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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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: 1. To calculate the mean sales for each region, you need to sum up the sales values in each region and divide by the number of data points.

For Region A:

Sales = [10, 15, 12, 8, 14]

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

For Region B:

Sales = [18, 20, 16, 22, 25]

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

Therefore, the mean sales for Region A is 11.8 and for Region B is 20.2.

2. To calculate the mode of the survey responses, you need to determine the value that appears most frequently in the data.

Survey Responses = [4, 5, 2, 3, 5, 4, 3, 2, 4, 5]

Mode = 4 (since 4 appears three times, which is more frequent than any other value)

Therefore, the mode of the survey responses is 4.

3. To calculate the median salary for each department, you need to arrange the salary data in ascending order and find the middle value. If there is an even number of data points, you take the average of the two middle values.

For Department A:

Salaries = [5000, 6000, 5500, 7000]

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

Median Salary = (5500 + 6000) / 2 = 11500 / 2 = 5750

For Department B:

Salaries = [4500, 5500, 5800, 6000, 5200]

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

Median Salary = 5500

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

4. To calculate the range of the stock prices, you need to find the difference between the highest and lowest values in the data.

Stock Prices = [25.5, 24.8, 26.1, 25.3, 24.9]

Range = highest value - lowest value = 26.1 - 24.8 = 1.3

Therefore, the range of the stock prices is 1.3.

5. To perform a t-test to determine if there is a significant difference in the mean scores between the two groups, you can use a two-sample independent t-test. The t-test compares the means of two groups to determine if they are significantly different from each other.

Group A: [85, 90, 92, 88, 91]

Group B: [82, 88, 90, 86, 87]

Performing a t-test will require the sample means, sample standard deviations, and sample sizes of both groups. Let's calculate them:

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) / (5-1)) = sqrt(8.8) ≈ 2.966

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) / (5-1)) = sqrt(5.2) ≈ 2.279

Using these values, you can perform a t-test to determine if there is a significant difference in the mean scores between the two groups. The t-test will provide a p-value, which indicates the probability of obtaining the observed difference (or a more extreme difference) if the null hypothesis is true (null hypothesis: there is no significant difference between the means of the two groups). If the p-value is below a chosen significance level (e.g., 0.05), then the difference is considered statistically significant.

6. To calculate the correlation coefficient between advertising expenditure and sales, you can use the Pearson correlation coefficient. The correlation coefficient measures the strength and direction of the linear relationship between two variables.

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

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

The correlation coefficient can be calculated using various formulas, but one common formula is:

Correlation coefficient = (n \* Σ(x \* y) - Σx \* Σy) / sqrt((n \* Σx^2 - (Σx)^2) \* (n \* Σy^2 - (Σy)^2))

Where n is the number of data points, Σ denotes summation, x represents advertising expenditure, and y represents sales.

Let's calculate the correlation coefficient using this formula:

n = 5

Σ(x \* y) = (10 \* 25) + (15 \* 30) + (12 \* 28) + (8 \* 20) + (14 \* 26) = 1050

Σx = 10 + 15 + 12 + 8 + 14 = 59

Σy = 25 + 30 + 28 + 20 + 26 = 129

Σx^2 = (10^2) + (15^2) + (12^2) + (8^2) + (14^2) = 445

Σy^2 = (25^2) + (30^2) + (28^2) + (20^2) + (26^2) = 2295

Correlation coefficient = (5 \* 1050 - 59 \* 129) / sqrt((5 \* 445 - 59^2) \* (5 \* 2295 - 129^2))

Calculate the numerator and denominator separately:

Numerator = 5 \* 1050 - 59 \* 129 = 5 \* 1050 - 7591 = 5250 - 7591 = -2341

Denominator = sqrt((5 \* 445 - 59^2) \* (5 \* 2295 - 129^2)) ≈ sqrt((2225 - 3481) \* (11475 - 16641)) ≈ sqrt((-1256) \* (-5166)) ≈ sqrt(6495696) ≈ 2549.48

Correlation coefficient ≈ -2341 / 2549.48 ≈ -0.917

Therefore, the correlation coefficient between advertising expenditure and sales is approximately -0.917, indicating a strong negative correlation.

