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.

[**6**] 1)

(a) We want to test

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{against} \quad H_1 : \mu_1 - \mu_2 < 0.$$

We have two independent normal population with equal variances. We will use a two-sample t-test to test these hypotheses.

The observed value of the test statistic is

$$t_0 = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{3.6 - 7.1}{0.53852\sqrt{1/7 + 1/7}} = -12.16,$$

where

$$\begin{aligned} s_p &= \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_1^2}{n_1+n_2-2}} \\ &= \sqrt{\frac{(7-1)(0.3)^2 + (7-1)(0.7)^2}{7+7-2}} = 0.53852. \end{aligned}$$

Since it is a left-sided alternative, then the p-value is $P(T < -12.16)$, where T has a $t(n_1 + n_2 - 2) = t(12)$ distribution. From our table for the t distribution, we conclude that the p-value is less than 0.005. Hence at a level of significance of 5%, we can conclude that the administration of R6G reduces the tumor growth rate on average.

(b) A 95% confidence interval for $\mu_1 - \mu_2$ is

$$\overline{x}_1 - \overline{x}_2 \pm t_{0.025;n_1+n_2-2}\, s_p \sqrt{1/n_1 + n_2} = [-4.13, -2.87],$$

where $t_{0.025;n_1+n_2-2} = t_{0.025;12} = 2.179$. Since all the values in the confidence interval are negative, then we are highly confident that $\mu_1 < \mu_2$. We can conclude that the administration of R6G reduces the tumor growth rate on average.

[4] 2) Assuming that the single factor hypothesis is true, then using a Punnett square (or a probability tree), we compute that the probability that a plant in the F_2 generation is iron-inefficient (i.e. ff) is $p = 0.25$. We want to test

$$H_0: \; p = 0.25 \quad \text{against } H_1: \; p \neq 0.25.$$

The observed value of the test statistic is

$$z_0 = \frac{\hat{p} - 0.25}{\sqrt{(0.25)(0.75)/245}} = -0.47,$$

where $\hat{p} = 58/245$. Since it is a two-sided alternative, then the p-value is

$$2\,P(Z > 0.47) = 2\,(1 - \Phi(0.47)) = 0.6384.$$

At a level of significance of 5%, the evidence against the single factor hypothesis is not significant.

[4] 3) Let μ be the mean weight of a trout post-eruption. We want to test

$$H_0 : \mu = 1.9 \quad \text{against} \quad H_1 : \mu > 1.9.$$

The observed value of the test statistic is

$$t_0 = \frac{\overline{x} - 1.9}{s/\sqrt{n}} = \frac{2.3 - 1.9}{4/\sqrt{30}} = 0.548.$$

Since it is a right-sided alternative, the p-value is $P(T > .548)$, where T has a $t(29)$ distribution. From our table for the t distribution, we obtain

$$0.25 < p\text{-value} < 0.4.$$

At a level of significance of 5%, we do not have significant evidence that the mean weight of the fish has increased after the eruption.

[6] 4)

(a) Let μ_1 be the mean plasma ascorbic acid level for the non-smokers and μ_2 be the mean plasma ascorbic acid level for the smokers. Since we are assuming independent normal populations with unequal variances, then a 95% confidence interval for $\mu_1 - \mu_2$ is

$$\begin{aligned} & \overline{x}_1 - \overline{x}_2 \pm t_{0.025;\nu}\sqrt{\frac{0.2144136^2}{24} + \frac{0.3914784^2}{8}} \\ = \; & 0.9158333 - 0.97625 \pm t_{0.025;\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ = \; & [-0.389, 0.268], \end{aligned}$$

where the degrees of freedom is computed with Welch's approximation

$$\nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)} = 8.44.$$

Rounding up gives $\nu = 9$ and so $t_{0.025;\nu} = t_{0.025;9} = 2.262$.

(b) Since 0 does belong the confidence interval computed in part (a), then it is plausible that $\mu_1 = \mu_2$. We should **not** 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.

