Question 1: Explain the different types of data (qualitative and quantitative) and provide examples of each. Discuss nominal, ordinal, interval, and ratio scales.

Ans-Qualitative Data: Non-numeric data that represent categories or qualities. Examples:

Colours of cars (red, blue, black).

Types of fruit (apple, banana).

Quantitative Data: Numeric data that represent amounts. Examples:

Age of people (25, 30, 35).

Heights of buildings (200m, 300m).

Nominal Scale: Categories with no inherent order. Example: Gender (male, female).

Ordinal Scale: Ordered categories, but differences between them are not measurable. Example: Customer satisfaction ratings (poor, fair, good).

Interval Scale: Ordered categories with measurable differences, but no true zero point. Example: Temperature in Celsius.

Ratio Scale: Like interval, but with a meaningful zero point. Example: Weight (0 kg means no weight).

Question 2: What are the measures of central tendency, and when should you use each? Discuss the mean, median, and mode with examples and situations where each is appropriate.

Mean: The arithmetic average. Used when data is symmetrically distributed. Example: Average test scores.

Median: The middle value when data is ordered. Used when data is skewed or contains outliers. Example: Median house prices.

Mode: The most frequent value. Used for categorical data. Example: Mode of clothing sizes in a store.

Question 3: Explain the concept of dispersion. How do variance and standard deviation measure the spread of data?

Dispersion refers to the spread or variability of data points. It shows how much the values differ from the mean.

Variance is the average of squared differences from the mean. It measures the overall spread.

Standard deviation is the square root of the variance, making it easier to interpret since it is in the same units as the data.

Question 4: What is a box plot, and what can it tell you about the distribution of data?

A **box plot** visually represents the distribution of data through quartiles. It shows:

- The median (centre line in the box).
- The interquartile range (the box's width).
- The minimum and maximum values (whiskers).
- Outliers (points outside the whiskers).

It helps in identifying skewness, spread, and potential outliers.

Question 5: Discuss the role of random sampling in making inferences about populations.

Random sampling involves selecting a subset of individuals from a population where each has an equal chance of being chosen. This method ensures the sample represents the population, allowing valid **inferences** about population parameters from the sample statistics, reducing bias.

Question 6: Explain the concept of skewness and its types. How does skewness affect the interpretation of data?

Skewness measures the asymmetry of a distribution. There are three types:

- **Positive Skew**: Tail on the right; data is concentrated on the left.
- **Negative Skew**: Tail on the left; data is concentrated on the right.
- **No Skew**: Symmetrical data.

Skewness affects data interpretation, particularly in using the mean and median. In positively skewed data, the mean is greater than the median, and in negatively skewed data, the median is greater than the mean.

Question 7: What is the interquartile range (IQR), and how is it used to detect outliers?

The **Interquartile Range (IQR)** is the range between the first quartile (Q1) and third quartile (Q3):

$$IQR=Q3−Q1$$

It measures the spread of the middle 50% of data. To detect outliers, you look for values that fall below $Q1−1.5×IQR$Q1 - 1.5 \times IQR

$$Q1−1.5×IQR \text{ or above } Q3+1.5×IQR$$Q3 + 1.5 * IQR

Question 8: Discuss the conditions under which the binomial distribution is used.

The **binomial distribution** is used under the following conditions:

1. The experiment has a fixed number of trials (n).
2. Each trial has only two outcomes: success or failure.
3. The probability of success (p) remains constant across all trials.
4. The trials are independent.

Example: Flipping a coin 10 times to count heads.

Question 9: Explain the properties of the normal distribution and the empirical rule (68-95-99.7 rule).

The **normal distribution** is symmetric and bell-shaped, cantered around the mean. Key properties:

- Mean = Median = Mode.
- The total area under the curve equals 1.

The **Empirical Rule** (68-95-99.7 rule) states that:

- 68% of data falls within 1 standard deviation of the mean.
- 95% falls within 2 standard deviations.
- 99.7% falls within 3 standard deviations.

Question 10: Provide a real-life example of a Poisson process and calculate the probability for a specific event.

A **Poisson process** models random events occurring at a constant rate over time. **Example**: Number of customer arrivals at a store.

Probability Example: If customers arrive at a rate of 5 per hour, what is the probability of exactly 3 customers arriving in an hour?

Using the Poisson probability formula:

$$P(X=k)= \lambda^k e^{-\lambda}/ k!$$

Where $\lambda=5$ and $k=3$, we calculate the probability.

Question 11: Explain what a random variable is and differentiate between discrete and continuous random variables.

A **random variable** assigns numerical values to outcomes of a random phenomenon.

- **Discrete Random Variable**: Takes on specific values (e.g., number of heads in coin tosses).
- **Continuous Random Variable**: Takes on an infinite number of values within a range (e.g., height, weight).

Question 12: Provide an example dataset, calculate both covariance and correlation, and interpret the results.

X	Y
1	2
2	3
3	4

To calculate **covariance** and **correlation**:

1. Covariance measures the direction of the relationship.
2. Correlation (Pearson's) measures both strength and direction, with values between -1 and 1.

For the dataset, the covariance will be positive (indicating a direct relationship), and correlation will be close to 1 (indicating a strong linear relationship).