Core Module-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.

   To calculate the mean sales for each region, you need to sum up the sales data for each region and divide it by the total number of observations in that region. Here are the calculations for the given data:

   Region A: [10, 15, 12, 8, 14]
   Mean sales in Region A = (10 + 15 + 12 + 8 + 14) / 5 = 59 / 5 = 11.8

   Region B: [18, 20, 16, 22, 25]
   Mean sales in Region B = (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. 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, you need to determine the value or values that appear most frequently in the data. In the given data set [4, 5, 2, 3, 5, 4, 3, 2, 4, 5], we can count the occurrences of each value and identify the one(s) with the highest frequency. Here's the calculation:

   Value 2: Appears 2 times
   Value 3: Appears 2 times
   Value 4: Appears 3 times
   Value 5: Appears 3 times

   Both 4 and 5 have the highest frequency of 3, indicating a tie for the mode. Therefore, in this survey data, the mode(s) 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.

   To calculate the median salary for each department, you need to arrange the salary data in each department in ascending order and find the middle value(s). Here are the calculations for the given data:

   Department A: [5000, 6000, 5500, 7000]
   Median salary in Department A: (5500 + 6000) / 2 = 11500 / 2 = 5750

   Department B: [4500, 5500, 5800, 6000, 5200]
   Median salary in Department B: 5500

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

4. Scenario: A data analyst wants to determine the variability in the daily stock prices of a company. The data collected is as follows:
   [25.5, 24.8, 26.1, 25.3, 24.9]
   Calculate the range of the stock prices.

   To calculate the range of the stock prices, you need to find the difference between the highest and lowest values in the given data. Here's the calculation for the provided stock price data [25.5, 24.8, 26.1, 25.3, 24.9]:

   Highest value: 26.1
   Lowest value: 24.8

   Range = Highest value - Lowest value = 26.1 - 24.8 = 1.3

   Therefore, the range of the stock prices is 1.3.

5. Scenario: A study is conducted to compare the performance of two different teaching methods. The test scores of the students in each group are as follows:
   Group A: [85, 90, 92, 88, 91]
   Group B: [82, 88, 90, 86, 87]
   Perform a t-test to determine if there is a significant difference in the mean scores between the two groups.

To perform a t-test and determine if there is a significant difference in the mean scores between the two groups (Group A and Group B), you can use the independent two-sample t-test. This test compares the means of two independent samples to evaluate if the difference between them is statistically significant. Here's how to perform the t-test using the provided test scores:

Step 1: State the Hypotheses:
- Null Hypothesis (H0): There is no significant difference in the mean scores between Group A and Group B.
- Alternative Hypothesis (Ha): There is a significant difference in the mean scores between Group A and Group B.

Step 2: Calculate the Test Statistic:
- Calculate the mean ($\bar{x}$) and standard deviation (s) for each group:
  - Group A: $\bar{x}A = (85 + 90 + 92 + 88 + 91) / 5 = 89.2$, sA = sqrt(((85-89.2)^2 + (90-89.2)^2 + (92-89.2)^2 + (88-89.2)^2 + (91-89.2)^2) / 4) = 2.48
  - Group B: $\bar{x}B = (82 + 88 + 90 + 86 + 87) / 5 = 86.6$, sB = sqrt(((82-86.6)^2 + (88-86.6)^2 + (90-86.6)^2 + (86-86.6)^2 + (87-86.6)^2) / 4) = 2.38

- Calculate the pooled standard deviation (sp) using the formula: sp = sqrt(((nA-1)sA^2 + (nB-1)sB^2) / (nA + nB - 2)) = sqrt(((4)(2.48)^2 + (4)(2.38)^2) / (4 + 4 - 2)) = 2.43

- Calculate the t-statistic using the formula: t = ($\bar{x}$A - $\bar{x}$B) / (sp * sqrt(1/nA + 1/nB)) = (89.2 - 86.6) / (2.43 * sqrt(1/5 + 1/5)) = 2.24

Step 3: Determine the Critical Value and p-value:
- With a specified significance level (e.g., $\alpha = 0.05$), determine the degrees of freedom (df) using df = nA + nB - 2 = 5 + 5 - 2 = 8.
- Look up the critical t-value from the t-distribution table or use statistical software for the given α and df.
- Calculate the p-value associated with the t-statistic using the t-distribution with the calculated df.

Step 4: Make a Decision:
- If the absolute value of the t-statistic is greater than the critical t-value or if the p-value is less than the specified significance level (α), reject the null hypothesis.
- If the absolute value of the t-statistic is less than the critical t-value or if the p-value is greater than the specified significance level (α), fail to reject the null hypothesis.

Based on the calculated t-statistic, the critical value, and the p-value, you can make a decision about whether 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.

