lab6.R

Feipeng Huang

2022-10-19

#lab6  
require(palmerpenguins)

## Loading required package: palmerpenguins

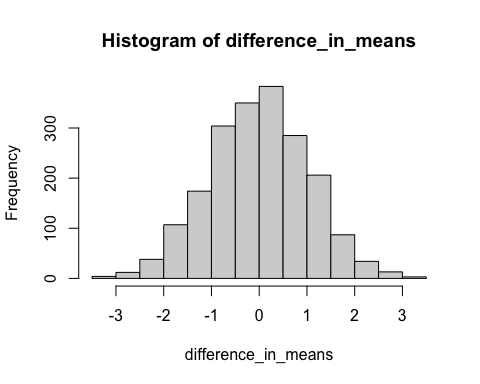
#Q1  
rm(list = ls())  
sse\_mean = function(x)  
{  
 sse = sd(x, na.rm = TRUE)/sqrt(length(x[!is.na(x)]))  
   
 return(sse)   
}  
sse\_mean(penguins$body\_mass\_g)

## [1] 43.36473

sse\_mean(mtcars$mpg)

## [1] 1.065424

#Q2  
two\_group\_resample\_diff = function(x, n\_1, n\_2)   
{  
 a = mean(sample(x, size = n\_1, replace = T), na.rm = T)  
 b = mean(sample(x, size = n\_2, replace = T), na.rm = T)   
 difference\_in\_means = a - b  
 return(difference\_in\_means)  
}  
#Q3  
#My function performs Monte Carlo resampling and simulates a null hypothesis because it breaks up associations in the data.  
#Q4  
dat\_pen = droplevels(subset(penguins, species != "Gentoo"))  
n = 2000  
difference\_in\_means = c()  
for (i in 1:n)  
{  
 difference\_in\_means = c(  
 difference\_in\_means,  
 two\_group\_resample\_diff(dat\_pen$flipper\_length\_mm, 68, 152)  
 )  
}  
hist(difference\_in\_means)

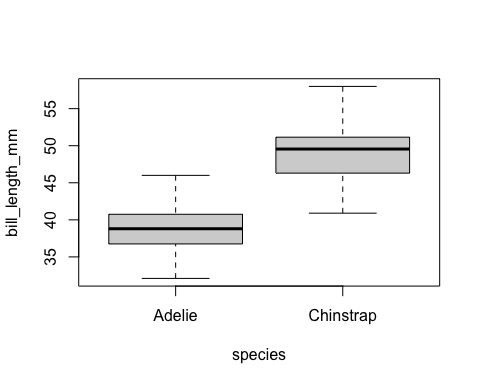


#Q5  
#0  
sum(abs(difference\_in\_means)>5.8)

## [1] 0

#Q6  
#more than 10 million  
#Q7

boxplot(  
 bill\_length\_mm ~ species, data = dat\_pen,  
 ylab = "bill\_length\_mm")



#Q8  
agg\_means = aggregate(  
 bill\_length\_mm ~ species,   
 data = dat\_pen,   
 FUN = "mean",   
 na.rm = TRUE)  
diff\_crit = diff(agg\_means[, 2])  
  
agg\_means

## species bill\_length\_mm  
## 1 Adelie 38.79139  
## 2 Chinstrap 48.83382

diff\_crit

## [1] 10.04243

#Q9  
t.test(dat\_pen$bill\_length\_mm ~ dat\_pen$species)

##   
## Welch Two Sample t-test  
##   
## data: dat\_pen$bill\_length\_mm by dat\_pen$species  
## t = -21.865, df = 106.97, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -10.952948 -9.131917  
## sample estimates:  
## mean in group Adelie mean in group Chinstrap   
## 38.79139 48.83382

#Chinstrap penguin has a longer bill than Adelie penguin. This relationship holds true for many individuals. There may be Adelie penguins with long bills, but that’s very rare.  
#Q10  
#0  
n = 1000  
difference\_in\_means = c()  
for (i in 1:n)  
{  
 difference\_in\_means = c(  
 difference\_in\_means,  
 two\_group\_resample\_diff(dat\_pen$bill\_length\_mm, 68, 152)  
 )  
}  
sum(abs(difference\_in\_means)>10.04243)

## [1] 0

#Q11  
hist(difference\_in\_means, main = "Histogram of difference in mean bill length of Adelie and Chinstrap penguins", cex.main = 0.7)

