CALIFORNIA STATE UNIVERSITY, SACRAMENTO

College of Engineering and Computer Science Department of
Civil Engineering

ENGR 115
Data Analysis Project

Nisha Joshi

Question 1:

Problem statement: Is there evidence to support the claim that the Voltage reading is on average, more than 12 V?

Test: The question involves testing only one sample, and since the claim is whether the reading is more than 12V, it is also one-tailed. A 95% confidence interval for the average number of voltage readings was constructed using $\alpha = 0.05$. This was chosen since a one-sample t-test is not available in Excel for data analysis.

Null Hypothesis: The average Voltage reading is 12V.

Alternative Hypothesis: The average voltage reading is more than 12V.

Descriptive Statistics: "Descriptive Statistics for voltage reading" retrieved by using Data Analysis Tool Kit add-in

Table 1.1: Voltage reading

Voltage (V)	
Mean	12.0686615
Standard Error	0.13820608
Median	12.107124
Mode	#N/A
Standard Deviation	1.38206084
Sample Variance	1.91009217
Kurtosis	-0.225888
Skewness	-0.042257
Range	6.8030904
Minimum	8.65502638
Maximum	15.4581168
Sum	1206.86615
Count	100
Confidence Level(95.0%)	0.27423086

Data Visualization: The figure below displays a histogram representing the frequency distribution of all voltage readings in the dataset. From the histogram, it is evident that the readings are concentrated around 12-13 volts, indicating that the average voltage is approximately in that range. This visualization helps provide a clear overview of the central tendency and spread of the voltage measurements.

Figure 1: Frequency of Voltage reading

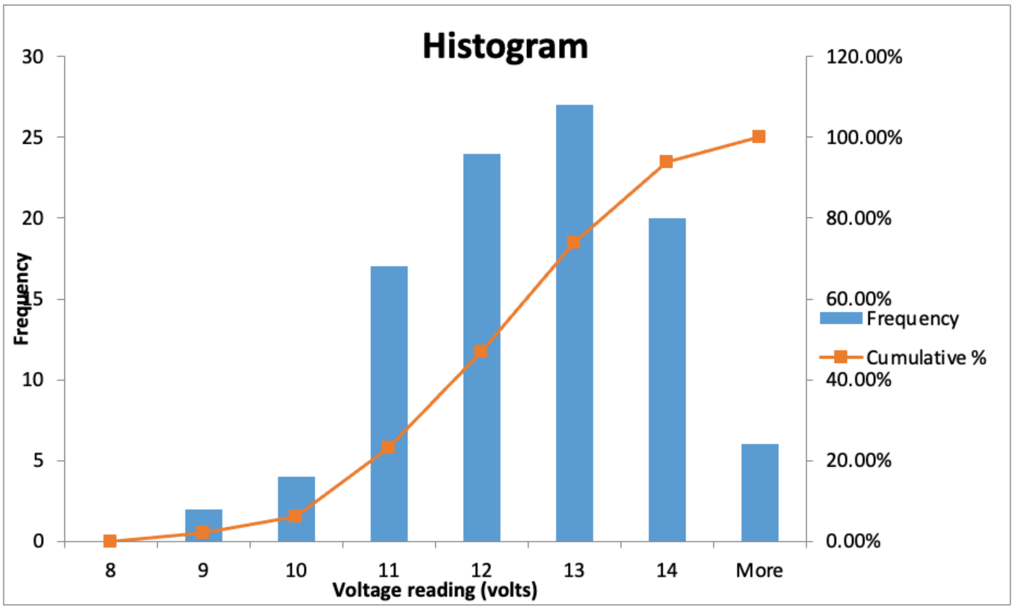


Table 1.2: Frequency table for voltage readings

Number	*Frequency*	*Cumulative %*
8	0	0.00%
9	2	2.00%
10	4	6.00%
11	17	23.00%
12	24	47.00%
13	27	74.00%
14	20	94.00%
More	6	100.00%

Table 1.2 illustrates the distribution of voltage readings across various value ranges. Most readings occur around **12-13 volts**, indicating that this is the most common voltage range in the dataset. The table supports the histogram by highlighting the central tendency of the data.

