Assignment 3

**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:** For Region A:

Add up all the sales values: 10 + 15 + 12 + 8 + 14 = 59

Divide the sum by the number of data points (in this case, 5): 59 / 5 = 11.8

The mean sales for Region A is 11.8.

For Region B:

Add up all the sales values: 18 + 20 + 16 + 22 + 25 = 101

Divide the sum by the number of data points (in this case, 5): 101 / 5 = 20.2

The mean sales for Region B is 20.2.

Therefore, 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.**

**Ans:** Value 2 appears 2 times.   
Value 3 appears 2 times.  
Value 4 appears 3 times.  
 Value 5 appears 3 times.

The mode is the value that appears most frequently. In this case, both 4 and 5 appear 3 times, so the mode of the survey responses is 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.**

**Ans**:

For Department A:

Arrange the salaries in ascending order: [5000, 5500, 6000, 7000]

Since the number of data points is even, take the average of the two middle values:

The middle values are 5500 and 6000.

Add the two values together and divide by 2: (5500 + 6000) / 2 = 5750

The median salary for Department A is 5750.

For Department B:

Arrange the salaries in ascending order: [4500, 5200, 5500, 5800, 6000]

Since the number of data points is odd, the median is the middle value:

The middle value is 5500.

The median salary for Department B is 5500.

Therefore, the median salary for Department A is 5750, and the median salary 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 highest and lowest values in the data set. Here's how you can do that:

Arrange the stock prices in ascending order: [24.8, 24.9, 25.3, 25.5, 26.1]

The lowest value is 24.8 and the highest value is 26.1.

Subtract the lowest value from the highest value: 26.1 - 24.8 = 1.3

The range of the stock prices is 1.3.

Therefore, the range of the daily stock prices for the company 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 and determine if there is a significant difference in the mean scores between Group A and Group B, you can follow these steps:

Step 1: Calculate the means of each group.

Group A mean: (85 + 90 + 92 + 88 + 91) / 5 = 89.2

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

Step 2: Calculate the standard deviations of each group.

Group A standard deviation (sA): 2.7 (rounded to one decimal place)

Group B standard deviation (sB): 2.7 (rounded to one decimal place)

Step 3: Calculate the t-value using the formula:

t = (meanA - meanB) / sqrt((sA^2 / nA) + (sB^2 / nB))

Where:

meanA and meanB are the means of Group A and Group B, respectively.

sA and sB are the standard deviations of Group A and Group B, respectively.

nA and nB are the sample sizes of Group A and Group B, respectively.

Plugging in the values:

t = (89.2 - 86.6) / sqrt((2.7^2 / 5) + (2.7^2 / 5))

Step 4: Determine the degrees of freedom (df). Since both groups have the same sample size, df = nA + nB - 2.

In this case, df = 5 + 5 - 2 = 8.

Step 5: Look up the critical t-value for your desired level of significance and degrees of freedom. Let's assume a significance level of 0.05 (5%) and a two-tailed test.

Using a t-table or statistical software, you can find the critical t-value. For df = 8 and a significance level of 0.05, the critical t-value is approximately 2.306.

Step 6: Compare the calculated t-value to the critical t-value. If the calculated t-value is greater than the critical t-value, there is a significant difference in the mean scores between the two groups.

If the calculated t-value is less than the critical t-value, there is not enough evidence to conclude a significant difference.

In this case, you can calculate the t-value and compare it to the critical t-value to determine if there is a 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 the advertising expenditure and sales data.

Advertising Expenditure:

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

Sales:

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

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

Advertising Expenditure deviations: [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]

Sales deviations: [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 data points.

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.

Advertising Expenditure squared deviations: [(-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]

Sales squared deviations: [(-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: Sum up the product of deviations, squared deviations for each variable, and their respective sums of squared deviations.

Sum of product of deviations = 1.44 + 13.44 + 0.44 + 22.04 + 0.44 = 37.8

Sum of squared deviations for Advertising Expenditure = 3.24 + 10.24 + 0.04 + 14.44 + 4.84 = 33.8

Sum of squared deviations for Sales = 0.64 + 17.64 + 4.84 + 33.64 + 0.04 = 57.8

Step 6: Calculate the correlation coefficient using the formula:

r = (n \* sum of product of deviations) / sqrt(sum of squared deviations for Advertising Expenditure \* sum of squared deviations for Sales)

Where n is the number of data points.

r = (5 \* 37.8) / sqrt(33.8 \* 57.8) ≈ 0.922

Step 7: Interpret the correlation coefficient.

The correlation coefficient (r) between advertising expenditure and sales is approximately 0.922. This indicates a strong positive correlation, suggesting that there is a significant relationship between the two variables. As the advertising expenditure increases, the sales tend to increase as well.

