

# HW5\_Lin

## IST 772 Homework 5

Due November 9, 2021 at 8:00AM EDT

Homework 5 by Nora Lin: I produced the material below with no assistance.

### Excercise 6 p.86:

```
#Doing some investigation first
mean(PlantGrowth$weight[PlantGrowth$group=="ctrl"])
```

```
## [1] 5.032
```

```
mean(PlantGrowth$weight[PlantGrowth$group=="trt1"])
```

```
## [1] 4.661
```

mean weight of control group is 5.032. #mean weight of treatment 1 group is 4.661. mean difference control-treatment1 = 0.371

```
t.test(PlantGrowth$weight[PlantGrowth$group=="ctrl"],PlantGrowth$weight[PlantGrowth$group=="trt1"])
```

```
##
## Welch Two Sample t-test
##
## data: PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weight[PlantGrowth$group == "trt1"]
## 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
```

The observed t value is 1.1913, the degrees of freedom are 16.524, and the p-value is 0.2504. Our null hypothesis would be that there is no difference in means between the two groups. Our alternative hypothesis would be that there is a difference in means between the two groups. Since our alpha is 0.05, we would fail to reject the null hypothesis that there is no difference in means between the control and treatment1 groups. The upper bound of the 95% CI is -0.2875162 and the lower bound of the 95% CI is 1.0295162.

### Excercise 7 p.86:

```
#install.packages("rjags")
#install.packages("BEST")
```

```
library(BEST)
```

```
## Loading required package: HDInterval
```

```
library(rjags)
```

```
## Loading required package: coda
```

```
## Linked to JAGS 4.0.1
```

```
## Loaded modules: basemod,bugs
```

```
BESTmcmc(PlantGrowth$weight[PlantGrowth$group=="ctrl"],PlantGrowth$weight[PlantGrowth$group=="trt1"])
```

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

```
## MCMC fit results for BEST analysis:
```

```
## 100002 simulations saved.
```

```
##           mean      sd  median  HDIlo  HDIup  Rhat n.eff
## mu1      5.0259  0.2250  5.0257  4.5701  5.468  1.000  54803
## mu2      4.6416  0.3082  4.6391  4.0255  5.252  1.000  53994
## nu       34.1601 29.4342 25.6500  1.2399 92.825  1.000  20583
## sigma1   0.6600  0.2040  0.6222  0.3378  1.059  1.001  26191
## sigma2   0.8956  0.2772  0.8458  0.4600  1.448  1.000  25592
##
```

```
## 'HDIlo' and 'HDIup' are the limits of a 95% HDI credible interval.
```

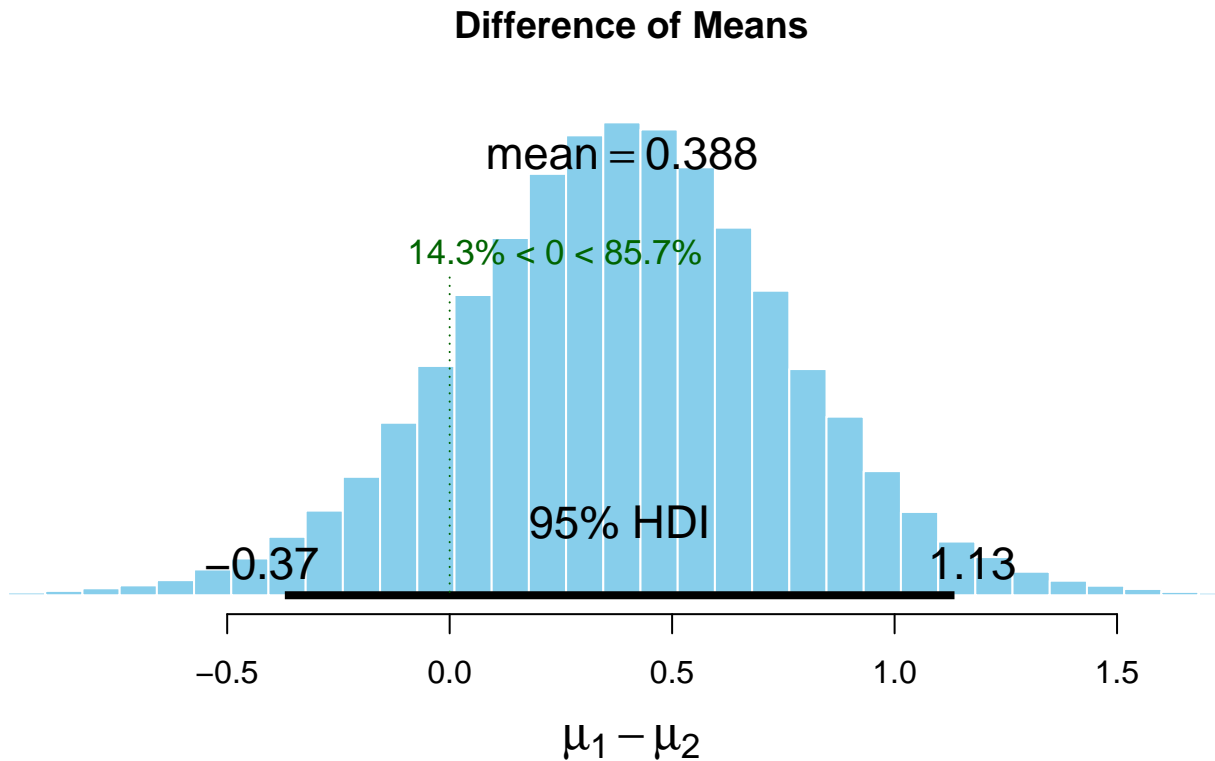
```
## 'Rhat' is the potential scale reduction factor (at convergence, Rhat=1).
```

```
## 'n.eff' is a crude measure of effective sample size.
```

```
con_trt1 <- BESTmcmc(PlantGrowth$weight[PlantGrowth$group=="ctrl"],PlantGrowth$weight[PlantGrowth$group=="trt1"])
```

