

- PREDICT 401 -

Assignment 2:

Further exploration of gaps in the Tasmanian abalone study

Ahmed Mustakim

Table of Contents:

Introduction.....	3
Results.....	3
Conclusions.....	10
References.....	11
Appendix (Code).....	12

Introduction:

In part 1, we exposed the flaw in the abalone age classification system. We showed that as abalones got heavier and larger, their age classification systems (A1 to A6, i.e. youngest to oldest based on ring count) did not consistently get higher. This is a continuation from part 1 and the purpose remains to further ascertain gaps in the Tasmanian abalone study, this time by using exploratory data analysis, analysis of variance and linear regression methods. In part 2, we begin by reusing the simple random sample of 500 observations from part 1 and create a matrix plot of 7 numeric variables (length, diameter, height, whole weight, shuck weight and viscera weight) and 1 categorical variable (sex). We do this to understand the basic correlations among these 7 numeric variables using the Pearson and Spearman Correlation Coefficients, depending on visible linear or non-linear patterns in the plots. Although the classification of infant as a separate gender is questionable, we then look at the dynamics of all three genders (female, infant and male) and six classes (A1 to A6) in shuck weight, volume and shuck weight to volume ratio individually to ascertain the impact of any unknown environmental factors on abalone growth and aging. We focus on the density variable (whole weight/volume) and measure the differences in mean densities among the different classes through one-way ANOVA. Our results then lead us to perform a chi-square test of independence to determine if volume is independent of shuck weight and subsequently develop a regression model between those correlated variables.

Results:

For now, we ignore the categorical variable (sex), which we will look at separately, and instead focus on the visible linear and non-linear correlations between the remaining seven variables: length (mm), diameter (mm) and height (mm), whole weight (gm), shuck weight (gm) and viscera weight (gm).

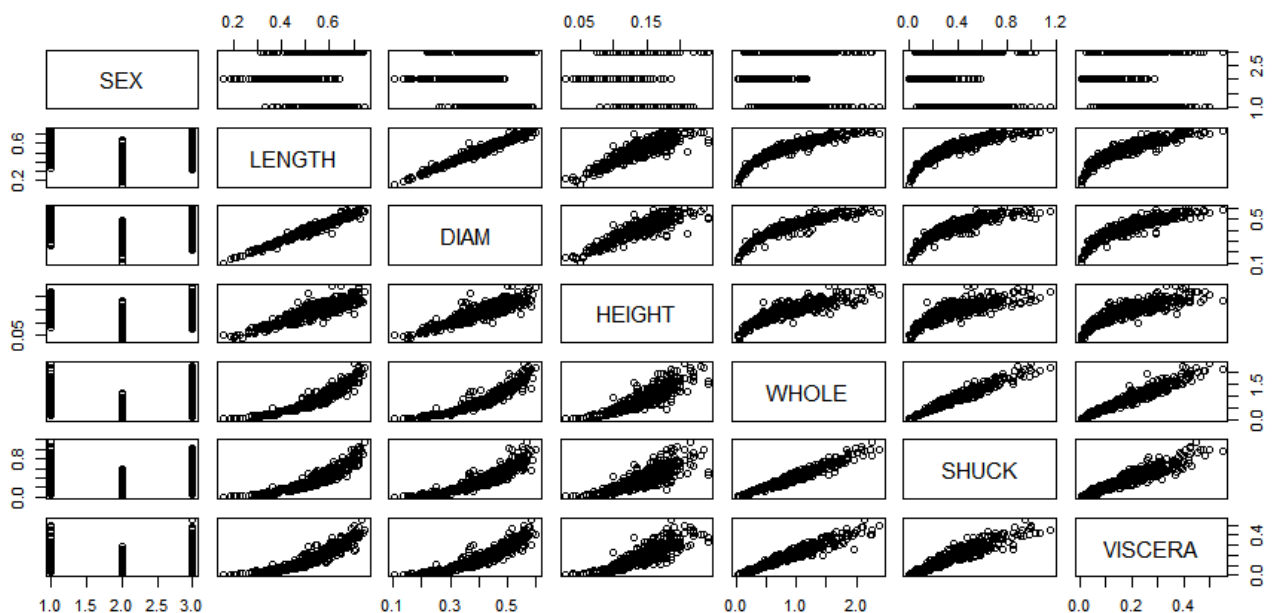


Figure 1: Matrix plot of sex, length, diameter, height, whole weight, shuck weight and viscera weight of abalones from a simple random sample of 500 observations, showing differential growth rates in some of the curved plots.

Overall, we see high degrees of correlations within these seven variables using either the Pearson or the Spearman correlation coefficient. The lowest degree of correlation is between shuck weight and height using either coefficient (0.84 using Pearson and 0.87 using Spearman). Among length, diameter and height we see very high linear correlations using the Pearson coefficient. Incidentally all 3 variables were combined into a single variable called volume (length x diameter x height) in part 1. We see high degrees of non-linear correlations using the Spearman coefficient between length, diameter, height and whole weight. All 4 variables were combined into the density variable (whole weight/ volume) in part 1. We also notice equal degrees of very high correlations (0.98) between shuck weight and whole weight using either the Pearson or Spearman coefficients, so we will investigate the ratio (shuck weight/ whole weight) between these two variables next.