Note: The correlation coefficient ranges between -1 and 1. A value close to 1 indicates a strong positive correlation, close to -1 indicates a strong negative correlation, and close to 0 indicates no correlation.

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

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

Mean = (160 + 170 + 165 + 155 + 175 + 180 + 170) / 7 = 166.43 (rounded to two decimal places)

Step 2: Calculate the deviations from the mean for each height.

Deviations from the mean: [160 - 166.43, 170 - 166.43, 165 - 166.43, 155 - 166.43, 175 - 166.43, 180 - 166.43, 170 - 166.43]

= [-6.43, 3.57, -1.43, -11.43, 8.57, 13.57, 3.57]

Step 3: Square each deviation.

Squared deviations: [(-6.43)^2, (3.57)^2, (-1.43)^2, (-11.43)^2, (8.57)^2, (13.57)^2, (3.57)^2]

= [41.3449, 12.7449, 2.0449, 130.6249, 73.5049, 184.3249, 12.7449]

Step 4: Calculate the average of the squared deviations (variance).

Variance = (41.3449 + 12.7449 + 2.0449 + 130.6249 + 73.5049 + 184.3249 + 12.7449) / 7 = 53.1832 (rounded to four decimal places)

Step 5: Take the square root of the variance to find the standard deviation.

Standard Deviation = √(53.1832) ≈ 7.2925 (rounded to four decimal places)

Therefore, the standard deviation of the heights is approximately 7.2925.

**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**: To perform a linear regression analysis to predict job satisfaction based on employee tenure, you can follow these steps:

Step 1: Organize the data into two variables: employee tenure and job satisfaction.

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]

Step 2: Calculate the means of both variables.

Mean of Employee Tenure = (2 + 3 + 5 + 4 + 6 + 2 + 4) / 7 = 3.5714 (rounded to four decimal places)

Mean of Job Satisfaction = (7 + 8 + 6 + 9 + 5 + 7 + 6) / 7 = 6.8571 (rounded to four decimal places)

Step 3: Calculate the deviations from the means for both variables.

Deviations from the mean of Employee Tenure: [-1.5714, -0.5714, 1.4286, 0.4286, 2.4286, -1.5714, 0.4286]

Deviations from the mean of Job Satisfaction: [0.1429, 1.1429, -0.8571, 2.1429, -1.8571, 0.1429, -0.8571]

Step 4: Calculate the product of the deviations for each pair of data points.

Product of deviations: [-1.5714 \* 0.1429, -0.5714 \* 1.1429, 1.4286 \* -0.8571, 0.4286 \* 2.1429, 2.4286 \* -1.8571, -1.5714 \* 0.1429, 0.4286 \* -0.8571]

= [-0.2245, -0.6531, -1.2245, 0.9184, -4.5143, -0.2245, -0.3671]

Step 5: Calculate the squared deviations for both variables.

Squared deviations of Employee Tenure: [(-1.5714)^2, (-0.5714)^2, (1.4286)^2, (0.4286)^2, (2.4286)^2, (-1.5714)^2, (0.4286)^2]

= [2.4640, 0.3265, 2.0408, 0.1837, 5.8980, 2.4640, 0.1837]

Squared deviations of Job Satisfaction: [(0.1429)^2, (1.1429)^2, (-0.8571)^2, (2.1429)^2, (-1.8571)^2, (0.1429)^2, (-0.8571)^2]

= [0.0204, 1.3061, 0.7347, 4.5918, 3.4489, 0.0204, 0.7347]

Step 6: Sum up the product of deviations and squared deviations for each variable.

Sum of product of deviations = -0.2245 + -0.6531 + -1.2245 + 0.9184 + -4.5143 + -0.2245 + -0.3671 = -6.1264

Sum of squared deviations of Employee Tenure = 2.4640 + 0.3265 + 2.0408 + 0.1837 + 5.8980 + 2.4640 + 0.1837 = 13.5607

Sum of squared deviations of Job Satisfaction = 0.0204 + 1.3061 + 0.7347 + 4.5918 + 3.4489 + 0.0204 + 0.7347 = 10.857

Step 7: Calculate the slope (β1) of the regression line using the formula:

β1 = (Sum of product of deviations) / (Sum of squared deviations of Employee Tenure)

β1 = -6.1264 / 13.5607 ≈ -0.4514 (rounded to four decimal places)

Step 8: Calculate the y-intercept (β0) of the regression line using the formula:

β0 = Mean of Job Satisfaction - (β1 \* Mean of Employee Tenure)

β0 = 6.8571 - (-0.4514 \* 3.5714) ≈ 8.4301 (rounded to four decimal places)

Step 9: Write the equation of the regression line.