7.

To calculate the standard deviation of the heights, you can use the following steps:

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

Step 1: Calculate the mean (average) of the heights.

Mean = (160 + 170 + 165 + 155 + 175 + 180 + 170) / 7 = 1175 / 7 = 167.86 (rounded to two decimal places)

Step 2: Calculate the difference between each height and the mean, and square the differences.

Differences from Mean = [160 - 167.86, 170 - 167.86, 165 - 167.86, 155 - 167.86, 175 - 167.86, 180 - 167.86, 170 - 167.86]

= [-7.86, 2.14, -2.86, -12.86, 7.14, 12.14, 2.14]

Squared Differences = [61.6396, 4.5796, 8.2096, 165.3796, 51.1396, 147.7156, 4.5796]

Step 3: Calculate the average of the squared differences.

Mean of Squared Differences = (61.6396 + 4.5796 + 8.2096 + 165.3796 + 51.1396 + 147.7156 + 4.5796) / 7 = 443.8436 / 7 = 63.4051 (rounded to four decimal places)

Step 4: Take the square root of the mean of squared differences to find the standard deviation.

Standard Deviation = sqrt(63.4051) ≈ 7.97 (rounded to two decimal places)

Therefore, the standard deviation of the heights is approximately 7.97.

8. To perform a linear regression analysis to predict job satisfaction based on employee tenure, you can use the given data:

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]

Linear regression aims to find a linear relationship between the predictor variable (employee tenure) and the response variable (job satisfaction). Here's how you can perform the analysis:

Step 1: Organize the data:

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

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

Step 2: Calculate the regression line:

- Use a statistical software or calculator to calculate the regression line, which represents the best-fit line for the given data. The regression line equation is of the form: y = mx + b, where m is the slope and b is the y-intercept.

- The slope (m) and y-intercept (b) can be calculated using formulas:

m = Σ((xi - x̄)(yi - ȳ)) / Σ((xi - x̄)²)

b = ȳ - m \* x̄

where Σ denotes summation, xi and yi are the individual data points, x̄ is the mean of x, and ȳ is the mean of y.

Step 3: Interpret the results:

- Once you have the regression line equation, you can use it to predict job satisfaction (y) based on employee tenure (x). The slope (m) indicates the change in job satisfaction for a one-unit increase in employee tenure, and the y-intercept (b) represents the expected job satisfaction when employee tenure is 0 (which might not make sense in this context).

Note: It's important to evaluate the assumptions of linear regression (e.g., linearity, independence, homoscedasticity, normality) and perform appropriate diagnostics to ensure the validity of the analysis.

9. To perform an analysis of variance (ANOVA) to determine if there is a significant difference in the mean recovery times between two medications, you can use the given data:

Medication A: [10, 12, 14, 11, 13]

Medication B: [15, 17, 16, 14, 18]

ANOVA compares the means of multiple groups to determine if there are significant differences between them. Here's how you can perform the analysis:

Step 1: Organize the data:

- Medication A: Group A = [10, 12, 14, 11, 13]

- Medication B: Group B = [15, 17, 16, 14, 18]

Step 2: Calculate the ANOVA:

- Use a statistical software or calculator to perform the ANOVA test, which calculates the F-statistic and associated p-value.

- The null hypothesis (H0) assumes that there is no significant difference in the mean recovery times between the two medications.

- The alternative hypothesis (Ha) assumes that there is a significant difference in the mean recovery times between the two medications.

Step 3: Interpret the results:

- If the p-value is below a chosen significance level (e.g., α = 0.05), you can reject the null hypothesis and conclude that there is a significant difference in the mean recovery times between the two medications.

- If the p-value is above the significance level, you fail to reject the null hypothesis, indicating that there is not enough evidence to suggest a significant difference in the mean recovery times.

Note: ANOVA assumes certain assumptions, including independence, normality, and homoscedasticity. It's important to verify these assumptions and conduct appropriate post-hoc tests if needed.

Please let me know if you'd like further guidance on any specific scenario.