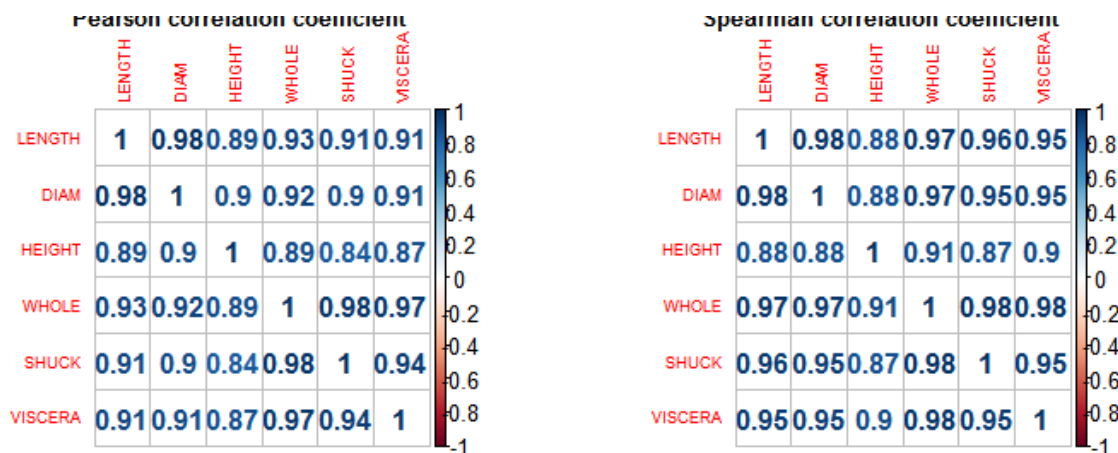


Figure 2: Pearson (left) and Spearman (right) correlation coefficients of seven abalone paired variables (length, diameter, height, whole weight, shuck weight and viscera weight)

There are outliers in the boxplots for the shuck weight to whole weight ratio boxplots in all three gender categories (indicating variability in data) and the infant ratio is the highest among all three genders.

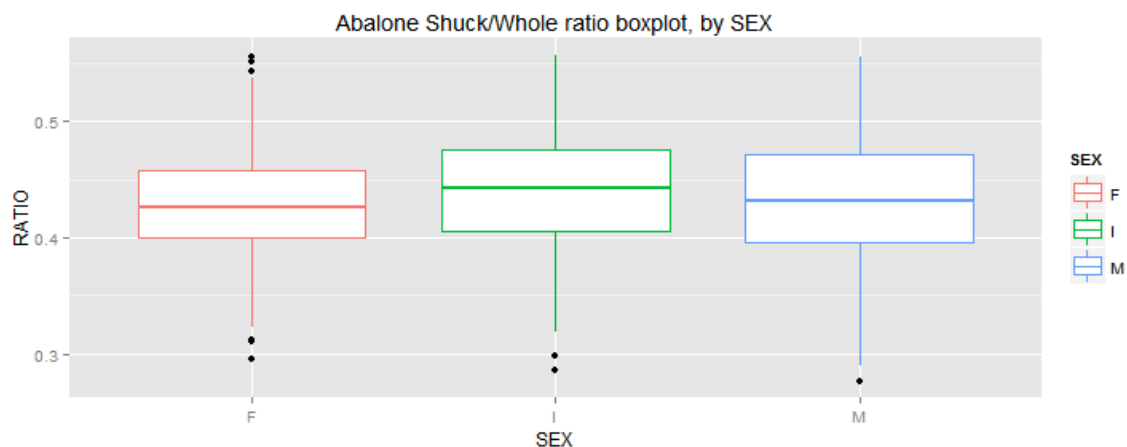


Figure 3: Box plots of abalone shuck/whole weight ratios differentiated by sex

The same general trend can be seen in the boxplots for classes – outliers in every class (variability in the data) and if we temporarily ignore the dip in A1 (to be revisited in page 7-8) then it seems younger abalones have higher mean ratios than older abalones. It appears from this boxplot that as abalones grow older, their densities tend to decrease. In part 1, we determined that infants tend to slightly dominate in the upper density quantiles (between 50% and 90%) indicating that infants are smaller and have more meat tightly packed as they are smaller in volume than adults.

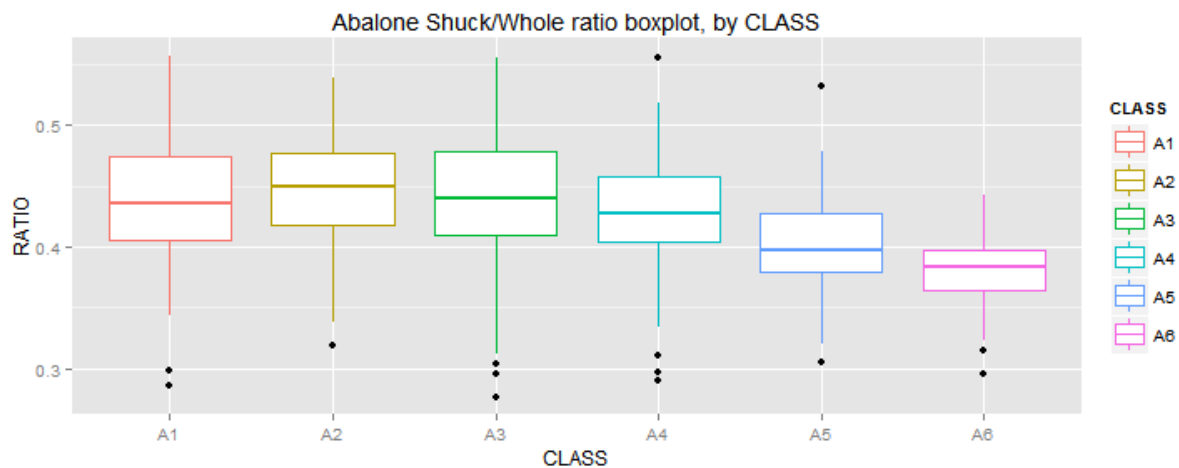


Figure 4: Box plots of abalone shuck/whole weight ratios differentiated by class

We now separately examine trends within shuck weight boxplots and density boxplots differentiated by sex and class in figure 5, 6, 7 and 8 below. There are outliers (variability) in the shuck weight boxplots and as expected, mean infant shuck weight is the lowest in figure 5 below.

