

Predictive Linear Models for Post-Molt to Pre-Molt Dungeness Crab Shell Sizes

James P. Donohue

February 13, 2022

Abstract

This study is aimed to determine the correlation between the Dungeness crab's post-molt shell size and its pre-molt size, whose linear model can be used to predict any pre-molt size. To test our hypothesis, a dataset was constructed using a 1989 study's sample observations and data description from 472 adult female Dungeness crabs collected in 1981, 1982, and 1992 (Hankin et al. [HDMI89]). While observing the residual plots for all three sub-groups of adult female Dungeness crabs, it can be 'eyeballed' that all three linear models are homoscedastic. It can still be noted that the sea-based crabshell residuals are visually dissimilar to the other two groups. Although this is not extreme enough to be considered heteroscedastic. It was demonstrated for all adult female Dungeness crabs with a post-molt shell size greater than 120mm that all residuals closely followed a gaussian distribution. It can be safely concluded that our linear models are accurate enough to make reliable backwards predictions of Dungeness Crab shell sizes.

Background and Significance

The Dungeness crab, *Metacarcinus magister*, is a species of crab that inhabits waters off the west coast of North America. It typically grows to 200 mm across the shell. Dungeness crabs have a wide, long, hard shell, which they must periodically molt to grow. The legal size for a Dungeness crab is larger than 165mm. Generally, fishermen when deciding to fish for these crabs will do a quick score of the beach checking the size of molts that lie there. This enables them to judge the probable size of crabs in the area, preventing wasted time throwing back undersized catches. Unless a large-scale investigation is undertaken, molts are generally found in groups of similar size ranging no more than 5mm difference. The goal of this study is to determine the correlation between the Dungeness crab's post-molt shell size and its pre-molt size, whose linear model can be used to predict any pre-molt size. The hypothesis of this study is that the size of a Dungeness crab's previous shell can be predicted using the size of their shell post-molt with reasonable accuracy.

