IST 772 Mid-Term Exam Key

All answers pertain to a study by the company comparing total dissolved solids between a control group and treatment group where the treatment was a new biofilm.

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

**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

The lower bound of the 95% confidence interval is -12.35187, the upper bound is 14.30250.

Question 2: What is the point estimate of the mean difference?

The point estimate is the point at the center of the confidence interval calculated as either:

([lower bound CI] + [upper bound CI]) / 2 (-12.35187 + 14.30250)/2 = 0.975315 or

[control sample mean] – [treatment sample mean]

32.05410 - 31.07879 = .97531

Question 3: Report the outcome of the null hypothesis significance test on the difference of means. Make sure to state the null hypothesis.

Null hypothesis: There is no mean difference between the control group the treatment group. t = 0.14925, p-value = 0.8823

Interpretation of results: The p-value is higher than the alpha that was set of .05, so we fail to reject the null hypothesis.

Question 4: Report the lower and upper bounds of the 95% Highest Density Interval for the difference of means.

Chart, histogram

Description automatically generated

The upper and lower bounds of the 95% Highest Density Interval (HDI) 0.0193 to 16.2 This means there is a 95% probability that the actual mean difference between the groups is between 0.0193 and 16.2.

Question 5: Report the percentages of values in the posterior distribution of mean differences that are above zero and below zero.



In the BEST test, 2.8% of observations fell below 0 while 97.2% were higher than 0.

Question 6: Technical Report

The company has done a study to compare the new biofilm product (treatment) to traditional mechanical filtering (control) to estimate whether the new film reduces total dissolved solids (TDS) in the filtered water that is the byproduct of the filtering process. Samples were taken from both the control and treatment groups and were measured using highly sensitive calibrated devices. As the costs of doing a sample run are high, the sample count (N) was limited to 32 from each population. A mean was calculated for TDS, calibrated in parts per million (ppm).

Confidence intervals, hypothesis testing, and Bayes simulation tests were run on the mean values collected for each sample run. The null hypothesis is that there is no difference in the means, the alternate hypothesis is there is a difference. It is a two tailed test strategy, looking for positive and negative difference between the sample sets. The objective of filtering is to have a lower TDS value. If the treatment group is estimated to have lower TDS, the mean difference would be positive.

**Summary Data on the Sample Groups**

> 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

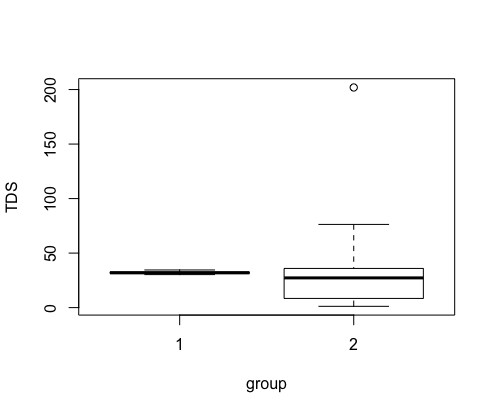
Min. :30.27 Min. : 1.178

1st Qu.:31.25 1st Qu.: 8.501

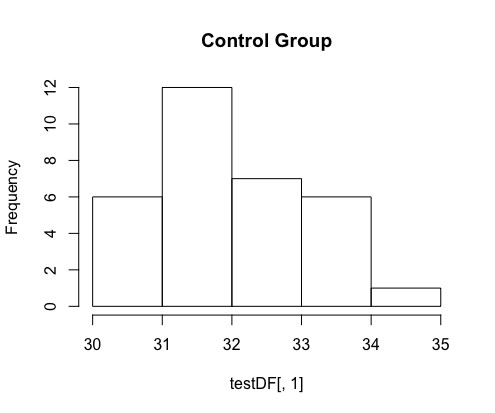
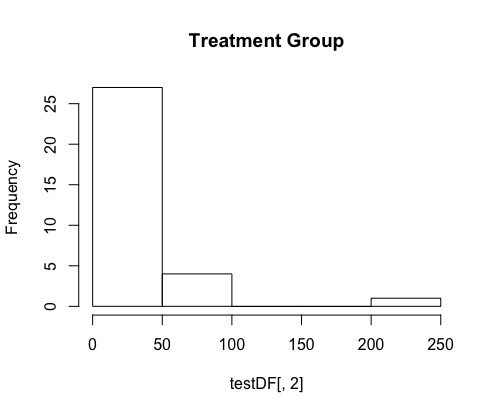
Median :31.90 Median : 27.259

