

PROC GENMOD

Required Statements

CLASS - As with PROC GLM and PROC ANOVA, the GENMOD procedure uses a CLASS statement for determining which variables in the model will define classification levels. These should be variables which code for terms such as replication id, treatment level, etc. They can be alphabetic or numeric in value. The form of the statement is:

`CLASS <variable list >;`

Please note: The data set does NOT need to be sorted in the order of the variable list prior to using GENMOD. While doing so will not affect the results in any manner, it is not necessary.

MODEL - The MODEL statement defines the statistical model to be evaluated. The syntax of the model itself is similar to that of PROC GLM, or PROC ANOVA. After the model specification, however, the options available to GENMOD differ from those of

other procedures.

The general form of the 'complete' MODEL statement is:

```
MODEL <response(s)> = <variable list> / DIST = <dist>  
LINK = <link><options>;
```

The <responses> represent any response variables which need to be included. Typically, this is the name of the variable, however, in the case of binomial data it can be a ratio of two variables (for more detailed information on the binomial case, see below). The <variable list> contains the regressors for the statistical model including any variables given in the CLASS statement. This list should appear in the correct order with all necessary crossed and nested terms.

Following the model specification, specialized options are used in PROC GENMOD. The <dist> option identifies the appropriate distribution for the data. These are given as keywords such as Gaussian, Binomial, Poisson, or Multinomial. If the DIST = option is omitted, SAS will assume the Gaussian distribution. Another option, <link>, refers to a transformation which is carried out on the <responses> before analysis. These are typically

related to the specified *<dist>*. Some examples would be the Identity, Log, and Logit links. More information on specific LINK functions for binomial and count data will be given below. When omitted, SAS will assume the Identity LINK function which results in no transformation. The remaining term, *<options>* allow for control over the estimation process as well as the type of statistics printed.

Additional Statements

LSMEANS - LSMEANS will produce mean values for specified effects. The statement is written as:

LSMEANS *<variable list>* / *<options>*;

where *<variable list>* are the effects of interest. These must be written in exactly the same way as they appear in the MODEL statement. The *<options>* give specific information relating to the means. Some examples are DIFF, which produces p-values for all possible pair-wise comparisons among the levels of the effects listed, and CL which produces confidence limits on the estimated means.

CONTRAST - The CONTRAST statement allows a linear combination of means to be tested for significance against 0. Specific hypotheses of interest regarding any effects listed in the model may be assessed in this manner. The statement is given as:

```
CONTRAST 'label' <effect> <coefficients>;
```

where *'label'* is an identifying description for the printout, <effect> is the model term of interest and <coefficients> are the appropriate coefficients for the linear combination.

ESTIMATE - The syntax for the ESTIMATE statement is exactly the same as that for the CONTRAST statement. While the CONTRAST statement tests whether a linear combination of means is significantly different from 0, the ESTIMATE statement will produce the estimated value of the linear combination along with its associated standard error and confidence limits.