hence τ_b is a perfect exponential (nonlinear) function of ϕ—a form of relationship not detected by linear correlation. One would suspect that if ϕ were squared prior to determining correlations with the other measures a strong PRE factor would emerge first. However, little departure from linearity should exist in the relationship between ϕ and τ_b, given the range of each measure. To illustrate, assume that the values of τ_b for the original seven tables range in equal intervals from .05 to .95 (this is not the case in Hunter's original tables, however, since the correlation between the criterion and τ_b is only .87). Further assume that the procedures used to generate the remaining 42 tables did not alter the original values of τ_b beyond $\pm$.005. In this situation the linear correlation between τ_b and ϕ would be circa .98. With such a high correlation, even assuming a linear relationship, it is puzzling why ϕ and τ_b did not load highly on the same factor. This question cannot be answered satisfactorily without careful inspection of the correlation matrix and the factor-analysis procedures.

I hope that Hunter resolves these problems and explores some of the questions I have raised before extending his analysis to more complicated situations than 2 $\times$ 2 contingency analysis. However, as it stands now, one redeeming feature of Hunter's paper is that it tends to support the more "objective" proportional reduction in error criteria for selecting association measures.

CARLTON A. HORNUNG

University of Maryland

REFERENCES

Costner, Herbert L. 1965. "Criteria for Measures of Association." *American Sociological Review* 30 (June): 341–53.

Goodman, L. A., and W. H. Kruskal. 1954. "Measures of Association for Cross-Classifications." *Journal of American Statistical Association* 49 (December): 732–64.

Hays, William L. 1973. *Statistics for the Social Sciences.* 2d. ed. New York: Holt, Rinehart & Winston.

Henkel, Ramon E. Forthcoming. "Part-Whole Correlations and the Treatment of Ordinal and Quasi-Interval Data as Interval Data." *Pacific Sociological Review.*

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

Labovitz, Sanford. 1970. "The Assignment of Numbers to Rank Order Categories." *American Sociological Review* 35 (June): 515–24.

McGinnis, Robert. 1958. "Logical Status of the Concept of Association." *Midwest Sociologist* 20 (May): 73–77.

ON "VALIDITY" AS A CRITERION

Hunter has proposed that we consider "validity" as one of the "most crucial criteria" for evaluating measures of association. His objective was to

enable sociologists to make a wiser choice in selecting the "best" measures of association from among the many possible. For the present, he restricted consideration to the nominal-nominal, 2×2 case and evaluated nine measures: the percentage difference, Goodman and Kruskal's τ_b, λ_a, λ_b, λ_{ab}, ϕ, Yule's Q, the contingency coefficient, and κ.

Hunter attempted to assess validity in two ways. First, he proposed that "the intuitive judgment of the level of association present in a set of data" be taken as the validity criterion. Accordingly, Hunter constructed artificial tables to represent the full range of association and had 19 sociologists " 'rank these tables from high to low, according to the degree of association you see between the two variables.' " All 19 ranked all but one table exactly the same, so the discrepant table was eliminated. All further analysis was performed on the tables for which there was perfect concordance and on other tables systematically derived from those. Hunter calculated first the actual associations in all of the tables for each of the nine measures and then the correlations between the values for each measure and the intuitive criterion as operationalized in the ranking of the tables by the 19 judges.

The product-moment correlations between the nine measures and the intuitive judgment criterion ranged from .98 to .56 (Hunter's table 1). Hunter considered some of these correlations "suprisingly low," and went on to interpret their strengths as indications of the extent to which the mathematical properties of each measure conform to "our intuitive conception" of association.

Although some readers may be put off by the use of intuitive judgments of fellow professionals, I would like to say that, for certain purposes, such consensual validation can be useful. It can, for example, establish that an interpretation is reasonable. Thus, it can be expedient in demonstrating that the "obvious" is indeed obvious, particularly when an interpretation is subject to challenge as to its prima facie plausibility by a recalcitrant and dogmatic opponent who might try to portray it as idiosyncratic. However, this technique is no substitute for ultimate empirical proof—all of the experts could be wrong.

In the present example, we observe that the only measure of association which can be calculated in one's head, the percentage difference, correlates almost perfectly (.98) with the intuitive criterion, rendering these two variables virtually identical for purposes of correlation with others. Consequently, the correlations of the other measures with Hunter's "intuitive criterion" would be virtually equivalent to their correlations with the percentage difference itself. We must ask, therefore, what sociologists rely on when asked to "eyeball" tables and rank them according to their degree of association. It is safe to assume that they do not quickly calculate ϕ or λ in their heads. I would suspect that they grasp at the simple

percentage difference, which is quite popular (Davis 1971, p. 64), and that all of the other measures studied by Hunter are related to the criterion only as they happen to be mathematically related to the percentage difference in Hunter's tables. It should be noted that this argument would not be seriously impaired even if the 19 sociologists were unaware of calculating the percentage difference. This would simply mean that the aspect of association most visible to the naked eye is that which happens to be embodied in the most visible measure.

