Abalones Research

Dharun Somalingam

Executive Summary

The given report provides a detailed analysis of the relationship between the age of Abalone(number of rings) and variables that might influence them. The Dataset consists of 4177 randomly selected samples of abalone(marine snails) is available, including the number of rings in it along with some physical estimations. A Multilinear Statistical model needs to be fit and visualize the given data which helps us to predict the Age of Abalones.

Introduction

Background

Abalones are marine gastropod mollusks belonging to the family Halictidae. They are known for their beautiful iridescent shells and are commonly referred to by various names in different parts of the world, including ear shells, sea ears, muttonfish or mutton shells in parts of Australia, ormer in the UK, perlemoen in South Africa, and pāua in New Zealand. Abalones are highly valued for their meat, which is considered a delicacy in many cuisines, and their shells are also used for various decorative and ornamental purposes.

Research Related

According to research conducted by Peter A. Cook (The University Of Western Australia, Albany, Australia),3 Global fish production continues to outpace world population growth, and aquaculture remains one of the fastest-growing food-producing sectors. In 2012, global aquaculture production was 90.4 million tonnes. Although, in terms of production tonnage, abalone contributes a relatively small proportion of this aquaculture production, it is one of the most highly prized seafood delicacies and, therefore, in terms of the value of production, is very important to many countries.

Source of DataSet

This dataset was collected from UCI Machine Learning Repository and donated by S.Waugh on 1995-12-01. The given data consist of 4177 samples of Abalones with 9 different variables. The first variable was sex which is categorical and the others were physical measurements(continuous variables). The last variable is the rings, plus 1.5 which equals the age of abalone according to the definition of abalone's age.

About Analysis and Data

This analysis aims to predict the age of abalones(number of rings) using 8 different variables. So, the classification of data into independent and dependent Variables. Independent variable: At first Sex, abalones have three different sexes. Those are Male, Female and Infant. The

Second is Length, which measures the longest shell of an abalone. Diameter is perpendicular to length and Height is the measure of meat. Then there is the whole weight of abalones with the shucked weight that measures the meat weight. Seventh, there is a Viscera weight to measure the gut weight. The final variable was shell weight measures the weight of abalones after being dried. Dependent variable: The dependent variable is Rings which we are going to make prediction is the age of abalone.

Methodology

Data processing

Check the data clearly, find any mismatching's or null values in the given data before performing the analysis and interpret the Summary. It was measured as the number of rings observed after cutting and examining an abalone. Although it does not denote the age of a given abalone directly. The age of an abalone equals Rings + 1.5 . Since this relationship holds reliably, Rings will be considered as the dependent variable

