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.

To calculate the mean sales for each region, you need to sum up the sales values for each region and then divide the sum by the total number of sales values in that region. Let's calculate the mean sales for each region using the provided data:

Region A: [10, 15, 12, 8, 14] To calculate the mean sales for Region A, you sum up the sales values and divide by the total count.

Sum of sales in Region A = 10 + 15 + 12 + 8 + 14 = 59 Number of sales in Region A = 5

Mean sales for Region A = Sum of sales / Number of sales = 59 / 5 = 11.8

Therefore, the mean sales for Region A is 11.8.

Region B: [18, 20, 16, 22, 25] To calculate the mean sales for Region B, you sum up the sales values and divide by the total count.

Sum of sales in Region B = 18 + 20 + 16 + 22 + 25 = 101 Number of sales in Region B = 5

Mean sales for Region B = Sum of sales / Number of sales = 101 / 5 = 20.2

Therefore, the mean sales for Region B is 20.2.

So, the mean sales for Region A is 11.8 and the mean sales for Region B is 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

To calculate the mode of the survey responses, we need to determine the value that appears most frequently in the data set. In this case, the data collected is: [4, 5, 2, 3, 5, 4, 3, 2, 4, 5].

Let's count the occurrences of each value:

- Value 2 appears twice.

- Value 3 appears twice.

- Value 4 appears three times.

- Value 5 appears three times.

Since both value 4 and value 5 appear the most frequently, the mode of the survey responses is a tie between 4 and 5.

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.

To calculate the median salary for each department, you need to find the middle value of the sorted salary data. If the number of data points is odd, the median is the middle value. If the number of data points is even, the median is the average of the two middle values.

Let's calculate the median salary for each department:

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

First, sort the salaries in ascending order:

[5000, 5500, 6000, 7000]

The number of data points is 4, which is even. So, to find the median, we take the average of the two middle values:

Median salary for Department A = (5500 + 6000) / 2 = 5750

Now, let's calculate the median salary for Department B:

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

Sort the salaries in ascending order:

[4500, 5200, 5500, 5800, 6000]

The number of data points is 5, which is odd. The median is the middle value:

Median salary for Department B = 5500

Therefore, the median salary for Department A is $5,750 and for Department B is $5,500.

 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.

To calculate the range of the stock prices, you need to find the difference between the highest and lowest values in the dataset. In this case, the data analyst wants to determine the variability in the daily stock prices of a company with the following data:

[25.5, 24.8, 26.1, 25.3, 24.9]

To calculate the range, follow these steps:

1. Find the highest value in the dataset:

- In this case, the highest value is 26.1.

2. Find the lowest value in the dataset:

- In this case, the lowest value is 24.8.

3. Calculate the range by subtracting the lowest value from the highest value:

- Range = Highest value - Lowest value

- Range = 26.1 - 24.8

- Range = 1.3

Therefore, the range of the stock prices is 1.3. This indicates the difference between the highest and lowest prices observed in the dataset.

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.

To perform a t-test and determine if there is a significant difference in the mean scores between Group A and Group B, we can use a two-sample t-test. The null hypothesis (H0) assumes that there is no significant difference between the means of the two groups, while the alternative hypothesis (Ha) assumes that there is a significant difference.

Let's calculate the t-statistic and the p-value using Python:

```python

import scipy.stats as stats

group\_a = [85, 90, 92, 88, 91]

group\_b = [82, 88, 90, 86, 87]

# Perform two-sample t-test

t\_statistic, p\_value = stats.ttest\_ind(group\_a, group\_b)

# Print the results

print("t-statistic:", t\_statistic)

print("p-value:", p\_value)

```

Running this code will give you the t-statistic and the p-value.

The t-statistic measures the difference between the means of the two groups relative to the variation within each group. The p-value represents the probability of observing the data if the null hypothesis is true. If the p-value is below a certain significance level (e.g., 0.05), we can reject the null hypothesis and conclude that there is a significant difference between the mean scores of the two groups.

Please note that for a t-test to be valid, certain assumptions need to be met, such as normality and equal variances.

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.

To calculate the correlation coefficient between advertising expenditure and sales, you can use the Pearson correlation coefficient formula. Here's how you can calculate it:

Step 1: Calculate the mean of the advertising expenditure and sales.

