PSYCHOMETRIKA—VOL. 50, NO. 1, 17–24
JUNE, 1985

A FAMILY OF ASSOCIATION COEFFICIENTS FOR METRIC SCALES

FRITS E. ZEGERS AND JOS M.F. TEN BERGE
UNIVERSITY OF GRONINGEN

Four types of metric scales are distinguished: the absolute scale, the ratio scale, the difference scale and the interval scale. A general coefficient of association for two variables of the same metric scale type is developed. Some properties of this general coefficient are discussed. It is shown that the matrix containing these coefficients between any number of variables is Gramian. The general coefficient reduces to specific coefficients of association for each of the four metric scales. Two of these coefficients are well known, the product-moment correlation and Tucker's congruence coefficient. Applications of the new coefficients are discussed.

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

The choice of an association (or similarity) coefficient between two variables depends on the scale type to which the variables belong. The scale type of a variable is defined by the class of admissible transformations. An admissible transformation is a transformation under which the variable remains invariant. In fundamental measurement the invariance properties are established by the uniqueness theorem (cf. Krantz, Luce, Suppes & Tversky, 1971, p. 9). In a less formal context the scale type of a variable may be determined by the class of transformations which leave the meaning of the variable unaffected, or, in an applied setting, by the class of transformations under which decisions made on the basis of the variable remain the same.

An association coefficient between two variables has to be invariant under admissible transformations of the variables, cf. Janson & Vegelius (1982). To this we can add the requirement that an association coefficient has to be sensitive to non-admissible transformations of the variables. For example, rank ordering two interval scaled variables changes their character, which should have an effect on the value of their association coefficient. Used as an association coefficient between two interval scaled variables the product moment correlation (PMC) meets both (in)variance requirements, whereas the rank correlation fails to meet the second requirement.

Four types of metric scales are distinguished here. Best known is the interval scale, which is invariant up to positive linear (affine) transformations. The other three metric scales are the absolute scale, for which the identity transformation is the only admissible transformation, the ratio scale, which is only invariant under similarity (positive multiplicative) transformations, and the difference scale (cf. Suppes & Zinnes, 1963), which is only invariant under additive transformations.

Association coefficients between two variables which meet the two (in)variance requirements are known for interval and for ratio scales. These are, respectively, the PMC and Tucker's congruence coefficient (Tucker, 1951), originally proposed by Burt (1948). In this paper a general formula for association coefficients for metric scales will be proposed. This general formula reduces to the PMC and Tucker's congruence coefficient in the case of two interval scaled and two ratio scaled variables, respectively. For the other two metric scale types the general formula leads to appropriate association coefficients. These new coefficients cannot be expressed as product moments in any obvious way. Neverthe-

Requests for reprints should be sent to Frits E. Zegers, University of Groningen, Dept. of Psychology, Grote Markt 31-32, 9712 HV Groningen, The Netherlands.

0033-3123/84/0600-6003$00.75/0

less, the associated matrices containing these coefficients between any number of variables will be shown to be Gramian. Some applications of the new coefficients will be discussed.

A General Formula for Association Coefficients for Metric Scales

In developing the general formula for association coefficients for metric scales we assume that we have scores of n subjects (or objects) on two variables, X_i and X_j. An individual score is denoted by X_{ih} or X_{jh}, $h = 1, 2, \ldots, n$. X_i and X_j are variables belonging to one of the four metric scale types. The association coefficients will be defined as sample statistics. The problem of generalizing to population characteristics will be taken up in the discussion section.

Throughout the paper the following notation will be used: The mean score on X_i is denoted by M_i. The sample variance of X_i is denoted by S_i^2, and the mean squared value of X_i is denoted by T_i^2:

$$T_i^2 = \frac{\sum X_{ih}^2}{n}, \tag{1}$$

where the summation sign, as in the remainder of this paper, denotes summation over $h = 1$ to n.

The general formula will be based on some kind of standardized version of the variables. Such a standardized version should be invariant under all admissible transformations of the variables and it should be sensitive to non-admissible transformations. In this way it is ensured that the resulting association coefficient meets the two (in)variance requirements. Because the term "standardized" version of a variable usually refers to a version with a mean of zero and a variance of one we shall, to avoid ambiguity, refer to the special kinds of standardized versions used in this paper as "uniformed" versions. A uniformed version of a variable is obtained by applying a "uniforming" transformation to the variable. A uniforming transformation for a certain scale type belongs to the class of admissible transformations of that scale type. If an additive transformation is allowed the uniforming transformation will center the variable around zero, if a multiplicative transformation is allowed the uniforming transformation will rescale the variable to obtain a mean squared value of one. The constants zero and one have been chosen for reasons of convenience.

