

# Estimating standard errors

## Bootstrap method

### Issues & findings:

*Our analysis of adult female Dungeness crabs revealed a noteworthy relationship between their pre- and post-molt carapace sizes. The data showed a strong linear correlation, with a correlation coefficient of 0.9903699 and a highly significant p-value of less than 0.001, indicating that pre-molt size is a good predictor of post-molt size. These findings provide strong evidence against the null hypothesis, indicating that there is a significant difference between pre-molt size and post-molt size in Dungeness crabs. We used the Bootstrap method to calculate the standard errors of the coefficients  $B_0$  and  $B_1$ , which were found to be 2.734877 and 0.01868114, respectively.*

### Discussions:

***To develop a linear model between pre-molt and post-molt carapace sizes in Dungeness crabs, we treated pre-molt size as the dependent variable and post-molt size as the independent variable. We then derived a linear equation to describe the relationship between the two variables, which was  $\text{pre-molt} = B_0 + B_1 \cdot \text{post-molt}$ .***

**To estimate the standard errors of the coefficients  $B_0$  and  $B_1$ , we used the Bootstrap method. This involved resampling the data and recalculating the coefficients multiple times to generate a distribution of estimates. From this distribution, we determined the standard errors of  $B_0$  and  $B_1$  to be 2.734877 and 0.01868114, respectively.**

$$\text{premol}t = \beta_0 \times \text{postmol}t + \beta_1 + \epsilon$$

## Appendix A: Method

**To analyze the relationship between pre-molt and post-molt carapace sizes in Dungeness crabs, we imported the relevant data into R-Studio. We then calculated the correlation coefficient between the two variables and developed a linear model to describe their relationship. We analyzed the summary of the model and found a very low p-value, indicating that the model is highly accurate.**

**To estimate the standard errors in coefficients  $B_0$  and  $B_1$ , we used the Bootstrap method. This involved resampling the data and recalculating the coefficients multiple times, generating a distribution of estimates. From this distribution, we determined the standard errors in  $B_0$  and  $B_1$  to further evaluate the accuracy of our model.**

# Appendix B: Results

Correlation on coefficient between the variables is 0.9903699.

```
cor<-cor.test(data$presize,data$postsize)
> cor
```

Pearson's product-moment correlation

```
data: data$presize and data$postsize
t = 155.08, df = 470, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval: 0.9884701
0.9919580 sample> estimates:
  cor
0.9903699
```

## Linear model summary

```
> mod1<-lm(presize~postsize , data= data)
> summary(mod1)
Call:
lm(formula = presize ~ postsize, data = data)
```

Residuals:

```
  Min   1Q Median   3Q   Max
-6.1557 -1.3052  0.0564  1.3174 14.6750
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)    (
Intercept) -25.21370   1.00089  -25.19 <2e-16 ***
postsize    1.07316   0.00692  155.08 <2e-16 ***
---
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.199 on 470 degrees of freedom  
Multiple R-squared: 0.9808, Adjusted R-squared: 0.9808  
F-statistic: 2.405e+04 on 1 and 470 DF, p-value: < 2.2e-16

## Bootstrap Method

```
> error1 <- sd(mat[,1])
> error2 <- sd(mat[,2])
> error1
[1] 2.767597
> error2
[1] 0.01892139
```

## Appendix C: Code

```
library(readxl)
```

```
file_path <- "C:\\Users\\hari\\OneDrive\\Desktop\\crab_molt.xls"
```

```
data <- read_excel(file_path, sheet = 1)
```

```
correlation <- cor.test(data$presize, data$postsize)
```

```
model1 <- lm(presize ~ postsize, data = data)
```

```
summary(model1)
```

```
bootstrap <- function(data) {
```

```
  bs_sample <- data[sample(nrow(data), replace = TRUE), ]
```

```
  model2 <- lm(presize ~ postsize, data = bs_sample)
```

```
  coef(model2)
```

```
}
```

```
n_bootstrap <- 1000
```

```
bootstrap_matrix <- matrix(nrow = n_bootstrap, ncol = 2)
```

```
for (i in 1:n_bootstrap) {  
  bootstrap_matrix[i,] <- bootstrap(data)  
}
```

```
standard_error1 <- sd(bootstrap_matrix[,1])  
standard_error2 <- sd(bootstrap_matrix[,2])
```

```
standard_error1  
standard_error2
```