An Investigation into Predicting the Pre Molt Carapace Length of Dungeness Crabs Using Post Molt Length Data

Gabriella Thomas

Professor Gary Davis, MTH 332 - Mathematical Statistics

University of Massachusetts - Dartmouth

Abstract

The purpose of this investigation is to determine a relationship between pre and post molt lengths of Dungeness crabs (*Cancer magister*) in order to aid lawmakers in creating female crab fishing restrictions. In order to determine and test the credibility of this relationship, simple linear regression analysis as well as graphical and comparative methods were used. It was found that the pre molt length of these crabs can be determined through the following equation:

$Y = 1.07316X - 25.21370$ where Y = Pre Molt Length and X = Post Molt Length.

Additionally, it was determined that this model was fairly accurate when actual pre molt data was compared to predicted pre molt data in two separate crab groups. Building on this, the model was used to determine the predicted frequency distribution of pre molt lengths based on post molt lengths of crabs following the May 1983 molting season. From this analysis it was determined that crabs which had not molted exhibited a greater pre molting season size and that there were more crabs which had experienced molting than those which had not. Overall, this model is fairly accurate at deterring pre molt length from post molt length, however, in the future, it will be critical to continue the experiment for a greater period of time and include data only from consecutive molting seasons in order to minimize external factors influencing crab length.

An Investigation into Predicting the Pre Molt Carapace Length of Dungeness Crabs Using Post Molt Length Data

Introduction

The marine fisheries industry provides food and economic stimulus for millions of people across the globe (Costello et al., 2012). With the great importance of the fisheries industry, it can be difficult to effectively manage the intake of the amount of fish each year. This has lead to the rise of overfishing. Overfishing has been a global issue since the founding of the commercial fisheries industry. Overfishing occurs when a population is unable to recover from the amount of fishing which takes place each year. This problem is not limited to just fish, but also other marine life such as the Dungeness crab (*Cancer magister*). On the Pacific coast of the United States, nearly the entire male population of the Dungeness crab is fished each year (Nolan, Speed, 2000). In order to combat the issues of fluctuating crab intake, the possibility of allowing female crabs to be caught is being considered. The purpose of this investigation is to determine a relationship between pre and post molt lengths in order to aid lawmakers in creating female crab fishing restrictions.

Methods

Data Collection. In order to create an accurate estimation for female crab fishing size restrictions, data was collected by scientists in combination with the California Department of Fish and Game and commercial crab fishers from northern California and southern Oregon (Nolan, Speed, 2000). The data was collected in two main sections. The first section details the pre and post molt widths of the crab carapaces of 472 female crabs at their widest points. These data were collected both in laboratory and field settings using recapture techniques. The data

were collected over a three season period (1981, 1982, 1992). The second portion of the data were collected in May of 1983 and contain information regarding carapace width as well as whether or not the crab had molted during that season.

Variable Determination. The variables which will be examined primarily in this investigation are the pre and post molt sizes of the Dungeness crab. These sizes were determined using the widest part of the carapace and did not include the spine of the crab (Nolan, Speed, 2000). The lengths were all measured in millimeters, regardless of whether the measurements were laboratory or field based. Building on this, the third variable which must be examined is the origin of the measurement. In the dataset, a "1" indicates that the crab molted in a laboratory and a "0" indicates that the crab molted in the wild. There can be varying size differences in terms of this location difference which solidifies its importance in this analysis. The final variable which will be examined in this analysis is whether or not the crab molted. In this case, a "0" in the dataset indicates that the crab did not molt whereas a "1" indicates that the crab did in fact molt.

Analytic Methods. The first step in the analytic process for this data set was to determine the relationship between pre and post molt lengths. Specifically, determining a method to predict pre molt length from post molt length. This was done by using R programming software to perform simple linear regression on the data provided. This provided a linear relationship between both of these variables.