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

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

Mean of Advertising Expenditure:

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

Mean of Sales:

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

Step 2: Calculate the deviations from the mean for each value in the advertising expenditure and sales.

Deviation from Mean of Advertising Expenditure:

[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]

Deviation from Mean of Sales:

[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 sum of the products of deviations.

Sum of Products of Deviations:

1.44 + 13.44 + 0.44 + 22.04 + 0.44 = 37.8

Step 5: Calculate the standard deviation for advertising expenditure and sales.

Standard Deviation of Advertising Expenditure:

sqrt(((-1.8)^2 + 3.2^2 + 0.2^2 + (-3.8)^2 + 2.2^2) / 5) = 2.424

Standard Deviation of Sales:

sqrt(((-0.8)^2 + 4.2^2 + 2.2^2 + (-5.8)^2 + 0.2^2) / 5) = 3.384

Step 6: Calculate the correlation coefficient.

Correlation Coefficient:

37.8 / (2.424 \* 3.384) = 4.988

The correlation coefficient between advertising expenditure and sales is approximately 4.988.

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.

To calculate the standard deviation of the heights, you can follow these steps:

1. Find the mean (average) of the heights.

2. Subtract the mean from each height value and square the result.

3. Find the mean of the squared differences obtained in step 2.

4. Take the square root of the mean from step 3.

Let's calculate the standard deviation using the given data:

Step 1: Find the mean

Sum of heights = 160 + 170 + 165 + 155 + 175 + 180 + 170 = 1175

Number of heights = 7

Mean = Sum of heights / Number of heights = 1175 / 7 = 167.857

Step 2: Calculate the squared differences

(160 - 167.857)^2 ≈ 0.062

(170 - 167.857)^2 ≈ 0.046

(165 - 167.857)^2 ≈ 0.081

(155 - 167.857)^2 ≈ 0.131

(175 - 167.857)^2 ≈ 0.052

(180 - 167.857)^2 ≈ 0.149

(170 - 167.857)^2 ≈ 0.046

Step 3: Find the mean of the squared differences

Sum of squared differences = 0.062 + 0.046 + 0.081 + 0.131 + 0.052 + 0.149 + 0.046 = 0.567

Mean = Sum of squared differences / Number of heights = 0.567 / 7 = 0.081

Step 4: Take the square root of the mean

Standard deviation ≈ √0.081 ≈ 0.285

Therefore, the standard deviation of the heights is approximately 0.285.

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.

To perform a linear regression analysis to predict job satisfaction based on employee tenure, we need to follow these steps:

Step 1: Set up the problem

- Define the variables:

- X: Employee Tenure (in years)

- Y: Job Satisfaction (on a scale of 1 to 10)

- Gather the data:

- X values: [2, 3, 5, 4, 6, 2, 4]

- Y values: [7, 8, 6, 9, 5, 7, 6]

- Determine the number of data points (n): n = 7

Step 2: Calculate the mean (average) of X and Y

- Mean of X (X̄):

- Sum of X values: 2 + 3 + 5 + 4 + 6 + 2 + 4 = 26

- X̄ = 26 / 7 = 3.714 (rounded to 3 decimal places)

- Mean of Y (Ȳ):

- Sum of Y values: 7 + 8 + 6 + 9 + 5 + 7 + 6 = 48

- Ȳ = 48 / 7 = 6.857 (rounded to 3 decimal places)

Step 3: Calculate the deviations from the mean for X and Y

- Deviations of X (X - X̄):

- X deviations: [2 - 3.714, 3 - 3.714, 5 - 3.714, 4 - 3.714, 6 - 3.714, 2 - 3.714, 4 - 3.714]

- X deviations: [-1.714, -0.714, 1.286, 0.286, 2.286, -1.714, 0.286]

- Deviations of Y (Y - Ȳ):

- Y deviations: [7 - 6.857, 8 - 6.857, 6 - 6.857, 9 - 6.857, 5 - 6.857, 7 - 6.857, 6 - 6.857]

- Y deviations: [0.143, 1.143, -0.857, 2.143, -1.857, 0.143, -0.857]

Step 4: Calculate the sum of squared deviations for X and Y

- Sum of squared deviations of X (Σ(X - X̄)²):

