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 using measures such as mean, median, mode, standard deviation, and graphical representations (like histograms and boxplots). For example, calculating the average marks of students in a class is descriptive statistics. Inferential statistics, on the other hand, involves making predictions or inferences about a larger population based on a sample of data. For example, conducting a survey on 1000 people to infer the voting preference of an entire country is inferential statistics.

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 individuals or observations from a larger population to estimate characteristics of the entire population. Random sampling is when every individual in the population has an equal chance of being selected. Stratified sampling involves dividing the population into subgroups (strata) based on specific characteristics (e.g., age, income) and then taking random samples from each stratum. This ensures representation from all important subgroups.

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

Answer: The mean is the arithmetic average of a dataset. The median is the middle value when the data is arranged in order. The mode is the value that occurs most frequently in the dataset. These measures are important because they summarize the dataset into a single representative value, making it easier to understand the central tendency of the data. For example, the mean gives the overall average, the median is useful for skewed distributions, and the mode identifies the most common value.

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

Answer: Skewness measures the asymmetry of the probability distribution of data. A positive skew indicates that the tail on the right side of the distribution is longer, meaning more values are concentrated on the lower side. Kurtosis measures the 'tailedness' of the distribution. High kurtosis means more data are in the tails, while low kurtosis means fewer extreme values. A positive skew implies that most values are below the mean, but a few very high values pull the mean to the right.

Question 5: Implement a Python program to compute the mean, median, and mode of a given list of numbers.

```
import statistics

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

mean_val = statistics.mean(numbers)
median_val = statistics.median(numbers)
mode_val = statistics.mode(numbers)
print("Mean:", mean_val)
```

```
print("Median:", median_val)
print("Mode:", mode_val)
```
Output: Mean: 20.0 Median: 19.0 Mode: 12

Question 6: Compute the covariance and correlation coefficient between the following two datasets provided as lists in Python.

```
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 Coefficient:", correlation)
```
Output: Covariance: 200.0 Correlation Coefficient: 0.9934

Question 7: Write a Python script to draw a boxplot for the following numeric list and identify its outliers. Explain the result:

```
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.show()
```
Explanation: The boxplot shows that most data points are within the interquartile range. The value 35 appears as an outlier because it lies significantly higher than the rest of the data.

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.

Answer: Covariance helps to understand whether advertising spend and sales move in the same direction (positive covariance) or opposite directions (negative covariance). Correlation measures the strength and direction of the linear relationship between the two variables, ranging from -1 to +1. A high positive correlation would suggest that increased advertising spend is associated with higher daily sales.

```
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)
```
Output: Correlation: 0.993

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.

Answer: To understand the distribution, summary statistics such as mean, median, mode, variance, and standard deviation can be calculated. A histogram provides a visual understanding of how scores are distributed. This helps the company gauge overall customer satisfaction trends.

```
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=6, edgecolor='black')
plt.title("Histogram of Customer Satisfaction Scores")
plt.xlabel("Scores")
plt.ylabel("Frequency")
plt.show()
```

The histogram would show most scores clustering around 7–9, indicating high satisfaction.