Statistics 402 A **Assignment 9** **Due April 14, 2004**

1. The SAS program *rcbd.sas* available on the course website provides example code for the analysis of data from an randomized complete block design (RCBD). Use the code and output to answer the following questions.

a) Was there any evidence for differences among treatment means in this experiment? Conduct one test to answer this question. Provide the test statistic, its degrees of freedom, an approximate *p*-value, and a brief conclusion.

b) Conduct tests for all pairwise comparisons between treatment means. Indicate which treatments are significantly different from one another using a method that keeps the probability of one or more type I errors at or below 0.05. (Your answer can simply be a list of pairs of treatments that are significantly different from one another and some explanation of which was the larger of the two treatment means for any treatment means that differ.)

c) Suppose treatment 1 was the control treatment. Which treatments were significantly different from the control treatment? Provide a *p*-value for each comparison of a treatment mean with a control mean that has been adjusted using Dunnett's method to account for multiple testing. Provide a brief conclusion for each test. (Note that the SAS output provides such *p*-values for you directly. Previously we used the *means* statement to get confidence intervals for differences between each treatment and the control that were adjusted for multiple comparisons using Dunnett's method. The code in *rcbd.sas* uses the *lsmeans* syntax to get adjusted *p*-values using Dunnett's method. A student in our class (Finn) asked about how to do this a long time ago. It is nice to be able to get the adjusted *p*-values, but the drawback to using *lsmeans* to get Dunnett's adjusted *p*-values is that you can't specify which treatment is the control like we did previously with the *means* syntax. SAS assumes the first treatment is the control, so you might need to reorganize your data accordingly to use this code in the future for your own problems.)

d) Estimate the difference between the mean for treatment 1 and the average of the other three treatment means (treatment 2, 3, and 4). You may do this by hand or by adding the appropriate estimate statement to the SAS code and examining the output. Along with your estimated difference, provide the standard error, a *t*-statistic testing whether the estimated difference is zero, the degrees of freedom for the *t*-statistic, a *p*-value, and a brief conclusion. (Note that the same formula that we used to get the standard errors of linear combinations in completely randomized designs applies for randomized complete block designs with blocks as fixed effects. If the linear combination is a contrast and the data are balanced, the formula is the same whether blocks are considered to be fixed or random.)

e) Suppose this data was from a field experiment with yield in bushels per acre as the response variable. Suppose the treatments were as follows:

Treatment 1: fertilizer type F1 with variety V1
Treatment 2: fertilizer type F1 with variety V2
Treatment 3: fertilizer type F2 with variety V1
Treatment 4: fertilizer type F2 with variety V2

Provide tests for fertilizer main effects, variety main effects, and the interaction between the two factors. These tests must account for the randomized complete block design used for the experiment. For each test, provide a test statistic, its degrees of freedom, a *p*-value, and a brief conclusion. (There are many ways to do this problem. One way would be to add columns for fertilizer type and variety to the data and use the code

```
proc glm;
  class block fertilizer variety;
```

*model y=block fertilizer variety fertilizer*variety;*
run;

Alternatively, you could compute the appropriate tests by hand fairly easily using the lsmeans and MSE from the output already available. The quickest method would probably be to add appropriate estimate statements to the current code. Please do this however you would like, but make sure you understand how to go about getting the answer using each of the methods.)

2. A Latin square design was used to compare three treatments A, B, and C.

a) Below is a partial sketch of the treatment assignment for the Latin square design with rows and columns as blocks. Determine the entire 3 X 3 assignment of treatments to experimental units

```
B C ?
C ? ?
? ? ?
```

b) Suppose the observations corresponding to each experimental unit were as noted below with the position of the numbers matching the arrangement in part (a). Determine the ANOVA table for the analysis of this data. Include columns for SOURCE, DF, SUM OF SQUARES, and MEAN SQUARES. (Please do this problem by hand calculation. You may use SAS to check your answers if you wish.)

```
4 5 9
2 7 3
6 3 0
```

c) Were there significant differences among treatment means? Conduct one test to answer this question. Provide the test statistic, its degrees of freedom, an approximate *p*-value, and a brief conclusion.

d) Provide a 95% confidence interval for the difference between the mean for treatment A and the mean for Treatment B.

3. Suppose a group of researchers at four universities is studying the effect of three treatments on plant growth. To share the workload and resources, the experiment was replicated once on each of the four university campuses. Two growth chambers were available for use on each campus. Thus a total of eight growth chambers were used for the experiment. Three trays of plants were kept in each growth chamber. The three treatments (call them A, B, and C) were randomly assigned to the three trays within each growth chamber. At the conclusion of the experiment, the total dry weight of plants in each tray was computed. These values were used for analysis. SAS code for this problem is contained in *multloc.sas* on the course website. Please use this code as an aid to answering the questions that follow.

The researchers are attempting to keep all the experimental protocols identical as much as possible, but variation will inevitably exist among university campuses, growth chambers, and trays of plants. Based on the SAS code, you should be able to see that the researchers would like to treat the effects associated with universities, chambers within universities, and treatment-by-university interaction as random effects. These may be reasonable assumptions. For example, the random assumption for universities says that effects associated with these four universities might be like a random sample of possible effects from some normal distribution of possible effects. Treating these effects as random effects will allow the researchers to generalize their conclusions beyond the four campuses where the experiment was conducted and beyond the eight growth chambers that happened to be used for this particular experiment. Any interaction that involves one or more

random effects is treated as random, so treatment-by-university interaction is random in this case. Examine the expected mean squares output produced by the *random* statement in *proc glm* before answering the following questions.

a) Was there any evidence for differences among treatment means in this experiment? Conduct one test to answer this question. Provide the test statistic, its degrees of freedom, an approximate *p*-value, and a brief conclusion.

b) Estimate the variance component corresponding to variability among the effects associated with trays of plants within growth chambers.

c) Estimate the variance component corresponding to the variability of growth chamber effects within universities. (Remember that the expected mean square for what SAS calls MSE is denoted by Var(Error) in SAS. You will need to use MSE and another mean square to answer this question.)

d) Estimate the variance component corresponding to the variation among the university effects. (Hint: You need to use four of the mean squares (including MSE) to obtain an estimate in this case.)

e) Examine the residual plot produced by the SAS code. Note that each point is labeled with the number of the university associated with that point. From this plot, do you see any reason to be concerned about combining data from all four universities?