Observation/Calculations: In Table 1.3, the results of the t-test calculations are provided. The obtained one-tailed p-value (0.31021405)is compared against the level of significance (0.05). And since the p-value is greater than the level of significance, we fail to reject the null hypothesis.

Table 1.3: t-Test: Two-Sample Assuming Unequal Variances

	Voltage (V)	*null value*
Mean	12.0686615	12
Variance	1.91009217	0
Observations	100	2
Hypothesized Mean Difference	0	
df	99	
t Stat	0.49680485	
P(T<=t) one-tail	0.31021405	
t Critical one-tail	1.66039116	
P(T<=t) two-tail	0.6204281	
t Critical two-tail	1.98421695	

Table 1.4 P-value

P Value table	
Level of Significance	0.05
P-value	0.310214048
Test	p>0.05
Result	Fail to reject the null hypothesis

Conclusion: Fail to reject the null hypothesis. There is insufficient evidence to reject the null hypothesis previously stated- the average Voltage reading is 12 volts.

Question 2:

Problem statement: Is there evidence to support the claim that Temperature reading is on average, more than 24 °C?

Test: The question involves testing only one sample, and since the claim is whether the reading is more than 24 °C, it is also one-tailed. A 95% confidence interval for the average temperature readings was constructed using $\alpha = 0.05$. This was chosen since a one-sample test is not available in Excel for data analysis.

Null Hypothesis: The average Temperature reading is 24°C.

Alternative Hypothesis: The average Temperature reading is more than 24°C.

Descriptive Statistics: "Descriptive Statistics for temperature reading" retrieved by using Data Analysis Tool Kit add-in

Table 2.1: Descriptive statistics

Temperature (°C)	
Mean	24.72212044
Standard Error	0.193463704
Median	24.52555455
Mode	#N/A
Standard Deviation	1.934637043
Sample Variance	3.742820486
Kurtosis	0.733116926
Skewness	0.582401265
Range	10.8543
Minimum	20.63746398
Maximum	31.49176398
Sum	2472.212044
Count	100
Confidence Level(95.0%)	0.383873961

Data Visualization: The figure below displays a histogram representing the frequency distribution of all temperature readings in the dataset. From the histogram, it is evident that the frequency goes up after the temperature goes above 26 degrees Celsius, indicating that the average temperature is approximately in that range. This visualization helps provide a clear overview of the central tendency and spread of the voltage measurements.