The equation of the regression line is: Job Satisfaction = -0.4514 \* Employee Tenure + 8.4301

Step 10: Interpret the results.

The slope (β1) of the regression line indicates that, on average, for each additional year of employee tenure, job satisfaction is expected to decrease by approximately 0.4514 units.

The y-intercept (β0) represents the predicted job satisfaction when the employee tenure is zero. In this case, it is approximately 8.4301.

Therefore, based on the linear regression analysis, the equation Job Satisfaction = -0.4514 \* Employee Tenure + 8.4301 can be used 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.**

**Ans:** To perform an analysis of variance (ANOVA) to determine if there is a significant difference in the mean recovery times between Medication A and Medication B, you can follow these steps:

Step 1: Organize the data into two groups: Medication A and Medication B.

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

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

Step 2: Calculate the means of each group.

Mean of Medication A: (10 + 12 + 14 + 11 + 13) / 5 = 12 (rounded to the nearest whole number)

Mean of Medication B: (15 + 17 + 16 + 14 + 18) / 5 = 16 (rounded to the nearest whole number)

Step 3: Calculate the sum of squares between groups (SSB).

SSB = (nA \* (meanA - meanTotal)^2) + (nB \* (meanB - meanTotal)^2)

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

= 20 + 20

= 40

Where nA and nB are the sample sizes of Medication A and Medication B, respectively.

meanA and meanB are the means of Medication A and Medication B, respectively.

meanTotal is the overall mean of the combined data.

Step 4: Calculate the sum of squares within groups (SSW).

SSW = (nA - 1) \* varianceA + (nB - 1) \* varianceB

= (5 - 1) \* ((10 - 12)^2 + (12 - 12)^2 + (14 - 12)^2 + (11 - 12)^2 + (13 - 12)^2)

+ (5 - 1) \* ((15 - 16)^2 + (17 - 16)^2 + (16 - 16)^2 + (14 - 16)^2 + (18 - 16)^2)

= 4 \* (4 + 0 + 4 + 1 + 1) + 4 \* (1 + 1 + 0 + 4 + 4)

= 4 \* 10 + 4 \* 10

= 40 + 40

= 80

Where varianceA and varianceB are the variances of Medication A and Medication B, respectively.

Step 5: Calculate the degrees of freedom between groups (dfB) and within groups (dfW).

dfB = k - 1

= 2 - 1

= 1

Where k is the number of groups.

dfW = N - k

= 10 - 2

= 8

Where N is the total number of data points.

Step 6: Calculate the mean squares between groups (MSB) and within groups (MSW).

MSB = SSB / dfB

= 40 / 1

= 40

MSW = SSW / dfW

= 80 / 8

= 10

Step 7: Calculate the F-statistic using the formula:

F = MSB / MSW

= 40 / 10

= 4

Step 8: Look up the critical F-value for your desired level of significance, degrees of freedom between groups (dfB), and degrees of freedom within groups (dfW). Let's assume a significance level of 0.05 (5%).

Using an F-table or statistical software, you can find the critical F-value. For dfB = 1 and dfW = 8, the critical F-value is approximately 5.32.

Step 9: Compare the calculated F-value to the critical F-value. If the calculated F-value is greater than the critical F-value, there is a significant difference in the mean recovery times between the two medications.

If the calculated F-value is less than the critical F-value, there is not enough evidence to conclude a significant difference.

In this case, the calculated F-value is 4, which is less than the critical F-value of 5.32.

Therefore, based on the ANOVA analysis, there is not enough evidence to conclude a significant difference in the mean recovery times between Medication A and Medication B at the 0.05 significance level.

**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:** To calculate the 75th percentile of the feedback ratings, you need to find the value below which 75% of the data falls. Here's how you can do that:

Step 1: Arrange the feedback ratings in ascending order: [6, 7, 7, 8, 8, 8, 8, 9, 9, 10]

Step 2: Calculate the index corresponding to the 75th percentile.

Index = (75 / 100) \* (N + 1)

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

= (75 / 100) \* 11

= 8.25

Since the index is not a whole number, we can interpolate to find the value. Interpolation involves taking the average of the values at the index and the next index.

The value at the 8th index is 8, and the value at the 9th index is 9.

Interpolated value = 8 + (0.25 \* (9 - 8))

= 8 + 0.25

= 8.25

Therefore, the 75th percentile of the feedback ratings is 8.25.

**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:** To perform a hypothesis test and determine if the mean weight differs significantly from 10 grams, you can follow these steps:

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

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

Alternative hypothesis (Ha): The mean weight is not equal to 10 grams.

