

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

Answer:

Data is broadly classified into two types: **Qualitative** and **Quantitative** data.

Qualitative Data (Categorical Data):

This type of data describes qualities or characteristics and cannot be measured numerically.

Example: Gender, Color, City names.

Types of qualitative data:

- **Nominal Scale:**

Data is categorized without any order.

Example: Gender (Male/Female), Blood group (A, B, O).

- **Ordinal Scale:**

Data has order or ranking but no fixed difference between categories.

Example: Satisfaction level (Low, Medium, High).

Quantitative Data (Numerical Data):

This type of data represents numbers and can be measured.

Types of quantitative data:

- **Interval Scale:**

Data has equal intervals but no true zero.

Example: Temperature in Celsius.

- **Ratio Scale:**

Data has equal intervals and a true zero.

Example: Height, Weight, Age.

2) What are the measures of central tendency? Discuss mean, median, and mode.

Answer:

Measures of central tendency represent the central or average value of a dataset.

- **Mean:**

Sum of all values divided by total number of values.

Best used when data has no extreme outliers.

Example:

Data = 2, 4, 6

Mean = $(2+4+6)/3 = 4$

- **Median:**

Middle value when data is arranged in order.

Best used when data has outliers.

Example:

Data = 2, 4, 100

Median = 4

- **Mode:**

Most frequently occurring value.

Best used for categorical data.

Example:

Data = 1, 2, 2, 3

Mode = 2

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

Answer:

Dispersion refers to how spread out the data values are around the mean.

- **Variance:**

It measures the average squared deviation from the mean.

Formula:

$$\text{Variance} = \sum (x_i - \text{mean})^2 / n$$

- **Standard Deviation:**

Square root of variance.

It shows how far values typically deviate from the mean.

Small standard deviation → Data is closely clustered.

Large standard deviation → Data is widely spread.

4) What is a box plot, and what does it tell about distribution?

Answer:

A box plot is a graphical representation of data based on five-number summary:

Minimum, Q1, Median, Q3, Maximum

It shows:

- Central tendency (median)
- Spread (IQR)
- Presence of outliers
- Skewness of data

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

Answer:

Random sampling ensures every individual in a population has an equal chance of being selected.

Importance:

- Reduces bias
- Provides representative sample
- Helps in making accurate population inferences
- Forms the basis of statistical tests

Without random sampling, conclusions may be incorrect.

6) Explain the concept of skewness and its types.

Answer:

Skewness measures the asymmetry of a distribution.

Types:

- **Positive Skew (Right Skew):**

Tail is longer on right side.

Mean > Median

- **Negative Skew (Left Skew):**

Tail is longer on left side.

Mean < Median

- **Zero Skew:**

Perfectly symmetrical distribution.

Skewness affects interpretation because mean gets influenced by extreme values.

7) What is Interquartile Range (IQR) and how is it used to detect outliers?

Answer:

$IQR = Q3 - Q1$

It measures the spread of the middle 50% of data.

Outlier detection rule:

Lower Bound = $Q1 - 1.5 \times IQR$

Upper Bound = $Q3 + 1.5 \times IQR$

Any value outside this range is considered an outlier.

8) Discuss the conditions under which binomial distribution is used.

Answer:

Binomial distribution is used when:

- Fixed number of trials (n)
- Only two outcomes (success/failure)

- Probability of success remains constant
- Trials are independent

Example: Tossing a coin 10 times.

9) Explain properties of normal distribution and empirical rule.

Answer:

Normal distribution is a bell-shaped symmetrical distribution.

Properties:

- Mean = Median = Mode
- Symmetrical about mean
- Total area under curve = 1

Empirical Rule (68-95-99.7 Rule):

- 68% data lies within 1 standard deviation
 - 95% within 2 standard deviations
 - 99.7% within 3 standard deviations
-

10) Provide a real-life example of Poisson process and calculate probability.

Answer:

Example: Number of calls received at a call center per minute.

Suppose average calls per minute (λ) = 2

Probability of receiving exactly 3 calls:

Formula:

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

$$\begin{aligned} P(3) &= (e^{-2} \times 2^3) / 3! \\ &= (0.1353 \times 8) / 6 \\ &= 0.1804 \end{aligned}$$

So probability ≈ 0.18

11) Explain random variable and differentiate between discrete and continuous.

Answer:

A random variable is a numerical outcome of a random experiment.

Types:

• **Discrete Random Variable:**

Takes countable values.

Example: Number of students.

- **Continuous Random Variable:**

Takes infinite values within range.

Example: Height, Weight.

12) Example dataset – Calculate covariance and correlation.

Answer:

Dataset:

$X = [1, 2, 3]$

$Y = [2, 4, 6]$

Covariance shows direction of relationship.

Since both increase together, covariance is positive.

Correlation measures strength of relationship.

Correlation coefficient (r) ranges from -1 to +1.

For this dataset:

Covariance > 0

Correlation = +1 (Perfect positive correlation)

Interpretation:

As X increases, Y increases proportionally.