1. Scenario: A company wants to analyze the sales performance of its products in different regions. They have collected the following data:

   Region A: [10, 15, 12, 8, 14]

   Region B: [18, 20, 16, 22, 25]

   Calculate the mean sales for each region.

Ans:

Region A:

Sales data: [10, 15, 12, 8, 14]

Mean = (10 + 15 + 12 + 8 + 14) / 5 = 59 / 5 = 11.8

Region B:

Sales data: [18, 20, 16, 22, 25]

Mean = (18 + 20 + 16 + 22 + 25) / 5 = 101 / 5 = 20.2

2. Scenario: A survey is conducted to measure customer satisfaction on a scale of 1 to 5. The data collected is as follows:

   [4, 5, 2, 3, 5, 4, 3, 2, 4, 5]

   Calculate the mode of the survey responses.

Ans: To calculate the mode of the survey responses, you need to find the value(s) that occur most frequently in the data. Here's how you can do it:

Survey responses: [4, 5, 2, 3, 5, 4, 3, 2, 4, 5] Count of each value:

* 2: 2 times
* 3: 2 times
* 4: 3 times
* 5: 3 times

Since both 4 and 5 occur the most frequently (3 times each), the mode of the survey responses is 4 and 5.

Note that in some cases, there can be multiple modes if multiple values have the same highest frequency.

3. Scenario: A company wants to compare the salaries of two departments. The salary data for Department A and Department B are as follows:

   Department A: [5000, 6000, 5500, 7000]

   Department B: [4500, 5500, 5800, 6000, 5200]

   Calculate the median salary for each department.

Ans:

Department A: [5000, 6000, 5500, 7000]

Arranging in ascending order: [5000, 5500, 6000, 7000]

Median: (5500 + 6000) / 2 = 5750

Department B: [4500, 5500, 5800, 6000, 5200]

Arranging in ascending order: [4500, 5200, 5500, 5800, 6000]

Median: 5500

Therefore, the median salary for Department A is 5750 and for Department B is 5500.

4. Scenario: A data analyst wants to determine the variability in the daily stock prices of a company. The data collected is as follows:

   [25.5, 24.8, 26.1, 25.3, 24.9]

   Calculate the range of the stock prices.

Ans:

To calculate the range of the stock prices, you need to find the difference between the maximum and minimum values in the dataset. Here's how you can calculate it for the given data:

Data: [25.5, 24.8, 26.1, 25.3, 24.9] Maximum value: 26.1 Minimum value: 24.8 Range: Maximum value - Minimum value = 26.1 - 24.8 = 1.3

Therefore, the range of the stock prices is 1.3.

5. Scenario: A study is conducted to compare the performance of two different teaching methods. The test scores of the students in each group are as follows:

   Group A: [85, 90, 92, 88, 91]

   Group B: [82, 88, 90, 86, 87]

   Perform a t-test to determine if there is a significant difference in the mean scores between the two groups.

Ans:

To perform a **t-test to determine if** there is a significant difference in the mean scores between Group A and Group B, you can follow these steps:

1. **Define** the **null *hypothesis (H0****)* and the **alternative hypothesis (Ha**):

- Null hypothesis (H0): There is no significant difference in the mean scores between Group A and Group B.

- Alternative hypothesis (Ha): There is a significant difference in the mean scores between Group A and Group B.

2. Calculate the **sample means** and **sample standard deviations** for each group:

- **Group A: mean\_A** = (85 + 90 + 92 + 88 + 91) / 5 = 88.4,

**standard deviation\_A** = sqrt(((85-88.4)^2 + (90-88.4)^2 + (92-88.4)^2 + (88-88.4)^2 + (91-88.4)^2) / 4) ≈ 2.607

- **Group B: mean\_B** = (82 + 88 + 90 + 86 + 87) / 5 = 86.6,

**standard deviation\_B** = sqrt(((82-86.6)^2 + (88-86.6)^2 + (90-86.6)^2 + (86-86.6)^2 + (87-86.6)^2) / 4) ≈ 2.607

3. **Calculate the t-value** using the formula:

**t = (mean\_A - mean\_B) / sqrt((standard deviation\_A^2 / n\_A) + (standard deviation\_B^2 / n\_B))**

t = (88.4 - 86.6) / sqrt((2.607^2 / 5) + (2.607^2 / 5))

t ≈ 0.768

4. Determine the **degrees of freedom (df)** using the formula:

df = n\_A + n\_B - 2 = 5 + 5 - 2 = 8

5. **Look up the critical t-value for the desired significance level and degrees of freedom**. For example, at a 95% confidence level with 8 degrees of freedom, the critical t-value is approximately 2.306.