Let U_i denote the uniformed version of X_i, then the uniforming transformations are:

$$U_i = X_i, \tag{2a}$$

for the absolute scale;

$$U_i = X_i - M_i, \tag{2b}$$

for the additive scale;

$$U_i = T_i^{-1} X_i, \tag{2c}$$

for the ratio scale, and

$$U_i = S_i^{-1}(X_i - M_i), \tag{2d}$$

for the interval scale, where T_i and S_i denote the square root of T_i^2 and S_i^2, respectively. From (2d) it is clear that for the interval scale the uniformed version of a variable is identical to the usual standardized version.

Let g_{ij} denote the association coefficient between two variables X_i and X_j. As mentioned above, g_{ij} will be based on the uniformed versions U_i and U_j. Therefore, g_{ij} will be some function f of U_i and U_j:

$$g_{ij} = f(U_i, U_j). \tag{3}$$

An association coefficient between two variables may be defined as the extent to which their uniformed versions are identical. The mean squared difference of the uniformed versions is a common indicator for this identity. A function of the mean squared difference which has the desirable property of attaining a maximum value of one if $U_i = U_j$ is

$$f(U_i, U_j) = 1 - \frac{c \sum (U_{ih} - U_{jh})^2}{n}, \tag{4}$$

where c denotes some positive constant. The constant c can be uniquely determined by requiring

$$f(U_i, -U_j) = -f(U_i, U_j), \tag{5}$$

which means that reflecting one of the variables, which results in reflection of the corresponding uniformed version, will merely change the sign of the association coefficient. From (4) and (5) we have

$$1 - cn^{-1} \sum (U_{ih} + U_{jh})^2 = -1 + cn^{-1} \sum (U_{ih} - U_{jh})^2, \tag{6}$$

from which it follows that

$$c = (n^{-1} \sum U_{ih}^2 + n^{-1} \sum U_{jh}^2)^{-1}. \tag{7}$$

Combining (3), (4) and (7) yields

$$g_{ij} = 1 - \frac{\sum (U_{ih} - U_{jh})^2}{(\sum U_{ih}^2 + \sum U_{jh}^2)} = \frac{2 \sum U_{ih} U_{jh}}{(\sum U_{ih}^2 + \sum U_{jh}^2)}, \tag{8}$$

which is the general formula of association coefficients for metric scales.

For each of the four metric scale types, the proper uniforming transformation may be inserted into (8). This yields four special coefficients of association. Before discussing these coefficients separately, we will prove that a matrix containing g coefficients between any number of variables is Gramian (symmetric and positive semidefinite). Three lemmas which are needed for this proof will be presented first.

Lemma 1: The Hadamard (element-wise) product of two Gramian matrices is Gramian.

Proof: See Schur (1911, p. 14) or Browne (1977, p. 208).

The proof of a closely related theorem, stating that the Hadamard product of two symmetric positive definite (SPD) matrices is SPD, may be found in Bellman (1960, p. 94). This proof can be easily modified into another proof of Lemma 1.

Lemma 2: A symmetric matrix is Gramian if and only if all principal minors are nonnegative.

Proof: See Gantmacher (1958, p. 282).

Lemma 3: Let $x_1, \ldots x_i, x_j, \ldots x_k$ and $y_1, \ldots y_i, y_j, \ldots y_k$ be numbers satisfying $x_i + y_j \neq 0$, $i, j = 1, \ldots k$. Let W_k be the $k \times k$ matrix, $k \geq 2$, with elements $w_{ij} =$

$(x_i + y_j)^{-1}$, $i, j = 1, \ldots k$. Then the determinant of W_k is given by

$$\det(W_k) = \frac{\prod_{i<j}^{k}(x_i - x_j)(y_i - y_j)}{\prod_{i,j}(x_i + y_j)}. \tag{9}$$

Proof: See Pólya & Szegö (1925, p. 299).