In order to confirm the accuracy of this relationship, the predicted pre molt size was calculated and then compared to the actual pre molt size in two different post molt size groups. These comparisons were illustrated graphically using scatterplots for both predicted and actual

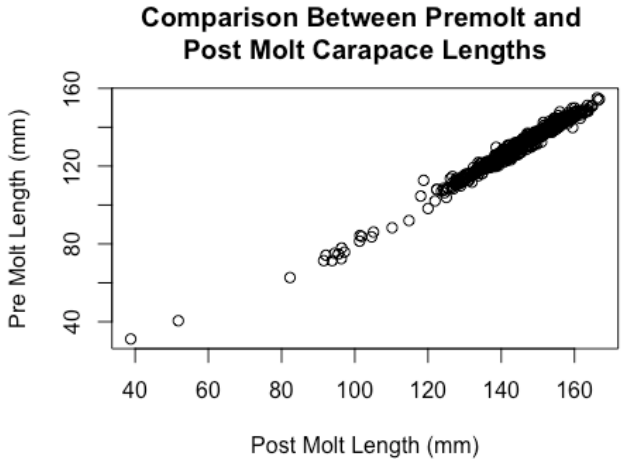
data as well as a combined scatterplot with a simple line graph for the predicted data. These plots were generated using R programming software.

The final form of analysis conducted was on completing a full prediction of the pre molt size for the crabs following the May, 1983 molting season. This was completed by first separating the second data set into two subsets: one with crabs which did molt and one with crabs which did not molt. Then, the linear relationship of pre and post molt size was applied to both of these subsets in order to determine the predicted pre molt sizes. The predicted data was then displayed in a combined histogram generated using R programming software in order to illustrate the frequencies of sizes.

Results

Firstly, the pre and post molt lengths were graphed in order to visually represent the relationship between both of these variables. This graph is shown in Figure 1 below. It was determined that there was a positive correlation between these variables with an R^2 value of 0.9808. This high value shows that there is a high association between these two variables overall.

Figure 1: Comparison between the Pre Molt and Post Molt Carapace Lengths (mm) of Dungeness crabs

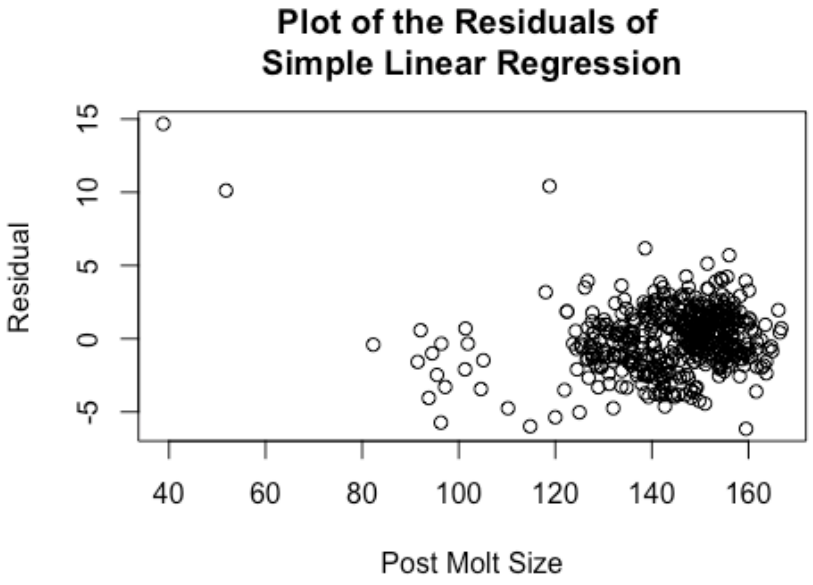


Building on this, the relationship between pre and post molt length was determined using simple linear regression techniques. This relationship can be illustrated using the linear equation below:

Y = 1.07316X - 25.21370 where Y = Pre Molt Length and X = Post Molt Length

The residual standard error for this equation was determined to be 2.199 on 470 degrees of freedom. Building on this, the residuals of the simple linear regression were plotted below in Figure 2. It was found that the residuals formed one main cluster in the range between -5 and 5 but there were some other outliers.

Figure 2: Scatter plot of the residuals from the simple linear regression analysis



The second form of analysis performed on this data was comparing the actual data of pre molt size versus the predicted pre molt size using the simple linear regression model developed

above. This was done by selecting two subgroups of post molt size data and illustrating both the predicted pre molt size and the actual pre molt size.

