**1) What is the purpose of descriptive statistics?**

**A)** Descriptive statistics are used to summarize and describe the main features of a dataset.  
They help you understand the central tendency, spread, and shape of your data using numerical and graphical methods.

**Main purposes:**

* Summarize large datasets in a simple way
* Identify patterns or outliers
* Prepare data for further analysis (like inferential statistics)
* Provide a snapshot of data characteristics (mean, median, mode, etc.).

**2) Can you explain the difference mean, median, and mode?**

**A) Mean (Average):**

* **Definition:** The sum of all values divided by the number of values.
* **Best used when:** The data is symmetrical and has no outliers.
* Sensitive to outliers (extreme values).

**Example:**Data = [10, 20, 30, 40, 100]  
Mean = (10 + 20 + 30 + 40 + 100) / 5 = 200 / 5 = 40

Notice how the 100 pulls the mean up

**Median (Middle value):**

* **Definition:** The middle number in an ordered dataset**.**
* **Best used when:** The data has outliers or is skewed.
* Not sensitive to outliers.

**Example:**Data = [10, 20, 30, 40, 100] (already ordered)  
Median = 30

Still reflects the center well, even with the 100 outlier.

**Mode (Most frequent value):**

* **Definition:** The value(s) that occur most often in a dataset.
* **Best used when:** You want to find the most common value.
* Can be no mode, one mode, or multiple modes.

**Example:**Data = [10, 20, 20, 30, 40, 100]  
Mode = 20 (because it appears twice).

**3) How do you interpret standard deviation in a dataset?**

**A**) Standard deviation (SD) measures the spread or dispersion of data points from the mean.

* A low SD means data points are close to the mean (less spread).
* A high SD means data points are more spread out from the mean.

Example:

* SD of **test scores** = 0 ➝ Everyone got the same score.
* SD = 15 ➝ Scores are widely spread, some did very well, others poorly.

**4) Describe the concept of skewness in statistics**

**A)** Skewness measures the asymmetry of the data distribution.

* **Symmetrical distribution** → Skewness = 0
* **Positive skew (Right-skewed):**  
  Long tail on the right. Mean > Median  
  (e.g., income distribution — a few people earn much more)
* **Negative skew (Left-skewed):**  
  Long tail on the left. Mean < Median  
  (e.g., test scores where most score high, a few score very low)

Skewness helps understand if the data is balanced or biased toward one side.

**5) What is the main goal of inferential statistics?**

**A)** Inferential statistics aim to:

Make conclusions, predictions, or generalizations about a population based on a sample of data.

Instead of studying an entire population (which is often impossible), we:

* Take a representative sample
* Perform analysis on that sample
* Use the results to infer characteristics of the whole population

**Main goals:**

* Estimate population parameters (like mean, proportion)
* Test hypotheses
* Predict future outcomes

**6) Explain the difference between a population and a sample**

**A) Population**

* The entire group you want to study or gather information about.
* It includes all possible individuals or items that fit your criteria.
* Often too large to study directly.

**Example:**All college students in India  
All cars manufactured by a company in 2024  
Every voter in a country

**Sample**

* A smaller subset selected from the population.
* Used to represent the population.
* More practical and cost-effective to study.

**Example**:  
1,000 students randomly chosen from colleges across India  
200 cars selected for quality testing  
500 voters surveyed before an election

**7) What is a confidence interval and how is it used in inferential statistics?**

**A)** A confidence interval (CI) gives a range of values that’s likely to contain the true population parameter (like the mean or proportion) with a certain level of confidence (usually 95% or 99%).

**In simpler terms:**

"We are 95% confident that the true average height of students is between 160 cm and 170 cm."

* **Wider intervals =** more uncertainty
* **Narrower intervals =** more precision

CI helps in decision making by showing the reliability of the estimate**.**

**8) Define p-score (p-value)**

**A)** The p-value (probability value) tells us how likely it is to get the observed results by chance, if the null hypothesis is true**.**

**Used in hypothesis testing:**

* A small p-value (< 0.05) suggests the results are statistically significant and not due to chance.
* A large p-value (> 0.05) means there is not enough evidence to reject the null hypothesis.

**Example:**p-value = 0.03  
➝ There's only a 3% chance the results occurred due to random variation ➝ likely significant.