

Cracking the Code: Unraveling Size Dynamics Before and After Molting

Authors: Veda Sahaja Bandi, Sindhuja Baikadi, Mythri Krpel Jayaram

The Issues:

Crabs undergo a fascinating natural process called molting, where they shed their old shells and grow larger exoskeletons. To delve deeper into this transformation, a dataset was compiled that captures the sizes of crab shells in Washington state before and after molting. By analyzing this data, we aim to answer captivating questions:

- What disparities exist in the sizes of crab shells before and after molting?
- Is there a statistically significant difference in the average shell sizes before and after molting?
- How does the difference between the mean size before and after molting, contribute to understanding the impact of molting on crab shell sizes?

Delving into these questions can illuminate the impact of molting on individual crabs and entire populations, while also revealing potential ecological consequences of size shifts within broader ecosystems. By thoroughly examining these questions and possible constraints, we gain valuable insights into the intricacies of crab molting and how it impacts their growth and ecological functions.

Findings:

In the analysis of crab molting data from Washington State, it's evident that a notable transformation occurs in the shell sizes of crabs before and after molting. Specifically, crabs display considerable alterations in shell dimensions post-molting, averaging around 15mm. This phenomenon highlights a significant aspect of crab development and adaptation within their natural habitat. The effect size, used to assess the disparity between pre-molt and post-molt, is estimated to be 0.96, which is large according to the general rule of thumb initially proposed by Cohen and expanded upon by Sawilowsky. This indicates that molting has a significant and consistent impact on crab growth across the studied population.

Statistical analysis suggests the possibility that the observed pre- and post-molt size difference is simply due to random chance. The statistical significance of this finding is akin to the rarity of obtaining 23 consecutive heads from flipping a fair coin. Our analysis sheds light on the necessity of comprehending the molting process's implications for crab populations and ecosystems, stressing the importance of ongoing research and conservation endeavors to safeguard crab habitats and populations.

Discussion:

The large effect size observed in our analysis underscores the profound biological significance of the molting process for crab growth. This substantial effect size, combined with the statistical significance of the findings, emphasizes the critical role of molting in facilitating crab development and adaptation within their habitats. The consistent increase in size across the

population suggests that molting is a critical growth strategy for crabs, likely tied to survival, competitive advantage, and reproductive success.

Factors such as environmental conditions, nutritional status, and genetic predispositions may influence the extent of growth during molting. Future research could explore these factors to understand their impact on the variability of growth outcomes. Additionally, the implications of these findings for crab population dynamics, fisheries management, and conservation strategies warrant further investigation.

Appendix A: Method

Data was downloaded as a comma-separated (.csv) file and imported into Jupyter Notebook. Header rows and all data entries with null values were removed. The analysis was done using the pre-molt and post-molt data.

The analysis utilized smooth approximations to visually and quantitatively compare the crab size distributions before and after molting. Cohen's d effect size further clarified the magnitude of change, revealing the extent and variability of growth across the crab population. This approach provided valuable insights into the impact of molting on crab size and growth patterns.

The average size of pre-molt and post-molt of crab shells were calculated, to assess whether the observed changes in crab size were statistically significant, and a paired t-test was conducted. This test compares the mean size before and after molting within the same individuals, controlling for individual variability.

If the distributions of crab shells by size are substantially non-normal a Monte Carlo procedure is used instead to compute a p-value, being an estimate of the probability of seeing the difference in size we do see if the null hypothesis of no statistical difference was true.

Appendix B: Results

From a database of 476 data points, we have used 472 data points after removing null values containing both pre and post-molting sizes of crabs.

The distribution of the crabs after molting is larger than the distribution of the crabs before molting. This means that after molting, the size of the crab increases as compared with the size of the crabs before the molting process.

There are a few possible explanations for this. One possibility is that the crabs lose some of their body mass during molting. Another possibility is that the molting process is more difficult for larger crabs so only the smaller crabs can molt successfully.

