Author: Min Ren

Lab 7: One-Way ANOVA
Objectives: Analyze data via the One-Way ANOVA

A. (50 pts.) Do isoflavones increase bone mineral density? (ex12-45bmd.txt) Kudzu is a plant that was imported to the United States from Japan and now covers over seven million acres in the South. The plant contains chemicals called isoflavones that have been shown to have beneficial effects on bones. One study used three groups of rats to compare a control group with rats that were fed either a low dose or a high dose of isoflavones from kudzu. One of the outcomes examined was the bone mineral density in the femur (in grams per square centimeter). Here are the data:

Treatment	Bone mineral density (g/cm^2)							
Control	0.228	0.207	0.234	0.220	0.217	0.228	0.209	0.221
	0.204	0.220	0.203	0.219	0.218	0.245	0.210	
Low dose	0.211	0.220	0.211	0.233	0.219	0.233	0.226	0.228
	0.216	0.225	0.200	0.208	0.198	0.208	0.203	
High dose	0.250	0.237	0.217	0.206	0.247	0.228	0.245	0.232
	0.267	0.261	0.221	0.219	0.232	0.209	0.255	

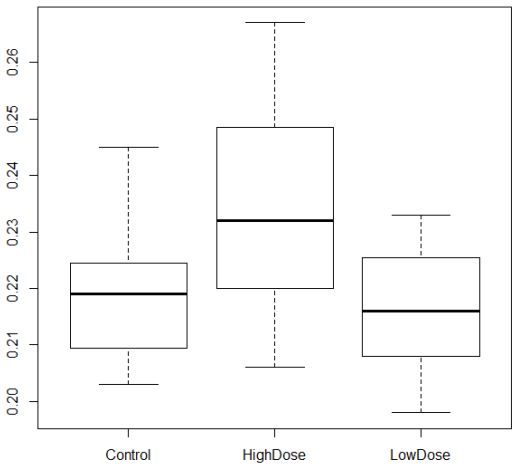
Solution:

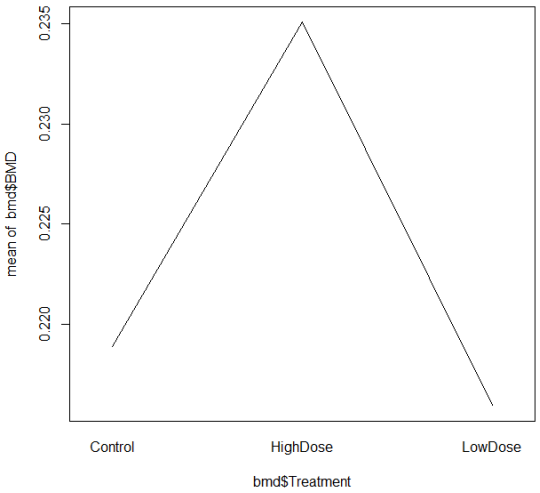
```
> bmd <- read.table("ex12-45bmd.txt",header=T)
> bmd
```

1. (10 pts.) Make side-by-side boxplots and an effects plot of the data. Also, make a table giving the sample size, mean, and standard deviation for each treatment group. From this information, do you think that all of the means are the same? Be sure to comment on each of the plots.

Solution:

```
> boxplot(BMD~Treatment,bmd)

> trace <- rep(1,length(bmd$Group))
> interaction.plot(bmd$Treatment,trace,bmd$BMD,fun=mean,legend=F)

> attach(bmd)

> tapply(BMD, Treatment, length)
 Control HighDose  LowDose
      15       15       15

> tapply(BMD, Treatment, mean)
  Control  HighDose   LowDose
0.2188667 0.2350667 0.2159333

> tapply(BMD, Treatment, sd)
   Control   HighDose    LowDose
0.01158735 0.01877105 0.01151066
```





We can see that although the means are different, we cannot say that they are significantly different since the boxplots overlap each other.	Here the means look different with HiDose being larger than the rest.

Group	n	sample mean	sample standard deviation
Control	15	0.2188667	0.01158735
HighDose	15	0.2350667	0.01877105
LowDose	15	0.2159333	0.01151066

It is hard to see if HighDose is large than the other two based on the means and the standard deviations.

All reasonable answers are acceptable as long as there is a comment for each plot/table.

