<div align="center">Data Analysis Assignment #2</div>

Introduction

As was stated in the previous data analysis assignment that data was made available from a previous study of abalones. The intent of the previous study was to predict the age and to a lesser extent sex of the abalone based on various physical measurements as the current method is time and labor-intensive requiring drilling into shells to count rings. We pick up on the Abalone data project this week focusing primarily on devising and evaluating binary decision rules for harvesting abalones. The assignment focuses mainly on the Ratio and Volume categories within the dataset, as another way to safeguard the abalone is through creating a cutoff based on several factors. Any above or below this cutoff will not be harvested, thus protecting the animal from over-harvesting. Later findings in this report state that Sex and Class play integral roles in the harvesting, or lack thereof, of the abalone.

Results

Figure 1 and Figure 2 are a histogram and a Q-Q plot using the Ratio. In addition, the skewness and kurtosis are calculated. The histogram is skewed to the right with the value for the skewness is given by 0.715 and kurtosis 1.67. For interpretation, a Shapiro-Wilks Test was calculated to check for normality. If the P-Value is >0.05 then the distribution is normal, though the test on Ratio in this dataset came back "Not Normal."

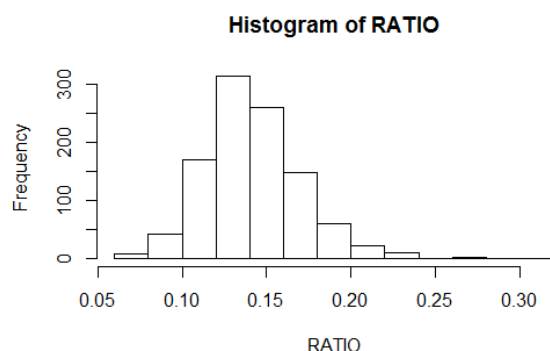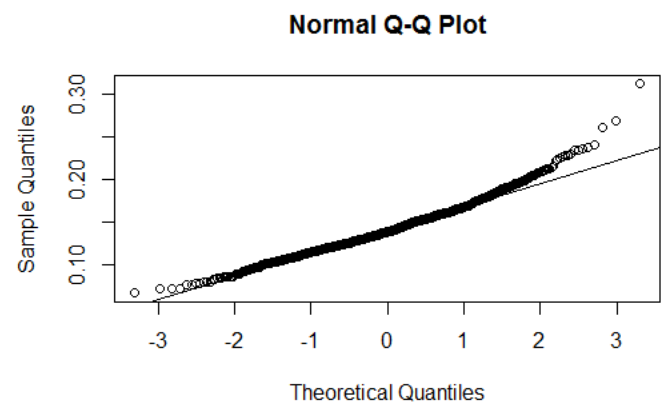Figure 1   Figure 2:


```
> skewness(RATIO, na.rm = TRUE, unbiased = TRUE)
[1] 0.7147056
> kurtosis(RATIO, na.rm = TRUE, unbiased = TRUE)
[1] 1.667298
```

Ratio was transformed using log10() to create L_RATIO and the above calculations were repeated forming a histogram and Q-Q plot. This has normalized the skewness as is evident by the Skewness of -.09 and a kurtosis of .54, though the distribution is still slightly skewed left. Figure 5 brings Class into the equation, noting several outliers in Class A3 against the Log Ratio
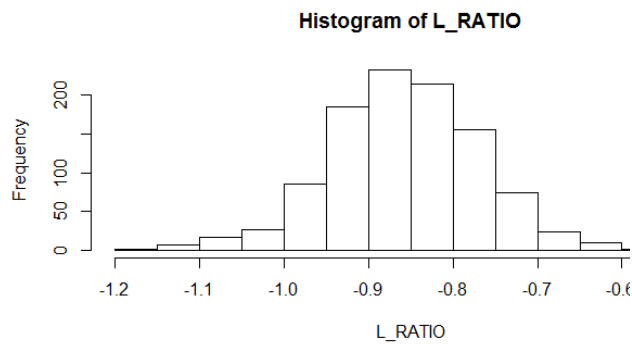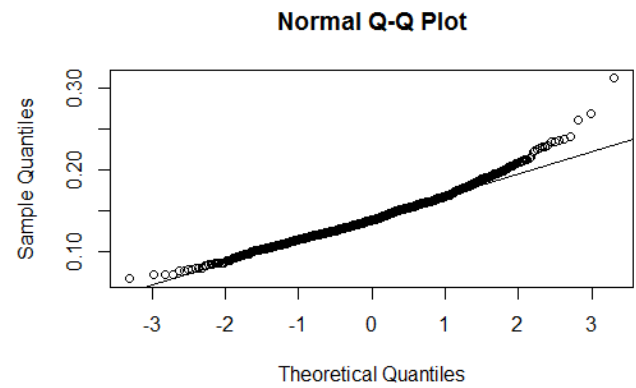
(L_Ratio).

Figure 3:   Figure 4:
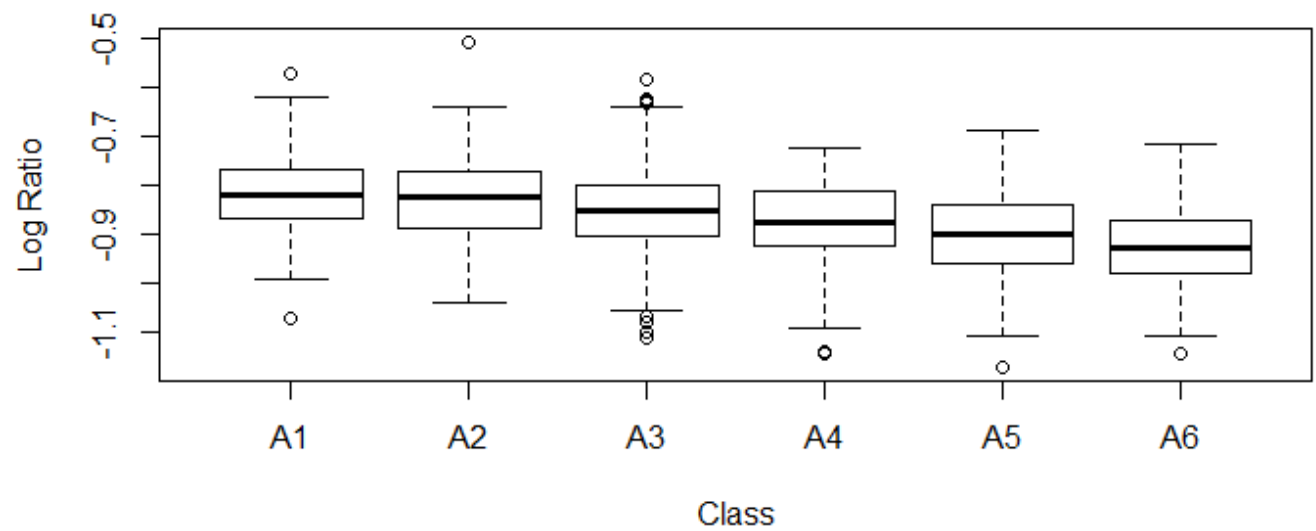
Histogram of L_RATIO