Effect Size

A comparison of smooth histogram approximations to the size distribution of crabs before and after molting shows a significant shift to the left and elevation after molting indicating that the size of crabs increased after the molting process.

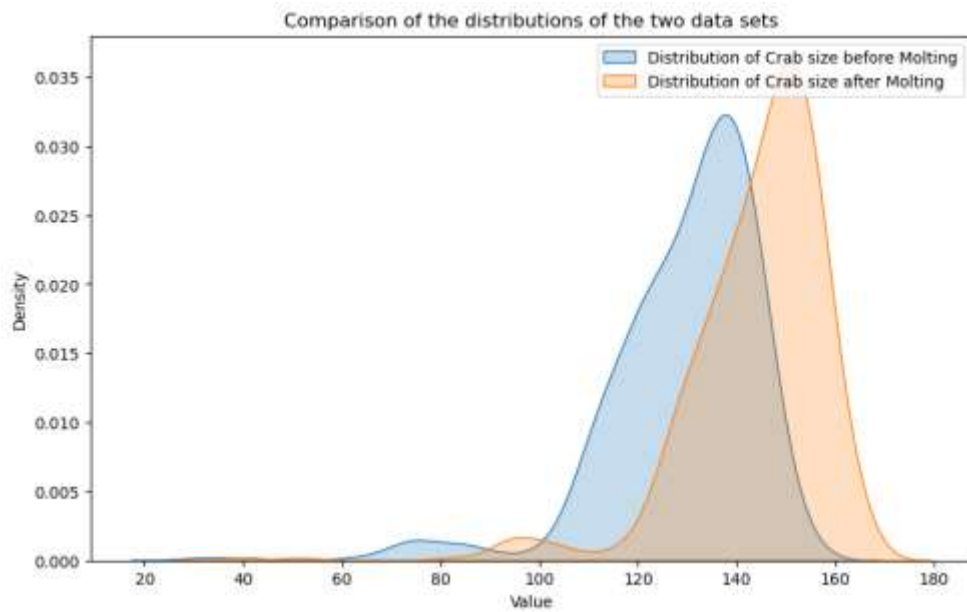


Figure 1: Crab Size distributions before and after molting

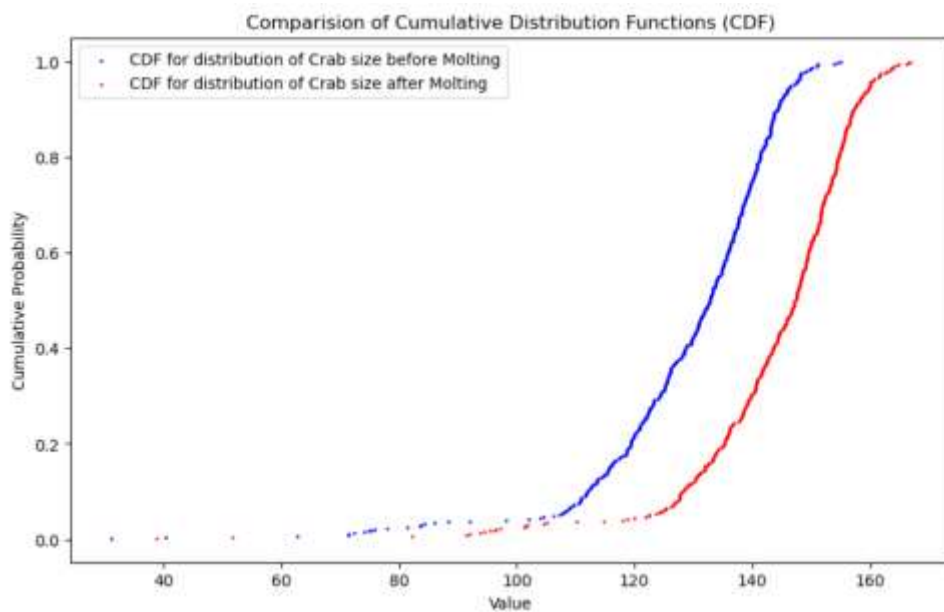


Figure 2: Comparison of Cumulative Distribution Functions (CDF) before and after molting

The calculated average size, rounded to two decimal places, for size before molting was 129.21 mm, while after molting it was 143.90 mm: a 14.69 mm average size difference.

