

# Analysis on Crab's Growth

## Post-Molt Vs Pre-Molt Dungeness Crab Growth

### Issues

Dungeness crabs are found in Pacific coast of North America, crabs are commercially fished between December and June. Fishing mainly focused on Male crabs, female crabs are not fished in order to maintain the viability of the population, this results to the imbalance in sex ratio of crabs. In order to overcome the issue, Size restrictions on male crabs are set to ensure that they at least one opportunity to mate before being fished. To understand the age of crabs a specific information on molting is required.

We address the questions:

- What is the relation between 'Pre-Molt' and 'Post-Molt' size?
- How the linear regression model fit to 'Pre-Molt' and 'Post-Molt' size?

### Findings

From the data, I found that the Pre-Molt size and Post-Molt size are linear related and R-squared value 0.9821. The mean of Post-Molt size is 144.1 and mean of Pre-Molt size is 129.3. The Median of Pre-Molt is 148.1 and median of Post-Molt is 133.9.

### Discussions

R-squared value is 0.9821 which means the size of crab can be explained by the Pre-Molt and Post-Molt variables. Pre-Molt and Post-Molt are linear related with coefficient 0.9910042.

## Appendix A: Method

Data was downloaded as a comma separated (.csv) file and imported into R-studio. The file contains the two columns which are Dungeness crabs' sizes (Post-Molt and Pre-Molt).

First, Summary of the variables such as minimum, maximum, mean, median, 1<sup>st</sup> Quartile, 3<sup>rd</sup> Quartile are extracted. By using the Probability density function plotted the histogram for Pre-Molt and Post-Molt variables.

Smooth Histograms for each variable, overlaid and the difference in distribution is observed. Scatter plot is plotted for Pre-Molt and Post-Molt variables and observed how they are related to each other. Plotted the linear regression with Pre-Molt and Post-Molt and calculated the Pearson's  $r^2$  for the regression.

Calculated the descriptive statistics (Minimum, 1<sup>st</sup> Quartile, median, Mean, 3<sup>rd</sup> Quartile, Maximum) of the residuals. Tested the Shapiro-Walks test for residuals.

## Appendix B: Results

There are 485 data points in the .csv which is imported to the R-studio. Summary of variables Post-molt and Pre-molt is obtained and can find in figure 1.

Post-molt		Pre-molt	
Min.	: 38.8	Min.	: 31.1
1st Qu.	:136.9	1st Qu.	:120.9
Median	:148.1	Median	:133.9
Mean	:144.1	Mean	:129.3
3rd Qu.	:153.8	3rd Qu.	:139.8
Max.	:166.8	Max.	:154.5

Figure 1. Summary of Post-molt and Pre-molt variables.

Standard deviation of Post-molt: 15.11685

Standard deviation of Pre-molt: 16.45753

Skewness of Post-molt: -2.268228

Skewness of Pre-molt: -1.979148

Kurtosis of Post-molt: 12.10619

Kurtosis of Pre-molt: 9.210475

Figure 2 shows the probability density function (PDF) histogram of Post-molt, it is left skewed and bell shaped with peak at 148.1, Figure 3 shows the probability density function (PDF) histogram of Pre-molt, it is left skewed and bell shaped with peak at 133.9.

Figure 4 shows the smooth histograms for each variable overlaid so the difference in distribution is visible. Difference in the size of crabs is clearly visible after molting.

Scatter plot is plotted between the Post-molt and Pre-molt, Figure 5 shows the plot. This plot shows that they are positively correlated to each other which means one variable increases the other variable also increases. Figure 6 shows that line has positive slope which means that one variable rises then other variable also rises. Pearson's  $r^2$  value is calculated 0.9910042.

Figure 7 shows the Quantile plot, in which the residuals are in a straight line that tells Pre-molt and Post-molt variables are normally distributed. Figure 8 shows the plot of residuals against the dependent variable.