Normal Q-Q Plot

```
> skewness(L_RATIO, na.rm = TRUE, unbiased = TRUE)
[1] -0.09391549
> kurtosis(L_RATIO, na.rm = TRUE, unbiased = TRUE)
[1] 0.535431
```

Figure 5:

We reject the null hypothesis that the L_Ratio dataset have homogenous variances across class because A1 and A2 are much higher than A5 and A6. Similarly, the p-value=0.6884 is higher than the set significant level of alpha=0.05. Therefore, we cannot assume from the Barlett's test that L_Ratio is normally distributed by class. The Bartlett test is included below:

```
        Bartlett test of homogeneity of variances

data:  L_RATIO by CLASS
Bartlett's K-squared = 3.0749, df = 5, p-value = 0.6884
```

The first analysis, Figure 6 below, shows that Sex has a p-value of .000995 with Class having an even smaller p-value conveying a very significant relationship. Though, the Class:Sex interaction is an extraneous variable/irrelevant variable that can be remove from the model. The second model output very similarly with the p-values, but has removed the Class:Sex component.

Figure 6:

```
               Df Sum Sq Mean Sq F value    Pr(>F)
CLASS           5  1.076 0.21512  31.313  < 2e-16 ***
SEX             2  0.096 0.04782   6.960 0.000995 ***
CLASS:SEX      10  0.029 0.00290   0.421 0.936789
Residuals    1018  6.994 0.00687
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 7:

```
               Df Sum Sq Mean Sq F value    Pr(>F)
CLASS           5  1.076 0.21512  31.490  < 2e-16 ***
SEX             2  0.096 0.04782   6.999 0.000957 ***
Residuals    1028  7.023 0.00683
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Though, the Turkey results in figure 8 show that male and female abalones cannot be combined within the model as the output for sex M-F is not significant with a P-value of 0.942. While male (P=0.092) and female(P=0.03) are significant on their own.

Figure 8:

```
        Tukey multiple comparisons of means
          95% family-wise confidence level

Fit: aov(formula = L_RATIO ~ CLASS + SEX, data = mydata)

$CLASS
             diff          lwr          upr       p adj
A2-A1 -0.01248831 -0.03990346  0.014926837 0.7848170
A3-A1 -0.03451323 -0.06067382 -0.008352646 0.0024066
A4-A1 -0.05863763 -0.08713038 -0.030144884 0.0000001
A5-A1 -0.08685165 -0.12129814 -0.052405154 0.0000000
A6-A1 -0.11174297 -0.14532240 -0.078163549 0.0000000
A3-A2 -0.02202492 -0.04214244 -0.001907396 0.0224189
A4-A2 -0.04614932 -0.06921824 -0.023080398 0.0000002
A5-A2 -0.07436334 -0.10447811 -0.044248565 0.0000000
A6-A2 -0.09925466 -0.12837366 -0.070135660 0.0000000
A4-A3 -0.02412440 -0.04568735 -0.002561445 0.0180550
A5-A3 -0.05233842 -0.08131574 -0.023361091 0.0000045
A6-A3 -0.07722974 -0.10517079 -0.049288694 0.0000000
A5-A4 -0.02821402 -0.05931298  0.002884949 0.1005227
A6-A4 -0.05310534 -0.08324107 -0.022969608 0.0000085
A6-A5 -0.02489132 -0.06070873  0.010926085 0.3520976

$SEX
            diff          lwr          upr       p adj
I-F -0.016277335 -0.031437534 -0.001117136 0.0318479
M-F  0.002062021 -0.012574216  0.016698257 0.9415134
M-I  0.018339356  0.003739124  0.032939587 0.0091596
```

Figure 9, the histogram for infant volume showed that the data was not normally distributed. It is asymmetrical and skewed to the right. While the adult volume seems to be normally distributed and symmetrical and slightly skewed to the right based on its histogram. The implications of separating the adult and infant volume is that we would have small negative outcome in volume because it is positively skewed.

