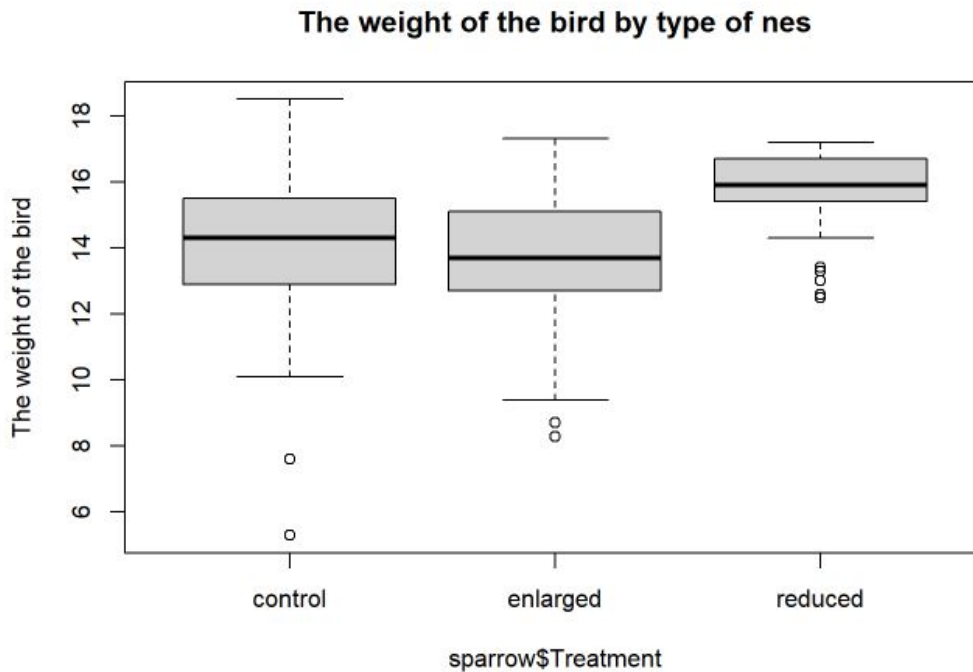
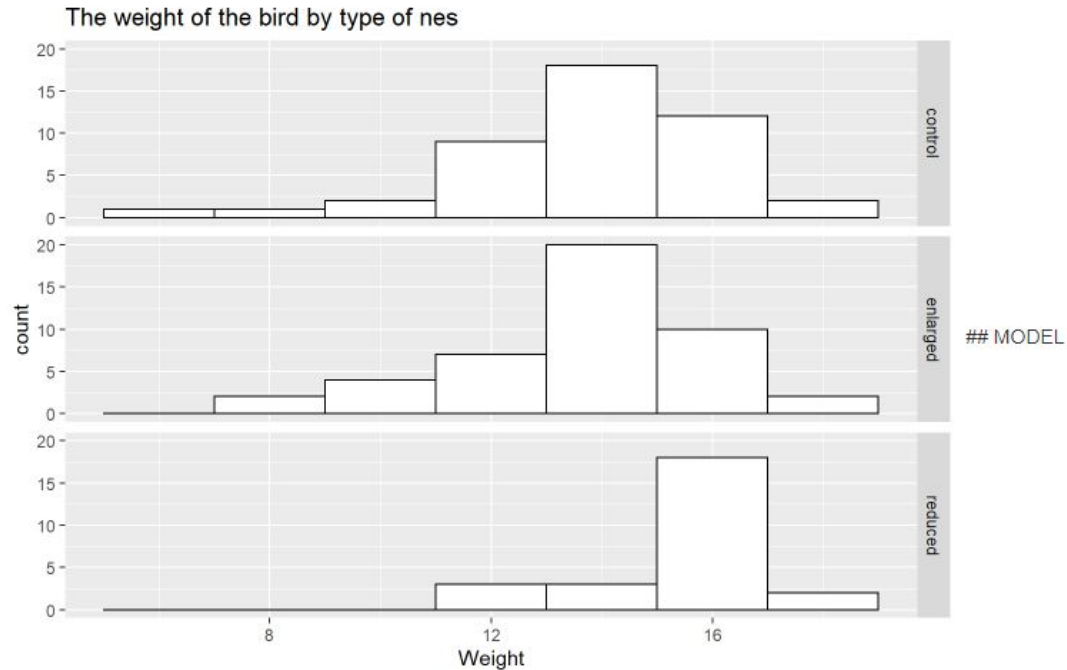