Figure 1: Frequency of Temperature readings

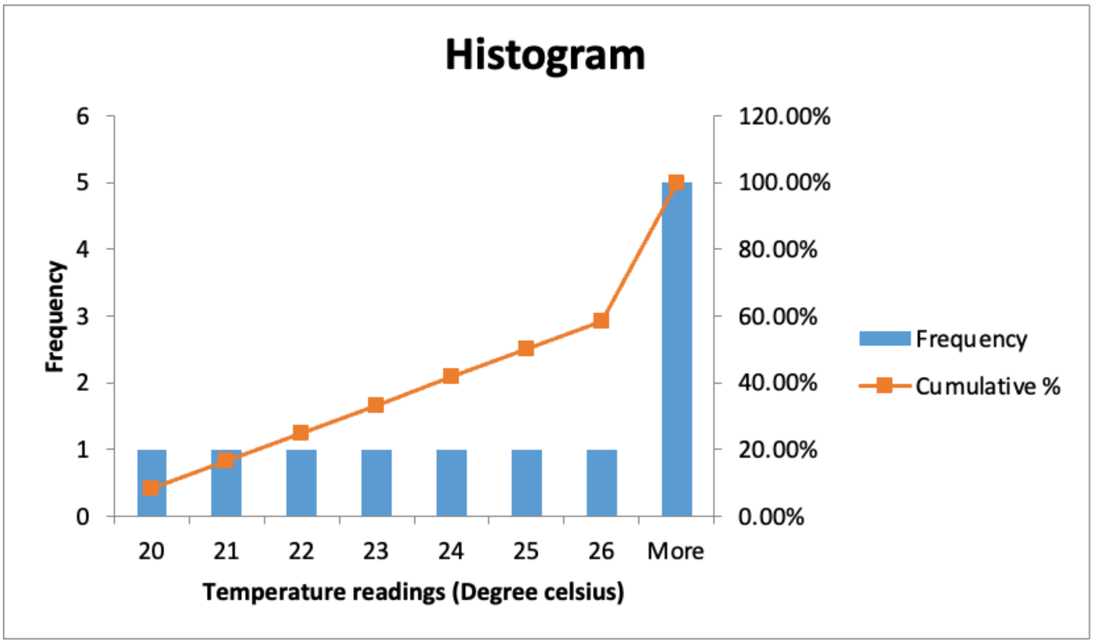


Table 2.2: Frequency table for voltage readings

Number	*Frequency*	*Cumulative %*
20	1	8.33%
21	1	16.67%
22	1	25.00%
23	1	33.33%
24	1	41.67%
25	1	50.00%
26	1	58.33%
More	5	100.00%

Observation/ Calculations: In Table 2.3, the results of the t-test calculations are provided. The obtained one-tailed p-value (0.31021405)is compared against the level of significance (0.05). And since the p-value is smaller than the level of significance, we reject the null hypothesis.

Table 2.3: t-Test: Two-Sample Assuming Unequal Variances:

	Temperature (°C)	*Null value*
Mean	24.72212044	24
Variance	3.742820486	0
Observations	100	2
Hypothesized Mean Difference	0	
df	99	
t Stat	3.732588728	
P(T<=t) one-tail	0.000158246	
t Critical one-tail	1.660391156	
P(T<=t) two-tail	0.000316491	
t Critical two-tail	1.984216952	

Table 2.4 P-value:

P Value table	
Level of Significance	0.05
P-value	0.000158246
Test	$p<0.05$
Result	Reject the null hypothesis

Conclusion: Reject the null hypothesis. There is strong evidence to reject the null hypothesis previously stated- the average temperature reading is 24 degrees celsius.

Question 3:

Problem statement: Is there evidence to support the claim that the Efficiency of Low Strength and High Strength products are different?

Test: The question involves testing two samples, and since the claim is whether the efficiency of low-strength and high-strength materials is different, it is also two-tailed. I will be performing a two-sample t-test, assuming unequal variances for the strength of the materials, with a significance level of 0.05.

Null Hypothesis: The efficiency of low and high-strength products is the same

$$(\mathbf{H_0}): \mu_{low} = \mu_{high}$$

Alternative Hypothesis: The efficiency of low and high-strength products is different

$$(\mathbf{H_1}): \mu_{low} \neq \mu_{high}$$

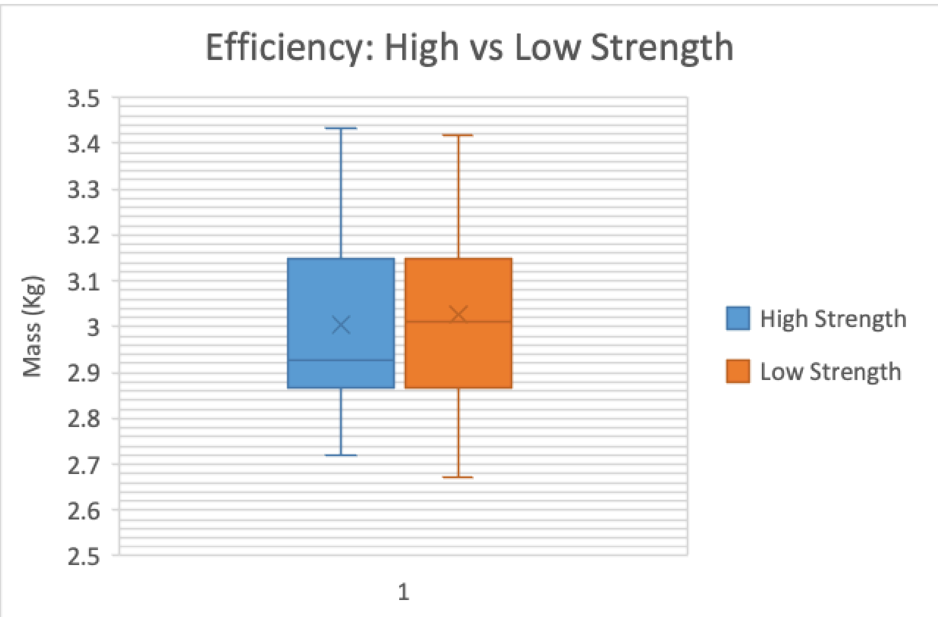
Descriptive Statistics: “Descriptive Statistics for the efficiency of strength of a material sample under tension” retrieved by using the Data Analysis Tool Kit add-in

