

IST 772 Homework 5

Maya Mileva

due date: Nov 7th, 2019

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

```
## Run these functions to get a clean test of homework code
## dev.off() # Clear the graph window
cat('\014') # Clear the console
```

```
rm(list=ls()) # Clear user objects from the environment
```

Exercises

```
## Activate the data set
data("PlantGrowth")

dim(PlantGrowth) # data frame of 30 cases, 2 variables
```

```
## [1] 30  2
```

```
head(PlantGrowth, n=3)
```

```
##   weight group
## 1   4.17  ctrl
## 2   5.58  ctrl
## 3   5.18  ctrl
```

```
summary(PlantGrowth)
```

```
##      weight      group
##  Min.    :3.590   ctrl:10
##  1st Qu.:4.550   trt1:10
##  Median :5.155   trt2:10
##  Mean    :5.073
##  3rd Qu.:5.530
##  Max.    :6.310
```

```
plant_table=data.table(PlantGrowth)
```

```
plant_summary=
  plant_table[,apply(.SD,function(x) list(mean=round(mean(x),3),sd=round(sd(x),3),
                                          se=round(sqrt(var(x))/length(x),3))),by=group]
```

```
## Change the names
plant_summary=setnames(plant_summary
```

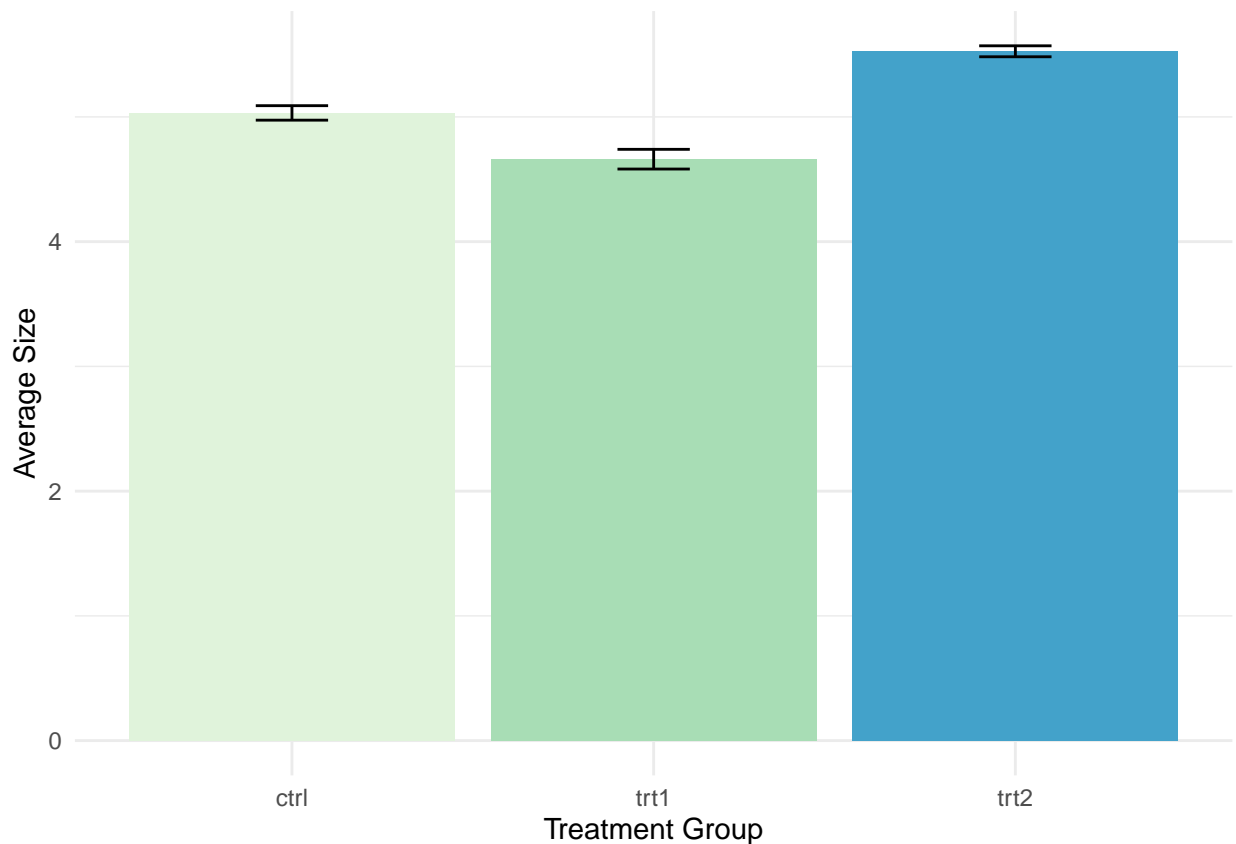
```

, c("TrmtGroup", "Mean", "Std.Dev", "Std.Err"))

## Create a bar chart to compare the different treatment groups
bar_plant = ggplot(plant_summary
, aes(x=TrmtGroup, y=Mean, fill=TrmtGroup)) +
  geom_bar(stat="identity") +
  guides(fill=FALSE) + xlab("Treatment Group") +
  ylab("Average Size") +
  geom_errorbar(aes(ymin=Mean-Std.Err, ymax=Mean+Std.Err), width=.2) +
  theme_minimal() + scale_fill_brewer(palette="GnBu")

bar_plant

