MULTIVARIATE ANALYSIS THROUGH A SYMMETRIC PARTIAL GAMMA

J. Sherwood Williams and B. Krishna Singh
Survey Research Center, Virginia Commonwealth University

The symmetrical measure of association proposed by Goodman and Kruskal (4) called gamma (γ) has proven to be a useful and increasingly popular measure of ordinal association in situations where x and y are ordered polytomies. This measure indicates how much more probable like orders are than unlike orders in two classifications when two cases are selected at random (4). Costner (2) has proposed an even clearer interpretation for gamma based upon the "proportional-reduction-in-error" (PRE) criterion.

Often researchers wish to examine the association between two polytomies while controlling for other polytomies. Goodman and Kruskal (4) suggested that measures of partial association for gamma might be developed for both asymmetrical and symmetrical situations. Davis (3), following Goodman and Kruskal, has developed two asymmetric measures of partial gamma. The first measure is based upon a weighted average of the conditional gamma coefficients in the different strata. The formula suggested by Davis may be written as:

$$\gamma_{xy.z} = \frac{\sum_{i=1}^{k} \gamma_{xy.z} (\Pi_s + \Pi_d)}{\sum_{i=1}^{k} (\Pi_s + \Pi_d)} \qquad (1)$$

where x, y, and z are ordered polytomies; Π_s denotes like orders and Π_d denotes unlike orders.

The other measures suggested by Davis is "based directly on probabilities of error." The second formula is:

$$\gamma_{xy.z} = \frac{\Pi_s xy.z - \Pi_d xy.z}{\Pi_s xy.z + \Pi_d xy.z} \qquad (2)$$

Formula 2 indicates how much more probable it is to get like orders in measures x and y when pairs of individuals differing on x and y and tied on z but unselected on any other measure are chosen at random from the population (3).

It is relatively simple to compute sample frequencies corresponding to $\Pi_s xy.z$ and $\Pi_d xy.z$ when the variables are dichotomies or trichotomies. However, difficulties arise in situations where the number of ordered categories in the polytomies increase beyond three or when the sample is not large. The risk of having zero cells, which would falsely raise the value of the partial coefficient, is increased as the number of categories in each variable increases. For example, if we have three ordered trichotomies, we have 27 cells in our cross-classification. If our three ordered polytomies have five categories each, the number of cells is increased to 125. Needless to say, the problems faced in terms of computation and interpretation increase rapidly as the number of cells in the cross-classification increase, particularly if we encounter zero cells. Further, the computation becomes even more cumbersome when it is necessary to go beyond the first order partials. Also, if the causal linkages and time order of the variables are unclear, it is not logical to compute an asymmetrical partial association.

The intent of this paper, then, is to arrive at a symmetrical partial coefficient for Goodman and Kruskal's gamma. Until now this task has not been done to the best of our knowledge. This is somewhat surprising since it was suggested in 1954 by Goodman and Kruskal.

THE SYMMETRIC PARTIAL GAMMA COEFFICIENT

It is clear that if we have two ordered polytomies, gamma may be used as a symmetrical measure of association, and it has a clear (PRE) interpretation. If a third ordered polytomy is added as a test variables, we may calculate three bivariate gamma coefficients, γ_{xy}, γ_{xz}, and γ_{yz}, all of which are symmetrical. Although the gamma coefficient is not ordinarily thought of as indicating the extent of linear association between variables, it may be considered as a general index of monotonicity of the underlying relationship. This being the case, gamma may indicate the tendency of the underlying rank orders that are to be related in a monotonic fashion (5). Gamma indicates the general tendency toward monotonicity.[1] Thus, if we desire a symmetrical partial gamma coefficient, the following measure is proposed:

$$\gamma_{xy.z} = \frac{\gamma_{xy} - (\gamma_{xz} \quad \gamma_{yz})}{(1 - \gamma^2_{xz}) \quad (1 - \gamma^2_{yz})} \qquad (3)$$

This formula may be expanded. The second order partial coefficient, where two test variables are used, would be:

$$\gamma_{xy.zw} = \frac{\gamma_{xy.z} - \gamma_{xw.z} \quad \gamma_{xy.w}}{(1 - \gamma^2_{xw.z}) \quad (1 - \gamma^2_{xy.w})} \qquad (4)$$

Since the proposed partial is symmetric, it does not necessitate assumptions concerning causal order or time sequence. This measure also takes into account the general monotonic tendencies of all the bivariate relationships. Unlike Davis' (3) asymmetric partial gamma coefficients, this measure is not affected by extended distributions of ordered categories. The proposed coefficient indicates the association between x and y adjusting both x and y for monotonicity on z. The monotonicity tendency between x and y adjusted for their monotonicity with z is represented in the relationships of the residuals when each variable is predicted from z. The symmetric partial association, then, is an association between errors in prediction (in terms of their original association). It should be noted that this partial coefficient is an indicator of monotonicity between x and y adjusting for z only if the initial bivariate relationship is monotonic. This partial coefficient is not appropriate if the zero order relationship between x and y is not monotonic.

AN EXAMPLE

We will use the data provided by Davis (3:192) to illustrate the differences in asymmetric and symmetric partial coefficients. Table 1 shows the distribution of three variables, i.e., age, education, and reading. The intent is to determine the degree of association between age and reading when the effects of education are partialled out.