- (Σ(X - X̄)²) = (-1.714)² + (-0.714)² + 1.286² + 0.286² + 2.286² + (-1.714)² + 0.286²

- (Σ(X - X̄)²) ≈ 10.857

- Sum of squared deviations of Y (Σ(Y - Ȳ)²):

- (Σ(Y - Ȳ)²) = 0.143² + 1.143² + (-0.857)² + 2.143² + (-1.857)² + 0.143² + (-0.857)²

- (Σ(Y - Ȳ)²) ≈ 11.571

Step 5: Calculate the product of deviations for X and Y

- Product of deviations (X - X̄)(Y - Ȳ):

- [(X - X̄)(Y - Ȳ)]: (-1.714 \* 0.143) + (-0.714 \* 1.143) + (1.286 \* -0.857) + (0.286 \* 2.143) + (2.286 \* -1.857) + (-1.714 \* 0.143) + (0.286 \* -0.857)

- [(X - X̄)(Y - Ȳ)] ≈ -3.571

Step 6: Calculate the slope (β) of the regression line

- β = [(Σ(X - X̄)(Y - Ȳ)) / (Σ(X - X̄)²)]

- β ≈ -3.571 / 10.857 ≈ -0.329 (rounded to 3 decimal places)

Step 7: Calculate the intercept (α) of the regression line

- α = Ȳ - βX̄

- α ≈ 6.857 - (-0.329 \* 3.714) ≈ 7.981 (rounded to 3 decimal places)

Step 8: Write the regression line equation

- The regression line equation is: Y = α + βX

- Substituting the values: Y = 7.981 - 0.329X

Therefore, the linear regression equation to predict job satisfaction based on employee tenure is: Y = 7.981 - 0.329X

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.

To perform an analysis of variance (ANOVA) to compare the mean recovery times between Medication A and Medication B, we can use the following steps:

Step 1: State the null and alternative hypotheses:

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

- Alternative hypothesis (H₁): There is a significant difference in the mean recovery times between Medication A and Medication B.

Step 2: Calculate the necessary statistics:

- Calculate the mean (average) recovery time for each group.

- Calculate the sum of squares between groups (SSB).

- Calculate the sum of squares within groups (SSW).

- Determine the degrees of freedom for both SSB and SSW.

- Calculate the mean square between groups (MSB) and mean square within groups (MSW).

- Calculate the F-statistics

Step 3: Determine the critical value or p-value:

- Look up the critical value for the desired significance level (e.g., α = 0.05) using the F-distribution table.

- Alternatively, calculate the p-value associated with the F-statistic.

Step 4: Compare the calculated F-statistic with the critical value or p-value:

- If the calculated F-statistic is greater than the critical value or the p-value is less than the significance level (α), reject the null hypothesis.

- Otherwise, fail to reject the null hypothesis.

Let's perform the calculations:

Step 1: State the null and alternative hypotheses:

- H₀: There is no significant difference in the mean recovery times between Medication A and Medication B.

- H₁: There is a significant difference in the mean recovery times between Medication A and Medication B.

Step 2: Calculate the necessary statistics:

- Mean recovery time for Medication A:

- Mean (μ₁) = (10 + 12 + 14 + 11 + 13) / 5 = 12

- Mean recovery time for Medication B:

- Mean (μ₂) = (15 + 17 + 16 + 14 + 18) / 5 = 16

- SSB = (n₁ \* (μ₁ - μ)²) + (n₂ \* (μ₂ - μ)²), where n₁ and n₂ are the sample sizes, μ₁ and μ₂ are the means, and μ is the grand mean.

- SSB = (5 \* (12 - 14)²) + (5 \* (16 - 14)²) = 80

- SSW = (Σ(x₁ - μ₁)²) + (Σ(x₂ - μ₂)²), where Σ is the sum of squares.

- SSW = (10 - 12)² + (12 - 12)² + (14 - 12)² + (11 - 12)² + (13 - 12)² + (15 - 16)² + (17 - 16)² + (16 - 16)² + (14 - 16)² + (18 - 16)² = 40

- Degrees of freedom (df):

- df₁ = k - 1 = 2 - 1 = 1 (between groups)

- df₂ = N - k = 10 - 2 = 8 (within groups)

- MSB = SSB / df₁ = 80 / 1 = 80

- MSW = SSW / df₂ = 40 / 8 = 5

- F-statistic = MSB / MSW = 80 / 5 = 16

Step 3: Determine the critical value or p-value:

- For α = 0.05 and df₁ = 1, df₂ = 8, the critical value from the F-distribution table is approximately 5.32.