```



6. Run a t-test to compare the means of the control group (“ctrl”) and treatment group 1 (“trt1”) in the PlantGrowth data. Report the observed value of t, the degrees of freedom, and the p-value associated with the observed value. Assuming an alpha threshold of .05, decide whether you should reject the null hypothesis or fail to reject the null hypothesis. In addition, report the upper and lower bound of the confidence interval.

```

## Inferential test
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
```

`t.test()` invokes the “Student’s t-Test”. We analyzed the difference in the results between control group $n=10$ and treatment 1 $n=10$. Results showed mean difference of -1.629, indicating that on average “trt1” reach the weight before the “ctrl” gr.

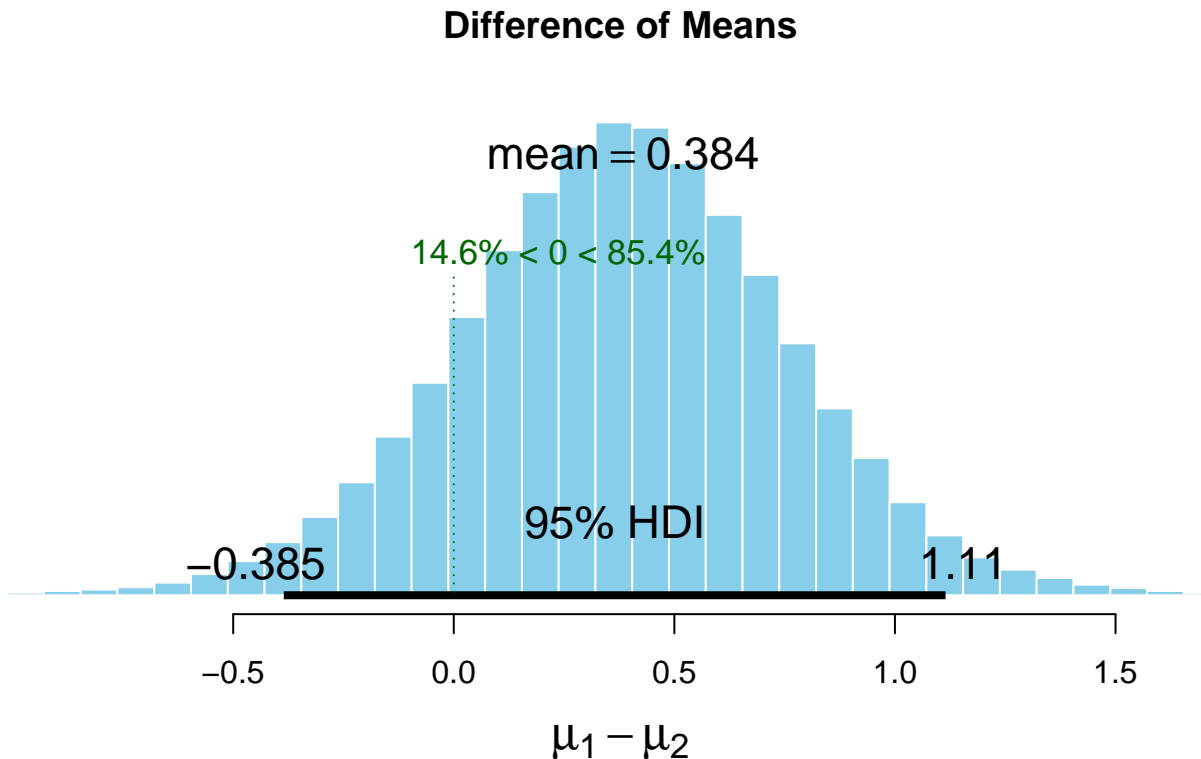
The t-test has used the two samples (ctrl and trt1), to calculate confidence interval ranging from a mean difference of -0.29 (lower end) to 1.03(upper end). We constructed 95% confidence interval around this mean difference. 95% chance that the confidence interval would contain pop mean difference in the long run (**This confidence interval may or may not contain the true population value.**) The width of the confidence band, about plus minus 0.37, gives some indication of the amount of uncertainty around the point of estimate(0.66). To reduce this uncertainty, we would have to increase sample size, reduce variability in within groups or both.

