SAMPLING THEORY AND ORDINAL MEASURES OF ASSOCIATION: A SOCIOLOGICAL PERSPECTIVE AND A CRITIQUE

B. Krishna Singh, Virginia Commonwealth University

Despite the emergence and reemergence of various ordinal measures of association, which have been extensively used in social research, little attention has been paid to explicit and implicit sampling theories behind such measures of association. This paper examines the nature of sampling theories associated with selected ordinal measures of association. It is assumed that if we have to continue using ordinal measures of association, it is better to use those measures which have known distributions as compared to those which have no known sampling distributions. Such an assumption seems consistent with the notion that theory construction is an implicit, if not explicit, aim of social research. In order to construct social theories (with ordinal measures of association), it is imperative that we deal with those measures of ordinal association which can provide some basis for significance testing rather than those which provide no grounds for such testing. If nothing else, measures with known sampling distributions are at least better for inferences concerning monotonic functions. As is usually the case, any inference is partly derived from statistical significance and mostly derived from substantive theoretical model to be used.

The Notion of Ordinality in Social Research

Social researchers have found that ordinal measures of association are not only appropriate in a variety of social research situations, they also involve fewer assumptions which have to be met. Even though somewhat less elegant that interval levels of relationships, ordinal associations provide appropriate and pragmatic measures which can be used meaningfully in many social research situations. The general notion of ordinality in associational terms implies that if there are two ordered politomies, they can be meaningfully associated provided there are some logical and empirical bases to assume such an association (c.f. Davis, 1967, Kim, 1971). Such a notion of association is based on the premise of transitivity asserting that if x is greater than y, and y is greater than z, then it follows that x is greater than z. The degree of greaterness is either immaterial or irrelevant or both (c.f. Coleman, 1964) in a truely ordinal sense. The notion of not knowing (and for that matter not caring to know) the magnitude of differences is usually based on the fact that sometimes it is either difficult or sometimes it is largely irrelevant to know these differences. The latter part of the above statement is a decision which has to be reached by the reseacher.

There are some prevailing controversies about the use of ordinal variables in general and use of ordinal measures of association in particular. The first controversy centers around the notion that since the basic intent of social research is to be as scientific as possible (and one way to achieve such an illusive goal is through the process of mathematization), the utility of ordinal variables is obviously limited since they cannot provide point estimations (c.f. Labovitz, 1970; Wilson, 1971). In fact, many suggestions have been made where ordinal variables should be treated as interval variables. The support for such an argument is based upon the assertion that better estimations become possible through such a treatment even though there may be some assumption violations. In addition, the argument also derives its support from the notion of statistical robustness if interval levels of measurement can be assumed.

The second controversy seems more germaine to this paper. It is based on a three dimensional view of interval measures of relationships, ordinal measures of association, and significance testing. Wilson (1971), for example, has seriously questioned the use of ordinal measures of association in the development and modifications of explanatory theories formulated in mathematical, axiomatic or deductive forms. In fact, he along with others advocate use of interval levels of measurement for constructing causal models. In addition, critiques of ordinal measures argue that since many of the ordinal measures have no known sampling distributions, there utility is of somewhat of a dubious nature in constructing models of any kind of model.

At the same time, a counterargument for such a position is proposed by stating that it is neither needed nor necessary to formulate social theories in axiomatic or deductive forms. In fact, a whole school within sociology and social sciences has recently advanced almost no quantification of social data much less go even as far as to get into the argument of ordinal versus interval levels of measurement and associations and relationships. Perhaps they may have a point but such a discussion is beyond the realm of this paper.

