murphy-lab6.R

William Murphy

Thu Nov 29 22:58:18 2018

setwd("/home/manwthglasses/Documents/IntroToDataScience/")  
  
set.seed(2) #this is to set seed. By doing so, the initiation point is always the same, not random.  
##The situation we are given tells us that we are recieving a sample of 30 students, therefore we must keep the sample size constant  
sampleSize <- 30  
  
# create normal distribution of 20000 observations with mean value 20 and standard deviation of 3 and set this as a student population   
##This is the population we are going to be basing our samples off of  
studentPop <- rnorm(20000,mean=20,sd=3)  
#investigate studentPop now. How many rows? What are the values look like? Are they close to the mean value 20?  
mean(studentPop)

## [1] 20.01569

##undergrads is a sample from the student population, with the set size of 30  
undergrads <- sample(studentPop,size=sampleSize,replace=TRUE)  
#create a sample of graduate students. Sample size is 30, mean is 25, standard deviation is 3. See the mean is 5 years older than the undergraduate sample apparently.   
grads <- rnorm(sampleSize,mean=25,sd=3)  
  
##the runif() function generates a random number from 0 to 1  
##the if statement is executed 50% of the time, because 0.5 is the halfway mark of the limit of the random number, therefore imitating the flipping of a coin.  
##there is a 50% chance a test sample is made from the undergrad sample and a 50% chance it is from the grad sample  
if (runif(1)>0.5){   
 testSample <- grads   
} else {   
 testSample <- undergrads   
}  
mean(testSample)

## [1] 25.54158

##a new sample is derived in order to compare it with the testSample variable  
popSample <- sample(studentPop, size=100, replace = TRUE)  
  
##the testSample is compared with the popSample to see if it falls within the extremes  
##the extremes are below 2.5% or above 97.5%  
if(mean(testSample)<quantile(popSample, probs = 0.025)| mean(testSample)>quantile(popSample, probs = .975)){  
 print("Sample mean is extreme")  
} else{  
 print("Sample mean is not extreme")  
}

## [1] "Sample mean is not extreme"