**STAT502 Homework #4**

due Monday, 7/24

1. (*adapted from textbook exercise 19.20*) A computer software firm was encountering difficulties in forecasting the programmer requirements for large-scale programming projects. As part of a study to remedy the difficulties, 24 programmers, classified into equal sized groups by type of experience (Factor A) and amount of experience (Factor B), were asked to predict the number of programmer-days required to complete a large project about to be initiated. After this project was completed, the prediction errors (actual minus predicted programmer-days) were determined as the response of interest. The data can be read into R with the commands

```
d = read.table('CH19PR20.txt')
y = d[,1]
A = as.factor(d[,2])
B = as.factor(d[,3])
```

   (a) Fit the two-way ANOVA model with interaction, and save the residuals and fitted values (see Lab 5). Plot the residuals versus fitted values, and comment on the assumption of constant variance.

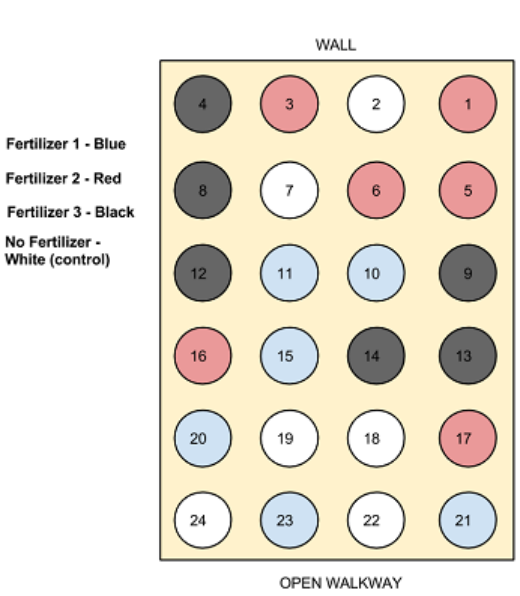
   (b) Obtain a QQ plot of the residuals, and comment on the assumption of normality.

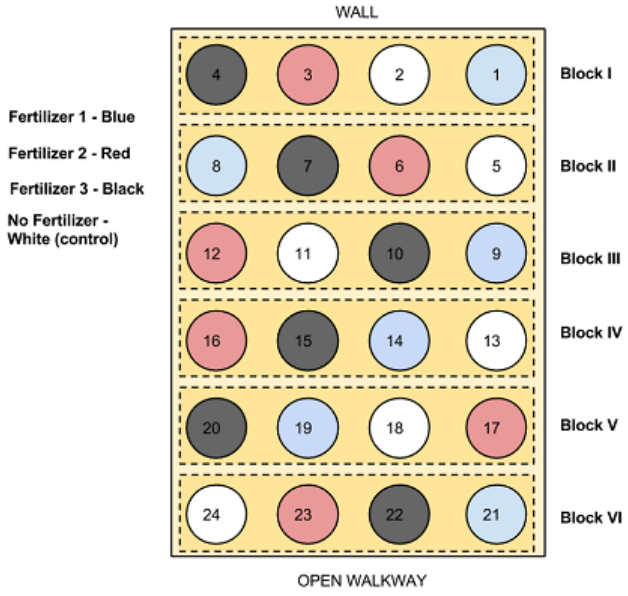
   (c) Obtain the ANOVA table, and test whether or not the two factors interact. Use $\alpha = 0.01$.

   (d) Test whether or not main effects for the two factors are present. Use $\alpha = 0.01$. Is it meaningful here to test for main factor effects? Explain.

   (e) Use Tukey to test all pairwise comparisons for Factor A at $\alpha = 0.05$ (see Lab 7).

2. Consider the following two scenarios involving 24 pots and four treatment groups (three fertilizers and a control group). The images illustrate how the treatments were assigned to the pots. The research question of interest is whether the mean yields of the treatment groups are equal or not.



(a) If $Y_{ij}$ represents the $i$th pot receiving the $j$th treatment, write the ANOVA model for each scenario above.

(b) Explain how the randomization of treatments to pots is different between the two scenarios.

(c) Give the sources and degrees of freedom in the ANOVA table for each scenario. Note: since no data is provided, you don't have to calculate any sum of squares.

3. In a study of three teaching methods, 30 instructors were selected. Since the performance of an instructor usually depends on his/her experience, the instructors were divided into 10 groups, each with three instructors, such that group 1 has the most teaching experience, and group 10 has the least teaching experience. The data is in the file "P1.txt".

```
data = read.table('P1.txt',header=T)
score = data[,1]
teacher = as.factor(data[,2])
method = as.factor(data[,3])
```

(a) Fit the RCBD ANOVA model, and report the ANOVA table.

(b) Test whether there is significant effect due to the teaching methods. State the hypotheses and conclusion with $\alpha = .05$.

(c) Use Tukey's method to examine the differences among the three teaching methods. What conclusions can be made?