Homework 5
Syracuse University
IST 772
Summer 2021

```
library(dplyr)
library(BEST)
```

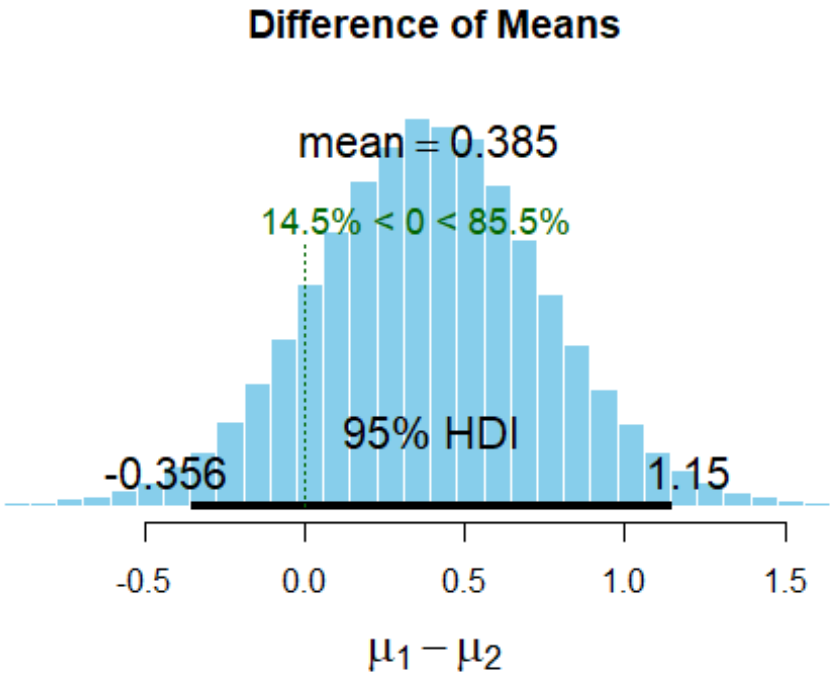
Question 6

```
# run a t test to compare the means of the control group and treatment group
1
t.test((PlantGrowth %>% filter(group == 'ctrl'))$weight,
       (PlantGrowth %>% filter(group == 'trt1'))$weight)

##
##  Welch Two Sample t-test
##
## data:  (PlantGrowth %>% filter(group == "ctrl"))$weight and (PlantGrowth
%>% filter(group == "trt1"))$weight
## 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

# t = 1.1913
# df = 16.5240
# p = 0.2504
# upper bound = 1.0295
# lower bound = -0.2875

# the null hypothesis is not rejected because zero is contained within the
# confidence interval. In addition, the results are not statistically
# significant given that the p-value of 0.2504 is > than the alpha of 0.05.
```

Question 7

```
# Use BESTmcmc to compare the PlantGrowth control group and treatment group 1
plantgrowthbest <- BESTmcmc((PlantGrowth %>% filter(group == 'ctrl'))$weight,
                            (PlantGrowth %>% filter(group == 'trt1'))$weight)

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

# print the results from the BESTmcmc simulation
print(plantgrowthbest)

## MCMC fit results for BEST analysis:
## 100002 simulations saved.
##           mean      sd  median  HDIlo  HDIup  Rhat n.eff
```

```
## mu1     5.0254  0.2238  5.0257 4.5816  5.475 1.000 54366
## mu2     4.6402  0.3078  4.6386 4.0269  5.249 1.000 52345
## nu     34.3926 29.8026 25.6490 1.2378 94.003 1.000 18396
## sigma1  0.6596  0.2023  0.6213 0.3415  1.062 1.002 27055
## sigma2  0.8958  0.2761  0.8448 0.4514  1.442 1.000 24960
##
## 'HDIlo' and 'HDIup' are the limits of a 95% HDI credible interval.
## 'Rhat' is the potential scale reduction factor (at convergence, Rhat=1).
## 'n.eff' is a crude measure of effective sample size.

# plot the results from the BESTmcmc simulation
plot(plantgrowthbest)
```




```
# upper boundary = 1.15
# lower boundary = -0.358

# the HDI is the highest density interval which means it is where the
# mean difference is going to land 95% of the time. Therefore we can be 95%
# confident that the mean difference is within the HDI.
```