Figure 9:
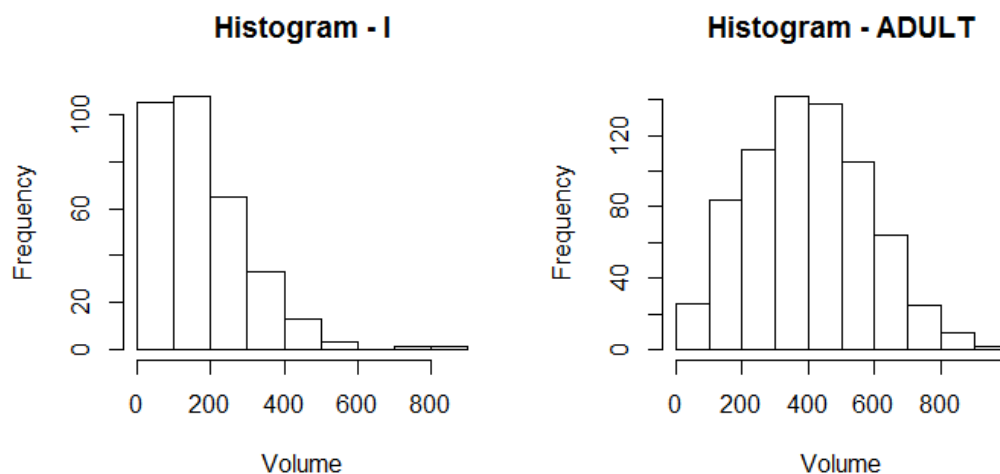
The scatter plot for SHUCK versus VOLUME, on the right in Figure 10, revealed wide variation from the mean especially for class A6 and A5. While adult compared to infants have more variability from the center line. In the log transformed dataset of Figure 10, the scatter plot revealed reduced variability within the dataset. The log transformed data helps to collapse the variation towards the mean value. The L_SHUCK versus L_VOLUME, on the left in Figure 11, demonstrates that most of the values for classes moving aggregating towards the center. Likewise, adults and infant values also moving towards the center line.
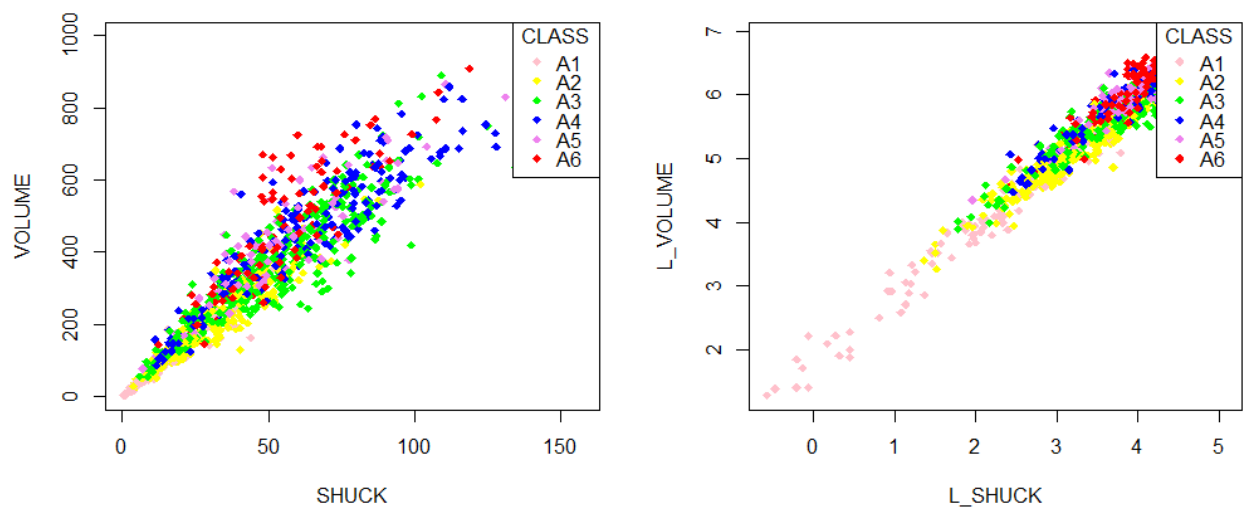
Figure 10:

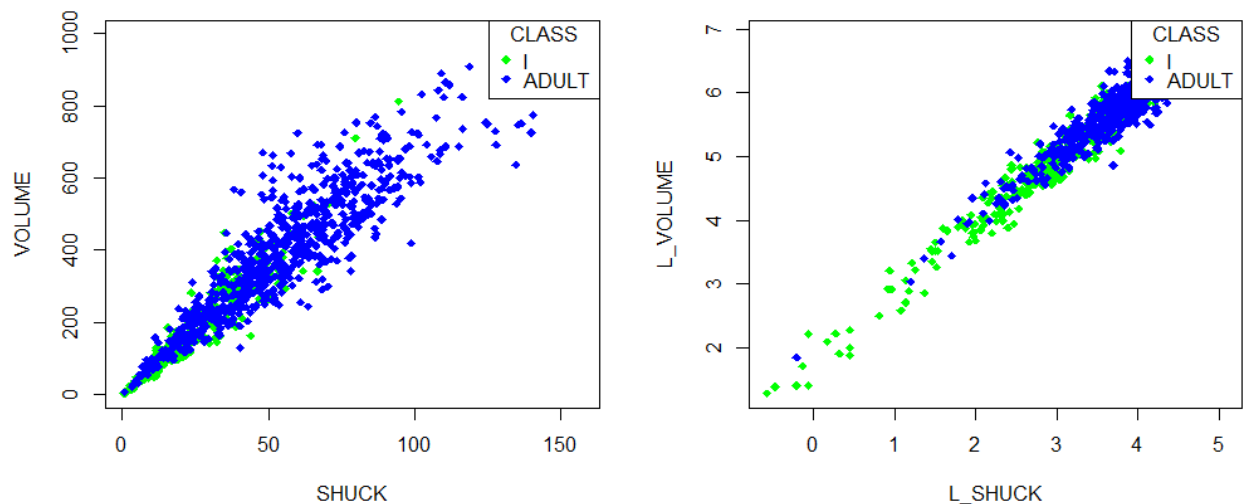Figure 11:

Data Analysis Assignment #2

Figure 12 is a regress of L_SHUCK as the dependent variable on L_VOLUME, CLASS and TYPE using a regression model.

Figure 12:

```
Call:
lm(formula = L_SHUCK ~ L_VOLUME + CLASS + TYPE, data = mydata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.63285 -0.12483 -0.00377  0.12889  0.70686

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.87058    0.04399 -42.528  < 2e-16 ***
L_VOLUME     0.99593    0.01031  96.554  < 2e-16 ***
CLASSA2     -0.03997    0.02519  -1.587 0.112927
CLASSA3     -0.10924    0.02824  -3.868 0.000117 ***
CLASSA4     -0.16894    0.03129  -5.399 8.30e-08 ***
CLASSA5     -0.23367    0.03458  -6.757 2.36e-11 ***
CLASSA6     -0.29244    0.03468  -8.433  < 2e-16 ***
TYPEADULT    0.05798    0.01570   3.693 0.000233 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1903 on 1028 degrees of freedom
Multiple R-squared:  0.9508,    Adjusted R-squared:  0.9505
F-statistic:  2841 on 7 and 1028 DF,  p-value: < 2.2e-16
```

Figure 13 shows that since abalone growth slows after class A3, infants in classes A4 and A5 are considered candidates for harvest. We have reclassified the infants in classes A4 and A5 as ADULTS and repeat the regression using the model from Figure 12.