Figure 9 shows the heteroscedasticity in the residuals.

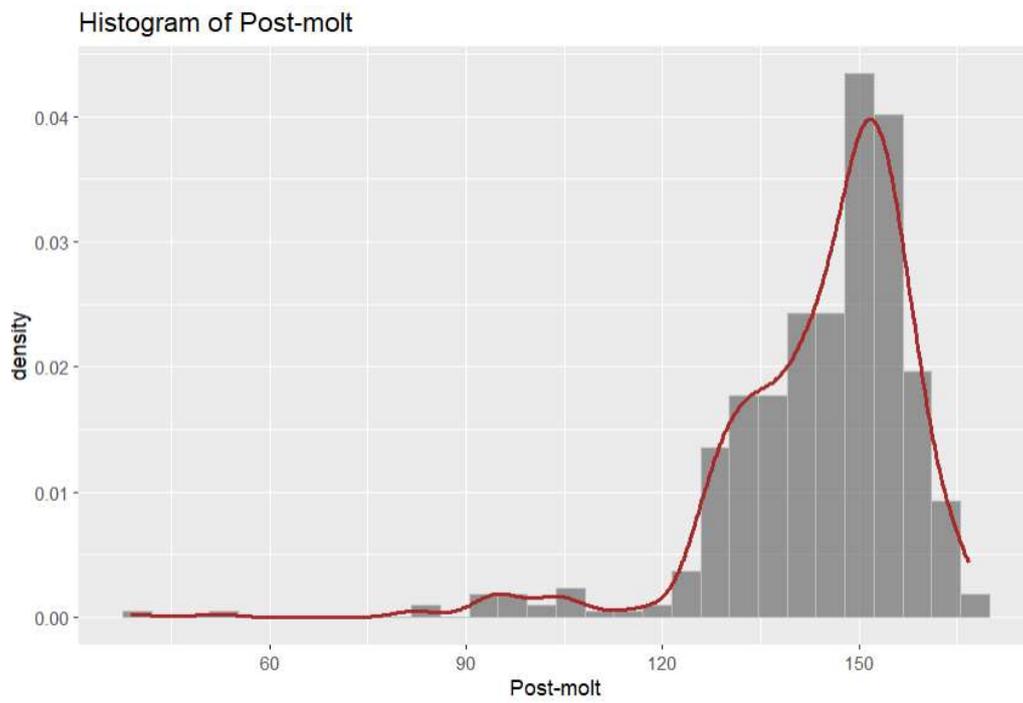


Figure 2 Histogram of Post-molt

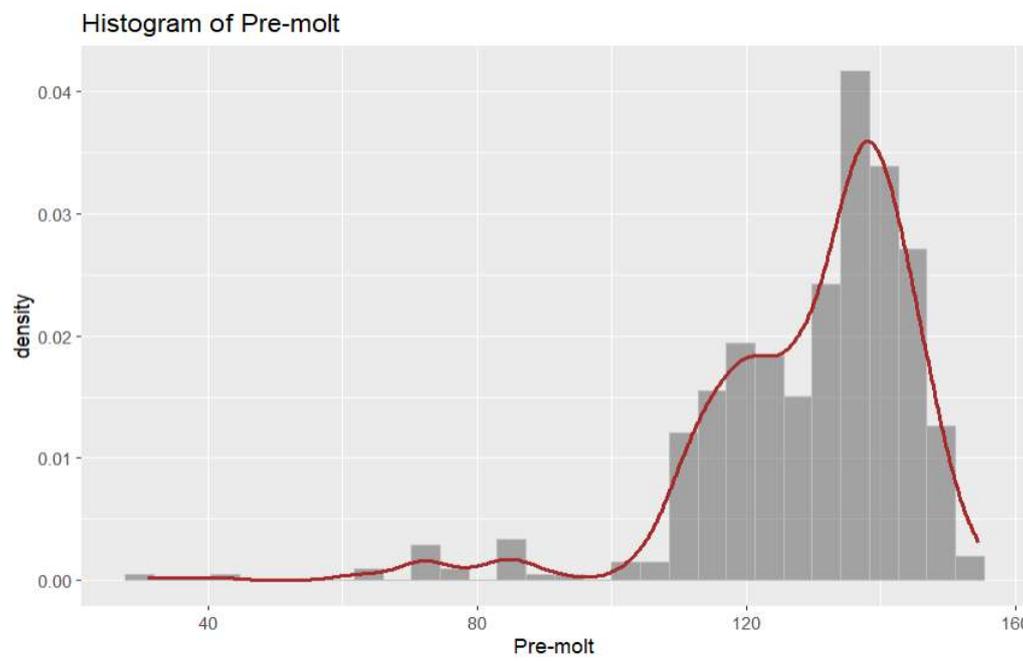