When t-test calculate the value of $t=1.1913$ from the sample data, it then position that value of t on the theoretical distribution of t values appropriate for the combined size of the two samples. Each of size $n=10$, the appropriate distribution is on about 17 degrees of freedom. The position of the observed t -value on the t -distribution divides the distribution into regions(quantiles). $T.p=0.25$ represents all of the area in the tails of the distribution, beyond the observed t -value(the probability of obtaining a value of t at least as high as what was actually observed).

We have to assert a null hypothesis: there is no mean difference between the means of the two groups. The opposite(there is difference) will be alternative hypothesis. Lets assume alpha level of .05. In reference to this alpha, p level shown above is $p=0.25$, noticeable larger than the alpha threshold. Based on the conventional rules of the Null Hypothesis Significance Test, we **fail to reject the null hypothesis**. Failing to reject the null hypothesis does not mean that we accept the null hypothesis, rather that we have no good evidence either way; likewise the p -value does not inform the question of how likely the null hypothesis is.

7. Install and library() the BEST package. Note that you may need to install a program called JAGS onto your computer before you try to install the BEST package inside of R. Use `BESTmcmc()` to compare the PlantGrowth control group (“ctrl”) to treatment group 1 (“trt1”). Plot the result and document the boundary values that `BESTmcmc()` calculated for the HDI. Write a brief definition of the meaning of the HDI and interpret the results from this comparison.

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



The Bayesian analysis shows that the “Highest Density Interval” or HDI ranges from -.37 up to 1.13(lb).

We can interpret HDI – there is a 95% probability that the population mean difference between the two groups falls within this range. The histogram shows that the most likely mean difference value is .38(lb). 14.4% of the mean differences in the distribution were negative and 85.6% positive(meaning that trt2 was more efficient). The mu symbol indicates population mean.

In conclusion: population mean difference is somewhere near .38lb, with 95% HDI ranging from -.37lb to 1.13lb. The likelihood of a population mean difference of 0 or larger is 85.6%.

8. Compare and contrast the results of Exercise 6 and Exercise 7. You have three types of evidence: the results of the null hypothesis test, the confidence interval, and the HDI from the BESTmcmc() procedure. Each one adds something, in turn, to the understanding of the difference between groups. Explain what information each test provides about the comparison of the control group (“ctrl”) and the treatment group 1 (“trt1”).

