Analysis of Abalone Rings

John Rempe Salvatore Palmeri Dr. Khyam Paneru (Mentor)
Department of Mathematics, The University of Tampa

Introduction

The current method of determining the age of abalone is by counting its number of rings in a long process involving cutting the shell, staining it, and examining the rings under a microscope. A dataset from researchers in a biological study collected physical measurements from blacklip abalone to find an alternative process for determining their age. With roughly 4 thousand observations in the dataset, our goal was to find the most significant attributes that accurately predict abalone age. After taking a random sample of 500 observations between male and female abalone, we tested to see if their physical attributes could replace the tedious process of determining age.

Table 1 – Variables for Data Exploration

Predictor/Response Variables	Details
Sex	0 for Male, 1 for Female
Length	Longest shell measurement (mm)
Diameter	Perpendicular to length (mm)
Height	With meat in shell (mm)
Whole Weight	Weight of whole abalone (g)
Shucked Weight	Weight of meat (g)
Viscera Weight	Gut weight after bleeding (g)
Shell Weight	Weight after being dried (g)
Rings	+1.5 gives the age in years

Figure 1 – Blacklip Abalone

Table 2 – Summary Statistics of Physical Measurements

Variable	Minimum	Q1	Median	Mean	Q3	Maximum	Std. Dev.
Length	0.18	0.515	0.59	0.527	0.64	0.775	0.093
Diameter	0.125	0.405	0.465	0.45	0.5	0.63	0.095
Height	0.05	0.135	0.155	0.155	0.175	0.25	0.04
Whole Weight	0.023	0.712	1.026	1.013	1.297	1.826	0.433
Shucked Weight	0.009	0.293	0.428	0.437	0.565	1.351	0.204
Viscera Weight	0.006	0.153	0.221	0.223	0.285	0.76	0.099
Shell Weight	0.01	0.208	0.292	0.293	0.375	0.897	0.124
Rings	3.00	9.00	10.00	10.9	12.00	25.00	2.94