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

```
plot(con_trt1)
```



There were 100,002 steps in the MCMC simulation. 5.0263 is a point estimate of the mean population value of the mean population value of the plant weight of the control group. 4.6446 is a point estimate of the mean population value of the mean population value of the plant weight of the treatment 1 group. These values are similar to the mean values we got in our investigation in Exercise #6.

SD, median, HDIlo and HDIup describes the distribution of the 100,002 estimates in the posterior distribution. 95% of the estimates of the population mean for the weight of the control group fell between 4.5859 and 6.476 with a median of 5.0268. Likewise 95% of the estimates of the population mean for the weight of the treatment 1 group fell between 4.0343 and 5.259 with a median of 4.6414.

### Exercise 8 p.86:

The results of our t-test showed that our p-value was 0.2504 which is greater than our set alpha value of 0.05, therefore we fail to reject the null hypothesis that there is no difference in means between the control and treatment 1 groups. The confidence interval can be interpreted as if we could replicate our whole study 100 times, 95 of those replication would contain the actual mean difference. Our high density interval showed that there is a 95% probability that the population mean difference between the two groups falls within our range.

### Exercise 9 p.87:

```
#Doing some investigation first
mean(PlantGrowth$weight[PlantGrowth$group=="ctrl"])
```

```
## [1] 5.032
```

```
mean(PlantGrowth$weight[PlantGrowth$group=="trt2"])
```

```
## [1] 5.526
```

mean weight of control group is 5.032. #mean weight of treatment 1 group is 5.526. mean difference control-treatment1 = -0.494

```
t.test(PlantGrowth$weight[PlantGrowth$group=="ctrl"],PlantGrowth$weight[PlantGrowth$group=="trt2"])
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weight[PlantGrowth$group == "trt2"]
```

```
## 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
```

The observed t value is -2.134, the degrees of freedom are 16.786, and the p-value is 0.0479. Our null hypothesis would be that there is no difference in means between the two groups. Our alternative hypothesis would be that there is a difference in means between the two groups. Since our alpha is 0.05, we would reject the null hypothesis that there is no difference in means between the control and treatment2 groups. The upper bound of the 95% CI is -0.98287213 and the lower bound of the 95% CI is -0.00512787.

```
BESTmcmc(PlantGrowth$weight[PlantGrowth$group=="ctrl"],PlantGrowth$weight[PlantGrowth$group=="trt2"])
```

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

```
## MCMC fit results for BEST analysis:
```

```
## 100002 simulations saved.
```

```
##      mean      sd  median HDIlo  HDIup  Rhat n.eff
```

```
## mu1    5.0287  0.2260  5.0278 4.5792  5.4787 1.000 54792
```

```
## mu2    5.5146  0.1711  5.5130 5.1702  5.8512 1.000 55069
```

```
## nu     34.8663 29.8497 26.1952 1.1136 94.9140 1.000 20803
```

```
## sigma1 0.6627  0.2026  0.6246 0.3373  1.0629 1.000 27489
```

```
## sigma2 0.5020  0.1558  0.4729 0.2579  0.8165 1.001 26379
```

```
##
```

```
## 'HDIlo' and 'HDIup' are the limits of a 95% HDI credible interval.
```