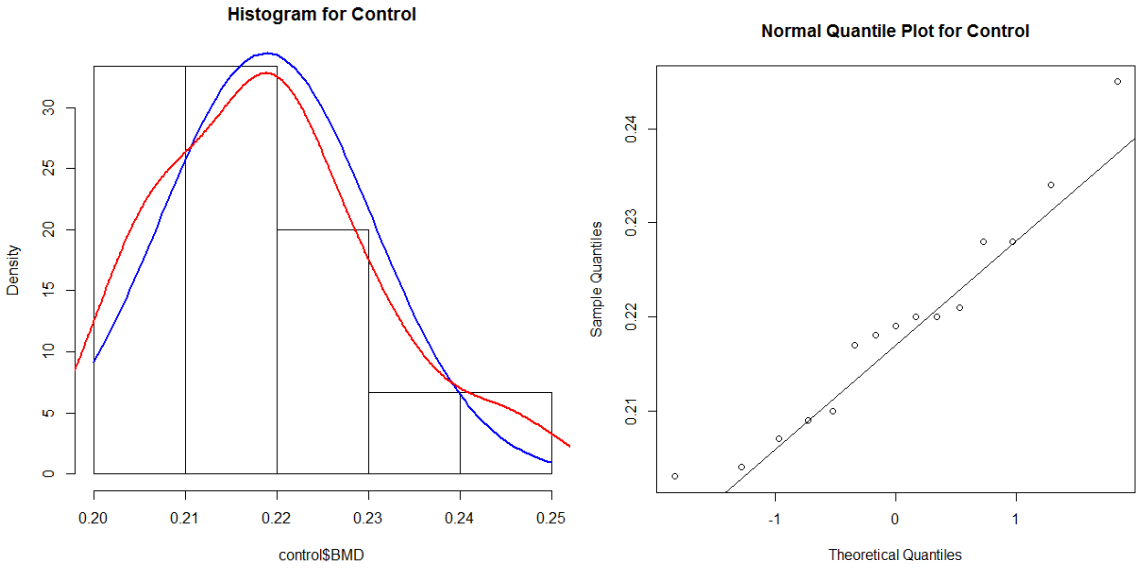
2. (10 pts.) Examine the assumptions necessary for ANOVA. Is it appropriate to continue the analysis? Be sure to state each of the assumptions and comment on each of them using the appropriate plots/data. Remember, you need to generate the normal probability plots and histograms for each population.

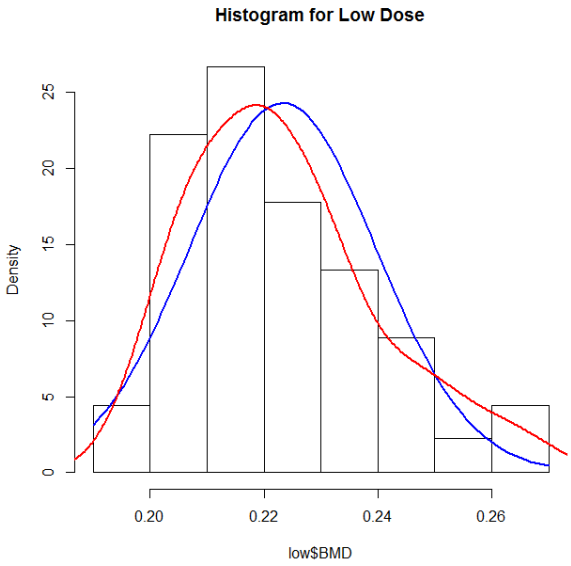
Solution:

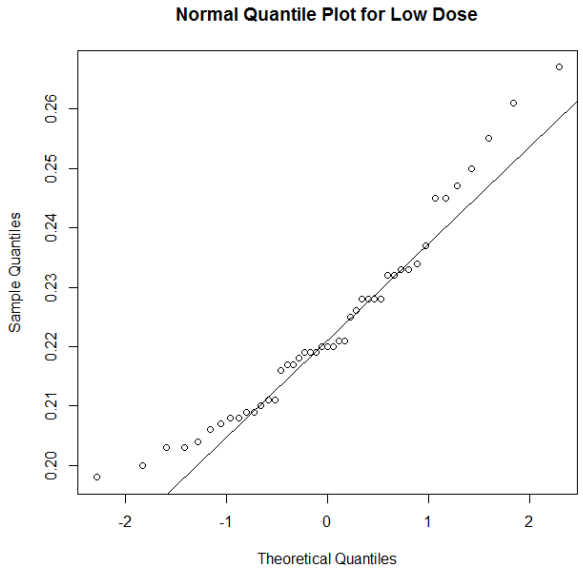
```
> control <- subset(bmd, Treatment == "Control")
> high <- subset(bmd, Treatment = "HighDose")
> low <- subset(bmd, Treatment = "LowDose")
>
> hist(control$BMD, freq = FALSE, main="Histogram for Control")
> curve(dnorm(x, mean=mean(control$BMD), sd=sd(control$BMD)), col="blue", lwd=2,
    add=TRUE)
> lines(density(control$BMD),col = "red", lwd=2)

> qqnorm (control$BMD, main = "Normal Quantile Plot for Control")
> qqline (control$BMD)

> hist(low$BMD, freq = FALSE, main="Histogram for Low Dose")
> curve(dnorm(x, mean=mean(low$BMD), sd=sd(low$BMD)), col="blue", lwd=2,
    add=TRUE)
> lines(density(low$BMD),col = "red", lwd=2)

> qqnorm (low$BMD, main = "Normal Quantile Plot for Low Dose")
> qqline (low$BMD)
```

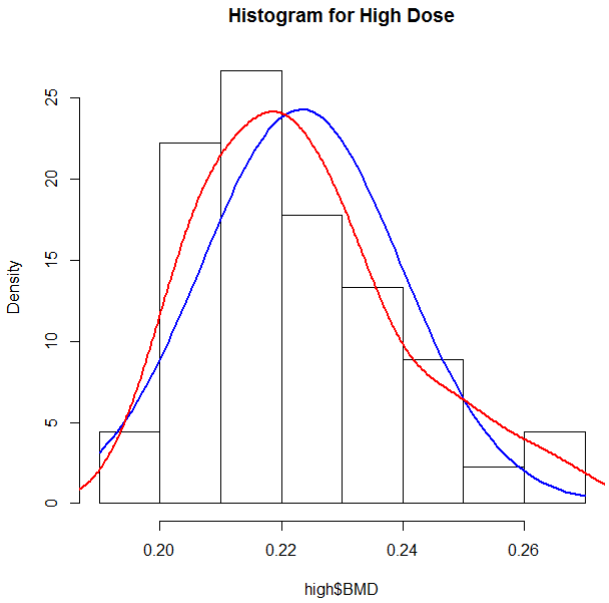
```
> hist(high$BMD, freq = FALSE, main="Histogram for High Dose")
> curve(dnorm(x, mean=mean(high$BMD), sd=sd(high$BMD)), col="blue", lwd=2,
    add=TRUE)
> lines(density(high$BMD),col = "red", lwd=2)

qqnorm (high$BMD, main = "Normal Quantile Plot for High Dose")
qqline (high$BMD)
```

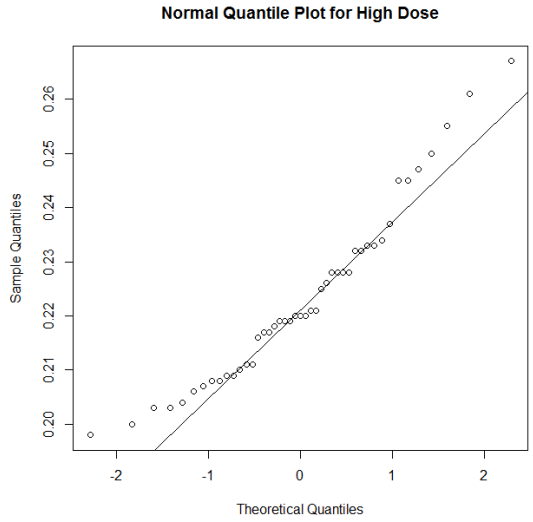
Normality:




Histogram for Low Dose
Density
0
5
10
15
20
25
0.20
0.22
0.24
0.26
low$BMD


Normal Quantile Plot for Low Dose
Sample Quantiles
0.20
0.21
0.22
0.23
0.24
0.25
0.26
-2
-1
0
1
2
Theoretical Quantiles





With a sample size of 15 *3 = 45, these distributions are close enough to being normal.