6. **Compare** the **calculated t-value with the critical t-value**. If the calculated t-value is greater than the critical t-value, reject the null hypothesis; otherwise, fail to reject the null hypothesis.

In this case, the calculated t-value (0.768) is less than the critical t-value (2.306), so we fail to reject the null hypothesis. Therefore, based on the t-test, there is no significant difference in the mean scores between Group A and Group B.

6. Scenario: A company wants to analyze the relationship between advertising expenditure and sales. The data collected is as follows:

   Advertising Expenditure (in thousands): [10, 15, 12, 8, 14]

   Sales (in thousands): [25, 30, 28, 20, 26]

   Calculate the correlation coefficient between advertising expenditure and sales.

Ans:

Using the provided data:

Advertising Expenditure (in thousands): [10, 15, 12, 8, 14]

   Sales (in thousands): [25, 30, 28, 20, 26]

* Mean of the advertising expenditure =:

x̄ = (10 + 15 + 12 + 8 + 14) / 5 = 11.8

* mean of the Sales:

ȳ = (25 + 30 + 28 + 20 + 26) / 5 = 25.8

* Calculate the differences:

(x - x̄) = [10 - 11.8, 15 - 11.8, 12 - 11.8, 8 - 11.8, 14 - 11.8]

(y - ȳ) = [25 - 25.8, 30 - 25.8, 28 - 25.8, 20 - 25.8, 26 - 25.8]

* Calculate the sum of the products:

Sum of [(x - x̄) \* (y - ȳ)] =37.8

**Now using all this to calculate Cov(x,y)= Sum of [(x - x̄) \* (y - ȳ)]/n**

**=37.8/5**

=**7.56**

* Subtract the mean from each data point:

[(10 - 11.8), (15 - 11.8), (12 - 11.8), (8 - 11.8), (14 - 11.8)] = [-1.8, 3.2, 0.2, -3.8, 2.2]

[(25 - 25.8), (30 - 25.8), (28 - 25.8), (20 - 25.8), (26 - 25.8)] = [-0.8, 4.2, 2.2, -5.8, 0.2]

* Square the differences

[(-1.8)^2, (3.2)^2, (0.2)^2, (-3.8)^2, (2.2)^2] = [3.24, 10.24, 0.04, 14.44, 4.84]

[(-0.8)^2, (4.2)^2, (2.2)^2, (-5.8)^2, (0.2)^2] = [0.64, 17.64, 4.84, 33.64, 0.04]

**Now calculating standard deviation of both** :

**Standard Deviation of Advertising Expenditure**= sqrt((3.24+10.24+0.04+14.44+4.84)/5)= **2.60**

**Standard Deviation of sales**= sqrt((0.64+ 17.64+ 4.84+ 33.64+ 0.04)/5)= **3.34**

Now final **correlation coefficient** = **Cov(x, y)** / **( Standard Deviation of Advertising Expenditure\* Standard Deviation of sales)**

**=7.56/2.60\*3.34**

**=0.89**

7. Scenario: A survey is conducted to measure the heights of a group of people. The data collected is as follows:

   [160, 170, 165, 155, 175, 180, 170]

   Calculate the standard deviation of the heights.

Ans :

Mean of height = 160 + 170 + 165 + 155 + 175 + 180 + 170) / 7 = 167.86

Calculate squared difference of each height from the mean= (160-167.86)^2 +(170-167.86)^2 + (165-167.86)^2 + (155-167.86)^2 + (175-167.86)^2 + (180-167.86)^2 + (170-167.86)^2= 311.2576

Now standard deviation of height = sqrt(311.2576/7)= 7.169

8. Scenario: A company wants to analyze the relationship between employee tenure and job satisfaction. The data collected is as follows:

   Employee Tenure (in years): [2, 3, 5, 4, 6, 2, 4]

   Job Satisfaction (on a scale of 1 to 10): [7, 8, 6, 9, 5, 7, 6]

   Perform a linear regression analysis to predict job satisfaction based on employee tenure.

