Hw8

Matthew Maslow

4/11/2023

library(dplyr)  
library(ggplot2)

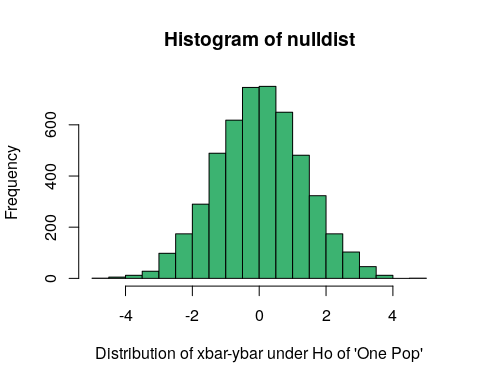
# Sec: 3.9

## 3.

library(resampledata)  
data("FlightDelays")

#### b.

MDclass <-   
 FlightDelays %>%  
 filter(Month == "May" | Month == "June") %>%  
 select(Month, Carrier, Delay)  
  
May = MDclass %>% filter(Month == "May") %>% select(Delay)  
May = May$Delay  
June = MDclass %>% filter(Month == "June") %>% select(Delay)  
June = June$Delay  
  
teststat = round(mean(June) - mean(May), 2)   
nMay = length(May)   
nJune = length(June)   
nulldist = NULL   
nPerm = 5000   
  
for (i in 1:nPerm)  
{  
 onepop = c(May, June)  
 Mayind = sample(1:(nMay + nJune), nMay)  
 arbMay = onepop[Mayind]   
 arbJune = onepop[-Mayind]  
 dstar = mean(arbJune) - mean(arbMay)  
 nulldist[i] = dstar  
}  
# visualize our null distribution  
hist(nulldist,col="mediumseagreen", xlab="Distribution of xbar-ybar under Ho of 'One Pop'")  
# visualize our test statistic on the null distribution  
abline(v = teststat, col = "red", lwd = 2)



# if Ha: mu\_May not equal mu\_June  
2\*(sum(nulldist > teststat)+1)/(nPerm+1)

## [1] 0.00039992

teststat

## [1] 5.66

In conclusion we reject the null hypothesis and accept the alternative, as the the p-value is significant enough, to say that the difference in the means of the two months for flight delays is significantly different.

## 5

#### i

AirLineclass <-   
 FlightDelays %>%   
 filter(Carrier == "AA" | Carrier == "UA")  
AA = AirLineclass %>% filter(Carrier == "AA") %>% select(Delay)  
AA = AA$Delay  
UA = AirLineclass %>% filter(Carrier == "UA") %>% select(Delay)  
UA = UA$Delay  
  
(teststatM = round(mean(UA), 2))

## [1] 15.98

(teststatS = round(sum(UA), 2))

## [1] 17949

(teststatD = round(mean(UA) - mean(AA), 2))

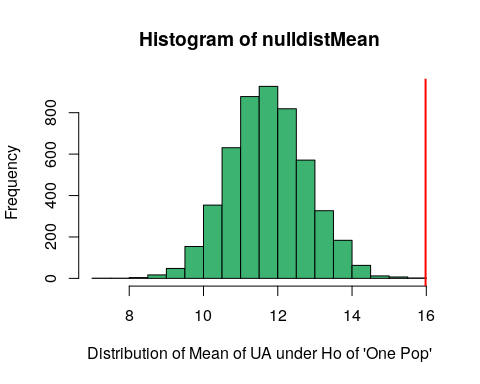
## [1] 5.89

nAA = length(AA)   
nUA = length(UA)

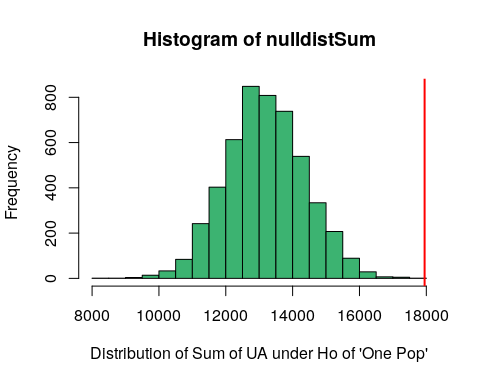
nulldistMean = NULL  
nulldistSum = NULL  
nulldistDiff = NULL  
nPerm = 5000   
  
for (i in 1:nPerm)  
{  
 onepop = c(AA,UA)   
 AAind = sample(1:(nUA + nAA), nAA)  
 arbAA = onepop[AAind]   
 arbUA = onepop[-AAind]

dstarMean = mean(arbUA)  
 nulldistMean[i] = dstarMean  
  
 dstarSum = sum(arbUA)  
 nulldistSum[i] = dstarSum  
   
 dstarDiff = mean(arbUA) - mean(arbAA)  
 nulldistDiff[i] = dstarDiff  
}

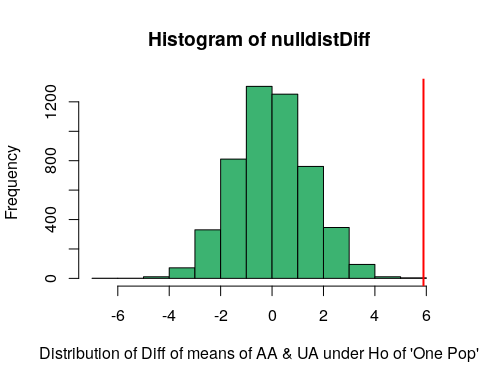
# mean  
hist(nulldistMean, col = "mediumseagreen",  
 xlab="Distribution of Mean of UA under Ho of 'One Pop'")  
abline(v = teststatM,col = "red", lwd = 2)



# sum  
hist(nulldistSum, col = "mediumseagreen",  
 xlab="Distribution of Sum of UA under Ho of 'One Pop'")  
abline(v=teststatS, col = "red", lwd = 2)



# difference  
hist(nulldistDiff, col = "mediumseagreen",   
 xlab="Distribution of Diff of means of AA & UA under Ho of 'One Pop'")  
abline(v=teststatD,col = "red", lwd = 2)



2\*((sum(nulldistMean > teststatM)+1)/(nPerm+1))

## [1] 0.00079984

2\*((sum(nulldistSum > teststatS)+1)/(nPerm+1))

## [1] 0.00079984

2\*((sum(nulldistDiff > teststatD)+1)/(nPerm+1))

## [1] 0.00079984

*For the mean of UA:*

*For the sum of UA:*

*For the diff of means:*

All of the ts’s do vary a bit but, the p-values are all the same, besides the one for the difference since it is looking at both tails, so its times two.

# Sec: 8.5

## 1

mu = 9.5  
xbar = 9.2  
s = 1.1  
n = 20  
ts = (xbar - mu) / (s/sqrt(n))  
ts

## [1] -1.219673

df = n - 1  
df

## [1] 19

p = 0.119 \* 2 # using t-dist. in statkey  
p

## [1] 0.238

We would accept our null hypothesis, and reject the alternative, as the evidence is not strong enough to say there is a significant difference in the community versus the sample, so it would not support his hypothesis.