Constant standard deviation

$$\frac{s_{max}}{s_{min}} = \frac{0.0187711}{0.0115107} = 1.63 < 2$$

Therefore the constant standard deviation assumption is valid.

The other two assumptions, the data comes from an SRS and that the populations are independent of each other are assumed to be correct.

Therefore, it is appropriate to continue the analysis.

3. (15 pts.) Run the ANOVA and report the results of the significance test (4 steps) using a significance level of 0.05. Are you results in this step consistent with part 1?.

Solution:

```
> fit <- aov(BMD ~ Treatment)
```

```
> summary(fit)
            Df   Sum Sq   Mean Sq F value Pr(>F)
Treatment    2 0.003186 0.0015928   7.718 0.0014 **
Residuals   42 0.008668 0.0002064
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step 0: Definition of the terms
μ_1 is the population mean BMD score for the Control group.
μ_2 is the population mean BMD score for the HighDose group.
μ_3 is the population mean BMD score for the LowDose group.

Step 1: State the hypotheses
H_0: $\mu_1 = \mu_2 = \mu_3$
H_a: at least one μ_i is different.

Step 2: Find the *Test Statistic*.
$F_t = 7.718$

Step 3: Find the *p-value, report DF.*
DF1 = 2, DF2 = 42
P-value = 0.0014

Step 4: Conclusion:
$\alpha = 0.05$
Since $0.0014 < 0.05$, we should reject H_0
The data provides sufficiently strong evidence (P-value = 0.0014) to the claim that the population mean values of at least one of the methods is different from the rest.

This answer depends on what you stated in part 1. However, I did state that the means were a little different but probably not significantly different. Therefore the results are not consistent.

4. (10 pts.) Use an appropriate multiple-comparison method to determine if isoflavones from kudzu have any effect on bones. Explain why you chose this method. Present a graphical representation of the results if appropriate. Write a short statement on your conclusion.

Solution:
Use only one of the following:

Bonforroni:

This method was chosen because we want to compare all of the means in a pairwise fashion.

```
> pairwise.t.test(BMD,Treatment,p.adjust="bon")

        Pairwise comparisons using t tests with pooled SD

data:  BMD and Treatment

         Control HighDose
HighDose 0.0107  -
LowDose  1.0000  0.0022

P value adjustment method: bonferroni
```

LowDose	Control	HighDose
0.2159333	0.2188667	0.2350667

The above results show that the Control and low dose level have similar effect on bones while high dose does increase BMD.

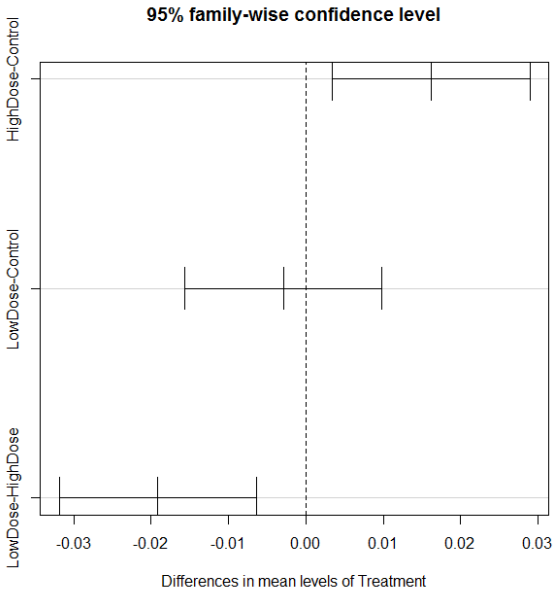
Tukey:

This method was chosen because we want to compare all of the means in a pairwise fashion.

```
> test.Tukey <- TukeyHSD(fit,conf.level=0.95)
> test.Tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = BMD ~ Treatment)

$Treatment
                          diff          lwr          upr     p adj
HighDose-Control  0.016200000  0.003455877  0.028944123 0.0097645
LowDose-Control  -0.002933333 -0.015677456  0.009810789 0.8423308
LowDose-HighDose -0.019133333 -0.031877456 -0.006389211 0.0020537
```

The following is optional:

```
> plot(test.Tukey)
```



LowDose	Control	HighDose
0.2159333	0.2188667	0.2350667

The above results show that the Control and low dose level have similar effect on bones while high dose does increase BMD.

5. (5 pts.) Write a short report explaining the effect of kudzu isoflavones on the femur of the rat. Be sure to answer the question posed in the initial statement and whether this data can be generalized to humans or not with a reason. This paragraph should be written in complete English sentences and should be understandable to someone who has not taken a course in Statistics.

Solution:
From the original question, we want to determine the effect of kudzu isoflavones on the bone density of femus of the rat. From the study above, we can conclude that it is appropriate to use ANOVA for this data and ANOVA does suggest that not all the means of the three groups are the same. Further analysis shows that a high dosage of kudzu isoflavones has a significant effect on the femur bone density of the rat. In addition, a low dosage doesn't have a significant effect on the femur bone density of the rat. Therefore, high doses of kudzu isoflavones do increase BMD. Knowing that the bones of rats and the bones of humans are similar, I would expect that kudzu isoflavones would have a similar effect on humans; however further studies would be warranted to be sure that this is true.

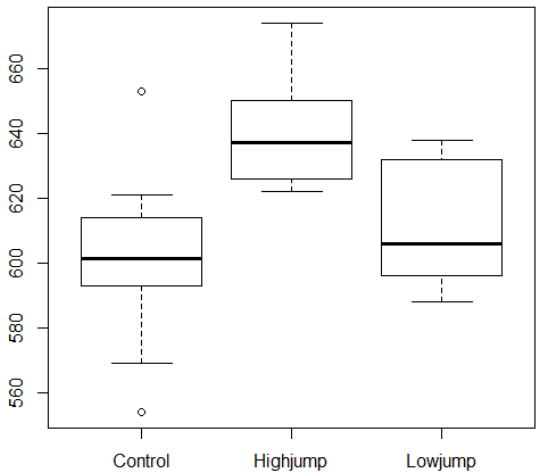
B. (50 pts.) Exercise and healthy bones. (ex12-47jump.txt) Many studies have suggested that there is a link between exercise and healthy bones. Exercise stresses the bones and this causes them to get stronger. One study examined the effect of jumping on the bone density of growing rats. There were three treatments: a control with no jumping, a low-jump condition (the jump height was 30 centimeters), and a high-jump condition (60 centimeters). After 8 weeks of 10 jumps per day, 5 days per week, the bone density of the rats (expressed in milligrams per cubic centimeter) was measured. Here are the data:

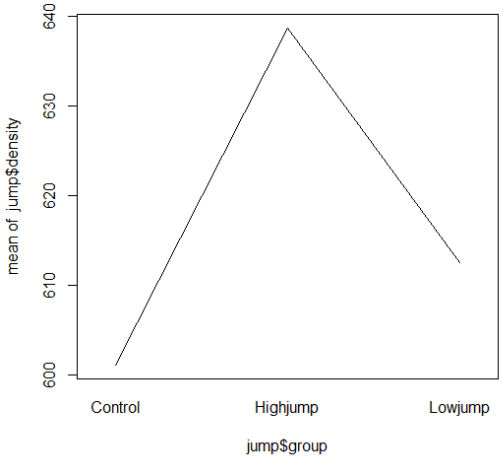
Group	Bone density (mg/cm^3)									
Control	611	621	614	593	593	653	600	554	603	569
Low jump	635	605	638	594	599	632	631	588	607	596
High jump	650	622	626	626	631	622	643	674	643	650

Solution:
```
> jump <- read.table("ex12-47jump.txt",header=T)
> jump
```

1. (10 pts.) Make side-by-side boxplots and an effects plot of the data. Also, make a table giving the sample size, mean, and standard deviation for each group of rats. From this information, do you think that all of the means are the same? Be sure to comment on each of the plots.

Solution:
```
> boxplot(density~group,jump)

> trace <- rep(1,length(jump$group))
> interaction.plot(jump$group,trace,jump$density,fun=mean,legend=F)

> attach(jump)

> tapply(density, group, length)
 Control Highjump  Lowjump
      10       10       10

> tapply(density, group, mean)
 Control Highjump  Lowjump
   601.1    638.7    612.5