Step 2: Set the significance level (α). Let's assume a 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 = 10.17 (rounded to two decimal places)

Sample standard deviation (s) = sqrt(((10.2 - 10.17)^2 + (9.8 - 10.17)^2 + (10.0 - 10.17)^2 + (10.5 - 10.17)^2 + (10.3 - 10.17)^2 + (10.1 - 10.17)^2) / (6 - 1))

= sqrt((0.0064 + 0.0064 + 0.0009 + 0.0081 + 0.0064 + 0.0064) / 5)

= sqrt(0.0346 / 5)

= sqrt(0.00692)

≈ 0.0832 (rounded to four decimal places)

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

t = (x̄ - μ) / (s / sqrt(n))

Where:

x̄ is the sample mean,

μ is the population mean (assumed to be 10 grams),

s is the sample standard deviation, and

n is the sample size.

Plugging in the values:

t = (10.17 - 10) / (0.0832 / sqrt(6))

≈ 0.17 / (0.0832 / sqrt(6))

≈ 0.17 / (0.0832 / 2.449)

≈ 0.17 / 0.03398

≈ 5.006 (rounded to three decimal places)

Step 5: Determine the degrees of freedom (df). The degrees of freedom is equal to the sample size minus 1.

In this case, df = 6 - 1 = 5.

Step 6: Look up the critical t-value for your desired level of significance (α) and degrees of freedom (df). For a two-tailed test and α = 0.05, the critical t-value is approximately ±2.571.

Step 7: Compare the calculated t-value to the critical t-value. If the calculated t-value falls within the rejection region (i.e., outside the range of the critical t-values), then the null hypothesis is rejected. Otherwise, if the calculated t-value falls within the acceptance region (i.e., within the range of the critical t-values), the null hypothesis cannot be rejected.

In this case, the calculated t-value is 5.006, which falls outside the range of the critical t-values of ±2.571.

Step 8: Make a decision. Since the calculated t-value falls in the rejection region, we reject the null hypothesis.

Therefore, based on the hypothesis test, there is enough evidence to suggest that the mean weight differs significantly from 10 grams for the given sample.

**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:

To perform a chi-square test and determine if there is a significant difference in the click-through rates between Design A and Design B, you can follow these steps:

Step 1: Organize the data into two categories: Design A and Design B.

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

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

Step 2: Calculate the observed frequencies for each category.

Observed frequency for Design A: 100, 120, 110, 90, 95

Observed frequency for Design B: 80, 85, 90, 95, 100

Step 3: Calculate the expected frequencies for each category assuming no difference between the designs. To do this, you can calculate the overall click-through rate and multiply it by the total number of observations for each design.

Total observations for Design A: 100 + 120 + 110 + 90 + 95 = 515

Total observations for Design B: 80 + 85 + 90 + 95 + 100 = 450

Overall click-through rate: (515 + 450) / (5 + 5) = 965 / 10 = 96.5

Expected frequency for Design A: 96.5 \* (515 / 965) ≈ 51.54

Expected frequency for Design B: 96.5 \* (450 / 965) ≈ 44.96

Step 4: Calculate the chi-square test statistic using the formula:

χ² = ∑ ((O - E)² / E)

Where O is the observed frequency and E is the expected frequency.

For Design A:

χ²(A) = ((100 - 51.54)² / 51.54) + ((120 - 51.54)² / 51.54) + ((110 - 51.54)² / 51.54) + ((90 - 51.54)² / 51.54) + ((95 - 51.54)² / 51.54)

For Design B:

χ²(B) = ((80 - 44.96)² / 44.96) + ((85 - 44.96)² / 44.96) + ((90 - 44.96)² / 44.96) + ((95 - 44.96)² / 44.96) + ((100 - 44.96)² / 44.96)

Step 5: Calculate the degrees of freedom (df). The degrees of freedom is equal to the number of categories minus 1.

In this case, df = 2 - 1 = 1.

Step 6: Look up the critical chi-square value for your desired level of significance (α) and degrees of freedom (df). Let's assume a significance level of 0.05 (5%).

Using a chi-square table or statistical software, you can find the critical chi-square value. For df = 1 and α = 0.05, the critical chi-square value is approximately 3.841.

Step 7: Compare the calculated chi-square value to the critical chi-square value. If the calculated chi-square value is greater than the critical chi-square value, there is a significant difference in the click-through rates between the designs.

If the calculated chi-square value is less than the critical chi-square value, there is not enough evidence to conclude a significant difference.

For Design A, calculate χ²(A) and compare it to the critical chi-square value.

For Design B, calculate χ²(B) and compare it to the critical chi-square value.

