36-402/608 Homework #7 due 10:30AM 3/4

1. Monkey memory (40 points)

   This problem uses the monkey memory data of Sleuth Chapter 16, case 1 from case1601.csv. See page 463 for a description.

   (a) Read in the data, produce a plot similar to the one on page 463, but with full (up and down) confidence bars. Each CIs should be based on the 7 or 11 values that constitute the corresponding mean; in other words don't pool any variances. The "multiplier" for each CIs will be either `qt(0.975,6)` or `qt(0.975,10)`. Turn in your R code. One nice variant you could try is to add a small fixed amount to the x-values for one of the treatments so the error bars don't overlap. It is not necessary to add the horizontal lines at the ends of the error bars or the second (lower right) legend box.

   (b) Examine the diagonals of the two separate 5 by 5 covariance matrices for the two treatments. Which time has the larges ratio of the variances for the two treatments, and what is that ratio? Examine the two separate 5 by 5 correlation matrices. What pair of times has the largest difference between corresponding correlations, and what is the difference?

   (c) Ignore the fact that we have fairly strong violations of the assumption of equal covariance across treatments, which will probably distort the results. Perform the MANOVA test of $\boldsymbol{\mu_C} = \boldsymbol{\mu_T}$ where each $\boldsymbol{\mu}$ represents all five individual times. Give the R code, the R result, and your interpretation. (Note that you need to run `aov()` with a matrix for the response, and then look at the `anova()` of that object. The `summary()` of the `aov()` object just gives the 5 one-way between-subjects ANOVAs.)

   (d) Turn in a quantile normal plot (using the qqn() function to get confidence bands) and a brief statement on the degree of violation of the normality assumption.

2. Nature / Nurture IQ Data (30 points)

   (a) Load the data from ex1605.csv, and make an EDA scatter plot that shows in four panels 1) the foster mothers education vs. age 2 IQ, 2) the birth mother's IQ vs. age 2 IQ, and the same two plots for age 13. Note: put the outcome on the y-axis. Use appropriate (non-default) labeling.

   (b) Load the "car" package, and run the following code (which we need to get more useful results in this more complex problem):

```
library(car)
idata = data.frame(time = ordered(1:4, labels=paste("age",c(2,4,8,13),sep="")))
rslt = Manova(lm(cbind(age2iq,age4iq,age8iq,age13iq)~fmed+tmiq,nn),
       idata=idata, idesign=~time)
print(rslt)
```

For this part, just turn in the R results for the above commands.

My interpretation: First we look at the interaction p-values, and conclude that we do not have sufficient evidence to reject the null hypothesis that the pattern of children's IQ across time stays the same (although possibly with changing overall level) as the foster mother's education and/or the birth mother's IQ changes.

The time p-value of < 0.0001 indicates that we have good evidence to reject the null hypothesis that the IQ of the children does not change over time. (Note that this is based on so-called "type 2" sum of squares so it is OK to interpret this as if the interaction were not in the model.)

(c) Give an interpretation of the fmed and tmiq p-values. Use your EDA plot and/or some calculated means to state the direction of the significant effect.

3. Fake beetles (10 points)

Load "fakebeet.csv". You might want to try:

```
with(fakebeet, plot(P, Q, pch=as.numeric(species),
                    main="Fake Beetle Study"))
legend("topleft",levels(fakebeet$species),pch=1:2)
```

as EDA. Turn in the separate t-tests for comparing the species (one vs. two) in terms of measurement P and in terms of measurement Q. Also turn in a T^2 test of the multivariate null hypothesis that

$$\begin{bmatrix} \mu_{P1} - \mu_{P2} \\ \mu_{Q1} - \mu_{Q2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Briefly comment on what it all means.

4. Writing Assignment (20 points) As in homework 4, you will write about the Algal Regrowth analysis. This time you are a newly hired junior statistician, and you must write a report for the chief statistician so that she can present your results to the CEO. You need to get your boss to understand what you did and why, but your boss is a busy person responsible for many projects, so you can only write a two-sided one page report.