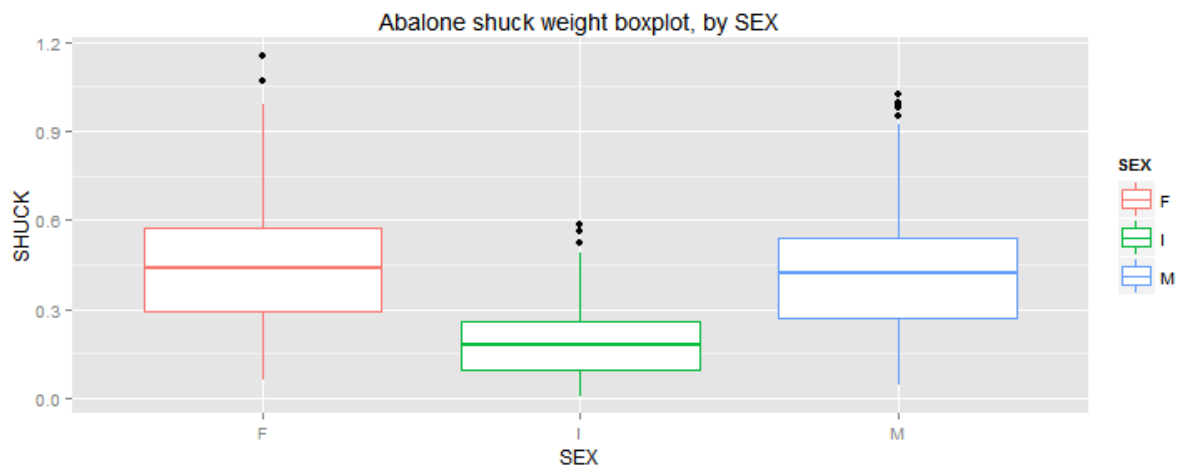


Figure 5: Box plots of abalone shuck weight (gm) differentiated by sex

However figure 6 is where we see a disruption of this trend. As the abalones get older, they do not consistently put on more shuck weight. We have already established the flaw in the ring count based age classification system. It is possible that “each class is a separate population due to environmental factors present during the life of the population”¹. According to a study at the University of Melbourne, various environmental factors such as water flow, dissolved oxygen, temperature, pH level, nitrogenous waste, food quantity and quality and stocking density can influence abalone growth.²

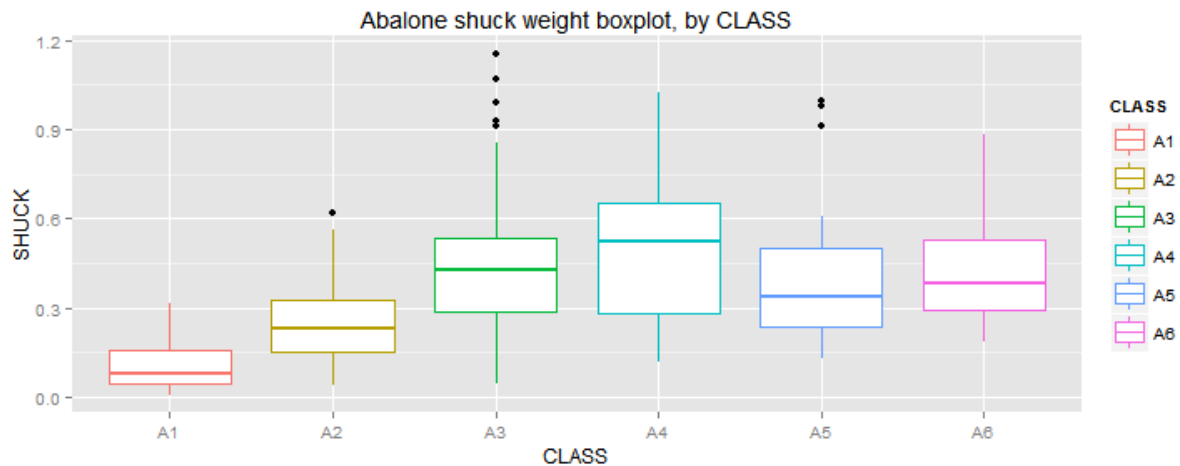


Figure 6: Box plots of abalone shuck weight (gm) ratios differentiated by class

In figure 7 below the mean of the densities for all three genders seem approximately the same and not surprisingly, there are plenty of outliers (variability in the data).

