MAT 2379 3X (Spring 2013)
Assignment 5
***Deadline:* Tuesday, July 23, 2013 (in class)**
There are a total of 6 questions.

Part I) Answer the following 4 questions **without** the use of R.

1) Rhodamine 6G (R6G) is a fluorochrome mitochondrial dye with potential use for cancer treatment. One of the objectives of a particular study was to show that the administration of R6G during a period of hypoglycemia reduces the growth rate of the Walker 256 tumor. A group of $n_1 = 7$ rats underwent implantation of 100 mg of viable fragments of Walker 256 carcinosarcoma, and after 48 hours they were administered R6G. The animals were fasted for 24 hours prior to the drug administration and 8 hours after. After a week, the tumors were weighed yielding a sample average and a sample standard deviation $\bar{x}_1 = 3.6g$ and $s_1 = 0.3$g. A control group of $n_2 = 7$ rats which received the same tumor transplant had the sample average and sample standard deviation $\bar{x}_2 = 7.1g$ and $s_2 = 0.7$g. Assume that the two populations are normally distributed with equal variances. Can we conclude that the administration of R6G reduces the tumor growth rate on average?

   (a) Justify your answer using a test of hypothesis at a level of significance of 5%.

   (b) Justify your answer with a 95% confidence interval.

2) In 1973, the authors of a particular study found that inefficient transport of iron from the root to the above-ground portion of a tomato mutant is controlled by a single recessive genetic factor. By crossing iron-inefficient mutants (ff) with normal plants (FF), they obtained a first generation F_1 of tomatoes in which all plants were normal (as expected). The plants in the F_1 generation were crossed with one another, to produce a second generation F_2. The F_2 generation consisted of 245 tomatoes, among which 58 were iron-inefficient. Using these data, is there enough evidence to reject the single factor hypothesis at a level of significance of 5%? (*Hint:* Formulate the hypotheses in terms of

the probability p that a plant in the F_2 generation is iron-inefficient, assuming that the phenotype is controlled by a single recessive genetic factor.)

3) Mount St.Helens is an active volcano situated in the Pacific Northwest region of the United States, which had a powerful eruption on May 18, 1980. The region around the volcano became a national park in 1982. Since the eruption, the size of the fish in Spirit Lake at the bottom of the volcano seems to have increased. In 2010, a sample of 30 rainbow trouts had an average weight of 2.3 lb and a sample standard deviation of 4 lb. Is there significant evidence that 30 years after the eruption, the average weight of the trouts is higher than the pre-eruption average weight of 1.9 lb? (Assume that the fish weight has a normal distribution.)

4) In a study conducted at Virginia Tech, the plasma ascorbic acid levels of pregnant women were compared for smokers and non-smokers. Thirty-two women in the last three months of pregnancy, free of major health disorders and ranging in age from 15 to 32 years, were selected for the study. Prior to the collection of 20 mL of blood, the participants were told to avoid breakfast, forgo their vitamin supplements, and avoid foods high in ascorbic acid content. From the blood samples, plasma ascorbic acid values were determined, in milligrams per 100 milliliters. Here are the command that we used to import the data into R and to assign one of the samples to x and the other sample to y.

```
> table = read.table(file.choose(),header=TRUE,sep="\t")
> names(table)
[1] "Plasma.Ascorbic.Acid" "Status"
> x=table$Plasma.Ascorbic.Acid[table$Status=="nonsmoker"]
> y=table$Plasma.Ascorbic.Acid[table$Status=="smoker"]
```

Here are some descriptive statistics.

```
> mean(x)
[1] 0.9158333
> sd(x)
[1] 0.2144136
```

```
> length(x)
[1] 24
> mean(y)
[1] 0.97625
> sd(y)
[1] 0.3914784
> length(y)
[1] 8
```

(a) Give a 95% confidence interval for the difference between the mean plasma ascorbic acid level for the non-smokers and the mean plasma ascorbic acid level for the smokers. Assume that the two sets of data come from independent normal populations with unequal variances.

(b) Based on the confidence interval from part (a), can we conclude that the mean plasma ascorbic acid level for the non-smokers is different compared to the mean level for smokers?

Part II) Answer the following 2 questions **with** the use of R.

Remarks:

- You must provide the R commands and output that were used in answering the question.
- The R output alone is not an answer to a question. The R output is used to support your answer.
- Please do not printout your whole R session. Only provide the R commands and output that are necessary to answer the question.

5) We wish to compare the effect of two preparations of a virus on tobacco plants. The study involves 8 plants. For each plant, half a leaf is inoculated with preparation 1 and the other half is inoculated with preparation 2. The number of lesions are measured and denoted as x_1 and x_2, respectively. The data are found in the file *TOBACCO.txt*.

   (a) Is it reasonable to assume that the difference of both measurements is normally distributed? (*Hint:* Produce a quantile-quantile plot for the differences.)

   (b) Do the preparations have a different effect on the tobacco plants? Justify your answer using a test of hypothesis of $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$, where μ_1 is the mean number of lesions when preparation 1 is used and μ_2 is the mean number of lesions when preparation 2 is used. Use a level of significance of $\alpha = 1\%$.

   (c) Give a 95% confidence interval for $\mu_1 - \mu_2$. Using this interval, write 1 or 2 sentences to compare the mean effect of the two preparations.

   (d) What is the conclusion of a test of hypothesis of $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$ with $\alpha = 1\%$, if we had assumed (incorrectly) that these are random samples from two independent normal populations with equal variances?

6) In a study conducted at Virginia Tech on the development of ectomycorrhizal, a symbiotic relationship between between the roots of trees and a fungus, in which minerals are transferred from the fungus to the trees and sugars from the trees to the fungus, 20 northern red oak seedlings exposed to the fungus *Pisolithus tinctorus* were grown in a greenhouse. All seedlings were planted in the same type of soil and received the same amount of sunshine and water. Half received no nitrogen at planting time, to served as a control, and the other half received 368 ppm of nitrogen in the form $NaNO_3$. The stem weights, in grams, at the end of 140 days were recorded and are found in the file *nitrogen.txt*.

   (a) Produce overlayed quantile-quantile plots for the two groups of stem weights and produce side-by-side boxplots for the two groups of stem weights. Is it reasonable to assume that the stem weight is

normally distributed? Is it reasonable to assume that the variance of the stem weight is the same for both groups?

(b) Does the use of nitrogen have a significant effect on the stem weights? (Use $\alpha = 5\%$.)

(c) Give a 95% confidence interval for the difference between the mean weight of the stems with nitrogen and the mean weight of the stems without nitrogen.