There are, of course, many meaningful aspects of association, and the better measures tend to capture one or more of them while neglecting one or more others. Thus, there can be no single conception of "association" (Goodman and Kruskal 1954). Kruskal (1958, p. 815), for example, states, "It is important to recognize that the question, 'Which single measure of association should I use?', is often unimportant. There may be no reason why two or more measures should not be used; the point I stress is that, whichever ones are used, they should have clear-cut . . . interpretations." From this perspective, the correlations between the nine measures and Hunter's intuitive criterion actually lead to a counterintuitive conclusion, for they appear to recommend the percentage difference over more sophisticated measures. Although the percentage difference has many virtues, its limitations are also well known (see, e.g., Davis 1971, pp. 63–71).

It should be evident by now that I am not accepting Hunter's argument, although I do appreciate its ingenuity. However, Kruskal's recognition that "two or more measures" might well be used has long intrigued me, and it brings us to Hunter's second method for assessing validity and to some possibly useful applications of it.

The second method was to factor analyze the matrix of correlations among the nine measures and rotate four factors. This strikes me as too many factors from only nine variables, although here I would remain open minded until I had inspected the solutions for fewer factors. With nine variables, the usual "eigenvalue of 1.0" criterion for extracting factors implies a cutoff at 11% (one-ninth) of the variance. This would have meant rotating only two factors. Hunter's attempt to account for virtually all of the variance (98%) gives rise to specific factors which may or may not be informative.

He intended this procedure to establish the "factorial validity" of the measures. What one gets out of a factor analysis, however, is intimately related to what goes into it. In the present case the matrix was seeded in advance with three variants of λ; it should not be surprising that they are prominent in defining the "general factor." Other measures that appear strongly related to this general factor probably behave much like λ. In an important sense, therefore, the general factor and the lesser factors as well

depend on how redundantly certain measurement concepts are represented among the available measures. If, for example, there had been included two or three more measures based on χ^2 (such as Tschuprow's coefficient and Cramér's coefficient), these might have dominated the general factor along with the contingency coefficient and ϕ which now define the second factor, Hunter's "χ^2 factor." Clearly, these outcomes depend more on the proliferation of families of measures and on Hunter's decisions concerning which measures to include than on any fundamental relationship between specific factors and a "real" meaning of association. The third factor, which featured Q and the percentage difference, and which Hunter did not attempt to name, is probably a "corner association" factor. This would emerge more clearly if the tetrachoric correlation were added to the analysis.

Parenthetically, it should be noted that Hunter inadvertently exaggerated the amount of variance accounted for by his rotated Factor I by attributing to it the amount of variance accounted for by Factor I (the first principal component, presumably) prior to rotation (75%). Because rotation redistributes the variance, his general factor actually accounts for 50% of the total variance and the second factor for 25% (rather than 13%).

Hunter's use of the factor analysis in his assessment of the validity of the various measures was apparently hindered by the fact that the percentage difference was not the variable with the highest loading on the general factor. Nevertheless, the percentage difference did have its highest loading on Factor I, as did the intuitive criterion when it was later incorporated into the factor analysis. These facts, along with the large fraction of variance accounted for by rotated Factor I, encouraged Hunter to prefer Factor I as the most valid factor and to prefer the variables loading most highly on Factor I as the most valid measures of association. This led to the spotlighting of τ_b and λ_b, since they were among the most valid measures according to both of Hunter's methods for assessing validity.

However, there is nothing in the concept of factorial validity that leads to a preference for one factor over another. According to Guilford (1965, pp. 471–72), for example, "The validity of a test as a measure of one of these factors is indicated by its correlation with the factor, which is its *factor loading*." Thus, factorial validity is a criterion for choosing among variables in a factor analysis as measures of a particular factor; it is not a criterion for choosing among factors.

Under some circumstances, one could agree with the decision to employ a general factor as the single best summary of a set of variables in one dimension. Generally, the first principal component would serve better than the strongest rotated factor, although under the conditions of Holzinger's now somewhat archaic "bi-factor" solution (described briefly in Gordon

[1968, p. 603]) a general factor might remain even after rotation. One might also justify preferring one factor over others if it is identified through inclusion of a "marker variable" as being more similar to a construct of interest than the remaining factors. Hunter's linkage of Factor I to his intuitive criterion might seem to fit this description.

However, I have already called attention to considerations that reveal Hunter's general factor to be subject to accidental aspects of the initial composition of variables. These aspects would generally be conceded to be insufficiently related to any fundamental concept of association to justify the favoring either of this factor over others or of particular variables (here, measures of association) as measures of this factor over other variables.

I have also commented on the obviously close connection between the intuitive criterion (our candidate as marker variable) and the simple percentage difference. (The negative loadings of the intuitive criterion on Factors II, III, and IV when it is included in the matrix cannot be interpreted as showing that it behaves differently from the percentage difference, which had positive loadings in the first factoring, without knowing whether those factors were reflected from their earlier positions in the course of the second factoring, as sometimes occurs.) I might add that if inspection were sufficient to provide an adequate sense of association in 2×2 tables, there would be no need to calculate measures of association. Hunter's reliance on inspection as the criterion seems to suggest that measures of association are superfluous—indeed, that most of them, with the exception of the doughty percentage difference, are inferior substitutes for the real thing.