The bivariate gamma coefficients are:

$$\gamma_{\text{age. reading}} = -.241$$

$$\gamma_{\text{age. education}} = -.416$$

$$\gamma_{\text{education. reading}} = .689$$

From the above coefficients we can compute the symmetric coefficient, which will indicate the degree of association between age and reading when the effect of education is partialled out. Substituting the gamma values in equation (3), we have:

$$\gamma_{ar.e} = \frac{-.241- (.689)\ (-.416)\}}{\{1-(.689)^2\}\ \{1-(-.416)^2\}} = -.067$$

The symmetric partial coefficient is slightly higher than the asymmetric partial coefficient (-.014) but essentially does not alter the interpretation suggested by Davis (that there is negligible association between age and reading). Similarly, we can examine the relationship between education and reading when the effect of age is partialled out. From the associations above, we can write:

$$\gamma_{er.a} = \frac{.689-\{\ (-.241)\ (-.416)\}}{\{1-(-.241)^2\}\ \{1-(-.416)^2\}} = .667$$

The asymmetric partials are not computed by Davis in his paper but can be computed easily from his data. The symmetric partial coefficient is .667 indicating that the original relationship be-

TABLE 1

AGE, EDUCATIONAL ATTAINMENT, AND BOOK READING IN A SAMPLE OF BALTIMORE WOMEN*

Education		Age		Book Reading: Low (-)	Book Reading: High (+)
College	(+)	45 or older	(+)	36	101
		Under 45	(-)	46	163
High School		45 or older	(+)	179	159
		Under 45	(-)	327	290
Less than high school	(-)	45 or older	(+)	335	54
		Under 45	(-)	133	24
TOTAL		45 or older	(+)	550	317
		Under 45	(-)	506	477

*Source: Davis (3)

tween education and reading holds regardless of age.

DISCUSSION

The gamma coefficient proposed by Goodman and Kruskal (4) has proven to be of considerable value to behavioral scientists. From this basic measure others have been developed. Somers (9) developed an asymmetric measure based upon the logic of gamma, which adjusted the gamma coefficient for ties. In 1967 Davis (3) developed an asymmetric partial for gamma. Morris (13) has explicated several ordinal measures of multiple correlation, among them gamma and gamma k. Others who have contributed to the development of ordinal measures of association, based upon the logic of gamma, are Leik (7), Leik and Gove (8), Kim (10), and Hawkes (11).

Until now a symmetrical partial coefficient for gamma has not been explicated. It should be noted that the proposed coefficient is similar to that proposed by Kendall (12) for his coefficient Tau-b. Kendall apparently was not aware of the reasons his partial coefficient was so much like that of the product-moment correlation coefficient. By showing how both variance and covariance can be estimated for ordinal data, Hawkes (11) has shown how several ordinal measures of association (including gamma) are analogs of product-moment correlation coefficient. A symmetrical partial gamma should prove useful in situations when we have ordinal data and where causal linkages or time order is not clear.

Although the sampling distribution for partial gamma is not yet known, it is known for zero-ordered gamma (Goodman and Kruskal, [14]). From the zero-order distribution it should be possible to generate a partial gamma sampling distribution.

FOOTNOTE

1. From Bishir and Drewes (1) and McGinnis (6) we know that monotone convergence implies that boundness is tantamount to convergence. If an increasing sequence (x_n) is bounded, it has a least upper bound u. Then u is an upper bound, but for any $r > 0$, the number $u - r$ is not an upper bound. Thus, for a member x_k:

$$u - r < x_k \leq u.$$

Since this is an increasing sequence, we have:

$$u - r < x_k \leq s_n \leq u \text{ for all } n \geq k.$$

This implies that $u = \lim (x_n)$. We should however keep in mind that when we deal with ordinal level measurements, we do not have strict monotones. By strict, we mean that the exact location of inequalities are within an interval, $I|a,b|$.

REFERENCES

1. Bishir, John W. and Donald W. Drewes. Mathematics in the Behavioral and Social Sciences. New York: Harcourt, Brace and World, Inc., Pp. 112-113.
2. Costner, Herbert L. Criteria for Measures of Association. American Sociological Review, 1965, 30, 341-353.
3. Davis, James A. A Partial Coefficient for Goodman and Kruskal's Gamma. Journal of the American Statistical Association, 1967, 62, 189-193.
4. Goodman, Leo A. and William H. Kruskal. Measures of Association for Cross Classifications. Journal of the American Statistical Association, 1954, 49, 732-764.
5. Hays, William L. Statistics for Psychologist. New York: Holt, Rinehart and Winston, 1964. Pp. 641-643.
6. McGinnis, Robert. Mathematical Foundations for Social Analysis. Indianapolis: Bobbs-Merrill, 1965. Pp. 248-250.
7. Leik, Robert K. A Measure of Association for Ordinal Variables. Paper read at the meeting of the American Sociological Association, 1966.
8. Leik, Robert K., and Walter R. Gove. The Conception and Measurement of Asymmetric Monotonic Relationships in Sociology. American Sociological Review, 1969, 74, 696-709.
9. Somers, Robert H. A New Asymmetric Measure of Association for Ordinal Variables. American Sociological Review, 1962, 27, 799-811.
10. Kim, Jae-On. Predictive Measures of Ordinal Association. American Journal of Sociology, 1971, 76, 891-907.
11. Hawkes, Roland K. The Multivariate Analysis of Ordinal Measures. American Journal of Sociology, 1971, 76, 908-926.
12. Kendall, Maurice G. Rank Correlation Methods. New York: Hafner, 1948.
13. Morris, Raymond N. Multiple Correlation and Ordinal Scaled Data. Social Forces, 1970, 48, 299-311.
14. Goodman, Leo A. and William H. Kruskal. Measures of Association from Cross-Classifications, III, Approximate Sample Theory. Journal of the American Statistical Association, 1963, 58, 319-364.