A Statistical Solution to Abalone Farming

**1. Introduction**

Abalones, a mollusk that resides in cold waters found off coasts such as Australia, South Africa, California, and Japan, are a highly demanded culinary delicacy. Abalones, like many higher-end dishes, require a lengthy and precise process to be correctly distributed into the market and are known for being the rarest and most valued shellfish in the world. Abalone farms exist to counter commercial farming of them and require one to two years of growth before the necessary maturity. Once farmed, abalones must be inspected for their age through opening their shell, staining them, and counting rings through a microscope. The age of these mollusks is positively correlated with their economic value on the market, so each measurement done to abalones is vital. Individual abalone farmers work alongside biologists and must have practice and experience to be considered reliable, therefore making the process often expensive and time consuming.

**2. Problem Statement**

The goal of any development in any industry is to maximize efficiency. As stated before, the processing of abalones takes time, skill, and resources. Data collection has become a global trend in optimizing many various industries and research in the abalone industry is no stranger. The tedious method of abalone processing may be limiting production efficiency of farms around the world. A better understanding of abalones’ age and corresponding value can be explained through a statistical approach. Researchers have collected data on particular shellfish to determine if there was a possibility to avoid the traditionally economically expensive processing of abalones. Using data collected on the abalone’s physical attributes, could a predictive model be constructed to determine the age of abalones before undergoing the individual measurement process?

**3. Data Set**

The data essential to answer this question was retrieved from the University of California Irvine Machine Learning Repository. The data set is multivariate with 4177 observations, 8 categories, and was originally donated to the repository in 1995. The categories consist of:

Name Data Type Meas. Description

---- --------- ----- -----------

Sex nominal M, F, and I (infant)

Length continuous mm Longest shell measurement

Diameter continuous mm perpendicular to length

Height continuous mm with meat in shell

Whole weight continuous grams whole abalone

Shucked weight continuous grams weight of meat

Viscera weight continuous grams gut weight (after bleeding)

Shell weight continuous grams after being dried

Rings integer +1.5 gives the age in years

**3. Proposed Method of Analysis**

With the main problem being determining the age (Rings) of abalones, the other variables will be used as predictor variables. The most likely approach will consist of multiple regression methodologies to accurately predict the Rings of individual abalones.