Figure 3: Histogram of Pre-molt

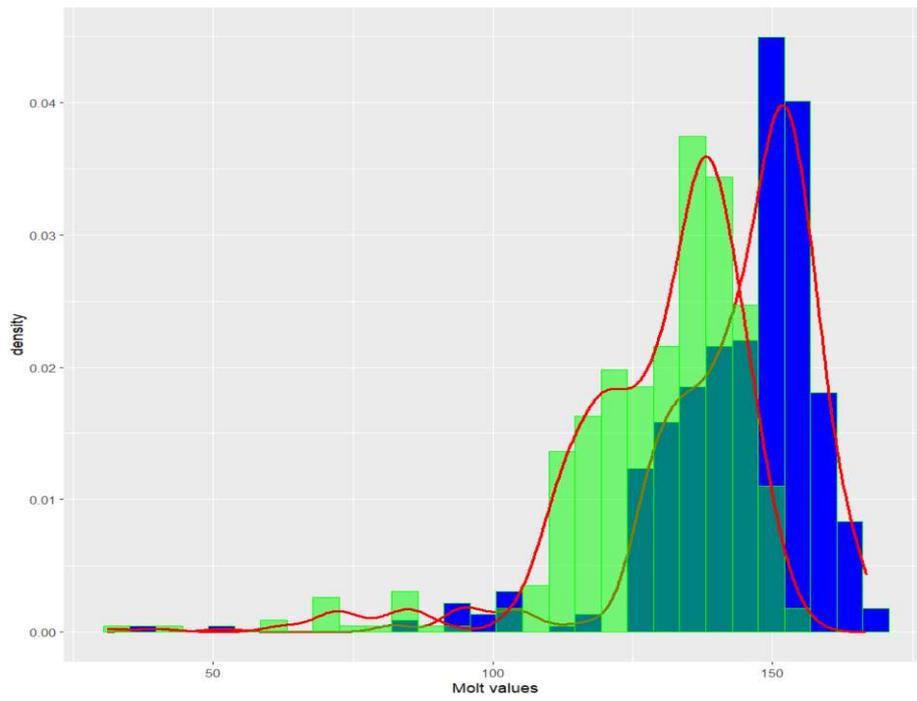


Figure 4: Overlaying of both histograms

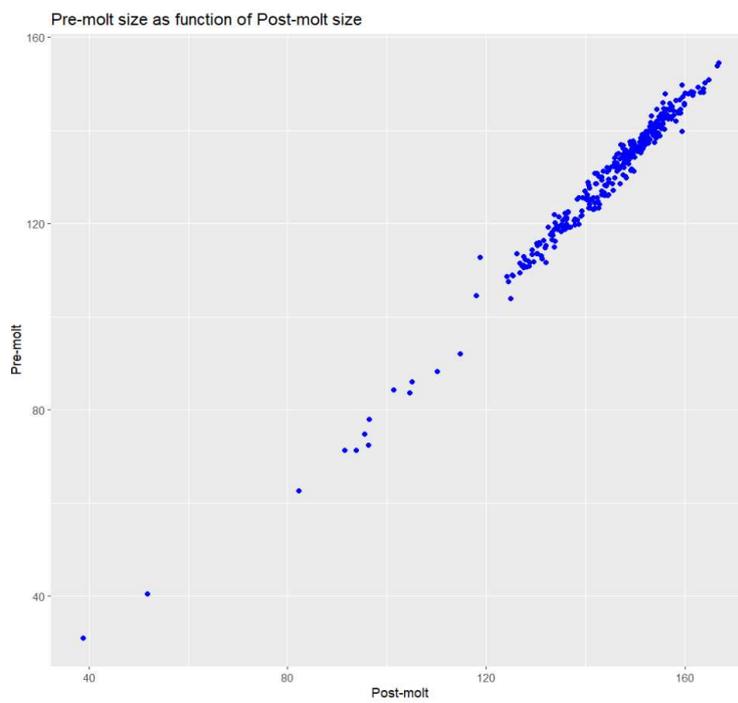


Figure 5: Scatter plot between the Post-molt and Pre-molt

