Behavior Research Methods & Instrumentation
1982, Vol. 14(3), 351-352

Two PASCAL programs for managing observational data bases and for performing multivariate information analysis and log-linear contingency table analysis of sequential and nonsequential data

DAVID G. SCHLUNDT
Indiana University, Bloomington, Indiana 47401

The sequential and nonsequential analysis of observational data requires the use of a mathematical model that is appropriate to nominal scale variables. Schlundt (Note 1) discusses this issue and presents multivariate information analysis of contingency tables as a general methodology for analyzing observational data. Information theory statistics (Garner & McGill, 1956; McGill, 1954) can be used to partition the uncertainty of a dependent variable in a manner similar to the analysis of variance. A proportional reduction of uncertainty measure can be computed that is analogous to a squared correlation or multiple correlation coefficient.

Log-linear models are also used to analyze contingency tables (Goodman, 1970). A log-linear analysis can be used to infer a set of social rules that describe a sample of behavioral interactions (Schlundt, Note 2). Essentially, the model expresses the logarithm of the probability of each cell in a contingency table as a linear combination of parameters. The parameters are analogous to the main effects and interaction effects of analysis of variance. There are procedures for testing the hypothesis that each parameter is equal to zero (Goodman, 1973). Parameters that differ from zero are interpreted as if-then relationships between independent and dependent variables and can be treated as hypothesized social rules.

Schlundt (1982) generated samples of role-played heterosexual conversational interactions in which the affect expressed between the partners was systematically varied. He used multivariate information analysis and log-linear contingency table analysis to construct a formal model of the rules governing topic management, turn taking, and the expression of affect in dating conversations. Person and environmental variables were analyzed in conjunction with sequential behavioral variables.

Program 1: Sequential Behavior Analysis (SBA). A PASCAL program performs multivariate information analysis and log-linear analysis of contingency tables. It extensively uses the PASCAL data types of sets, records, and lists. The input conventions are designed to be easy to use.

The program allows for the creation and analysis of contingency tables with from two to seven dimensions. Each dimension and the categories within a dimension are extensively labeled on the output. The program has facilities for lumping categories at run time.

Each stream of behavioral observations may be preceded by a series of nonsequential variables. The SBA program takes alphanumeric input in a free format. Commas are used to separate behavioral observations, and a semicolon is used to mark the end of a stream of observations. A lag parameter can be set and used to create Markov models of different orders. Each lag creates a new dimension of the contingency table and is labeled behavior at t + 1, behavior at t + 2, and so on.

The multivariate information analysis procedure computes an uncertainty measure for all possible marginal tables. A complete partitioning of the uncertainty of each dimension is computed. The uncertainty of a dimension is expressed as a function of a set of main effects (contingent uncertainties), interaction effects (interaction uncertainties), and a residual (conditional uncertainty). Statistical tests and proportional reduction in uncertainty measures are also computed. The program has options that allow the computation of other user-specified information statistics.

The log-linear contingency table analysis procedure has two modes: model fitting and model testing. Model fitting computes all the parameters of a completely saturated log-linear model, the standard deviation, and a z test of the hypothesis that each parameter is equal to zero. These parameters are used to form a guided hypothesis as to the best-fitting log-linear model (Goodman, 1973). The hypothesized model consists of a set of marginal tables that are sufficient to account for the observed cell frequencies. The expected values under this hypothesized model are computed using an iterative procedure, and these values are compared with the actual frequencies using a chi-square statistic.

The model testing mode allows the user to specify a series of log-linear models to be tested. A model is specified by giving the program a list of the minimally sufficient set of marginal tables under that model. The expected values are computed and compared with the actual values using the chi-square statistic.

The program has a set of routines that allow for the inclusion of two sequential variables in the data. It is based on the methodological suggestions of Van den Bercken and Cools (1980). Essentially, it allows the prediction of the behavior of Individual A at Time t using the behavior of A at Time t − 1, the behavior of B at Time t − 1, and the unique combination of A and B's behavior at t − 1. If a Lag 1 structure is specified and no nonbehavioral variables are included, then the Van den Bercken and Cools option generates a four-dimensional contingency table (A_{t-1}, B_{t-1}, A_t, B_t). This table can be analyzed using either the information statistics or the log-linear contingency table analysis.