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

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

Since you have 10 ratings, the 75th percentile can be calculated as follows:

Position = (75 / 100) \* (10 + 1)

Position = 7.5

The 75th percentile falls between the 7th and 8th ratings. To determine the value at the 75th percentile, you can take the average of these two ratings:

(7 + 8) / 2 = 7.5

Therefore, the 75th percentile of the feedback ratings is 7.5.

11. To perform a hypothesis test to determine if the mean weight differs significantly from 10 grams, you can use a one-sample t-test. Here are the steps for the hypothesis test:

Step 1: State the null and alternative hypotheses:

Null Hypothesis (H0): The mean weight is equal to 10 grams.

Alternative Hypothesis (Ha): The mean weight differs significantly from 10 grams.

Step 2: Set the significance level (alpha). Let's assume alpha = 0.05, which is a common choice.

Step 3: Calculate the sample mean and standard deviation from the given weights:

Sample Mean (x̄) = (10.2 + 9.8 + 10.0 + 10.5 + 10.3 + 10.1) / 6 = 10.17

Sample Standard Deviation (s) = √[(Σ(xi - x̄)²) / (n - 1)] = √[(0.07² + (-0.37)² + (-0.17)² + (0.33)² + (0.13)² + (-0.07)²) / (6 - 1)] ≈ 0.201

Step 4: Calculate the t-value using the formula:

t = (x̄ - μ) / (s / √n)

where μ is the hypothesized mean (10 grams), x̄ is the sample mean, s is the sample standard deviation, and n is the sample size.

t = (10.17 - 10) / (0.201 / √6) ≈ 1.695

Step 5: Determine the critical t-value. Since this is a two-tailed test, we need to find the t-value corresponding to the alpha/2 level of significance with (n - 1) degrees of freedom. In this case, (n - 1) = (6 - 1) = 5 degrees of freedom. Using a t-distribution table or statistical software, the critical t-value for alpha/2 = 0.025 and 5 degrees of freedom is approximately 2.571.

Step 6: Compare the calculated t-value with the critical t-value. If the calculated t-value is greater than the critical t-value or less than the negative of the critical t-value, we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

Since |1.695| < 2.571, we fail to reject the null hypothesis.

Therefore, based on the given sample, there is not enough evidence to conclude that the mean weight differs significantly from 10 grams.

12. To perform a chi-square test to determine if there is a significant difference in the click-through rates between the two designs, we first need to set up a contingency table. The contingency table will have two rows (Design A and Design B) and five columns (representing the number of clicks).

The contingency table looks like this:

Design A Design B

-------------------------------

Click | 100 80

Click | 120 85

Click | 110 90

Click | 90 95

Click | 95 100

To perform the chi-square test, we'll use the chi-square test for independence. Here's how you can calculate it:

Step 1: Set up the null and alternative hypotheses:

- Null Hypothesis (H0): There is no significant difference in click-through rates between the two designs.

- Alternative Hypothesis (HA): There is a significant difference in click-through rates between the two designs.

Step 2: Calculate the expected frequencies for each cell. The expected frequency for each cell is calculated by multiplying the row total and column total, then dividing by the grand total.

Step 3: Calculate the chi-square test statistic. This is done by summing the ((observed frequency - expected frequency)^2 / expected frequency) for each cell.

Step 4: Determine the degrees of freedom (df). The degrees of freedom is calculated as (number of rows - 1) \* (number of columns - 1).

Step 5: Determine the critical value for the chi-square test at a given significance level (e.g., α = 0.05). You can use a chi-square distribution table or a statistical software for this.

Step 6: Compare the calculated chi-square test statistic with the critical value. If the calculated chi-square value is greater than the critical value, we reject the null hypothesis and conclude that there is a significant difference in click-through rates between the two designs. Otherwise, we fail to reject the null hypothesis.

Perform these calculations to obtain the results of the chi-square test.

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

Confidence Interval = X̄ ± (t \* (s / √n))

Where:

- X̄ is the sample mean.

- t is the critical value for the t-distribution corresponding to the desired confidence level and degrees of freedom.

- s is the sample standard deviation.

- n is the sample size.

