Week 6 - Midterm Exam

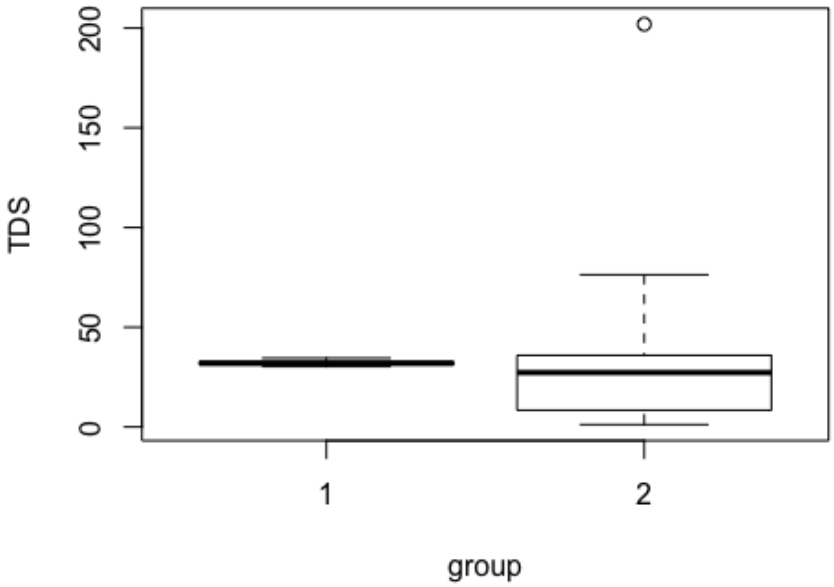
General Instructions: This is an honor system exam that is open book and open notes. You may consult any of the feedback I have provided to you on homework or practice exams. You may not confer or collaborate with any human besides me.

Problem Scenario: A startup company has developed an inexpensive and environmentally friendly biofilm to remove dissolved solids in water treatment plants. TDS is an abbreviation that refers to “total dissolved solids” and is measured in parts per million (PPM). Lower TDS is better – it means the water is cleaner.

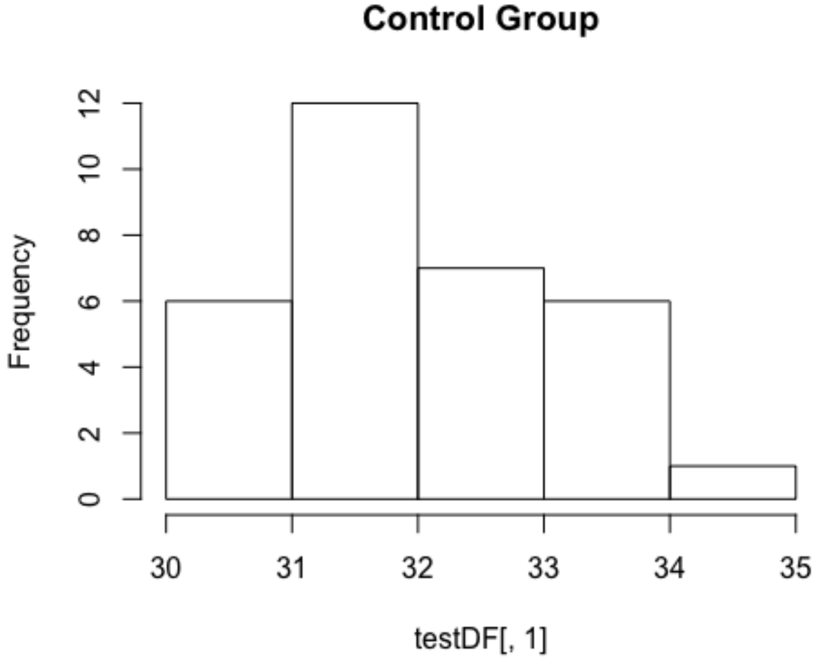
The startup conducts a comparison of batches of dirty water with and without their new treatment. The control group contains batches of water processed using industry standard mechanical filtering methods. The treatment group contains batches of water filtered with the biofilm. The research (alternative) hypothesis is that the mean TDS in the treatment group will be lower than the mean TDS in the control group. Specially calibrated, highly sensitive devices are used to measure TDS, so each control and treatment batch costs a lot of money to run.