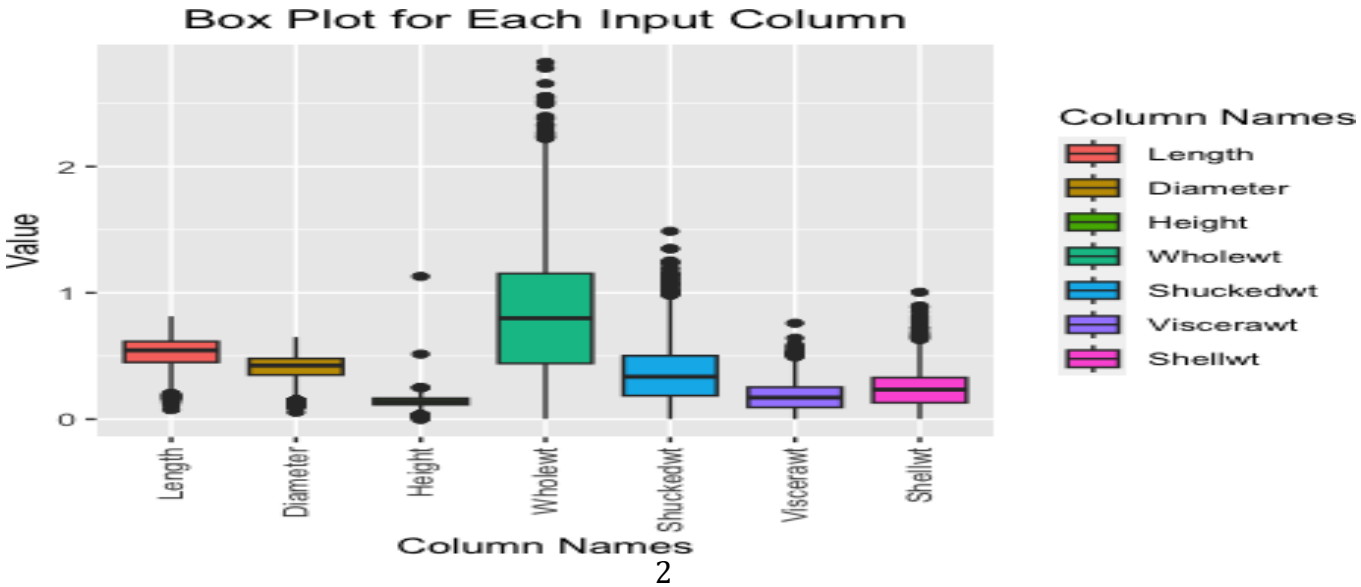
Linear Model*:*

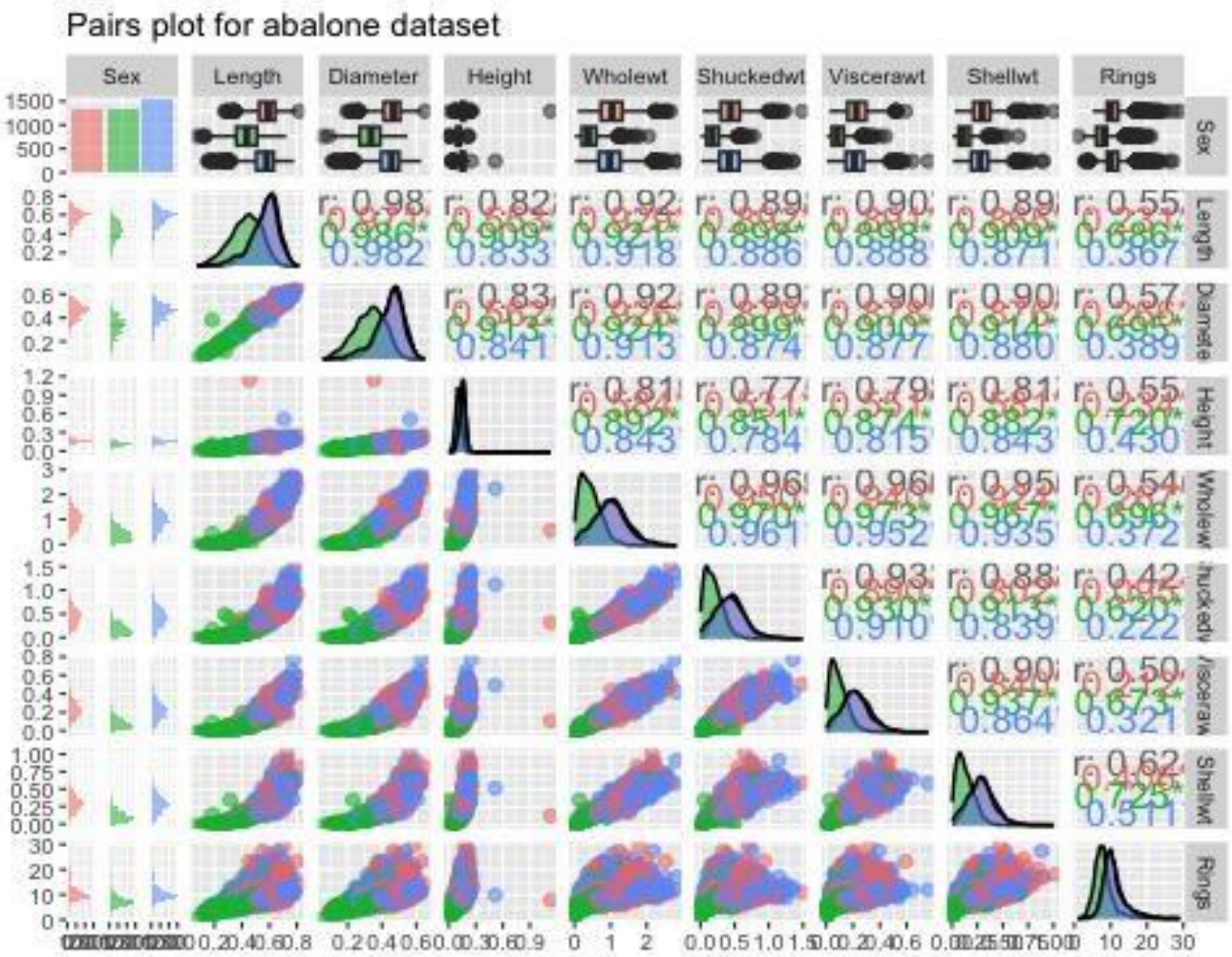
We have to fit the model with all variables and look for the significance of each parameter. Based on this, our model will be re-fitted again. For now, we are fitting Multi Linear Model which includes all variables.

In the Interaction model, we are going to fit the model with some interactions between the variables Length * Diameter Interaction: Interaction between the length and diameter of the shell. Explore whether the combination of length and diameter has a different impact on the number of rings.