   To calculate the correlation coefficient between advertising expenditure and sales, you can use the Pearson correlation coefficient formula. The Pearson correlation coefficient measures the strength and direction of the linear relationship between two variables. Here's how to calculate it using the provided data:

   Step 1: Calculate the mean of advertising expenditure ($\bar{x}$) and sales ($\bar{y}$):
   - Advertising Expenditure: $\bar{x}$ = (10 + 15 + 12 + 8 + 14) / 5 = 11.8
   - Sales: $\bar{y}$ = (25 + 30 + 28 + 20 + 26) / 5 = 25.8

   Step 2: Calculate the deviation of each value from the mean for both variables:
   - Advertising Expenditure (x):
     - (10 - 11.8), (15 - 11.8), (12 - 11.8), (8 - 11.8), (14 - 11.8)
     - (-1.8, 3.2, 0.2, -3.8, 2.2)
   - Sales (y):
     - (25 - 25.8), (30 - 25.8), (28 - 25.8), (20 - 25.8), (26 - 25.8)
     - (-0.8, 4.2, 2.2, -5.8, 0.2)

   Step 3: Calculate the product of the deviations for each pair of values:
   - (-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 deviation of each variable:
   - Advertising Expenditure (x):
     - (-1.8)^2, (3.2)^2, (0.2)^2, (-3.8)^2, (2.2)^2
     - (3.24, 10.24, 0.04, 14.44, 4.84)
   - Sales (y):
     - (-0.8)^2, (4.2)^2, (2.2)^2, (-5.8)^2, (0.2)^2
     - (0.64, 17.64, 4.84, 33.64, 0.04)

   Step 5: Calculate the sum of the product of the deviations:
   - Sum of (product of deviations) = 1.44 + 13.44 + 0.44 + 22.04 + 0.44 = 37.8

Step 6: Calculate the square root of the sum of squared deviations for both variables:
- Square root of (sum of squared deviations for x) = sqrt(3.24 + 10.24 + 0.04 + 14.44 + 4.84) = sqrt(32.8) = 5.73
- Square root of (sum of squared deviations for y) = sqrt(0.64 + 17.64 + 4.84 + 33.64 + 0.04) = sqrt(57.8) = 7.61

Step 7: Calculate the correlation coefficient (r):
- r = (Sum of (product of deviations)) / (Square root of (sum of squared deviations for x) * Square root of (sum of squared deviations for y))
- r = 37.8 / (5.73 * 7.61) = 0.89 (rounded to two decimal places)

Therefore, the correlation coefficient between advertising expenditure and sales is approximately 0.89, indicating a strong positive correlation between the two variables.

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 use the following steps:

Step 1: Calculate the mean (x̄) 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 deviation of each height from the mean:
Deviations: (160 - 167.86), (170 - 167.86), (165 - 167.86), (155 - 167.86), (175 - 167.86), (180 - 167.86), (170 - 167.86)
Deviations: (-7.86, 2.14, -2.86, -12.86, 7.14, 12.14, 2.14)

Step 3: Square each deviation:
Squared deviations: (-7.86)^2, (2.14)^2, (-2.86)^2, (-12.86)^2, (7.14)^2, (12.14)^2, (2.14)^2
Squared deviations: (61.70, 4.58, 8.20, 165.08, 51.05, 147.61, 4.58)

Step 4: Calculate the mean of the squared deviations:
Mean of squared deviations = (61.70 + 4.58 + 8.20 + 165.08 + 51.05 + 147.61 + 4.58) / 7 = 442.80 / 7 = 63.26 (rounded to two decimal places)

Step 5: Calculate the square root of the mean of squared deviations:
  Standard Deviation = sqrt(63.26) = 7.95 (rounded to two decimal places)

Therefore, the standard deviation of the heights is approximately 7.95. The standard deviation provides a measure of the average amount by which the heights deviate from the mean height.

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, you can use the least squares method to fit a linear regression model to the given data. Here's how you can perform the analysis using the provided data:

   Step 1: Define the Variables:
   - Employee Tenure (in years): [2, 3, 5, 4, 6, 2, 4] (denoted as X)
   - Job Satisfaction (on a scale of 1 to 10): [7, 8, 6, 9, 5, 7, 6] (denoted as Y)

   Step 2: Calculate the Means of X and Y:
   - Mean of X (x̄) = (2 + 3 + 5 + 4 + 6 + 2 + 4) / 7 = 26 / 7 = 3.71
   - Mean of Y (ȳ) = (7 + 8 + 6 + 9 + 5 + 7 + 6) / 7 = 48 / 7 = 6.86

   Step 3: Calculate the Deviations of X and Y from their Means:
   - Deviations of X from x̄: [-1.71, -0.71, 1.29, 0.29, 2.29, -1.71, 0.29]
   - Deviations of Y from ȳ: [0.14, 1.14, -0.86, 2.14, -1.86, 0.14, -0.86]

   Step 4: Calculate the Sum of the Products of the Deviations:
   - Sum of (Deviations of X from x̄ * Deviations of Y from ȳ):
     = (-1.71 * 0.14) + (-0.71 * 1.14) + (1.29 * -0.86) + (0.29 * 2.14) + (2.29 * -1.86) + (-1.71 * 0.14) + (0.29 * -0.86)
     = -0.24 - 0.81 - 1.11 + 0.62 - 4.26 - 0.24 - 0.25
     = -5.29