> tapply(density, group, sd)
 Control Highjump  Lowjump
27.36360 16.59351 19.32902
```





We can see that the group means are significantly different, especially between HighJump and the other two levels of jumping.	We can see that the group means are significantly different, especially between HighJump and the other two levels of jumping.

Group	n	sample mean	sample standard deviation
Control	10	601.1	27.36360
HighJump	10	638.7	16.59351
LowJump	10	612.5	19.32902

We can see that the group means are significantly different, especially between HighJump and the other two levels of jumping.

All reasonable answers are acceptable as long as there is a comment for each plot/table.

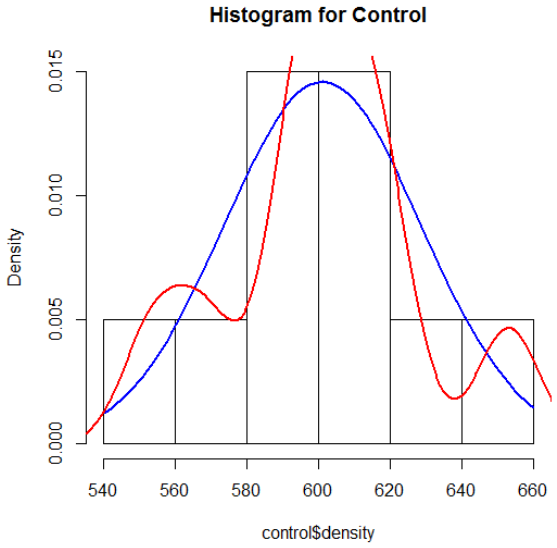
2. (10 pts.) Examine the assumptions necessary for ANOVA. Is it appropriate to continue the analysis? Be sure to state each of the assumptions and comment on each of them using the appropriate plots/data. Remember, you need to generate the normal probability plots and histograms for each population.

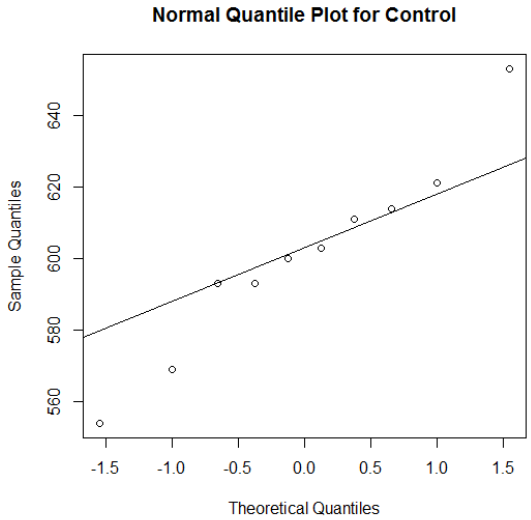
Solution:

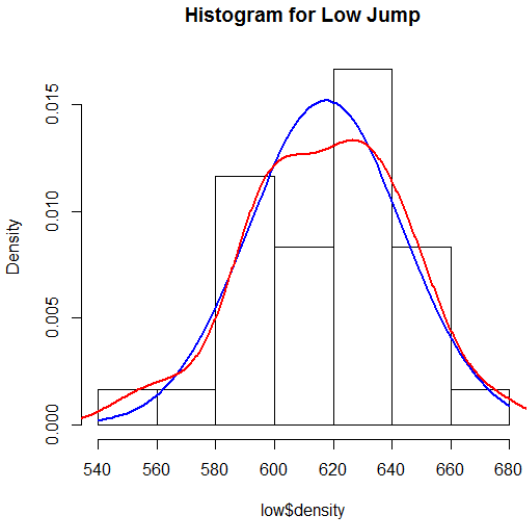
```
> control <- subset(jump, group == "Control")
> high <- subset(jump, group = "Highjump")
> low <- subset(jump, group = "Lowjump")

> hist(control$density, freq = FALSE, main="Histogram for Control")
> curve(dnorm(x, mean=mean(control$density), sd=sd(control$density)),
    col="blue", lwd=2, add=TRUE)
> lines(density(control$density),col = "red", lwd=2)

> qqnorm (control$density, main = "Normal Quantile Plot for Control")
> qqline (control$density)

> hist(low$density, freq = FALSE, main="Histogram for Low Jump")
> curve(dnorm(x, mean=mean(low$density), sd=sd(low$density)), col="blue", lwd=2,
    add=TRUE)
> lines(density(low$density),col = "red", lwd=2)

> qqnorm (low$density, main = "Normal Quantile Plot for Low Jump")
> qqline (low$density)