Question 8

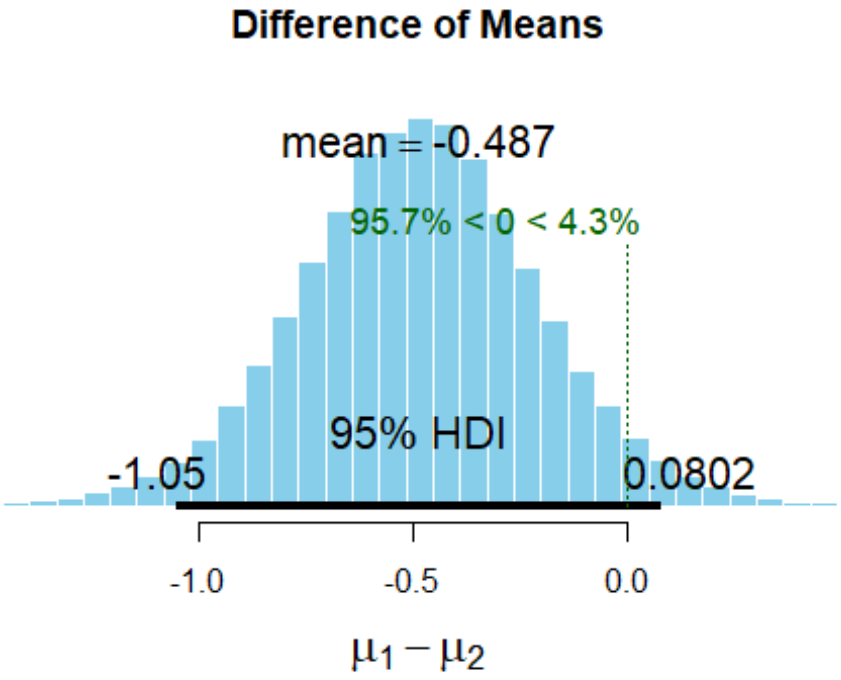
```
# null hypothesis test

# the result of the null hypothesis test was that the null hypothesis could
# not be rejected, that is that there is not a statistically significant
# mean difference in the weight between the control group and treatment group
1.
```

```
# the confidence interval

# we are 95% confident that the population mean difference in weight between
the control
# group and treatment group 1 is between -0.2875 and 1.0295.

# BESTmcmc HDI

# we can be 95% confident that the population mean difference in weight
between
# the control group and treatment group 1 is between -0.358 and 1.15.
```

Question 9

```
# run a t test to compare the means of the control group and treatment group
2
t.test((PlantGrowth %>% filter(group == 'ctrl'))$weight,
       (PlantGrowth %>% filter(group == 'trt2'))$weight)

##
##  Welch Two Sample t-test
##
## data:  (PlantGrowth %>% filter(group == "ctrl"))$weight and (PlantGrowth
%>% filter(group == "trt2"))$weight
## 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

# Use BESTmcmc to compare the PlantGrowth control group and treatment group 1
plantgrowthbest <- BESTmcmc((PlantGrowth %>% filter(group == 'ctrl'))$weight,
                            (PlantGrowth %>% filter(group == 'trt2'))$weight)

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

# print the results from the BESTmcmc simulation
print(plantgrowthbest)

## MCMC fit results for BEST analysis:
## 100002 simulations saved.
##           mean      sd  median  HDIlo   HDIup  Rhat n.eff
## mu1     5.0279  0.2261  5.0274 4.5813  5.4797 1.000 56296
## mu2     5.5146  0.1714  5.5125 5.1770  5.8637 1.000 50534
## nu     34.7301 29.8236 25.9987 1.2184 94.9641 1.002 19959
## sigma1  0.6614  0.2028  0.6225 0.3467  1.0751 1.000 27899
## sigma2  0.5005  0.1540  0.4719 0.2557  0.8107 1.000 26608
##
## 'HDIlo' and 'HDIup' are the limits of a 95% HDI credible interval.
```

```
## 'Rhat' is the potential scale reduction factor (at convergence, Rhat=1).
## 'n.eff' is a crude measure of effective sample size.

# plot the results from the BESTmcmc simulation
plot(plantgrowthbest)
```




```
# null hypothesis test

# the result of the null hypothesis test is that the null hypothesis is
# rejected, that is that there is a statistically significant mean difference
# in the weight between the control group and treatment group 1. Zero is not
# contained in the confidence interval and the p-value of 0.0479 is less than
# the alpha value of 0.05.

# the confidence interval

# we are 95% confident that the population mean difference in weight between
the control
# group and treatment group 1 is between -0.9829 and -0.0051.

# BESTmcmc HDI

# we can be 95% confident that the population mean difference in weight
between
# the control group and treatment group 1 is between -1.06 and 0.0578.
```

Question 10

```r
# perform the t test
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 = -6.2051, df = 2e+05, p-value = 5.477e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.13888904 -0.07221015
## sample estimates:
## mean of x mean of y
##  17.10231  17.20786

# comment on the results of the t test

# the result of the null hypothesis test is that the null hypothesis is
# rejected, that is that there is a statistically significant mean difference
# in the values between the two groups. Zero is not contained in the
# confidence interval and the p-value of 0.000000001588.

# what are the implications of using NHST on very large data sets?

# going back to the law of large numbers, we know that the results will
# converge on a certain value. Given that the mean of the data set is
# set up to be 17.1, it makes sense that the t.test estimates the mean
# to be 17.09. It is not exactly 17.1 due to randomness but it is very
# close.
```