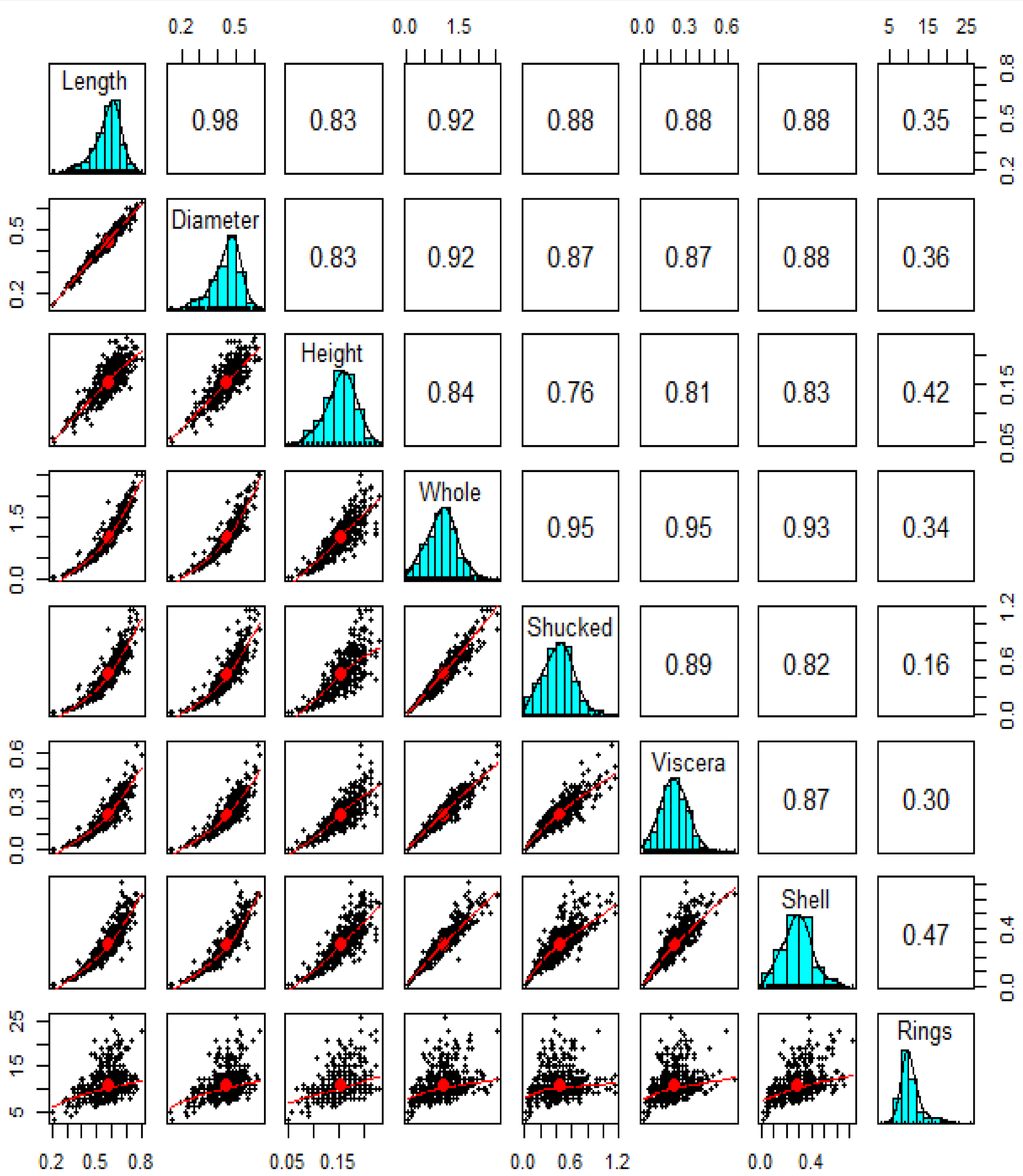

Figure 2 – Pairwise Scatterplots, Distributions, and Correlations

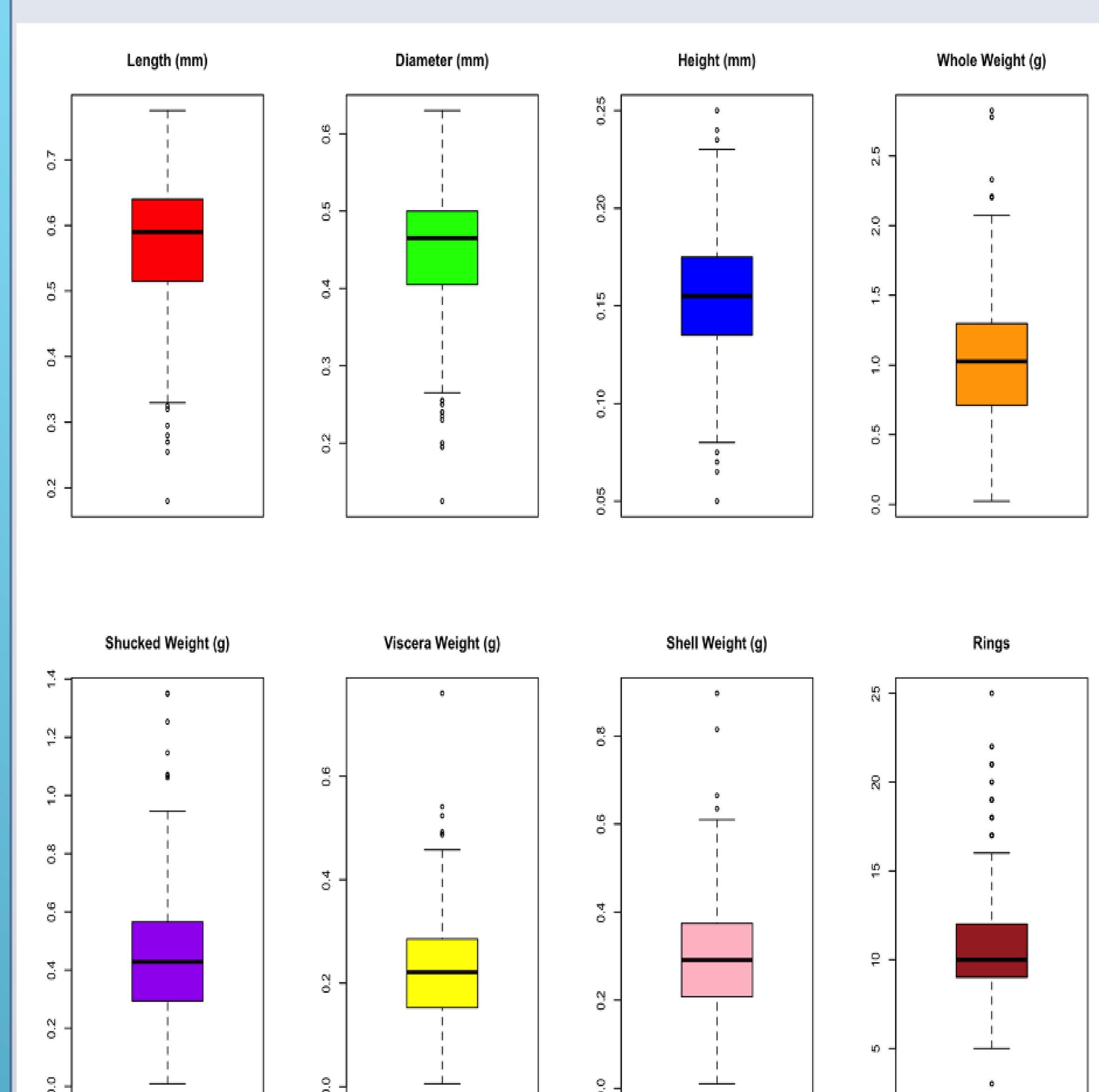

Figure 3 – Boxplots of Physical Measurements

The means and medians of physical measurements are very similar, indicating a symmetric distribution. The scatterplots for each pair of variables show that many of the relationships are linear. The histograms show mostly right-skewed distributions, indicating that most abalones are light in weight. The correlations show that the physical measurements have a strong positive correlation with each other, but only a moderate correlation with Rings. All physical measurements display a few outliers in their plots, but this represents the range of sizes between the abalone.

Regression Model of Rings

Overall, the model is highly significant, with an F-statistic of 42.53 and low p-value. This indicates that the predictors explain a meaningful portion of the variation.

Table 3 – Model with Transformation and Influential Values

Variable	Estimate	Standard Error	T Value	P(> |t|)	Significance
Intercept	1.84	0.102	18.063	0.000	***
Length	-0.28	0.514	-0.546	0.586	
Diameter	0.719	0.62	1.16	0.247	
Height	1.924	0.566	3.4	0.000	***
Whole	0.727	0.168	4.331	0.000	***
Shucked	-1.518	0.184	-8.269	0.000	***
Viscera	-0.767	0.291	-2.639	0.009	**
Shell	0.568	0.266	2.138	0.033	*
Sex	-0.029	0.018	-1.67	0.096	
R- Squared: 40.93% Adjusted R-Squared: 39.97%			F = 42.53 on 8 and 491 DF p-value < 0.000		p-value < 0.000

