**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. Mean= Summation of all observation/ Total no of observation

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

Mean of Region B= [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 Mode is defined as the observation with most number of occurrences

Here in the given observations mode is 4&5 because both of them has occurred 2 times.

**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. Arrange data in the ascending order:

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

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

Median for even numbers=[(n/2) term +(n/2+1) term ]/2

Median for dep A: [4/2 +6/2]/2

= [2nd term + 3rd term]/2

=[5500+6000]/2

=5750 is the median for dep A

Median for odd numbers=[n+1]/2

Median for dep B= [5+1]/2

=6/2

=3rd term = 5500 is the median for dep B

**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** Range= Max number- Min number

Range for the given observations= 26.1-24.8= 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**

Step 1: State the null hypothesis (H0) and alternative hypothesis (H1):

H0: There is no significant difference in the mean scores between Group A and Group B.

H1: There is a significant difference in the mean scores between Group A and Group B.

Step 2: Set the significance level (alpha). Let's assume a common significance level of 0.05 (5%).

Step 3: Calculate the means and standard deviations of both groups:

Group A: MeanA = (85 + 90 + 92 + 88 + 91) / 5 = 88.4

Standard deviationA = sqrt(((85 - 88.4)^2 + (90 - 88.4)^2 + (92 - 88.4)^2 + (88 - 88.4)^2 + (91 - 88.4)^2) / (5 - 1))

= sqrt((8.8 + 0.8 + 13.6 + 0.8 + 6.4) / 4) = sqrt(30.4 / 4) = sqrt(7.6) ≈ 2.76

Group B: MeanB = (82 + 88 + 90 + 86 + 87) / 5 = 86.6

Standard deviationB = sqrt(((82 - 86.6)^2 + (88 - 86.6)^2 + (90 - 86.6)^2 + (86 - 86.6)^2 + (87 - 86.6)^2) / (5 - 1))

= sqrt((18.56 + 1.44 + 11.56 + 0.16 + 0.16) / 4) = sqrt(32.88 / 4) = sqrt(8.22) ≈ 2.87

Step 4: Calculate the t-statistic:

t = (MeanA - MeanB) / sqrt((Standard deviationA^2 / nA) + (Standard deviationB^2 / nB))

= (88.4 - 86.6) / sqrt((2.76^2 / 5) + (2.87^2 / 5))

= 1.8 / sqrt(1.5176 + 1.6544)

= 1.8 / sqrt(3.172)

≈ 1.8 / 1.782

≈ 1.01

Step 5: Determine the degrees of freedom (df):

df = nA + nB - 2 = 5 + 5 - 2 = 8

Step 6: Find the critical value or p-value:

Since we're performing a two-tailed t-test, we need to find the critical value for a significance level of 0.05 and 8 degrees of freedom.

Using a t-table or a statistical software, the critical t-value is approximately ±2.306.

Step 7: Compare the calculated t-statistic with the critical value:

If the calculated t-statistic falls within the critical region (outside the range of ±2.306), we reject the null hypothesis and conclude that there is a significant difference in the mean scores between the two groups. Otherwise, if the calculated t-statistic falls within the non-critical region, we fail to reject the null hypothesis and conclude that there is no significant difference in the mean scores.

In this case, the calculated t-statistic (1.01) falls within the non-critical region (-2.306 to 2.306). Therefore, we fail to reject the null hypothesis and conclude that 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**

Step 1: Calculate the means of both variables:

Mean of Advertising Expenditure (X) = (10 + 15 + 12 + 8 + 14) / 5 = 11.8

Mean of Sales (Y) = (25 + 30 + 28 + 20 + 26) / 5 = 25.8

Step 2: Calculate the deviations from the means for both variables:

Deviations of X: [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]

Deviations of Y: [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]

Step 3: Calculate the product of the deviations for each pair of values:

Product of deviations: [-1.8 \* -0.8, 3.2 \* 4.2, 0.2 \* 2.2, -3.8 \* -5.8, 2.2 \* 0.2] = [1.44, 13.44, 0.44, 22.04, 0.44]

Step 4: Calculate the squared deviations for each variable:

Squared deviations of X: [(-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]

Squared deviations of Y: [(-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]

Step 5: Calculate the sum of the product of deviations and the sum of the squared deviations for each variable:

Sum of product of deviations (ΣXY) = 1.44 + 13.44 + 0.44 + 22.04 + 0.44 = 37.8

Sum of squared deviations of X (ΣX^2) = 3.24 + 10.24 + 0.04 + 14.44 + 4.84 = 33.8

Sum of squared deviations of Y (ΣY^2) = 0.64 + 17.64 + 4.84 + 33.64 + 0.04 = 57.8

Step 6: Calculate the correlation coefficient (r):

r = ΣXY / sqrt(ΣX^2 \* ΣY^2)

= 37.8 / sqrt(33.8 \* 57.8)

≈ 37.8 / sqrt(1952.04)

≈ 37.8 / 44.17

≈ 0.857

Therefore, the correlation coefficient between advertising expenditure and sales is approximately 0.857. The positive value indicates a strong positive linear relationship between the two variables, suggesting that as advertising expenditure increases, sales also tend to increase.

**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.** Std Deviation= sqrt[(x-mu)2/N]

Here mu= mean, x= data points and N=total no of points

Step 1: Find the mean (average) of the data set:

Mean =[160+170+165+155+175+180+170]/7=167.86

Step 2: Calculate the deviation of each data point from the mean:

Deviation of 160 = 160 - 167.86 = -7.86

Deviation of 170 = 170 - 167.86 = 2.14

Deviation of 165 = 165 - 167.86 = -2.86

Deviation of 155 = 155 - 167.86 = -12.86

Deviation of 175 = 175 - 167.86 = 7.14

Deviation of 180 = 180 - 167.86 = 12.14

Deviation of 170 = 170 - 167.86 = 2.14

Step 3: Square each deviation:

Squared deviation of -7.86 = (-7.86)^2 = 61.6996

Squared deviation of 2.14 = (2.14)^2 = 4.5796

Squared deviation of -2.86 = (-2.86)^2 = 8.1796

Squared deviation of -12.86 = (-12.86)^2 = 165.6196

Squared deviation of 7.14 = (7.14)^2 = 51.1396

Squared deviation of 12.14 = (12.14)^2 = 147.6196

Squared deviation of 2.14 = (2.14)^2 = 4.5796

Step 4: Calculate the mean of the squared deviations:

Mean of squared deviations = (61.6996 + 4.5796 + 8.1796 + 165.6196 + 51.1396 + 147.6196 + 4.5796) / 7

= 443.9092 / 7

= 63.4156

Step 5: Taking the sqrt of the mean of standard deviation= √63.4156

≈ 7.96

**8. 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** To calculate the 75th percentile of a dataset, you can follow these steps:

Step 1: Arrange the data in ascending order:

[6, 7, 7, 8, 8, 8, 8, 9, 9, 10]

Step 2: Determine the index of the percentile:

The 75th percentile corresponds to the index position that is 75% of the total number of data points. Since there are 10 data points in this dataset, 75% of 10 is 7.5. To find the index position, round up to the next whole number, which is 8.

Step 3: Find the value at the determined index:

The value at the 8th index position in the sorted dataset is 9.

Therefore, the 75th percentile of the dataset [8, 9, 7, 6, 8, 10, 9, 8, 7, 8] is 9.

**9. 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. To perform a hypothesis test to determine if the mean weight differs significantly from 10 grams, we can use a one-sample t-test. This test will assess whether the sample mean is significantly different from a given population mean.

Here are the steps to perform the hypothesis test:

Step 1: State the null hypothesis (H0) and alternative hypothesis (H1):

H0: The mean weight is equal to 10 grams.

H1: The mean weight is significantly different from 10 grams.

Step 2: Set the significance level (alpha). Let's assume a common significance level of 0.05 (5%).