In this case, χ²(A) and χ²(B) are calculated using the formulas in Step 4.

Step 8: Make a decision. If both χ²(A) and χ²(B) are greater than the critical chi-square value, we can reject the null hypothesis and conclude that there is a significant difference in the click-through rates between the designs.

Therefore, you can calculate χ²(A) and χ²(B) and compare them to the critical chi-square value of 3.841 to determine if there is a significant difference in the click-through rates between Design A and Design B.

**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, you can follow these steps:

Step 1: Organize the data into a sample.

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

Step 2: Calculate the sample mean (x̄) and the sample standard deviation (s).

Sample mean (x̄) = (7 + 9 + 6 + 8 + 10 + 7 + 8 + 9 + 7 + 8) / 10 = 7.9 (rounded to one decimal place)

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 + 2.89 + 0.01 + 4.41 + 0.81 + 0.01 + 1.21 + 0.81 + 0.01) / 9)

= sqrt(12.07 / 9)

≈ sqrt(1.34)

≈ 1.16 (rounded to two decimal places)

Step 3: Determine the sample size (n).

Sample size (n) = 10

Step 4: Determine the desired level of confidence (1 - α). In this case, the confidence level is 95%, so α = 0.05.

Step 5: Calculate the standard error (SE) using the formula:

SE = s / sqrt(n)

SE = 1.16 / sqrt(10)

≈ 0.366 (rounded to three decimal places)

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

ME = critical value \* SE

The critical value depends on the desired level of confidence and the distribution (t-distribution in this case). For a 95% confidence level and 9 degrees of freedom (n - 1), the critical value can be obtained from a t-table or statistical software. In this case, the critical value is approximately 2.262.

ME = 2.262 \* 0.366

≈ 0.828 (rounded to three decimal places)

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

Lower bound = x̄ - ME

= 7.9 - 0.828

≈ 7.072 (rounded to three decimal places)

Upper bound = x̄ + ME

= 7.9 + 0.828

≈ 8.728 (rounded to three decimal places)

Step 8: Write the confidence interval.

The 95% confidence interval for the population mean satisfaction score is approximately 7.072 to 8.728.

Therefore, based on the given data, we can be 95% confident that the true population mean satisfaction score falls within this range.

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: To perform a simple linear regression and predict performance based on temperature, you can follow these steps:

Step 1: Organize the data into two variables: temperature and performance.

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

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

Step 2: Calculate the means of both variables.

Mean of Temperature = (20 + 22 + 23 + 19 + 21) / 5 = 21

Mean of Performance = (8 + 7 + 9 + 6 + 8) / 5 = 7.6

Step 3: Calculate the deviations from the means for each variable.

Deviations from the mean of Temperature: [-1, 1, 2, -2, 0]

Deviations from the mean of Performance: [0.4, -0.6, 1.4, -1.6, 0.4]

Step 4: Calculate the product of the deviations for each pair of data points.

Product of deviations: [-1 \* 0.4, 1 \* -0.6, 2 \* 1.4, -2 \* -1.6, 0 \* 0.4]

= [-0.4, -0.6, 2.8, 3.2, 0]

Step 5: Calculate the squared deviations for each variable.

Squared deviations of Temperature: [(-1)^2, 1^2, 2^2, (-2)^2, 0^2]

= [1, 1, 4, 4, 0]

Squared deviations of Performance: [0.4^2, (-0.6)^2, 1.4^2, (-1.6)^2, 0.4^2]

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

Step 6: Calculate the sums of the product of deviations and squared deviations.

Sum of product of deviations = -0.4 + -0.6 + 2.8 + 3.2 + 0 = 5

Sum of squared deviations of Temperature = 1 + 1 + 4 + 4 + 0 = 10

Sum of squared deviations of Performance = 0.16 + 0.36 + 1.96 + 2.56 + 0.16 = 5.2

Step 7: Calculate the slope (β1) of the regression line using the formula:

β1 = (Sum of product of deviations) / (Sum of squared deviations of Temperature)

β1 = 5 / 10 = 0.5

Step 8: Calculate the y-intercept (β0) of the regression line using the formula:

β0 = Mean of Performance - (β1 \* Mean of Temperature)

β0 = 7.6 - (0.5 \* 21) = -1

Step 9: Write the equation of the regression line.

The equation of the regression line is: Performance = 0.5 \* Temperature - 1

Step 10: Interpret the results.

The slope (β1) of the regression line indicates that, on average, for each increase of 1 degree Celsius in temperature, performance is expected to increase by 0.5 units.

The y-intercept (β0) represents the predicted performance when the temperature is 0 degrees Celsius. In this case, it is -1.

