R Assignment - Soybean Stem Length Analysis

Sean Parrell

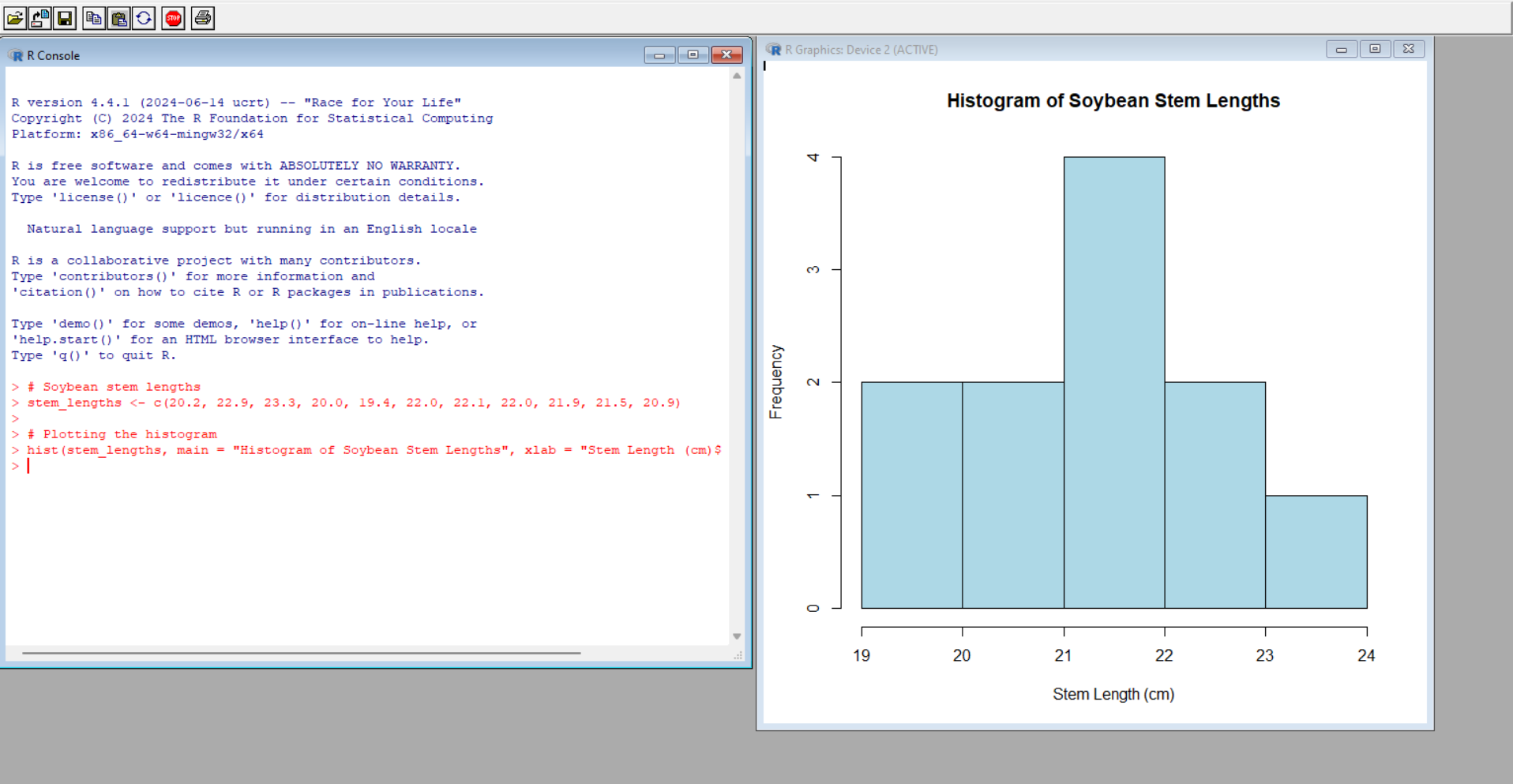
October 20, 2024

Problem Overview:

We are given the stem lengths of soybeans from an experiment, and our objective is to visualize the data, test if the population mean is different from 23, obtain a confidence interval, and assess the assumption of normality using a QQ plot.

1. Visualize the Data - Histogram

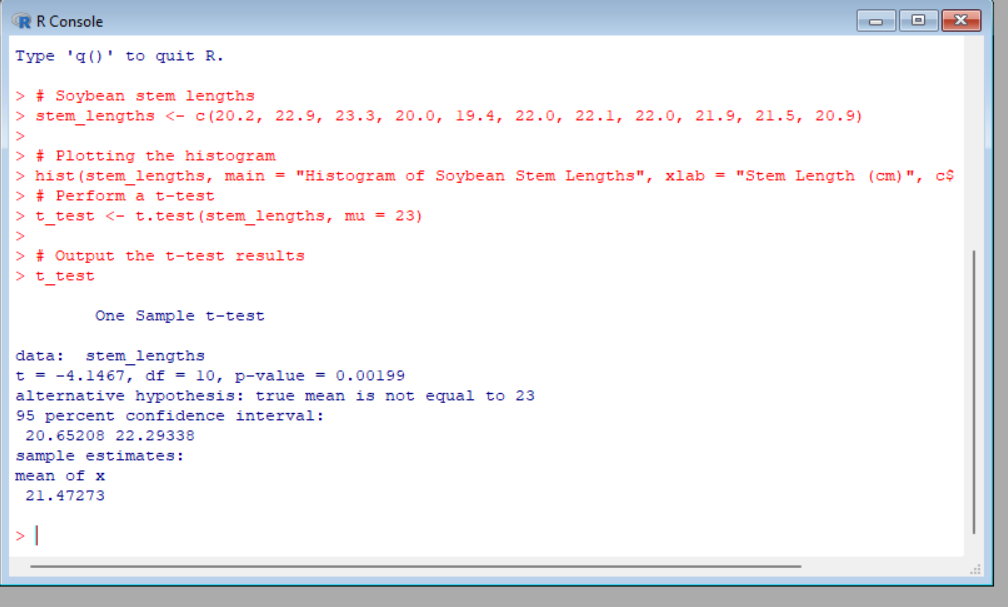
We will first create a histogram to visualize the distribution of the soybean stem lengths.