Quantile plots for the distributions of sizes pre and post-molting suggest that molting affects the size distribution of crabs indicating relatively serious departures from normality:

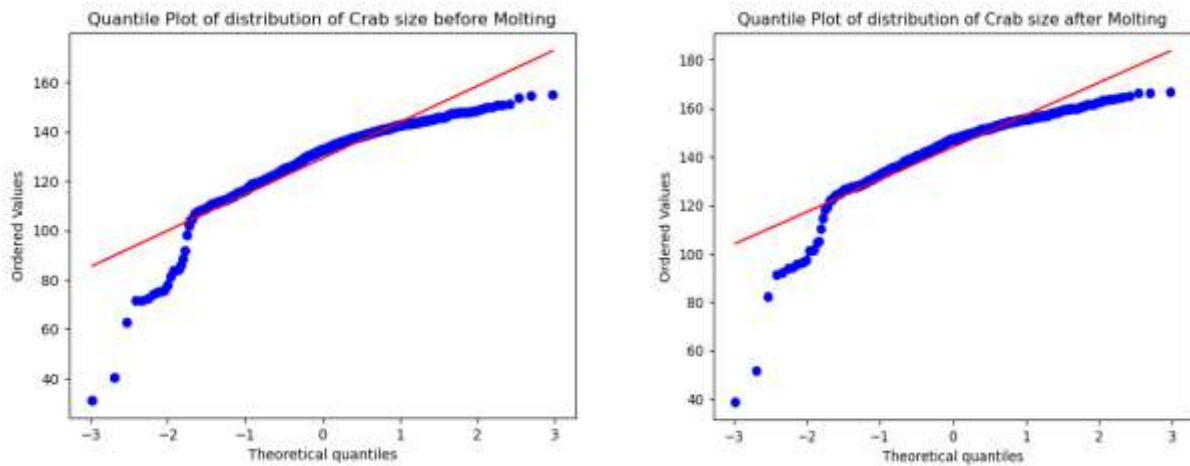


Figure 3: Quantile plots of Crab size distributions before and after molting

The calculated Cohen's d of 0.96 suggests a large effect size. In biological terms, this indicates that molting has a significant and consistent impact on crab growth across the studied population. It suggests that the molting process is a crucial period for crab development.

Statistical Significance:

The paired t-test resulted in a p-value significantly less than 0.05 ($p < 0.0001$), demonstrating that the observed increase in size is statistically significant. This strong statistical evidence confirms that the size change is a consistent outcome of the molting process rather than a result of random variation.

The T-test assumes that the data is approximately normally distributed. However, our data is negatively skewed. So, we will calculate the p-value using the Monte Carlo method.

Given substantial departures from normality in the size distributions, this method was chosen over traditional t-tests. The maximum number of simulations was set to 10,000,000. Pooled sizes, total, and lengths of pre- and post-molt groups were calculated. The observed difference in means, representing the change in crab size due to molting, was determined.

During each iteration of the Monte Carlo simulation, a random sample without replacement was drawn from the pooled sizes. The means of the pre-and post-molt groups were computed based on the sample, and if the difference in means exceeded the observed difference, it was counted.

After completing the simulations, the p-value was estimated as the proportion of occurrences where the average difference in means from the random sampling was at least as large as the observed difference. The resulting p-value from the Monte Carlo simulation was found to be 9.999×10^{-8}

These findings suggested that the departures from normality were significant enough to invalidate a p-value obtained from a t-test. Therefore, the Monte Carlo procedure provided a more suitable approach for estimating the p-value.