The first group which was studied was defined as being crabs with post molt carapace lengths between 147.5 and 152.5mm. The actual data provided from the data set is illustrated in the scatterplot of Figure 3 below. Building on this, the predicted data from the regression model is illustrated in Figure 4 below. In order to compare these two results, a combined graph was created, as seen in Figure 5. It was determined that this relationship provides a somewhat accurate prediction of the pre molt size, however, there was a high residual present in many cases. This can be observed through the high number of actual data points which lie outside of the predicted data line.

Figure 3: Scatter plot of the actual pre and post molt lengths of Group 1 [147.5 to 152.5 mm] crabs

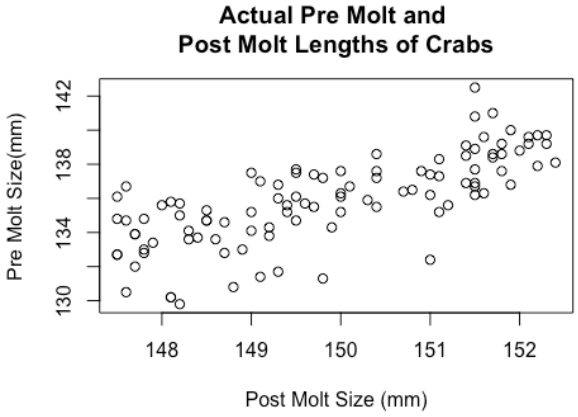


Figure 4: Scatter plot of the predicted pre molt length of Group 1 [147.5 to 152.5 mm] crabs using the Y = 1.07316X - 25.21370 relationship

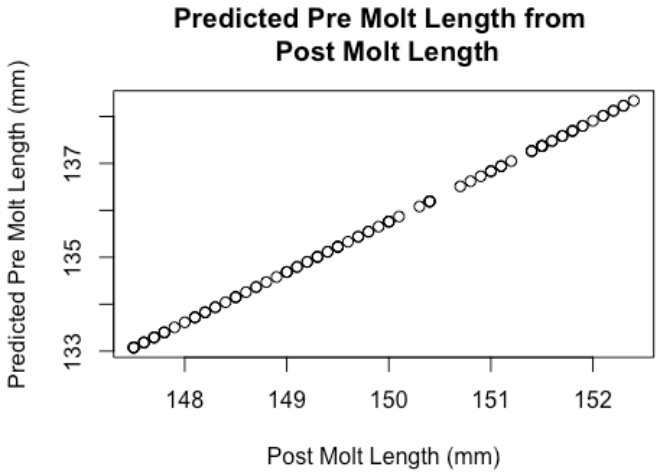
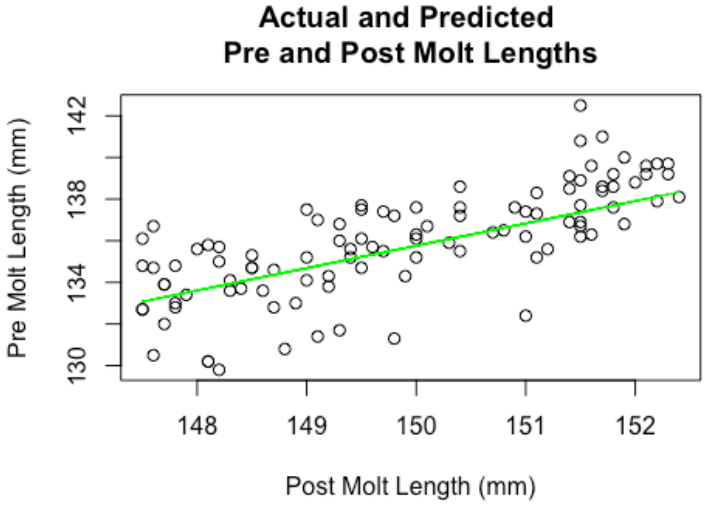


Figure 5: Combined scatter and line plots showing the actual pre molt data (scatter) and the predicted pre molt data (green line) versus the post molt length for Group 1 [147.5 to 152.5 mm] crabs



The second group which was studied was crabs with the post molt length between 125.0 and 135.0 mm. The same analysis was performed on Group 2 as in Group 1. Firstly, the actual pre and post molt lengths of Group 2 crabs were illustrated in a scatter plot depicted in Figure 6 below. Following this, the predicted pre molt lengths using the regression relationship was illustrated in Figure 7. In order to compare the actual and predicted pre molt lengths, the figures were combined in Figure 8 where the predicted lengths are shown in the blue line. It was determined that this model was more accurate at predicting the actual pre molt lengths in comparison to the results of the Group 1 analysis. This can be shown through the smaller number of data points which lie far outside of the predicted line.