A slightly different proof can be given by partitioning W_k as

$$W_k = \left[\begin{array}{c|c} W_{k-1} & w \\ \hline v' & \dfrac{1}{x_k + y_k} \end{array}\right] \tag{10}$$

and using a standard result on determinants

$$\det(W_k) = (x_k + y_k)^{-1} \det[W_{k-1} - (x_k + y_k)wv']. \tag{11}$$

After some algebra we obtain the recurrent formula

$$\det(W_k) = \frac{1}{(x_k + y_k)} \prod_{i=1}^{k-1} \frac{(x_i - x_k)(y_i - y_k)}{(x_i + y_k)(x_k + y_i)} \det(W_{k-1}), \tag{12}$$

from which Lemma 3 can be deduced.

The Matrix of g Coefficients is Gramian

Given a set of r variables, the g coefficients between the variables can be collected in a symmetric matrix G, of order r, with elements

$$g_{ij} = 2v_{ij}(v_{ii} + v_{jj})^{-1}, \tag{13}$$

where

$$v_{ij} = n^{-1} \sum (U_{ih} U_{jh}), \qquad i, j = 1, \ldots r. \tag{14}$$

This matrix G may be expressed as the Hadamard product

$$G = U * V, \tag{15}$$

where U denotes the symmetric matrix with elements v_{ij} and V is the symmetric matrix with elements $2(v_{ii} + v_{jj})^{-1}$.

By Lemma 1, G is Gramian if both U and V are Gramian. Clearly, U is Gramian, being a product-moment matrix. It will be shown that V is Gramian too.

Let V_k, $1 \le k \le r$, denote the symmetric submatrix of V, obtained by deleting all but the first k rows and columns of V. Then $\det(V_k)$ is the k-th leading principal minor of V. Because $v_{ii} = n^{-1} \sum U_{ih}^2 > 0$, we have

$$\det(V_1) = v_{11}^{-1} > 0. \tag{16}$$

For $2 \le k \le r$, V_k is a matrix of the type defined in Lemma 3, multiplied by 2, with $x_i = y_i = v_{ii}$. Therefore,

$$\det(V_k) = \frac{\prod_{i<j}(v_{ii} - v_{jj})^2}{\prod_{i,j}(v_{ii} + v_{jj})} \ge 0 \tag{17}$$

with equality iff $v_{ii} = v_{jj}$ for at least one pair (i, j). Equations (16) and (17) show that all leading principal minors of V are nonnegative. Because no use has been made of the numbering of the r variables, (16) and (17) hold for any permutation and renumbering of the variables. Therefore, it can be concluded that all principal minors of V are nonnegative, which, by Lemma 2, proves that V is Gramian.

The Gramian property of G is a sufficient condition for a representation of the variables in Euclidean space with distances proportional to $(1 - g_{ij})^{1/2}$, $i, j = 1, \ldots k$ (Gower, 1966), which implies that G can be used in various metric multivariate techniques.

The Gramian property of G together with the fact that all diagonal entries of G are one also implies that G can be considered as a product moment correlation matrix. Therefore the g coefficient has all the properties of the ordinary PMC. Some of these properties have been imposed in the process of developing g, i.c. $g_{ij} \leq 1$, $g_{ii} = 1$ and $g_{ij} = g_{ji}$. Other properties of the g coefficient which follow immediately from the fact that G is a correlation matrix are:

a. $g_{ij} \geq -1$.
b. if $g_{ij} = 1$, then $g_{im} = g_{jm}$, for every third variable m.
c. if $g_{ij} = -1$, then $g_{im} = -g_{jm}$, for every third variable m.

The Four Coefficients of Association for Metric Scales

Inserting the proper uniforming transformation (2) into the general formula (8), yields a coefficient of association for each of the four metric scales. These coefficients will now be derived separately.

The Coefficient of Identity

The association coefficient for absolute scales reflects the degree to which two variables are identical and, therefore, it will be called the coefficient of identity. Inserting the identity transformation (2a) into the general formula (8) yields

$$e_{ij} = 2 \sum X_{ih} X_{jh} \left(\sum X_{ih}^2 + \sum X_{jh}^2\right)^{-1}, \tag{18}$$

where e_{ij} denotes the coefficient of identity.