[**10**] 5)

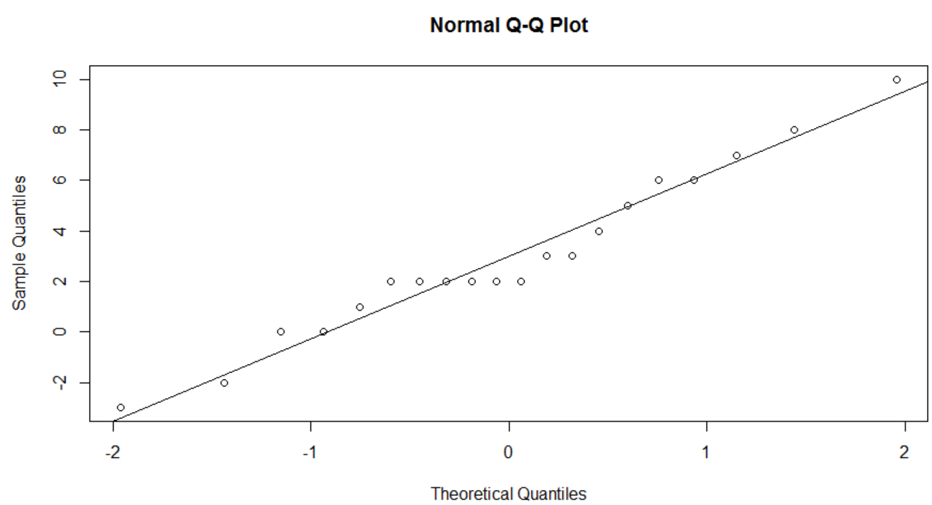
We will start by assigning the data to the dataframe *table*. We display the names of the columns in this dataframe and define a variable d which will be the difference of the measurement for preparation 1 and the measurement for preparation 2. Here are the commands and output.

```
> table = read.table(file.choose(),header=TRUE,sep="\t")
> names(table)
[1] "preparation.1" "preparation.2"
> d=table$preparation.1-table$preparation.2
```

(a) Here are the commands that we used to produce a quantile-quantile plot for the differences.

```
> qqnorm(d)
> abline(mean(d),sd(d))
```

The quantile-quantile plot is found below. Since there is a linear tendency in the plot, then it is reasonable to assume that the differences are normally distributed.



(b) To test $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$, we use a t-test for paired measurements. Here is the command and corresponding output.

```
> t.test(table$preparation.1,table$preparation.2,paired=TRUE)

        Paired t-test

data:  table$preparation.1 and table$preparation.2
t = 4.1147, df = 19, p-value = 0.0005896
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.473987 4.526013
sample estimates:
mean of the differences
                      3
```

Since the p-value (which is 0.0005896) is less than the level of significance of 1%, then we have significant evidence the number number of lesions are different when comparing the two preparations. We can conclude that the preparations have a different effect on the tobacco plants.

(c) A 95% confidence interval for $\mu_1 - \mu_2 = \mu_D$ is [1.473987, 4.526013], refer to the R output from part (b). We are highly confident that on average there are between 1.5 to 4.5 more lesions on a plant when using preparation 1 compared to preparation 2.

(d) If we had incorrectly considered the two groups of measurements as being independent, then we would have used a two-sample t-test for the comparison. The R output is found below. The p-value of 0.05898 is larger than the level of significance of 1%. This means that we would have concluded that the evidence against the equality of means is not significant.

```
> t.test(table$preparation.1,table$preparation.2,var.equal=TRUE)

        Two Sample t-test

data:  table$preparation.1 and table$preparation.2
```

```
t = 1.9468, df = 38, p-value = 0.05898
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1196281  6.1196281
sample estimates:
mean of x mean of y
     15.7      12.7
```

[6] 6) We have two samples from independent populations. We assign the data to the dataframe *table* and display the names of the columns.

```
> table = read.table(file.choose(),header=TRUE,sep="\t")
> names(table)
[1] "Stem.Weight" "Nitrogen"
```

We assign the stew weights with no nitrogen to x and the stem weights with nitrogen to y.

```
> x=table$Stem.Weight[table$Nitrogen=="no"]
> y=table$Stem.Weight[table$Nitrogen=="yes"]
```