Ans: GitHub link:

[Machine\_Learning\_assignments/ML\_practice\_questions/ppt\_assignment3\_Q8.ipynb at main · Rajesh29049495/Machine\_Learning\_assignments · GitHub](https://github.com/Rajesh29049495/Machine_Learning_assignments/blob/main/ML_practice_questions/ppt_assignment3_Q8.ipynb)

9. Scenario: A study is conducted to compare the effectiveness of two different medications. The recovery times of the patients in each group are as follows:

   Medication A: [10, 12, 14, 11, 13]

   Medication B: [15, 17, 16, 14, 18]

   Perform an analysis of variance (ANOVA) to determine if there is a significant difference in the mean recovery times between the two medications.

Ans: GitHub link:

[Statistics/statistics\_practiceQs/ppt\_Assignment3\_Q9.ipynb at main · Rajesh29049495/Statistics · GitHub](https://github.com/Rajesh29049495/Statistics/blob/main/statistics_practiceQs/ppt_Assignment3_Q9.ipynb)

10. Scenario: A company wants to analyze customer feedback ratings on a scale of 1 to 10. The data collected is

 as follows:

    [8, 9, 7, 6, 8, 10, 9, 8, 7, 8]

    Calculate the 75th percentile of the feedback ratings.

Ans:

index corresponding to the 75th percentile= (75/100)\*n ## considering index start from 1

= 0.75\*10

= 8

Therefore, the 75th percentile of the feedback ratings is 9.

11. Scenario: A quality control department wants to test the weight consistency of a product. The weights of a **sample** of products are as follows:

    [10.2, 9.8, 10.0, 10.5, 10.3, 10.1]

    Perform a hypothesis test to determine if the mean weight differs significantly from 10 grams.

Ans: Ho = mean weight equak to 10 grams

Ha = mean weight different from 10 grams

Set significance level to determine the threshold for statistical significance, lets considered 0.05

Sample mean = 60.9/6= 10.2

Standard deviation = sqrt((10.2-10.2)^2+……/5)= 0.187

Use the t-test formula to calculate the t-value:

t = (sample\_mean - population\_mean) / (sample\_standard\_deviation / sqrt(sample\_size))

= (10.2-10)/(0.187/sqrt(6))

= 2.618

Df = 5

critical t-value for a two-tailed test with a significance level of 0.05 and 5 degrees of freedom is approximately 2.571.

calculated t-value falls in the rejection region (exceeding the critical t-value), reject the null hypothesis. It suggests that **there is enough evidence to conclude that the mean weight differs significantly from 10 grams.**

12. Scenario: A company wants to analyze the click-through rates of two different website designs. The number of clicks for each design is as follows:

    Design A: [100, 120, 110, 90, 95]

    Design B: [80, 85, 90, 95, 100]

    Perform a chi-square test to determine if there is a significant difference in the click-through rates between the two designs.

Ans:

Chi-square test is used for analysing categorical data rather than continuous data.

So we will apply independent samples t-test to compare the click-through rates between Design A and Design B.

Ho= no significant difference

Ha= there is significant difference

Chosen the significance level for the t-test as 0.05

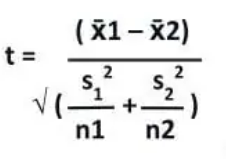
**Mean of design A**= (515)/5=103

**Mean of design B** = (450)/5 = 90

**Sample standard deviation of design A**= sqrt(((103-100)^2+(103-120)^2+ (103-110)^2+ (103-90)^2+(103-95)^2)/4)

=12.04

**Sample standard deviation of design B**= sqrt(((90-80)^2+(90-85)^2+ (90-90)^2+ (90-95)^2+(90-100)^2)/4)= 7.9



t= 13/6.439=2.01

degree of freedom = 4

critical t-value for a two-tailed test with a significance level of 0.05 and 4 degrees of freedom is approximately 2.776

as t- value falls in rejection region as it is less than 2.776, so there is a significant difference between click through rates of two designs.

13. Scenario: A survey is conducted to measure customer satisfaction with a product on a scale of 1 to 10. The data collected is as follows:

    [7, 9, 6, 8, 10, 7, 8, 9, 7, 8]

    Calculate the 95% confidence interval for the population mean satisfaction score

Ans:

**To calculate the 95% confidence interval for the population mean satisfaction score**, the formula:

**Confidence Interval = (sample mean) ± (critical value) \* (standard error)**

Step 1: Calculate the sample mean.

