

Ethan Ducharme

Report 2.3

MTH 332 Gary Davis

Analysis of Crab Dungeness Data

from Stat Labs Data Site A Report

As Evaluation

### **Abstract**

Crab shell sizes pre molt correlate linearly with post molt sizes the correlation is a tight fit to the regression line.

### **Intro**

This data analyzed was cloned from a stat labs data site. The data was cleaned, plotted, and analyzed with regression analysis. The linear regression showed a strong fit to the line.

### **Analysis**

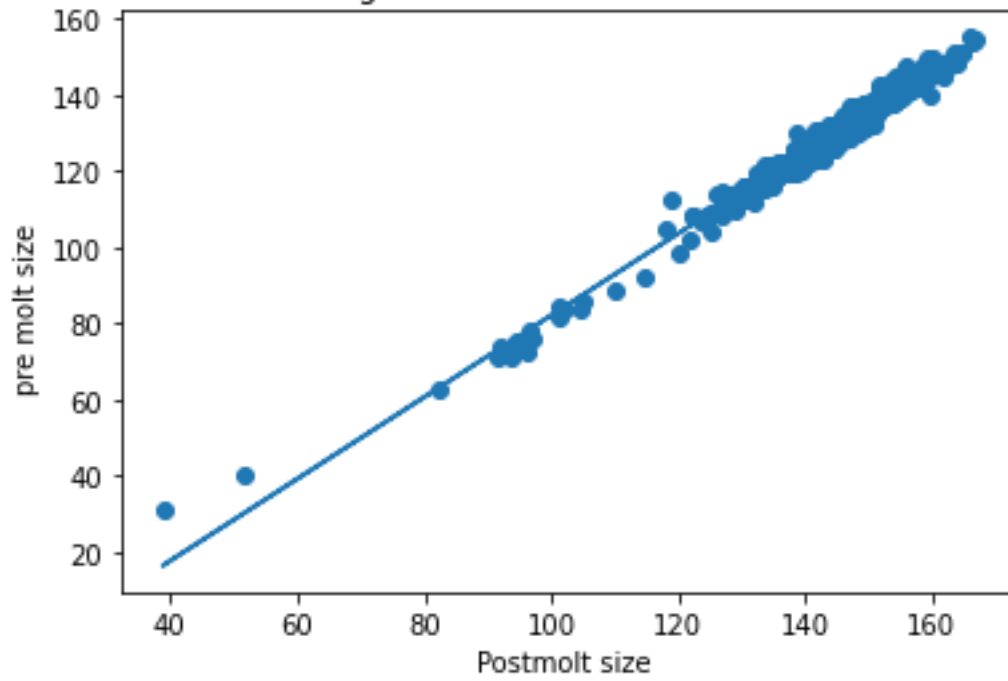
Data found on the stat labs site was parsed by the python function “pandas.readcsv”, it was parsed into a pandas Data Frame. The first column of the pandas Data frame was the scalar values for the pre-molt size Dungeness crab shell in both the lab and field denominations. The second column held scalar values for the Post molt size of the crab. These two fields were plotted on a 2-dimensional plane using matplotlib. With x corresponding to the first column

(Pre-molt) and  $y$  to the second(post-molt). The plot was entitled "Dungeness Crab Shell sizes overall". The data overall combines the data from the field and the lab, those were plotted separately depending on their values in the fifth column of the data frame for distinction between the field(0) and lab(1) with 1 and 0 as numerical flags. The three datasets were then fit to regression lines of degree 1, using python library "numpy.polynomial.polynomial". The lines were then plotted using "matplotlib.pyplot". The residuals were returned from "numpy.polynomial.polynomial" the residuals were for overall, field and lab respectively. The residuals show that linear models do not hold very well to this particular data set as the sum of the error squared (residual) is correlated relatively strongly. I then chose to plot a histogram of the same residuals in the programming language R and the hist function (to plot) and simple.fit function (to regress), leading to the fourth plot in gray. This histogram of residuals shows the line is a good fit to the data.

### **Conclusion**

A linear model is correlated strongly enough to show that pre molt size is a linear determinant of post molt size there is a slope and the data is tightly fit.

Dungness Crab Shell sizes overall



Dungness Crab Shell sizes in field