Given the data collected: [7, 9, 6, 8, 10, 7, 8, 9, 7, 8]

Step 1: Calculate the sample mean (X̄):

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

Step 2: Calculate the sample standard deviation (s):

s = √[(Σ(x - X̄)^2) / (n - 1)]

= √[((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)]

= √[(0.81 + 1.21 + 2.89 + 0.01 + 4.41 + 0.81 + 0.01 + 1.21 + 0.81 + 0.01) / 9]

≈ √(12.27 / 9)

≈ √1.363 ≈ 1.17

Step 3: Determine the critical value for the t-distribution for a 95% confidence level with 9 degrees of freedom. You can use a t-distribution table or a statistical software for this. For a two-tailed test, the critical value is approximately 2.262.

Step 4: Calculate the confidence interval:

Confidence Interval = 7.9 ± (2.262 \* (1.17 / √10))

= 7.9 ± (2.262 \* 0.369)

= 7.9 ± 0.835

≈ (7.065, 8.735)

Therefore, the 95% confidence interval for the population mean satisfaction score is approximately (7.065, 8.735).

14. To perform a simple linear regression to predict performance based on temperature, we'll use the given data:

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

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

Step 1: Calculate the mean of temperature (X̄) and performance (Ȳ):

X̄ = (20 + 22 + 23 + 19 + 21) / 5 = 105 / 5 = 21

Ȳ = (8 + 7 + 9 + 6 + 8) / 5 = 38 / 5 = 7.6

Step 2: Calculate the deviation of each temperature (X) from the mean:

X = [20 - 21, 22 - 21, 23 - 21, 19 - 21, 21 - 21]

= [-1, 1, 2, -2, 0]

Step 3: Calculate the deviation of each performance (Y) from the mean:

Y = [8 - 7.6, 7 - 7.6, 9 - 7.6, 6 - 7.6, 8 - 7.6]

= [0.4, -0.6, 1.4, -1.6, 0.4]

Step 4: Calculate the sum of the products of the deviations:

Σ(X \* Y) = (-1 \* 0.4) + (1 \* -0.6) + (2 \* 1.4) + (-2 \* -1.6) + (0 \* 0.4)

= -0.4 - 0.6 + 2.8 + 3.2 + 0

= 5

Step 5: Calculate the sum of the squared deviations of X:

Σ(X^2) = (-1)^2 + 1^2 + 2^2 + (-2)^2 + 0^2

= 1 + 1 + 4 + 4 + 0

= 10

Step 6: Calculate the slope (β1):

β1 = Σ(X \* Y) / Σ(X^2)

= 5 / 10

= 0.5

Step 7: Calculate the intercept (β0):

β

0 = Ȳ - (β1 \* X̄)

= 7.6 - (0.5 \* 21)

= 7.6 - 10.5

= -2.9

The equation of the regression line is: Performance = -2.9 + 0.5 \* Temperature.

15. To perform a Mann-Whitney U test to determine if there is a significant difference in the median preferences between the two groups, we can use the given data:

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

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

Step 1: Combine the data from both groups and rank the observations from lowest to highest, assigning the same rank for ties.

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

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

Step 2: Calculate the sum of ranks for each group.

Sum of ranks for Group A = 7 + 3.5 + 10 + 1 + 7 = 28.5

Sum of ranks for Group B = 3.5 + 1 + 7 + 3.5 + 3.5 = 19

Step 3: Calculate the U statistic for each group.

U statistic for Group A = n1 \* n2 + (n1 \* (n1 + 1)) / 2 - sum of ranks for Group A

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

= 25 + (5 \* 6) / 2 - 28.5

= 25 + 15 - 28.5

= 11.5

U statistic for Group B = n1 \* n2 + (n2 \* (n2 + 1)) / 2 - sum of ranks for Group B

= 5 \* 5 + (5 \* (5 + 1)) / 2 - 19

= 25 + (5 \* 6) / 2 - 19

= 25 + 15 - 19

= 21

Step 4: Determine the smaller U value (Umin) between the two groups. In this case, Umin = 11.5.

Step 5: Calculate the critical U value for a given significance level (e.g., α = 0.05) and sample sizes n1 and n2. You can use a Mann-Whitney U critical value table or a statistical software for this.

