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

**Answer:**

|  |
| --- |
| Descriptive statistics are used to **summarize, organize, and present** data in a meaningful way. It provides simple summaries about the sample and the measures.  **Example:**  Suppose you survey 100 students and find the **average height** is **165 cm**. This average is descriptive—it tells you about your current dataset.  ***Inferential statistics*** use data from a sample to make predictions or inferences about a larger population.  Example - find the average height whole population of india which is impossible . |

**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 (a sample) from a larger group (population) to estimate characteristics of the whole population.  1. Random Sampling (Simple Random Sampling)  In random sampling, every member of the population has an equal chance of being selected.  2. Stratified Sampling  In stratified sampling, the population is divided into subgroups based on a specific characteristic (e.g., gender, age, income), and then a random sample is taken from each subgroup. |

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

**Answer:**

|  |
| --- |
| Mean - The average of all values.  Mean=Sum of all values/Number of values  Median - The middle value when the data is arranged in order. It divides the data into two equal parts.  Mode - The most frequently occurring value in the dataset.  Importance of the central tendency in data   1. Summarizes the Data 2. Helps in comparion 3. Supports in decision making |

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

**Answer:**

|  |
| --- |
| Skewness –  Skewness measures the **asymmetry** of the data distribution.   * **Symmetric Distribution** → Skewness = 0  (Mean ≈ Median ≈ Mode) * **Positive Skew (Right Skew):** Tail is longer on the **right** side.   Mean>Median>Mode   * **Negative Skew (Left Skew):** Tail is longer on the **left** side.   Mean<Median<Mode  Kurtosis - Kurtosis measures the **tailedness** or **peakedness** of the distribution.  Importance of the Positive of skew :-  In Positive skew data, tail is present on the right side and most of the data is present in the left side but some large value present on the right side. |

**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:**

|  |
| --- |
| **Output :** |

**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:**

|  |
| --- |
| Output : |

**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:**

|  |
| --- |
|  |

**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:**   * **Measures the direction of the relationship between two variables.** * **Positive → both variables increase together** * **Negative → one increases, the other decreases** * **But it's scale-dependent, so hard to interpret by itself.**   **🔹 Correlation:**   * **Standardized version of covariance, ranges from -1 to 1.** * **+1 → perfect positive relationship** * **0 → no relationship** * **–1 → perfect negative relationship** * **Easier to compare across different datasets.**   ** Covariance will tell you if advertising spend and daily sales move in the same direction.**  ** Correlation will tell you how strongly they're related (independent of scale).** |

**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 (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]

|  |
| --- |
| Output |