & Stegun, A. (1972). *Handbook of mathematical functions.* New York: Dover.

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

Janson, S., & Vegelius, J. (1978). *On constructions of E-correlation coefficients,* (Research Report 78-11). Uppsala: University of Uppsala, Department of Statistics.

Jobson, J. D. (1976). A coefficient for questionnaire items with interval scales. *Educational and Psychological Measurement, 36,* 271–274.

Kendall, M. G., & Stuart, A. (1961). *The advanced theory of statistics* (Vol. I). London: Griffin.

Pinneau, S. R., & Newhouse, A. (1964). Measures of invariance and comparability in factor analysis for fixed variables. *Psychometrika, 29,* 271–281.

Schur, J. (1911). Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen [Some notes on the theory of bounded bilinear forms with an infinite number of variables]. *Journal für die reine und angewandte Mathematik, 140,* 1–28.

Sjöberg, L., & Holley, J. W. (1967). A measure of similarity between individuals when scoring directions are arbitrary. *Multivariate Behavioral Research, 2,* 377–384.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington DC: Department of the Army.

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

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