Figure 13:

```
Call:
lm(formula = L_SHUCK ~ L_VOLUME + CLASS + TYPE, data = mydata)

Residuals:
     Min      1Q   Median      3Q      Max
-0.62328 -0.12552 -0.00056  0.12850  0.71269

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.88198    0.04383 -42.937  < 2e-16 ***
L_VOLUME     0.99917    0.01026  97.381  < 2e-16 ***
CLASSA2     -0.04151    0.02529  -1.641 0.101135
CLASSA3     -0.10963    0.02860  -3.834 0.000134 ***
CLASSA4     -0.17485    0.03218  -5.434 6.87e-08 ***
CLASSA5     -0.23987    0.03544  -6.768 2.19e-11 ***
CLASSA6     -0.29276    0.03508  -8.344 2.28e-16 ***
TYPEADULT    0.04932    0.01716   2.874 0.004131 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1908 on 1028 degrees of freedom
Multiple R-squared:  0.9506,    Adjusted R-squared:  0.9503
F-statistic:  2825 on 7 and 1028 DF,  p-value: < 2.2e-16
```

The class levels are inversely related to L_SHUCK, for example in CLASSA3, a decrease of 0.11 units would lead to 1-unit increase in L_SHUCK. The values for the CLASS revealed an increasing negative or inverse trend to CLASS A6. When compared to earlier display, tend to demonstrate similar trends.

Type is an important predictor of L-SHUCK and has a direct relationship with the outcome variable when compared to CLASS categorical variable that has a negative relationship with the outcome. For every 0.04-unit increase in the value of TYPE, there is a corresponding 1-unit increase in the L_SHUCK variable. The variable is also statistically significant at a P-value of 0.004.

The results for the histogram and Q-Q plot for the residuals in Figure 14 reveal a normally distributed dataset and is symmetrical around the center. The skewness is slightly negative and close to zero. The values in QQ plot are well fitted and linearly distributed towards the center.

Figure 14:

Histogram of Model2$residuals

Normal Q-Q Plot
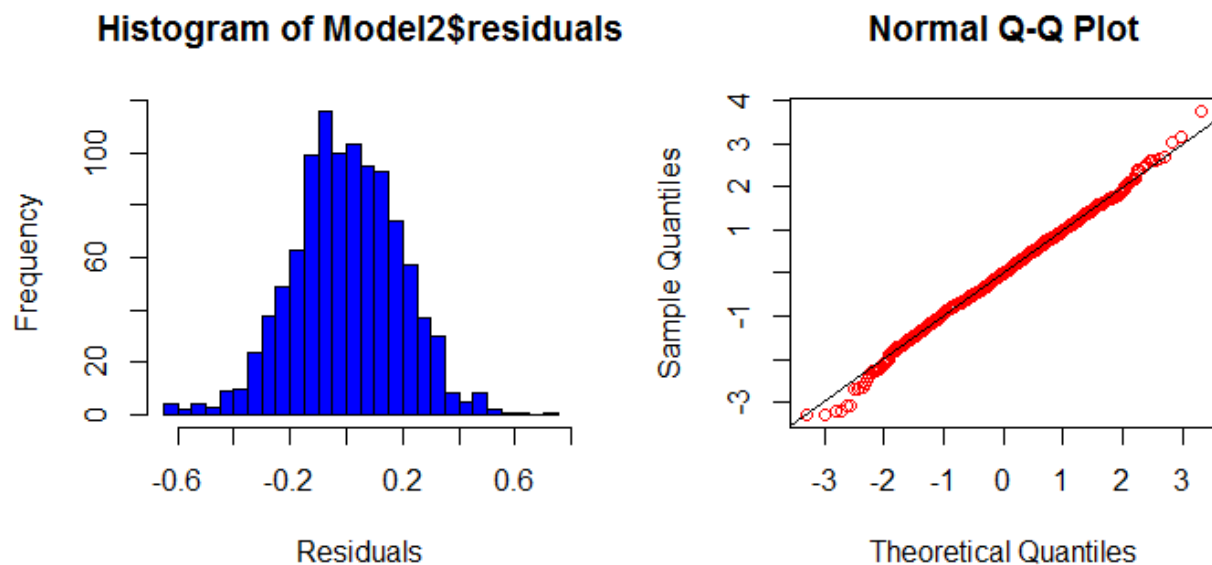
```
> skewness(x, na.rm = TRUE, unbiased = TRUE)
[1] -0.06164053
> kurtosis(x, na.rm = TRUE, unbiased = TRUE)
[1] 0.3514734
```