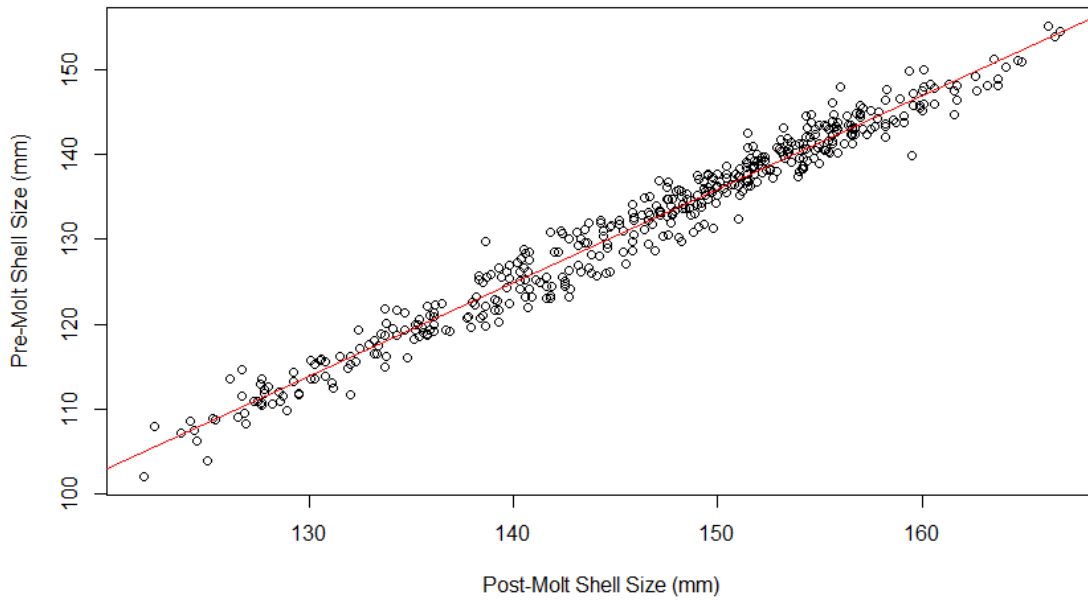
Methods

To test our hypothesis, a dataset was constructed using a 1989 study's sample observations and data description from 472 adult female Dungeness crabs collected in 1981, 1982, and 1992 (Hankin et al. [HDMI89]). The data contains five variables: 'presz,' representing the size of an adult female Dungeness crab before molting, 'postsz,' representing the size of an adult female Dungeness crab after molting, 'inc,' representing the disparity of size between the pre and post-molt shells, 'year,' representing the year a datapoint was collected, and lastly 'lf,' which represents whether or not the Dungeness crab was caught live in the ocean or hatched in a laboratory. For our linear model, our independent variable will be post-molt shell size, so naturally we consider 'postsz' our X-variable, and 'presz' our Y-variable. Next, the dungeness crabs will be divided into three distinct sub-groups: a sub-group of all crabs whose shell size is above 120mm (to account for outliers), and two further divided sub-groups: ocean-caught, and lab-grown. For each group we then will construct a linear regression model which provides for us the equation for our line-of-best-fit. For each line-of-best-fit, we then construct and subsequently plot their residuals. Through observation, we use our residual plots to determine the scedasticity of each group's regression model. If each group is observed to be homoscedastic, we can safely conclude our models' accuracy.

Results

The linear regression model of all crab shell sizes greater than 120mm generates coefficients which allow us to construct a line-of-best-fit of $y = 1.10x - 29.19$, where x is our post-molt shell size, and our y is our predicted pre-molt shell size.

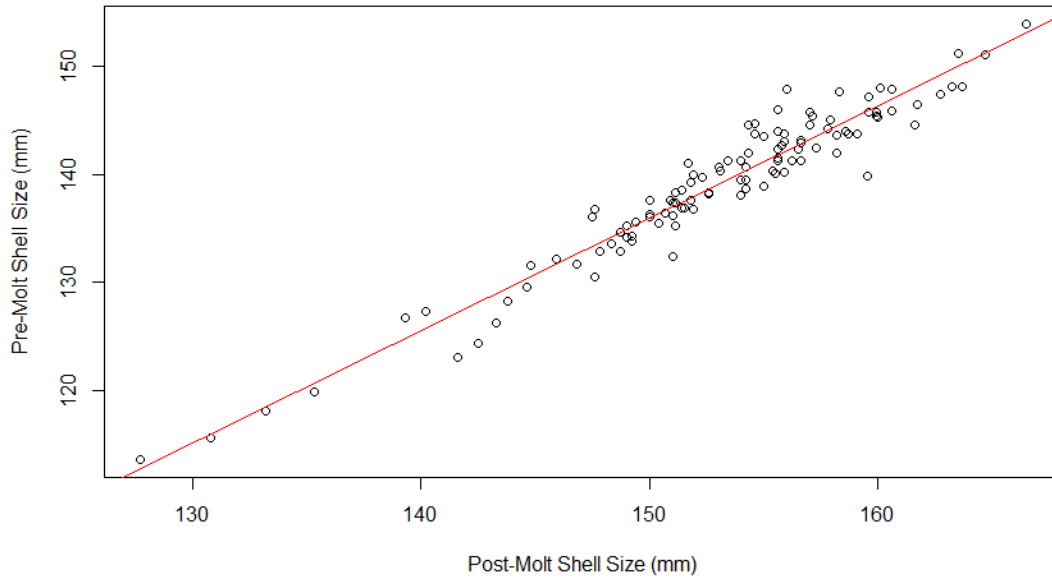
Linear Regression Model of Crabshell Sizes



Post-Molt (mm)	127.7	133.2	154.8	142.5	120.0	134.1	133.8
Pre-Molt (mm)	113.6	118.1	142.3	125.1	98.2	119.5	116.2
Source	Ocean	Ocean	Lab	Lab	Lab	Lab	Lab
Predicted Pre-Molt (mm)	111.28	117.33	141.09	127.56	102.81	118.32	117.99
Difference (mm)	2.32	0.77	1.21	2.46	4.61	1.18	1.79