Step 3: Calculate the sample mean and sample standard deviation:

Sample mean (X̄) = (10.2 + 9.8 + 10.0 + 10.5 + 10.3 + 10.1) / 6 = 60.9 / 6 ≈ 10.15

Sample standard deviation (s) = sqrt(((10.2 - 10.15)^2 + (9.8 - 10.15)^2 + (10.0 - 10.15)^2 + (10.5 - 10.15)^2 + (10.3 - 10.15)^2 + (10.1 - 10.15)^2) / (6 - 1))

= sqrt((0.0025 + 0.0225 + 0.0025 + 0.1025 + 0.0075 + 0.0025) / 5)

= sqrt(0.14 / 5)

≈ sqrt(0.028)

≈ 0.167 (rounded to three decimal places)

Step 4: Calculate the t-statistic:

t = (X̄ - μ) / (s / sqrt(n))

= (10.15 - 10) / (0.167 / sqrt(6))

= 0.15 / (0.167 / sqrt(6))

≈ 0.15 / (0.167 / 2.449)

≈ 0.15 / 0.068

≈ 2.206 (rounded to three decimal places)

Step 5: Determine the degrees of freedom (df):

df = n - 1 = 6 - 1 = 5

Step 6: Find the critical value or p-value:

Since we're performing a two-tailed t-test, we need to find the critical t-value for a significance level of 0.05 and 5 degrees of freedom. Using a t-table or a statistical software, the critical t-value is approximately ±2.571.

Step 7: Compare the calculated t-statistic with the critical value:

If the absolute value of the calculated t-statistic falls within the critical region (outside the range of ±2.571), we reject the null hypothesis and conclude that the mean weight is significantly different from 10 grams. Otherwise, if the absolute value of the calculated t-statistic falls within the non-critical region, we fail to reject the null hypothesis and conclude that there is not enough evidence to suggest a significant difference in the mean weight.

In this case, the absolute value of the calculated t-statistic (2.206) is less than the critical t-value (2.571). Therefore, we fail to reject the null hypothesis and conclude that there is not enough evidence to suggest a significant difference in the mean weight from 10 grams.

**10. 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, we can use the t-distribution since the population standard deviation is not known.

Here are the steps to calculate the confidence interval:

Step 1: Calculate the sample mean and sample standard deviation:

Sample mean (X̄) = (7 + 9 + 6 + 8 + 10 + 7 + 8 + 9 + 7 + 8) / 10 = 79 / 10 = 7.9

Sample standard deviation (s) = sqrt(((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) / (10 - 1))

= sqrt((0.81 + 1.21 + 3.61 + 0.01 + 4.41 + 0.81 + 0.01 + 1.21 + 0.81 + 0.01) / 9)

= sqrt(13.19 / 9)

≈ sqrt(1.4656)

≈ 1.21 (rounded to two decimal places)

Step 2: Determine the sample size (n):

The sample size is 10.

Step 3: Determine the degrees of freedom (df):

Degrees of freedom (df) = n - 1 = 10 - 1 = 9

Step 4: Find the t-value for a 95% confidence level and the degrees of freedom:

Using a t-table or a statistical software, the t-value for a 95% confidence level and 9 degrees of freedom is approximately 2.262.

Step 5: Calculate the standard error (SE):

Standard error (SE) = s / sqrt(n) = 1.21 / sqrt(10) ≈ 0.383 (rounded to three decimal places)

Step 6: Calculate the margin of error (ME):

Margin of error (ME) = t-value \* SE = 2.262 \* 0.383 ≈ 0.867 (rounded to three decimal places)

Step 7: Calculate the lower and upper bounds of the confidence interval:

Lower bound = X̄ - ME = 7.9 - 0.867 ≈ 7.033 (rounded to three decimal places)

Upper bound = X̄ + ME = 7.9 + 0.867 ≈ 8.767 (rounded to three decimal places)

Step 8: Write the 95% confidence interval for the population mean satisfaction score:

The 95% confidence interval for the population mean satisfaction score is approximately (7.033, 8.767).

Therefore, we can be 95% confident that the true population mean satisfaction score falls within the interval of (7.033, 8.767).