0005-7878/82/030351-02$00.45/0

Program 2: SBA PREPROCESSOR. The data input conventions for the SBA program proved to be somewhat inflexible. Therefore, a preprocessor program was written that enables the user to set up a data base of behavioral observations, to flexibly retrieve information from this data base, and to automatically generate a job stream for the SBA program. The preprocessor generates a large number of time-lagged variables that can be combined flexibly with nonsequential variables into multidimensional contingency tables.

The preprocessor expects a fixed-format numeric data base in which each observation is punched on a separate card. The order of the records in the data base is assumed to be their order of occurrence. The conventions for declaring and labeling variables, for lumping categories, and for invoking options are exactly the same as those for the SBA program. Additional commands allow for recoding variables, combining variables, creating lagged variables, excluding certain observations, and selecting a subset of variables for analysis. In addition, a variable named LAG is implicitly defined. The use of this variable in conjunction with exclusion criteria allows the user to perform a lag-sequential analysis (Sackett, Holm, Crowley, & Henkins, 1979). An advantage of this program over the one written by Sackett et al. is that the information and log-linear statistics can be used to relate lag structures to experimental and personological variables.

The fact that any number of variables can have temporal parameters allows the user to perform functional analysis and multichannel communication systems analysis.

Implementation. The SBA and SBA PREPROCESSOR programs were implemented in standard PASCAL. The programs have been run successfully on a Control Data 6600 and a Control Data CYBER 172. A few features may be implementation dependent (such as the base size for the SET data type). However, every effort was made to code the program so that it would be usable on other machines. The amount of data storage space required depends upon the maximum number of categories allowed for each dimension and the total size of the contingency table. Setting the maximum number of categories to 25 and the maximum size of the contingency table to 2,500 cells requires about 60,000 words of memory on the CDC 6600. The preprocessor can handle 25 variables with a maximum category size of 25 using about 40,000 words of memory.

Availability. Listings of the SBA and SBA PREPROCESSOR, along with manuals documenting their use, are available at no cost from David G. Schlundt, Department of Psychiatry and Human Behavior, University of Mississippi Medical School, Jackson, Mississippi 39216. Copies of the articles cited in the reference notes are also available.

REFERENCE NOTES

1. Schlundt, D. G. *Mathematical models of interpersonal behavior: An information theory approach.* Manuscript submitted for publication, 1982.
2. Schlundt, D. G. *Mathematical models of the rules of interpersonal behavior: A log-linear approach.* Manuscript submitted for publication, 1982.

REFERENCES

Garner, W. R., & McGill, W. J. The relation between information and variance analysis. *Psychometrika,* 1956, **21**, 219-228.

Goodman, L. A. The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association,* 1970, **65**, 226-256.

Goodman, L. A. Guided and unguided methods for the selection of models for a set of T dimensional contingency tables. *Journal of the American Statistical Association,* 1973, **68**, 165-175.

McGill, W. J. Multivariate information transmission. *Psychometrika,* 1954, **19**, 97-116.

Sackett, G. P., Holm, R., Crowley, C., & Henkins, A. A FORTRAN program for lag sequential analysis of contingency and cyclicity in behavioral interaction data. *Behavior Research Methods & Instrumentation,* 1979, **11**, 366-378.

Schlundt, D. G. *An observational study of the behavioral components of social competence: Topic management, speaking turn regulation, and the communication of affect in heterosexual dyadic conversations.* Unpublished doctoral dissertation, Indiana University, Department of Psychology, 1982.

Van den Bercken, J. H. L., & Cools, A. R. Information statistical analysis of social interaction and communication: An analysis of variance approach. *Animal Behavior,* 1980, **11**, 548-560.

(Accepted for publication February 13, 1982.)