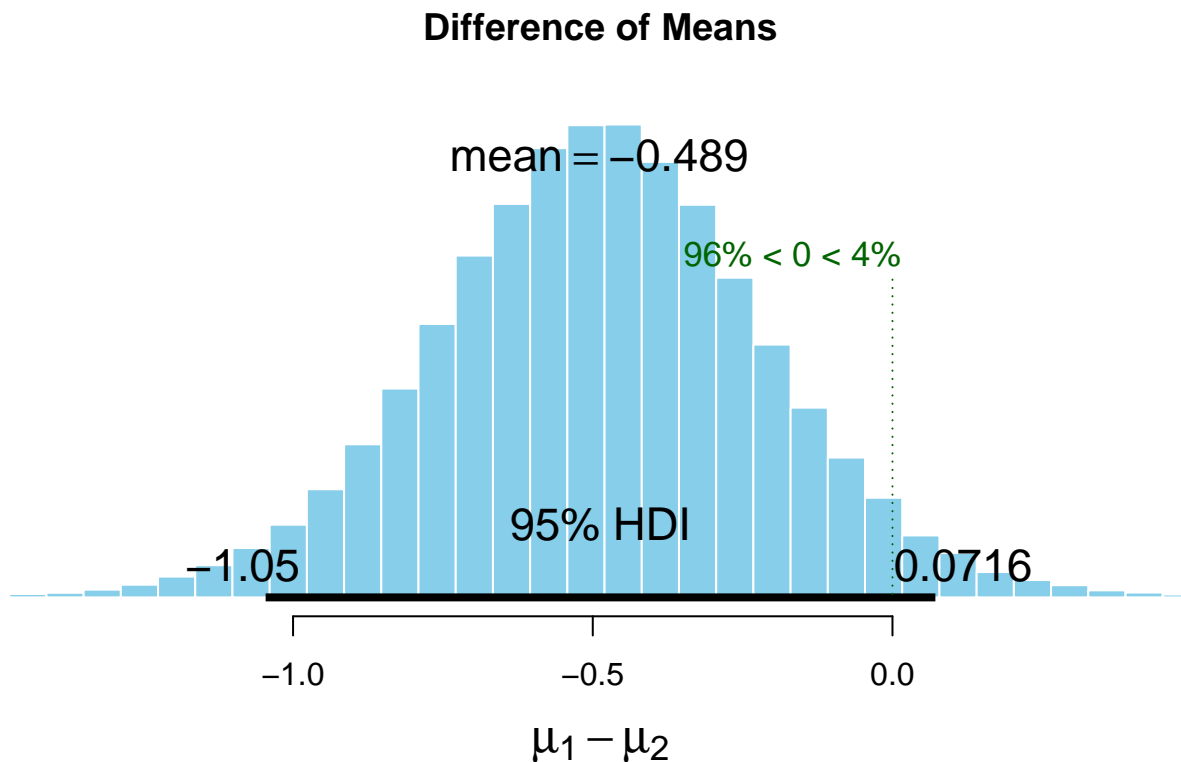
```
## 'Rhat' is the potential scale reduction factor (at convergence, Rhat=1).
```

```
## 'n.eff' is a crude measure of effective sample size.
```

```
con_trt2 <- BESTmcmc(PlantGrowth$weight[PlantGrowth$group=="ctrl"],PlantGrowth$weight[PlantGrowth$group=="trt2"])
```

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

```
plot(con_trt2)
```



There were 100,002 steps in the MCMC simulation. 5.0255 is a point estimate of the mean population value of the mean population value of the plant weight of the control group. 5.5139 is a point estimate of the mean population value of the mean population value of the plant weight of the treatment 2 group. These values are similar to the mean values we got in our investigation.

SD, median, HDIlo and HDIup describes the distribution of the 100,002 estimates in the posterior distribution. 95% of the estimates of the population mean for the weight of the control group fell between 4.5803 and 5.4776 with a median of 5.0257. Likewise 95% of the estimates of the population mean for the weight of the treatment 2 group fell between 5.1762 and 5.8603 with a median of 5.5122.

### Exercercise 10 p.87:

```
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 = -6.0592, df = 2e+05, p-value = 1.371e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.13616646 -0.06960512
```

```
## sample estimates:  
## mean of x mean of y  
## 17.08777 17.19065
```

```
#investigate:  
t.test(rnorm(10,mean=17.1,sd=3.8),rnorm(10,mean=17.2,sd=3.8))
```

```
##  
## Welch Two Sample t-test  
##  
## data: rnorm(10, mean = 17.1, sd = 3.8) and rnorm(10, mean = 17.2, sd = 3.8)  
## t = 0.20639, df = 17.481, p-value = 0.8389  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3.144134 3.827548  
## sample estimates:  
## mean of x mean of y  
## 16.41222 16.07051
```

The results of running  $n=100,000$  observations in comparison too  $n=10$  observations has a huge influence on whether we would reject or fail to reject the null hypothesis. If we were discussing fuel economy, then our null hypothesis would be that there is no difference between the fuel economy between the two groups. In the case of  $n=100,000$  observations, if our alpha value is 0.05, then we would reject the null because our p-value is smaller than alpha. However, in the case of  $n=10$  observations, given an alpha value of 0.05, we would fail to reject the null hypothesis because our p-value is greater than alpha. This shows that there are consequences to using NHST on very large datasets.

End of Homework 5