Statistical Programs
College of Agriculture and Life Sciences
http://www.uidaho.edu/ag/statprog/

Analysis of Variance (ANOVA)

Required Statements:

CLASS: The CLASS statement is used to define variables which represent groupings or classifications of the data. Examples would be treatment and replication ID numbers or letters. Since the values of these variables represent the levels of an effect, SAS will accept either numeric or alphabetic data in the CLASS statement variables. Note: Data do not have to be sorted according to the CLASS variables.

MODEL: The model statement specifies the response and independent effects to be used. The general form is MODEL Dependent var. = Independent var., where the *Independent var.* is a list of all or some of the variable names given in the CLASS statement. Like GLM, MIXED provides the ability to easily state interaction as well as nested terms in the MODEL statement. Interaction effects can be specified by the * symbol between two or more terms (e.g. A*B). This notation can become cumbersome if many interactions are present so a shorthand version exists. The vertical bar between effects signifies all possible interactions. An example would be: A|B|C which produces the effects A, B, A*B, C, A*C, B*C, and A*B*C, in that order. Nesting occurs when one effect exists completely within another effect either in time or space. These effects are denoted with parentheses such as REP(A) which is read as 'REP within A'. The *, |, and () can be used alone or in combination to give the desired results.

Additional Statements and Options:

RANDOM: PROC MIXED derives its name from the ability to incorporate random effects into the model, i.e. a mixture of fixed and random effects. The syntax for implementing a mixed model is: RANDOM *Independent var.* / <options>, where *Independent var.* is a list of variables that should be considered as random effects in the model. Variables listed may appear in the CLASS statement, although it is not required. Interactions and nested effects may also be used in a random statement. Typically in ANOVA models, variables which are used in the RANDOM statement are not also used in the model statement. Common <options> available in the RANDOM statement are the SUBJECT= and TYPE= options. The SUBJECT= option specifies the effect that represents the basic experimental unit from which random effects are to be computed. The TYPE= option will determine form of the variance-covariance structure to be estimated. Typical examples would be UN for unstructured (all

possible variances and covariances) and VC or Variance Component which estimates a diagonal variance-covariance matrix. While many TYPE= structures are available in MIXED, care should be taken when specifying structures with numerous components, e.g. TYPE=UN.

REPEATED: Repeated measures designs may be incorporated into the model estimation using the REPEATED statement. The syntax and options are similar to the RANDOM statement above, i.e. REPEATED *Independent var.* / <options>. The *Independent var.* in this case is optional, although specifying the effect can avoid problems when missing values are present. As with the RANDOM statement, the SUBJECT= and TYPE= options may be used to specify the repeated effect and correlation structure, respectively. The correlation structures available are numerous and cover both time as well as spatial effects. Typical examples of correlational structures for time effects would be UN for unstructured and AR(1) for auto-regressive correlations. Spatial correlation may take on the most common semi-variogram models such as SP(POW), SP(GAUS), and SP(EXP) for power, Gaussian, and exponential variogram models. As before, care should be taken when specifying correlational structure with large numbers of components.

LSMEANS: Although PROC MIXED estimates models utilizing the Maximum Likelihood technique, SAS has retained the nomenclature LSMEANS or Least Squares Means for estimating mean treatment effects. A short explanation of LSMEANS in general is given in the GLM handout # 2.1. SAS provides for comparison of LSMEANS by the DIFF option which produces a table of all possible pair-wise comparisons. It is important to remember that the probabilities associated with DIFF are applicable to a limited number of preplanned comparisons only. Standard errors for LSMEANS are automatically produced. LSMEANS can also be used for interaction and nested effects.

CONTRAST: Single degree of freedom contrasts which test specific hypotheses of interest are an alternative to pair-wise comparisons. SAS provides for these with the CONTRAST statement. A single df contrast has the general form of: CONTRAST 'any label' Factor name {coefficients}. The label portion allows you to give a name to the hypothesis being tested so it can be identified on the printout. The factor name identifies what the contrast is working on and the coefficients specify the contrast itself. An example might be:

CONTRAST ' Trts vs Ctrl' A 2 -1 -1.

Here the average of the last 2 levels of A are contrasted to the first level. Some contrasts test more than one hypothesis at a time (multiple DF contrasts) and these are separated by a comma in the CONTRAST statement. For example:

CONTRAST 'Testit' A 2 -1 -1 , A 0 1 -1

would be a 2-degree of freedom contrast.

Common Experimental Designs

Example 1: Randomized Complete Block (RCB)

```
PROC MIXED;
        CLASS BLOCK VAR FERT;
        MODEL YIELD = VAR FERT VAR*FERT;
        RANDOM BLOCK;
        LSMEANS VAR FERT VAR*FERT / DIFF;
        CONTRAST 'FERT LINEAR' FERT -1 0 1;
        CONTRAST 'FERT QUAD' FERT -1 2 -1;
```

NOTES: The effect of blocks is random and does not appear in the model statement. In this simple case, SAS will compute the desired degrees of freedom to test VAR, FERT and the interaction VAR*FERT. Two contrasts are used to test for linear and quadratic trends in the FERT means.

Example 2 - Split Plot:

```
PROC MIXED;
        CLASS BLOCK VAR FERT;
        MODEL YIELD = VAR FERT VAR*FERT;
        RANDOM BLOCK BLOCK*VAR ;
        LSMEANS VAR FERT / DIFF;
```

NOTES: An error term for main plots (VAR) has been added to the RANDOM statement (BLOCK*VAR). A TEST statement is not required. SAS will test the main plot effect VAR with the appropriate error term, BLOCK*VAR, with 3 degrees of freedom. The LSMEANS will also be computed and tested using the appropriate error terms and DF.

Example 3 - Split Block:

```
PROC MIXED;
        CLASS BLOCK VAR FERT;
        MODEL YIELD = VAR FERT VAR*FERT;
        RANDOM BLOCK BLOCK*VAR  BLOCK*FERT;
        LSMEANS VAR FERT VAR*FERT;
```

NOTES: In this example the number of error terms in the RANDOM statement has been increased one more level due to an additional restriction on the randomization. As before, the error terms and DF will be appropriately assigned.

Example 4 - Repeated Measures RCB:

```
PROC MIXED;
    CLASS VAR FERT BLOCK;
    MODEL YIELD = BLOCK VAR FERT VAR*FERT/DDF=15,15,15;
    RANDOM BLOCK;
    REPEATED TIME / SUBJECT=BLOCK*VAR*FERT TYPE=AR(1);
    LSMEANS VAR FERT VAR*FERT / DIFF;
```

NOTES: Here a repeated factor, TIME, has been added to the data. Because this variable represents subsampling of the experimental units, the denominator degrees of freedom must be specified with the “DDF=” option. The REPEATED statement specifies the variable TIME as a repeated measure with each block-variety-fertilizer combination as the experimental units. Correlation between time points is specified as an autoregressive process with a lag period of 1. The LSMEANS statement will utilize the DF as specified in the MODEL statement.