Therefore, based on the simple linear regression analysis, the equation Performance = 0.5 \* Temperature - 1 can be used 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.**

Ans: 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, you can follow these steps:

Step 1: Organize the data into two groups: Group A and Group B.

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

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

Step 2: Rank the data from both groups together, considering ties.

Ranks for Group A: [4, 2, 5, 1, 4]

Ranks for Group B: [2, 1, 4, 2, 2]

In case of ties, assign the average rank to the tied values. For example, the two 4s in Group A are assigned ranks 4 and 5, so their average rank is (4 + 5) / 2 = 4.5.

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

Sum of ranks for Group A = 4 + 2 + 5 + 1 + 4 = 16

Sum of ranks for Group B = 2 + 1 + 4 + 2 + 2 = 11

Step 4: Calculate the U statistic for each group.

U statistic for Group A = nA \* nB + (nA \* (nA + 1)) / 2 - Sum of ranks for Group A

= 5 \* 5 + (5 \* (5 + 1)) / 2 - 16

= 25 + (5 \* 6) / 2 - 16

= 25 + 15 - 16

= 24

U statistic for Group B = nA \* nB + (nB \* (nB + 1)) / 2 - Sum of ranks for Group B

= 5 \* 5 + (5 \* (5 + 1)) / 2 - 11

= 25 + (5 \* 6) / 2 - 11

= 25 + 15 - 11

= 29

Step 5: Determine the smaller U statistic (Umin) and the total number of observations (N).

Umin = min(U statistic for Group A, U statistic for Group B)

= min(24, 29)

= 24

N = nA + nB

= 5 + 5

= 10

Step 6: Calculate the expected value of U (E(U)) using the formula:

E(U) = (nA \* nB) / 2

E(U) = (5 \* 5) / 2

= 25 / 2

= 12.5

Step 7: Calculate the standard deviation of U (SD(U)) using the formula:

SD(U) = sqrt((nA \* nB \* (nA + nB + 1)) / 12)

SD(U) = sqrt((5 \* 5 \* (5 + 5 + 1)) / 12)

= sqrt((25 \* 11) / 12)

≈ sqrt(22.92)

≈ 4.79 (rounded to two decimal places)

Step 8: Calculate the z-score using the formula:

z = (Umin - E(U)) / SD(U)

z = (24 - 12.5) / 4.79

= 11.5 / 4.79

≈ 2.40 (rounded to two decimal places)

Step 9: Look up the critical z-value for your desired level of significance (α). Let's assume a significance level of 0.05 (5%).

Using a standard normal distribution table or statistical software, you can find the critical z-value. For α = 0.05, the critical z-value is approximately ±1.96 for a two-tailed test.

Step 10: Compare the calculated z-score to the critical z-value. If the calculated z-score falls within the rejection region (i.e., outside the range of the critical z-values), then the null hypothesis is rejected. Otherwise, if the calculated z-score falls within the acceptance region (i.e., within the range of the critical z-values), the null hypothesis cannot be rejected.

In this case, the calculated z-score is 2.40, which is greater than the critical z-value of 1.96.

Step 11: Make a decision. Since the calculated z-score falls in the rejection region, we reject the null hypothesis.

Therefore, based on the Mann-Whitney U test, there is enough evidence to suggest that there is a significant difference in the median preferences between Group A and Group B.

**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 : To calculate the interquartile range (IQR) of the ages, you can follow these steps:

Step 1: Organize the data in ascending order: [25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

Step 2: Determine the position of the first quartile (Q1) and the third quartile (Q3).

Q1 position = (1 \* (n + 1)) / 4

= (1 \* (10 + 1)) / 4

= 2.75

≈ 2.75 (rounded to two decimal places)

Q3 position = (3 \* (n + 1)) / 4

= (3 \* (10 + 1)) / 4

= 8.25

≈ 8.25 (rounded to two decimal places)

Step 3: Calculate the values of Q1 and Q3 using interpolation.

Q1 = Value at index 2 + (0.75 \* (Value at index 3 - Value at index 2))

= 30 + (0.75 \* (35 - 30))

= 30 + (0.75 \* 5)

= 30 + 3.75

= 33.75

Q3 = Value at index 8 + (0.25 \* (Value at index 9 - Value at index 8))

= 60 + (0.25 \* (65 - 60))

= 60 + (0.25 \* 5)

= 60 + 1.25

= 61.25

Step 4: Calculate the interquartile range (IQR) using the formula:

IQR = Q3 - Q1

= 61.25 - 33.75

= 27.5

Therefore, the interquartile range (IQR) of the ages is 27.5.

