**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.**

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

Mean of Region B = (18+20+16+22+25)/5 = 20.2

**2. Scenario: A survey is conducted to measure customer satisfaction on a scale of 1 to 5. The data collected is as follows:**

**[4, 5, 2, 3, 5, 4, 3, 2, 4, 5]**

**Calculate the mode of the survey responses.**

4 and 5 are the modes

**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.**

Median of Department A is 5750

Median of 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.**

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

**5. Scenario: A study is conducted to compare the performance of two different teaching methods. The test scores of the students in each group are as follows:**

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

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

**Perform a t-test to determine if there is a significant difference in the mean scores between the two groups.**

Mean of group A and group B = 89.2 and 86.6 and Variance of group

**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.**

Step 1. calculating mean of Advertising and sales = 11.8 and 25.8

Step 2. calculating sd of both = [-1.8,3.2,0.2,-3.8,2.2] and [-0.8,4.2,2.2,-5.8,0.2]

Step 3. product of sd = [1.44,13.44,0.44,22.04,0.44]

Step 4. calculating squared of sd of each variable = [3.24, 10.24, 0.04, 14.44, 4.84] and [0.64, 17.64, 4.84, 33.64, 0.04]

Step 5. calculating sum of squared of sd of each variable = 37.8 and 57.8

Step 6. Calculating square root of sum of squared deviations of each variable = 5.81 and 7.61

Step 7. Calculating correlation coefficient = sum of products of sd/(square root of sum of squared sd of both) = 37.8/(5.81\*7.61) = ***0.856***

**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.**

Mean of height = 166.43

Devation of each height = [-6.43,3.57,-1.43,-11.43,8.59,13.57,3.57]

Squared of each deviation = [41.344,12.744,2.044,130.62,73.44,183.78,12.74]

Sum of squared of devation = 456.73

Calculate variance = 76.122

SD = square root of variance(76.12) = 8.72

**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.**

Index of 75th percentile = (75/100)\*(10+1) = 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.**

**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.**

Caluculating mean and sd then set critical value for 95%ci

**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.**

Q1 = 37.5 and Q3 = 57.5

**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.**

**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.**

Calculating sample mean and sd = 7.7 and 0.8  
Standard error of mean = SD-square root of sample size = 0.246

**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.**