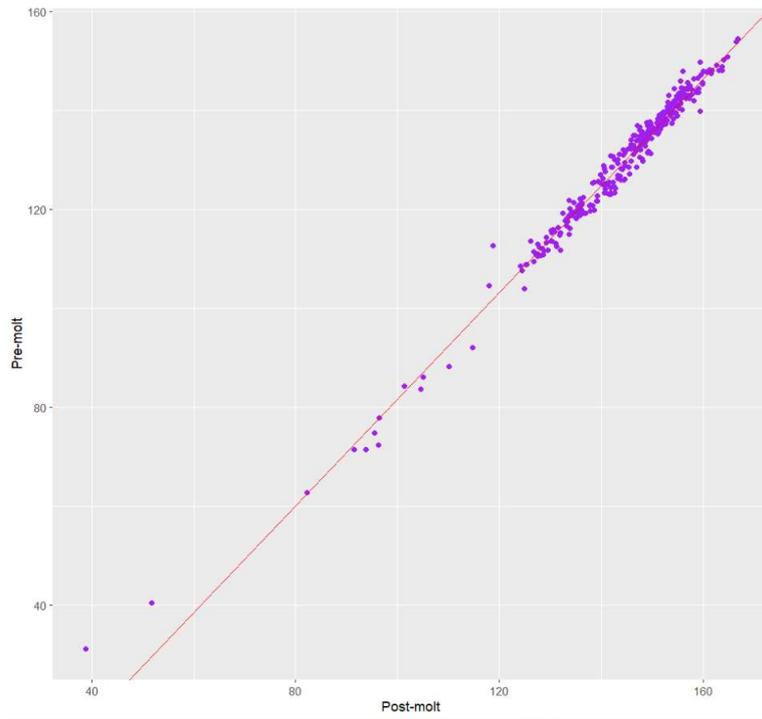


Figure 6: Regression line

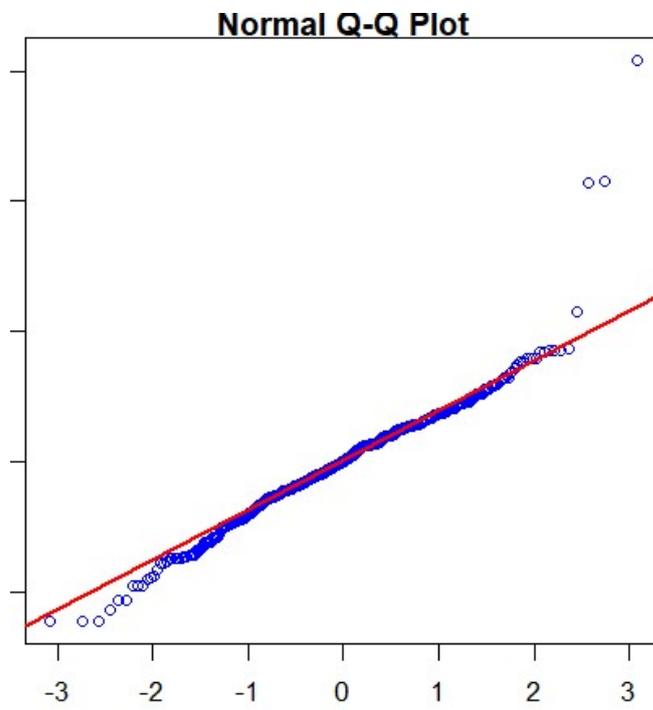


Figure 7: Quantile plot

**Residuals against the dependent