Despite conducting 10 million random samplings, there were zero instances where the average difference in means was as large as the observed difference in crab size before and after molting. The estimated p-value, calculated using the method advocated by Davison & Hinkley

(1997), and described in North, Curtis & Sham (2002, p. 439), was 10^{-7} indicating strong evidence against the null hypothesis of no statistical difference in crab size before and after molting. This is approximately the probability of getting 23 heads in a row from tosses of a fair coin.

Appendix C: Code

In this appendix, we document the Python code for the comparison of cumulative distributions of crab sizes, Cohen's d, T-test, and Monte Carlo procedure to estimate the effect size and statistical significance for the Crab Molt dataset.

Plotting the distribution of crab sizes

```
plt.figure(figsize=(10, 6))
sns.kdeplot(pre_molt, fill=True, label='Distribution of Crab size before Molting')
sns.kdeplot(post_molt, fill=True, label='Distribution of Crab size after Molting')
plt.xlabel('Value')
plt.ylabel('Density')
plt.title('Comparison of the distributions of the two data sets')
plt.legend()
plt.show()
```

Plotting the Cumulative Distributions of Crab sizes

```
sorted_pre_molt = np.sort(pre_molt)
sorted_post_molt = np.sort(post_molt)
cumulative_prob1 = np.arange(1, len(sorted_pre_molt) + 1) / len(sorted_pre_molt)
cumulative_prob2 = np.arange(1, len(sorted_post_molt) + 1) / len(sorted_post_molt)
plt.figure(figsize=(10, 6))
plt.scatter(sorted_pre_molt, cumulative_prob1, label='CDF for distribution of Crab size before Molting', s=1, color='blue')
plt.scatter(sorted_post_molt, cumulative_prob2, label='CDF for distribution of Crab size after Molting', s=1, color='red')
plt.xlabel('Value')
plt.ylabel('Cumulative Probability')
plt.title('Comparison of Cumulative Distribution Functions (CDF)')
plt.legend()
plt.show()
```

Cohen's d for Effect Size

```
def cohend(d1, d2):  
    n1, n2 = len(d1), len(d2)  
    s1, s2 = np.var(d1, ddof=1), np.var(d2, ddof=1)  
    s = np.sqrt(((n1 - 1) * s1 + (n2 - 1) * s2) / (n1 + n2 - 2))  
    u1, u2 = np.mean(d1), np.mean(d2)  
    return (u2 - u1) / s  
  
cohend_value = cohend(pre_molt, post_molt)  
print(f'Cohen's d: {cohend_value}')
```

Output

Cohen's d: 0.9620678692342478

T-test for Statistical Significance

```
t_stat, p_value_t = ttest_rel(pre_molt, post_molt)  
print(f't-test: {t_stat}')
```

```
print(f'p-value (T-Test): {p_value_t}')
```

Output:

t-test: -130.56558198217652

p-value (T-Test): 0.0

Monte Carlo Method

```
max_simulations = 10000000  
# Pooled ages, total, and lengths  
PooledSizes = np.concatenate((pre_molt, post_molt))  
total = np.sum(PooledSizes)  
w = len(post_molt)  
b = len(pre_molt)  
# Observed difference in means  
diff = np.mean(post_molt) - np.mean(pre_molt)  
# Monte Carlo simulation  
r = 0  
for n in range(1, max_simulations + 1):  
    A = np.random.choice(PooledSizes, size=w, replace=False) # Random sample without replacement
```

```
x = np.mean(A)
y = (total - np.sum(A)) / b
if x - y > diff:
    r += 1
# Calculate p-value
p_value_mc = (r + 1) / (n + 1)
print("p-value (Monte Carlo)=", p_value_mc)
```

Output:

p-value (Monte Carlo)= 9.9999990000001e-08

Contributions:

Sindhuja Baikadi - 02128756: Worked on the Discussion, Method, and Results sections. Also self-plotted the graphs to analyze the data using the various methods discussed in the report.

Veda Sahaja Bandi - 02105111: Outlining key observations and insights, worked on the coding portion of the project to implement necessary functionalities and features.

Mythri Krpet Jayaram - 02127296: Worked on the Issues and Findings section of the report, addressing various issues and identifying insights