- Alternatively, you can calculate the p-value associated with the F-statistic using statistical software or online tools.

Step 4: Compare the calculated F-statistic with the critical value or p-value:

- The calculated F-statistic (16) is greater than the critical value (5.32).

- Therefore, we reject the null hypothesis.

Conclusion:

Based on the ANOVA results, there is a significant difference in the mean recovery times between Medication A and Medication B (p < 0.05).

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.

To calculate the 75th percentile of the feedback ratings, you first need to arrange the data in ascending order:

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

The 75th percentile represents the value below which 75% of the data falls. In this case, since we have 10 data points, we want to find the value that is below 75% of the data, which corresponds to the 7.5th (10 \* 0.75) data point.

To find the 75th percentile, we can use the following formula:

75th percentile = L + (0.75 \* (U - L))

where L is the lower value, and U is the upper value, between which the desired percentile falls.

In our case, the 7.5th data point falls between the values 7 and 8. To calculate the 75th percentile:

L = 7 (lower value)

U = 8 (upper value)

75th percentile = 7 + (0.75 \* (8 - 7))

= 7 + (0.75 \* 1)

= 7 + 0.75

= 7.75

Therefore, the 75th percentile of the feedback ratings is 7.75.

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.

To perform a hypothesis test to determine if the mean weight differs significantly from 10 grams, we can use a one-sample t-test. The null hypothesis (H0) states that the mean weight is equal to 10 grams, while the alternative hypothesis (Ha) states that the mean weight is different from 10 grams.

Here are the steps to conduct the hypothesis test:

Step 1: Define the hypotheses:

H0: The population mean weight = 10 grams

Ha: The population mean weight ≠ 10 grams

Step 2: Set the significance level (α):

Let's assume a significance level of α = 0.05. This means we want to be 95% confident in our decision.

Step 3: Collect the sample data:

The sample weights are as follows: [10.2, 9.8, 10.0, 10.5, 10.3, 10.1]

Step 4: Compute the test statistic:

We'll calculate the t-statistic using the sample data and the formula:

t = (sample mean - hypothesized mean) / (sample standard deviation / sqrt(sample size))

Step 5: Determine the critical value:

Since we have a two-tailed test (the alternative hypothesis is not specific about the direction), we need to find the critical values corresponding to the chosen significance level. For a significance level of α = 0.05 and a sample size of 6, the critical t-values can be obtained from a t-distribution table or a statistical software. Let's assume the critical values are -2.571 and 2.571.

Step 6: Compare the test statistic with the critical value:

If the test statistic falls outside the critical region (i.e., it is less than the negative critical value or greater than the positive critical value), we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

Step 7: Draw a conclusion:

Based on the comparison in Step 6, we will either reject or fail to reject the null hypothesis and make a conclusion.

Now, let's calculate the test statistic and draw a conclusion:

Step 4: Compute the test statistic:

The sample mean is (10.2 + 9.8 + 10.0 + 10.5 + 10.3 + 10.1) / 6 = 10.1667

The sample standard deviation can be calculated as follows:

s = sqrt((Σ(xi - x̄)^2) / (n - 1))

where xi represents each individual weight, x̄ is the sample mean, and n is the sample size.

Using the given weights, we have:

s = sqrt(((10.2 - 10.1667)^2 + (9.8 - 10.1667)^2 + (10.0 - 10.1667)^2 + (10.5 - 10.1667)^2 + (10.3 - 10.1667)^2 + (10.1 - 10.1667)^2) / (6 - 1))

Calculating this expression yields s ≈ 0.2208.

The test statistic t can be calculated as:

t = (10.1667 - 10) / (0.2208 / sqrt(6))

Calculating this expression yields t ≈ 1.536.

Step 5: Determine the critical value:

The critical t-values for a significance level of α = 0.05 and a two-tailed test with a sample size of 6 are -2.571 and 2.571.