   Step 5: Calculate the Sum of Squares of X Deviations:
   - Sum of Squares of (Deviations of X from x̄):
     = (-1.71)^2 + (-0.71)^2 + (1.29)^2 + (0.29)^2 + (2.29)^2 + (-1.71)^2 + (0.29)^2
     = 2.92 + 0.50 + 1.67 + 0.08 + 5.26 + 2.92 + 0.08

= 13.43

Step 6: Calculate the Regression Coefficients:
- Slope (β1) = (Sum of (Deviations of X from x̄ * Deviations of Y from ȳ)) / (Sum of Squares of (Deviations of X from x̄))
  = -5.29 / 13.43
  = -0.394 (rounded to three decimal places)
- Intercept (β0) = ȳ - (β1 * x̄)
  = 6.86 - (-0.394 * 3.71)
  = 6.86 + 1.46
  = 8.32 (rounded to two decimal places)

Step 7: Formulate the Linear Regression Equation:
- The linear regression equation is given by: Y = β0 + β1*X
  - Y = 8.32 - 0.394*X

Therefore, based on the linear regression analysis, the predicted job satisfaction (Y) can be estimated using the equation Y = 8.32 - 0.394*X, where X represents the employee tenure in years. The slope coefficient (β1) of -0.394 indicates a negative relationship between employee tenure and job satisfaction. The intercept (β0) of 8.32 represents the estimated job satisfaction when the employee tenure is zero (which may not have practical meaning in this context).

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) and determine if there is a significant difference in the mean recovery times between Medication A and Medication B, you can use a one-way ANOVA. The one-way ANOVA compares the means of three or more groups to evaluate if there are significant differences among them. Here's how you can perform the ANOVA using the provided recovery time data:

   Step 1: State the Hypotheses:
   - Null Hypothesis (H0): There is no significant difference in the mean recovery times between Medication A and Medication B.

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

Step 2: Calculate the Sum of Squares:

- Calculate the grand mean ($\bar{Y}$) by taking the mean of all the recovery times:
  - $\bar{Y} = (10 + 12 + 14 + 11 + 13 + 15 + 17 + 16 + 14 + 18) / 10 = 140 / 10 = 14$
- Calculate the sum of squares between groups (SSB):
  - $SSB = nA * (\bar{x}A - \bar{Y})^2 + nB * (\bar{x}B - \bar{Y})^2$, where nA and nB are the number of observations in each group, and $\bar{x}A$ and $\bar{x}B$ are the means of Medication A and Medication B, respectively.
  - $SSB = 5 * (12 - 14)^2 + 5 * (16 - 14)^2 = 2 * 2 + 2 * 2 = 8$
- Calculate the sum of squares within groups (SSW):
  - $SSW = \Sigma(xi - \bar{x}i)^2$, where xi is an observation and $\bar{x}i$ is the mean of each group.
  - $SSW = (10 - 12)^2 + (12 - 12)^2 + (14 - 12)^2 + (11 - 12)^2 + (13 - 12)^2 + (15 - 16)^2 + (17 - 16)^2 + (16 - 16)^2 + (14 - 16)^2 + (18 - 16)^2$
  - $SSW = 4 + 0 + 4 + 1 + 1 + 1 + 1 + 0 + 4 + 4 = 20$
- Calculate the total sum of squares (SST):
  - $SST = SSB + SSW = 8 + 20 = 28$

Step 3: Calculate the Degrees of Freedom:

- Degrees of Freedom between groups (dfB) = number of groups - 1 = 2 - 1 = 1
- Degrees of Freedom within groups (dfW) = total number of observations - number of groups = 10 - 2 = 8
- Total Degrees of Freedom (dfT) = dfB + dfW = 1 + 8 = 9

Step 4: Calculate the Mean Squares:

- Mean Square between groups (MSB) = SSB / dfB = 8 / 1 = 8
- Mean Square within groups (MSW) = SSW / dfW = 20 / 8 = 2.5

Step 5: Calculate the F-statistic:

- F-statistic = MSB / MSW = 8 / 2.5 = 3.2

Step 6: Determine the Critical Value and p-value:

- With a specified significance level (e.g., $\alpha = 0.05$), determine the critical F-value from the F-distribution table or use statistical software for the given degrees of freedom.
- Calculate the p-value associated with the F-statistic using the F-distribution with the calculated degrees of freedom.

Step 7: Make a Decision:

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

Based on the calculated F-statistic, the critical value, and the p-value, you can make a decision about whether there is a significant difference in the mean recovery times between Medication A and Medication B.