The regression model from the residual plot show some level of heteroscedasticity for both class and type variables. With a Skewness of -.06 and kurtosis of .35 show much more of a linear relationship. In addition, moving into Figure 15, we can see that the lower classes of abalone evenly correspond with the Infant scatter plot of Figure 16. It is clear that the lower Class of Abalone have a relationship with Infant as opposed to Adult. This has been previously shown in several other graphs.
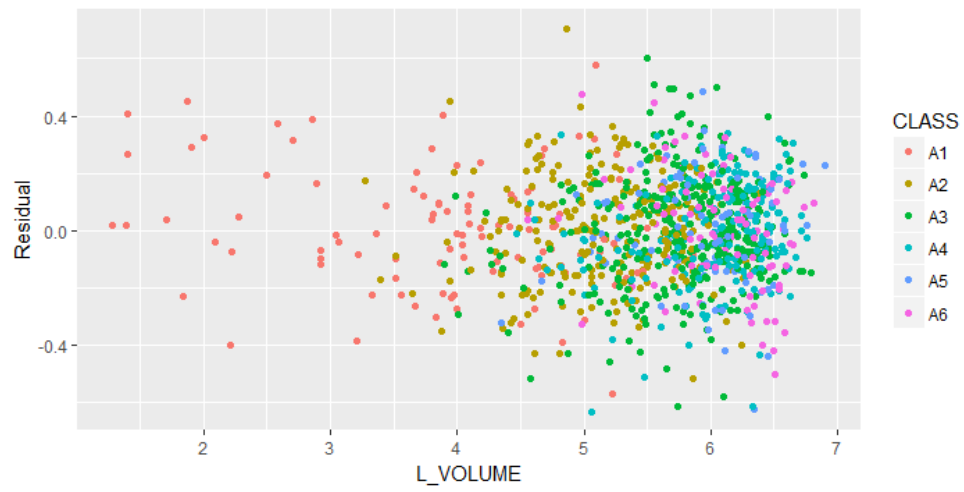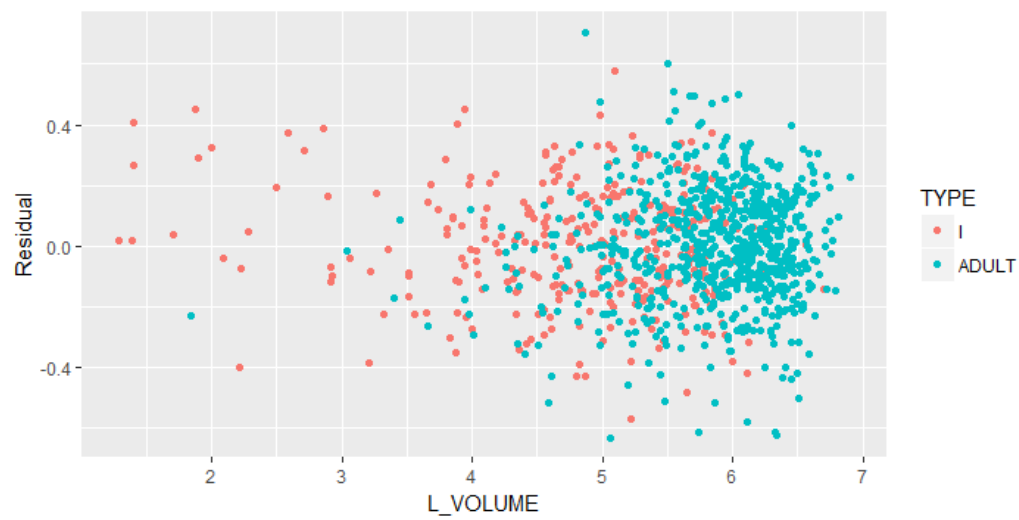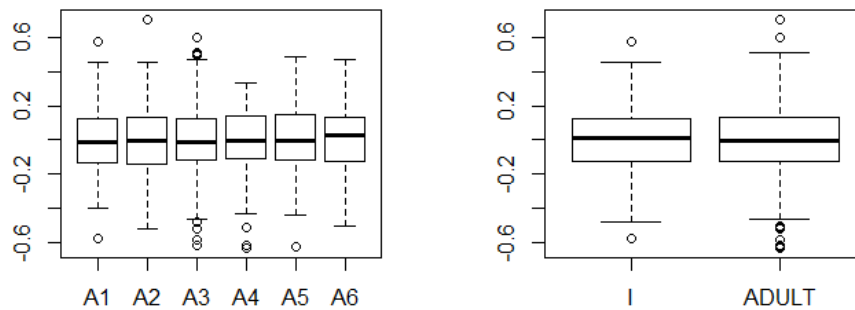
Figure 15:

Figure 16:

Figure 17:

There is a tradeoff faced in managing the abalone harvest. The infant population must be protected since that represents future harvests. On the other hand, the harvest should be designed to be efficient with a yield to justify the effort. VOLUME is used for binary decision rules. If VOLUME is below a "cutoff" (i.e. specified volume), that individual will not. Below is the calculated proportion of infant abalones and adult abalones which fall beneath a specified volume or "cutoff". A series of volumes covering the range from minimum to maximum abalone volume will be used in a "for loop" to determine how the harvest proportions change as the "cutoff" changes.

```
> head(volume.value)
[1] 4.603851 5.595913 6.587974 7.580036 8.572097 9.564159
> head(prop.adults)
[1] 0.000000000 0.000000000 0.001355014 0.001355014 0.001355014 0.001355014
> head(prop.infants)
[1] 0.01342282 0.01677852 0.02013423 0.02684564 0.03020134 0.03691275
```

Figure 18 is a plot showing the infant proportions and the adult proportions versus volume with a 50% "split" volume.value for each shown on the plot. This visually shows the difference between the two populations. The two split points suggest an interval within which potential cutpoints may be located.

Data Analysis Assignment #2

Figure 18:

Proportion of Adults and Infants Protected
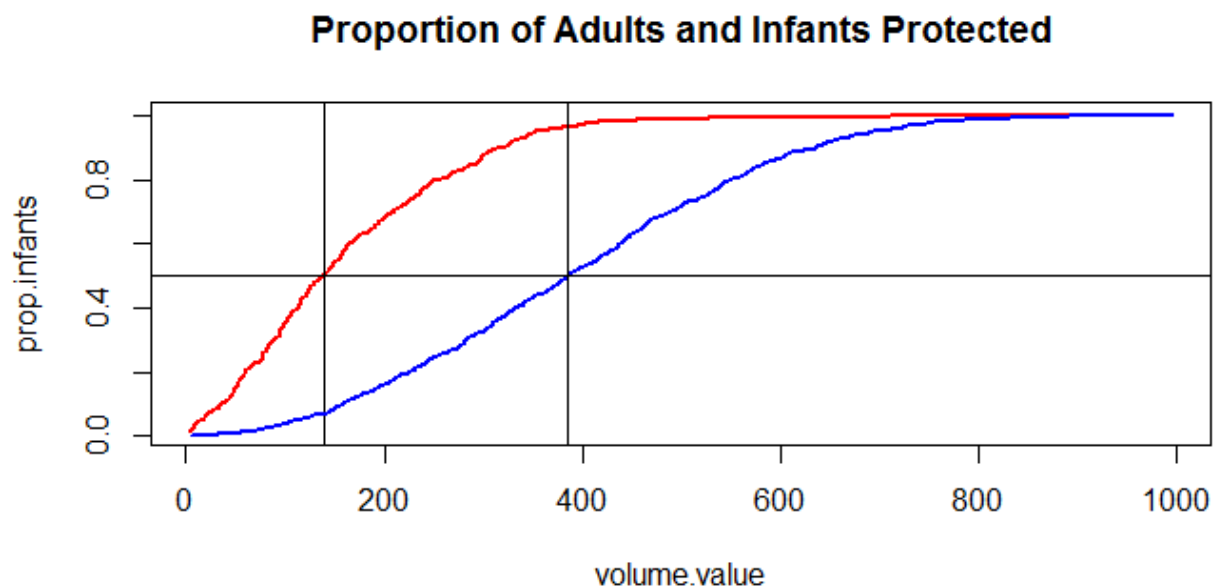
This part will address the determination of a volume.value corresponding to the observed maximum difference in harvest percentages of adults and infants. These proportions are converted from "not harvested" proportions to "harvested" proportions by using for infants, and for adults. The reason the proportion for infants drops sooner than adults, is that infants are maturing and becoming adults with larger volumes.

There is considerable variability present in the peak area of this plot. The observed "peak" difference may not be the best representation of the data. One solution is to smooth the data to determine a more representative estimate of the maximum difference.

```
> head(difference)
[1] 0.01342282 0.01677852 0.01877921 0.02549062 0.02884633 0.03555774
```

Curve smoothing calculated information is provided below:

```
> volume.value[volume.value > max(mydata[mydata$CLASS == "A1" &
+                                   mydata$TYPE == "I", "VOLUME"])][1]
[1] 206.9844
```

Figure 19 presents a plot of the difference versus volume.value with the variable smooth.difference superimposed. This shows peak location corresponding to the cutoff determined.

Figure 19:

Difference in Harvest Proportion
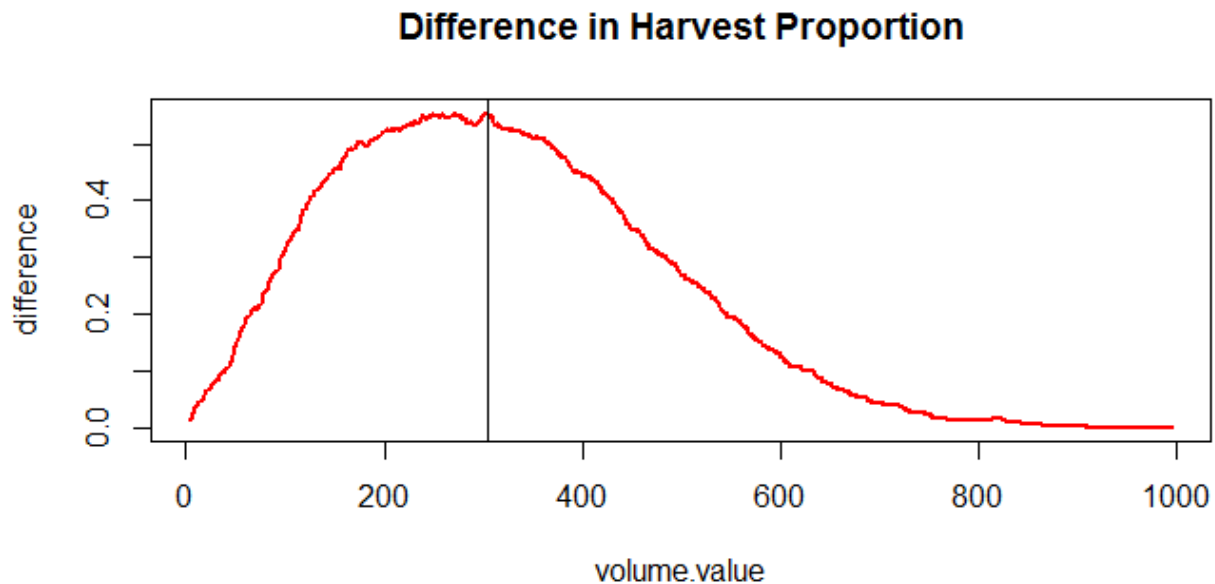
From Figure 19, the largest difference between adults and infants harvested occur between the volume values of approximately 200 and 400. This range of values is similar to the range we calculated for the 50% split cutoff.

```
> volume.value[volume.value > max(mydata[mydata$CLASS == "A1" &
+                                    mydata$TYPE == "I", "VOLUME"])][1]
[1] 206.9844
> sum(mydata[mydata$TYPE == "I", "VOLUME"] > 206.9844) /
+    sum(mydata$TYPE == "I")
[1] 0.3020134
> volume.value[which.min(abs(prop.adults - (1-prop.infants)))]
[1] 241.7065
> sum(mydata[mydata$TYPE == "I", "VOLUME"] > 241.7065) /
+    sum(mydata$TYPE == "I")
[1] 0.2248322
```

Figure 20 is a construct of an ROC curve by plotting Adult harvest populations versus Infant harvest populations. Each point which appears corresponds to a particular volume.value. The locations of the cutoffs determined in (7) and (8) are on this plot. Using auc() for the chart we were able to calculate the area under the curve of .848.

```
> auc(a,b)
[1] 0.8480611
```