In certain crucial respects Hunter has lost sight of the fact that there is no single meaning for "association," that the various measures operationalize various meanings, and that all or most of the meanings and measures are intuitively accessible to one degree or another through study of their mathematical formulas, comments by experts, and experience. Indeed, the main thrust of recent work has been directed toward providing intuitively meaningful interpretations for measures and toward devising measures that lend themselves to such interpretations (e.g., Goodman and Kruskal 1954). The percentage difference simply happens to be the most intuitively accessible measure; it is not necessarily the most intuitively meaningful. In view of its virtual identity with Hunter's criterion, the same would apply also to the latter. Thus both Hunter's first method and his second insofar as it depends on the first have revealed, not the intuitively more correct over the intuitively less correct measures, but simply the intuitively more accessible over the intuitively less accessible. This turns out to be a step backward rather than forward. The fact that Hunter was

able to emerge from these analyses with the spotlight on the two relatively sophisticated measures τ_b and λ_b rather than on the more primitive percentage difference is beside the point and for the most part accidental. By the same token, other measures of association are unfairly impugned by his results.

On a deeper level, I doubt that the concept of validity, which applies to empirical and therefore synthetic propositions, is an appropriate criterion for judging measures (definitions) of association, which are analytic propositions. The problem for users, after having gained sufficient intuitive understanding even of the more difficult measures, has been to decide which of the various meanings of association they wish to elicit for their data. There has apparently been a tendency to regard these meanings as absolute and therefore mutually exclusive. However, a social scientist would rarely be interested in association in only one sense and willing to disregard its presence in other senses if the first failed to be revealed. This harks back to my earlier quotation from Kruskal concerning the simultaneous use of more than one measure. Another quotation, from Cramér, may serve to underscore the point: ". . . there is no absolutely general measure of the degree of dependence. Every attempt to measure a conception like this by a single number must necessarily contain a certain amount of arbitrariness and suffer from certain inconveniences" ([1924], p. 226, quoted in Goodman and Kruskal [1959], p. 140). In the physical sciences there is no hesitation about describing a state by more than one number. Why not in the social sciences?

Hunter's factor analysis points the way toward a basis for deciding which additional measures (numbers) to apply. Obviously, measures that behave much the same way as a given measure across the domain of tables provide no additional information. But a properly executed factor analysis based on a thorough sampling of tables (either contrived or genuine) could identify families of measures that respond to different aspects of association. Certainly the results of such an analysis would enhance our intuitive understanding of the various measures and of the relations between them, perhaps in ways not easily foreseen by the mathematical statistician.

Some potential benefits are not hard to anticipate. For one thing, data analysts would not have to waste time deciding between two equally applicable measures from the same factor. For another, if there proved to be only two major factors, data analysts would be aware that they could represent all of the major aspects of association in two basic measures, one from each factor. I do not want to oversimplify by seeming to imply that if a measure failed to load on the major factors it would have no purpose, or that there could be no special purposes even for measures with the same factorial makeup. Such details would have to be settled by other means.

It is quite conceivable, however, that a factor analysis along the lines of Hunter's could provide important new insights into the behavior of measures of association.

ROBERT A. GORDON

Johns Hopkins University

REFERENCES

Cramér, Harald. 1924. "Remarks on Correlation." *Skandinavisk aktuarietidskrift* 7:220–40.

Davis, James A. 1971. *Elementary Survey Analysis*. Englewood Cliffs, N.J.: Prentice-Hall.

Goodman, Leo A., and William H. Kruskal. 1954. "Measures of Association for Cross Classifications." *Journal of the American Statistical Association* 49 (December): 732–64.

———. 1959. "Measures of Association for Cross Classifications. II: Further Discussion and References." *Journal of the American Statistical Association* 54 (March): 123–63.

Gordon, Robert A. 1968. "On the Interpretation of Oblique Factors." *American Sociological Review* 33 (August): 601–20.

Guilford, J. P. 1965. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill.

Kruskal, William H. 1958. "Ordinal Measures of Association." *Journal of the American Statistical Association* 53 (December): 814–61.

METHODOLOGICAL AND OPERATIONAL PROBLEMS

My reaction to the Hunter article is one of extreme ambivalence. The author is to be commended for bringing attention to an apparently novel and sensible method for evaluating measures of association, but he is somewhat overenthusiastic about the validity criterion at the expense of others. More important, two fundamental methodological problems characterize his research design, and one of his basic recommendations suffers from an operational difficulty. I recommend that his work be reexamined in the light of the following observations.

Hunter (p. 99) stresses that the increasing number of measures of association constitutes a worsening and plaguing problem for social scientists. Although the advantage of having a limited number of measures to utilize in given situations is not at issue, the existence of a larger set is functional in some respects. For instance, Yule's Q, although it does run high, has been recommended as particularly useful in situations where the analyst wishes to maximize the probability of locating associations (Davis 1971); some researchers have found this advice helpful (e.g., Reeder and Berka-