Homework 5

Kyle Walter

5/16/2021

################################################

# IST772, Standard Homework Heading

#

# Student name: Kyle Walter

# Homework number: 5

# Date due: 2021-05-16

#

# Attribution statement: (choose only one)

# I did this homework by myself, with help from the book and the professor

## Exercise 6

data("PlantGrowth")  
ctrl.weights = PlantGrowth$weight[PlantGrowth$group == 'ctrl']  
trt1.weights = PlantGrowth$weight[PlantGrowth$group == 'trt1']  
trt2.weights = PlantGrowth$weight[PlantGrowth$group == 'trt2']

t.test(ctrl.weights,trt1.weights)

##   
## Welch Two Sample t-test  
##   
## data: ctrl.weights and trt1.weights  
## t = 1.1913, df = 16.524, p-value = 0.2504  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.2875162 1.0295162  
## sample estimates:  
## mean of x mean of y   
## 5.032 4.661

From the output above, the T value is 1.19193 and the Degress of Freedom are 16.524. The alpha is .05 while the p value is .2504. Since the P value is greater than .05 we would *fail* to reject the null hypothesis. In addition the confidence interval runs from -.28 to 1.03 passing through zero, or basically indicating that that mean difference between control and treatment 1 could be zero.

## Exercise 7

library(rjags)

## Loading required package: coda

## Linked to JAGS 4.3.0

## Loaded modules: basemod,bugs

library(BEST)

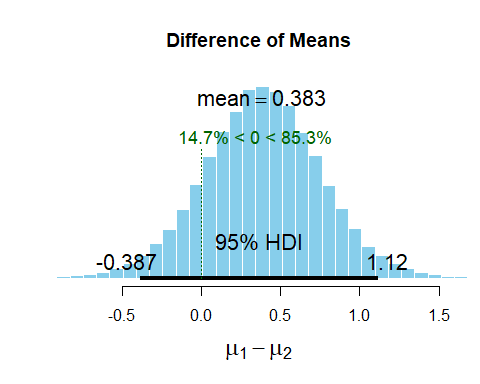
## Loading required package: HDInterval

PGBest <- BESTmcmc(ctrl.weights,trt1.weights)

## Waiting for parallel processing to complete...

## done.

plot(PGBest, main=NULL)



The HDI, or highest density interval falls between -.36 and 1.15. This MCMC simulation shows us with 95% confidence there the population mean is likely be with in the range of mean differences in this case it is likely to be .388, meaning the control group is showing a slightly higher population weight than the treat 1 group.

## Exercise 8 - Explain the findings from the null hypothesis test, the confidence interval, and the HDI.

The T test from question 6 shows us that the signficant T value given the sample is like to appear 25% of the time. As we want the value to be rarer as a way of demonstrating the population means of the control and sample are likely no similiar, we would fail to reject the null hypothesis.

The confidence interval from the T test, in this case -.28 to 1.03 says that 95 out of the 100 times if sampling was repeated our population mean would likely fall in this range. Whenever the mean crosses zero there is a probability that the mean difference is zero and thus there is no statistical difference detected between the two groups.

HDI on the other hand has a range of -.352 and 1.16 where 95% probability that that population mean is around .388, however; as there is still probability that the population mean difference is zero thus reasonable suspicion that the two groups do not have a mean difference.

#Exercise 9 -Perform a t-test between TRT2 and the control group.

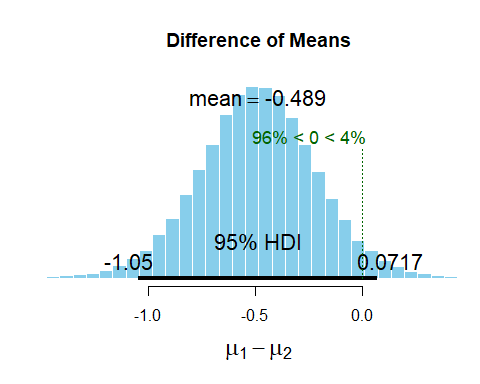
t.test(ctrl.weights,trt2.weights)

##   
## Welch Two Sample t-test  
##   
## data: ctrl.weights and trt2.weights  
## t = -2.134, df = 16.786, p-value = 0.0479  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.98287213 -0.00512787  
## sample estimates:  
## mean of x mean of y   
## 5.032 5.526

PGBest2 <- BESTmcmc(ctrl.weights,trt2.weights)

## Waiting for parallel processing to complete...done.

plot(PGBest2, main=NULL)

 Rerunning the T Test between the control and treatmen 2 group. we see a different outcome. The P value is now 0.0479 which is less than .05 suggesting the event is rare enough that we would reject the null hypothesis.

The confidence interval ranges from -0.982 to -0.00512 and does not pass through zero. We would expect if we ran the T test 100 times with samples from these two groups that the population mean would fall some where in this range 95 out of 100 times and likely not pass through zero, though there is still possibility it could.

The HDI shows the show a range of -1.06 to .059 with the population mean around -.489 and less than 5% chance that the population mean is above zero. While the data does suggest this is likely the case it is close to the threshold in all three. Running with a larger sample set would provide higher statitical significance to make that assertion.

## Exercise 10 - Examine the results of running a t-test on a very large sample size.

set.seed(1234)  
t.test(rnorm(100000,mean=17.1,sd=3.8),rnorm(100000,mean=17.2,sd=3.8))

##   
## Welch Two Sample t-test  
##   
## data: rnorm(1e+05, mean = 17.1, sd = 3.8) and rnorm(1e+05, mean = 17.2, sd = 3.8)  
## t = -5.9629, df = 199994, p-value = 2.483e-09  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.13427906 -0.06784225  
## sample estimates:  
## mean of x mean of y   
## 17.11113 17.21219

As the sample size grows, there is evidence to detect small differences in the population mean between two samples. The P values 2.483e-09 is nearing zero, meaning the chance of observing these values is very small.

The confidence interval as well indicates that the population mean difference is betwee -.13 and -.067 also not passing zero. If rerun 95 times we would expect the population mean to be somewhere in this range.