Step 6: Compare the test statistic with the critical value:

Since the test statistic t = 1.536 falls within the range of -2.571 to 2.571, it does not fall into the critical region.

Step 7: Draw a conclusion:

Since the test statistic does not fall into the critical region, we fail to reject the null hypothesis. There is not enough evidence to conclude that the mean weight differs significantly from 10 grams based on the given sample.

In summary, based on the hypothesis test, we cannot determine that the mean weight differs significantly from 10 grams with the provided sample data.

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.

To perform a chi-square test to determine if there is a significant difference in the click-through rates between the two designs (Design A and Design B), we need to follow these steps:

Step 1: State the hypotheses.

The null hypothesis (H₀): There is no significant difference in click-through rates between Design A and Design B.

The alternative hypothesis (H₁): There is a significant difference in click-through rates between Design A and Design B.

Step 2: Set the significance level.

Let's assume a significance level (α) of 0.05, which is a common choice.

Step 3: Calculate the expected frequencies.

We need to calculate the expected frequencies for each design. The expected frequency for each cell is calculated as (row total \* column total) / grand total.

The observed frequencies for each design are as follows:

Design A: [100, 120, 110, 90, 95]

Design B: [80, 85, 90, 95, 100]

The total number of observations (grand total) is 5 (the number of elements in each design).

The expected frequencies for each design can be calculated as:

Design A (expected): [92, 110, 102, 83, 87]

Design B (expected): [88, 105, 97, 79, 83]

Step 4: Calculate the chi-square test statistic.

The chi-square test statistic (χ²) can be calculated using the formula:

χ² = Σ [(Oᵢ - Eᵢ)² / Eᵢ]

Where Oᵢ is the observed frequency and Eᵢ is the expected frequency for each cell.

For our example, the calculations are as follows:

χ² = [(100-92)²/92] + [(120-110)²/110] + [(110-102)²/102] + [(90-83)²/83] + [(95-87)²/87] + [(80-88)²/88] + [(85-105)²/105] + [(90-97)²/97] + [(95-79)²/79] + [(100-83)²/83]

After calculating the above expression, we get the chi-square test statistic (χ²).

Step 5: Determine the degrees of freedom.

The degrees of freedom (df) for a chi-square test of independence can be calculated as (number of rows - 1) \* (number of columns - 1).

In our case, the number of rows and columns are both 2, so the degrees of freedom is (2 - 1) \* (2 - 1) = 1.

Step 6: Determine the critical value.

Using the significance level and degrees of freedom, we can determine the critical value from the chi-square distribution table or using statistical software. For α = 0.05 and df = 1, the critical value is approximately 3.841.

Step 7: Compare the test statistic with the critical value.

If the test statistic is greater than the critical value, we reject the null hypothesis and conclude that there is a significant difference in click-through rates between Design A and Design B.

Step 8: Make a conclusion.

Based on the comparison of the test statistic and critical value, we can make a conclusion about the hypotheses.

Please note that the calculations for the chi-square test statistic can be a bit tedious. To simplify the process, you can use statistical software or online calculators that perform the chi-square test automatically.

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.

To calculate the 95% confidence interval for the population mean satisfaction score, we can use the following formula:

Confidence Interval = Sample Mean ± Margin of Error

First, let's calculate the sample mean:

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

= 79 / 10

= 7.9

Next, we need to calculate the margin of error. To do this, we need to determine the standard error of the mean. The standard error can be calculated using the following formula:

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

To find the sample standard deviation, we first calculate the sum of the squared differences between each data point and the sample mean, and then divide it by the sample size minus 1 (for an unbiased estimate):

Sum of Squared Differences = (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

= 0.81 + 1.21 + 3.61 + 0.01 + 4.41 + 0.81 + 0.01 + 1.21 + 0.81 + 0.01

= 13.09

Sample Standard Deviation = √(Sum of Squared Differences / (Sample Size - 1))

= √(13.09 / (10 - 1))

= √(13.09 / 9)

= √1.454444...