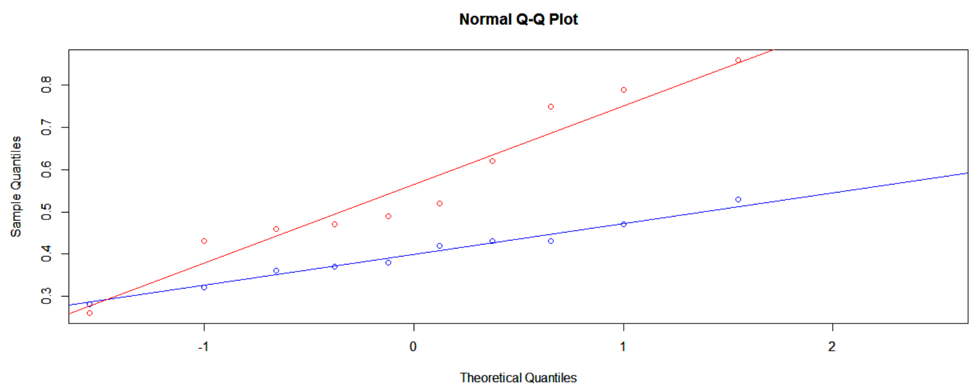
(a) Here are the commands to produce the overlayed quantile-quantile plot of the stem weights.

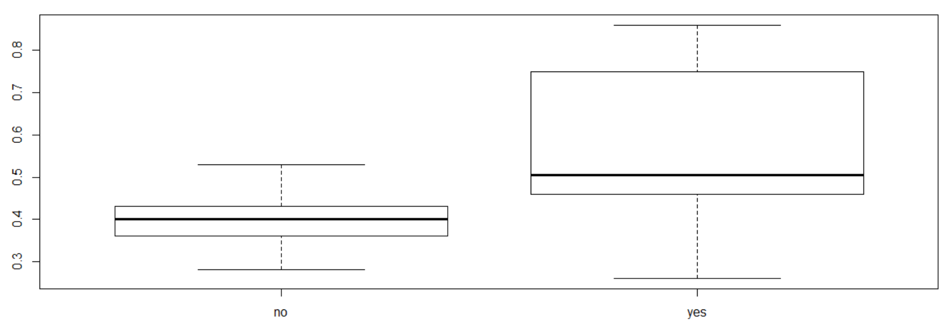
```
> vlmts=range(x,y)
> hlmts=range(qnorm(ppoints(x),ppoints(y)))
> qqnorm(x,ylim=vlmts,xlim=hlmts,col="blue")
> abline(mean(x),sd(x),col="blue")
> par(new=T) # begin overlay
> qqnorm(y,ylim=vlmts,xlim=hlmts,col="red")
> abline(mean(y),sd(y),col="red")
> par(new=F) # end overlay
```

Here is the command to produce the side-by-side boxplot of the stem weights.

```
> boxplot(table$Stem.Weight~table$Nitrogen)
```

Below are the graphs.





Since the tendencies in the quantile-quantile plots are both linear, then it is reasonable to assume that the populations are normal. However these tendencies are far from being parallel, so it is not reasonable to assume that the population variances are equal. We observe in the boxplots that the stem weights with nitrogen are much more dispersed than the stem weights without nitrogen.

(b) Since we have samples from normal populations with unequal variances, we will use the Welch two-sample t-test to compare the means. Let μ_1 be the mean stem weight with nitrogen and μ_2 be the mean stem weight without nitrogen. We want to test $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$. The p-value is 0.02286. Since the p-value is less than 5%, then there are significant evidence that $\mu_1 \neq \mu_2$. So the use of nitrogen has a significant effect on the stem weights.

```
> t.test(y,x)

        Welch Two Sample t-test

data:  y and x
t = 2.6191, df = 11.673, p-value = 0.02286
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.02747562 0.30452438
sample estimates:
mean of x mean of y
    0.565     0.399
```

(c) We are 95% confident that the difference between the mean weight of the stems with nitrogen and the mean weight of the stems without nitrogen is between 0.03 to 0.30 grams, refer to the R output from part (b).

[/**36**]