The question which is of immediate concern is whether some of our variables make more sense as ordinal variables or whether they should be treated as interval variables. It may be pointed out that it is usually not enough on the researchers' part to assume that his or her data meet the criterion of interval level of measurement but in fact it is necessary to demonstrate the nature if reflexiveness and transitivity in precise and accurate terms. Leaving aside the question of scaling technique, there are some questions as to whether there are indeed true interval levels of measurement in social variables and whether those variables (which for argument sake can be considered at interval levels of measurement) make more conceptual sense at ordinal variables. It is one thing to ask for more precise and accurate levels of measurement, it is another thing whether such measurements do in fact exist in a real sense. At the same time, it could be also argued that in fact we have only two levels of measurement i.e. nominal and ratio, and the distinctions between ordinal and interval levels of measurement are not very meaningful.

The most common form of use of ordinal associations in making predictions about the dependent variable y, from an independent

variable x, rests on the assumption that the independent variable predict the order properties of the dependent variable. Such order properties can be tied, concordant or discordant. The general model of such predictions is known as PRE or "Proportionate Reduction in Error" of the form:

$$PRE = \frac{E(1) - E(2)}{E(1)}$$

Distributions of Ordinal Measures of Associations:

In this section, a brief examination of five measures of association will be made. These measures are (1) tau, (2) gamma, (3) dyx (4) ordinal consensus (5) Robinson's A. It can be safely assumed that ordered polytomies with two or more categories conform to multinomial distributions. The basic postulate underlying the binomial distribution can be generalized to situations with two or more classes. Such a generalization follows the rule that: "if there are C classes, mutually exclusive and exhaustive, and with probability of $p_1, p_2, \ldots p_n$. If N observations are made independently and at random, then the probability that exactly n_1 will be of kind 1, n_2 of kind 2 and n_c of kind c is given by:

$$\frac{N!}{n_1!\, n_2! \ldots n_c!} (p_1)^{n_1} (p_2)^{n_2} \ldots (p_c)^{n_c}$$

Given any discrete probability distribution, one can easily workout the probability of all possible samples. Then, in terms of these probabilities of the disagreement of sample and theoretical distribution can be evaluated. However, the development of distributions for ordinal measures of association has been hampered by the sheer number of computations involved in multinomial distributions (c.f. Hays, 1963).

As has been suggested by Goodman and Kruskal (1963), any cross-classification follows either a multinomial distribution or a hypergeometric distribution. If the sampling is with replacement, the distribution is multinomial and if the sampling is without replacement, the resulting distribution is a hypergeometric distribution. Although, with large samples, the resulting probabilities make very little difference, the differences are hightened when we deal with samples of small sizes.

Kruskal (1958:844) has suggested that a measure of association should contain (1) simplicity of interpretation, (2) reasonable sensitivity to form of distribution, and (3) relative simplicity of sampling theory. It can be said with reasonable degree of certainty that most measures of association meet the first criterion. However, the problems arise at the last two criteria. It would appear that only measures which can meet all three criteria are (1) Gamma, and (2) Morris' adaptation of Sommer's Dyx for the bivariate case.

Ordinal Measures of Association and Theory Building in Social Research.

Up to this point, the effort has been directed toward assessing some of the problems we encounter when we use ordinal measures of association and especially those measures which have no known sampling distributions. It has been suggested that we should attempt to use only those measures which have known distributions for the simple fact that PRE interpretations are much more clearer in such measures of association.

We will operate with the premise that often times it is neither reasonable nor necessary to treat ordinal variables as interval variables and for that matter the practice can be of questionable value under most circumstances. It is not being suggested that we abandon our approach toward mathematization but it should be kept in mind that a faith blinded by trust in mathematical jargon rather than the logic of mathematics is no panacea for constructing causal models. It might be added that the process of mathematization and for that matter use of higher levels of statistical techniques is not only commendable but a necessary first step toward our eventual goal of theory construction from axiomatic and deductive perspectives. But we must be aware of what our inputs are in constructing such models.

The argument that there are certain social variables which will never meet the assumption of interval levels of measurement and consequently be never subject to higher levels of statistical analysis is a valid if not necessarily a comfortable argument. As is clear from various methodological discussions, making mere assumption about interval levels of measurement and arriving at point estimations do not necessarily mean that we are dealing with any more real phenomena than using ordinal measures of associations for constructing monotone increasing or decreasing models.