## 1. Introduction

We investigate the first problem to see if different nests for sparrows on the island attracted different size sparrows. And we also have the data which contains 3 treatments and weight for those different conditions. In this way, if we try to detect the mean for each condition are the same or not so that we can know which condition will increase the weight or which one can have less weight for the bird. And we are going to use .. , anova table, diagnostic test, and so on to detect whether the mean for each group are the same or not.





```
## Treatment Weight
## 1 control 13.92444
## 2 enlarged 13.51556
## 3 reduced 15.56923
```

```
## Treatment Weight
## 1 control 2.419631
## 2 enlarged 2.103996
## 3 reduced 1.459252
```

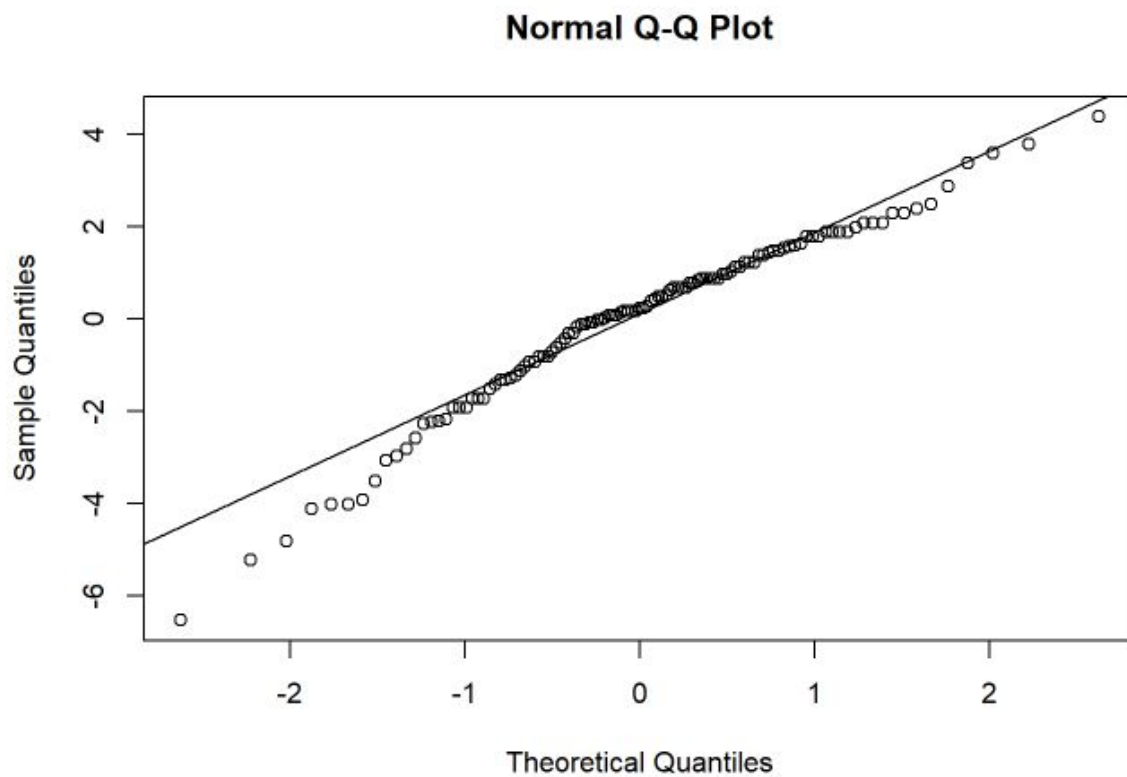
```
## Treatment Weight
## 1 control 14.3
## 2 enlarged 13.7
## 3 reduced 15.9
```