The Coefficient of Additivity

The association coefficient for difference scales reflects the degree to which two variables are identical up to an additive transformation. This coefficient will be called the coefficient of additivity. Inserting the additive transformation (2b) into the general formula (8) yields

$$a_{ij} = 2S_{ij}(S_i^2 + S_j^2)^{-1}, \tag{19}$$

where S_{ij} is the covariance between X_i and X_j and a_{ij} denotes the coefficient of additivity. The coefficient of additivity is the special case of Winer's "anchor point" intraclass correlation (Winer, 1971, p. 289–296) for only two variables. The coefficient of additivity is also identical to a statistic which plays a role in various estimation problems of parameters in bivariate normal distributions, cf. Cureton (1958, p. 722), Mehta & Gurland (1969) and Kristof (1972).

The Coefficient of Proportionality

The association coefficient for ratio scales reflects the degree to which two variables are identical up to a positive multiplicative transformation, that is, the degree to which the variables are proportional. Therefore, this coefficient may be called the coefficient of proportionality. Inserting the similarity transformation (2c) into the general formula (8) yields, after some algebra,

$$p_{ij} = \sum X_{ih} X_{jh} (\sum X_{ih}^2 \sum X_{jh}^2)^{-1/2}, \tag{20}$$

where p_{ij} denotes the coefficient of proportionality. It is identical to Tucker's congruence coefficient. The coefficient of proportionality will be called congruence coefficient in the remainder of this paper.

The Coefficient of Linearity

The coefficient of association for interval scales reflects the degree to which two variables are identical up to a positive linear transformation. This coefficient may be called the coefficient of linearity. The uniforming transformation (2d) equals the usual standardization, transforming X_i into Z_i. Inserting $U_i = Z_i$ and $U_j = Z_j$ into the general formula (8) yields

$$r_{ij} = 2 \sum Z_{ih} Z_{jh} (\sum Z_{ih}^2 + \sum Z_{jh}^2)^{-1}, \tag{21}$$

where r_{ij} is the coefficient of linearity. By observing that $\sum Z_{ih}^2 = \sum Z_{jh}^2 = n$, it is clear that r_{ij} is identical to the PMC.

Relations Between the Four Coefficients

Each of the four g coefficients has been obtained by substituting the appropriate uniforming transformation (2) into the general formula (8). Each of the coefficients reflects the degree of identity of the uniformed versions of the variables. The more kinds of transformations a scale type allows, the more freedom there is to transform the variables to a high degree of identity. Therefore one could expect that for a given pair of variables the PMC should always exceed or equal (in absolute value) the other three coefficients, and that the coefficient of additivity and the congruence coefficient should exceed or equal (in absolute value) the coefficient of identity. This, however, appears not to be true. Only two inequalities can be shown to exist. Specifically, we have

$$r_{ij}^2 \geq a_{ij}^2, \tag{22}$$

and

$$p_{ij}^2 \geq e_{ij}^2. \tag{23}$$

Both inequalities (22) and (23) rely on the fact that the arithmetic mean of two positive numbers exceeds or equals their geometric mean. From this we have, for instance,

$$2(S_i^2 S_j^2)^{1/2} \leq S_i^2 + S_j^2, \tag{24}$$

and therefore

$$S_{ij}^2 (S_i^2 S_j^2)^{-1} \geq 4 S_{ij}^2 (S_i^2 + S_j^2)^{-2}, \tag{25}$$

which is equivalent to (22). Similarly (23) results from

$$2(\sum X_{ih}^2 \sum X_{jh}^2)^{1/2} \leq \sum X_{ih}^2 + \sum X_{jh}^2. \tag{26}$$

In some situations two or more of the four g coefficients are identical. Equality in (24) is obtained iff $S_i^2 = S_j^2$. In that case the coefficient of additivity is identical to the PMC. Similarly, the coefficient of identity equals the congruence coefficient iff $\sum X_i^2 = \sum X_j^2$. From (18) and (19) it is clear that the coefficient of identity and the coefficient of additivity are identical if both variables have mean zero. Combining these results shows that all four g coefficients are identical if the variables have mean zero and equal sums of squares.

Applications

Test Theory

If a set of items satisfies the requirements of the one parameter logistic (Rasch) model, item and person parameters can be determined on a difference scale. To compare the item parameters resulting from different studies with the same items, the coefficient of additivity is an appropriate association measure. In this context McDonald (1982) has suggested using the coefficient

$$\tau_x = 1 - \frac{\mathrm{Var}\,(X_i - X_j)}{\frac{1}{2}\,[\mathrm{Var}\,(X_i) + \mathrm{Var}\,(X_j)]}. \tag{27}$$

It can be verified that McDonald's coefficient is identical to minus one plus twice the coefficient of additivity. As a result McDonald's coefficient lacks most of the desirable properties of the coefficient of additivity.