Figure 6: : Scatter plot of the actual pre and post molt lengths of Group 2 [125.0 to135.0] crabs

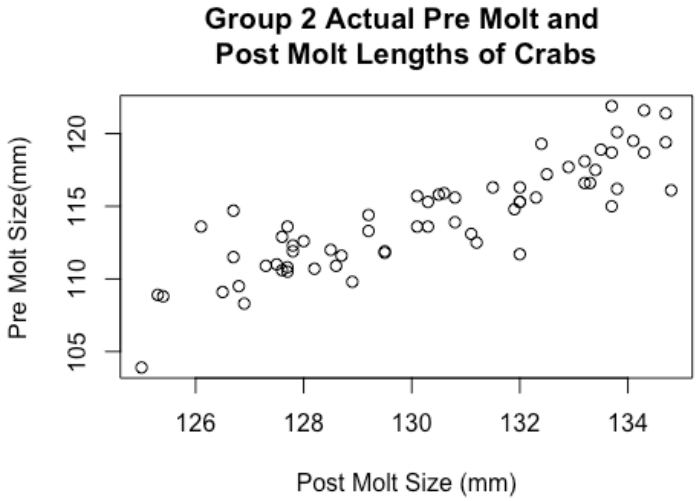


Figure 7: Scatter plot of the predicted pre molt length of Group 2 [125.0 to135.0] crabs using the Y = 1.07316X - 25.21370 relationship

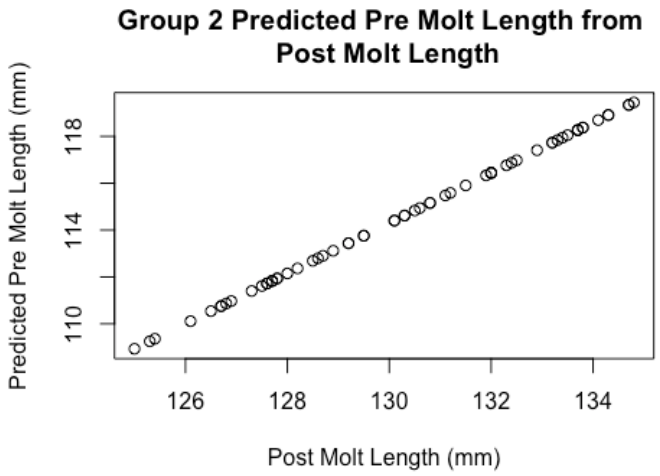
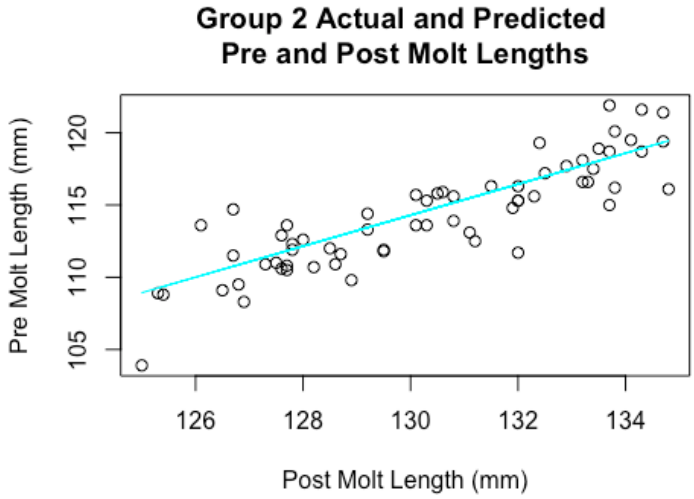
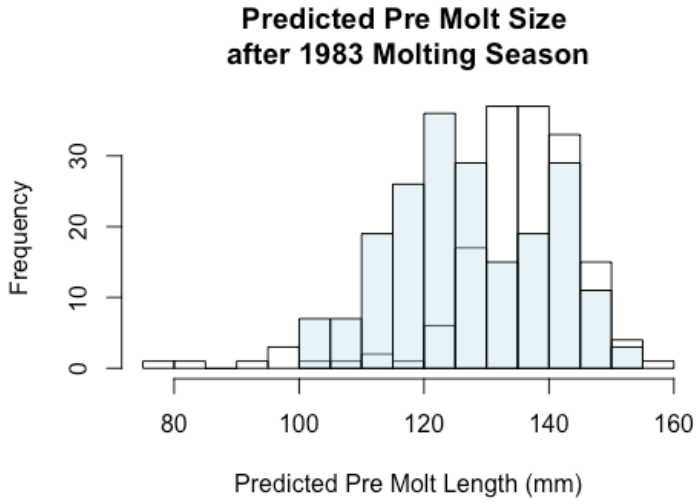