Step 6: Compare the Umin value with the critical U value. If the Umin value is less than or equal to the critical U value, we reject the null hypothesis and conclude that there is a significant difference in the median preferences between the two groups. Otherwise, we fail to reject the null hypothesis.

Perform these calculations to obtain the results of the Mann-Whitney U test.

16. To calculate the interquartile range (IQR) of the ages, we can use the given data:

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

Step 1: Sort the data in ascending order:

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

Step 2: Calculate the first quartile (Q1) and the third quartile (Q3).

Q1: The median of the lower half of the data.

Q3: The median of the upper half of the data.

Since the dataset has an even number of values, we'll take the average of the two middle values to calculate the quartiles.

Q1: (30 + 35) / 2 = 32.5

Q3: (55 + 60) / 2 = 57.5

Step 3: Calculate the interquartile range (IQR):

IQR = Q3 - Q1 = 57.5 - 32.5 = 25

Therefore, the interquartile range (IQR) of the ages is 25.

17. To perform a Kruskal-Wallis test, we compare the median accuracy scores between the three machine learning algorithms. Here are the steps to conduct the test:

Step 1: Define the hypotheses:

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

- Alternative hypothesis (H1): There is a significant difference in the median accuracy scores between the algorithms.

Step 2: Rank the data:

- Combine all the accuracy scores from the three algorithms into a single list and assign ranks to each score, ignoring ties.

Algorithm A: [0.85, 0.80, 0.82, 0.87, 0.83] (Ranked as 3, 1, 2, 5, 4)

Algorithm B: [0.78, 0.82, 0.84, 0.80, 0.79] (Ranked as 1, 3, 4, 2, 5)

Algorithm C: [0.90, 0.88, 0.89, 0.86, 0.87] (Ranked as 5, 3, 4, 1, 2)

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

Algorithm A: 3 + 1 + 2 + 5 + 4 = 15

Algorithm B: 1 + 3 + 4 + 2 + 5 = 15

Algorithm C: 5 + 3 + 4 + 1 + 2 = 15

Step 4: Calculate the test statistic:

- The test statistic for the Kruskal-Wallis test is calculated using the following formula:

H = (12 / (n \* (n + 1))) \* ∑((Ri^2) / ni) - 3 \* (n + 1)

where n is the total number of observations, Ri is the sum of ranks for the ith algorithm, and ni is the number of observations for the ith algorithm.

For our example, n = 5 (since we have 5 observations in each algorithm).

H = (12 / (5 \* (5 + 1))) \* ((15^2 / 5) + (15^2 / 5) + (15^2 / 5)) - 3 \* (5 + 1)

= (12 / 30) \* (45 + 45 + 45) - 3 \* 6

= (2 / 5) \* 135 - 18

= 54 - 18

= 36

Step 5: Determine the critical value:

- The critical value for the Kruskal-Wallis test depends on the significance level (α) and the degrees of freedom (df).

- With three algorithms, df = k - 1 = 3 - 1 = 2, where k is the number of algorithms.

- Look up the critical value in the Kruskal-Wallis table or use a statistical software tool. Let's assume a significance level of α = 0.05.

- In this case, the critical value is 5.991 (for α = 0.05 and df = 2).

Step 6: Make a decision:

- If the test statistic (H) is greater than the critical value, reject the null hypothesis. Otherwise, fail to reject the null hypothesis.

In our example, H = 36 and the critical value is 5.991. Since H > critical value, we reject the null hypothesis.

Therefore, there is a significant difference in the median accuracy scores between the algorithms.

18. To perform a simple linear regression analysis to predict sales based on price, follow these steps:

Step 1: Plot the data:

- Create a scatter plot with Price on the x-axis and Sales on the y-axis to visualize the relationship between the variables.

Step 2: Calculate the regression line:

- Use the least squares method to estimate the regression line equation: y = mx + b.

- Calculate the slope (m) and intercept (b) of the regression line using the following formulas:

m = (nΣxy - ΣxΣy) / (nΣx^2 - (Σx)^2)

b = (Σy - mΣx) / n

where n is the number of observations, Σxy is the sum of the product of x and y, Σx is the sum of x, and Σy is the sum of y.