Figure 20:
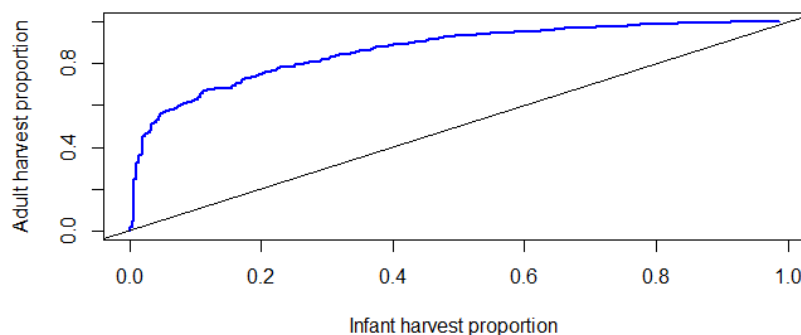
Areas-under-curve, or AUCs, greater than 0.8 are taken to indicate good discrimination potential. I agree with the general rule that the concordance index of 0.8 validates how good our predictive model was. Receiver Operating Characteristic (ROC) graphics are useful for organizing binary classifiers and visualizing their performance. Is it possible to make a final selection at this time? What additional information would be helpful?

```
> head(TruePostiveRate)
[1] 1.000000 1.000000 0.998645 0.998645 0.998645 0.998645
> head(TotalHarvestPopulation)
[1] 0.004443872 0.005401460 0.006359049 0.007316637 0.008274225
[6] 0.009231814
> head(df)
  TruePostiveRate FalsePositiveRate TotalHarvestPopulation
1        1.000000         0.9865772            0.004443872
2        1.000000         0.9832215            0.005401460
3        0.998645         0.9798658            0.006359049
4        0.998645         0.9731544            0.007316637
5        0.998645         0.9697987            0.008274225
6        0.998645         0.9630872            0.009231814
```

Conclusions (Address each part)

Assuming we are making a presentation to the investigators for the purposes of determining a cutoff to implement based on the analysis, I would initially present Figure 9, the histogram for infant volume showing that the data was is normally distributed, thus leaving a large portion of the Infant population with lower shuck volumes. I would then follow up with Figure 17 and 18, showing outliers of Infants within the later Classes show that there is a lower cutoff of Infant abalone. I would finalize the presentation with the constructed ROC graph to determine cut off between sensitivity and specificity for the dataset.

There were several qualifications and considerations to present regarding the analysis. The Class of the abalone would need to meet Type, as in "Adult" or "Infant" in addition to meeting Volume qualifications within each. Those cutoff points would be determined based on the above calculations.

There would be several specific recommendations along with outlined choices. Specifically using the cut off value volumes for the Adult and Infant abalone, I would recommend Class and Volume parameters for harvesting.

All of the cutoff information listed above, the main question to the investigators would be whether or not applying a cutoff would be feasible. As was mentioned in the first data assignment, gendering an abalone and finding its age was very labor intensive and still resulted in harvesting abalones. In addition, it was clearly stated that we would not be able to reasonably class the abalone based on shuck volume or other characteristics. So, with that stated, is a cutoff feasible?

Taking the above and moving into suggestions along with moving forward, I would suggest continuing research and requesting additional data. This initial investigation was to re-expose a past study into the abalone. These investigations or consideration of alternative approaches because there is no optimal cut off. The more data, the better our prediction model and cut -off will become.

Code:
```
#1a
install.packages("rockchalk")
library(rockchalk)
mydata <- read.csv(file.path("c:/Rabalone/", "mydata.csv"), sep = ",")
RATIO <- mydata$RATIO
hist(RATIO)
qqnorm(RATIO)
qqline(RATIO)
skewness(RATIO, na.rm = TRUE, unbiased = TRUE)
kurtosis(RATIO, na.rm = TRUE, unbiased = TRUE)
# Checking for normal distribution of RATIO
normality <- shapiro.test(RATIO)
# Shapiro-Wilks Test is used to check for normality. If P-Value is >0.05 then the
distribution is normal
ifelse(normality$p.value > 0.05,"Normal","Not Normal")
> Not Normal
#1b
L_RATIO=log10(RATIO)
hist(L_RATIO)
qqnorm(L_RATIO)
qqline(L_RATIO)
skewness(L_RATIO, na.rm = TRUE, unbiased = TRUE)
kurtosis(L_RATIO, na.rm = TRUE, unbiased = TRUE)
par(mfrow = c(1,1)) #changed from 6 to 1 so that only one chart is displayed
plot(L_RATIO~mydata$CLASS ,ylab = "Log Ratio",xlab="Class") #added x axis and y-axis
label
#1c
bartlett.test(L_RATIO ~ CLASS, data=mydata)
p-value is 0.6884
Thus, we cannot reject the null hypothesis that the variance is same across all CLASS
#2a
anova1 <- aov(L_RATIO ~ CLASS*SEX, data = mydata)
summary(anova1)
anova2 <- aov(L_RATIO ~ CLASS+SEX, data = mydata)
summary(anova2)
#2b
TukeyHSD(anova2)
# The results suggest that the difference between Male and Female is not significant.
#Hence they can be combined
#3a
mydata$TYPE <- combineLevels(mydata$SEX, levs = c("M","F"), "ADULT")
TYPE <- mydata$TYPE
par(mfrow=c(1,2))
```