> hist(high$density, freq = FALSE, main="Histogram for High Jump")
```

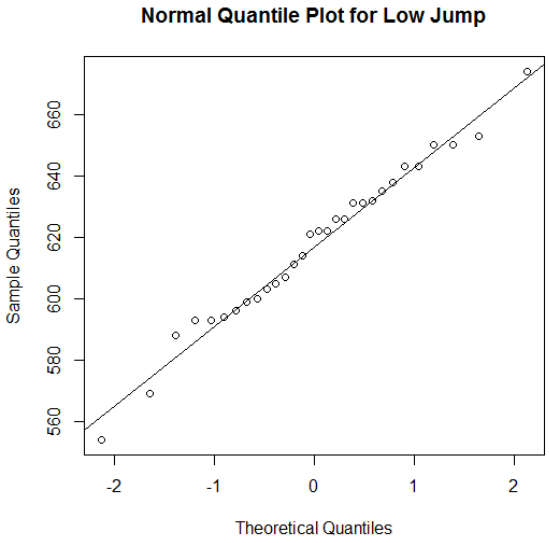
```
> curve(dnorm(x, mean=mean(high$density), sd=sd(high$density)), col="blue",
    lwd=2, add=TRUE)
> lines(density(high$density),col = "red", lwd=2)

> qqnorm (high$density, main = "Normal Quantile Plot for High Jump")
> qqline (high$density)
```

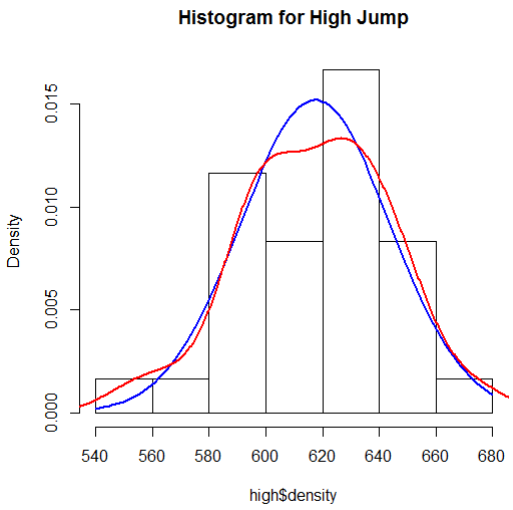
Normality

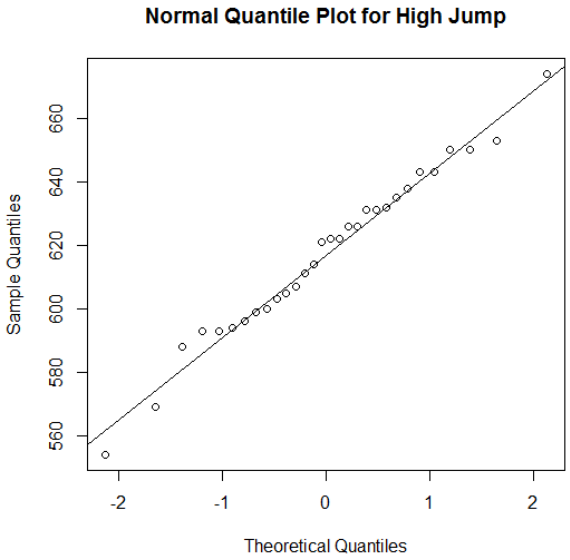





Histogram for Low Jump
Density
0.000
0.005
0.010
0.015
540
560
580
600
620
640
660
680
low$density


Normal Quantile Plot for Low Jump
Sample Quantiles
560
580
600
620
640
660
-2
-1
0
1
2
Theoretical Quantiles





With a sample size of 10 * 3 = 40, even though these distributions are close to normal, they are still acceptable.

Constant standard deviation

$$\frac{s_{max}}{s_{min}} = \frac{27.36360}{16.59351} = 1.65 < 2$$

Therefore the constant standard deviation assumption is valid.

The other two assumptions, the data comes from an SRS and that the populations are independent of each other are assumed to be correct.

Therefore, it is appropriate to continue the analysis.

3. (15 pts.) Run the ANOVA and report the results of the significance test (4 steps) using a significance level of 0.01. Are you results in this step consistent with part 1?

Solution:

```
> fit <- aov(density ~ group)
> summary(fit)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        2   7434    3717   7.978 0.0019 **
Residuals   27  12579     466
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step 0: Definition of the terms
μ_1 is the population mean density score for the Control group.
μ_2 is the population mean density score for the HighJump group.
μ_3 is the population mean densigy score for the LowJump group.

