Predicting Meat Production Weights in Abalone

Caleb McWhorter

April 25, 2017

1 Introduction

One of the fastest growing subsections of the food industry is the seafood industry. Among these industries is the abalone meat industry. Abalone are a large sea snail. While their shell is greatly admired for its distinct colorations, its meat is also consumed. A vibrant abalone fishing industry, in addition to a farming industry, can even be found on the west coast of the United States. The purpose of our study is to predict the meat production from the weight of the abalone itself as well as from its dimensions. We also investigate whether there is a difference between male and female abalone weight or meat production.

2 Data Collection

The data set consists of measurements from 4,177 different abalone at various stages of development. Each abalone had the following characteristics examined:

Table 1: Data set characteristics and measurement details.

Characteristic	Unit	Detail
Sex	M, F, I	Male, Female, or Infant.
Length	*mm*	Longest shell measurement.
Diameter	*mm*	Perpendicular to length.
Height	*mm*	Including meat in shell.
Whole Weight	*g*	Whole abalone.
Shucked Weight	*g*	Meat Weight.
Viscera Weight	*g*	Gut Weight (post bleeding).
Shell Weight	*g*	After drying
Rings	–	+1.5, gives age in years.

Any continuous variables were scaled by a factor of 200 for use with an ANN. The age of an abalone is determined by cutting through the cone of the shell, staining it, and under a microscope counting the number of rings. All other measurements follow from simple weighings. Any missing values were removed before the data set was compiled. Note the original study suggested that introducing the variables "weather patterns" and "location" (relating to food availability) may be necessary to accurately predict age. This remark replies equally well to our models as it is clear that both variables directly affects the development of abalones. Future data collections of abalone involving the above variables should try to include this data for possibly more accurate modeling. Since the variables in this data set are easily directly observed/measured and their values hold little impact, there is no reason to believe any bias in the data collection. Hence, any statistical analyses conducted in this paper influenced by such factors should not be effected.

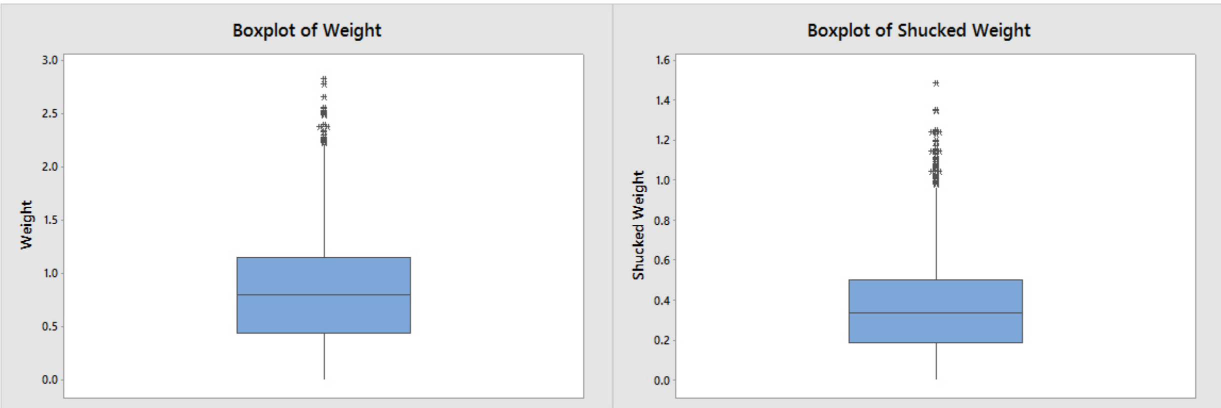
3 Analysis & Results

A basic statistical analyses of the examines variables is found in the table below.

Table 2: Basic statistics of the examined variables.

	Length	Diameter	Height	Whole W.	Shucked W.	Viscera W.	Shell W.	Rings
Min	0.075	0.055	0.000	0.002	0.002	0.001	0.002	1
Max	0.815	0.650	1.130	2.826	1.488	0.760	1.005	29
Mean	0.524	0.408	0.140	0.829	0.359	0.181	0.239	9.934
SD	0.120	0.099	0.042	0.490	0.222	0.110	0.139	3.224
Correl	0.557	0.575	0.557	0.540	0.421	0.504	0.628	1.0

Since the primary variable of interest for this study is weight and shucked weight, we investigate these variables further. Box plots of weight and shucked weight are found in Figure 1. The mean weight of the abalones was 0.82874 g while the average shucked weight was 0.3597 g. Thus, abalone meat producers can expect on average a 56.6% decrease in meat weight from the whole abalone. Observe that both weight and shucked weight are non–normal, each having a distinct right skew. This is perhaps a happy fortune of nature for those in the industry.