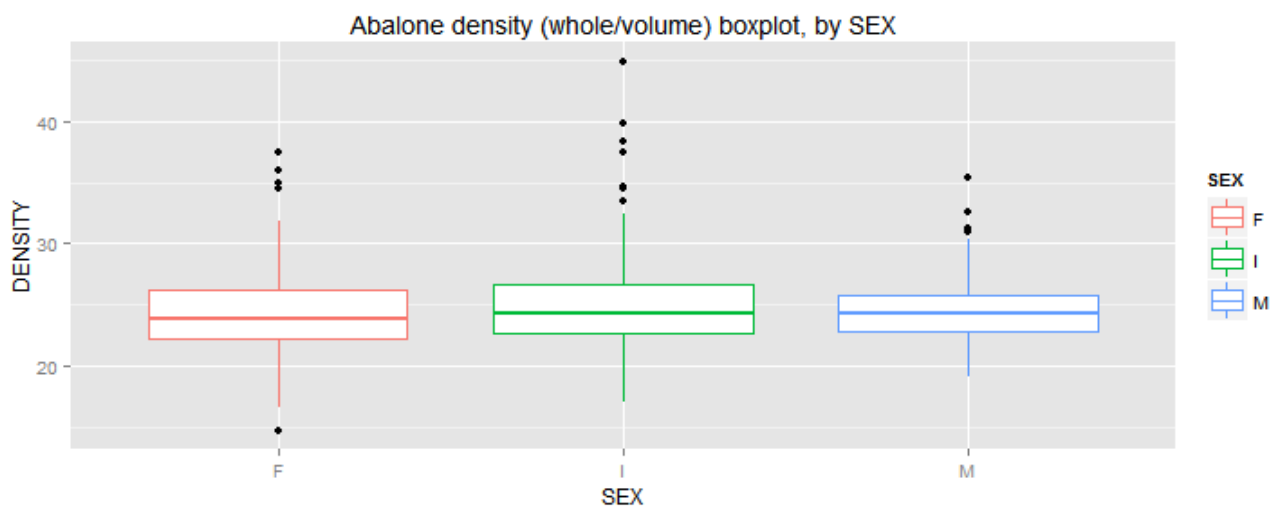


Figure 7: Box plots of abalone density (gm/mm^3) differentiated by sex

In figure 8 we also notice marginal differences between the mean densities of the different classes. We already know the classification system is flawed and there are unaccounted environmental factors (confounding) which could influence abalone growth rates. We must also remind ourselves that in part 1, we observed densities are not normally distributed and any differences between mean densities between the classes are overshadowed by variability in the data. So this is to be expected.

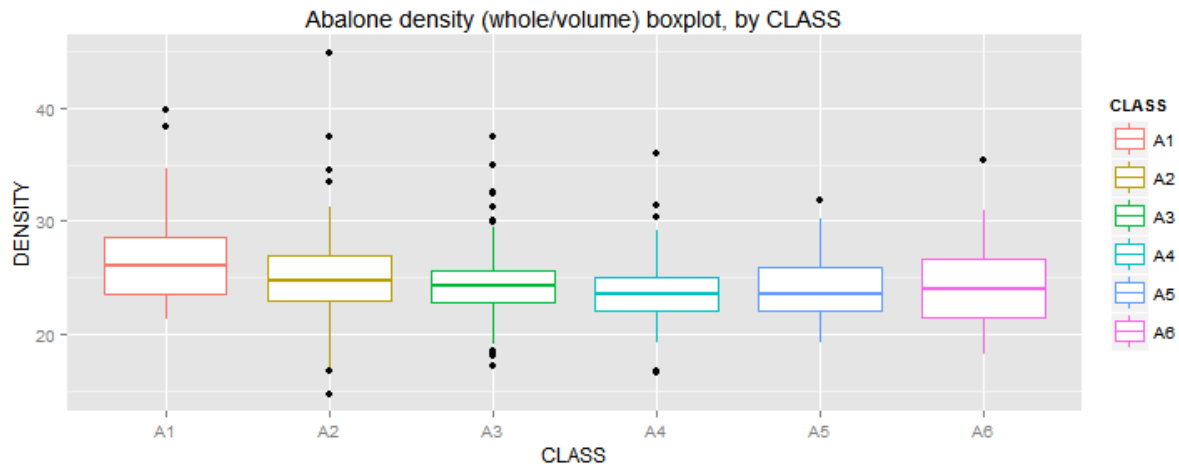


Figure 8: Box plots of abalone density (gm/mm^3) differentiated by class

However, it still begs the question: are the mean densities between the different classes (i.e. different populations) the same? A one-way analysis of variance test on density using class as the grouping variable returns a p-value of $3.973\text{e-}05$, which is significantly less than 0.05. Therefore we reject the null hypothesis that the means from the different classes are the same. In other words not all the means are the same. A Tukey's honest significant test show differences in the mean densities between the highlighted classes based on the highlighted p-values which are less than 0.05 (significance level):

```
Tukey multiple comparisons of means
95% family-wise confidence level
Fit: aov(formula = my_model, data = mydata)
$CLASS
```

Difference exist in mean densities between these classes:		diff	lwr	upr	p adj	
	A2-A1	-1.71196923	-3.347367	-0.07657194	0.0340866	p-values < 0.05
	A3-A1	-2.42861754	-3.975000	-0.88223541	0.0001271	
	A4-A1	-2.90632276	-4.602711	-1.20993464	0.0000191	
	A5-A1	-2.51479496	-4.633583	-0.39600689	0.0095901	
	A6-A1	-2.33626002	-4.455048	-0.21747195	0.0210115	
	A3-A2	-0.71664831	-1.867635	0.43433805	0.4787967	
	A4-A2	-1.19435353	-2.540164	0.15145646	0.1149493	
	A5-A2	-0.80282573	-2.652889	1.04723773	0.8161748	
	A6-A2	-0.62429079	-2.474354	1.22577266	0.9286259	
	A4-A3	-0.47770521	-1.713819	0.75840882	0.8790089	
	A5-A3	-0.08617742	-1.858043	1.68568832	0.9999927	
	A6-A3	0.09235752	-1.679508	1.86422326	0.9999897	
	A5-A4	0.39152780	-1.512663	2.29571873	0.9917961	
	A6-A4	0.57006273	-1.334128	2.47425367	0.9564431	
	A6-A5	0.17853494	-2.110018	2.46708762	0.9999240	

Remember the dip in A1 in figure 4 (box plots of abalone shuck/whole weight ratios differentiated by class) in page 5? Perhaps, in light of the difference of the density means between A1 and all the other classes, this now make sense. It certainly looks like A1 is from a different population group, not age group.