**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.**

Ans: To perform a Kruskal-Wallis test and determine if there is a significant difference in the median accuracy scores between Algorithm A, Algorithm B, and Algorithm C, you can follow these steps:

Step 1: Organize the data into three groups: Algorithm A, Algorithm B, and Algorithm C.

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]

Step 2: Rank the data from all groups together, considering ties.

Ranks for all data: [7, 2, 4, 10, 5, 1, 3, 8, 6, 9, 12, 14, 11, 13]

In case of ties, assign the average rank to the tied values. For example, the two 0.82 values in Algorithm A are assigned ranks 4 and 5, so their average rank is (4 + 5) / 2 = 4.5.

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

Sum of ranks for Algorithm A = 7 + 2 + 4.5 + 10 + 5 = 28.5

Sum of ranks for Algorithm B = 1 + 3 + 6 + 8 + 4.5 = 22.5

Sum of ranks for Algorithm C = 14 + 11 + 13 + 9 + 12 = 59

Step 4: Calculate the grand total of ranks (T).

T = (Number of groups) \* (Number of observations per group) \* ((Number of observations per group) + 1) / 2

T = 3 \* 5 \* (5 + 1) / 2

= 3 \* 5 \* 6 / 2

= 45

Step 5: Calculate the test statistic (H) using the formula:

H = (12 \* (T - (Sum of ranks for all groups)^2 / (Number of observations) - (Number of observations per group) \* (Number of groups + 1) / 12)) / ((Number of groups - 1) \* (Number of groups + 1))

H = (12 \* (45 - (28.5^2 + 22.5^2 + 59^2) / 15 - 5 \* (3 + 1) / 12)) / ((3 - 1) \* (3 + 1))

≈ 1.71 (rounded to two decimal places)

Step 6: Determine the degrees of freedom (df). The degrees of freedom is equal to the number of groups minus 1.

In this case, df = 3 - 1 = 2.

Step 7: Look up the critical chi-square value for your desired level of significance (α) and degrees of freedom (df). Let's assume a significance level of 0.05 (5%).

Using a chi-square table or statistical software, you can find the critical chi-square value. For df = 2 and α = 0.05, the critical chi-square value is approximately 5.991.

Step 8: Compare the calculated test statistic to the critical chi-square value. If the calculated test statistic is greater than the critical chi-square value, there is a significant difference in the median accuracy scores between the algorithms.

In this case, the calculated test statistic (H) is approximately 1.71, which is less than the critical chi-square value of 5.991.

Step 9: Make a decision. Since the calculated test statistic is less than the critical chi-square value, we fail to reject the null hypothesis.

Therefore, based on the Kruskal-Wallis test, there is not enough evidence to suggest a significant difference in the median accuracy scores between Algorithm A, Algorithm B, and Algorithm C.

**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**

 Ans: To perform a simple linear regression and predict the effect of price on sales, you can follow these steps:

Step 1: Organize the data into two variables: price and sales.

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

Sales: [100, 80, 90, 110, 95]

Step 2: Calculate the means of both variables.

Mean of Price = (10 + 15 + 12 + 8 + 14) / 5 = 11.8 (rounded to one decimal place)

Mean of Sales = (100 + 80 + 90 + 110 + 95) / 5 = 95 (rounded to one decimal place)

Step 3: Calculate the deviations from the means for each variable.

Deviations from the mean of Price: [-1.8, 3.2, 0.2, -3.8, 2.2]

Deviations from the mean of Sales: [5, -15, -5, 15, 0]

Step 4: Calculate the product of the deviations for each pair of data points.

Product of deviations: [-1.8 \* 5, 3.2 \* -15, 0.2 \* -5, -3.8 \* 15, 2.2 \* 0]

= [-9, -48, -1, -57, 0]

Step 5: Calculate the squared deviations for each variable.

Squared deviations of Price: [(-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 Sales: [5^2, (-15)^2, (-5)^2, 15^2, 0^2]

= [25, 225, 25, 225, 0]

Step 6: Calculate the sums of the product of deviations and squared deviations.

Sum of product of deviations = -9 + -48 + -1 + -57 + 0

= -115

Sum of squared deviations of Price = 3.24 + 10.24 + 0.04 + 14.44 + 4.84

= 32.8

Sum of squared deviations of Sales = 25 + 225 + 25 + 225 + 0

= 500

Step 7: Calculate the slope (β1) of the regression line using the formula:

β1 = (Sum of product of deviations) / (Sum of squared deviations of Price)

β1 = -115 / 32.8

≈ -3.51 (rounded to two decimal places)

Step 8: Calculate the y-intercept (β0) of the regression line using the formula:

β0 = Mean of Sales - (β1 \* Mean of Price)

β0 = 95 - (-3.51 \* 11.8)

≈ 133.09 (rounded to two decimal places)

Step 9: Write the equation of the regression line.