HDI show us the likely position of the population value, CI - shows uncertainty(-0.28 to 1.02). The highest density interval showed graphically where the mean difference in weight was more likely to lie: a region surrounding .38lb and favoring the weight gain of trt1. The results seem similar to the idea of CI but there is difference: the Bayesian output showed a distributional model of population parameter of interest, while the CI provided just a single interval estimate of that parameter with no guarantee that the particular CI constructed, actually contained the population parameter. HDI gives us information about the posterior distribution of possible value of the mean difference, CI- long run possibilities, not about the accuracy of this particular CI.

The NHST uses a model that assumes zero difference between the two populations and seeks to find results so extreme as to make assumption improbable. When the null hypothesis is rejected $p > \alpha(.05)$, we consider evidence in favor of an alternative hypothesis, though the results of that significance test do not say what the alternative hypothesis might be or the probability that might be correct.

We should always look for additional evidence after we see that the results meet the minimum bar.

9. Using the same `PlantGrowth` data set, compare the “ctrl” group to the “trt2” group. Use all of the methods described earlier (t-test, confidence interval, and Bayesian method) and explain all of the results.

```
## Inferential test
```

```
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:
```

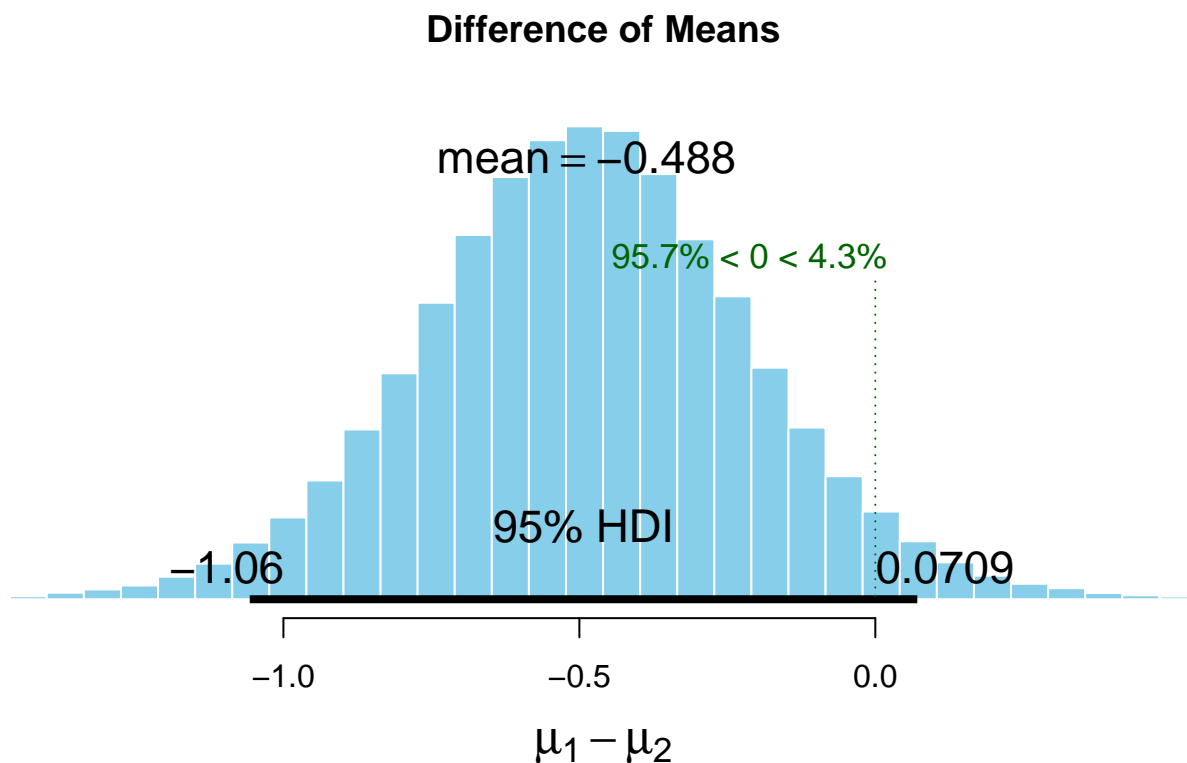
```
## mean of x mean of y
```

```
## 5.032 5.526
```

```
## BEST
```

```
plantBest1 <- BESTmcmc(PlantGrowth$weight[PlantGrowth$group=="ctrl"],  
                      PlantGrowth$weight[PlantGrowth$group=="trt2"])
```

```
plot(plantBest1)
```



