DATA ANALYSIS (40 points)
Sophomore Engineering Clinic II
Homework Due: Monday, 2/28/11

Problem 1:
In class, we collected data regarding handedness, gender, and height. Divide the data from class into four categories: left-handed males, right-handed males, left-handed females, and right-handed females. Create a figure showing the data from these groups and include 95% confidence intervals. Are there any statistically significant differences between groups? (Check your answers with the ttest function in Excel.)

Problem 2:[A]
The data in the table below is a random sample of 15 observations from two experiments using two different machines.

Sample	Machine A	Machine B
1	12.03	13.74
2	13.01	13.59
3	9.75	10.75
4	11.03	12.95
5	5.81	7.12
6	9.28	11.38
7	7.63	8.69
8	5.70	6.39
9	11.75	12.01
10	6.28	7.15
11	12.53	13.47
12	10.22	11.57
13	7.17	8.81
14	11.36	13.10
15	9.16	11.32

a) Test the null hypothesis that the population means from the two machines are equal against the alternative hypothesis that they are not equal. Assume that the two population variances are equal. Report the p-value and write a short statement interpreting this result.
b) Repeat part a without assuming that the two population variances are equal.
c) Test the null hypothesis that the population means from the two machines are equal against the alternative hypothesis that the readings from Machine B are greater than those from Machine A. Assume that the two population means are equal. Report the p-value and write a short statement interpreting this result.

[A] Problem adapted from Ogunnaike, *Random Phenomena: Fundamentals of Probability and Statistics for Engineers,* 2010.

d) The statistical tests from parts a-c assumed that the 15 samples were randomly collected. If, however, these measurements were collected in a manner where they are correlated, a paired t-test is necessary. For example, imagine that the measurements are taken throughout a process where each machine takes a measurement every 10 minutes and the measurements are expected to change with time. Repeat the statistical test from part a using a paired t-test. Report the p-value and write a short statement interpreting this result. Is there a statistically significant difference between machines at a 95% confidence level?

Problem 3:

Zebrafish are a popular model system for studying development. Bilotta et al. (2004)[B] studied the effect of alcohol on zebrafish embryos and obtained the following survival data. The data represent the percentage survival.

Days Postfertilization	Control	1.5% Alcohol	2.9% Alcohol
0	100	100	100
1	98	90	98
2	98	90	89
3	97	90	33
4	97	87	33
5	97	87	27
6	97	87	27
7	97	85	1

Plot this data and perform a Chi-squared goodness of fit test to determine whether alcohol affects embryo survival. Report the p-values and your interpretation of the results.

Problem 4:

Sometimes, when examining experimental data, a particular set of data seems to increase variability considerably. If there is an experimental reason why this seemingly errant data might be different, the data can be excluded from further analysis. Re-examine the algae growth data and identify any potential outliers (either entire experiments where a group's data seem to be very different or individual measurements that seem to be very different). Repeat your initial statistical analysis without including these "outliers." Write a short paragraph describing your rationale, your results, and your hypotheses as to what may have led to the outliers.

[B] Bilotta, J., J.A. Barnett, L. Hancock, S. Saszik, (2004) *Ethanol exposure alters zebrafish development: A novel model of fetal alcohol syndrome*. Neurotox and Teratol 26: 737-743.