STAT502 Lab #12

1. In this study, eight people were given drug A, and another eight people were given drug B. Each person's heart rate was measured at four time points (after 5, 10, 15, and 20 minutes) from when the drug was given. The data are available in the "drugs.dat" and can be read into R with the commands below. Columns correspond to person, heart rate measurements one through four, and drug type.

```
data = read.table('drugs.dat')
t1 = data[,2]
t2 = data[,3]
t3 = data[,4]
t4 = data[,5]
drug = as.factor(data[,6])
```

Notice that the data are "unstacked" in that each time measurement is in its own column. To "stack" these columns in way suitable for R, we use

```
s=8; a=2; b=4
y = c(t1,t2,t3,t4)
subj = as.factor(rep(1:(a*s),b))
drug = rep(drug,b)
time = as.factor(rep(1:b,each=a*s))
```

To see the difference, compare `data` with `cbind(y,subj,drug,time)`. Also, notice that we're refining the subjects to values 1 through 16 (why?)

(a) Recall the repeated measures model for this situation:

$$Y_{ijk} = \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$$

Which terms are random, and which are fixed? Which are nested, and which are crossed?

(b) Use `interaction.plot(time,drug,y)`, and comment on the effect of drug, the effect of time, and their interaction.

(c) Use the following commands (recall the '/' symbol to indicate nesting) to get the ANOVA table for this model.

```
fit = lm(y ~ drug/subj + drug + time + drug:time)
anova(fit)
```

Note that although the sums of squares and degrees of freedom are correct, the F statistics are all dividing by the residual error (MSE), which is not correct for the main effect for drugs. Which mean square should be used for this?

(d) Plot the residuals versus the fitted values, and also construct a QQ plot for the residuals. Comment on the assumption that $\epsilon_{ijk} \sim N(0, \sigma^2)$.

(e) Carry out the test for a drug effect with $\alpha = .05$. State the hypotheses in terms of parameters in the model above. Comment on the appropriateness of this test in light of the interaction between drug and time.

(f) Repeat part (e) above for the time effect.

(g) To make comparisons among different time points, we first fix a particular drug. For drug A, the estimate for the difference in mean heart rates between 5min and 10min is $\hat{D}_{12} = \overline{Y}_{\cdot 11} - \overline{Y}_{\cdot 12}$, and the variance is given by (recall s is the number of subjects)

$$Var(\hat{D}_{12}) = \sigma^2 \left(\frac{1}{s} + \frac{1}{s} \right)$$

It follows that the standard error of D_{12} is the sqrt of the above with MSE substituted for σ^2. Report the numeric value of both $\hat{D}_{12}$ and its standard error.

Hint: use `tapply(y,drug:time,mean)` *to get the means for drug/time combinations.*

(h) Carry out pairwise comparisons among the time points for drug A. Use Tukey's LSD to control the family-wise error rate at .05.

2. For each of the following descriptions, identify the factors and state if they are random or fixed. Also, state the relationship between them if there is any nesting.

(a) Hardwood chips are believed to differ from softwood chips in nitrogen concentration. Researchers were specifically interested in six tree species currently used in biomass fuel production (three species of hardwoods, three species of softwoods). Four samples of each species were tested for N concentration.

(b) Consider an experiment wherein four methods of tablet compaction are being tested, and the measurement of interest is the elastic modulus (em). The researchers think that the method of compaction will affect the em. Four methods (m1, m2, m3, and m4) were used, and there were five replications of the test for each method.

(c) A survey of health awareness in households (an index) is made in three states. Three cities within each state are randomly chosen, and sampling five households within each city are surveyed. Researchers are interested in how much variation is due to cities and how much is due to states.

(d) In an agricultural field experiment an investigator wants to evaluate four disease treatments on four varieties of wheat. The disease treatment levels and varieties were specifically chosen and of interest to the researchers. They can apply the disease treatment levels to small sub-sections of a field but need to plant the varieties with a mechanized seeder on entire field. They chose eight fields (two per variety) and then split the fields into four sections to apply the disease treatments.