The difference in the weight results between “ctrl” group $n=10$, and “trt2” $n=10$ was analyzed. Results showed mean difference of 0.49lb, indicating that on average “trt2” reach the weight before the “ctrl” gr.

The t-test has used the two samples (ctrl and trt2), to calculate confidence interval ranging from a mean difference of -0.982 (lower end) to -0.005(upper end). We constructed 95% confidence interval(CI) around this mean difference. (**This confidence interval may or may not contain the true population value.**). The width of the confidence band, about plus minus 0.49lb, gives some indication of the amount of uncertainty around the point of estimate (-0.49). To reduce this uncertainty, we would have to increase sample sizes, reduce variability in weight within groups or both. We are observing 16.8 degree of freedom (2 samples, each $n=10$). The position of the t-value (-2.134) on the t-distribution divides the distribution into regions. The p-value represent all of the area in the tails of the distribution, beyond the observed t-value – the probability of obtaining a value of t at least as high as what was actually observed. If we assume alpha is equal to .05, the p-value from the t-test is smaller alpha, but really close to it. The bigger the difference, the more convincing the results are. Still we can reject the null hypothesis, stating that there is no mean difference between the mean of the two groups. That rejection can be considered evidence in favor of some unspecified alternative hypothesis(opposite to the null).

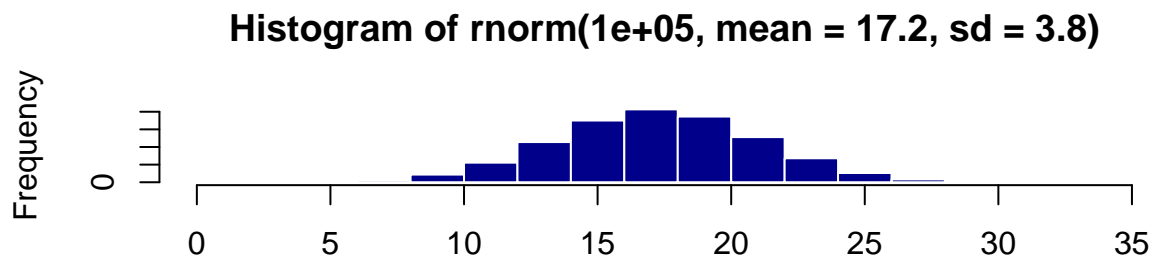
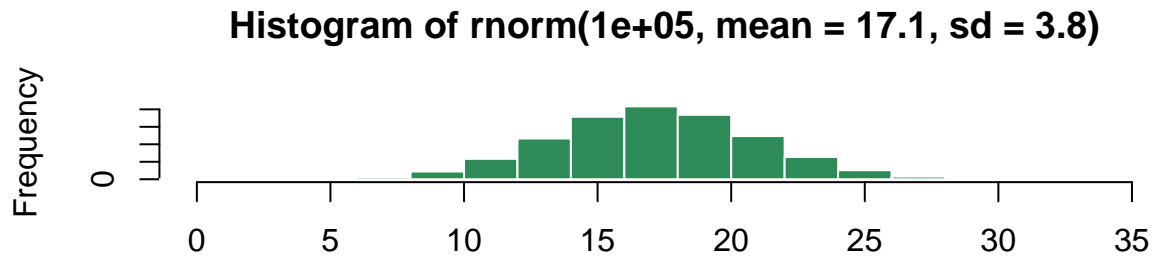
The BESTmcmc() function computes a probability distribution for the mean difference between the two groups, using the full information available in the two samples of plant data. The bell-shaped histogram shows at its peak the most highly likely mean difference value as -0.489 95.9% < 0 < 4.1% - shows the proportion of mean difference that were negative versus the proportion that were positive. We observed that 95.9% of the mean differences in the distribution were negative (meaning that “trt2” was more efficient) and 4.1% were positive. The chances “ctrl” group is equal to or better than “trt2” are really close to 0. We can also see under the plot that Bayesian approached used samples of the two groups to provide a model of the difference in population means. HDI here is between -1.04 and 0.08 (HDI is explained as the 95% probability that the population mean difference between the two groups falls within this range). According to this reasoning 95% of the likely values of the population mean difference lie in the bell-shaped area between -1.04 and 0.08(the greatest likelihood is around -0.49). The likelihood of a population mean difference of 0 or larger is 4.1%. In contrast to HD, HDI is build gradually from more than 100,000 steps , with each step depicting a possible combination of the population parameter.