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 need to arrange the ratings in ascending order and find the value below which 75% of the data falls. Here's how you can calculate the 75th percentile using the provided data:

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

Step 2: Calculate the index position corresponding to the 75th percentile:
Index = (75/100) * (n+1) = (0.75) * (10+1) = 8.25

Step 3: Determine the value at the calculated index position:
Since the index is not a whole number, we need to interpolate the value.
The value at the 8th index position is 9, and the value at the 9th index position is also 9.
Interpolation: (8.25 - 8) * (9 - 9) = 0,
Therefore, the 75th percentile value is 9.

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

11. Scenario: A quality control department wants to test the weight consistency of a product. The weights of a sample of products are as follows:
[10.2, 9.8, 10.0, 10.5, 10.3, 10.1]
Perform a hypothesis test to determine if the mean weight differs significantly from 10 grams.

To perform a hypothesis test and determine if the mean weight of the product differs significantly from 10 grams, you can use a one-sample t-test. The one-sample t-test compares the mean of a sample to a known population mean to evaluate if there is a significant difference. Here's how you can perform the hypothesis test using the provided weight data:

Step 1: State the Hypotheses:
- Null Hypothesis (H0): The mean weight of the product is equal to 10 grams.
- Alternative Hypothesis (Ha): The mean weight of the product differs significantly from 10 grams.

Step 2: Set the Significance Level:
- Choose a significance level (e.g., $\alpha = 0.05$) to determine the level of significance at which you will accept or reject the null hypothesis.

Step 3: Calculate the Test Statistic:
- Calculate the sample mean ($\bar{x}$) and the sample standard deviation (s) of the weight data:
  - Sample mean ($\bar{x}$) = (10.2 + 9.8 + 10.0 + 10.5 + 10.3 + 10.1) / 6 = 60.9 / 6 = 10.15
  - Sample standard deviation (s) = sqrt(((10.2-10.15)^2 + (9.8-10.15)^2 + (10.0-10.15)^2 + (10.5-10.15)^2 + (10.3-10.15)^2 + (10.1-10.15)^2) / 5) = sqrt(0.008) = 0.089

- Calculate the t-statistic using the formula: t = ($\bar{x}$ - μ) / (s / sqrt(n)), where μ is the population mean, s is the sample standard deviation, and n is the sample size.
  - Population mean (μ) = 10 (given)
  - Sample size (n) = 6
  - t-statistic = (10.15 - 10) / (0.089 / sqrt(6)) = 0.15 / (0.089 / sqrt(6)) ≈ 0.15 / 0.0365 ≈ 4.11

Step 4: Determine the Critical Value and p-value:
- With a specified significance level ($\alpha = 0.05$) and the degrees of freedom (df = n - 1 = 6 - 1 = 5), determine the critical t-value from the t-distribution table or use statistical software.
- Calculate the p-value associated with the t-statistic using the t-distribution with the calculated degrees of freedom.

Step 5: Make a Decision:

- If the absolute value of the t-statistic is greater than the critical t-value or if the p-value is less than the specified significance level (α), reject the null hypothesis.
- If the absolute value of the t-statistic is less than the critical t-value or if the p-value is greater than the specified significance level (α), fail to reject the null hypothesis.

Based on the calculated t-statistic, the critical value, and the p-value, you can make a decision about whether 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.

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 use a chi-square test of independence. The chi-square test of independence evaluates if there is an association between two categorical variables. In this case, the variables are the website designs and the click-through rates. Here's how you can perform the chi-square test using the provided click data:

Step 1: State the Hypotheses:

Null Hypothesis (H0): There is no significant difference in the click-through rates between Design A and Design B.
Alternative Hypothesis (Ha): There is a significant difference in the click-through rates between Design A and Design B.

Step 2: Create a Contingency Table:

Create a 2x2 contingency table representing the observed frequencies of the click-through rates for each design:

	Design A	Design B
Clicked	100	80
Not Clicked	120	85

Step 3: Calculate the Expected Frequencies:

Calculate the expected frequencies under the assumption of independence between the variables.

The expected frequencies can be calculated using the formula: E = (row total * column total) / grand total.

	Design A	Design B
Clicked	(100+120) * (100+80) / (100+120+80+85)	
Not Clicked	(120+85) * (100+80) / (100+120+80+85)	

Step 4: Calculate the Chi-Square Test Statistic:

Calculate the chi-square test statistic using the formula: $\chi^2 = \Sigma((O - E)^2 / E)$, where O is the observed frequency and E is the expected frequency.

Step 5: Determine the Degrees of Freedom:

The degrees of freedom (df) for a 2x2 contingency table is calculated as: df = (number of rows - 1) * (number of columns - 1).

Step 6: Determine the Critical Value and p-value:

With a specified significance level (e.g., $\alpha = 0.05$) and the calculated degrees of freedom, determine the critical chi-square value from the chi-square distribution table or use statistical software.

Calculate the p-value associated with the chi-square test statistic using the chi-square distribution with the calculated degrees of freedom.

Step 7: Make a Decision:

If the chi-square test statistic is greater than the critical chi-square value or if the p-value is less than the specified significance level (α), reject the null hypothesis.