The equation of the regression line is: Sales = -3.51 \* Price + 133.09

Step 10: Interpret the results.

The slope (β1) of the regression line indicates that, on average, for each increase of 1 dollar in price, sales are expected to decrease by 3.51 units.

The y-intercept (β0) represents the predicted sales when the price is 0 dollars. In this case, it is 133.09. However, it is important to note that this interpretation may not be meaningful in the context of the problem.

Therefore, based on the simple linear regression analysis, the equation Sales = -3.51 \* Price + 133.09 can be used 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.**

Ans: To calculate the standard error of the mean satisfaction score, you can follow these steps:

Step 1: Organize the data into a sample.

Sample: [7, 8, 9, 6, 8, 7, 9, 7, 8, 7]

Step 2: Calculate the sample mean (x̄) and the sample standard deviation (s).

Sample mean (x̄) = (7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7) / 10 = 7.7 (rounded to one decimal place)

Sample standard deviation (s) = sqrt(((7 - 7.7)^2 + (8 - 7.7)^2 + (9 - 7.7)^2 + (6 - 7.7)^2 + (8 - 7.7)^2 + (7 - 7.7)^2 + (9 - 7.7)^2 + (7 - 7.7)^2 + (8 - 7.7)^2 + (7 - 7.7)^2) / (10 - 1))

= sqrt((0.49 + 0.09 + 1.69 + 2.89 + 0.09 + 0.09 + 1.69 + 0.09 + 0.09 + 0.09) / 9)

= sqrt(7.31 / 9)

≈ sqrt(0.812)

≈ 0.90 (rounded to two decimal places)

Step 3: Determine the sample size (n).

Sample size (n) = 10

Step 4: Calculate the standard error (SE) using the formula:

SE = s / sqrt(n)

SE = 0.90 / sqrt(10)

≈ 0.28 (rounded to two decimal places)

Therefore, the standard error of the mean satisfaction score is approximately 0.28.

**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: To perform a multiple regression analysis and predict sales based on advertising expenditure, you can follow these steps:

Step 1: Organize the data into two variables: advertising expenditure and sales.

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

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

Step 2: Calculate the means of both variables.

Mean of Advertising Expenditure = (10 + 15 + 12 + 8 + 14) / 5 = 11.8 (rounded to one decimal place)

Mean of Sales = (25 + 30 + 28 + 20 + 26) / 5 = 25.8 (rounded to one decimal place)

Step 3: Calculate the deviations from the means for each variable.

Deviations from the mean of Advertising Expenditure: [-1.8, 3.2, 0.2, -3.8, 2.2]

Deviations from the mean of Sales: [-0.8, 4.2, 2.2, -5.8, 0.2]

Step 4: Calculate the product of the deviations for each pair of data points.

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 5: Calculate the squared deviations for each variable.

Squared deviations of Advertising Expenditure: [(-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 Sales: [(-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 6: Calculate the sums of the product of deviations and squared deviations.

Sum of product of deviations = 1.44 + 13.44 + 0.44 + 22.04 + 0.44

= 37.8

Sum of squared deviations of Advertising Expenditure = 3.24 + 10.24 + 0.04 + 14.44 + 4.84

= 32.8

Sum of squared deviations of Sales = 0.64 + 17.64 + 4.84 + 33.64 + 0.04

= 57.8

Step 7: Calculate the slope (β1) of the regression line using the formula:

β1 = (Sum of product of deviations) / (Sum of squared deviations of Advertising Expenditure)

β1 = 37.8 / 32.8

≈ 1.15 (rounded to two decimal places)

Step 8: Calculate the y-intercept (β0) of the regression line using the formula:

β0 = Mean of Sales - (β1 \* Mean of Advertising Expenditure)

β0 = 25.8 - (1.15 \* 11.8)

≈ 12.38 (rounded to two decimal places)

Step 9: Write the equation of the regression line.

The equation of the regression line is: Sales = 1.15 \* Advertising Expenditure + 12.38

Step 10: Interpret the results.

The slope (β1) of the regression line indicates that, on average, for each increase of 1 thousand dollars in advertising expenditure, sales are expected to increase by 1.15 thousand dollars.

The y-intercept (β0) represents the predicted sales when the advertising expenditure is 0 thousand dollars. In this case, it is 12.38. However, it is important to note that this interpretation may not be meaningful in the context of the problem.

Therefore, based on the multiple regression analysis, the equation Sales = 1.15 \* Advertising Expenditure + 12.38 can be used to predict sales based on advertising expenditure.