Sample Mean = (7 + 9 + 6 + 8 + 10 + 7 + 8 + 9 + 7 + 8) / 10 = 7.9

Step 2: Calculate the standard error.

Standard Error = (sample standard deviation) / √(sample size)

First, calculate the sample standard deviation:

Sample Standard Deviation = √[((7 - 7.9)^2 + (9 - 7.9)^2 + (6 - 7.9)^2 + (8 - 7.9)^2 + (10 - 7.9)^2 + (7 - 7.9)^2 + (8 - 7.9)^2 + (9 - 7.9)^2 + (7 - 7.9)^2 + (8 - 7.9)^2) / 9]

= √[1.9]

≈ 1.378

Next, calculate the standard error:

Standard Error = 1.378 / √10

≈ 0.436

Step 3: Determine the critical value.

The critical value is based on the desired confidence level and the sample size. For a 95% confidence level and a sample size of 10, the critical value can be obtained from the t-distribution table or using a statistical software. Let's assume the critical value is 2.262 (for a two-tailed test).

Step 4: Calculate the confidence interval.

Confidence Interval = 7.9 ± (2.262 \* 0.436)

≈ 7.9 ± 0.986

≈ (6.914, 8.886)

Therefore, **the 95% confidence interval for the population mean satisfaction score is approximately (6.914, 8.886).** This **means** that we can be **95% confident that the true population mean satisfaction score falls within this range based on the given sample data**.

14. Scenario: A company wants to analyze the effect of temperature on product performance. The data collected is as follows:

    Temperature (in degrees Celsius): [20, 22, 23, 19, 21]

    Performance (on a scale of 1 to 10): [8, 7, 9, 6, 8]

    Perform a simple linear regression to predict performance based on temperature.

Ans: GitHub link:

[Machine\_Learning\_assignments/ML\_practice\_questions/ppt\_assignment3\_Q14.ipynb at main · Rajesh29049495/Machine\_Learning\_assignments · GitHub](https://github.com/Rajesh29049495/Machine_Learning_assignments/blob/main/ML_practice_questions/ppt_assignment3_Q14.ipynb)

15. Scenario: A study is conducted to compare the preferences of two groups of participants. The preferences are measured on a Likert scale from 1 to 5. The data collected is as follows:

    Group A: [4, 3, 5, 2, 4]

    Group B: [3, 2, 4, 3, 3]

    Perform a Mann-Whitney U test to determine if there is a significant difference in the median preferences between the two groups.

Ans:

1. Formulate the null and alternative hypotheses:

* **Null hypothesis (H0**): There is no significant difference in the median preferences between Group A and Group B.
* **Alternative hypothesis** **(HA**): There is a significant difference in the median preferences between Group A and Group B.

1. Here, our level of significance is 0.05
2. Now, we rank the samples according to batches,  if two samples have same rank then we will average the rank

Arranging them in ascending order : [2,2,3,3,3,3,4,4,4,5]

| **Group A** | **Rank (Group A)** | **Group B** | **Rank (Group B)** |
| --- | --- | --- | --- |
| 4 | 8 | 3 | 4.5 |
| 3 | 4.5 | 2 | 1.5 |
| 5 | 10 | 4 | 8 |
| 2 | 1.5 | 3 | 4.5 |
| 4 | 8 | 3 | 4.5 |
| **Rank Sum** | 32 | **Rank Sum** | 23 |

1. Now calculate the U-statistics:

UA= 5\*5+(5\*(5+1)/2)-32 = 8

UB= 5\*5 + (5(5+1)/2)-23= 17

So, our test statistics U = min ( UA, UB) = min (8,17) =8

1. Now, we look into the U-statistics table for n1 = 5 and n2 = 5  and the level of significance. The critical values is :

Uo= 2

Here, U>Uo, then we accept the null hypothesis.