Data Analysis Assignment #2

```
hist(mydata$VOLUME[mydata$TYPE == "I"],xlab = "Volume",main="Histogram - I")
hist(mydata$VOLUME[mydata$TYPE == "ADULT"],xlab = "Volume",main="Histogram -
ADULT")
# ADULT RESULTED IN A LESS SKEWNESS OF THE DATA
#3b
install.packages('ggplot2')
library(ggplot2)
mycolors = c('pink','yellow','green','blue','violet','red')
with(mydata, plot(SHUCK,VOLUME, col = mycolors[CLASS],pch = 18))
with(mydata, legend('topright',legend=levels(CLASS),col=mycolors, pch = 18,
title='CLASS'))
mydata$L_SHUCK <- log(mydata$SHUCK)
mydata$L_VOLUME <- log(mydata$VOLUME)
with(mydata, plot(L_SHUCK,L_VOLUME, col = mycolors[CLASS],pch = 18))
with(mydata, legend('topright',legend=levels(CLASS),col=mycolors, pch = 18,
title='CLASS'))
# Based on the classes using log each of them is segregated. Class A1 appears in (0,3) to
(0,3) range. Other classes as one goes towards top right.
mycolors = c('green','blue')
with(mydata, plot(SHUCK,VOLUME, col = mycolors[TYPE],pch = 18))
with(mydata, legend('topright',legend=levels(TYPE),col=mycolors, pch = 18,
title='CLASS'))
mydata$L_SHUCK <- log(mydata$SHUCK)
mydata$L_VOLUME <- log(mydata$VOLUME)
with(mydata, plot(L_SHUCK,L_VOLUME, col = mycolors[TYPE],pch = 18))
with(mydata, legend('topright',legend=levels(TYPE),col=mycolors, pch = 18,
title='CLASS'))
#4a
Model1 <- lm(L_SHUCK~L_VOLUME+CLASS+TYPE, data = mydata)
summary(Model1)
#4b
index <- (mydata$CLASS == "A5")|(mydata$CLASS == "A4")
mydata$TYPE[index] <- combineLevels(mydata$TYPE[index],
 levs = c("I", "ADULT"), "ADULT")
Model2 <- lm(L_SHUCK~L_VOLUME+CLASS+TYPE, data = mydata)
summary(Model2)
# The coefficients of classA4 and classA5 have decreased in value
#5a
Model2.stdres = rstandard(Model2)
x = Model2$residuals
h <- hist(Model2$residuals, breaks=20, col="blue",xlab="Residuals")
qqnorm(Model2.stdres, ylab="Sample Quantiles", xlab="Theoretical Quantiles", col =
"red")
qqline(Model2.stdres,ylab="Sample Quantiles", xlab="Theoretical Quantiles", col =
```

```
"black")
library(rockchalk)
skewness(x, na.rm = TRUE, unbiased = TRUE)
kurtosis(x, na.rm = TRUE, unbiased = TRUE)
#5b
#Plot the residuals versus L_VOLUME coloring the data points by CLASS
ggplot(Model1, aes(x = L_VOLUME,y = Model1$residuals)) + geom_point(aes(color =
CLASS)) + labs(x = "L_VOLUME", y = "Residual")
#second time coloring the data points by SEX
ggplot(Model1, aes(x = L_VOLUME,y = Model1$residuals)) + geom_point(aes(color =
TYPE)) + labs(x = "L_VOLUME", y = "Residual")
#boxplot
install.packages('RColorBrewer')
library(RColorBrewer)
boxplot(split(Model1$residuals, mydata$CLASS))
boxplot(split(Model1$residuals, mydata$TYPE))
bartlett.test(c ~ CLASS, data = mydata)
#6a
idxi <- mydata$TYPE=="I"
idxa <- mydata$TYPE=="ADULT"
max.v <- max(mydata$VOLUME)
min.v <- min(mydata$VOLUME)
delta <- (max.v - min.v)/1000
prop.infants <- numeric(0)
prop.adults <- numeric(0)
volume.value <- numeric(0)
total.infants <- length(mydata$TYPE[idxi])
total.adults <- length(mydata$TYPE[idxa])
for (k in 1:1000) {
  value <- min.v + k*delta
  volume.value[k] <- value
  prop.infants[k] <- sum(mydata$VOLUME[idxi] <= value)/total.infants
  prop.adults[k] <- sum(mydata$VOLUME[idxa] <= value)/total.adults
}
n.infants <- sum(prop.infants <= 0.5)
split.infants <- min.v + (n.infants + 0.5)*delta  # This estimates the desired volume.
n.adults <- sum(prop.adults <= 0.5) split.adults <- min.v + (n.adults + 0.5)*delta
#6b
par(mfrow = c(1,1))
plot(volume.value, prop.infants, col = "red", main = "Proportion of Infants Not
Harvested",
    type = "l", lwd = 2)
abline(h=0.5)
abline(v = split.infants)
```

Data Analysis Assignment #2

```
lines(volume.value, prop.adults, col = "blue", main = "Proportion of adults Not
Harvested",
     type = "l", lwd = 2)
abline(h=0.5)
abline(v = split.adults)
#7a
difference <- (1-prop.adults) - (1-prop.infants)
plot(difference,volume.value)
# ?plot(), ?abline(), ?text() to review documentation pages
#7b
# loess, local polynomial regression fitting
y.loess.a <- loess(1-prop.adults ~ volume.value, span = 0.25, family = c("symmetric"))
y.loess.i <- loess(1-prop.infants ~ volume.value, span = 0.25, family = c("symmetric"))
smooth.difference <- predict(y.loess.a) - predict(y.loess.i)
#7c
plot(volume.value, difference, col = "red", main = "Difference in Harvest Proportion",
     type = "l", lwd = 2)
abline(v = volume.value[which(difference == max (difference))])
#7d
(1-prop.infants)[which.max(smooth.difference)]
#8a
volume.value[volume.value > max(mydata[mydata$CLASS == "A1" &
 mydata$TYPE == "I", "VOLUME"])][1]
# Now, to determine the proportions harvested, we can look to the proportions
# of infants and adults with VOLUMEs greater than this threshold.
# For example, for infants:
sum(mydata[mydata$TYPE == "I", "VOLUME"] > 211.9447) /
 sum(mydata$TYPE == "I")
#8b
volume.value[which.min(abs(prop.adults - (1-prop.infants)))]
sum(mydata[mydata$TYPE == "I", "VOLUME"] > 211.9447) /
 sum(mydata$TYPE == "I")
#9
plot(a,b, col = "blue", xlab = "Infant harvest proportion", ylab = "Adult harvest
proportion",
type = "l", lwd = 2)
abline(0,1)
install.packages("flux")
library(flux)
auc(a,b)
#10
TruePostiveRate <- 1-prop.adults
 FalsePositiveRate <- 1-prop.infants
 TotalHarvestPopulation <- volume.value/ (total.adults + total.infants)
```

Data Analysis Assignment #2

```
df <- data.frame(TruePostiveRate,FalsePositiveRate,TotalHarvestPopulation)
```