The company will not release the raw data because they consider it a trade secret, but they have provided the following statistical outputs for you. Your job is to produce a report that will guide their biologists and investors on the next steps for this project. As such, the company wants you to evaluate the research hypothesis and write an interpretation of it that can be understood by non-statisticians. Here’s the output that they provided to you. You can feel free to cut and paste any of the graphics that appear below into your report, as appropriate for the audience:

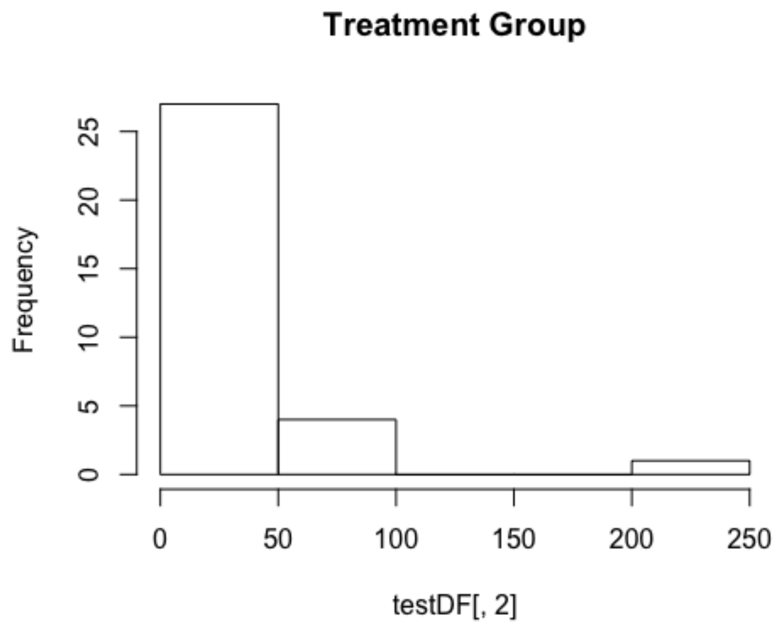
```
> str(testDF)
'data.frame':   32 obs. of  2 variables:
 $ Control  : num  32.8 33.2 30.3 33.2 31.9 ...
 $ Treatment: num  35.1 16.5 31.6 11.8 201.9 ...
> summary(testDF)
    Control        Treatment
 Min.   :30.27   Min.   :  1.178
 1st Qu.:31.25   1st Qu.:  8.501
Median :31.90
Mean   :32.05
3rd Qu.:32.81
Max.   :34.53   Max.   :201.908
Median : 27.259
Mean   : 31.079
3rd Qu.: 35.533
> boxplot(list(testDF[,1],
testDF[,2]),ylab="TDS",xlab="group")
```



```
> hist(testDF[,1], main="Control Group")
```