If the chi-square test statistic is less than the critical chi-square value or if the p-value is greater than the specified significance level (α), fail to reject the null hypothesis.

Based on the calculated chi-square test statistic, the critical value, and the p-value, you can make a decision about whether 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.

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

Confidence Interval = $\bar{x} \pm (t * (s / \sqrt{n}))$

Where:
- $\bar{x}$ is the sample mean,
- t is the critical value for the desired confidence level and degrees of freedom (df),
- s is the sample standard deviation,
- n is the sample size.

Step 1: Calculate the sample mean ($\bar{x}$) and sample standard deviation (s):
- Sample mean ($\bar{x}$) = (7 + 9 + 6 + 8 + 10 + 7 + 8 + 9 + 7 + 8) / 10 = 79 / 10 = 7.9
- Calculate the deviations from the mean and then the squared deviations:
  Deviations from the mean: (7 - 7.9), (9 - 7.9), (6 - 7.9), (8 - 7.9), (10 - 7.9), (7 - 7.9), (8 - 7.9), (9 - 7.9), (7 - 7.9), (8 - 7.9)
  Squared deviations: (0.81, 1.21, 3.61, 0.01, 4.41, 0.81, 0.01, 1.21, 0.81, 0.01)
- Calculate the sample variance (s^2) by summing the squared deviations and dividing by (n-1):
  Sample variance (s^2) = (0.81 + 1.21 + 3.61 + 0.01 + 4.41 + 0.81 + 0.01 + 1.21 + 0.81 + 0.01) / (10-1) = 13.08 / 9 = 1.453
- Calculate the sample standard deviation (s) by taking the square root of the sample variance:
  Sample standard deviation (s) = sqrt(1.453) ≈ 1.205 (rounded to three decimal places)

Step 2: Determine the critical value (t) for a 95% confidence level and the degrees of freedom (df = n - 1 = 10 - 1 = 9):
- Using a t-table or statistical software, the critical value for a 95% confidence level with 9 degrees of freedom is approximately 2.262.

Step 3: Calculate the standard error (SE) of the mean:
- Standard Error (SE) = s / √n = 1.205 / √10 ≈ 0.381 (rounded to three decimal places)

Step 4: Calculate the lower and upper bounds of the confidence interval:

- Lower bound = x̄ - (t * SE) = 7.9 - (2.262 * 0.381) ≈ 7.9 - 0.863 ≈ 7.037 (rounded to three decimal places)
- Upper bound = x̄ + (t * SE) = 7.9 + (2.262 * 0.381) ≈ 7.9 + 0.863 ≈ 7.963 (rounded to three decimal places)

Step 5: Write the 95% confidence interval:
- The 95% confidence interval for the population mean satisfaction score is approximately 7.037 to 7.963.

Therefore, we can say with 95% confidence that the true population mean satisfaction score falls within the range of 7.037 to 7.963 based on the given sample.

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 and predict the performance based on temperature, you can use the following steps:

Step 1: Organize the data:
- 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:
- Calculate the mean of the temperature data (x̄):
  x̄ = (20 + 22 + 23 + 19 + 21) / 5 = 105 / 5 = 21
- Calculate the mean of the performance data (ȳ):
  ȳ = (8 + 7 + 9 + 6 + 8) / 5 = 38 / 5 = 7.6

Step 3: Calculate the deviations:
- Calculate the deviations of the temperature data from the mean (x - x̄):
  (20 - 21), (22 - 21), (23 - 21), (19 - 21), (21 - 21) = -1, 1, 2, -2, 0
- Calculate the deviations of the performance data from the mean (y - ȳ):
  (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 product of deviations:
- Calculate the product of the deviations of temperature and performance (x - x̄)(y - ȳ):

(-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:
- Calculate the squared deviations of the temperature data (x - x̄)^2:
  (-1)^2, 1^2, 2^2, (-2)^2, 0^2 = 1, 1, 4, 4, 0
- Calculate the squared deviations of the performance data (y - ȳ)^2:
  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:
- Calculate the sum of the product of deviations (Σ[(x - x̄)(y - ȳ)]):
  -0.4 + -0.6 + 2.8 + 3.2 + 0 = 5
- Calculate the sum of the squared deviations of the temperature data (Σ[(x - x̄)^2]):
  1 + 1 + 4 + 4 + 0 = 10
- Calculate the sum of the squared deviations of the performance data (Σ[(y - ȳ)^2]):
  0.16 + 0.36 + 1.96 + 2.56 + 0.16 = 5.2

Step 7: Calculate the slope (β1):
- Calculate the slope of the regression line (β1) using the formula:
  β1 = Σ[(x - x̄)(y - ȳ)] / Σ[(x - x̄)^2]
  β1 = 5 / 10 = 0.5

Step 8: Calculate the intercept (β0):
- Calculate the intercept of the regression line (β0) using the formula:
  β0 = ȳ - β1 * x̄
  β0 = 7.6 - 0.5 * 21 = 7.6 - 10.5 = -2.9

Step 9: Write the regression equation:
- The regression equation is y = β0 + β1 * x, where y represents performance and x represents temperature.
- In this case, the regression equation is y = -2.9 + 0.5 * x.

