Exploratory Data Analysis, Confidence Intervals, and the t-test

Amartya Dutta

1 Problem 1:

1.1 Part A:

Histogram:

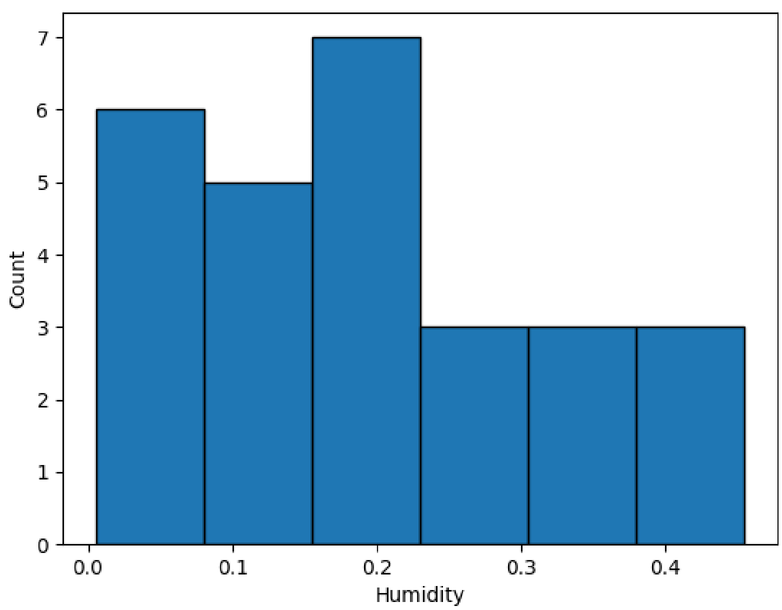

Figure 1: Histogram showing bins with respective Humidity values

Box Plot:

5 Number Summary:
Min: 0.005
Q1: 0.100
Median: 0.205
Q3: 0.282
Max: 0.455

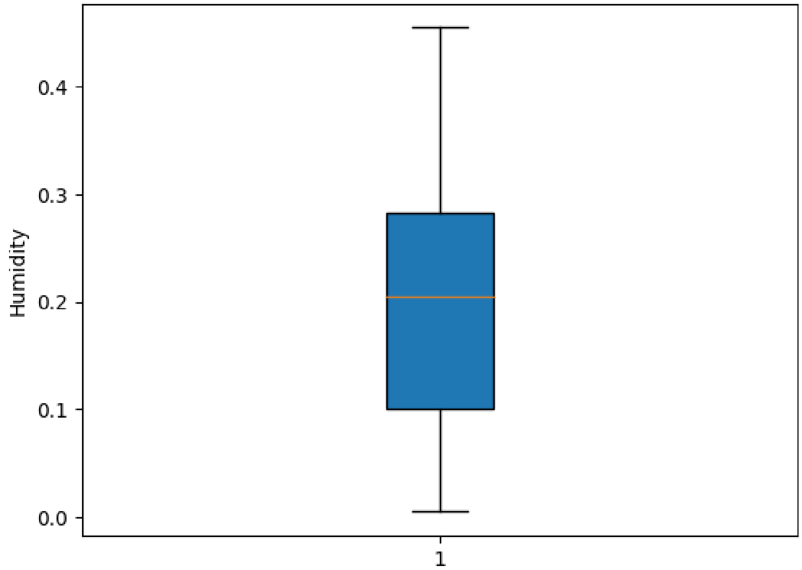

Figure 2: Box Plot showing the distribution of Humidity values

1.2 Part B:

Sample Mean: 0.198
Sample Median: 0.205
Standard Deviation: 0.130

2 Problem 2:

2.1 Part A:

Since the population is normally distributed, the sample has a Normal Distribution with a mean = 60, and a standard deviation $(\sqrt{\frac{5}{16}}) = 1.25$. This is in accordance with the central limit theorem.

2.2 Part B:

$$\alpha = 1 - 0.95 \tag{1}$$

$$\alpha = 0.05 \tag{2}$$

$$\alpha/2 = 0.025 \tag{3}$$

From JMP Software, Critical Value: t_{n-1}, $\alpha/2 = t_{15,0.025} = 1.96$
Confidence Interval = y ± Z · $\sigma/\sqrt{n}$

$$60 \pm (1.96 \cdot 1.25) \tag{4}$$

$$60 \pm 2.45 \tag{5}$$

$$(57.55, 62.45) \tag{6}$$

3 Problem 3:

3.1 Part A:

The random variable is the number of months it takes a new laptop battery to die. This is a continuous random variable whose parameter of interest is the mean. Our hypothesis, therefore, is about the mean. This tells us that this is a two-tailed test: if the battery's lifetime is 45 or not. The null and alternative hypotheses are thus:

$$H_0 : \mu = 45 \tag{7}$$

$$H_a : \mu \neq 45 \tag{8}$$

3.2 Part B:

Given:

Hypothesized Mean: $\mu_0 = 45$
Sample Mean $\bar{y}$=39.8
Sample STD $\sigma = 10.13$
Sample Size n = 5
Significance Level $\alpha = 0.1$

Using JMP software, we have the following test statistics results:
T score = -1.1478
P value = 0.315
Critical Values = +/- 2.1318
STD error of mean = 4.5303
The conclusion, therefore, is that it failed to reject Null Hypothesis. Since the p-value is greater than the significance level, the null hypothesis cannot be rejected. However, enough sample proof isn't present to say that mean of the true population differs from 45.

3.3 Part C:

Confidence Interval = 0.90
Using JMP software to calculate Confidence Intervals for One Mean, we get:
T score = 2.13185
Lower Limit = 30.01422
Upper Limit = 49.4578
Since the hypothesized mean, 45 is between the intervals, the null hypothesis cannot be rejected and thus it is consistent with the previous results from Part B.

4 Problem 4:

4.1 Part A:

According to the Rule of Thumb if the ratio of maximum standard deviation to minimum standard deviation in the two distributions is greater than 2 then the variances of the two distributions are different and we cannot use the pooled standard deviation for them.
Summary of Statistics:

Substrate Concentration = 1.5
Sample Mean = 7.1
Sample STD = 1.316561
N = 15

Substrate Concentration = 2.0
Sample Mean = 8.409091
Sample STD = 1.746922
N = 11
Max STD/ Min STD = $\frac{1.746922}{1.316561}$=1.3268

Since the ratio of max std/ min std < 2, we can assume that variances of both the distributions have a similar standard deviation and we can use the pooled standard deviation method to perform the t-test.

4.2 Part B:

Since the researchers expect to see a difference in the velocity, the hypothesis is as follows:

$$H_0 : \mu_{1.5} = \mu_2 \tag{9}$$

$$H_a : \mu_{1.5} \neq \mu_2 \tag{10}$$

Using the JMP software we eventually find the following:
T score = 2.1827
T Critical values = +/-2.0639
P value = 0.0391
Since the p-value is less than the significance level ($\alpha = 0.05$), we can reject the null hypothesis and conclude that changing concentration affects velocity.