**11. 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. Arrange the data in ascending order:

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

First quartile(Q1) will be the median from the first half i.e, [25,30,35,40,45] = 35

Third quartile(Q3) will be the median from the second half i.e, [50, 55, 60, 65, 70]= 60

Interquartile range= Q3-Q1

=60-35

=25

**12. Scenario: A study is conducted to compare the performance of three different machine learning algorithms. The accuracy scores for each algorithm are as follows:**

**Algorithm A: [0.85, 0.80, 0.82, 0.87, 0.83]**

**Algorithm B: [0.78, 0.82, 0.84, 0.80, 0.79]**

**Algorithm C: [0.90, 0.88, 0.89, 0.86, 0.87]**

**Perform a Kruskal-Wallis test to determine if there is a significant difference in the median accuracy scores between the algorithms.**

Ans Here are the steps to perform the Kruskal-Wallis test:

Step 1: State the null hypothesis (H0) and alternative hypothesis (H1):

H0: There is no significant difference in the median accuracy scores between the algorithms.

H1: There is a significant difference in the median accuracy scores between the algorithms.

Step 2: Set the significance level (alpha). Let's assume a common significance level of 0.05 (5%).

Step 3: Combine the accuracy scores from all algorithms:

Combined data: [0.85, 0.80, 0.82, 0.87, 0.83, 0.78, 0.82, 0.84, 0.80, 0.79, 0.90, 0.88, 0.89, 0.86, 0.87]

Step 4: Rank the combined data:

Rank the combined data in ascending order: [0.78, 0.79, 0.80, 0.80, 0.82, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.87, 0.88, 0.89, 0.90]

Step 5: Calculate the sum of ranks for each algorithm:

Sum of ranks for Algorithm A = 8 + 10 + 11 + 13 + 9 = 51

Sum of ranks for Algorithm B = 1 + 5 + 7 + 3 + 2 = 18

Sum of ranks for Algorithm C = 15 + 14 + 12 + 6 + 4 = 51

Step 6: Calculate the test statistic (H):

H = [12 \* (Sum of ranks for each algorithm)^2] / (n \* (n + 1)) - 3 \* (n + 1)

= [12 \* ((51)^2 + (18)^2 + (51)^2)] / (15 \* (15 + 1)) - 3 \* (15 + 1)

≈ 1266

Step 7: Determine the degrees of freedom (df):

Degrees of freedom (df) = Number of groups - 1 = 3 - 1 = 2

Step 8: Find the critical value or p-value:

Using a Kruskal-Wallis table or a statistical software, find the critical value or p-value for a significance level of 0.05 and 2 degrees of freedom. In this case, the critical value is approximately 5.991.

Step 9: Compare the calculated test statistic with the critical value:

If the calculated test statistic is greater than the critical value, we reject the null hypothesis and conclude that there is a significant difference in the median accuracy scores between the algorithms. Otherwise, if the calculated test statistic is less than the critical value, we fail to reject the null hypothesis and conclude that there is no significant difference in the median accuracy scores.

In this case, the calculated test statistic (1266) is greater than the critical value (5.991). Therefore, we reject the null hypothesis and conclude that there is a significant difference in the median accuracy scores between the algorithms.

**13. 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** Step 1: Calculate the sample mean:

Sample mean (X̄) = (7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7) / 10 = 76 / 10 = 7.6

Step 2: Calculate the sample standard deviation:

Sample standard deviation (s) = sqrt(((7 - 7.6)^2 + (8 - 7.6)^2 + (9 - 7.6)^2 + (6 - 7.6)^2 + (8 - 7.6)^2 + (7 - 7.6)^2 + (9 - 7.6)^2 + (7 - 7.6)^2 + (8 - 7.6)^2 + (7 - 7.6)^2) / (10 - 1))

= sqrt((0.36 + 0.16 + 1.96 + 2.56 + 0.16 + 0.36 + 1.96 + 0.36 + 0.16 + 0.36) / 9)

= sqrt(8.52 / 9)

≈ sqrt(0.947)

≈ 0.973 (rounded to three decimal places)

Step 3: Calculate the standard error (SE) of the mean:

Standard error (SE) = s / sqrt(n) = 0.973 / sqrt(10) ≈ 0.308 (rounded to three decimal places)

Therefore, the standard error of the mean satisfaction score is approximately 0.308**.**