Step 10: Interpret the results:
- The slope (β1) of 0.5 indicates that for every one-unit increase in temperature, the predicted performance increases by 0.5 units.
- The intercept (β0) of -2.9 represents the predicted performance when the temperature is zero, which may not be meaningful in this context.
- The regression equation can be used to predict the performance based on a given temperature value.

Therefore, the simple linear regression equation to predict performance based on temperature is y = -2.9 + 0.5 * x.

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, you can use the following steps:

Step 1: State the Hypotheses:
- Null Hypothesis (H0): There is no significant difference in the median preferences between Group A and Group B.
- Alternative Hypothesis (Ha): There is a significant difference in the median preferences between Group A and Group B.

Step 2: Rank the Data:
- Combine the preferences from both groups and rank them in ascending order, assigning the same rank to tied values.

Step 3: Calculate the U Statistic:
- Calculate the U statistic for each group using the formula:
  - Ua = n1 * n2 + (n1 * (n1 + 1)) / 2 - Ra
  - Ub = n1 * n2 + (n2 * (n2 + 1)) / 2 - Rb

  where n1 is the sample size of Group A, n2 is the sample size of Group B, Ra is the sum of ranks for Group A, and Rb is the sum of ranks for Group B.

Step 4: Calculate the Expected U Statistic:
- Calculate the expected U statistic (E[U]) using the formula:
  - E[U] = n1 * n2 / 2

Step 5: Calculate the Variance of U (Var[U]):
- Calculate the variance of U (Var[U]) using the formula:
  - Var[U] = (n1 * n2 * (n1 + n2 + 1)) / 12

Step 6: Calculate the Z Score:
- Calculate the Z score using the formula:
  - Z = (U - E[U]) / sqrt(Var[U])

Step 7: Determine the Critical Z Value and p-value:
 - With a specified significance level (e.g., $\alpha = 0.05$), determine the critical Z value from the standard normal distribution table or use statistical software.
 - Calculate the p-value associated with the Z score.

Step 8: Make a Decision:
 - If the absolute value of the Z score is greater than the critical Z value or if the p-value is less than the specified significance level (α), reject the null hypothesis.
 - If the absolute value of the Z score is less than the critical Z value or if the p-value is greater than the specified significance level (α), fail to reject the null hypothesis.

Based on the calculated U statistic, expected U statistic, Z score, critical Z value, and p-value, you can make a decision about whether 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.

To calculate the interquartile range (IQR) of the ages, you can use the following steps:

Step 1: Organize the data:
- Customer ages: [25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

Step 2: Sort the data:
- Sort the ages in ascending order: [25, 30, 35, 40, 45, 50, 55, 60, 65, 70]

Step 3: Calculate the first quartile (Q1):
- Q1 is the median of the lower half of the data. To calculate Q1, find the median of the lower values in the sorted data.
- In this case, the lower half of the data is [25, 30, 35, 40]. The median of this lower half is $(30 + 35) / 2 = 32.5$.

Step 4: Calculate the third quartile (Q3):
- Q3 is the median of the upper half of the data. To calculate Q3, find the median of the higher values in the sorted data.

- In this case, the upper half of the data is [50, 55, 60, 65, 70]. The median of this upper half is (55 + 60) / 2 = 57.5.

Step 5: Calculate the interquartile range (IQR):
- IQR is the difference between the third quartile (Q3) and the first quartile (Q1).
- In this case, IQR = Q3 - Q1 = 57.5 - 32.5 = 25.

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

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 and determine if there is a significant difference in the median accuracy scores between the three algorithms (Algorithm A, Algorithm B, and Algorithm C), you can use the following steps:

Step 1: State the Hypotheses:
- Null Hypothesis (H0): There is no significant difference in the median accuracy scores between the algorithms.
- Alternative Hypothesis (Ha): There is a significant difference in the median accuracy scores between the algorithms.

Step 2: Combine the Data:
- Combine the accuracy scores from all algorithms into a single dataset.

Step 3: Rank the Data:
- Rank the combined accuracy scores in ascending order, assigning the same rank to tied values.

Step 4: Calculate the Rank Sums:
- Calculate the sum of ranks for each algorithm.
- Calculate the sum of ranks for each group (Algorithm A, Algorithm B, and Algorithm C).

Step 5: Calculate the Kruskal-Wallis Test Statistic (H):
- Calculate the Kruskal-Wallis test statistic (H) using the formula:
  - H = ((12 / (n * (n + 1))) * Σ((R^2 / ni) - (n + 1) / 2)) / (1 - (Σ(Ti^2) / (n * (n^2 - 1))))

  where n is the total number of observations, R is the sum of ranks for each group, ni is the number of observations in each group, and Ti is the sum of tied ranks in each group.