(a) Boxplot of weight with outliers.

(b) Boxplot of shucked weight with outliers.

Figure 1: A comparison of weight and shucked weight using boxplots.

It is not known from the data set source whether the measured abalones were wild or captive–though it is assumed that were wild abalone. It is important to note that the right–skewness of the data could be easily affected by this condition. If this region is rich in food source or provides other benefits to the abalone, this could explain the presence of overly–large abalone. On the other hand, if these abalone were raised in captivity, the restricted conditions could stunt their growth and we should expect to see even larger variety of abalone, meaning our mean is simply too low. Equally, abalone raised in captivity could receive more resources than those available to wild abalone. Though given the size of the data set and assuming these are wild abalone, it is more likely that this region is a rich environment for the abalone or that deficiency in size for abalone is rare. A 99% confidence interval for the weight and shucked weight of the examined abalone was $(0.80919, 0.84830)$ grams and $(0.35052, 0.36822)$ grams, respectively. Note that since the number of test subjects, $n = 4177$, normality conditions as well as skewness/outlier issues are not an issue in the construction of these intervals. Thus, means of 0.829 g and 0.359 g are accurate approximate predictors of the weight of abalone and average meat weight of the abalone, respectively.

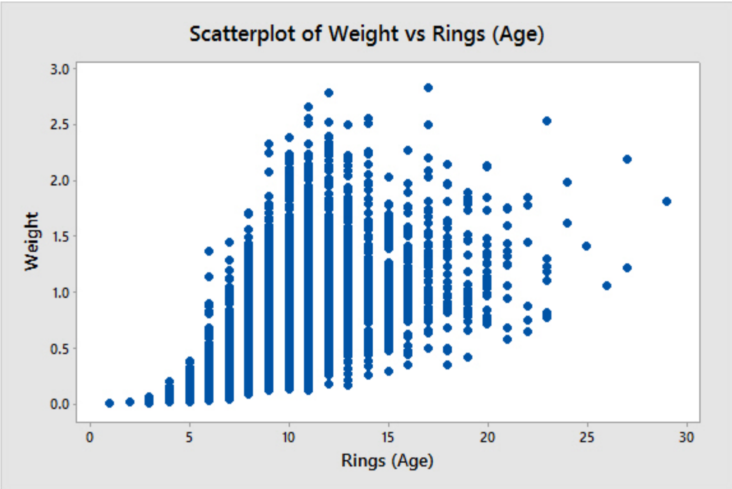


Figure 2: Comparing age and weight of abalone.

Certainly, the age and weight of abalone are related. The exact relationship between the two is certainly of interest to those in the abalone industry. Those that raise abalone in captivity would be benefitted by knowing the best age at which to harvest abalone meat. Vice versa, it could be useful to predict the age of an abalone solely based on its weight. This would avoid counting rings–thus killing the abalone. However, upon examining Figure 2, there seems little predictive relationship between the two.

Clearly, there is no hope of a linear relationship between the two–a linear regression produces $W(A) = 0.01227 + 0.08219A$, where A is the number of rings and W is the weight in g, with an R^2 value of 29.2%. Though the lack of predictive power of the model can be seen directly from the regression itself. While the constant 0.01227 g, predicting the birth weight of the abalone, is believable, the slope of the model 0.08218 g/ring is a clear under–predictor for the weight on average. An abalone possessing 15 rings would have a predicted weight of 1.245 g. This clearly places such an abalone among the lighter abalone with 15 rings.

However, we can see not only a positive correlation between the two, but an increasing trend. Meaning, the rate at which the weight increases with age is increasing. There even appears to be an exponential trend. But this cannot be a physical reality. Reversing the relationship, we produce Figure 3. The situation for possibly creating a model here is better given an apparent logistic relationship. However, the model

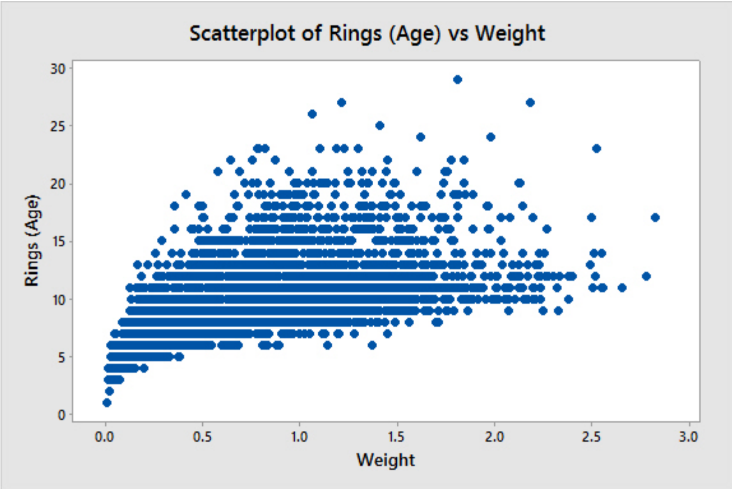


Figure 3: Comparing age and weight of abalone.