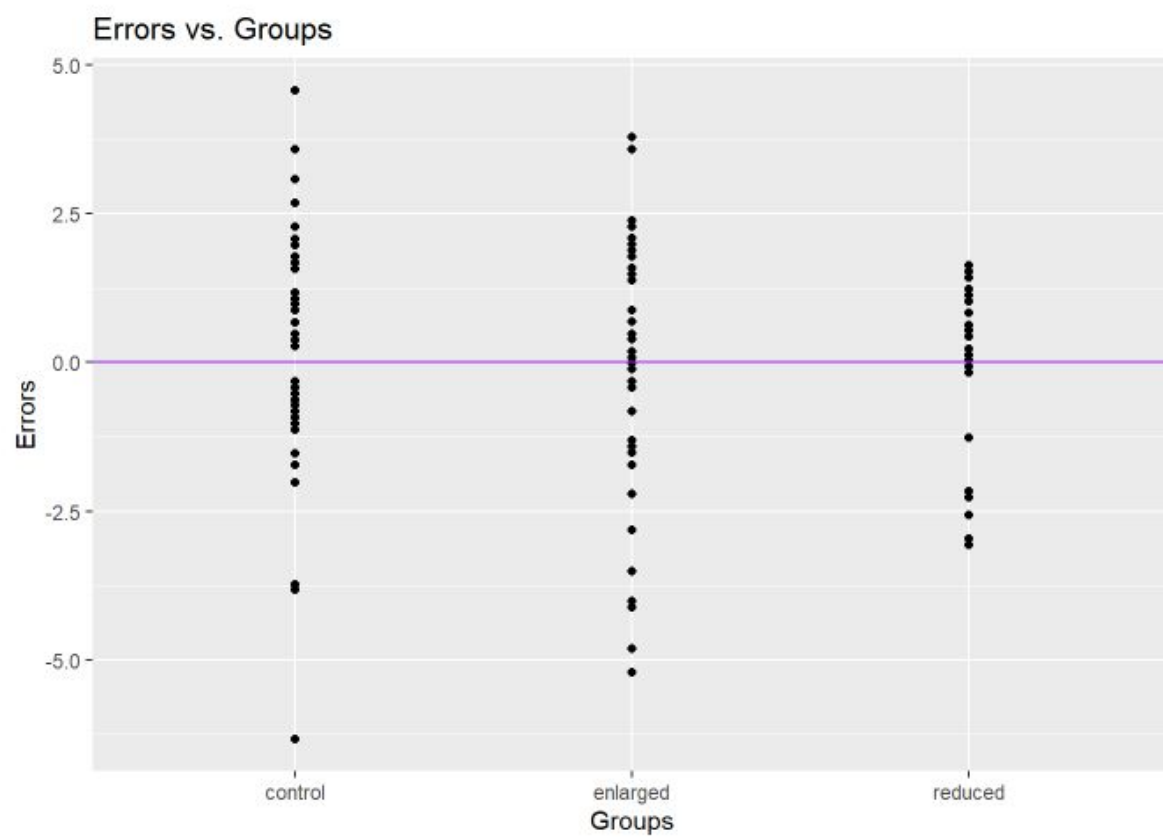
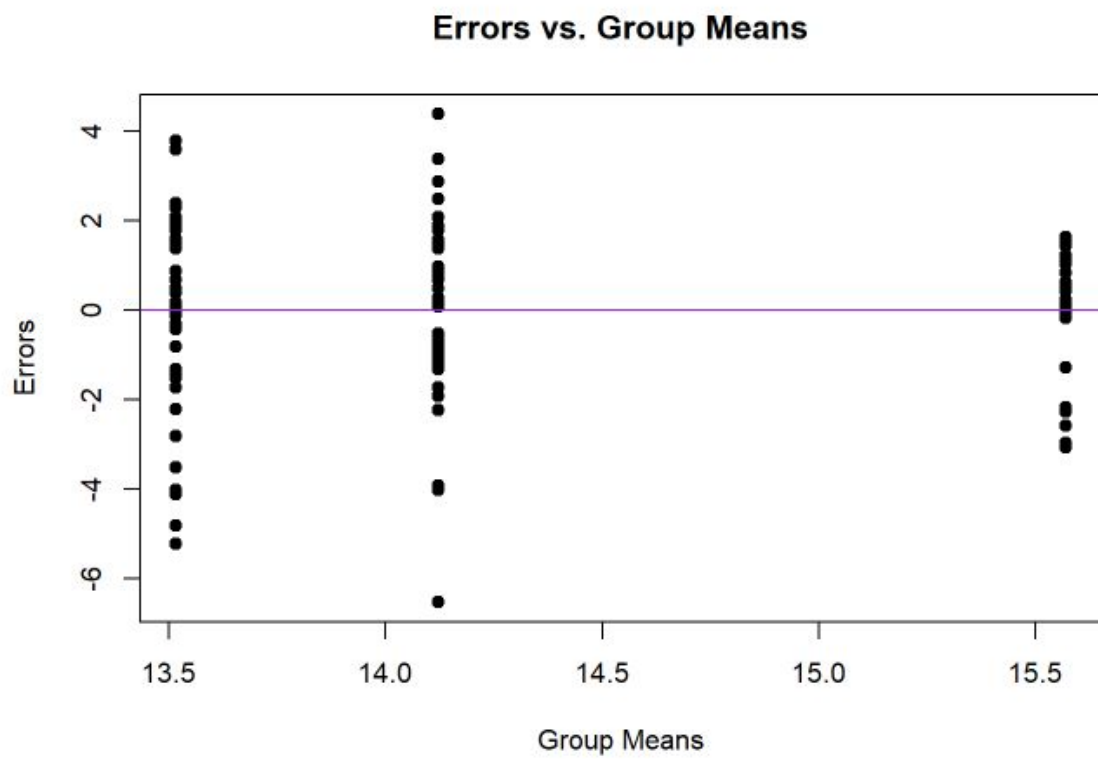
## 2. Summary of Data

