Appendix 9: Indicator Method

The indicator method is used in the GENLOG and the GLM procedures to generate the design matrix corresponding to the design specified. Under this method, each parameter (either non-redundant or redundant) in the model corresponds uniquely to a column in the design matrix. Therefore, the terms *parameter* and *design matrix column* are often used interchangeably without ambiguity.

Notation

The following notation is used throughout this appendix unless otherwise stated:

n	Number of valid observations
p	Number of parameters
$\mathbf{X}$	$n \times p$ design matrix (also known as model matrix)
x_{ij}	Elements of $\mathbf{X}$

Design Matrix

Row Dimension

The design matrix has as many rows as the number of valid observations. In the GLM procedure, an observation is a case in the data file. In the GENLOG procedure, an observation is a cell. In both procedures, the observations are uniquely identified by the factor-level combination. Therefore, rows of the design matrix are also uniquely identified by the factor-level combination.

Column Dimension

The design matrix has as many columns as the number of parameters in the model. Columns of the design matrix are uniquely indexed by the parameters, which are in turn related to factor-level combinations.

Elements

A factor-level combination is contained in another factor-level combination if the following conditions are true:

- All factor levels in the former combination appear in the latter combination.
- There are factor levels in the latter combination which do not appear in the former combination.

For example, the combination [A=1] is contained in [A=1]*[B=3] and so is the combination [B=3]. However, neither [A=3] nor [C=1] is contained in [A=1]*[B=3].

The design matrix **X** is generated by rows. Elements of the *i*th row are generated as follows:

- If the *j*th column corresponds to the intercept term, then $x_{ij} = 1$.
- If the *j*th column is a parameter of a factorial effect which is constituted of factors only, then $x_{ij} = 1$ if the factor-level combination of the *j*th column is contained in that of the *i*th row. Otherwise $x_{ij} = 0$.
- If the *j*th column is a parameter of an effect involving covariates (or, in the GLM procedure, a product of covariates), then x_{ij} is equal to the covariate value (or the product of the covariate values in GLM) of the *i*th row if the levels combination of the factors of the *j*th column is contained in that of the *i*th row. Otherwise $x_{ij} = 0$.

Redundancy

A parameter is redundant if the corresponding column in the design matrix is linearly dependent on other columns. Linear dependent columns are detected using the SWEEP algorithm by Clarke (1982) and Ridout and Cobby (1989). Redundant parameters are permanently set to zero and their standard errors are set to system missing.

References

Clarke, M. R. B. 1982. Algorithm AS 178: The Gauss-Jordan sweep operator with detection of collinearity. *Applied Statistics*, Vol. 31, No. 2: 166–168.

Ridout, M. S., and Cobby, J. M. 1989. A remark on algorithm AS 178. *Applied Statistics*, 38: 420–422.