Estimating standard errors using Bootstrap method

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 used the dataset which contains the two variables (Post-size, Pre- size) with 472 samples. Developed the Linear equation between these two variables and estimated the standard errors of coefficients.

Findings

By using the 472 samples in dataset we developed the linear equation between the two variables pre-size and post-size by taking the pre-size as dependent variable and post-size as independent variable. And calculated the standard errors of coefficients using Bootstrap method.

The standard error for coefficient 0 is 2.763058.

The standard error for coefficient 1 is 0.01891015.

Discussions

In finding the linear equation between the variables (pre-size and post-size), we considered the pre-size variable as dependent and post-size as independent variable. Developed the linear model between the two variables and observed the R-squared value which is 0.9808.

$$premolt = \beta_0 \times postmolt + \beta_1 + \epsilon$$

By using the above linear equation we estimated the standard errors in coefficients β0 and β1 by using the Bootstrap method.

Appendix A: Method

The first step involves loading the crab-molt dataset into R studio, followed by creating a linear model using the dataset. Once the model has been built, its summary can be obtained for further analysis.

Our objective is to create a custom function that takes data and index as inputs and employs a linear model. Subsequently, we will generate 1000 bootstrap samples to record the coefficients. Finally, we will compute and display the standard errors of beta0 and beta1.

Appendix B: Results

Linear model was built between the two variables and obtained the summary.

```
> #Linear model
> model1<-lm(presize~postsize , data= data)
> summary(model1)

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
```

Standard error in coefficients of β0, β1.

```
> err_1
[1] 2.763058
> err_2
[1] 0.01891015
```

Appendix C: Code

install.packages('readxl')

library(readxl)

file <-"E:\\Assignments\\MTH 522\\Project 3\\crab_molt.xls"

data <- read_excel(file, sheet = 1)

data

str(data)

```
summary(data)

colnames(data)

ncol(data)

nrow(data)


#Linear model

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

summary(model1)

coef(model1)[1]

coef(model1)[2]


#Using bootstrap

bs <- function(data)

  {

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

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

  coef(model2)

}


n_bs <- 1000

mat <- matrix(nrow = n_bs, ncol = 2)

for (i in 1:n_bs) {
```

```
  mat[i,] <- bs(data)
}

# standard error of the coefficients
err_1 <- sd(mat[,1])
err_2 <- sd(mat[,2])
err_1
err_2
```