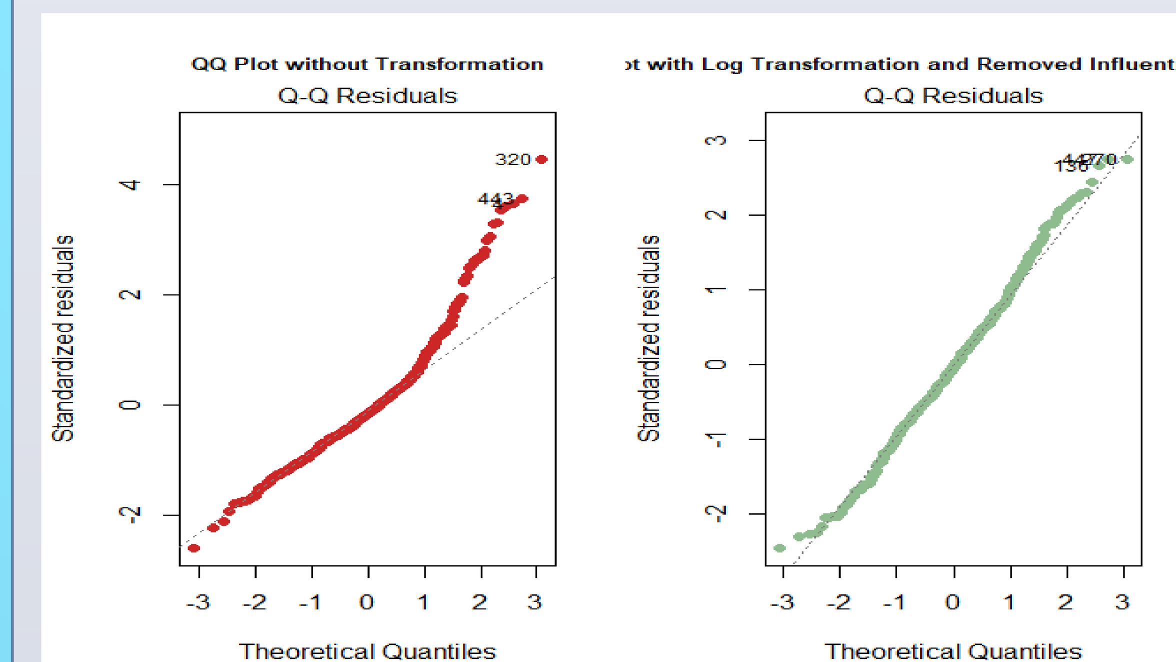

Figure 4 – Transformation of response variable

The original model's QQ plot shows deviations from normality, indicating skewed and unstable residuals. After performing a log transformation on Rings, the normality improved slightly. Finally, removing the influential points with the log transformation produced the best QQ plot. It confirmed that those observations were affecting the residual distribution.

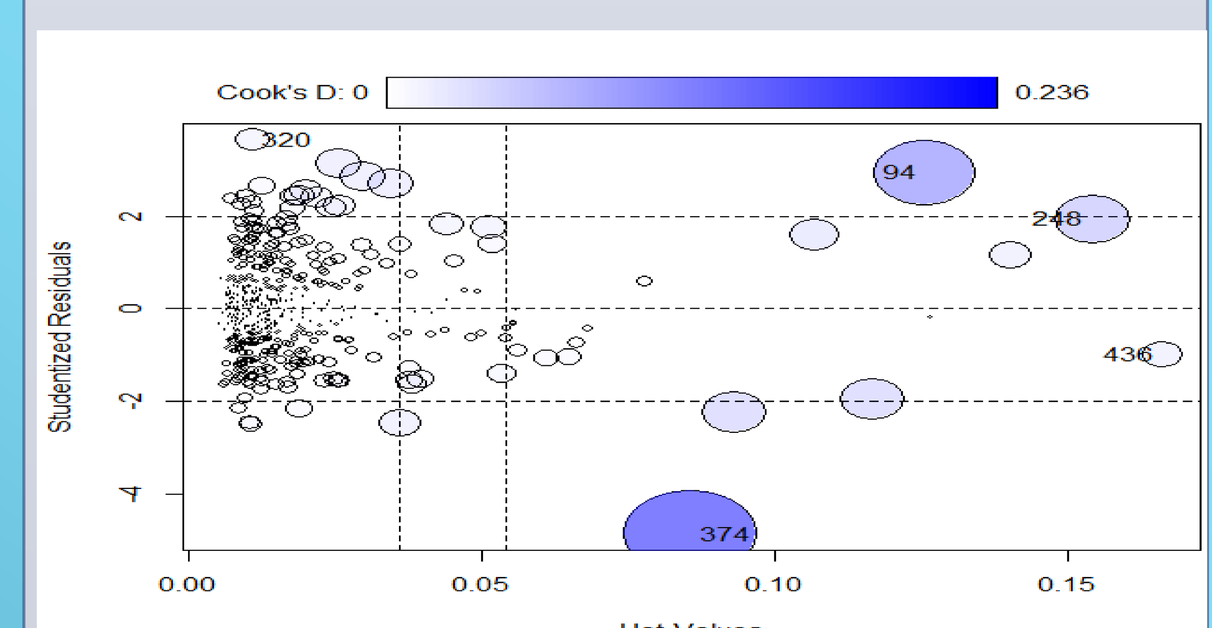

Figure 5 – Influential Values Plot

Transformed and Removed Model

Table 4 – Summary of Model with Transformation and Removed Influential Values

Variable	Estimate	Standard Error	T Value	P(> |t|)	Significance
Intercept	2.142	0.111	19.365	0.000	***
Length	-0.674	0.485	-1.388	0.166	
Diameter	0.268	0.598	0.449	0.654	
Height	1.627	0.544	2.99	0.003	**
Whole	0.861	0.215	4.006	0.000	***
Shucked	-1.526	0.226	-6.761	0.000	***
Viscera	-0.878	0.341	-2.577	0.010	*
Shell	0.803	0.312	2.577	0.010	*
Sex	-0.049	0.016	-3.121	0.002	**
R-Squared: 40.95% Adjusted R-Squared: 39.88%			F = 38.94 on 8 and 444 DF		p-value < 0.000