From the R code, we can know that the median for the whole group is and the mean for the whole group. and the median for different conditions are also different. The mean for the control group is 13.92444, and standard deviation for this control group is 2.419631, and median for the control is 14.3. The mean for enlarged group is 13.51556, and standard deviation for enlarged group is 2.103996, and median for this is 13.7. For the last reduced group, the mean for is 15.56923, the standard deviation for this is 1.459252, and median for this is 15.9.

From the boxplot, we can see for each group there are some outliers, and the median for the control group is similar to the enlarged group. For the reduced group, the total area is less than the other two groups. And we also can see in the histogram that the reduced group is skewed to the left, the main data are located in the right. The control group and enlarged group also have some trend that mainly increase and locate at right, which is in a larger weight area.

### 3. Diagnostics





For those plots, we can know that some groups do have some outliers because the plots represent whether the data are normally distributed or not. Because for the qq plot, there is not a close line to the fit line. There are some trends. So we need to remove some outliers for some groups.

#### 4. Analysis

Our null hypothesis is that the mean for those three groups are the same. alternative hypothesis is that at least of them, their mean is different from others. The F value for this case is 8.1288, and the p value for this case is 0.0005031. In this way, if we have 99% confidence that alpha is equal to 1%. The p value is still less than the alpha. Hence, if we have 95% confidence or 90% confidence, the p value is still less than the alpha. Therefore, we will reject the null hypothesis. As a result, there is at least one group of group's that is different from others.

```
## Analysis of Variance Table
##
## Response: Weight
##      Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment    2   72.74    36.372    8.1288 0.0005031 ***
## Residuals  113  505.62     4.474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The power for this test is 0.9547259.

```
##POWER
```

```
## [1] 0.9547259
```

For those in additional questions, if we have 95% confidence level, we will get the control group from 13,29971 to 14.54917. And in the same confidence level, we have 12.89083 to 14.14028 for the enlarged group. For the reduced group, we have an interval from 14.74735 to 16.39111. And we also compare the difference for the control group with the enlarged group, the difference is from -0.4746072 to 1.292385. For difference with control group to reduced group, the difference is from -2.67149 to -0.612424.

## confidence interval

##	Estimate	Lower Bound	Upper Bound
##	13.92444	13.29972	14.54917

##	Estimate	Lower Bound	Upper Bound
##	13.51556	12.89083	14.14028

##	Estimate	Lower Bound	Upper Bound
##	15.56923	14.74735	16.39111

##	Estimate	Lower Bound	Upper Bound
##	0.4088889	-0.4746072	1.2923850

##	Estimate	Lower Bound	Upper Bound
##	-1.644786	-2.677149	-0.612424

## 5. Interpretation :

Because we calculated the p value in the last question, and the p value is really small. In this way, even though we have 99% confidence, we still will reject the null hypothesis. In other words, there is at least one group mean for those groups that are different from others.

confidence interval?

For those additional questions, first, we will have 95% confidence that we will observe the true mean or more extreme value for the control group from 13,29971 to 14.54917. And we will have 95% confidence that we will observe the true mean or more extreme value for the enlarged group from 12.89083 to 14.14028. we also will have 95% confidence that we will observe the true mean or more extreme value for the reduced group from 14.74735 to 16.39111.

## 6. Conclusion :

According to the R code, those tables, plots, we reject the null hypothesis that there are at least one group mean that is different from others. Besides, we use diagnostics to determine that there are some outliers in those data. In this way, if we use transformation to those data, we can make this result to be more precise. And median for control and enlarged are similar to each other, but for reduced, this is mainly located at a large weight area.

And the confidence interval for the nest that tends to have the largest sparrow is the reduced group. A confidence interval comparing the control nest to the enlarged nest is from -0.4746072 to 1.292385. Besides, A confidence interval comparing the control nest to the reduced nest is from -2.677149 to -0.612424.

## Appendix

```
## Plot
sparrow <- read.csv("E:/UCDAVIS/2020 Winter Quarter/STA 106/sparrow.csv")
boxplot(sparrow$Weight ~ sparrow$Treatment, main = "The weight of the bird by type of nes", ylab = "The weight of
the bird")
library(ggplot2)
ggplot(sparrow, aes(x = Weight)) + geom_histogram(binwidth = 2, color = "black", fill = "white") + facet_grid(Trea
tment ~.) + ggtitle("The weight of the bird by type of nes")