Step 1: State the hypotheses
H_0: $\mu_1 = \mu_2 = \mu_3$
H_a: at least one μ_i is different.

Step 2: Find the *Test Statistic*.
$F_t = 7.978$

Step 3: Find the *p-value, report DF.*
DF1 = 2, DF2 = 27
P-value = 0.0019

Step 4: Conclusion:
$\alpha = 0.01$
Since $0.0019 < 0.01$, we should reject H_0
The data provides sufficiently strong evidence (P-value = 0.0019) to the claim that the population mean values of at least one of the methods is different from the rest.

Yes, the results are consistent with part 1) because in both cases, there is at least one mean which is different from the rest.

4. (10 pts.) Use an appropriate multiple-comparison method to determine if the height of jumping effects the bone density of growing rats. Present a graphical representation of the results if appropriate. Write a short statement on your conclusion.

Solution:

Use only one of the following:

Bonforroni:

```
> pairwise.t.test(density,group,p.adjust="bon")

        Pairwise comparisons using t tests with pooled SD

data:  density and group

         Control Highjump
Highjump 0.0018  -
Lowjump  0.7437  0.0343

P value adjustment method: bonferroni
```

Control	LowJump	HighJump
601.1	612.5	638.7

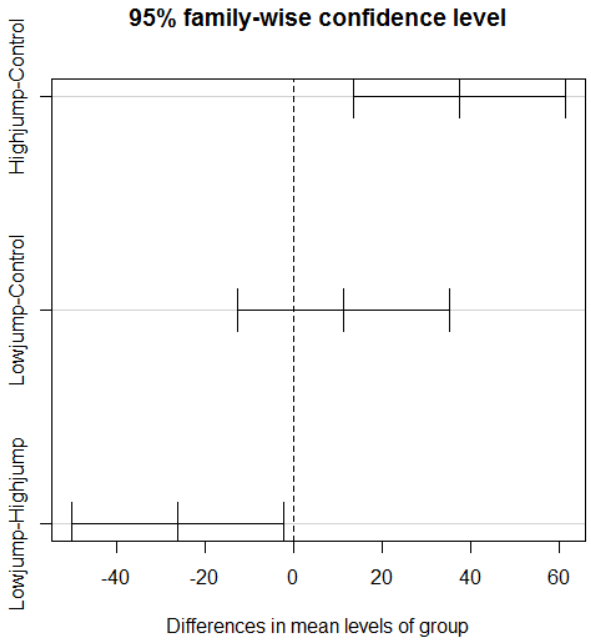
We can see that the mean bone density for high jump and low jump is the same. In addition, the mean bone density for control and low jump are the same.

Tukey

```
> test.Tukey <- TukeyHSD(fit,conf.level=0.95)
> test.Tukey
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = density ~ group)

$group
                    diff       lwr       upr     p adj
Highjump-Control   37.6  13.66604 61.533957 0.0016388
Lowjump-Control    11.4 -12.53396 35.333957 0.4744032
Lowjump-Highjump -26.2 -50.13396 -2.266043 0.0297843
```

The following is optional:

```
> plot(test.Tukey)
```



Control	LowJump	HighJump
601.1	612.5	638.7

We can see that the mean bone density for high jump and low jump is the same. In addition, the mean bone density for control and low jump are the same.

5. (5 pts.) Write a short report explaining the effect of the height of jumping on the bone density of growing rats. Be sure to answer the question of the study concerning whether there is a link between exercise (jumping) and the development of healthy bones and whether this data can be generalized to humans or not with a reason. This paragraph should be written in complete English sentences and should be understandable to someone who has not taken a course in Statistics.

Solution:
From the original question, we want to determine the effect of height of jumping on the bone density of growing rats. From the study above, we conclude that it is appropriate to use ANOVA for this data and ANOVA does suggest that not all the means of the three groups are the same. Further analysis shows that while bone density remains nearly unchanged for low jump and control group, the high jump does increase bone density. Even though the bones of rats and humans are similar, I do not know if the growth dynamics are the same. Therefore, further tests would need to be performed to determine if there is a similar effect in humans.