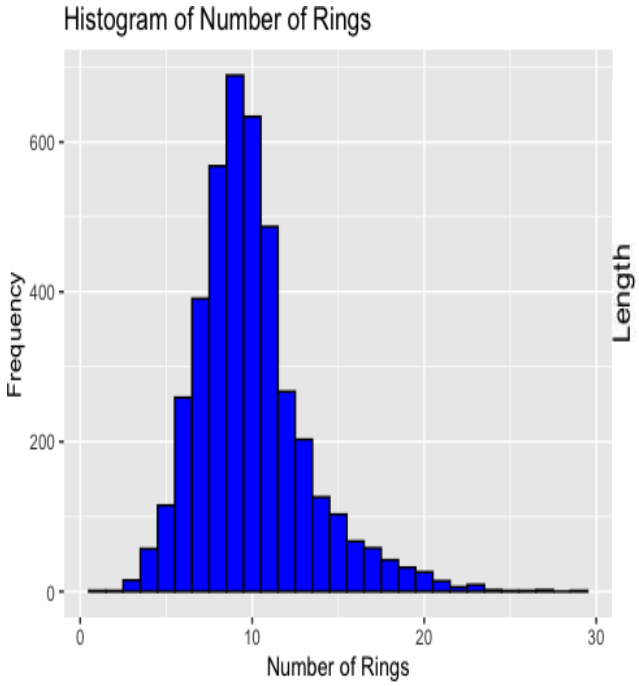
Results

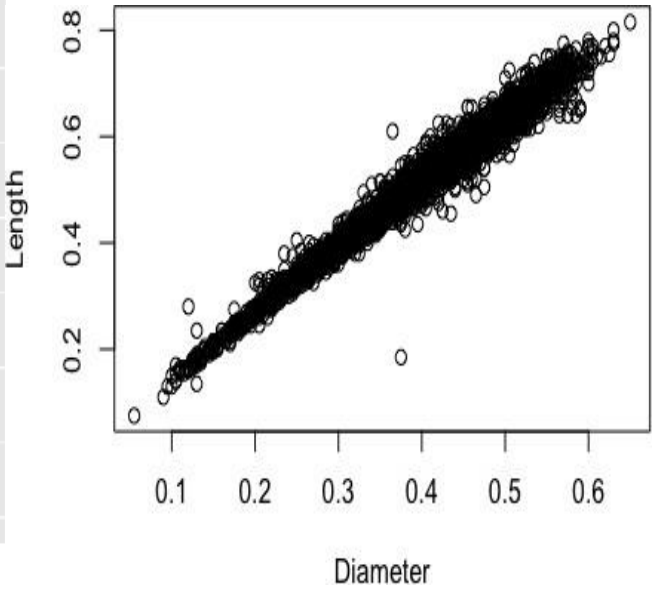
The first thing is to note the correlations within the variables in the given data. The correlation between the Length and Diameter is extremely high about 98.7. Similarly, Whole





weight seems to be highly correlated with other weight predictors and is the sum of Shucked weight, Viscera weight and Shell weight.


Histogram of Number of Rings
Frequency
600
400
200
0
0
10
20
30
Number of Rings

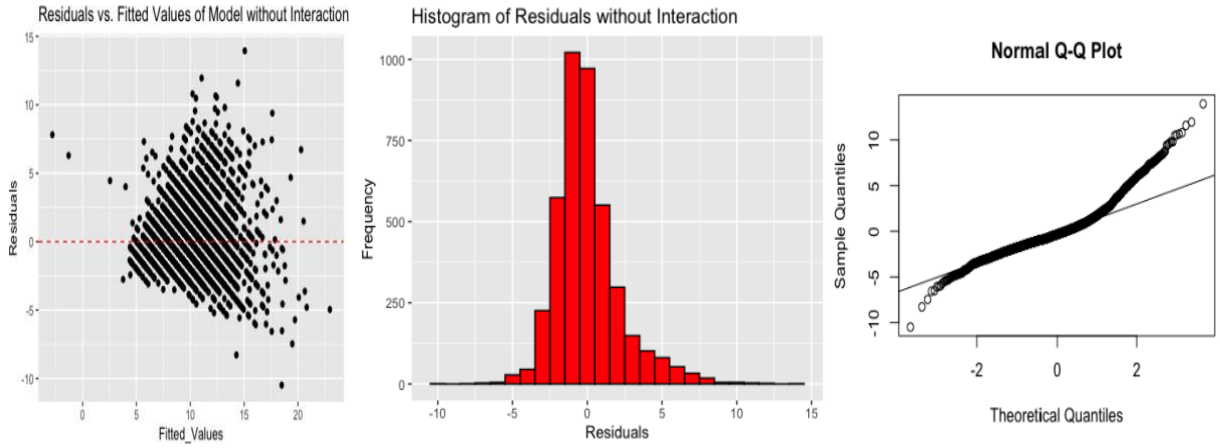

Length
0.8
0.6
0.4
0.2
0.1
0.2
0.3
0.4
0.5
0.6
Diameter

Secondly, the distribution of predictor Sex with factor level value of female and male are very similar concerning all other predictor. Most of the abalone rings are between 5 and 15

