1)What is the difference between descriptive statistics and inferential statistics? Explain with examples.

→ Descriptive statistics summarizes and describes the characteristics of a dataset, while inferential statistics uses sample data to draw conclusions and make predictions about a larger population.

Example of descriptive statistics:
>>Calculating the average score on a math test for all students in a classroom (the whole population).

Example of inferential statistics:
>>Conducting a poll to estimate the percentage of voters in a city who support a particular candidate, based on a survey of a random sample of voters.

2) What is sampling in statistics? Explain the differences between random and stratified sampling.

→ the process of selecting a subset of individuals from a larger population to study and draw conclusions about the entire population.

Difference between random and stratified sampling:
>>Simple random sampling gives every member of the population an equal chance of being chosen, while stratified random sampling divides the population into subgroups (strata) and then randomly samples within each subgroups.

3)Define mean, median, and mode. Explain why these measures of central tendency are important.

→Mean, median, and mode are measures of central tendency, which describe the "typical" value in a dataset. Mean is the average, calculated by summing all values and dividing by the count. Median is the middle value when data is ordered. Mode is the most frequent value.
Explaination:
>> They are important because they summarize large datasets with a single value, making it easier to understand patterns and make comparisons.

4)Explain skewness and kurtosis. What does a positive skew imply about the data?

→ Skewness and kurtosis are statistical measures that describe the shape of a data distribution. Skewness indicates the asymmetry of the distribution, while kurtosis measures the "tailedness" or the presence of outliers. Positive skewness, also known as right skewness, means the data is more concentrated on the lower values, with a longer tail extending towards higher values.

5) Implement a Python program to compute the mean, median, and mode of a given list of numbers. numbers = [12, 15, 12, 18, 19, 12, 20, 22, 19, 19, 24, 24, 24, 26, 28] .

→ **mean:**

```
Import numpy as np
Df =  [12, 15, 12, 18, 19, 12, 20, 22, 19, 19, 24, 24, 24, 26, 28]
np.mean(df)
```

Output:> (19.6)

Median:

```
np.median(df)
```

Output: > (19.0)

Mode:

```
from scipy import stats
stats.mode(df)
```

output : > `ModeResult(mode=np.int64(12), count=np.int64(3))`

6)Compute the covariance and correlation coefficient between the following two datasets provided as lists in Python: list_x = [10, 20, 30, 40, 50] list_y = [15, 25, 35, 45, 60]

→**covariance :**

```
import numpy as np
Covariance_mat = np.cov(list_x,list_y)
Covariance = covariance_mat[0,1]
```

output :

(275.0)

Correlation coefficient :

```
Correlation_coefficient = np.corrcoef(list_x,list_y)[0,1]
print(correlation_coefficient)
```

output:> `0.995893206467704`

7)Write a Python script to draw a boxplot for the following numeric list and identify its outliers. Explain the result: data = [12, 14, 14, 15, 18, 19, 19, 21, 22, 22, 23, 23, 24, 26, 29, 35]

```
→ import matplotlib.pyplot as plt
Import numpy as np
plt.figure(figsize = (8,7))
plt.boxplot(data)
plt.title("boxplot")
plt.ylabel("value")
plt.grid(True)
plt.show()

Q1 = np.percentile(data, 25)
Q3 = np.percentile(data,75)

Iqr = q3-q1
 Lower_bound = q1 -1.5*Iqr
Upper_bound = q3 +1.5*Iqr

Outlier = [x for x in data if x > upper_bound or x < lower_bound]
```

Output:

```
print("q1 value",Q1) >>   17.25
print("q3 value",Q3) >>    23.25
print("iqr",Iq) >>    6.0
print("lower_bound",Lower_bound)>>  8.25
print("Upper_bound",Upper_bound)>>  32.25
print("outlier",Outlier)>>  35
```

8) You are working as a data analyst in an e-commerce company. The marketing team wants to know if there is a relationship between advertising spend and daily sales. ● Explain how you would use covariance and correlation to explore this relationship. ● Write Python code to compute the correlation between the two lists: advertising_spend = [200, 250, 300, 400, 500] daily_sales = [2200, 2450, 2750, 3200, 4000]

→To analyze the relationship between advertising spend and daily sales, covariance and correlation can be used. Covariance indicates the direction of the relationship (positive, negative, or zero), while correlation measures the strength and direction of the linear relationship

```
Import numpy as np
Correlation_coefficient = corrcoef(advertising_spend,daily_sales)[0,1]
print(correlation_coefficient)
```

Output:

```
np.float64(0.9935824101653329)
```

9) Your team has collected customer satisfaction survey data on a scale of 1-10 and wants to understand its distribution before launching a new product. ● Explain which summary statistics and visualizations (e.g. mean, standard deviation, histogram) you'd use. ● Write Python code to create a histogram using Matplotlib for the survey data: survey_scores = [7, 8, 5, 9, 6, 7, 8, 9, 10, 4, 7, 6, 9, 8, 7]

→To understand the distribution of customer satisfaction data, you should use a combination of summary statistics like mean and standard deviation, and visualizations like a histogram. The mean provides a measure of central tendency, while the standard deviation indicates the spread of the data. A histogram visually represents the frequency distribution of scores, revealing patterns and potential outliers.

```
Import matplotlib.pyplot as plt
Import numpy as np

np.mean(survey_scores)  >> np.float64(7.333333333333333)
np.var(survey_scores)  >> np.float64(2.488888888888889)
np.std(survey_scores) >> np.float64(1.577621275493231)
```

```
plt.hist(survey_scores, bins = range(1,15),edgecolor = ‘white’,color = ‘blue’)
plt.xlabel(“customer satisfaction”)
plt.ylabel(“frequency”)
plt.title(“survey”)
plt.xticks(range(1,15))
plt.show()
```