Table 3.1: Descriptive Statistics for the efficiency of strength of a material sample under tension

	High Strength	Low. Strength
Mean	3.00258284	3.027898242
Standard Error	0.02651995	0.027476718
Median	2.92811377	3.011019735
Mode	#N/A	#N/A
Standard Deviation	0.18752439	0.194289737
Sample Variance	0.0351654	0.037748502
Kurtosis	-0.3635206	-0.700247048
Skewness	0.70336703	0.25812014
Range	0.71227345	0.744489947
Minimum	2.71991272	2.671116533
Maximum	3.43218616	3.41560648
Sum	150.129142	151.3949121
Count	50	50
Confidence Level(95.0%)	0.05329384	0.055216532

Data Visualization: In the figure below, we observe that the efficiency of both the Low Strength and High Strength products appears very similar. It is difficult to determine if one is consistently greater than the other based solely on visual inspection. However, the box-and-whisker plot provides useful insight into the spread and central tendency of each group, allowing us to infer that any differences in efficiency are likely minimal and may not be statistically significant.

Figure 1: Box & Whisker:- Efficiency: High vs Low Strength



Observation/ Calculations: In Table 3.2, the results of the t-test calculations are provided. The obtained two-tailed p-value (0.508934556) is compared against the level of significance (0.05). And since the p-value is greater than the level of significance, we can conclude that the null hypothesis cannot be rejected.

Table 3.2: t-Test: Two-Sample Assuming Unequal Variances

	High Strength	*Low Strength*
Mean	3.002582839	3.027898242
Variance	0.035165398	0.037748502
Observations	50	50
Hypothesized Mean Difference	0	
df	98	
t Stat	-0.662925438	
P(T<=t) one-tail	0.254467278	
t Critical one-tail	1.660551217	
P(T<=t) two-tail	0.508934556	
t Critical two-tail	1.984467455	

Table 3.3 P-value:

P Value table	
Level of Significance	0.05
P-value	0.508934556
Test	p>0.05
Result	Fail to reject the null hypothesis

Conclusion: Based on the visualized data and the t-test results, we conclude that the Low Strength and High Strength products do not differ significantly in efficiency. In other words, we fail to reject the null hypothesis.

Question 4:

Problem statement: Is there evidence to support the claim that the Mass of High Strength products is more than that of Low Strength products?

Test: The question involves testing two samples, and since the claim is whether the mass of high-strength products is more than low-strength products, it is also one-tailed. I will be performing a two-sample t-test, assuming unequal variances for the strength of the materials with a significance level of 0.05.