Code Implementation:

# code for Mann-Whitney U test

**from** scipy.stats **import** mannwhitneyu

# Take Group A and Group B data as per above example

batch\_1 **=**[4, 3, 5, 2, 4]

batch\_2 **=**[3, 2, 4, 3, 3]

# perform mann whitney test

stat, p\_value **=** mannwhitneyu(batch\_1, batch\_2)

**print**('Statistics=%.2f, p=%.2f' **%** (stat, p\_value))

# Level of significance

alpha **=** 0.05

# conclusion

**if** p\_value < alpha:

**print**('Reject Null Hypothesis (Significant difference between two samples)')

**else**:

    print('Do not Reject Null Hypothesis (No significant difference between two samples)')

16. Scenario: A company wants to analyze the distribution of customer ages. The data collected is as follows:

    [25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

    Calculate the interquartile range (IQR) of the ages.

Ans:

IQR = Q3-Q1

Q3= (75/100)\*10=7.5

So Q3 falls between 7th and 8th datapoints, which are 55 and 60

Therefore, Q3 = (55+60)/2= 57.5

Q1= (25/100)\*10=2.5

So Q1 falls between 2nd and 3rd datapoints, which are 30 and 35

Therefore, Q3 = (30+35)/2= 32.5

Now IQR = 57.5-32.5

= 25

18. Scenario: A company wants to analyze the effect of price on sales. The data collected is as follows:

    Price (in dollars): [10, 15, 12, 8, 14]

    Sales: [100, 80, 90, 110, 95]

    Perform a simple linear regression to predict sales based on price.

Ans: Github link:

[Machine\_Learning\_assignments/ML\_practice\_questions/ppt\_assignment3\_Q18.ipynb at main · Rajesh29049495/Machine\_Learning\_assignments · GitHub](https://github.com/Rajesh29049495/Machine_Learning_assignments/blob/main/ML_practice_questions/ppt_assignment3_Q18.ipynb)

19. Scenario: A survey is conducted to measure the satisfaction levels of customers with a new product. The data collected is as follows:

    [7, 8, 9, 6, 8, 7, 9, 7, 8, 7]

    Calculate the standard error of the mean satisfaction score.

Ans:

To calculate the standard error of the mean, you can use the following formula:

**Standard Error = Standard Deviation / √(Sample Size)**

data collected is [7, 8, 9, 6, 8, 7, 9, 7, 8, 7].

Let's calculate the standard error of the mean:

Step 1: Calculate the mean:

Mean = (7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7) / 10

= 77 / 10

= 7.7

Step 2: Calculate the deviation of each value from the mean:

Deviation = [7 - 7.7, 8 - 7.7, 9 - 7.7, 6 - 7.7, 8 - 7.7, 7 - 7.7, 9 - 7.7, 7 - 7.7, 8 - 7.7, 7 - 7.7]

= [-0.7, 0.3, 1.3, -1.7, 0.3, -0.7, 1.3, -0.7, 0.3, -0.7]

Step 3: Calculate the squared deviation of each value:

Squared Deviation = [(-0.7)^2, 0.3^2, 1.3^2, (-1.7)^2, 0.3^2, (-0.7)^2, 1.3^2, (-0.7)^2, 0.3^2, (-0.7)^2]

= [0.49, 0.09, 1.69, 2.89, 0.09, 0.49, 1.69, 0.49, 0.09, 0.49]

Step 4: Calculate the variance:

Variance = (0.49 + 0.09 + 1.69 + 2.89 + 0.09 + 0.49 + 1.69 + 0.49 + 0.09 + 0.49) / 10

= 8.01 / 10

= 0.801

Step 5: Calculate the standard deviation:

Standard Deviation = √Variance

= √0.801

≈ 0.895

Step 6: Calculate the standard error of the mean:

Standard Error = Standard Deviation / √(Sample Size)

= 0.895 / √10

≈ 0.283

Therefore, the **standard error of the mean satisfaction score is approximately 0.283.**

20. Scenario: A company wants to analyze the relationship between advertising expenditure and sales. The data collected is as follows:

    Advertising Expenditure (in thousands): [10, 15, 12, 8, 14]

    Sales (in thousands): [25, 30, 28, 20, 26]

    Perform a multiple regression analysis to predict sales based on advertising expenditure.

Ans: Github link:

[Machine\_Learning\_assignments/ML\_practice\_questions/ppt\_assignment3\_Q20.ipynb at main · Rajesh29049495/Machine\_Learning\_assignments · GitHub](https://github.com/Rajesh29049495/Machine_Learning_assignments/blob/main/ML_practice_questions/ppt_assignment3_Q20.ipynb)