Step 6: Determine the Critical Value and p-value:
- With a specified significance level (e.g., α = 0.05), determine the critical value from the chi-square distribution table or use statistical software.
- Calculate the p-value associated with the Kruskal-Wallis test statistic.

Step 7: Make a Decision:
- If the Kruskal-Wallis test statistic (H) is greater than the critical value or if the p-value is less than the specified significance level (α), reject the null hypothesis.
- If the Kruskal-Wallis test statistic (H) is less than the critical value or if the p-value is greater than the specified significance level (α), fail to reject the null hypothesis.

Based on the calculated Kruskal-Wallis test statistic, the critical value, and the p-value, you can make a decision about whether there is a significant difference in the median accuracy scores between the three algorithms (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 sales based on price.

    To perform a simple linear regression and predict sales based on price, you can use the following steps:

    Step 1: Organize the data:
    - Price (in dollars): [10, 15, 12, 8, 14]
    - Sales: [100, 80, 90, 110, 95]

    Step 2: Calculate the means:
    - Calculate the mean of the price data (x̄):
      x̄ = (10 + 15 + 12 + 8 + 14) / 5 = 59 / 5 = 11.8

- Calculate the mean of the sales data (ȳ):
  ȳ = (100 + 80 + 90 + 110 + 95) / 5 = 475 / 5 = 95

Step 3: Calculate the deviations:
- Calculate the deviations of the price data from the mean (x - x̄):
  (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
- Calculate the deviations of the sales data from the mean (y - ȳ):
  (100 - 95), (80 - 95), (90 - 95), (110 - 95), (95 - 95) = 5, -15, -5, 15, 0

Step 4: Calculate the product of deviations:
- Calculate the product of the deviations of price and sales (x - x̄)(y - ȳ):
  (-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:
- Calculate the squared deviations of the price data (x - x̄)^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
- Calculate the squared deviations of the sales data (y - ȳ)^2:
  5^2, (-15)^2, (-5)^2, 15^2, 0^2 = 25, 225, 25, 225, 0

Step 6: Calculate the sums:
- Calculate the sum of the product of deviations (Σ[(x - x̄)(y - ȳ)]):
  -9 + -48 + -1 + -57 + 0 = -115
- Calculate the sum of the squared deviations of the price data (Σ[(x - x̄)^2]):
  3.24 + 10.24 + 0.04 + 14.44 + 4.84 = 33.8
- Calculate the sum of the squared deviations of the sales data (Σ[(y - ȳ)^2]):
  25 + 225 + 25 + 225 + 0 = 500

Step 7: Calculate the slope (β1):
- Calculate the slope of the regression line (β1) using the formula:
  β1 = Σ[(x - x̄)(y - ȳ)] / Σ[(x - x̄)^2]
  β1 = -115 / 33.8 ≈ -3.404 (rounded to three decimal places)

Step 8: Calculate the intercept (β0):
- Calculate the intercept of the regression line (β0) using the formula:
  β0 = ȳ - β1 * x̄
  β0 = 95 - (-3.404 * 11.8) ≈ 132.899 (rounded to three decimal places)

Step 9: Write the regression equation:
- The regression equation is y = β0 + β1 * x, where y represents sales and x represents price.
- In this case, the regression equation is y = 132.899 - 3.404 * x.

Step 10: Interpret the results:

- The slope ($\beta 1$) of -3.404 indicates that for every one-unit increase in price, the predicted sales decrease by 3.404 units.
- The intercept ($\beta 0$) of 132.899 represents the predicted sales when the price is zero, which may not be meaningful in this context.
- The regression equation can be used to predict the sales based on a given price value.

Therefore, the simple linear regression equation to predict sales based on price is y = 132.899 - 3.404 * x.

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 satisfaction score, you can use the following steps:

Step 1: Organize the data:
- Customer satisfaction scores: [7, 8, 9, 6, 8, 7, 9, 7, 8, 7]

Step 2: Calculate the mean:
- Calculate the mean of the satisfaction scores ($\bar{x}$):
  $\bar{x}$ = (7 + 8 + 9 + 6 + 8 + 7 + 9 + 7 + 8 + 7) / 10 = 76 / 10 = 7.6

Step 3: Calculate the differences from the mean:
- Calculate the differences of each satisfaction score from the mean (x - $\bar{x}$):
  (7 - 7.6), (8 - 7.6), (9 - 7.6), (6 - 7.6), (8 - 7.6), (7 - 7.6), (9 - 7.6), (7 - 7.6), (8 - 7.6), (7 - 7.6) = -0.6, 0.4, 1.4, -1.6, 0.4, -0.6, 1.4, -0.6, 0.4, -0.6

Step 4: Calculate the squared differences:
- Calculate the squared differences of each score from the mean ((x - $\bar{x}$)^2):
  (-0.6)^2, (0.4)^2, (1.4)^2, (-1.6)^2, (0.4)^2, (-0.6)^2, (1.4)^2, (-0.6)^2, (0.4)^2, (-0.6)^2 = 0.36, 0.16, 1.96, 2.56, 0.16, 0.36, 1.96, 0.36, 0.16, 0.36

