**Problem**

The stem length of soybeans from an experiment are:

20.2, 22.9, 23.3, 20.0, 19.4, 22.0, 22.1, 22.0, 21.9, 21.5, 20.9

1. Create a histogram to visualize the data.
2. Test "t.test" whether the population mean is different from 23 - finding p-value and discussing whether it is considered a "small" p-value.
3. Obtain a 2 sided 98% confidence interval on the true mean using "t.test".
4. The researcher, by using "t.test" on a sample size of 11 was assuming that the data was normally distributed.  Is that a valid claim?  Create a QQ plot and interpret.

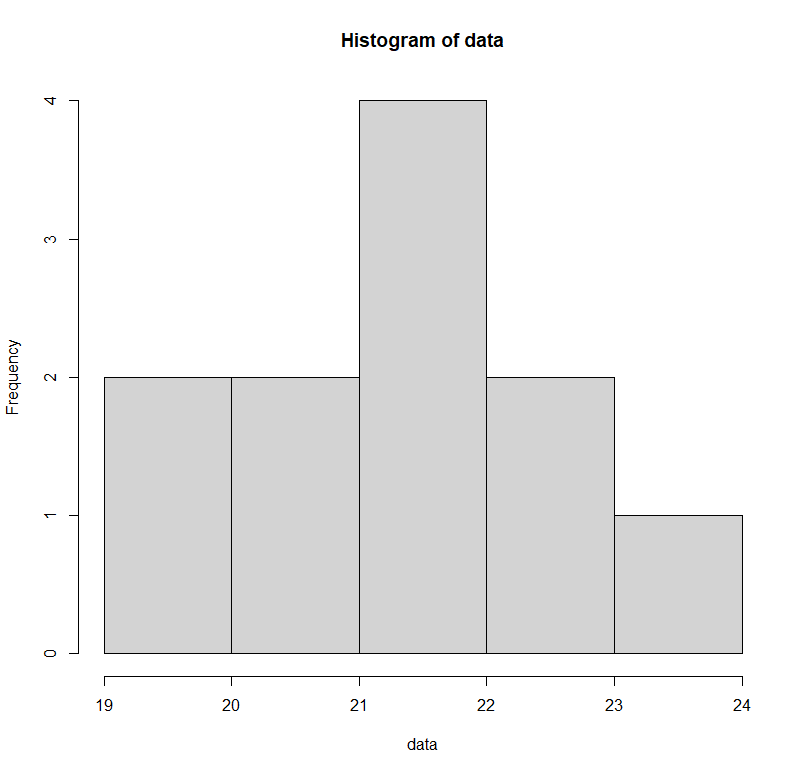
**Solution**

**Part (1) – Histogram**

Data is saved to the program and a histogram created using the following code:



The resulting histogram is below:



**Part (2) – t.test**

The code and results for t.test, run with a two sided alternative hypothesis and a population mean of 23 are shown below:



The p-value of the above test is 0.00199. This is significantly smaller than the typical significance value of 0.05, and therefore would be considered small. Additionally, the mean value for the sample is over four standard deviations away from the assumed mean, meaning that it is very unlikely that this sample would occur naturally with a population mean of 23.

**Part (3) – 98% CI t.test**

A 09% confidence interval can be obtained by adding the “conf.level” argument to the code used above. This and the results are shown below:



Per the above test the 98% confidence interval is 20.45480 - 22.49065.

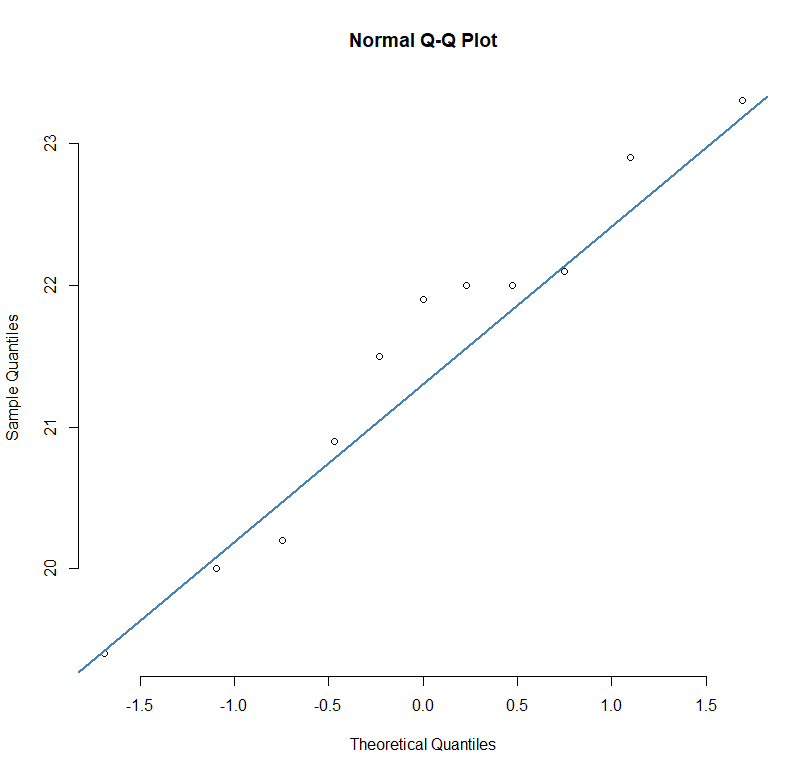
**Part (4) – Normality**

In general, we can assume that the underlying data is normally distributed if the sample size is greater than or equal to 30 or if the collected data is itself normally distributed. Our sample size in this problem is less than 30, so we will need to inspect the data using a QQ-plot.

Creating a QQ-plot from the existing data is done with the following code:



The resulting plot is shown below:



Most points in the plot are close to the QQ-line, though it is certainly far from ideal. Combining this with the distribution shown in the histogram plot earlier, the claim that the data is normally distributed is likely valid.

**Summary**

In this four-part exercise we created a histogram of the provided data, ran two t-tests to determine first the p-value of the data and a proposed population mean and then to find a 98% confidence interval, and finally we created a QQ-plot of the data and analyzed whether the assumption of normality is valid for this dataset.