≈ 1.206

Now, we can calculate the standard error:

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

= 1.206 / √10

≈ 0.381

To find the margin of error, we need to multiply the standard error by the critical value. For a 95% confidence level, the critical value is approximately 1.96 (assuming a large sample size):

Margin of Error = Critical Value \* Standard Error

≈ 1.96 \* 0.381

≈ 0.747

Finally, we can calculate the confidence interval:

Confidence Interval = Sample Mean ± Margin of Error

= 7.9 ± 0.747

≈ (7.153, 8.647)

Therefore, the 95% confidence interval for the population mean satisfaction score is approximately (7.153, 8.647).

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.

To perform a simple linear regression to predict performance based on temperature, we can use the collected data:

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

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

Let's go through the steps to perform the regression analysis:

Step 1: Import the necessary libraries

You'll need to import the required libraries for performing the regression analysis. In this case, we'll use the NumPy and scikit-learn libraries.

```python

import numpy as np

from sklearn.linear\_model import LinearRegression

```

Step 2: Prepare the data

Create NumPy arrays for the temperature and performance values:

```python

temperature = np.array([20, 22, 23, 19, 21]).reshape(-1, 1)

performance = np.array([8, 7, 9, 6, 8])

```

Reshape the `temperature` array using `.reshape(-1, 1)` to convert it into a column vector. This step is necessary because scikit-learn expects the input features to be a 2D array.

Step 3: Create and fit the regression model

Create an instance of the `LinearRegression` model and fit it to the data:

```python

model = LinearRegression()

model.fit(temperature, performance)

```

Step 4: Obtain the regression results

Retrieve the regression coefficients and the coefficient of determination (R-squared) to evaluate the model's performance:

```python

intercept = model.intercept\_

slope = model.coef\_[0]

r\_squared = model.score(temperature, performance)

```

The `intercept` represents the y-intercept of the regression line, and the `slope` represents the coefficient for the temperature variable.

Step 5: Predict performance based on temperature

You can use the trained model to make predictions. Let's say you want to predict the performance at a temperature of 25 degrees Celsius:

```python

new\_temperature = np.array([25]).reshape(-1, 1)

predicted\_performance = model.predict(new\_temperature)

```

The `predicted\_performance` variable will contain the predicted value.

Step 6: Print the results

Finally, print the regression results:

```python

print("Intercept:", intercept)

print("Slope:", slope)

print("R-squared:", r\_squared)

print("Predicted performance at 25 degrees Celsius:", predicted\_performance)

```

This will display the intercept, slope, R-squared value, and the predicted performance at 25 degrees Celsius.

Keep in mind that this is a basic example of performing a simple linear regression. Additional considerations such as model assumptions, model evaluation, and data preprocessing might be required for more complex analyses.

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.

To perform a Mann-Whitney U test and determine if there is a significant difference in the median preferences between Group A and Group B, we can follow these steps:

Step 1: Define 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 (H1): There is a significant difference in the median preferences between Group A and Group B.

Step 2: Rank the data:

Combine the data from both groups and rank them in ascending order, assigning ranks to the values. Ties should be assigned average ranks.

Group A: [4, 3, 5, 2, 4]

Group B: [3, 2, 4, 3, 3]

Combined data: [4, 3, 5, 2, 4, 3, 2, 4, 3, 3]

Ranking: [7, 3.5, 10, 1, 7, 3.5, 1, 7, 3.5, 3.5]

Step 3: Calculate the U statistic:

Now, we calculate the U statistic, which represents the sum of the ranks for one of the groups. We can use the formula:

U = R1 - (n1 \* (n1 + 1)) / 2

where R1 is the sum of ranks for Group A and n1 is the number of participants in Group A.

For Group A:

U1 = 7 + 3.5 + 10 + 1 + 7 = 28.5

n1 = 5

U1 = 28.5 - (5 \* (5 + 1)) / 2 = 28.5 - 15 = 13.5

Step 4: Calculate the U critical value:

The U critical value depends on the sample sizes of both groups. Since both groups have 5 participants, we can refer to a Mann-Whitney U table or use a statistical software to determine the critical value.

For a two-tailed test at a significance level of α = 0.05, the critical value for U with n1 = n2 = 5 is 5.

Step 5: Compare the U statistic with the critical value:

If U ≤ U critical, we fail to reject the null hypothesis; otherwise, we reject the null hypothesis in favor of the alternative hypothesis.

U1 = 13.5

U critical = 5

Since U1 > U critical, we reject the null hypothesis.