The histogram provides an overview of the frequency distribution of soybean stem lengths. We can observe whether the data appears normally distributed or skewed based on the shape.

2. T-Test for Population Mean ≠ 23

We will use a t-test to determine whether the population mean is significantly different from 23.

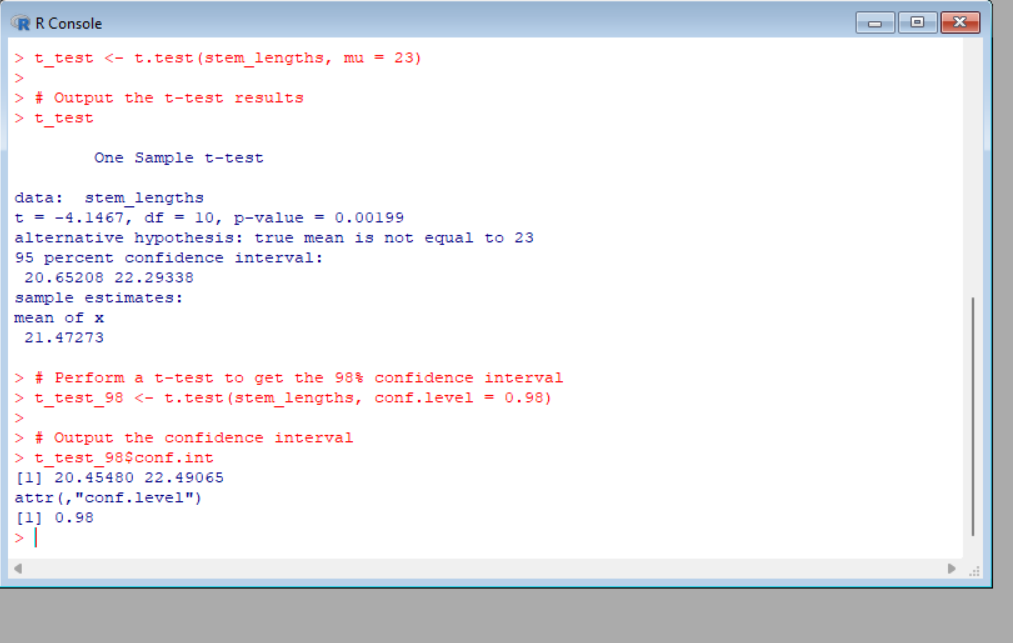


P-Value Interpretation:

Evaluating the p-value obtained from the t-test to determine whether it is considered "small" (typically a p-value < 0.05 would be small). The smaller the p-value, the stronger the evidence against the null hypothesis (population mean = 23).

3. 98% Confidence Interval on the True Mean

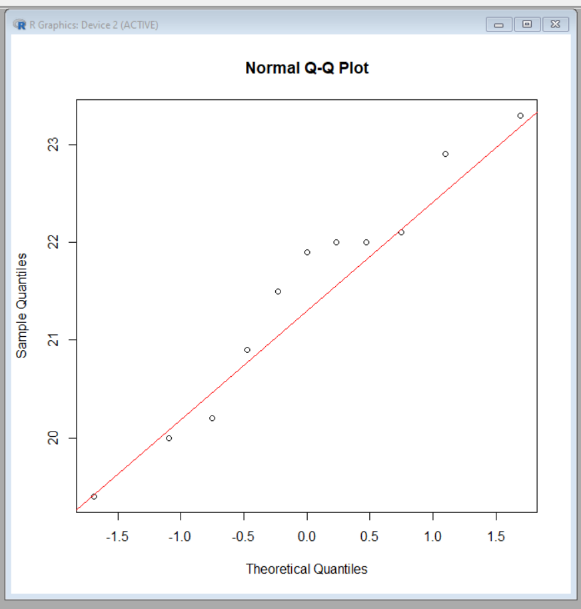
Next, we calculate a 98% confidence interval for the true mean stem length.



The results of the t-test indicate a t-value of -4.1467 with 10 degrees of freedom and a p-value of 0.00199, which is significantly smaller than 0.05. This provides strong evidence to reject the null hypothesis that the population mean is 23. Additionally, the 95% confidence interval (20.65208, 22.29338) and the 98% confidence interval (20.45480, 22.49065) both exclude 23, further supporting the conclusion that the true mean soybean stem length is significantly different from 23. Therefore, the data strongly suggests that the population mean differs from 23.

4. Normality Assumption - QQ Plot

The researcher assumed that the data was normally distributed. We will assess this by creating a QQ plot.



Interpretation: the QQ plot suggests that the assumption of normality is relatively valid for this data, but there may be some minor deviations, particularly in the tails of the distribution. This mild deviation is not necessarily problematic, especially with a small sample size (n = 11), but it's something to consider when interpreting the results of any statistical tests.