ABALONE PROJECT

Introduction

This report outlines, for the purposes of sustainability, methods to be used in order to identify the sex of an abalone correctly as well be able distinguish an infant abalone from the other abalone.

Additionally, for commercial benefit, this report also provides methods of estimating the viscera weight and shucked weight of the abalone.

Sustainability

Welch Two sample test with 95% confidence shows that the length, diameter, and height of an infant abalone is significantly different from other adult abalone. We can therefore use these dimensions to distinguish between an Infant and other abalone.

Table 1 below shows the mean and range of length, diameter, and height of Infant abalone at confidence level of 95%

	Mean	**95% Confidence Range**
Length	85.54 mm	84.38 mm to 86.71 mm
Diameter	65.3 mm	64.35 mm to 66.24 mm
Height	21.6 mm	21.25 mm to 21.94 mm

Table 1

Therefore, any abalone having a length lesser than 86.71 mm, diameter lesser than 66.24 mm and height less than 21.94 mm can be considered to be an infant abalone.

For abalones beyond these dimensions a classifier model is used. For this purpose, we compare models developed by Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)

As the three independent variables are not univariate or multivariate normal, SVM could the preferable classifier model

For LDA and QDA we have calculated prior probabilities from the given data, as this data is quite large and therefore can be assumed to population data.

Table 2 compares the accuracies of the three classifier models

	SVM	**LDA**	**QDA**
Female	11.15%	22.34%	14.38%
Infant	74.89%	69.30%	69.29%
Male	64.68%	61.85%	61.84%
Overall	50.76%	51.88%	52.15%
Misclassification Rate	49.24 %	48.12%	47.85%

Table 2

The table shows that the three models have similar overall accuracies. In fact, there are very accurate in identifying a male and an infant abalone. They do struggle with identifying a female abalone correctly. LDA model is the most accurate among the three in classifying a female abalone correctly.

This makes an LDA model the preferable classifier, instead of SVM as previously mentioned.

Profitability

We try an estimate the shucked weight for meat and the viscera weight from length, diameter, and height of the abalone. We first test correlations between the independent variables' length, diameter and height. All the three variable show very high and positive correlation between each other. Therefore, as the Length increases the diameter as well the height of the abalone also increases and vice a versa is also true. Table 3 shows the correlation between three variables

Dimension	**Correlation**
Length and Diameter	0.987
Diameter and Height	0.833
Height and Length	0.827

Table 3

Similarly, there is also a high degree of positive correlation between shucked weight and viscera weight. The correlation coefficient between the two is 0.932

We use canonical correlation analysis to make a weighted dimension index and a weighted index for weight of the abalone. We use raw coefficients as standardized coefficient index is not necessary in this case. This is because all the three independent variables, Length, Diameter and Height are in the same unit, mm. Similarly, the two dependent variables, shucked weight and viscera weight are all in the same unit grams.

Using the first canonical correlation

The weighted dimension index (X) is

X = - 0.0266*Length - 0.0124*Diameter - 0.0162*Height

Ans the weighted index for abalone weight (Y) is

Y = - 0.01*Shucked Weight - 0.02*Viscera Weight

Equating the two sides gives the linear combination between the dependent and independent variables

- 0.01*Shucked Weight - 0.02*Viscera Weight = - 0.0266*Length - 0.0124*Diameter - 0.0162*Height

0.01*Shucked Weight + 0.02*Viscera Weight = 0.0266*Length + 0.0124*Diameter + 0.0162*Height

The above equation shows as the length, diameter, and height of abalone increases, the shucked weight and viscera weight of the abalone.

Asymptotic Hypothesis Tests show that the first canonical correlation is statistically very significant.

Assuming, a linear relationship between the dependent and independent variables, we can use multivariate linear model.

The linear model for Shucked Weight is

Shucked Weight = -99.107 + 1.352 * Length + 0.474 * Diameter + 0.476 * Height

The adjusted R^2 value for this linear model is 0.816, this shows that in this model Shucked weight can account for 81.6% of variance seen in length, diameter, and height of abalone.

ANOVA test for this model gives a p value of less than 0.05 for the intercept and the coefficients of in length, diameter and height showing that these values are statistically significant for this linear model

The linear model for Viscera Weight is

Viscera Weight = -48.339 + 0.502 * Length + 0.255 * Diameter + 0.391 * Height

The adjusted R^2 value for this linear model is 0.8249, this shows that in this model Shucked weight can account for 82.49% of variance seen in length, diameter, and height of abalone.

ANOVA test for this model gives a p value of less than 0.05 for the intercept and the coefficients of in length, diameter and height showing that these values are statistically significant for this linear model.

The estimate of the value of that abalone: $S = V_{shucked} \times X_{shucked} + V_{viscera} \times X_{viscera}$ we can substitute for $X_{shucked}$ and $X_{viscera}$ to get

$S = V_{shucked} \times$ (-99.107 + 1.352 * Length + 0.474 * Diameter + 0.476 * Height) + $V_{viscera} \times$ (-48.339 + 0.502 * Length + 0.255 * Diameter + 0.391 * Height)

Note

1. All values in this report is stated with a confidence interval level of 95%
2. All the variables, dependent and independent variable did not have univariate or multivariate normality as they contained outliers

,