Profile Similarity

The introduction of the coefficients of additivity and identity provides some new perspectives on profile analysis, that is, the analysis of sets of scores of individuals on a number of variables. Nunnally (1978, p. 439) has argued that profiles contain three major types of information: level (mean), dispersion (variance) and shape (distribution form). Choosing a coefficient of profile similarity implies choosing to respect or ignore certain types of information. For instance, adopting the PMC as a measure of profile similarity implies that level and dispersion are ignored.

Adopting the coefficient of additivity as a measure of profile similarity implies ignoring levels only. This can be useful as a means of eliminating the effects of certain response tendencies, e.g., acquiescence on a vocational interest questionnaire, or leniency in the case where 'profiles' of judges who evaluated a set of objects are compared.

When both levels and dispersions are to be respected then one may adopt the coefficient of identity as a measure of profile similarity. Nunnally (1978, p. 444) advocated raw score factor analysis of cross-products (i.e. numerators of the identity coefficients) as a method of clustering profiles. Factoring coefficients of identity instead seems to be an attractive alternative, since the latter are constrained to be between -1 and $+1$.

Sjöberg and Holley (1967) have discussed a measure H of profile similarity which is insensitive to changing the polarity (sign) of one or more variables in the profiles. Their H coefficient is identical to the congruence coefficient. It may be noted that the coefficient of identity is also insensitive to changing the polarity of one or more variables in the profiles, which means that in this respect it is an alternative to the H coefficient.

Discussion

The association coefficients described in this paper have been derived as sample statistics. The population equivalents of the coefficient of additivity and the PMC can be obtained by replacing sample variances and covariances by their corresponding population parameters, degenerate cases excluded. The population equivalents of the coefficient of identity and the congruence coefficient can be obtained in a parallel fashion, by replacing mean cross-products and mean squares by corresponding expectations.

Sampling properties of the PMC are well-known. In addition, Kristof (1972) has obtained some results for the statistic u (Mehta & Gurland, 1969), which is identical to our coefficient of additivity. Sampling properties of the congruence coefficient and the coefficient of identity seem not to be available, to the best of the authors' knowledge.

The family of association measures, described in this paper, is by no means the only "family" to which its members belong. For instance, the PMC and the congruence coefficient also belong to a family of coefficients r_c derived by Cohen (1969, p. 282).

References

Bellman, R. (1960). *Introduction to matrix analysis.* New York: McGraw-Hill.

Browne, M. W. (1977). Generalized least-squares estimators in the analysis of covariance structures. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models.* Amsterdam: North-Holland.

Burt, C. (1948). The factorial study of temperamental traits. *British Journal of Psychology* (Statistical Section), *1*, 178–203.

Cohen, J. (1969). r_c: A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin, 71*, 281–284.

Cureton, E. E. (1958). The definition and estimation of test reliability. *Educational and Psychological Measurement, 18*, 715–738.

Gantmacher, F. R. (1958). *Matrizenrechnung* (Teil 1). Berlin: VEB Deutscher Verlag der Wissenschaften.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika, 53*, 315–328.

Janson, S., & Vegelius, J. (1982). Correlation coefficients for more than one scale type. *Multivariate Behavioral Research, 17*, 271–284.

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

Kristof, W. (1972). On a statistic arising in testing correlation. *Psychometrika, 37*, 377–384.

McDonald, R. M. (1982). Linear versus nonlinear models. *Applied Psychological Measurement, 6*, 379–396.

Mehta, J. S., & Gurland, J. (1969). Some properties and an application of a statistic arising in testing correlation. *Annals of Mathematical Statistics, 40*, 1736–1745.

Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw-Hill.

Pólya, G. & Szegö, G. (1925). *Aufgaben und Lehrsätze aus der Analysis* (Band 2). Berlin: Springer.

Schur, J. (1911). Bemerkungen zur Theorie der Beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *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.

Suppes, P., & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. II). New York: Wiley.

Tucker, L. R. (1951). *A method for synthesis of factor analytic "s" studies* (Personnel Research Section Report No. 984). Washington, D. C.: Department of the Army.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

Manuscript received 7/18/83
Final version received 1/9/84