NONMET II: A program for the analysis of contingency tables and other types of nonmetric data by weighted least squares

HERBERT M. KRITZER

Department of Political Science, Rice University
Houston, Texas 77001

The program described here for the analysis of contingency tables by weighted least squares is a revision and combination of several earlier programs which performed this analysis (Forthofer & Koch, 1974; Forthofer, Starmer, & Grizzle, 1971; Kritzer, 1973). In addition, extensive changes have been made in the input and output features of the program, and an interactive capability has been added. As a result of the use of dynamic storage allocation techniques, there is no arbitrary limit on the rank of the input matrices; the program is limited only by the time parameter and the memory capacity of the machine. Furthermore, the use of dynamic storage allocation allows the user to request a region that is not larger than needed for his particular problem; this region may be as small as 128K on IBM 370 systems.

Method. The NONMET program provides a general package for the general linear approach to the analysis of nonmetric (nominal and ordinal) data (Forthofer & Koch, 1973; Grizzle, Starmer, & Koch, 1969; Kritzer, 1976; Lehnen & Koch, 1974). This statistical technique involves the fitting of a linear model to data arrayed in a complex contingency table by the use of weighted least squares. The input for NONMET permits the analysis of linear, log-linear, or more complex functions of the response variables.

The user either supplies a vector representing a dependent variable and its associated variance-covariance matrix, or generates a dependent variable (response) function from the row conditional proportions (probabilities) corresponding to a contingency table (see Table 1).

The probability matrix P can be rewritten as a vector p:

$$\underline{p} = \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \\ p_{22} \\ \bullet \\ \bullet \\ \bullet \\ p_{sr} \end{bmatrix} \quad (1)$$

By supplying one or more operator matrices, the user may generate a response function vector f:

$$f_{linear} = A p \quad (2)$$

$$f_{log} = K \ln (A p) \quad (3)$$

$$f_{exp} = Q \exp [K \ln (A p)], \quad (4)$$

$$f_{log-log} = L \ln \left\{ Q \exp [K \ln (A p)] \right\}. \quad (5)$$

For any of these functions, a corresponding variance-covariance matrix V_f can be generated (Forthofer & Koch, 1974, pp. 238-239).

Table 1
Hypothetical Data and Corresponding Probabilities of Occurrence

Subpopulation	Response Category		Total
	$j = 1, 2, \cdots r$		
$i = 1$	$n_{11} (p_{11}), n_{12} (p_{12}), \cdots n_{1r} (p_{1r})$		n_1
2	$n_{21} (p_{21}), n_{22} (p_{22}), \cdots n_{2r} (p_{2r})$		n_2
$\cdots$	$\cdots \quad \cdots \quad \cdots \quad \cdots$		$\cdots$
s	$n_{s1} (p_{s1}), n_{s2} (p_{s2}), \cdots n_{sr} (p_{sr})$		n_s

Note—$p_{ij} = (n_{ij}/n_i), \sum_{j=1}^{r} p_{ij} = 1.00.$

The user then specifies a series of dummy variables to represent the effect of the independent variables (factors) that define the original table. The program will then obtain estimates of model parameters from the general linear model:

$$f = X b \quad (6)$$

by means of weighted least squares,

$$b = (X' V_f^{-1} X)^{-1} (X' V_f^{-1} f). \quad (7)$$

The variance-covariance of the vector b is given by:

$$V_b = (X' V_f^{-1} X)^{-1}. \quad (8)$$

A test of the fit of the dummy variable model may be obtained by:

$$\chi^2_{GOF} = [f' V_f^{-1} f] - [b' V_b^{-1} b]. \quad (9)$$

If the χ^2_{GOF} indicates that the dummy variable model adequately represents the data (i.e., all significant specification error has been removed), complex hypotheses based upon the model parameters may be tested by means of the general linear hypothesis:

$$C b = O. \quad (10)$$

The χ^2 to test the hypothesis is obtained by:

$$\chi^2 = (C b)' (C V_b C')^{-1} (C b). \quad (11)$$

The program. The NONMET program is written in standard FORTRAN IV and is designed to be easily implemented on a variety of computer systems. The program has been used extensively on an IBM 370/155-II system running SVS.

The program makes available to the user a wide variety of options for operating mode (interactive or batch), control card input (free-field, keyword control statements are used), matrix input (fixed format or free-field format are included among the options), matrix generation (the program will automatically generate a variety of the operator matrices used to create the response function, and a variety of the dummy variable models that can be fitted to the response function), and for data input (card or tape, fixed format or free-field, contingency table or previously created response function).

Program availability. To obtain the program, contact Herbert M. Kritzer, Social Science Program Library, Department of Political Science, Rice University, Houston, Texas 77001. A copy of the program manual is available at a cost of $3.50.

REFERENCES

FORTHOFER, R., & KOCH, G. G. An analysis for compounded functions of categorical data by linear models. *Biometrics*, 1973, **29**, 143-157.

FORTHOFER, R., & KOCH, G. G. A program for the analysis for compounded functions of categorical data. *Computer Programs in Biomedicine*, 1974, **3**, 237-248.

FORTHOFER, R. C., STARMER, F., & GRIZZLE, J. A program for the analysis of categorical data by linear models. *Journal of Biomedical Systems*, 1971, **2**, 3-48.

GRIZZLE, J. C., STARMER, F., & KOCH, G. G. Analysis of categorical data by linear models. *Biometrics*, 1969, **25**, 489-504.

KRITZER, H. NONMET: A program for the analysis of nonmetric data by linear models. *Behavioral Science*, 1973, **18**, 74-75.

KRITZER, H. *Analyzing contingency tables by weighted least squares: An alternative to the Goodman approach.* Unpublished manuscript. Department of Political Science, Rice University, Houston, Texas, 1976.

LEHNEN, R. G., & KOCH, G. G. A general linear approach to the analysis of non-metric data: Applications for political science. *American Journal of Political Science*, 1974, **18**, 283-313.