Density seems to be a key variable. In fact, the afore-mentioned study from University of Melbourne claims “Grow-out density is a key factor to consider in abalone mariculture... In aquaculture, density may affect growth directly through competition for food or space, which is similar to the natural environment, or indirectly through the accumulation of excretory products.”²

Let’s deconstruct the density variable, which was calculated by dividing whole weight by volume, and look at its component parts, i.e. shuck weight (highly correlated to whole weight, as seen in the correlation plots in figure 2 in page 4) and volume. We look at shuck weight instead of whole weight because as we saw in figure 6 in page 6, shuck weight data has more variability and inexplicability and thus should be explored more. So what is the relationship between volume and shuck weight? Are those variables independent of each other?

Let’s perform a chi-square test of independence where:

H₀: Volume is independent of shuck weight

H₁: Volume and shuck weight are dependent

Our contingency table is:

	Volume ABOVE median (0.03257888 mm ³)	Volume BELOW median (0.03257888 mm ³)
Shuck weight ABOVE median (0.3285 gm)	225	24
Shuck weight BELOW median (0.3285 gm)	25	226

We see that p-value is less than the 0.05 significance level in our chi-square test result:

```
Number of cases in table: 500
Number of factors: 2
Test for independence of all factors:
  Chisq = 323.2, df = 1, p-value = 2.891e-72
```

Therefore we reject the null hypothesis and it appears volume is dependent on the shuck weight.

This dependency is further evidenced by the linear correlations in the scatterplots between those 2 variables in figure 9 below. At log scale the linear correlation is even more apparent as the points are more tightly packed.

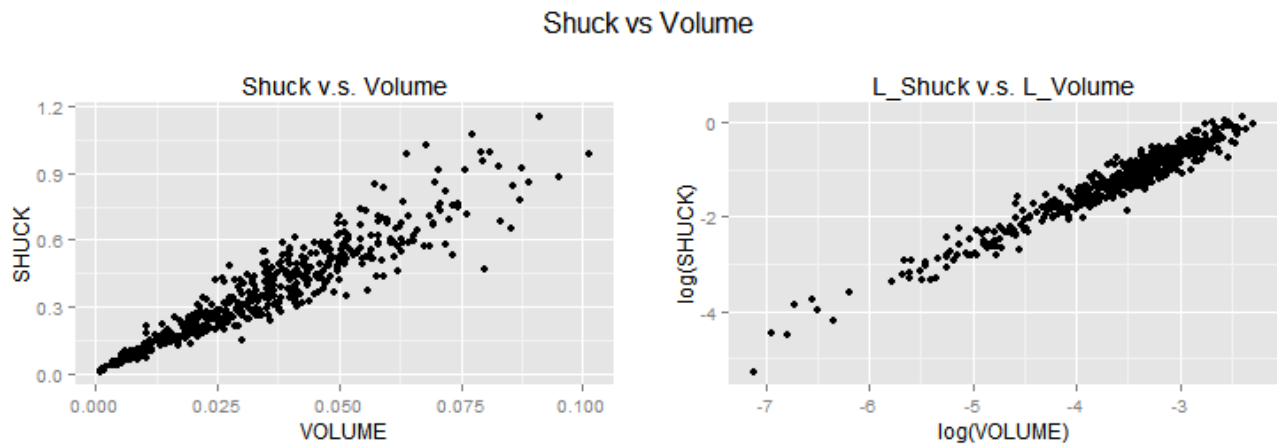


Figure 9: Scatterplots of abalone shuck weight (gm) v.s. volume (mm³) and in log scale.

So let's develop a simple linear regression model for these correlated and dependent variables at log scale, using a 95% confidence limit:

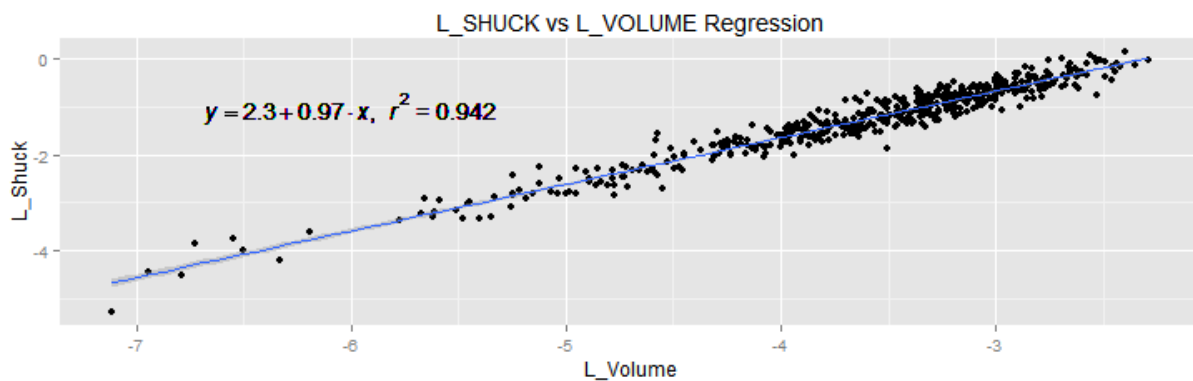


Figure 10: Regression plot of abalone shuck weight v.s. volume in log scale.

