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

Answer:
• Descriptive statistics summarize and describe the main features of a dataset.
• They include measures such as mean, median, mode, standard deviation, and graphs.
• Example: Calculating the average height of students in a class.

• Inferential statistics draw conclusions or make predictions about a population based on a sample.
• They use techniques like hypothesis testing, confidence intervals, and regression.
• Example: Predicting the average height of all students in India using a sample of 1,000 students.

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

Answer:
• Sampling is the process of selecting a subset of a population for analysis.
• It helps collect data efficiently when studying an entire population is not practical.

• Random sampling:
– Every individual in the population has an equal chance of being selected.
– Simple to implement.
– Can lead to uneven representation of subgroups.

• Stratified sampling:
– Population is divided into groups (strata) based on characteristics.
– Samples are taken from each stratum.
– Ensures better representation of all groups.

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

Answer:
- Mean: The arithmetic average of values.
- Median: The middle value when data is arranged in order.
- Mode: The value that appears most frequently.

- Importance:
  – They summarize the central point of a dataset.
  – They simplify large datasets for analysis.
  – They help compare different groups or distributions.
  – They assist in making informed decisions based on typical values.

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

Answer:
- Skewness measures the asymmetry in the distribution of data.
- Kurtosis measures how peaked or flat the distribution is compared to a normal distribution.

- Positive skew (right skew) implies:
  – The tail of the distribution extends toward the right side.
  – Mean is greater than median.
  – Most values are lower, with a few unusually high values.

Question 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]

Answer:

```
from statistics import mean, median, mode

numbers = [12, 15, 12, 18, 19, 12, 20, 22, 19, 19, 24, 24, 24, 26, 28]

print("Mean:", mean(numbers))
print("Median:", median(numbers))
print("Mode:", mode(numbers))
```

Expected output:
- Mean ≈ 19.6
- Median = 19
- Mode = 12

Question 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]

Answer:

```
import numpy as np

list_x = [10, 20, 30, 40, 50]
```

```
list_y = [15, 25, 35, 45, 60]

covariance = np.cov(list_x, list_y, bias=True)[0][1]
correlation = np.corrcoef(list_x, list_y)[0][1]

print("Covariance:", covariance)
print("Correlation:", correlation)
```

Expected output:
- Covariance = 187.5
- Correlation ≈ 0.993

Question 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]
```

Answer:

```
import matplotlib.pyplot as plt

data = [12, 14, 14, 15, 18, 19, 19, 21, 22, 22, 23, 23, 24, 26, 29, 35]

plt.boxplot(data)
plt.title("Boxplot of Data")
plt.xlabel("Index")
plt.ylabel("Value")
plt.show()
```

Explanation:
- The value 35 appears as an outlier.
- It lies above the upper whisker, which is beyond 1.5 times the interquartile range (IQR).
- This indicates that 35 is unusually high compared to other data points.

Question 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]

Answer:
• Covariance shows whether the relationship is positive or negative.
• A positive covariance means both advertising spend and sales increase together.
• Correlation measures the strength of the relationship on a scale of -1 to +1.
• A value close to +1 indicates a strong positive relationship.

Python code:

```
import numpy as np

advertising_spend = [200, 250, 300, 400, 500]
daily_sales = [2200, 2450, 2750, 3200, 4000]

correlation = np.corrcoef(advertising_spend, daily_sales)[0][1]
print("Correlation:", correlation)
```

Expected output:
• Correlation ≈ 0.98 (very strong positive relationship)

Question 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 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]

Answer:
• Summary statistics used: mean, median, mode, standard deviation, range.
• Visualizations used: histogram (to see distribution), boxplot (to detect outliers), density plot (to observe shape).

Python code:

```
import matplotlib.pyplot as plt

survey_scores = [7, 8, 5, 9, 6, 7, 8, 9, 10, 4, 7, 6, 9, 8, 7]

plt.hist(survey_scores, bins=5)
plt.title("Histogram of Survey Scores")
plt.xlabel("Scores")
plt.ylabel("Frequency")
plt.show()
```

.