



BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. # 7

DATE: April 12, 1988

SUBJECT: Logistic Regression Analysis - Model statements in PROC CATMOD

Logistic regression can be used to analyze data where the response variable has 2 or more categories. There are two methods of specifying the model in PROC CATMOD for the analysis:

METHOD 1: as a type of log-linear model (an extension of contingency tables)

METHOD 2: with regression-like syntax.

An example will demonstrate the differences. Suppose that an experiment has two treatments, T and G, applied factorially and in a completely randomized fashion to individual trees. The response variable, C, is an assessment of the number of cones produced with four categories: none, low, medium and high. The following SAS code would perform a logistic regression analysis of the data:

METHOD 1:

```
PROC CATMOD;  
  MODEL C*T*G = _RESPONSE_/ML NOGLS PRED=PROB;  
  REPEATED / _RESPONSE_ = T G T*G C C*T C*G C*T*G;
```

METHOD 2:

```
PROC CATMOD;  
  POPULATION T G;  
  MODEL C = T G T*G / ML NOGLS PRED=PROB;
```

Both methods may require a WEIGHT statement if the dataset contains records of the number of trees (or experimental units) falling into each category for each treatment.

The `_RESPONSE_` = statement of the first method MUST include the response variable C and all its interactions for a logistic regression analysis to be correctly performed.

Note that PROC CATMOD will not pool into a high order interaction any lower order terms that are missing from the model statement as PROC ANOVA and GLM do (see Biometrics Information pamphlet #3).

CONTACT: Wendy Bergerud
387-5676

NOTE: This is the first in a series of pamphlets that will discuss logistic regression.