Homework 3:

Note: This and other homeworks are intended as learning tools and take you step by step. The quizzes are assessment tools and are more challenging. Be sure to look at the old quizzes.

1. Here we choose 4 locations in NC at random in which to plant maple trees. Location effects will be represented by L_j, j=1,2,3,4, There are three methods of planting. These methods, with effects M_i, i=1,2,3, are our three treatments. In each location, three plots are laid out and the methods randomly assigned to them. Our response Y_{ij} under treatment i in location j is the tree growth after some time period so we assume

$$Y_{ij} = \mu + M_i + L_j + e_{ij} \text{ where } L_i \sim N(0, \sigma_L^2) \text{ and } e_{ij} \sim N(0, \sigma^2)$$

and all random effects are independent of each other.

(A) Explain why this is or is not properly considered to be a randomized complete block design. If it is not, explain your reasoning. If it is, what are the blocks? Should they be considered fixed or random?

(B) Suppose I average the 4 location effects (L's). What is the variance of that average (a formula involving σ_L^2 of course)? What is the variance of the average of the four e_{2j} values? These are the e's that go with method 2 because i=2 in the subscript. Also take the square root to get the standard deviation formula for this average of e's.

Now suppose I average all four responses Y for method 2. The expected value of this average is obviously $\mu + M_2$. What is the variance of this average (a formula related to what you've already done). Also take the square root to show the standard deviation formula for this average of 4 Y's.

(C) Our answers so far have involved some unknown variances that we need to estimate. Suppose we have this data where L = location, M=method and Y=yield:

```
L M  Y
1 1 77
1 2 38
1 3 49
2 1 81
2 2 50
2 3 62
3 1 96
3 2 58
3 3 70
4 1 80
4 2 51
4 3 66
```

Run the analysis of variance on this data in PROC GLM. Include these statements in your GLM run (after your MODEL statement):

```
RANDOM LOCATION;
ESTIMATE "Method 2 Mean" INTERCEPT 1 METHOD 0 1 0;
MEANS METHOD;
LSMEANS METHOD; run;
```

(i) Are the means from GLM just the ordinary averages of 4 Y's? How do the LSMEANS differ from them, if at all, in this simple model?
(ii) What happens if you omit INTERCEPT 1 in the ESTIMATE statement?
(iii) The RANDOM statement produces expected mean squares from which we can compute estimates of σ_L^2 (called Var(LOCATION) in the printout) and of σ^2. I am sure you remember how to estimate variance components from ST 512 but just in case…

(a) Set the LOCATION mean square (a number) equal to its expected mean square.
(b) Set the mean squared error equal to its expected value (σ^2 or Var(error) in the SAS printout).
(c) Solve these two equations for the variance components desired.

In general, setting estimates equal to their expected values and solving is called the "method of moments" in statistics. Record the estimates of the two variance components.

(iv) Insert these estimates into the standard deviation formula you got in part (B) for the average of 4 Y's. Inserting estimates into this formula gives you what we call the "standard error" of the average of the 4 Y's. Does it match the standard error from your PROC GLM ESTIMATE statement? Is it just the standard error for the average of 4 e's? (If so it is treating the locations as fixed effects and thus incorrectly ignoring the random locations' contributions to the standard error).

(D) Now try this code on the data:

```
PROC MIXED; CLASS LOCATION METHOD;
MODEL Y = METHOD;
RANDOM LOCATION;
ESTIMATE "Method 2 Mean" INTERCEPT 1 METHOD 0 1 0;
LSMEANS METHOD; run;
```

In PROC MIXED, only the fixed effects appear in the MODEL statement and the variance components are given directly – no expected mean squares or hand computations are involved.

(i) Is the estimate in the ESTIMATE statement the same as in GLM? How about the standard error?
(ii) Does the standard error for the treatment 2 mean match what you computed by hand (in C, iv) now that you are in PROC MIXED?
(iii) Are the estimated variance components from PROC MIXED the same as those you computed from GLM using the method of moments? (note: this is only because the design is simple and completely balanced)