$A(W) = 10.93 + 5.187 \log_{10} W$ does little better with $R^2 = 34.1\%$. This says there is little hope in predicting the age of an abalone solely based on its weight.

There is, however, more hope in predicting the shucked weight of abalone solely based on its length, diameter, height, and weight. Though this should come as no surprise given these dimensions certainly determine the 'size' of an abalone which clearly is related to the total amount of meat an abalone could provide. Running a multiple linear regression yields the following data:

Table 3: Multiple linear regression model output.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	193.750	48.4375	16852.59	0.000
Length	1	0.175	0.1749	60.84	0.000
Diameter	1	0.135	0.1351	46.99	0.000
Height	1	0.212	0.2118	73.69	0.000
Weight	1	26.973	26.9728	9384.49	0.000
Error	4172	11.991	0.0029		
Total	4176	205.741			

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
0.0536115	94.17%	94.17%	93.94%

```
Coefficients

            Term          Coef  SE Coef  T-Value  P-Value    VIF
            Constant  -0.00026  0.00640    -0.04    0.968
            Length      0.3395   0.0435     7.80    0.000  39.69
            Diameter   -0.3653   0.0533    -6.85    0.000  40.63
            Height     -0.3167   0.0369    -8.58    0.000   3.46
            Weight     0.45240  0.00467    96.87    0.000   7.62
```

Observe the extraordinarily large F and t–values indicate fair predictive power for the model. Furthermore, we have $R^2 = 94.17\%$ so that despite the lack of linear relationship between weight and age, there is a fair linear relationship between the selected variables and the shucked weight.

Finally, it may be useful for producers in the abalone meat industry, especially those producers which farm abalone in their own facility, to know whether there is a difference between average total meat production between male and female abalone. Conducting a two–sample t–test for the mean weight yields a 95% confidence interval of $(0.0219, 0.0883)$ with $p = 0.001$. [Again, note that while the data may skewed the number of data points, 1307 females and 1528 males, means the effect on the t–test should be negligible.] While the difference is slight, there is a statistical difference. However, is the difference in weight a result of the male abalones possessing more meat or is there something else at hand? Conducting the same test for shucked weight between males and females yields a 95% confidence interval of $(-0.00229, 0.02877)$. This interval includes 0 so that it is possible that there is no difference in meat production between the genders. Indeed, the p–value for this test is 0.095. Why then is there a difference in weight between the genders? Again conducting a two–sample t–test between males and females shell weight gives a 95% confidence interval of $(0.01058, 0.02950)$ with p–value 0.000. This suggests the increased weight of male abalones is a result of males possessing a larger shell. Indeed, we can compare the production between males and females on a more refined scale via a chi–squared test. We declare small/large abalones to have weight two standard deviations below/above the mean for all abalones, respectively. This produces the data found in Figure 4.

Table 4: Counts of small, medium, and large abalones.

	Male	Female
Small	530	402
Medium	37	32
Large	961	873

A chi–squared test yields a p–value of 0.083, not statistically significant. There seems to be little difference between male and female abalones in size, except perhaps in shell size.

4 Conclusion

Our study finds a definite linear relationship between an abalones dimensions along with weight and its total meat production. However, the analyses fail to find a useful relationship between age and meat production. Thus while we cannot say what may be the best age to harvest the abalone, we can make predictions about total meat production given the weight of a collection of abalone. This study also finds no difference between male and female abalone meat production. Hence, there need not be a focus in abalone harvesting/farming between the genders to maximize meat production. Future studies should try to refine this data by collecting abalone from different regions, keeping track of weather conditions and also compare captive versus wild abalone. Future studies should also include more analyses investigating the relationship between age and weight of abalone as this would be a more useful predictor for the abalone farming industry.

References

[1] David S. Moore, George P. McCabe, Bruce A. Craig. *Introduction to the Practice of Statistics*. W.H. Freedman, 2014.

[2] M. Lichman. UCI Machine Learning Repository. 2013.

[3] Warwick J. Nash, Tracy L. Sellers, Simon R. Talbot, Andrew J. Cawthorn and Wes B. Ford. The Population Biology of Abalone (Haliotis Species) in Tasmania: Blacklip abalone (H. rubra) from the north coast and the islands of Bass Strait. 48, 1994.