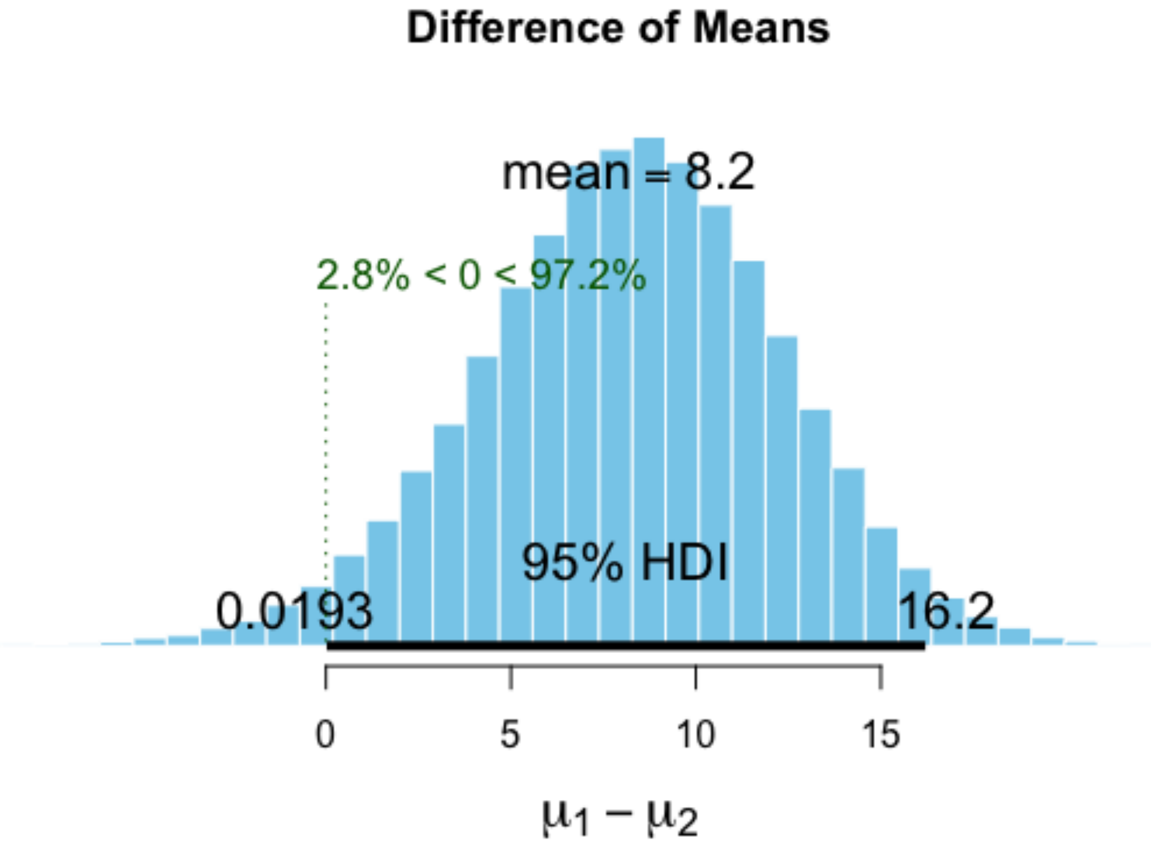
```
> hist(testDF[,2], main="Treatment Group")
```



```
> t.test(x=testDF[1],y=testDF[2])
     Welch Two Sample t-test
data:  testDF[1] and testDF[2]
t = 0.14925, df = 31.048, p-value = 0.8823
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 -12.35187  14.30250
sample estimates:
mean of x mean of y
 32.05410  31.07879
> bestOut <- BESTmcmc(y1=testDF[,1],y2=testDF[,2])
Waiting for parallel processing to complete...done.
> print(bestOut)
MCMC fit results for BEST analysis:
100002 simulations saved.
   mean     sd  median   HDIlo  HDIup  Rhat n.eff
32.0202 0.1969 32.0183 31.6371 32.409 1.000 57403
23.8196 4.1250 23.7048 15.7920 31.971 1.000 47821
mu1
mu2
nu
sigma1  0.9308 0.1582  0.9182  0.6327  1.246 1.000 39628
sigma2 19.5777 4.1796 19.1006 12.2217 28.019 1.001 20224
```

```
4.9958 2.9804  4.2830  1.5269 10.157 1.012  9307
```

```
'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(bestOut)
```



Report Components: Make sure your report includes all of the following elements.

1. (1 point) What are the lower bound and upper bounds of the (frequentist) 95% confidence interval of the mean difference?

The lower and upper bounds of the frequentist 95% confidence interval of the mean difference is as follows:

-12.35187 and 14.30250

2. (1 point) What is the point estimate of the mean difference?

   The point-estimate is approx. 0.98 (32.05410 - 31.07879).

3. (1 point) Report the outcome of the null hypothesis significance test on the difference of means. Make sure to state the null hypothesis.

   The null hypothesis indicates that there is no significant difference between the two comparison groups. A t-value is observed to be 0.14925 while the degrees of freedom value is observed to equal 31.048; lastly, the p-value = 0.8823. Since this p-value is significantly higher than a typical/conventional academic alpha value of 0.05, conductors of this t-test/NHST can conclude that they fail to reject the null-hypothesis, meaning that there is no credible evidence, from this test, that indicates a significant mean difference between the two groups.