Step 6: Interpret the results:

The Mann-Whitney U test indicates that there is a significant difference in the median preferences between Group A and Group B.

In conclusion, based on the provided data, we have evidence to suggest that the preferences of the two groups significantly differ.

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.

To calculate the interquartile range (IQR) of the customer ages, you first need to find the values of the first quartile (Q1) and the third quartile (Q3).

Step 1: Sort the data in ascending order:

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

Step 2: Calculate the median (second quartile, Q2) of the data. Since the data set has an even number of elements, the median is the average of the two middle values:

Q2 = (45 + 50) / 2 = 47.5

Step 3: Find the position of the first quartile (Q1) by taking the median of the lower half of the data set. In this case, the lower half is [25, 30, 35, 40]:

Q1 = (30 + 35) / 2 = 32.5

Step 4: Find the position of the third quartile (Q3) by taking the median of the upper half of the data set. In this case, the upper half is [55, 60, 65, 70]:

Q3 = (60 + 65) / 2 = 62.5

Step 5: Calculate the interquartile range (IQR) by subtracting Q1 from Q3:

IQR = Q3 - Q1 = 62.5 - 32.5 = 30

Therefore, the interquartile range (IQR) of the customer ages is 30.

17. 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.

To perform a Kruskal-Wallis test to determine if there is a significant difference in the median accuracy scores between the algorithms, we need to follow these steps:

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

- H0 (null hypothesis): There is no significant difference in the median accuracy scores between the algorithms.

- Ha (alternative hypothesis): There is a significant difference in the median accuracy scores between the algorithms.

Step 2: Calculate the ranks for all the accuracy scores combined. Assign lower ranks to lower scores and higher ranks to higher scores, with tied scores receiving the average rank.

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

The ranks would be: [4, 1, 2, 5, 3]

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

The ranks would be: [1, 3, 5, 2, 4]

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

The ranks would be: [5, 3, 4, 2, 1]

Step 3: Calculate the sum of ranks for each algorithm.

Algorithm A: Sum of ranks = 4 + 1 + 2 + 5 + 3 = 15

Algorithm B: Sum of ranks = 1 + 3 + 5 + 2 + 4 = 15

Algorithm C: Sum of ranks = 5 + 3 + 4 + 2 + 1 = 15

Step 4: Calculate the test statistic (H) using the formula:

H = (12 / (N(N+1))) \* (sum(rank^2) - (N(N+1)^2)/4)

Where N is the total number of observations (15 in this case).

H = (12 / (15(15+1))) \* ((15^2) - (15(15+1)^2)/4)

H = (12 / (15(16))) \* (225 - (15(16)^2)/4)

H = (12 / 240) \* (225 - 2400/4)

H = 0.05 \* (225 - 600)

H = 0.05 \* (-375)

H = -18.75

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

df = (k - 1)

Where k is the number of algorithms (3 in this case).

df = (3 - 1) = 2

Step 6: Look up the critical value for the Kruskal-Wallis test in a chi-square distribution table or use a statistical software. For a significance level (α) of 0.05 and df = 2, the critical value is approximately 5.99.

Step 7: Compare the test statistic (H) to the critical value. If the test statistic is greater than the critical value, we reject the null hypothesis; otherwise, we fail to reject the null hypothesis.

Since the test statistic (-18.75) is less than the critical value (5.99), we fail to reject the null hypothesis.

Therefore, based on the Kruskal-Wallis test, there is not enough evidence to conclude that there is a significant difference in the median accuracy scores between the algorithms.

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.

To perform a simple linear regression analysis, we can use the given data to create a model that predicts sales based on price. In this case, price will be the independent variable, and sales will be the dependent variable.