10. Consider this t-test, which compares two groups of $n = 100,000$ observations each:

```
t.test(rnorm(100000,mean=17.1,sd=3.8),rnorm(100000,mean=17.2,sd=3.8))
```

For each of the groups, the rnorm() command was used to generate a random normal distribution of observations similar to those for the automatic transmission group in the mtcars database (compare the programmed standard deviation for the random normal data to the actual mtcars data). The only difference between the two groups is that in the first rnorm() call, the mean is set to 17.1 mpg and in the second it is set to 17.2 mpg. I think you would agree that this is a negligible difference, if we are discuss- ing fuel economy. Run this line of code and comment on the results of the t-test. What are the implications in terms of using the NHST on very large data sets?

```
par(mfrow = c(2,1))
#par(mar =c(1,1,1,1))
hist(rnorm(100000,mean=17.1,sd=3.8),
     xlim = c(0,35),
     col = "seagreen",
     border = "white",
     xlab = NA)
hist(rnorm(100000,mean=17.2,sd=3.8),
     xlim = c(0,35),
     col = "darkblue",
     border = "white",
     xlab = NA)
```



```
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 = -5.6584, df = 2e+05, p-value = 1.53e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.12956776 -0.06289992
## sample estimates:
## mean of x mean of y
## 17.10407 17.20030
```

Small confidence interval - small band of uncertainty. With large sample sizes we have small standard error. In large data sets almost every difference is statistically significant.

The results showed a mean difference of 0.1. The 95% CI is really small - lower bound: -0.13 and upper bound: -0.06. That confidence interval may or may not contain the true population value. The width of the confidence band shows only plus, minus 0.035, which is the amount of the uncertainty around the point estimate of -0.09. That might be due to the big sample sizes of the samples (we reduce uncertainty by increasing sample sizes). p-value is really small, way smaller than .05 so we can reject the null hypothesis (the difference between the two means is 0). So under the assumption of the null hypothesis and with $df = 199989$, the probability of observing a value of t with an absolute value greater than or equal to -5.3522 is $8.7e-08$.

There are some problems that statisticians encountered working with NHST. One of them is large data sets. Once you have more than about $n=1000$ observation (like our example with `rnorm`), the small difference in the means is statistically significant.

Another one is “effect size” :refers to the strength/magnitude of the statistical findings. Some effect sizes appear on scale such as 0 to 1, so they can be compared in different analysis. The point is, the differences between group means have to be maximized.

```
cohen.d(rnorm(100000,mean=17.1,sd=3.8),rnorm(100000,mean=17.2,sd=3.8))
```

```
##
## Cohen's d
##
## d estimate: -0.02985324 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.03861900 -0.02108747
```

Cohen's d showed really small effect size of -0.021.