Modelling:

Model 1:

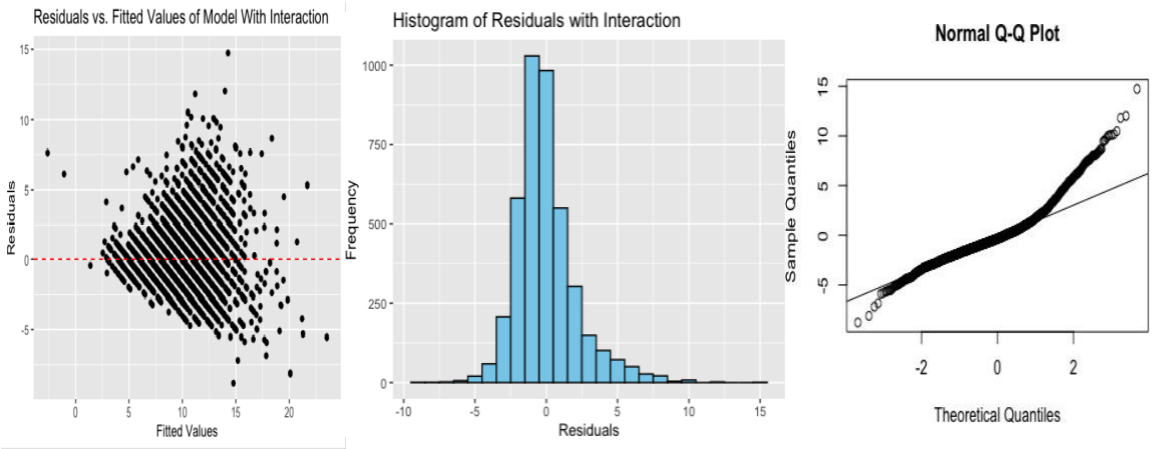
Rings (Age)=3.89−0.82×SexI+0.05×SexM−0.45×Length+11.07×Diameter+10.761×Height+8.975×Wholewt−19.786×Shuckedwt−10.581×Viscerawt+8.741×Shellwt



We have started with fitting an additive model with all variables and will look at the significance of the parameters. For now we will work with original value for variable Sex which have factor levels of F, I and M. The model explains approximately 53.79% of the variability in the number of rings, and the adjusted version of R square is 53.69% with standard error in residual of 2.194.

Model 2:

Rings (Age)=0.165−0.76×SexI+0.02×SexM+7.98×Length+26.43×Diameter+7.02×Height+10.03×Wholewt−18.62×Shuckedwt−8.52×Viscerawt+10.70×Shellwt−38.03×Length:Diameter



Similarly, we will introduce an Interaction in the model within Length and Diameter. Based on that, we will analyses the model. Whereas, the Multiple R square of this model is 55.06% and the adjusted R square is 54.95% with a Residual standard error of 2.164.

Discussion

Within the two linear models fitted with and without interactions, based on the above metric, Model 2 seems to be a bit better with a Higher adjusted R square and with a lower Standard error. But also, using some model diagnostics and AIC values supports Model 2 compared to Model 1. However, it's important to note that the choice of the "better" model also depends on the specific goals of your analysis and the context of your study.

However, it's important to acknowledge potential limitations, such as the quality and representativeness of the dataset, and the assumptions made during modeling. We found that there were some instances where the information was not accurate. For example, In 2 observations Height of abalones is 0 which is practically not possible and when we look at other variables we find some Outliners

Reference

1. https://en.wikipedia.org/wiki/Abalone
2. Cook, P. (2014) The Worldwide Abalone Industry. Modern Economy, 5, 1181-1186. doi: 10.4236/me.2014.513110.
3. Nash,Warwick, Sellers,Tracy, Talbot,Simon, Cawthorn,Andrew, and Ford,Wes. (1995). Abalone. UCI Machine Learning Repository. https://doi.org/10.24432/C55C7W.
4. AIlhan Uysal and H. Altay Guvenir. Instance-Based Regression by Partitioning Feature Projections. Applied. 2004.
5. Linghui Fu, & Wang Huiwen. (2004). A comparative study of modeling methods for polynomial regression. Journal of Mathematical Statistics and Management, 23(1), 5.
6. Miles, J. (2005). R-Squared, Adjusted R-Squared. In Encyclopedia of Statistics in Behavioral Science (eds B.S. Everitt and D.C. Howell). https://doi.org/10.1002/0470013192.bsa526