



BIOMETRICS INFORMATION

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

PAMPHLET NO. # 6

DATE: April 12, 1988

SUBJECT: ANOVA: Using Plot means

When using SAS to analyse an unbalanced analysis of variance, it is usually necessary to use PROC GLM . Theoretically, this approach should be fine, however, in practice, it often fails due to lack of memory. Complex ANOVA designs cause PROC GLM to construct matrices which exceed the 640 K memory limit of DOS (in some cases, they will even exceed the available memory on the mainframe). If the lack of balance is at the pure error level (that is to say different numbers of trees receive each treatment combination), a simple solution is to use plot means. Instead of using actual measurements from individual trees, the mean measurement for trees in each treatment or experimental unit is used instead. This results in a balanced design, and allows the use of PROC ANOVA (which assumes a balanced design) instead of PROC GLM . Since PROC ANOVA requires far less memory than PROC GLM , the analysis proceeds smoothly.

Consider the following split-split plot analysis of variance design, with 50 trees per split-split plot.

Source of Variation		Degrees of Freedom	Error Term
Block	B	2	Pure Error
Fenced	F	1	Main Plot Error
BxF	(Main Plot Error)	2	Pure Error
Treatment	T	3	Split Plot Error
TxF		3	
Species	S	3	Split-Split Plot Error
SxF		3	
SxT		9	
SxFxT		9	
BxT	(Split Plot Error)	6	Pure Error
BxFxT		6	
BxS	(Split-Split Plot Error)	6	Pure Error
BxSxF		6	
BxSxT		18	
BxSxFxT		18	
Pure Error (Trees)		4704	
Total		4799	

As it stands, this design is balanced. However, if there were significant mortality in some of the plots, the design would become unbalanced. The SAS statements required to perform this

ANOVA are as follows.

```
PROC GLM;
  CLASS B F T S;
  MODEL HEIGHT = B F B*F T T*F S S*F S*T S*F*T B*F*T B*S*F*T;
  TEST H=F E=B*F;
  TEST H=T T*F E=B*F*T;
  TEST H=S S*F S*T S*F*T E=B*S*F*T;
  TITLE 'UNBALANCED ANOVA USING PROC GLM';
```

See Biometrics Information Pamphlet #4 (BI #4) for a discussion of how SAS pools error terms.

This design may be too complex for the memory limits of SAS on the PC, and the program would be unable to execute. To solve the problem, the analysis can be performed using plot means, which balances the design and allows the use of PROC ANOVA. The use of plot means removes the pure error term from the design and reduces the total degrees of freedom to 95.

To use plot means, first sort the data by all category variables in the CLASS statement. Next, process the data set as shown, with PROC MEANS. This produces an output data set called PLOTMEAN with one data record for each split-split plot. These records will contain the variables B, F, T, S and PMHEIGHT. The last variable, PMHEIGHT (plot mean height) is the mean height for trees in each split-split plot. PMHEIGHT is then used as the dependent variable in the MODEL statement. The SAS commands necessary for analysis change to the following:

```
PROC SORT;
  BY B F T S;
PROC MEAN NOPRINT;
  BY B F T S;
  VAR HEIGHT;
  OUTPUT OUT=PLOTMEAN MEAN=PMHEIGHT;
PROC ANOVA DATA=PLOTMEAN;
  CLASS B F T S;
  MODEL PMHEIGHT=B F B*F T T*F S S*F S*T S*F*T B*F*T;
  TEST H=F E=B*F;
  TEST H=T T*F E=B*F*T;
  TITLE 'BALANCED ANOVA USING PLOT MEANS';
```

Notice that the split-split plot error term was not coded in the model statement. With the removal of the pure error term, the split-split plot error term becomes the lowest level of error in the design. Excluding it from the MODEL statement causes SAS to use it as the default error term. Since there is no pure error term, it is not possible to test BLOCK or any of the BLOCK interactions.

CONTACT: John Thornton
387-3021