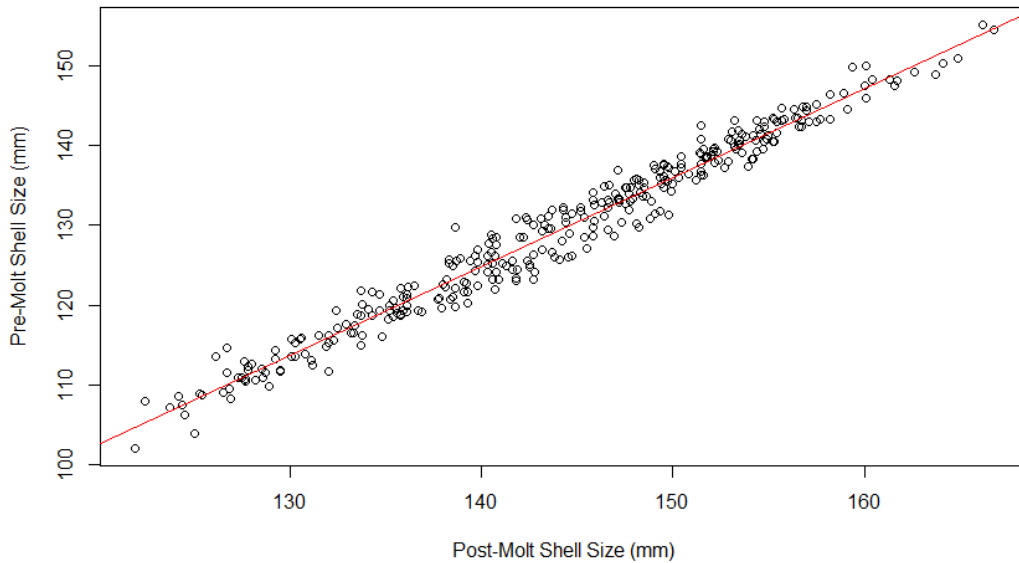
The linear regression model of all sea-caught crab shell sizes greater than 120mm generates coefficients which allow us to construct a line-of-best-fit of $y = 1.042x - 20.402$.

Linear Regression Model of Sea-based Crab Shell Sizes



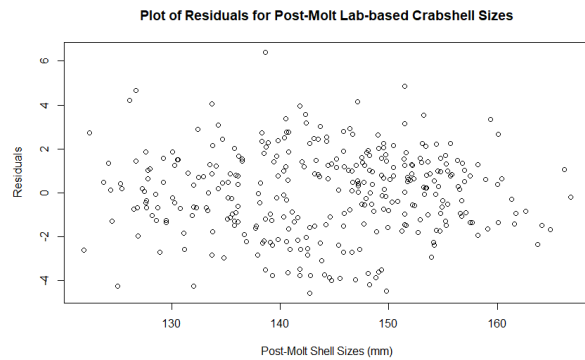
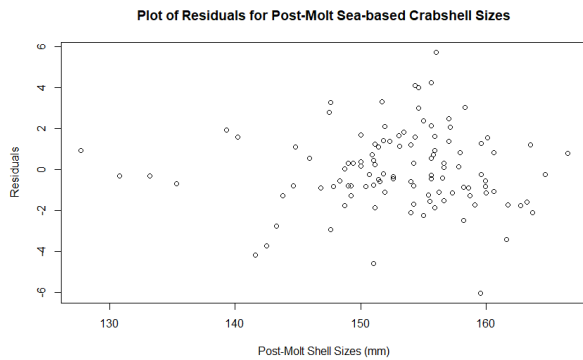
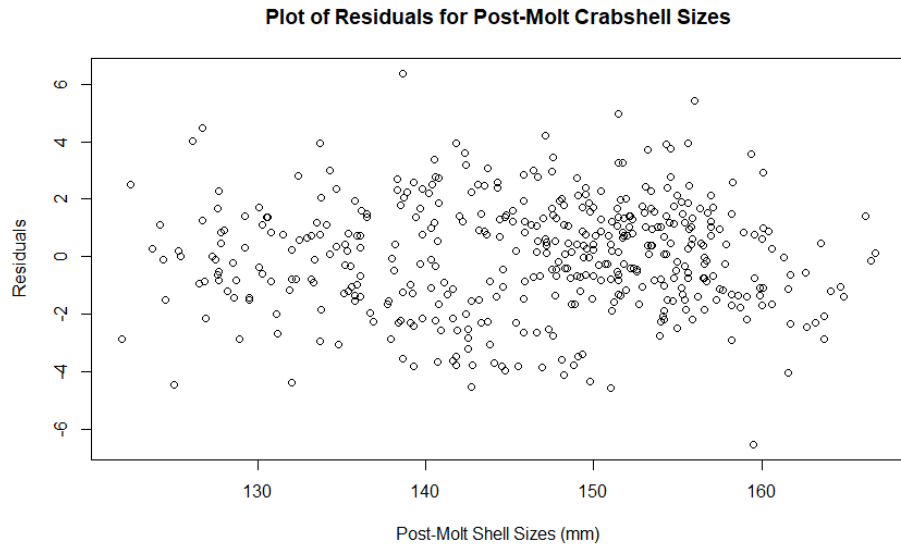
The linear regression model of all lab-hatched crab shell sizes greater than 120mm generates coefficients which allow us to construct a line-of-best-fit of $y = 1.113x - 31.025$.