Figure 8: Combined scatter and line plots showing the actual pre molt data (scatter) and the predicted pre molt data (blue line) versus the post molt length for Group 2 [125.0 to 135.0 mm] crabs



The final analysis performed was conducted on the second set of data of post molt lengths of 362 female crabs following the May 1983 molting season. The pre molt lengths of these female crabs were calculated using the linear regression model calculated in the initial analysis above. The frequency distribution of the predicted pre molt lengths of crabs which both did and did not molt in that season is illustrated in Figure 9 below. It was determined firstly that the average pre molt lengths of crabs which did not molt (white bars) was greater than those which did molt (blue bars). Additionally, it was determined that there was a greater overall frequency of crabs which did molt (201 crabs) compared to those which did not molt (161 crabs). Finally, it can be said that both of the distributions of both crabs which did and did not molt were fairly left skewed suggesting that there is a greater frequency of crabs with higher pre molt lengths than those with smaller pre molt lengths.

Figure 9: Frequency distribution of predicted pre molt lengths of Dungeness crabs which molted (blue) and did not molt (white) following the 1983 molting season.



Discussion and Conclusions

The purpose of this investigation was to determine a relationship between the post molt size and the pre molt size of Dungeness crabs caught off of the Pacific Coast of the United States Using linear regression analysis, it was determined that the general relationship between these two variables was Y = 1.07316X - 25.21370 where Y is the pre molt size and X is the post molt size. It was found that this relationship was fairly accurate when the predicted and actual pre molt sizes of different groups of crabs based on their post length sizes were compared. The relationship was a closer fit for the second group (125.0 mm to 135.0 mm) compared to the first group (147.5 to 152.5) based on the higher average residuals. When this relationship was applied to crabs caught following the May 1983 molting season, the prediction showed that the crabs which did not molt during that season were larger overall in comparison to those which did molt that season. Additionally, it was found that there were a greater total number of crabs which had molted in comparison to those which had not molted. Overall, this model would be beneficial to policy makers as it would allow for them to predict the pre molt size fairly accurately in order to determine the actual size restrictions for fishing female crabs.

In the future, it will be critical to continue to collect data from additional seasons in order to gain a stronger idea both of the timeline of crab growth but also on potential molting and breeding yearly behaviors. One set back of this particular dataset was that some of the data was collected in 1992 when the rest was collected in the early 1980's. In future studies, it will be imperative to ensure that the data collected is from consecutive molting seasons in order to minimize the external influences such as environmental change which may have an impact on the length of the crabs.

Overall, the relationship determined in this investigation will prove helpful as an initial estimate for pre molt size based on post molt length in order to aid stakeholders in creating female crab fishing restrictions. In the future, studies conducted over longer periods of consecutive years will be beneficial in continuing to develop the most beneficial restrictions for this fisheries industry.

Works Cited

Costello, C., Ovando, D., Hilborn, R., Gaines, S. D., Deschenes, O., & Lester, S. E. (2012). Status and Solutions for the Worlds Unassessed Fisheries. *Science*, *338*(6106), 517–520. doi: 10.1126/science.1223389

Nolan, D., & Speed, T. (2000). Stat Labs: Mathematical Statistics Through Applications. New York: Springer.