According to our simple linear regression model, for every mm³ increment in volume we can expect a 0.97 gm increment in shuck weight. For practical purposes we ignore the intercept, because it represents 2.3 gm of shuck weight when there is zero volume, which is not possible.

Conclusions:

From our two-part observational study, we have learned that the ring count based age classification system of Tasmanian abalones was flawed because it did not correlate with the abalone growth rates. Abalones did not get heavier or larger as they aged, based on this classification system. It seems the researchers failed to control for confounding as they did not account for various lurking environmental variables, such as water flow, dissolved oxygen, etc., which could have plausibly explained the differential abalone growth rates. It may have also explained the variability in the data.

Another issue was classifying infants into a different gender. This was a serious experiment-design problem that may have made age-prediction all the more difficult for the researchers. In fact without careful experiment design, data collection methods, sampling methods, controlling for confounding etc., it's difficult to establish a predictive model with a high degree of certainty. There are many environmental factors that affect abalone growth rates, which the researchers did not account for. It's possible that each of the 6 classes which is a grouping of ring count is from a separate population group of abalones rather than from different age bracket. We tackled this question in our analysis.

In fact, given the existing data set, we asked a different set of questions than the original researchers. Rather than trying to establish an age classification model or trying to predict age, we analyzed the density variable, which is a key factor in abalone growth rates. We asked: do the mean densities differ between the existing classes (which we assumed are different populations of abalones)? If so, is shuck weight (gm) dependent on volume (mm³)? In other words can abalone sizes predict abalone shuck weights?

The answers to our questions were quite interesting. We used a 95% confidence level for all tests. It turned out the mean densities did differ between A1 and the remaining 5 classes (populations). This showed that A1 is plausibly from a different population of abalones, not from a different age bracket. We also found that shuck weight is dependent on volume and based on a simple linear regression model we can now predict shuck weight (in gm) given a non-zero volume (in mm³) with a 95% confidence level:

Shuck weight in gm = 2.3 + (0.97 x Volume in gm)

References

¹courses.northwestern.edu – *Data Analysis Project Assignment#2.pdf*. Retrieved March 8, 2015 from https://courses.northwestern.edu/bbcswebdav/pid-5534688-dt-content-rid-16766761_1/xid-16766761_1

²researchgate.net – *Growth of juvenile blacklip abalone (Haliotis rubra) in aquaculture tanks: effects of density and ammonia* (Sylvain M.H. Huchettea, C.S. Koh , Rob W. Daya), Department of Zoology, The University of Melbourne. Retrieved March 8, 2015 from https://www.researchgate.net/profile/Robert_Day2/publication/249967970_Growth_of_juvenile_blacklip_abalone_%28Haliotis_rubra%29_in_aquaculture_tanks_effects_of_density_and_ammonia/links/545b10fc0cf25c508c31a659.pdf?ev=pub_int_doc_dl&origin=publication_detail&inViewer=true

stackoverflow.com – *Adding Regression Line Equation and R2 on graph*. Retrieved March 8, 2015 from <http://stackoverflow.com/questions/7549694/ggplot2-adding-regression-line-equation-and-r2-on-graph>

gist.github.com – *GitHub Gist (Gist is a simple way to share snippets and pastes with others. All gists are Git repositories, so they are automatically versioned, forkable and usable from Git.)* Retrieved March 8, 2015 from <https://gist.github.com/>