Null Hypothesis: The mass of low and high-strength products is the same

$$(\mathbf{H_0}): \mu_{low} = \mu_{high}$$

Alternative Hypothesis: The mass of high-strength products is more than low-strength products.

$$(\mathbf{H_1}): \mu_{high} > \mu_{low}$$

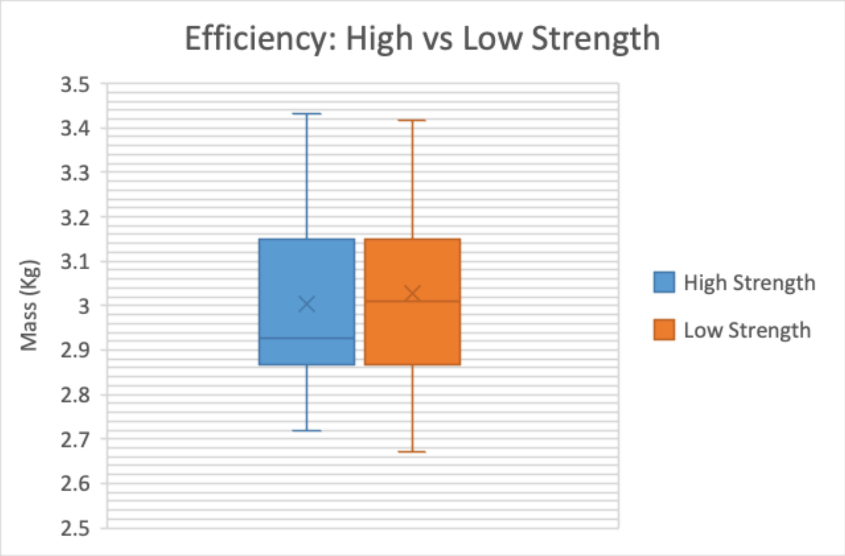
Descriptive Statistics: "Descriptive Statistics for the mass of strength of a material sample under tension" retrieved by using the Data Analysis Tool Kit add-in.

Table 4.1: Descriptive Statistics for the mass of low and high-strength products

	High	Low
Mean	2.00847267	1.85598215
Standard Error	0.03263818	0.04112836
Median	1.97008797	1.8535777
Mode	#N/A	#N/A
Standard Deviation	0.23078681	0.29082145
Sample Variance	0.05326255	0.08457712
Kurtosis	0.50882724	0.23045608
Skewness	-0.2195361	-0.1905317
Range	1.1453473	1.4580363
Minimum	1.35889043	1.05810844
Maximum	2.50423773	2.51614474
Sum	100.423634	92.7991075
Count	50	50
Confidence Level(95.0%)	0.06558889	0.08265054

Data Visualization: In the figure below, we observe that the mass of high-strength products is more than that of low-strength products. The median values of high-strength products lie above the median value of the low-strength products.

Figure 1:Box & Whisker:- Mass: High vs Low Strength



Observation/ Calculations: In Table 4.2, the results of the t-test calculations are provided. The obtained two-tailed p-value (0.002299048) is compared against the level of significance (0.05). And since the p-value is smaller than the level of significance, we reject the null hypothesis.

Table 4.2: t-Test: Two-Sample Assuming Unequal Variances

	High	*Low*
Mean	2.008472674	1.855982151
Variance	0.053262553	0.084577117
Observations	50	50
Hypothesized Mean Difference	0	
df	93	
t Stat	2.904295133	
P(T<=t) one-tail	0.002299048	
t Critical one-tail	1.661403674	
P(T<=t) two-tail	0.004598096	
t Critical two-tail	1.985801814	

Table 4.3: P-Value table

P Value table	
Level of Significance	0.05
P-value	0.002299048
Test	$p<0.05$
Result	Reject the null hypothesis

Conclusion: Based on the visualized data and the t-test results, we conclude that the mass of high-strength products is more than that of low-strength products. In other words, we reject the null hypothesis.

Question 5: Is there evidence to support the claim that the flow rate of Low Strength and High Strength products is different?

Test: The question involves testing two samples, and since the claim is whether the flow rate of low-strength and high-strength materials is different, it is also two-tailed. I will be performing a two-sample t-test, assuming unequal variances for the strength of the materials, with a significance level of 0.05.