## MODEL
the.model = lm(Weight ~ Treatment, data = sparrow)
anova.table = anova(the.model)
anova.table
MSE = anova.table[2,3]

group.means = aggregate(Weight ~ Treatment, data = sparrow, mean)
group.sd = aggregate(Weight ~ Treatment, data = sparrow, "sd")
group.median = aggregate(Weight ~ Treatment, data = sparrow, "median")
group.nis = by(sparrow$Weight, sparrow$Treatment, length)
group.means
group.sd
group.median
MSE = anova.table[2,3]

## POWER
give.me.power = function(ybar, ni, MSE, alpha){
  a = length(ybar) # Finds a
  nt = sum(ni) # Finds the overall sample size
  overall.mean = sum(ni*ybar)/nt # Finds the overall mean
  phi = (1/sqrt(MSE))*sqrt( sum(ni*(ybar - overall.mean)^2)/a) # Finds the books value of phi
  phi.star = a *phi^2 # Finds the value of phi we will use for R
  Fc = qf(1-alpha, a-1, nt-a) # The critical value of F, use in R's function
  power = 1 - pf(Fc, a-1, nt-a, phi.star) # The power, calculated using a non-central F
  return(power)
}
group.means1 = by(sparrow$Weight, sparrow$Treatment, mean)
group.nis1 = by(sparrow$Weight, sparrow$Treatment, length)
the.power = give.me.power(group.means1, group.nis1, MSE, 0.05)
the.power
```

```

## Diagnostics
the.model = lm(Weight ~ Treatment, data = sparrow)
sparrow$ei = the.model$residuals
nt = nrow(sparrow)
a = length(unique(sparrow$Treatment))
SSE = sum(sparrow$ei^2)
MSE = SSE/(nt-a)
eij.star = the.model$residuals/sqrt(MSE)
alpha = 0.05
t.cutoff = qt(1-alpha/(2*nt), nt-a)
CO.eij = which(abs(eij.star) > t.cutoff)
rij = rstandard(the.model)
CO.rij = which(abs(rij) > t.cutoff)

CO1 = which(sparrow$Weight < 6)
outliers = CO1
new.data = sparrow[-outliers,]
new.model = lm(Weight ~ Treatment, data = new.data)
qqnorm(new.model$residuals)
qqline(new.model$residuals)

plot(new.model$fitted.values, new.model$residuals, main = "Errors vs. Group Means", xlab = "Group Means", ylab = "Errors", pch = 19)
abline(h = 0, col = "purple")

library(ggplot2)
ggplot(Treatment, ei, data = new.data) + ggtitle("Errors vs. Groups") + xlab("Groups") + ylab("Errors") + geom_hline(yintercept = 0, col = "purple")

```



```

## confidence interval
give.me.CI = function(ybar,ni,ci,MSE,multiplier){
  if(sum(ci) != 0 & sum(ci !=0 ) != 1){
    return("Error - you did not input a valid contrast")
  } else if(length(ci) != length(ni)){
    return("Error - not enough contrasts given")
  }
  else{
    estimate = sum(ybar*ci)
    SE = sqrt(MSE*sum(ci^2/ni))
    CI = estimate + c(-1,1)*multiplier*SE
    result = c(estimate,CI)
    names(result) = c("Estimate","Lower Bound","Upper Bound")
    return(result)
  }
}

group.means = by(sparrow$Weight,sparrow$Treatment,mean)
group.sd = aggregate(Weight ~ Treatment, data = sparrow,"sd")
group.nis = by(sparrow$Weight,sparrow$Treatment,length)
the.model = lm(Weight ~ Treatment, data = sparrow)
anova.table = anova(the.model)
MSE = anova.table[2,3]

t.value = qt(1-0.05/2, sum(group.nis) - length(group.nis))
##control group
ci.1 = c(1,0,0)
CI1 = give.me.CI(group.means,group.nis,ci.1,MSE,t.value)
CI1
##enlarged group
ci.2 = c(0,1,0)
CI2 = give.me.CI(group.means,group.nis,ci.2,MSE,t.value)
CI2
##reduced group
ci.3 = c(0,0,1)
CI3 = give.me.CI(group.means,group.nis,ci.3,MSE,t.value)
CI3
##control and enlarged
ci.4 = c(1,-1,0)
CI4 = give.me.CI(group.means,group.nis,ci.4,MSE,t.value)
CI4
##control and reduced
ci.5 = c(1,0,-1)
CI5 = give.me.CI(group.means,group.nis,ci.5,MSE,t.value)
CI5

```