Appendix (Code):

```
#Assumption: You have ggplot2, plyr, grid, gridExtra, corrplot
require(ggplot2)
require(plyr)
require(grid)
require(gridExtra)
require(corrplot)

# STEP 1
#####

#Go to directory where abalone.csv and mydata.csv file is saved
setwd("C:/Users/AhmedToshiba/Documents/Education/Northwestern/PREDICT 401/Assignment2")

#Read abalone and sample files into data frame
abalone <- read.csv("abalone.csv", sep = " ")
mydata <- read.csv("mydata.csv")

#Make sure data frames were correctle imported

str(mydata)
# 'data.frame':   500 obs. of  11 variables:
#   $ X      : int  1196 3294 1711 3687 3926 190 2206 3725 2303 1906 ...
#   $ SEX     : Factor w/ 3 levels "F","I","M": 1 3 3 1 1 1 2 1 3 2 ...
#   $ LENGTH : num  0.705 0.575 0.645 0.625 0.47 0.585 0.275 0.485 0.46 0.575 ...
#   $ DIAM   : num  0.55 0.455 0.51 0.485 0.35 0.46 0.22 0.365 0.375 0.45 ...
#   $ HEIGHT : num  0.17 0.185 0.195 0.135 0.115 0.17 0.08 0.15 0.13 0.135 ...
#   $ WHOLE  : num  1.219 1.156 1.226 1.302 0.487 ...
#   $ SHUCK  : num  0.639 0.552 0.589 0.61 0.196 ...
#   $ VISCERA: num  0.236 0.243 0.222 0.268 0.127 ...
#   $ SHELL  : num  0.301 0.295 0.374 0.36 0.155 ...
#   $ RINGS  : int   9 13 10 14 8 9 6 7 9 10 ...
#   $ CLASS  : Factor w/ 6 levels "A1","A2","A3",...: 3 5 3 5 2 3 1 2 3 3 ...

str(abalone)
# 'data.frame':   4141 obs. of  10 variables:
#   $ SEX     : Factor w/ 3 levels "F","I","M": 3 3 1 3 2 2 1 1 3 1 ...
#   $ LENGTH : num  0.455 0.35 0.53 0.44 0.33 0.425 0.53 0.545 0.475 0.55 ...
#   $ DIAM   : num  0.365 0.265 0.42 0.365 0.255 0.3 0.415 0.425 0.37 0.44 ...
#   $ HEIGHT : num  0.095 0.09 0.135 0.125 0.08 0.095 0.15 0.125 0.125 0.15 ...
#   $ WHOLE  : num  0.514 0.226 0.677 0.516 0.205 ...
#   $ SHUCK  : num  0.2245 0.0995 0.2565 0.2155 0.0895 ...
#   $ VISCERA: num  0.101 0.0485 0.1415 0.114 0.0395 ...
#   $ SHELL  : num  0.15 0.07 0.21 0.155 0.055 0.12 0.33 0.26 0.165 0.32 ...
#   $ RINGS  : int  15 7 9 10 7 8 20 16 9 19 ...
#   $ CLASS  : Factor w/ 6 levels "A1","A2","A3",...: 6 2 3 3 2 2 6 6 3 6 ...

# Construct a plot matrix
plot(mydata[,2:8])
```

```
#Display Pearson and Spearman correlation coefficients
library(corrplot)

#to display to graphs side by side
#and set outer margin area so that title does not get cut off
par(mfrow=c(1,2), mar=c(0,0,4,0))
corrplot( cor(mydata[,3:8]),method = "number", title="Pearson correlation coefficient",
tl.cex=0.55, cex.main=0.8) #Pearson
corrplot(cor(mydata[,3:8], method="spearman"), method = "number", title="Spearman correlation
coefficient", tl.cex=0.55, cex.main=0.8) #spearman

#turn of side by side display setting
# and change back to default
par(mfrow=c(1,1), oma=c(0,0,0,0))

# STEP 2
#####

# New variable RATIO (SHUCK / WHOLE)
mydata$RATIO <- mydata$SHUCK/mydata$WHOLE

#Side-by-side boxplots for RATIO differentiated by SEX
library(ggplot2)
ggplot(mydata, aes(SEX, RATIO, color=SEX)) + ggtitle("Abalone Shuck/Whole ratio boxplot, by
SEX") + geom_boxplot()

#Side-by-side boxplots for RATIO differentiated by CLASS
ggplot(mydata, aes(CLASS, RATIO, color=CLASS)) + ggtitle("Abalone Shuck/Whole ratio boxplot,
by CLASS") + geom_boxplot()

# STEP 3
#####

# New variable VOLUME (LENGTH x DIAM x HEIGHT)
mydata$VOLUME <- mydata$LENGTH*mydata$DIAM*mydata$HEIGHT

# New variable DENSITY (WHOLE/VOLUME)
mydata$DENSITY <- mydata$WHOLE/mydata$VOLUME

#Side-by-side boxplots for SHUCK differentiated by SEX
ggplot(mydata, aes(SEX, SHUCK, color=SEX)) + ggtitle("Abalone shuck weight boxplot, by SEX") +
geom_boxplot()

#Side-by-side boxplots for SHUCK differentiated by CLASS
ggplot(mydata, aes(CLASS, SHUCK, color=CLASS)) + ggtitle("Abalone shuck weight boxplot, by
CLASS") + geom_boxplot()
```

```
#Side-by-side boxplots for DENSITY differentiated by SEX
ggplot(mydata, aes(SEX, DENSITY, color=SEX)) + ggtitle("Abalone density (whole/volume)
boxplot, by SEX") + geom_boxplot()

#Side-by-side boxplots for SHUCK differentiated by CLASS
ggplot(mydata, aes(CLASS, DENSITY, color=CLASS)) + ggtitle("Abalone density (whole/volume)
boxplot, by CLASS") + geom_boxplot()

# STEP 4
#####

# A one-way ANOVA on DENSITY using CLASS
my_model <- {DENSITY ~ CLASS}

#Use oneway.test() for the first analysis of variance
oneway.test(my_model, mydata, var.equal = TRUE)
#One-way analysis of means

#data: DENSITY and CLASS
#F = 5.7029, num df = 5, denom df = 494, p-value = 3.973e-05

#Use aov() for another analysis of variance.
anova <- aov(my_model, mydata)

anova$coefficients
# (Intercept)      CLASSA2      CLASSA3      CLASSA4      CLASSA5      CLASSA6
# 26.745528   -1.711969   -2.428618   -2.906323   -2.514795   -2.336260

summary(anova)
