# Project Work Statement

Participants

Xiaohan Yang, `xyang39@jhu.edu`

Potential Participants

Bruce Buckson, `pr.webmaster@noaa.gov`

Tracy Dunn, `habitat.conservation@noaa.gov`

Date: October 16, 2012

# 1 Background

National Oceanic and Atmospheric Administration's National Marine Fisheries Service is a federal agency and a division of the Department of Commerce. It is responsible for the management, conservation and protection of living marine resources and their habitats within the United States' Exclusive Economic Zone (water three to 200 mile offshore). It assesses and predicts the status of fish stocks, makes sure that people obey fisheries regulations, and works to reduce wasteful fishing practices. Under the law of fishing and living things protection, it recovers protected marine species (i.e. whales and turtles) without unnecessarily preventing economic and recreational opportunities. It has six regional offices and eight councils, which help the National Marine Fisheries Service to work with communities on fishery management issues. It endeavors to promote sustainable fisheries and to prevent overfishing to avoid the lost of economic potential, declining of species and damage of habitats. Also, it makes its best to balance competing public needs [2].

# 2 Problem Statement

Dungeness crabs are a kind of species of crab that are fished along the Pacific coast of North America during December and June. People only fish male crabs and female crabs are left to make sure that the crabs can reproduce. But this leads to great fluctuations in the catches of crabs. During December and June, there are a lot of male crabs being caught while in other months of the year, there may be none. Also, it is thought that the large ratio of female to male crabs may result in the decrease of crab population along the central California coast. So National Marine Fisheries Service is considering about changing the law so that female Dungeness crabs can also be fished.

As a result, size restrictions on female crabs have to be set so that crabs will not be caught if they are too young. For male crabs, size restrictions are made so that they can mate at least once before being fished so that the species can reproduce. It is similar for the restrictions on female crabs. Crabs mate in April and May when the female crabs molt. Molting is a kind of activity that crabs and many other species have during which the animals cast off their shells and grow new and larger shells so that their bodies can grow. So the growth of female crabs needs to be studied. More specifically, biologists from National Marine Fisheries Service would like to know how they can predict the size of a female crab before molting given the size after molting [1].

My project is to find a good way to predict the changes of shell sizes given the sizes after molting. In another way of saying it, I'm going to provide a way that

can predict the sizes of shells before molting according to the sizes after molting. There are a lot of factors that influence the changes in sizes, such as the year when the crabs molt, the place where they molt and the types of the shells, but I will only take into consideration the place where they molt. By specifying where the crabs molt, I will explore how the shell sizes change given different postmolt sizes of shells.

After knowing how to predict premolt sizes of shells from postmolt sizes, the biologists from National Marine Fisheries Service will be able to find a proper size of the nets so that the female crabs caught by the nets have already mated at least once.

# 3  Approach

The data used in my project were collected by scientists and commercial fisheries from northern California and southern Oregon and they were obtained over three fishing seasons: Year 1981, 1982 and 1992. Among these data, there are two kinds with respect to the way how they were got: laboratory data and capture-recapture data. The laboratory data were collected purely in the lab. Female crabs were caught just before molting and were fed in the lab. The premolt shell width was measured when the crab was caught, and the postmolt size was measured three to four days after the crab developed the new shell. The whole process of molting was made in the lab. The capture-recapture data were collected by tagging 12,000 crabs. These crabs were caught, measured and tagged with an identification number and then returned to the water. All these procedures were done before the molting season. After the molting season, crabs were fished and those with tags were brought to lab for second measurements. The whole process of molting was done in the nature [1].

In this way, a large sample of premolt and postmolt shells sizes was got corresponding with the time when the data were got and how they were measured.

The data for this project were collected for a study of the adult female Dungeness crab instead of my project. So they are secondary data and need some adjustments. We pair the premolt sizes with the postmolt sizes and specify the year when they were obtained and whether they are laboratory data or capture-recapture data. Then we check the data and abandon those missing premolt size or postmolt size or failing to specify whether they are laboratory or capture-recapture data. There are some data without time, but we will still use them because time is beyond our consideration. Also, we will delete the outliers by drawing a graph with postmolt size as the horizontal axis and premolt size as the vertical axis. After these adjustments, there are 472 data in our sample with three quarters of them laboratory

data. Part of the data are shown in Figure 1.

| Premolt Size | Postmolt Size | Year | Type of Data |
|---|---|---|---|
| 113.6 | 127.7 | NA | 0 |
| 142.3 | 154.8 | 81 | 1 |
| 120.3 | 139.3 | 82 | 1 |
| 111.8 | 129.5 | 92 | 1 |

Figure 1: Pattern of Data after Adjustments

Then we will focus on how premolt size changes with postmolt size for laboratory data and capture-recapture data.

I will use the software R to analysis the data and build the statistical model. To begin with, I will first build a linear regression model with the postmolt size and the type of data (laboratory or capture-recapture data) as the independent variables and premolt size as the dependent variable. The type of data is categorical data. 0 represents capture-recapture data and 1 represents laboratory data. Then I will find the coefficients of the two independent variables and test how significant they are to see whether the two independent variables should be included in the more. Then I will get an expression in the following form:

$$Premolt\ Size = \beta_0 + \beta_1(Postmolt\ Size) + \beta_2(Type\ of\ Data).$$

I will also use other methods to decide whether I should include the type of the data in the model.

After finishing building this simplest linear model, I will check how the model performs to predict the dependent variable according to some criteria. If it performs well, it may be a good time to stop. But if not, I will consider about some other complex forms of the regression model to see how they work. For example, I may consider the form that includes the cross term:

$$\begin{aligned} Premolt\ Size = \ & \beta_0 + \beta_1(Postmolt\ Size) + \beta_2(Type\ of\ Data) \\ & + \beta_3(Postmolt\ Size) * (Type\ of\ Data), \end{aligned}$$

or the form containing the squared term:

$$\begin{aligned} Premolt\ Size = \ & \beta_0 + \beta_1(Postmolt\ Size) + \beta_1(Postmolt\ Size)^2 \\ & + \beta_3(Type\ of\ Data). \end{aligned}$$

Stop when the performance of the model is satisfactory.

# 4 Milestones

We have the following major deadlines:

- Work Statement due date, Sep 28, 2012,

- Midterm Presentation due date, Oct 12, 2012,

- Progress Report due date, Oct 26, 2012,

- Final Presentation due date, Nov 6, 2012,

- Final Report due date, Nov 30, 2012.

# 5 Deliverable

## 5.1 From Team to Sponsor

The following outputs are expected from this project:

- A regression formula predicting premolt sizes from postmolt sizes of female Dungeness crabs taken the place of molting into consideration

- A R programming package which can predict the premolt sizes of shells and produce a graph for the distribution of premolt sizes when people input the data of postmolt sizes of shells and the place of molting in the program

## 5.2 From Sponsor to Team

In order for our project to be of successful one, we will need:

- Data of shell sizes before molting and corresponding postmolt sizes specifying the molting environment and measurement year. The data should be delivered to me by Oct 12, 2012. If my sponsor fails to do so, I will only develop a method for the prediction problem without an explicit regression formula.

- Computing resources

- Timely responses to inquiries,

- Symposium attendance travel expenses.

# References

[1] Terry Speed Deborah Nolan. *Stat Labs: Mathematical Statistics Through Applications*. Springer, 2000.

[2] National Oceanic And Atmospheric Administration. *About National Marine Fisheries Service*. `http://www.nmfs.noaa.gov/aboutus/aboutus.html`. Accessed: September 30, 2012.