

Student name: Ryan Dean

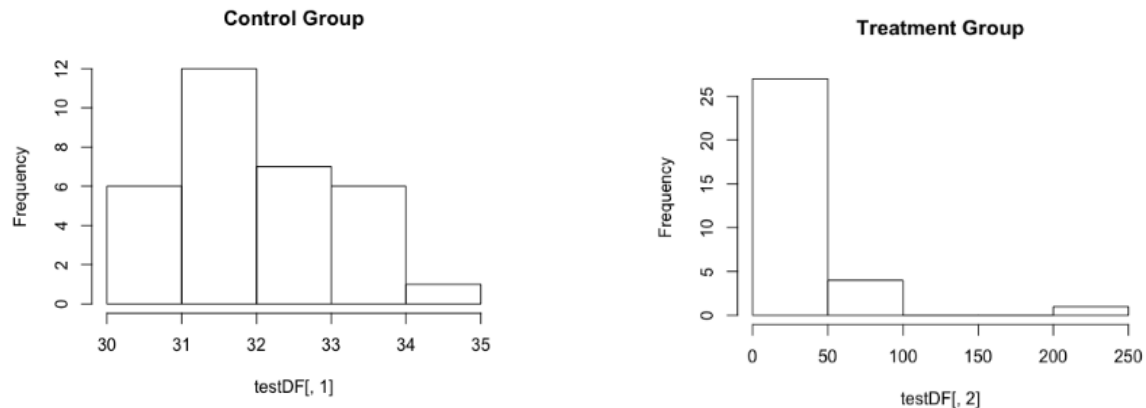
Midterm Exam

Date due: 5/24/21

1. The confidence interval ranges from -12.35187 to 14.30250.
2. The point estimate of the mean difference is .975315.
3. The null hypothesis is that the mean TDS in the biofilm treatment group is greater than or equal to the control group (Industry wide) TDS. The alternative hypothesis states that the mean in the TDS biofilm treatment group is less than the mean in the control group (industry wide). The large p-value of .8823 tells us we would not reject the null hypothesis as we cannot reject the null hypothesis at an alpha of .05. This is also apparent visually using the box plot, the majority of the datapoints between distributions overlap. (one tailed t-test) Would Still need to see a p-value $< .1$
4. The Bayesian HDI interval for difference in means is .0193 to 16.2.
5. The percentage values in the posterior distributions are $2.8\% < 0 < 97.2\%$.

6.

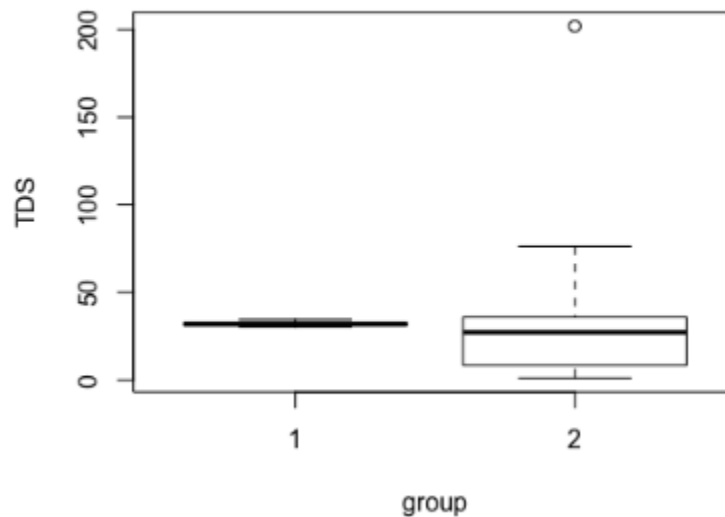
Ultimately from the sample data we have found that there is not a significant difference in water purity when using the biofilm water treatment. The t-value of 0.14925 is fairly low. The confidence interval ranges from -12.351 to 14.302. This confidence interval may not include the true population mean difference, but every 95/100 observations it should. The mean of the control group (water filtered by the industry standard) is 32.05, and the mean of the treatment (water treated by biofilm) is 31.079. According to the original t-test to find the difference of means the p-value was extremely high. The value of .8823 tells us that the observed difference could have just occurred by chance 88% of the time, which is not significant to an industry standard of an alpha of .05. Therefore, we would fail to reject the null hypothesis that the biofilm treatment group is actually less than the industry standard treated water, even though visually the mean of the biofilm group is lower, we can't prove it statistically with this dataset.



Within the histogram of the control group, we see a slight skew to the right. But the way it is scaled most datapoints lie between 30 and 35. When viewing the treatment group, there is a much more extreme skew to the right, however more than 78% of the datapoints are located between 0-50, with some outliers between 50-100 and even as large as 200-250. The outliers skew the biofilm TDS data significantly. While the HDI's mean is 8.2, showing a high probability of the sampled mean difference lying close to this value, it still overlaps zero (2.8%<0<97.2%), so some of the means suggest there is a mean difference of zero. The interval for the HDI was .0193 to 16.2 Unfortunately, with the size of the sample we are unable to make any definitive conclusions other than that it is not a significant difference, especially with the large p-value from the t-test.

7.

Ultimately from the sample data provided we found that there was no significant difference in the control group of industry standard filtered water and the group treated with bio film. Or, the TDS between both groups were not significantly different whereas you could not tell a difference in the biofilm's efficacy.



In the above boxplot the solid black line is called the “median” or middle value of the data, which would be TDS. The black box around the medians are the interquartile range, this is where approximately 50% of the data points lie. There is also much larger variation (data points distanced from the mean) in the biofilm group, this is represented by more whitespace in the box. The overlap of the median from the boxplots suggests the two-sample means aren’t different, but these require more precise testing, even though the median is visually lower in the biofilm.

In leu of boring you with the details, there have been formulas built and refined for giving us the likelihood of a significant difference, or in this case a difference between our control group and the treated group(that the TDS is actually lower in the biofilm group). They take the mean, or the average values in each group. Those means are then compared to each point in the dataset. With each of these differences we are able to calculate the likelihood the calculated differences are “statistically significant.” Unfortunately, the probability that the differences found were generated by random chance alone was very high. Therefore, we would not suggest that the biofilm improves the quality of the water in this test. In the future an experiment may benefit from adding another group, such as non-treated water to test its efficacy.