lab6.R

Feipeng Huang

2022-10-19

```r
#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)
  )
```
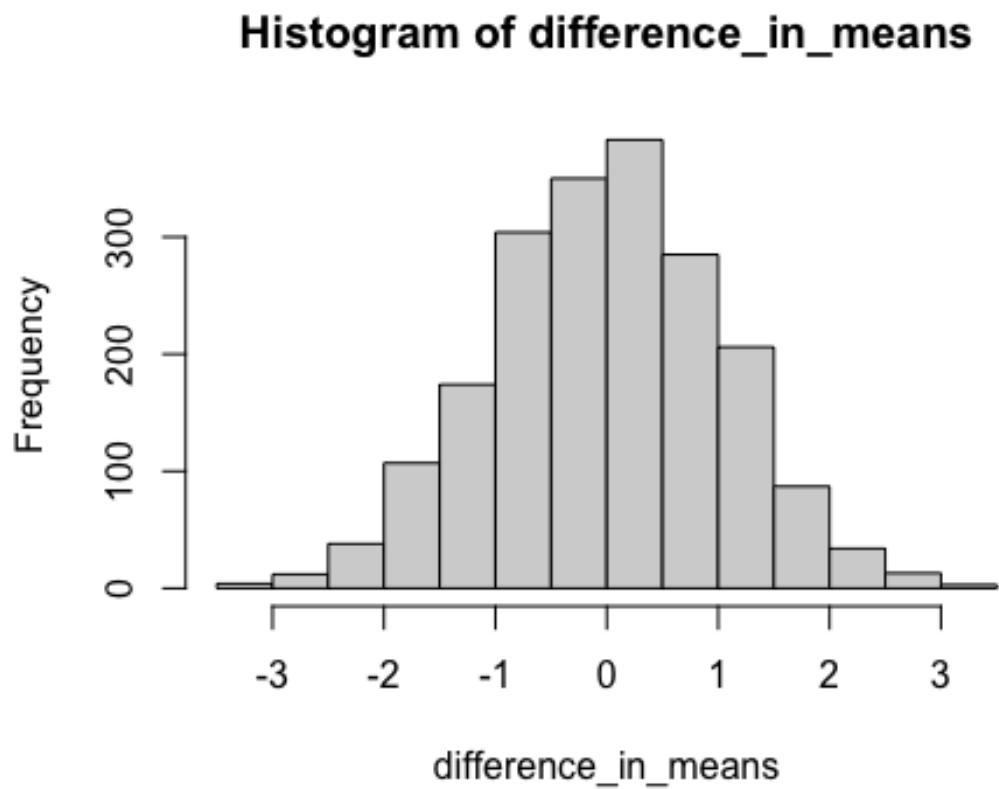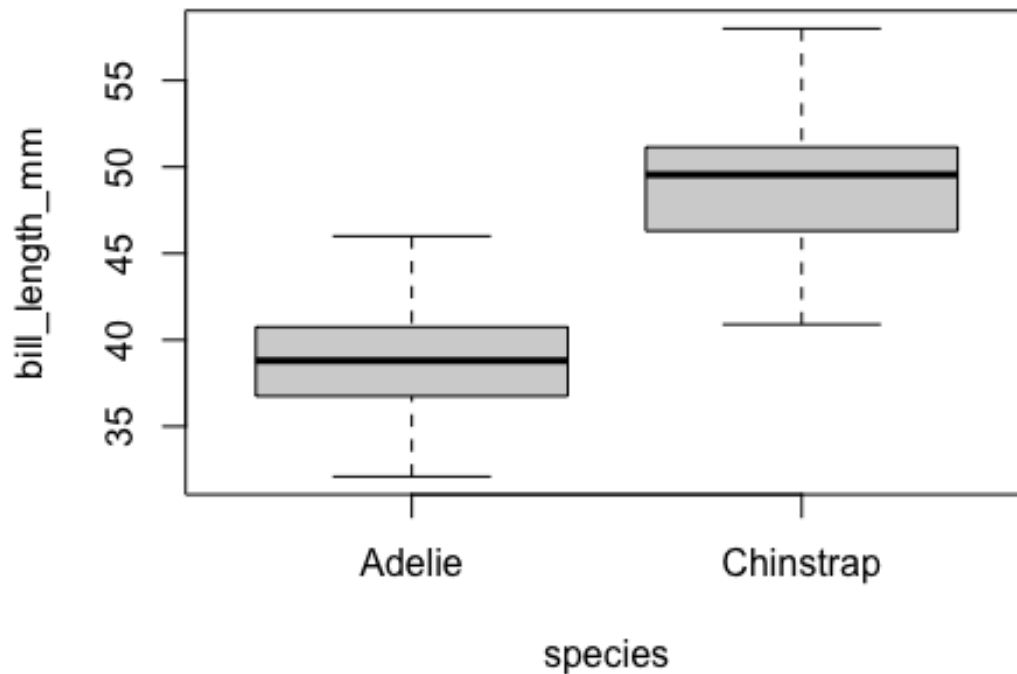
```
}
hist(difference_in_means)
```

Histogram of 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)
```

**Histogram of difference in mean bill length of Adelie and Chinstrap penguins**