Mean :32.05 Mean : 31.079

3rd Qu.:32.81 3rd Qu.: 35.533 Max. :34.53 Max. :201.908

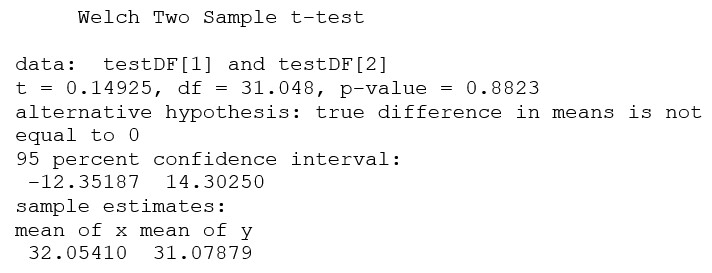


The summary data and box plots show that the control group had a visibly narrower distribution and no outliers. The treatment group was visibly more dispersed with an outlier that was 7.4 times higher than the median (201.9 ppm compared to the median of 27.26 ppm). A recommendation on the outlier will be made at the end. The following histograms show the difference in dispersion. Note that the X axis are on a different scale, the control group from 30 to 35 and the treatment group from 0 to 250. The Y-axis scales are also different, 0 to 12 vs. 0 to 25 for the treatment group.



The 95% calculated confidence interval between the two groups was -32.3519 to 14.30250 with a point estimate of 0.9753. Of note is that the confidence interval spans 0. The meaning of the confidence interval is that, if the experiment were repeated 100 times, the confidence interval in approximately 95 of those hundred tests would contain the population mean difference.

The null hypothesis statistical test (NHST) resulted in a tvalue of 0.14925 and a p-value of 0.8823. As our alpha was .05, a p-value higher than that would fail to reject the null hypothesis that there is no mean difference between the groups.



The final test on the two groups was a BEST Markov Curve Monte Carlo (MCMC) simulation of mean differences, using 100,002 observations. The central tendency for the observations was a mean of 8.2 with a high-density interval (HDI) range of 0.0193 to 16.2. This means, using Bayesian logic, there is a 95% probability that the mean difference between the control and treatment groups is between 0.0193 and 16.2. Of all observations, only 2.8% were less than 0 with 97.2% above. In isolation, this would suggest a high probability that the group means are different. In context with the NHST, there is a contradiction that makes the findings inconclusive.

Although expensive, the recommendation to the team is to rerun the experiment with new samples. If possible, the outlier should be investigated as being a potential issue with measuring or sampling. If outliers are found with either group deemed a result of normal variation, they should be removed before tests are performed.

Question 7: Business Report

The company has been developing a new biofilm for use in water filtration. This product will be an environmentally friendly alternative to mechanical filtration and is designed to be more effective at removing dissolved solids than the mechanical process. A team of scientists has performed an experiment that compares water filtered mechanically with water filtered with the biofilm. Sensitive equipment was used to measure dissolved solids to parts per million and an average was determined for samples of the mechanically filtered water and the biofilm treated water. Although an experiment using samples cannot prove if there is a difference, it can determine if there is a high probability there is a difference between the two methods. Since the objective is have lower numbers in the treatment group, then a positive difference in averages would be the desired outcome.

Three types of tests were performed on the two groups. One experiment measured uncertainty in the results by determining a range that might contain the actual value. In fact, if done 100 times, 95 of those times it would contain the actual value. The results showed a wide range of uncertainty that went from a difference of -12.35 to 14.30. The second test has a statistical test that provides a value to judge the reliability of a difference in average dissolved solids. The test returned a value that did not support there was a difference between the two groups.

The third test is a simulation that creates a collection of observations of a statistical process that tries to determine the average difference between the mechanical and biofilm groups. This test supports that there is a difference. In fact, the test showed a probability of 95% that the difference falls between 0.02 and 16.2 with an average estimate of 8.2. In isolation, this supports that the biofilm may provide better filtration. However, the second test had a contrasting position that suggests there is no difference.

The team of scientists recommends rerunning the tests, understanding that the company has made a significant investment in the current experiment. It is important that if the company brings the biofilm to market it can confidently support the improved filtration with statistical evidence. Before conducting the tests, the team will revisit how the measuring is performed including reviewing measuring equipment. The team feels that if it can narrow the range of measures, the goal of sufficient evidence on effectiveness will be achieved.