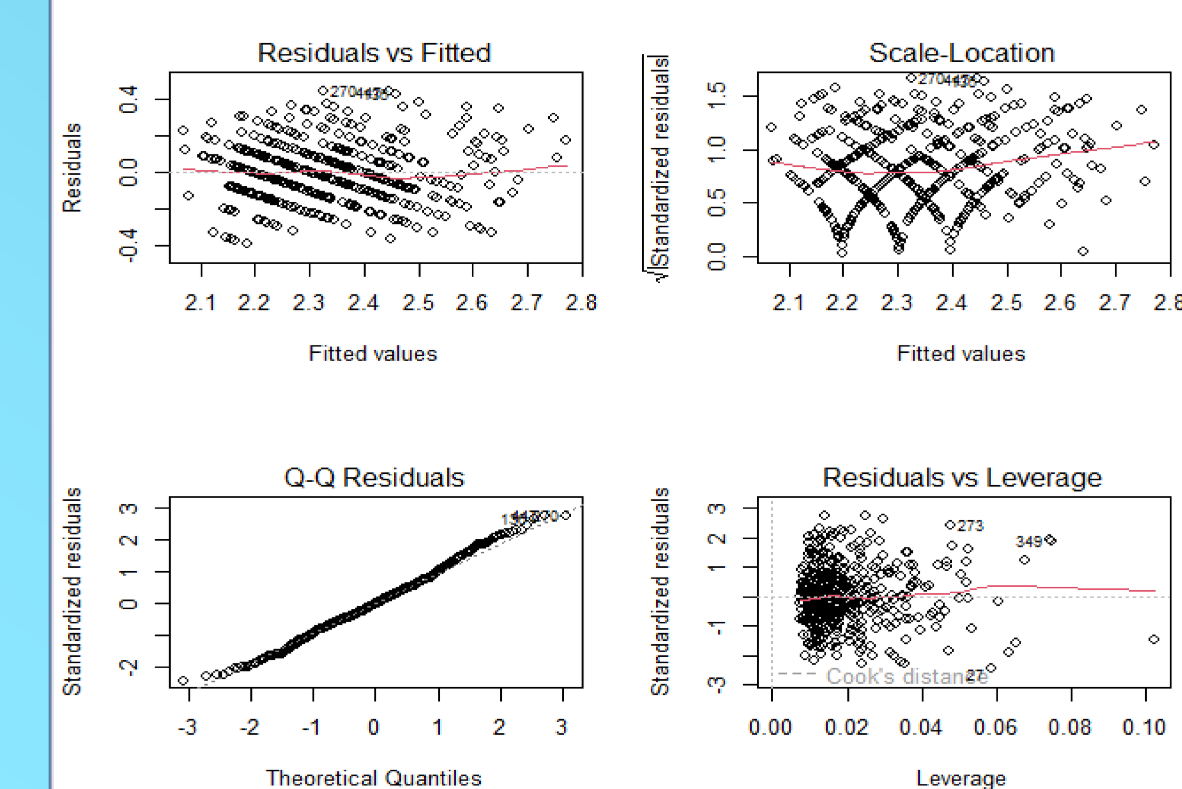

Figure 6 – Regression Diagnostic for Plots Final Model

In the final model, the same predictors remained statistically significant, with the addition of Sex becoming statistically significant and levels of significance varying. The estimates are more stable, and the model improved as seen by the 40.95% R-squared value.

Conclusion:

Although our original and final model did not produce the highest R-squared values, it is completely normal given that a trait such as age is influenced by many unmeasured biological and environmental factors. The model still identifies several statistically significant predictors meaning it captures meaningful relationships even if it doesn't explain all the variability. Overall, this provides a strong starting model for predicting age of abalone without relying on the traditional time-consuming methods. With additional data beyond physical features such as genetics, environment, or growth conditions, this model could be further improved to make age prediction more accurate and reliable.

FishFiles. (n.d.). *Abalone fishery*. FishFiles: Fisheries & Farms – VIC. Retrieved November 21, 2025, from https://www.fishfiles.com.au/fisheries-and-farms/vic-fisheries-and-farms/abalone-fishery

Nash, W., Sellers, T., Talbot, S., Cawthorn, A., & Ford, W. (1994). Abalone [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C55C7W.