#              Df    Sum Sq   Mean Sq    F value    Pr(>F)
# CLASS          5      319      63.86      5.703    3.97e-05 ***
# Residuals    494     5532      11.20
# ---
#      Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(anova)
# Tukey multiple comparisons of means
# 95% family-wise confidence level
#
# Fit: aov(formula = my_model, data = mydata)
#
# $CLASS
#              diff          lwr          upr      p adj
# A2-A1 -1.71196923 -3.347367 -0.07657194 0.0340866
# A3-A1 -2.42861754 -3.975000 -0.88223541 0.0001271
# A4-A1 -2.90632276 -4.602711 -1.20993464 0.0000191
# A5-A1 -2.51479496 -4.633583 -0.39600689 0.0095901
# A6-A1 -2.33626002 -4.455048 -0.21747195 0.0210115
# A3-A2 -0.71664831 -1.867635  0.43433805 0.4787967
# A4-A2 -1.19435353 -2.540164  0.15145646 0.1149493
# A5-A2 -0.80282573 -2.652889  1.04723773 0.8161748
```

```
# A6-A2 -0.62429079 -2.474354 1.22577266 0.9286259
# A4-A3 -0.47770521 -1.713819 0.75840882 0.8790089
# A5-A3 -0.08617742 -1.858043 1.68568832 0.9999927
# A6-A3 0.09235752 -1.679508 1.86422326 0.9999897
# A5-A4 0.39152780 -1.512663 2.29571873 0.9917961
# A6-A4 0.57006273 -1.334128 2.47425367 0.9564431
# A6-A5 0.17853494 -2.110018 2.46708762 0.9999240

# STEP 5
#####
library(plyr)

#Dichotomize SHUCK and VOLUME about their median
median(mydata$SHUCK)
#0.3285
shuck <- factor(mydata$SHUCK > 0.3285)
shuck <- as.character(shuck)
shuck <- revalue(shuck, c("TRUE"="ABOVE", "FALSE"="BELOW"))

median(mydata$VOLUME)
#0.03257888
volume <- factor(mydata$VOLUME > 0.03257888)
volume <- as.character(volume)
volume <- revalue(volume, c("TRUE"="ABOVE", "FALSE"="BELOW"))

shuck_volume <- table(shuck, volume)

#Test for independence using a chi square test.
#State the hypothesis
#-----
# H0: Volume is independent of shuck weight
# H1: Volume and shuck weight are dependent

#Show the table
#-----
shuck_volume
#           volume
# shuck  ABOVE BELOW
# ABOVE   225    24
# BELOW   25    226

#The results
#-----
summary(shuck_volume)
# Number of cases in table: 500
# Number of factors: 2
# Test for independence of all factors:
#      Chisq = 323.2, df = 1, p-value = 2.891e-72

#Conclusion
#-----
```

```
#The p-value is less than 0.05 so we reject the null hypothesis.
#It appears volume is dependent on the shuck weight.

# STEP 6
#####

library(grid)
library(gridExtra)

#Plot VOLUME, WHOLE and DENSITY versus RINGS using ggplot().
#In these plots use SEX as a facet to color the plot.
shuckwholePlot <- ggplot(mydata, aes(VOLUME, SHUCK)) + ggtitle("Shuck v.s. Volume") +
  geom_jitter()
shuckwholeLPlot <- ggplot(mydata, aes(log(VOLUME), log(SHUCK))) + ggtitle("L_Shuck v.s.
L_Volume") + geom_jitter()

grid.arrange(shuckwholePlot, shuckwholeLPlot, ncol=2, main="Shuck vs Volume")

# STEP 7
#####
#Regress L_SHUCK as the dependent variable on L_VOLUME.
#Plot the regression line with 95% confidence limits on a scatterplot

reg <- ggplot(mydata, aes(x=log(VOLUME), y=log(SHUCK))) +
  geom_point() +
  geom_smooth(method="lm", level = 0.95) +
  labs(x="L_Volume", y="L_Shuck") +
  ggtitle("L_SHUCK vs L_VOLUME Regression")

#Plot regression line equation on plot
#Source: http://stackoverflow.com/questions/7549694/ggplot2-adding-regression-line-equation-and-r2-on-graph

df <- data.frame(cbind(log(mydata$SHUCK), log(mydata$VOLUME)))
colnames(df) <- c("y", "x")

lm_eqn <- function(df){
  m = lm(y ~ x, df);
  eq <- substitute(italic(y) == a + b %.% italic(x)*", "~italic(r)^2~"=="~r2,
    list(a = format(coef(m)[1], digits = 2),
          b = format(coef(m)[2], digits = 2),
          r2 = format(summary(m)$r.squared, digits = 3)))
  as.character(as.expression(eq));
}

#final plot
regplot <- reg + geom_text(aes(x = -6,
  y = -1,
  label = lm_eqn(df))
```



```
),  
parse = TRUE)
```

```
regplot
```

```
#How well does the regression model fit the data?  
LMshuckVolume <- lm ( log(SHUCK) ~ log(VOLUME), data = mydata )  
LMshuckVolume  
#  
# Call:  
# lm(formula = log(SHUCK) ~ log(VOLUME), data = mydata)  
#  
# Coefficients:  
# (Intercept) log(VOLUME)  
# 2.2510 0.9728  
#  
summary(LMshuckVolume)  
#  
# Call:  
# lm(formula = log(SHUCK) ~ log(VOLUME), data = mydata)  
#  
# Residuals:  
# Min 1Q Median 3Q Max  
# -0.72164 -0.12503 -0.00179 0.13078 0.63459  
#  
# Coefficients:  
# Estimate Std. Error t value Pr(>|t|)  
# (Intercept) 2.25102 0.03989 56.43 <2e-16 ***  
# log(VOLUME) 0.97276 0.01079 90.16 <2e-16 ***  
# ---  
# Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
#  
# Residual standard error: 0.1937 on 498 degrees of freedom  
# Multiple R-squared: 0.9423, Adjusted R-squared: 0.9422  
# F-statistic: 8129 on 1 and 498 DF, p-value: < 2.2e-16
```