Linear Regression Model of Lab-based Crab Shell Sizes

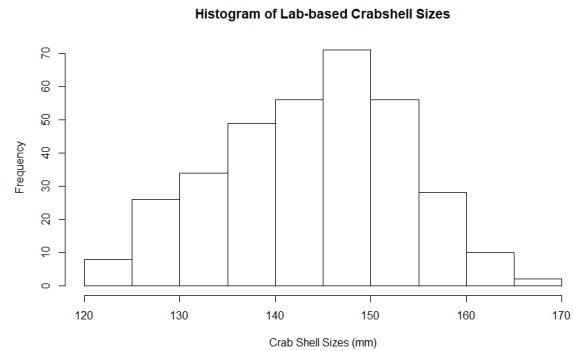
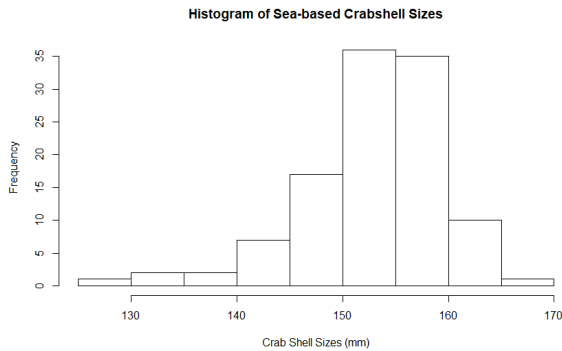
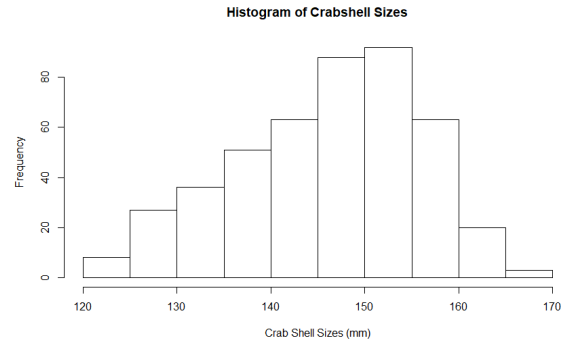


While observing the residual plots for all three sub-groups of adult female Dungeness crabs, it can be 'eyeballed' that all three linear models are homoscedastic. It can still be noted that the sea-based crabshell

residuals are visually dissimilar to the other two groups. Although this is not extreme enough to be considered heteroscedastic.



The histograms of all three groups reveals that there is a noticeably larger skew in the distribution of sea-based crabshell sizes than the other groups. Sea-based crabshell sizes appear to also peak at a much higher size than the other groups. The general crabshell sizes and lab-based crab-shell sizes appear to be much closer to normal distributions than that of the sea-based crabshell sizes as well.



This idea is supported by all three groups' residuals having a kurtosis at or around 3, save for sea-based crabshell sizes for which a kurtosis of 4 gives way for a visual difference.

[table of descriptive statistics goes here]

Conclusion

The goal of this study was to determine the correlation between the Dungeness crab's post-molt shell size and its pre-molt size, whose linear model can be used to predict any pre-molt size. It was demonstrated for all adult female Dungeness crabs with a post-molt shell size greater than 120mm that all residuals closely followed a gaussian distribution. All residual plots appeared to follow a homoscedastic model. Among all groups, it appeared the aggregate/general crab group and lab-hatched crab group followed very similar residual distributions. However, the sea-caught dungeness crabs followed a less consistent distribution, higher kurtosis, and larger peak size, the far latter of which may be causal of these apparent differences. Nonetheless it can be safely concluded that our linear models are accurate enough to make reliable backwards predictions of Dungeness Crab shell sizes. Perhaps the controlled environment of the lab removed any aggravating variables which would impact the molting process, encouraging greater growth? Perhaps the lack of predators impacted the amount of crabs of smaller recorded size? Perhaps a

study looking further into these aggravating variables that are absent in the lab environment could benefit future studies.

