A FORTRAN program for evaluating the extent of missing data in multivariate observations (program BLANKS)

MILTON R. GOLDSAMT*

*Naval Personnel Research and Development Laboratory
Washington, D.C. 20390*

This program is intended as a flexible, user-oriented data screening aid prior to utilizing statistical software which lack provisions for exclusion of missing values. Individual data records are checked for the extent of omitted multivariate observations (i.e., questionnaire items, predictor variables, rating scale item components, etc.). Individual and/or group information is printed. The user can bypass this overview and directly refine the data set by creating another data set which excludes all data records having more than a user-supplied number of missing responses. (Where no information is available about the fullness of the data, the user may desire to first obtain the overview, then refine the data accordingly.) Multiple data sets can be handled by the program in one job submittal. Tape, disk, or card variable-format input is feasible on whatever units are available at the user installation.

Input. A title card, parameter card, A-type variable format cards (for input, and output if options for generating a refined data set are utilized).

Output. Regardless of the option selected, (1) number of cases, (2) number of items, (3) input format used.

For *group* statistics only, (1) number of cases without any blanks, (2) average number of blanks, (3) average percentage of blanks.

For *group* statistics and *individual* statistics, those listed directly above and, for each case, (1) number of blanks, (2) percentage of blanks, (3) item numbers of blank items.

For creation of a refined data set, (1) output format selected, (2) case acceptance criterion—number of blank items permitted by user in new data set, (3) item numbers of blank items for those cases entering refined data set and for those cases screened out, (4) number of blank items for those cases screened out (i.e., exceeding case acceptance criterion), (5) number of cases entering refined data set.

Computer and Language. This program was written in FORTRAN IV for the IBM 360/Model 91 computer and requires 98 K. Program BLANKS is not installation-specific since it contains its own FORTRAN subroutine to check for blanks.

Restrictions. Up to 500 observations per data record, 10,000 records, and 10 input variable-format cards.

*Now at System Development Corporation, Falls Church, Virginia.

Running time depends upon the number of observations and data records to be checked. For example, processing of 374 cases and 200 items, using the maximum amount of data overview options, and compilation time, requires 11.9 sec on the IBM 360/Model 91 computer.

Availability. A copy of source program, documentation and sample output can be obtained free of charge by writing to Milton R. Goldsamt, 2702 Loma Street, Silver Spring, Maryland 20902.

A FORTRAN program for linear log odds analysis

LAWRENCE S. MAYER and PHILLIP J. PICHOTTA

*Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061*

This program is designed to perform a linear log odds analysis of a multiway contingency table, provided the variables are dichotomized and a dependent variable is specified. The program aids the user in determining a simple model which explains the odds on the dependent variable. The (saturated) model with all parameters present is fitted and the parameters are tested. Then user specified hierarchical (unsaturated) models are fitted (models such that if AB is in the model, the model must include A, B) with some parameters set equal to zero, the expected frequencies are computed, the chi-squared statistic is computed, and the parameters are estimated. The program is user oriented and extremely easy to run.

Input. The observed frequencies with cell descriptions are read in; also the highest order interactions for the user specified (unsaturated) hierarchical models are read in.

Output. The statistics discussed above as well as the odds and log odds.

Computer and Language. The program is written in FORTRAN IV for the IBM 370/158.

Restrictions. The program has a region size of 46K and will analyze up to 7-way tables.

Availability. For a free copy and description of the program, contact Lawrence S. Mayer, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

REFERENCES

- 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. A modified multiple regression approach to the analysis of dichotomous variables. *American Sociological Review*, 1972, 37, 28-46.
- Grizzle, J. E., Starmer, C. F., & Koch, G. G. Analysis of categorical data by linear models. *Biometrics*, 1969, 25, 489-504.