Let's start by representing the data in tabular form:

Price (in dollars) | Sales

-------------------|-------

10 | 100

15 | 80

12 | 90

8 | 110

14 | 95

Next, we'll calculate the slope (b) and the y-intercept (a) of the regression line using the following formulas:

b = Σ((x - mean(x))(y - mean(y))) / Σ((x - mean(x))^2)

a = mean(y) - b \* mean(x)

Where:

- x represents the price values

- y represents the sales values

- mean(x) is the mean of the price values

- mean(y) is the mean of the sales values

Let's calculate the slope (b) and the y-intercept (a) using these formulas:

Step 1: Calculate the means of x and y

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

mean(y) = (100 + 80 + 90 + 110 + 95) / 5 = 95

Step 2: Calculate the numerator and denominator for b

numerator = ((10 - 11.8)(100 - 95)) + ((15 - 11.8)(80 - 95)) + ((12 - 11.8)(90 - 95)) + ((8 - 11.8)(110 - 95)) + ((14 - 11.8)(95 - 95))

= (-1.8)(5) + (3.2)(-15) + (0.2)(-5) + (-3.8)(15) + (2.2)(0)

= -9 + (-48) + (-1) + (-57) + 0

= -115

denominator = ((10 - 11.8)^2) + ((15 - 11.8)^2) + ((12 - 11.8)^2) + ((8 - 11.8)^2) + ((14 - 11.8)^2)

= (-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

= 32.8

Step 3: Calculate b

b = numerator / denominator = -115 / 32.8 ≈ -3.506

Step 4: Calculate a

a = mean(y) - b \* mean(x) = 95 - (-3.506) \* 11.8 ≈ 134.758

Therefore, the regression line equation is:

sales = 134.758 - 3.506 \* price

This equation represents the relationship between price and sales based on the given data. Using this equation, you can predict sales for any given price value.

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.

To calculate the standard error of the mean, you need to follow these steps:

1. Find the mean (average) of the data set. Add up all the numbers and divide the sum by the total count of numbers in the data set.

The sum of the numbers in the data set is: 7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7 = 78

The total count of numbers is 10.

The mean is: 78 / 10 = 7.8

2. Find the difference between each data point and the mean, and square each difference.

The squared differences are:

(7 - 7.8)^2 = 0.64

(8 - 7.8)^2 = 0.04

(9 - 7.8)^2 = 1.44

(6 - 7.8)^2 = 3.24

(8 - 7.8)^2 = 0.04

(7 - 7.8)^2 = 0.64

(9 - 7.8)^2 = 1.44

(7 - 7.8)^2 = 0.64

(8 - 7.8)^2 = 0.04

(7 - 7.8)^2 = 0.64

3. Calculate the variance by finding the average of the squared differences.

The sum of the squared differences is: 0.64 + 0.04 + 1.44 + 3.24 + 0.04 + 0.64 + 1.44 + 0.64 + 0.04 + 0.64 = 9.6

The variance is: 9.6 / 10 = 0.96

4. Finally, calculate the standard error of the mean by taking the square root of the variance divided by the total count of numbers.

The standard error of the mean is: √(0.96 / 10) ≈ 0.309

Therefore, the standard error of the mean satisfaction score is approximately 0.309.

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.

To perform a multiple regression analysis to predict sales based on advertising expenditure, we'll use the data you provided:

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

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

In multiple regression, we assume that there is a linear relationship between the independent variable (advertising expenditure) and the dependent variable (sales) with some random error. We can use statistical software or tools like Python, R, or Excel to calculate the regression coefficients and predict sales based on advertising expenditure.

Here, I'll demonstrate how to perform the analysis using Python and the popular data analysis library, pandas and statsmodels. If you don't have Python and the required libraries installed, you can install them using pip or refer to the documentation for detailed instructions.

```python

import pandas as pd

import statsmodels.api as sm

# Create a DataFrame from the provided data

data = {

'Advertising Expenditure': [10, 15, 12, 8, 14],

'Sales': [25, 30, 28, 20, 26]

}

df = pd.DataFrame(data)

# Add a constant column for the intercept term

df = sm.add\_constant(df)

# Perform the multiple regression analysis

model = sm.OLS(df['Sales'], df[['const', 'Advertising Expenditure']])

results = model.fit()

# Print the regression results

print(results.summary())

```

The code above creates a DataFrame from the given data, adds a constant column for the intercept term, and then performs the multiple regression analysis using the Ordinary Least Squares (OLS) method. Finally, it prints the summary of the regression results.

The regression results will provide information about the coefficients (intercept and slope) of the regression equation, their statistical significance, goodness-of-fit measures, and other useful statistics. With these results, you can make predictions of sales based on new values of advertising expenditure.

Please note that the results might differ slightly depending on the statistical software or tool used, but the general approach remains the same.