4. (1 point) Report the lower and upper bounds of the 95% Highest Density Interval for the difference of means.

   The lower and upper bounds of the 95% HDI (Highest Density Interval) is as follows: 0.0193 and 16.2.

5. (1 point) Report the percentages of values in the posterior distribution of mean differences that are above zero and below zero.

   The percentage value in the posterior distribution of the mean difference that is above zero is 97.2% while the percentage value below zero is 2.8%.

6. (5 points) Write a 1-2 paragraph technical report. The technical report should contain the detailed information that it would be *important for other statisticians to know* about the data, about the analytical results, about any anomalies you observed, and about how any such anomalies may have affected the reported results. You can cut and paste any of the graphics included above, as long as you provide a 2-3 sentence explanation of what the graphic means.

   Ultimately, two groups of data were compared in hopes to shed light on possible differences on which group, if any, performed better in the experiment. From a technical perspective, both the Frequentist and Bayesian approaches were conducted.

Before the statistical tests were ran, a mere boxplot was constructed to visually observe the differences in the two groups. The control group was observed to not have widespread variability while the treatment group had significantly more variability (as indicated by the wider 'boxes' and 'whiskers'). By observing the boxplot, a clear difference could not be determined. When looking at the individual histograms, we see that some outliers occurred on the 200-250 range.

Moreover, the frequentist approach consisted of performing a Null Hypothesis Significance Test (NHST), in which a t-test was done. The findings indicated that a confidence interval of -12.35187 and 14.30250, which indicates that, if this test was conducted 100 times, we would expect the mean difference to reside in this range in 95 of those tests. A point-estimate, or the most likely difference value, of 0.98 (32.05410 - 31.07879) was observed. Since the confidence interval spans across zero, there is a possibility that there is no difference between the two observed groups of data. A t-value of 0.14925 was calculated while the degrees of freedom value came out to be 31.048.

With that said, we also observed a p-value from these calculations, which is drastically higher than a typical/conventional alpha comparison value of 0.05. This p-value equates to 0.8823, meaning that we fail to reject the null hypothesis. In other words, we cannot conclude that there is a significant difference between the two groups.

Additionally, the Bayesian perspective was also conducted within this experiment. All in all, the Highest Density Interval was conducted and an interval of 0.0193 and 16.2 was observed. Interestingly enough, we still see that the distribution still spans across 0, with 2.8% of the data showing below 0.

By using both the NHST and Bayesian/mcmc techniques, we can conclude that a possibility exists that there is no difference between the two groups despite evidence from the 95% HDI of the Bayesian test. Before we can conclude that the Bayesian results are significant, we would need to conduct a Bayes Factor test to see if the evidence for the alternative hypothesis is in excess of 150 to 1 odds (to be considered extremely strong evidence). Because we do not have access to this Bayes Factor test, we must revert back to the results from NHST and indicate that there is no clear, strong evidence for a difference between the two groups.

7. (5 points) Write a 1-2 paragraph report of the results of your analysis for presentation to the company's biologists and investors. *This report should be in plain language, interpretable by non-statisticians*. Make sure to integrate the Bayesian evidence, the frequentist confidence interval, and the results of the null hypothesis significance test. The biologists and investors need to decide what the startup should do next: The essential question they want to answer is whether or not the biofilm shows promise as an alternative to traditional filtering techniques. Use the results of these statistical analysis to provide them with guidance.

Ultimately, we cannot conclude that the alternative to traditional filtering is significantly different than the conventional method.

The NHST indicated that no significant difference was observed (due to the p-value equaling a value above our 0.05 alpha standard). Additionally, the confidence interval from our t-test experiment spanned across zero, which means the difference may be just that – zero. Moreover, the Bayesian test indicated that the interval also spanned zero (despite it being outside of the 95% HDI), with about 2.8% of probable differences showing a negative advantage for the control group. Further testing would need to be done to confirm our Bayesian tests.

To reiterate, at this time, there is not enough clear evidence to indicate one method is better than the other in terms of filtration techniques; however, group 1 seems to have less variability and can have a more concrete prediction of performance.