Step 5: Calculate the sum of squared differences:
- Calculate the sum of the squared differences:
  0.36 + 0.16 + 1.96 + 2.56 + 0.16 + 0.36 + 1.96 + 0.36 + 0.16 + 0.36 = 8.32

Step 6: Calculate the variance:

- Calculate the variance of the satisfaction scores using the formula:
  variance = sum of squared differences / (n - 1)
  where n is the number of observations.
  variance = 8.32 / (10 - 1) = 8.32 / 9 ≈ 0.924 (rounded to three decimal places)

Step 7: Calculate the standard error of the mean:
- Calculate the standard error of the mean using the formula:
  standard error = sqrt(variance / n)
  where n is the number of observations.
  standard error = sqrt(0.924 / 10) ≈ 0.304 (rounded to three decimal places)

Therefore, the standard error of the mean satisfaction score is approximately 0.304.

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 and predict sales based on advertising expenditure, you can use the following steps:

    Step 1: Organize the data:
    - Advertising Expenditure (in thousands): [10, 15, 12, 8, 14]
    - Sales (in thousands): [25, 30, 28, 20, 26]

    Step 2: Calculate the means:
    - Calculate the mean of the advertising expenditure ($\bar{x}$):
      $\bar{x}$ = (10 + 15 + 12 + 8 + 14) / 5 = 59 / 5 = 11.8
    - Calculate the mean of the sales ($\bar{y}$):
      $\bar{y}$ = (25 + 30 + 28 + 20 + 26) / 5 = 129 / 5 = 25.8

    Step 3: Calculate the deviations:
    - Calculate the deviations of the advertising expenditure from the mean (x - $\bar{x}$):
      (10 - 11.8), (15 - 11.8), (12 - 11.8), (8 - 11.8), (14 - 11.8) = -1.8, 3.2, 0.2, -3.8, 2.2
    - Calculate the deviations of the sales from the mean (y - $\bar{y}$):
      (25 - 25.8), (30 - 25.8), (28 - 25.8), (20 - 25.8), (26 - 25.8) = -0.8, 4.2, 2.2, -5.8, 0.2

    Step 4: Calculate the product of deviations:

- Calculate the product of the deviations of advertising expenditure and sales $(x - \bar{x})(y - \bar{y})$:
  (-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:
- Calculate the squared deviations of the advertising expenditure $(x - \bar{x})^2$:
  $(-1.8)^2$, $(3.2)^2$, $(0.2)^2$, $(-3.8)^2$, $(2.2)^2$ = 3.24, 10.24, 0.04, 14.44, 4.84
- Calculate the squared deviations of the sales $(y - \bar{y})^2$:
  $(-0.8)^2$, $(4.2)^2$, $(2.2)^2$, $(-5.8)^2$, $(0.2)^2$ = 0.64, 17.64, 4.84, 33.64, 0.04

Step 6: Calculate the sums:
- Calculate the sum of the product of deviations $(\Sigma[(x - \bar{x})(y - \bar{y})])$:
  1.44 + 13.44 + 0.44 + 22.04 + 0.44 = 37.8
- Calculate the sum of the squared deviations of the advertising expenditure $(\Sigma[(x - \bar{x})^2])$:
  3.24 + 10.24 + 0.04 + 14.44 + 4.84 = 32.8
- Calculate the sum of the squared deviations of the sales $(\Sigma[(y - \bar{y})^2])$:
  0.64 + 17.64 + 4.84 + 33.64 + 0.04 = 57.8

Step 7: Calculate the regression coefficients:
- Calculate the regression coefficient for advertising expenditure ($\beta1$) using the formula:
  $\beta1 = \Sigma[(x - \bar{x})(y - \bar{y})] / \Sigma[(x - \bar{x})^2]$
  $\beta1 = 37.8 / 32.8 \approx 1.15$ (rounded to two decimal places)

Step 8: Calculate the intercept ($\beta0$):
- Calculate the intercept of the regression line ($\beta0$) using the formula:
  $\beta0 = \bar{y} - \beta1 * \bar{x}$
  $\beta0 = 25.8 - (1.15 * 11.8) \approx 12.35$ (rounded to two decimal places)

Step 9: Write the regression equation:
- The regression equation is $y = \beta0 + \beta1 * x$, where y represents sales and x represents advertising expenditure.
- In this case, the regression equation is $y = 12.35 + 1.15 * x$.

Step 10: Interpret the results:
- The regression coefficient ($\beta1$) of 1.15 indicates that for every one-unit increase in advertising expenditure (in thousands), the predicted sales increase by 1.15 units (in thousands).
- The intercept ($\beta0$) of 12.35 represents the predicted sales when the advertising expenditure is zero (which may not be meaningful in this context).

- The regression equation can be used to predict the sales based on a given advertising expenditure value.

Therefore, the multiple regression equation to predict sales based on advertising expenditure is y = 12.35 + 1.15 * x.