In our example:

n = 5

Σx = 10 + 15 + 12 + 8 + 14 = 59

Σy = 100 + 80 + 90 + 110 + 95 = 475

Σxy = (10 \* 100) + (15 \* 80) + (12 \* 90) + (8 \* 110) + (14 \* 95) = 4250

Σx^2 = (10^2) + (15^2) + (12^2) + (8^2) + (14^2) = 715

m = (5 \* 4250 - 59 \* 475) / (5 \* 715 - 59^2)

= (21250 - 28025) / (3575 - 3481)

= -675 / 94

≈ -7.18

b = (475 - (-7.18 \* 59)) / 5

= (475 + 424.02) / 5

≈ 179.00 / 5

≈ 35.80

The regression line equation is: y = -7.18x + 35.80

Step 3: Interpret the results:

- The slope (m) represents the change in Sales for each unit increase in Price. In this case, it indicates that for each dollar increase in Price, Sales decrease by approximately $7.18.

- The intercept (b) represents the Sales value when Price is zero. However, in this context, it may not have a meaningful interpretation since Price is unlikely to be zero.

Step 4: Evaluate the model:

- Assess the goodness of fit of the linear regression model by calculating the coefficient of determination (R-squared) and examining the residuals. Additional statistical tests can be conducted to determine the statistical significance of the relationship.

19. To calculate the standard error of the mean satisfaction score, follow these steps:

Step 1: Calculate the mean (average) of the satisfaction scores:

- Add up all the scores and divide by the total number of scores.

In our example:

Satisfaction scores: [7, 8, 9, 6, 8, 7, 9, 7, 8, 7]

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

= 76 / 10

= 7.6

Step 2: Calculate the standard deviation (SD) of the satisfaction scores:

- Calculate the square root of the average squared deviation from the mean.

- The formula for the sample standard deviation is slightly different from the population standard

deviation when calculating it for a sample.

In our example:

Squared deviations: [(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]

= [0.36, 0.16, 1.96, 2.56, 0.16, 0.36, 1.96, 0.36, 0.16, 0.36]

Sum of squared deviations = 0.36 + 0.16 + 1.96 + 2.56 + 0.16 + 0.36 + 1.96 + 0.36 + 0.16 + 0.36

= 8.92

Sample variance = Sum of squared deviations / (n - 1) = 8.92 / (10 - 1) = 1.1125

Sample standard deviation = √(Sample variance) = √(1.1125) ≈ 1.055

Step 3: Calculate the standard error of the mean:

- Divide the sample standard deviation by the square root of the sample size (n) to obtain the standard error of the mean.

In our example:

Standard error of the mean = Sample standard deviation / √(n) = 1.055 / √(10) ≈ 0.333

Therefore, the standard error of the mean satisfaction score is approximately 0.333.

20. To perform a multiple regression analysis to predict sales based on advertising expenditure, follow these steps:

Step 1: Organize the data:

- Create a dataset with two variables: Advertising Expenditure (independent variable) and Sales (dependent variable).

- The dataset should include multiple observations (rows) with corresponding values for both variables.

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

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

Step 2: Build the multiple regression model:

- Use statistical software or tools to perform the regression analysis. The software will estimate the regression coefficients and provide other statistical measures.

Step 3: Interpret the results:

- The regression model will provide the estimated coefficients for each independent variable (advertising expenditure) and an intercept term.

- The coefficients represent the relationship between the independent variable and the dependent variable, controlling for other variables in the model.

- In this case, the coefficient for advertising expenditure will indicate the expected change in sales for each unit increase in advertising expenditure, assuming other factors remain constant.

Step 4: Assess the model fit:

- Evaluate the overall fit of the multiple regression model by examining measures such as R-squared, adjusted R-squared, F-statistic, and p-values.

- These measures provide information about the strength of the relationship between the independent variable(s) and the dependent variable, as well as the overall significance of the model.

Step 5: Validate the assumptions:

- Multiple regression analysis relies on certain assumptions, including linearity, independence, normality, and homoscedasticity.

- It's important to assess whether these assumptions hold for the given dataset. Residual analysis and statistical tests can help in this evaluation.

By following these steps, you can perform a multiple regression analysis to predict sales based on advertising expenditure.