In using ordinal measures of association in theory construction, there are certain rules that we can abide by. The first rule concerns the nature of explanation on the basis of ordinal measures of association. The term "explanation" implies the nature of description which can be offered from a variable about another variable. It would appear that PRE interpretations are not only consistent but are actually well within the criterion of what we generally mean by scientific explanation.

The second rule concerns the notion of prediction. As has been noted by Wilson (1971) and Kim (1971), most measures of ordinal association are totally predictive at least at the bivariate level. However, models of ordinal predictiveness usually fail to minimize errors or expected errors of predictions. Thus, we need to be careful about how such predictions are used in theory construction. It may be noted that most of the ordinal measures of association tend to predict in the direction of concordance (depending upon the nature of scaling). One should be careful in terms of making predictions from ordinal measures of association in that they do not offer what we generally refer to as unbiased estimators.

The third rule is concerned with the nature of substantive theory building which is to be pursued. It is a cardinal fact that measures of association or relationship exist as aids rather than determinants of theory construction. When one uses ordinal measures of association for theory construction, one is simply offering building blocks for a theoretical framework which can be tested in more precise and accurate terms provided measurement problems can be

resolved. In other words, the level of theoretical abstractions that we will be able to deal with are going to be somewhat less precise but necessary steps toward a greater formalization of theoretical model.

Discussion

The major concern of this paper is that if one must use ordinal measures of association in social research, one should try to use only those measures which have known distributions. The suggestion is based on the premise that such measures provide a clearer interpretation as PRE measures and provide some notion concern-the power of statistics. In addition, with known distributions, it also becomes possible to do significance testing.

It is further suggested that from a theory construction point of view, ordinal measures of association can be as effective as any other measure if they are used carefully in conjunction with substantive theoretical efforts. While the use of ordinal measures of association may not be totally consistent with our professed goals of deriving axiomatic and deductive models, it would be imperative that we keep in mind the utility of such measures till we have arrived at ways and means of defining social variables in ratio or interval terms.

What perhaps is needed is an effort toward the development of ordinal measures of association which can take into account the nature of prediction usually associated with interval levels of relationships without loosing the three benefits proposed by Kruskal (1958).

References

Davis, J. A.
1967 "A Partial Coefficient for Goodman and Kruskal's gamma." Journal of the American Statistical Association. 62(March):189-193.

Goodman, Leo A. and William H. Kruskal
1954 "Measures of Association for Cross Classification." Journal of the American Statistical Association. 49(December):732-763.

Goodman, Leo A. and William H. Kruskal
1963 "Measures of Association for Cross Classifications. III Approximate Sample Theory." Journal of the American Statistical Association. 58(March): 310-364.

Hays, William
1963 Statistics: New York:Holt, Rinehart and Winston.

Kendall, Maurice G.
1959 Rank Correlation Methods. 2nd ed. London:Griffin.

Kim, Jae-on
1971 "Predictive Measures of Ordinal Association." American Journal of Sociology. 76(March):891-907.

Kruskal, William H.
1958 "Ordinal Measures of Association." Journal of the American Statistical Association. 53(December):814-861.

Labovitz, S.
1970 "The Assignment of Numbers to Rank Order Categories." American Sociological Review. 35(June):515-524.

Leik, Robert K.
1969 "A Measure of Ordinal Consensus." Pacific Sociological Review. 9(fall): 85-90.

Morris, Raymond N.
1970 "Multiple Correlation and Ordinally Scaled Data." Social Forces. 48(March) :299-311.

Robinson, W. J.
1957 "The Statistical Measurement of Agreement." American Sociological Review. 22(February):17-25.

Sommer, Robert H.
1962 "A New Asymmetric Measure of Association for Ordinal Variables." American Sociological Review. 27(December):799-811.

Wilson, Thomas P.
1971 "Critique of Ordinal Variables." Social Forces. 49(March):432-444.