

MULTIVARIATE MODEL BUILDING: The Validation of a Search Strategy

by John A. Sohnquist

(Ann Arbor: Institute for Social Research, University of Michigan, 1970, 244 pp. \$5.00).

Sohnquist has attempted to shine light into the twilight zone between the platonic heights of empirically grounded theory and the more typical depths of idiosyncratic data-grubbing. Asserting that "theory ought to emerge from...data..." (p. iii) he and his colleagues (much of the work reported on was done collaboratively) have sought to develop a procedure which can be used to "reveal the structure of relationships implicit in a set of data," (p.iii). Dependent variables in this procedure must be either dichotomies or interval level variables; independent variables must be coded as nominal (or possibly ordinal).

AID (Automatic Interaction Detection), a computer algorithm based on one-way analysis of variance, is to be used to sequentially partition the population into non-overlapping subgroups such that each split "provides the largest possible reduction in the unexplained sum of squares" (p. 20), with certain constraints which prevent the process from simply generating all possible subgroups. The end product is a tree structure. In the absence of severe skewness in one or more of the independent variables, this tree will be symmetric when the impact of the independent variables can be represented additively; asymmetries in the structure suggest interaction effects.

Then, the subgroups generated by the AID analysis are to be used to create new composite variables for each major "interaction" effect. The transformed set of variables is then entered into an MCA (Multiple Classification Analysis) program.⁽¹⁾ (MCA is equivalent to a one-zero dummy variable multiple regression in which coefficients of each class reflect deviations from the grand mean).

The bulk of MMB gives a rationale for these recommendations. ~~One~~ ^{One} chapter comprehensively surveying the literature on "contextual" (i.e. interaction) effects for nominal level variables (although omitting important work by James Coleman), reveals that sociologists and political scientists have been slow to grasp the proposition that "variables don't always add." The next chapters compare results of AID and other methods for previously analyzed data and evaluate the ability of AID and MCA to "recover," despite superimposed stochastic noise, the actual forms of relationships of artificially generated processes. In an admirable final chapter Sohnquist is honest about possible lack of replicability of the AID partitions, recognizing that the more complex the statistical representation the more likely it is to be picking up idiosyncratic features of a particular data base. (Fudge's Second Law). His discussion of a variety of graphical, tabular

(1)

There is a typo in the defining equation of the MCA Algorithm in equation (1.7), p. 18 - n_{12} should be n_{21} .

and other special techniques for data inspection is excellent, though abbreviated, and even the statistically sophisticated reader could benefit from it.

AID is basically a technique for describing, i.e. parsimoniously classifying, a population in terms of its value on a criterion variable. As such it may be useful preliminary to theory building. However, AID tends to use variables which "divide the sample into subgroups of approximately the same size" (p. 131); thus theoretically significant effects may be overlooked for factors rare in the sample, a problem which Sohnquist dismisses far too cavalierly. Also, since the end product of an AID classification is a nominal level variable, AID seems appropriate to theory building only where level of measurement is weak.

MMB sets an extremely high standard for care and candor in data analysis and concern for the appropriate use of statistical techniques, and is in no way a statistical cookbook. However, not unlike the children of Israel worshipping the Golden Calf, some political scientists may ^{mistake} ~~worship~~ the computer for ~~their~~ God (or a sausage grinder), and cry out to it for AID so that their articles may multiply and they may flourish. If, however, they read Sohnquist with the care which he deserves, they should at least know better.

Bernard Grofman

State University of New York

at Stony Brook