Null Hypothesis: There is no difference in average flow rate between the two products.

$$(\mathbf{H_0}): \mu_{low} = \mu_{high}$$

Alternative Hypothesis: There is a difference in average flow rate.

$$(\mathbf{H_1}): \mu_{low} \neq \mu_{high}$$

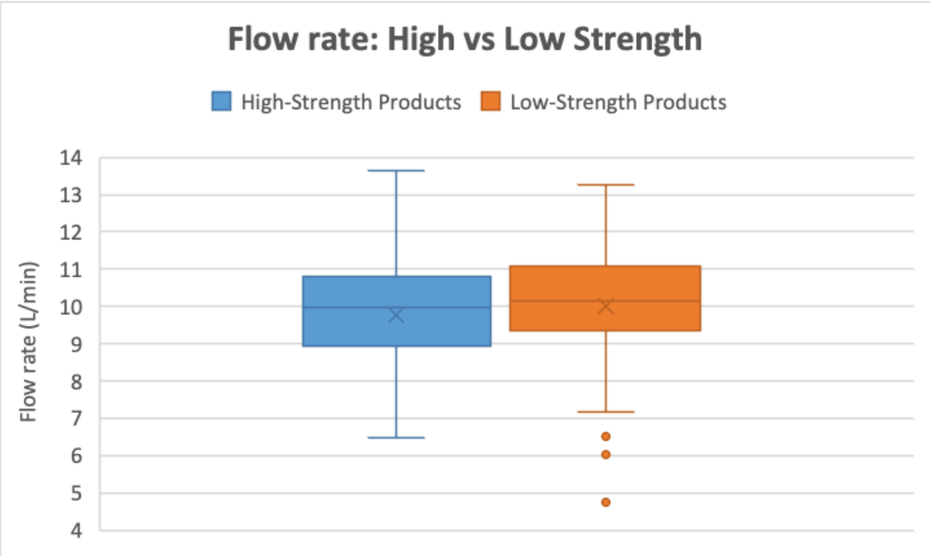
Descriptive Statistics: "Descriptive Statistics for the flow rate of high-strength and low-strength products." retrieved by using the Data Analysis Tool Kit add-in

Table 5.1: Descriptive Statistics for the flow rate of low and high strength products

Values	*High-Strength Products*	*Low-Strength Products*
Mean	9.776073484	10.00656734
Standard Error	0.209781198	0.243204947
Median	9.979482672	10.15705843
Mode	#N/A	#N/A
Standard Deviation	1.483377077	1.719718672
Sample Variance	2.200407552	2.95743231
Kurtosis	0.175564182	1.21632859
Skewness	-0.206523728	-0.819088045
Range	7.176939849	8.505589128
Minimum	6.476884091	4.764517929
Maximum	13.65382394	13.27010706
Sum	488.8036742	500.3283672
Count	50	50
Confidence Level(95.0%)	0.421571101	0.488738639

Data Visualization: In the figure below, we observe that the flow rate of low-strength products is slightly more than that of high-strength products, but it does not show any major differences.Though the median of low-strength products lies above the median of the high-strength products.

Figure 1:Box & Whisker:- Flow rate: High vs Low Strength



Observations/Calculations: In Table 5.2, the results of the t-test calculations are provided. The obtained two-tailed p-value (0.474718116) is compared against the level of significance (0.05). And since the p-value is greater than the level of significance, we can conclude that the null hypothesis cannot be rejected.

Table 5.2 t-Test: Two-Sample Assuming Unequal Variances

	High-Strength Products	*Low-Strength Products*
Mean	9.776073484	10.00656734
Variance	2.200407552	2.95743231
Observations	50	50
Hypothesized Mean Difference	0	
df	96	
t Stat	-0.71764628	
P(T<=t) one-tail	0.237359058	
t Critical one-tail	1.66088144	
P(T<=t) two-tail	0.474718116	
t Critical two-tail	1.984984312	

Table 5.3 P-Value

P-value table	
Level of Significance	0.05
P-value	0.474718116
Test	p>0.05
Result	Fail to reject the null hypothesis

Conclusion: Based on the visualized data and the t-test results, we conclude that the Low Strength and High Strength products do not differ significantly in efficiency. In other words, we fail to reject the null hypothesis.