Homework (Due Tuesday 1st December)

Do all tests at the 5% level and quote p-values when possible. When answering each question uses sentences and include the relevant JMP output and plots (do not include the raw data in your solutions and make sure your plots and output are embedded in the solution).

All students must produce their own solution.

1. There has been speculation that breakfast cereal manufacturers add more sugar to cereals targeted at children than adults.

The sugar content of 24 varieties of cereals are given in

http://www.stat.tamu.edu/~suhasini/teaching651/HW/Cereal.csv (note that the target audience is also given in this data set).

(a) Based on the discussion at the start of this question, what is the hypothesis of interest.

(b) Based on this data set, which cereals contain more sugar, what is the difference between them?

(c) Do an independent sample t-test to test this assertion.

(d) Make a QQplot of the residuals to check for normality.

(e) Give a 95% CI for the mean difference in sugar levels.

2. Who are faster at paper mazes, Adults or Children?

A random sample of 10 adults and 10 children (between the ages 6-10) was taken. Each was asked to complete a maze and their times recorded. The data is given in


(a) For this data set, who are faster at completing a maze, children or adults, give the average difference in times.

(b) Give a 95% confidence interval for the mean difference in maze times.

(c) Use an independent sample t-test to test the research hypothesis that there is a difference between the mean adult and child times.

(d) The sample sizes are small. Therefore do a Wilcoxon sum rank test too. The results in (c) and (d) concur?

3. A group of dieticians are investigating the efficacy of a certain diet in losing weight.

A small group of 6 individuals are placed on a diet. The weights before going on the diet and after being on the diet for two months are recorded. For these 6 individuals we see the average weight loss is 11 pounds.

(a) First apply an independent sample t-test to see whether the difference is statistically significant (remember to state the hypothesis of interest).

The data is given in

(b) There is a clear matching in the data (since the same person is weighed before and after diet). The same data, but with the appropriate matching is given in http://www.stat.tamu.edu/~suhasini/teaching651/HW/Diet_Matched.csv. Use the Fit Y by X option to make a scatter plot of the weights before and after the diet. Is there a clear dependence between these variables (give the plot and discuss what you see)?

(c) Apply the Matched paired t-test to this data (remember do the appropriate one-sided test).

(d) By comparing the outputs of the independent sample t-test and matched paired t-test explain what (i) are the main differences and (ii) why are the conclusion of the tests different?

(e) Apply the Sign test to this data (remember do a one-sided test), you can use JMP to do this.

(f) Apply the Wilcoxon sign test (remember do a one-sided test). Please do this by hand.

(g) Briefly compare your conclusions from the three matched procedures in (c), (e) and (f).

(4) Do have fertilizers have an impact on the height of marigolds?

150 marigolds (each in boxes of size 50) were given three different fertilizer treatments (Control, Fertilizer A and Fertilizer B) from seed until day 90. At Day 90 they were measured.

The data is given in the link (it is tiny in order to fit it on one line, else the link does not work):

http://www.stat.tamu.edu/~suhasini/teaching651/HW/MarigoldFertilizerComparison.csv.

(a) What is the average height of marigolds in each of the three groups?

(b) Use ANOVA to test the hypothesis that the treatment has no influence on the mean height $H_0 : \mu_C = \mu_A = \mu_B$ vs $H_A :$ At least one mean is different.

(c) Make a QQplot of the residuals.

(d) Use ANOVA to test for equality of means on the residuals. Is the result what you expect?