Bootstrap Method to estimate standard errors

The issue:

We are using.xls-formatted data on crab molt for this project. Our aim is to generate a linear equation from the above dataset. Next, we use bootstrap to estimate the standard error of the coefficients 0 and 1. There will be one independent variable and one predictor variable in the equation.

Findings:

According to the model we created, we found that a crab's post size plays a large role in forecasting its pre size. We determined the standard error for 0 to be 2.8 and for 1 to be 0.01914 using the bootstrap approach.

Discussions:

The findings state that post-molt size is a significant predictor of pre-molt size compared to other variables. It is possible that unmeasured or unanalyzed confounding factors may exist. Thus, further research is necessary to examine these potential factors. Their impact on the relationship between pre-molt and post-molt sizes is also necessary.

It's important to take into account the practical significance of the relationship between pre-molt and post-molt sizes, despite statistical significance suggesting a connection. It's crucial to take into account the sample size and its representativeness of the population of interest. When the sample size is small or unrepresentative, the estimated coefficients and standard errors may not provide accurate reflections of the true population values.

Appendix A: Method

Initially crab-molt dataset is loaded into R studio, and then we develop a linear model from the given dataset crab-molt and get the summary of the model.

We will develop a function that utilizes a linear model and accepts data and index as inputs. Afterwards, we will produce 1000 bootstrap samples and record the coefficients. The beta0 and beta1 standard errors is then computed and printed.

Appendix B: Results

From the model we built, following is the result obtained.

```
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
```

Below we can see the final result i.e, standard error in coefficients of β_0, β_1.

```
> cat("Standard error of beta0:", se_beta0, "\n")
Standard error of beta0: 2.721713
> cat("Standard error of beta1:", se_beta1, "\n")
Standard error of beta1: 0.01859958
```

Appendix C: Code

We will write our code in R studio using R language

Code:

```
library(readxl)
library(readxl)
library(tidyverse)
library(magrittr)
library(broom)
library(boot)

faf<- "C:/Users/DELL/Downloads/crab_molt.xls"
data<- readxl::read_xls(faf)
View(data)
str(data)

mod<-lm(presize~postsize , data= data)
summary(mod)
coef(mod)[2]
coef(mod)[1]

#writing bootstrap function
bootstrap <- function(data) {
  bootstrap_sample <- data[sample(nrow(data), replace = TRUE), ]
  model <- lm(presize ~ postsize, data = bootstrap_sample)
  coef(model)
}

# Generate 1000 bootstrap samples and store the coefficients
n_bootstrap <- 1000
```

```
betas <- matrix(nrow = n_bootstrap, ncol = 2)
for (i in 1:n_bootstrap) {
  betas[i,] <- bootstrap(data)
}

# Calculate the standard errors of beta0 and beta1
se_beta0 <- sd(betas[,1])
se_beta1 <- sd(betas[,2])

# Print the standard errors
cat("Standard error of beta0:", se_beta0, "\n")
cat("Standard error of beta1:", se_beta1, "\n")
```