



BIOMETRICS INFORMATION

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

PAMPHLET NO. # 9

DATE: April 12, 1988

SUBJECT: Reading Category variables in statistics software

When performing analysis of variance with software packages, it is necessary to indicate all main effect variables. This information is supplied to the program by a CLASS statement in SAS, a CATEGORY statement in SYSTAT, or by some similar command with other software packages.

Consider a design with the following main effect variables:

BLOCK	- 3 levels, coded as 1, 2 or 3
FERTILIZER	- 4 levels, coded as A, B, C or D
TEMPERATURE	- 3 levels, coded as 12, 18 or 23

In SAS, the CLASS statement is simple. It would be coded as:

```
CLASS BLOCK FERT TEMP;
```

The SAS program is 'clever' enough to determine the number of levels of each CATEGORY variable from the coded data. Other software packages are not as friendly. SYSTAT for example, would require the following CATEGORY command:

```
CATEGORY BLOCK=3 , FERT=4 , TEMP=3
```

Not only does SYSTAT require the user to know and provide the number of levels for each CATEGORY variable, it also demands that the CATEGORY variables are actually coded as integers from 1 to K in the input data file. Thus while SAS will read TEMP values of 12, 18 and 23 and determine for itself that there are 3 levels of TEMP, SYSTAT would require that TEMP be coded as 1, 2 or 3.

Many statistical packages have requirements similar to SYSTAT. An awareness of such quirks can make data analysis considerably less painful.

CONTACT: John Thornton
387-3021