

Estimating standard errors

Bootstrap method

Issues

One of the biggest and most prevalent crab species is the huge crustacean known as the Dungeness crab, which is frequently seen along the Pacific coast. The sizes of the pre- and post-molt carapaces were recorded in a database that was created for the analysis of adult female Dungeness crabs. Both a lab setting and an ocean habitat were used to molt the crabs in the database. Our goal is to look into the relationship between the size of Dungeness crabs' carapaces before and after molting. We'll then summarize the results and include any standard errors for the coefficients β_0 and β_1 .

Findings

A striking link between the two variables was found in our research of the pre- and post-molt carapace sizes in Dungeness crabs. We discovered a strong linear association between pre-molt size and post-molt size, as demonstrated by an impressive correlation coefficient of 0.9903699 and a remarkably small p-value of $2.2e-16$ (less than 0.001). These outstanding statistical findings show that there is a significant difference between pre-molt size and post-molt size in Dungeness crabs, and they give strong evidence against the null hypothesis. Any by using the Bootstrap method calculated standard errors of the coefficients β_0 , β_1 .

- The standard error for coefficient β_0 is 2.734877.
- The standard error for coefficient β_1 is 0.01868114.

Discussions

In designing the linear model between the two variables, we considered pre-size variable as dependent variable and post-size as independent variable and developed the linear equation between the two variables.

$$pre\text{molt} = \beta_0 \times post\text{molt} + \beta_1 + \epsilon$$

By using the linear equation estimated the standard errors in coefficient β_0 , β_1 by Bootstrap method.

Appendix A: Method

We imported the data on post-molt and pre-molt carapace sizes of Dungeness crabs into R-Studio. Pre-molt size, which represents the size of the crab's shell before molting, was analyzed as a function of post molt size, the size of the shell after molting.

Calculated the correlation coefficient the two variables and developed the linear model between the variables, analyzed the summary of the model and found that p-value is very less which shows that model is accurate.

By using the Bootstrap method calculated the standard errors in coefficient β_0 , β_1 .

Appendix B: Results

Correlation 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

```
install.packages('readxl')

library(readxl)

file<-"C:\\Users\\sasik\\OneDrive\\Desktop\\crab_molt.xls"

data <- read_excel(file, sheet = 1)

data

str(data)

summary(data)

colnames(data)

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

cor

mod1<-lm(presize~postsize , data= data)

summary(mod1)

# bootstrap

b_s <- function(data)
{
bs_sample <- data[sample(nrow(data), replace = TRUE), ]
model2 <- lm(presize ~ postsize, data = bs_sample)
coef(model2)
}
```

```
n_b_s <- 1000
mat <- matrix(nrow = n_b_s, ncol = 2)
for (i in 1:n_b_s) {
  mat[i,] <- b_s(data)
}
```

```
error1 <- sd(mat[,1])
```

```
error2 <- sd(mat[,2])
```

```
error1
```

```
error2
```