Appendix

Handwritten mathematical derivations on lined paper:

$$\begin{matrix} ax & a^2 & abx \\ +b & abx & b^2 \end{matrix}$$

$$(y - (ax + b))$$

$$y^2 + (ax + b)^2 - 2y(ax + b)$$

$$y^2 + ax^2 + 2bx + b^2 - 2y(ax + b)$$

$$y^2 + ax^2 + 2abx + b^2 - 2yax - 2yb$$

w/ repeat to a

$$2x^2a + 2bx - 2yx$$

w/ repeat to b

$$2ax + 2b - 2y$$

$$2x^2a + 2bx - 2yx = 0$$

$$2ax + 2b - 2y = 0$$

$$x^2a + bx - yx = 0$$

$$ax + b - y = 0$$

$$b = y - xa$$

relates a & b

Descriptive statistics:

//////////////////////FOR RESIDUALS

```

> mean(residSubCrabs)
[1] 7.039934e-17
> median(residSubCrabs)
[1] 0.1057197
> sd(residSubCrabs)
[1] 1.901232
> skewness(residSubCrabs)

```

```
[1] -0.08152442
> kurtosis(residSubCrabs)
[1] 3.098374
> summary(residSubCrabs)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
-6.5264 -1.2921  0.1057  0.0000  1.3125  6.3719
```

```
mean(residSeaCrabs)
[1] 4.719698e-18
> median(residSeaCrabs)
[1] -0.2496302
> sd(residSeaCrabs)
[1] 1.880067
> skewness(residSeaCrabs)
[1] 0.03555623
> kurtosis(residSeaCrabs)
[1] 3.912915
> summary(residSeaCrabs)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
-6.0205 -1.0908 -0.2496  0.0000  1.2111  5.7270
```

```
mean(residLabCrabs)
[1] 1.157609e-17
> median(residLabCrabs)
[1] 0.1753904
> sd(residLabCrabs)
[1] 1.891766
> skewness(residLabCrabs)
[1] -0.07855221
> kurtosis(residLabCrabs)
[1] 2.931417
> summary(residLabCrabs)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
-4.5654 -1.2878  0.1754  0.0000  1.2767  6.3998
```

//////////////////////////////////FOR POST MOLT

```
mean(subCrabsX)
[1] 146.1051
> median(subCrabsX)
[1] 147.8
> sd(subCrabsX)
```



```
[1] 9.857574
> skewness(subCrabsX)
[1] -0.3894415
> kurtosis(subCrabsX)
[1] 2.388866
> summary(subCrabsX)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
121.9  139.2  147.8  146.1  153.8  166.8
```

```
mean(seaCrabsX)
[1] 152.964
> median(seaCrabsX)
[1] 154
> sd(seaCrabsX)
[1] 6.719967
> skewness(seaCrabsX)
[1] -1.119064
> kurtosis(seaCrabsX)
[1] 5.240706
> summary(seaCrabsX)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
127.7  150.0  154.0  153.0  157.0  166.5
```

```
> mean(labCrabsX)
[1] 143.8659
> median(labCrabsX)
[1] 144.55
> sd(labCrabsX)
[1] 9.690375
> skewness(labCrabsX)
[1] -0.1651182
> kurtosis(labCrabsX)
[1] 2.319648
> summary(labCrabsX)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
121.9  136.2  144.6  143.9  151.5  166.8
```

```
//////////////////////////////////FOR PRE MOLT
```

```
> mean(subCrabsY)
[1] 131.5867
> median(subCrabsY)
[1] 133.3
> sd(subCrabsY)
```

```
[1] 11.01258
> skewness(subCrabsY)
[1] -0.402511
> kurtosis(subCrabsY)
[1] 2.325261
> summary(subCrabsY)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
102.1  123.3  133.3  131.6  140.3  155.1
```

```
> mean(seaCrabsY)
[1] 139.009
> median(seaCrabsY)
[1] 140.1
> sd(seaCrabsY)
[1] 7.251151
> skewness(seaCrabsY)
[1] -1.110875
> kurtosis(seaCrabsY)
[1] 4.761443
> summary(seaCrabsY)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
113.6  136.1  140.1  139.0  143.8  153.9
```

```
> mean(labCrabsY)
[1] 129.1635
> median(labCrabsY)
[1] 130.2
> sd(labCrabsY)
[1] 10.95437
> skewness(labCrabsY)
[1] -0.1548048
> kurtosis(labCrabsY)
[1] 2.247076
> summary(labCrabsY)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
102.1  120.9  130.2  129.2  137.7  155.1
```