variable**

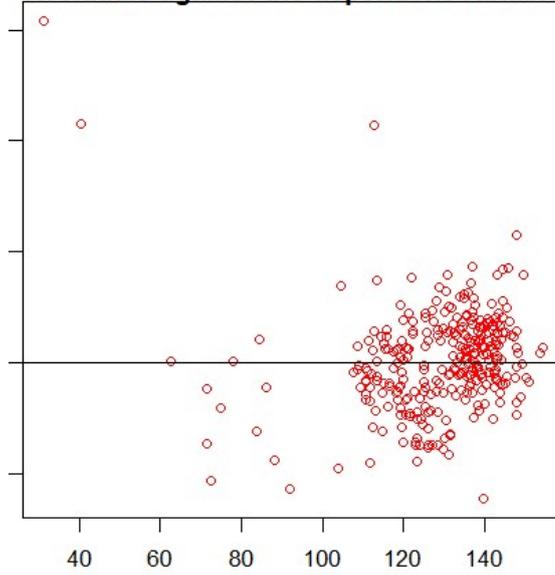


Figure 8 Residuals against the dependent variable

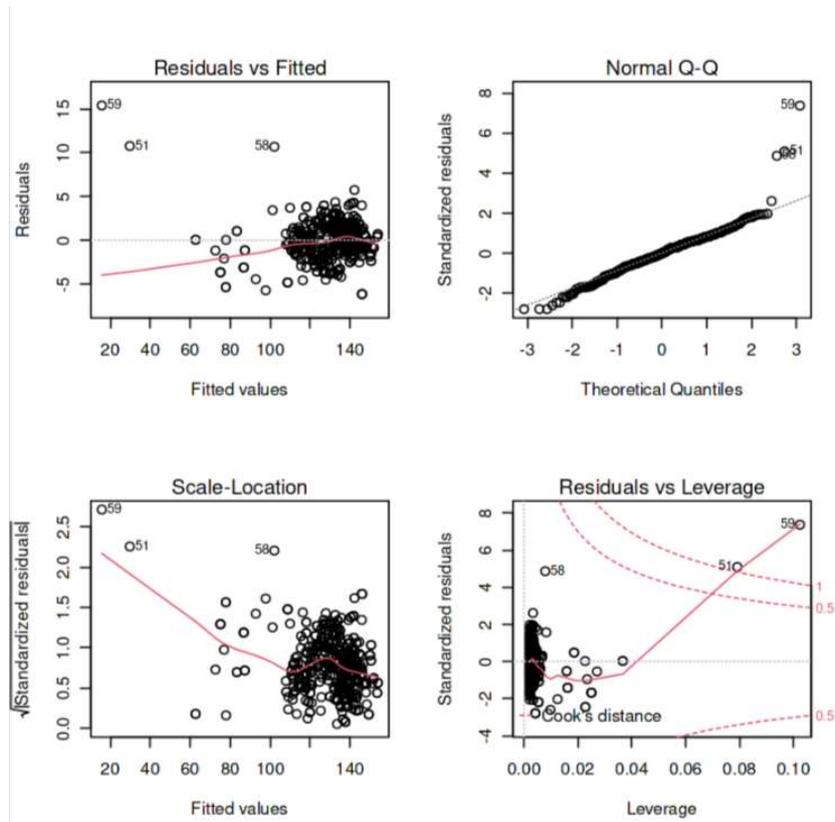


Figure 9: Heteroscedasticity in the residuals.

