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. A= [10, 15, 12, 8, 14]

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

res = sum(A + B) / len(A + B)

print('the mean sales for each region' , res)

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. import statistics

l = [4, 5, 2, 3, 5, 4, 3, 2, 4, 5]

print("mode of the survey", statistics.mode(l))

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. import statistics as st

A = [5000, 6000, 5500, 7000]

median\_value = st.median(A)

print(median\_value)

import statistics as st

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

median\_value = st.median(A)

print(median\_value)

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. l=   [25.5, 24.8, 26.1, 25.3, 24.9]

range = max(l)-min(l)

print(range)

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. import scipy.stats as stats

import numpy as np

expenditure = [85, 90, 92, 88, 91]

sales = [82, 88, 90, 86, 87]

print(np.var(expenditure), np.var(sales))

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. import numpy as np

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

sales = [25, 30, 28, 20, 26]

corr\_coeff = np.corrcoef(expenditure, sales)

print(corr\_coeff)

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. import statistics

print(statistics.stdev([160, 170, 165, 155, 175, 180, 170]))

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.

Ans. from scipy.stats import f\_oneway

medicationA = [10, 12, 14, 11, 13]

medicationB = [15, 17, 16, 14, 18]

f\_oneway(medicationA, medicationB)

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. import numpy as np

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

print(np.percentile(arr, 75))

11. Scenario: A quality control department wants to test the weight consistency of a product. The weights of a sample of products are as follows:

[10.2, 9.8, 10.0, 10.5, 10.3, 10.1]

Perform a hypothesis test to determine if the mean weight differs significantly from 10 grams.

Ans. from scipy.stats import ttest\_1samp

import numpy as np

weight = [10.2, 9.8, 10.0, 10.5, 10.3, 10.1]

print(ages)

mean = np.mean(weight)

print(mean)

t\_test, p\_val = ttest\_1samp(weight, 30)

print("P-value is: ", p\_val)

if p\_val < 0.05:

    print(" We can reject the null hypothesis")

else:

    print("We can accept the null hypothesis")

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. from scipy.stats import chi2\_contingency

# defining the table

data = [[100, 120, 110, 90, 95], [80, 85, 90, 95, 100]]

stat, p, dof, expected = chi2\_contingency(data)

# interpret p-value

alpha = 0.05

print("p value is " + str(p))

if p <= alpha:

  print('Dependent (reject H0)')

else:

  print('Independent (H0 holds true)')

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. import numpy as np

import scipy.stats as st

# define sample data

gfg\_data = [7, 9, 6, 8, 10, 7, 8, 9, 7, 8]

# create 95% confidence interval

st.t.interval(alpha=0.95, df=len(gfg\_data)-1,

      loc=np.mean(gfg\_data),

      scale=st.sem(gfg\_data))

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. import numpy as np

import matplotlib.pyplot as plt

def estimate\_coef(x, y):

  # number of observations/points

  n = np.size(x)

  # mean of x and y vector

  m\_x = np.mean(x)

  m\_y = np.mean(y)

  # calculating cross-deviation and deviation about x

  SS\_xy = np.sum(y\*x) - n\*m\_y\*m\_x

  SS\_xx = np.sum(x\*x) - n\*m\_x\*m\_x

  # calculating regression coefficients

  b\_1 = SS\_xy / SS\_xx

  b\_0 = m\_y - b\_1\*m\_x

  return (b\_0, b\_1)

def plot\_regression\_line(x, y, b):

  # plotting the actual points as scatter plot

  plt.scatter(x, y, color = "m",

      marker = "o", s = 30)

  # predicted response vector

  y\_pred = b[0] + b[1]\*x

  # plotting the regression line

  plt.plot(x, y\_pred, color = "g")

  # putting labels

  plt.xlabel('x')

  plt.ylabel('y')

  # function to show plot

  plt.show()

def main():

  # observations / data

  x = np.array([20, 22, 23, 19, 21])

  y = np.array([8, 7, 9, 6, 8])

  # estimating coefficients

  b = estimate\_coef(x, y)

  print("Estimated coefficients:\nb\_0 = {} \

    \nb\_1 = {}".format(b[0], b[1]))

  # plotting regression line

  plot\_regression\_line(x, y, b)

if \_\_name\_\_ == "\_\_main\_\_":

  main()

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. GroupA= [4, 3, 5, 2, 4]

GroupB= [3, 2, 4, 3, 3]

import scipy.stats as stats

#perform the Mann-Whitney U test

stats.mannwhitneyu(GroupA, GroupB, alternative='two-sided')

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. import numpy as np

#define array of data

data = np.array([25, 30, 35, 40, 45, 50, 55, 60, 65, 70])

#calculate interquartile range

q3, q1 = np.percentile(data, [75 ,25])

iqr = q3 - q1

#display interquartile range

iqr

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. 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]

from scipy import stats

# Conduct the Kruskal-Wallis Test

result = stats.kruskal(Algorithm\_A, Algorithm\_B, Algorithm\_C)

print(result)

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.

Ans. import numpy as np

import matplotlib.pyplot as plt

def estimate\_coef(x, y):

  # number of observations/points

  n = np.size(x)

  # mean of x and y vector

  m\_x = np.mean(x)

  m\_y = np.mean(y)

  # calculating cross-deviation and deviation about x

  SS\_xy = np.sum(y\*x) - n\*m\_y\*m\_x

  SS\_xx = np.sum(x\*x) - n\*m\_x\*m\_x

  # calculating regression coefficients

  b\_1 = SS\_xy / SS\_xx

  b\_0 = m\_y - b\_1\*m\_x

  return (b\_0, b\_1)

def plot\_regression\_line(x, y, b):

  # plotting the actual points as scatter plot

  plt.scatter(x, y, color = "m",

      marker = "o", s = 30)

  # predicted response vector

  y\_pred = b[0] + b[1]\*x

  # plotting the regression line

  plt.plot(x, y\_pred, color = "g")

  # putting labels

  plt.xlabel('x')

  plt.ylabel('y')

  # function to show plot

  plt.show()

def main():

  # observations / data

  x = np.array([10, 15, 12, 8, 14])

  y = np.array([100, 80, 90, 110, 95])

  # estimating coefficients

  b = estimate\_coef(x, y)

  print("Estimated coefficients:\nb\_0 = {} \

    \nb\_1 = {}".format(b[0], b[1]))

  # plotting regression line

  plot\_regression\_line(x, y, b)

if \_\_name\_\_ == "\_\_main\_\_":

  main()

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. import pandas as pd

data= [7, 8, 9, 6, 8, 7, 9, 7, 8, 7]

df = pd.DataFrame(data)

print(df.sem())

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.