## Appendix C: Code

```
install.packages('readxl')
library(readxl)

file <- "E:\\Assignments\\MTH 522\\crab_molt_data_muchinapalli_nagendra.xls"
crab <- read_excel(file, sheet = 1)

#summary of variables
summary(crab)

#standard deviation of Post-molt
sd(crab$`Post-molt`)

#standard deviation of Pre-molt
sd(crab$`Pre-molt`)

install.packages("moments")
library(moments)

# Skewness of pre-molt and pre-molt
skewness(crab$`Post-molt`)
skewness(crab$`Pre-molt`)

#Kurtosis of post-molt and Pre-molt
kurtosis(crab$`Post-molt`)
kurtosis(crab$`Pre-molt`)
```

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
install.packages("dplyr")
```

```
library(dplyr)
```

```
#Probability density function of each variable
```

```
crab %>% ggplot(aes(x=`Post-molt`)) +  
  geom_histogram(aes(y=..density..),alpha = 0.6,col='grey') +  
  geom_density(aes(y=..density..),col = 'brown',lwd = 1) +  
  ggtitle("Histogram of Post-molt")
```

```
crab %>% ggplot(aes(x=`Pre-molt`)) +  
  geom_histogram(aes(y=..density..),alpha = 0.5,col='grey') +  
  geom_density(aes(y=..density..),col = 'brown',lwd = 1) +  
  ggtitle("Histogram of Pre-molt")
```

```
#overlaying of both histograms
```

```
crab %>% ggplot(aes(x=`Post-molt`)) +  
  geom_histogram(aes(y=..density..),color='green',fill='blue')+  
  geom_density(aes(y=..density..),color='red', lwd=1)+  
  geom_histogram(aes(x=`Pre-molt`,y=..density..),color='green', fill='green',alpha=0.5)+  
  geom_density(aes(x= `Pre-molt`,y=..density..),color='red', lty=1,lwd=1) +  
  labs(x="Molt values")  
ggtitle("Overlaying of both histograms ")
```

```
# Pre-molt size as function of Post-molt size
```

```
crab %>% ggplot(aes(x = `Post-molt`, y = `Pre-molt`)) +  
  geom_point(col = 'blue', pch = 19) +  
  ggtitle("Pre-molt size as function of Post-molt size")
```

```
#Simple linear Regression
```

```
linear_regression <- lm(crab$`Pre-molt` ~ crab$`Post-molt`)  
summary(linear_regression)  
crab %>% ggplot(aes(x = `Post-molt`, y = `Pre-molt`)) +  
  geom_point(col = 'purple') +  
  geom_abline(intercept = -26.16211, slope = 1.07889, col = 'red' )
```

```
# Pearson's calculation
```

```
pearson_cor <- cor(crab$`Post-molt`, crab$`Pre-molt`, method = "pearson")  
print(pearson_cor)
```

```
#statistics of the residuals and quantile
```

```
residuals <- resid(linear_regression)  
summary(residuals)
```

```
#Q-Q plot
```

```
qqnorm(residuals, col = 'blue')  
qqline(residuals, col = 'red', lwd = 2)
```

```
#Shapiro-Walks test.
```

```
shapiro.test(residuals)
```

```
#Plot the residuals against the dependent variable
```

```
plot(crab$`Pre-molt`, residuals, xlab="Pre-molt values" , ylab="residuals",col='red' , main  
="Residuals against the dependent variable")
```

```
abline(0,0)
```

```
#Heteroscedasticity
```

```
par(mfrow=c(2,2))
```

```
plot(linear_regression)
```