1. **Types of Data: Qualitative and Quantitative**

- **Qualitative (Categorical)**: Non-numeric data representing categories or labels.
  - **Examples**: Colors of cars (red, blue, green), types of cuisine (Italian, Chinese).
- **Quantitative (Numerical)**: Numeric data representing measurable quantities.
  - **Examples**: Height (in cm), test scores (out of 100).
- **Scales of Measurement**:
  - **Nominal**: Categories without a meaningful order (e.g., types of fruits: apple, banana).
  - **Ordinal**: Categories with a meaningful order, but no consistent difference between ranks (e.g., satisfaction levels: poor, good, excellent).
  - **Interval**: Numeric scales with equal intervals but no true zero (e.g., temperature in Celsius).
  - **Ratio**: Numeric scales with a true zero, allowing ratios to be meaningful (e.g., weight in kg).

---

2. **Measures of Central Tendency**

- **Mean**: Average value. Use when data is symmetric and without extreme outliers.
  - **Example**: Average test score.
- **Median**: Middle value when data is sorted. Use when data has outliers.
  - **Example**: Median income in a region.
- **Mode**: Most frequently occurring value. Use for categorical data or to identify peaks in numerical data.
  - **Example**: Most common shoe size.

---

3. **Concept of Dispersion**

- **Dispersion**: Indicates the spread of data points.
  - **Variance**: Average of squared deviations from the mean.
  - **Standard Deviation**: Square root of variance; shows spread in the same units as the data.
- **Example**: In a dataset of exam scores, a high standard deviation indicates widely varying scores.

---

4. **Box Plot**

- **Definition**: A graphical summary using five-number summary (minimum, Q1, median, Q3, maximum).

- **Insights**: Identifies the spread, central tendency, and outliers.

---

5. **Random Sampling**

- Ensures every individual in the population has an equal chance of selection.
- **Role**: Reduces bias and enables generalization to the population.

---

6. **Skewness**

- **Concept**: Measures data asymmetry.
  - **Positive Skew**: Tail on the right (e.g., income data with a few very high incomes).
  - **Negative Skew**: Tail on the left (e.g., exam scores where most scored high).
- **Impact**: Skewness affects the choice of central tendency (e.g., median over mean in skewed data).

---

7. **Interquartile Range (IQR)**

- **Definition**: Difference between Q3 (75th percentile) and Q1 (25th percentile).
- **Use**: Identifies outliers using the rule: values below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$.

---

8. **Binomial Distribution**

- **Conditions**:
  - Fixed number of trials.
  - Each trial has two outcomes (success/failure).
  - Constant probability of success.
- **Example**: Flipping a coin 10 times to count heads.

---

9. **Normal Distribution and Empirical Rule**

- **Properties**: Symmetrical, bell-shaped curve with mean = median = mode.
- **Empirical Rule**:
  - 68% of data within 1 standard deviation.
  - 95% within 2 standard deviations.
  - 99.7% within 3 standard deviations.

---

10. **Poisson Process**

- **Example**: Number of customer arrivals at a store per hour.
- **Calculation**: If the average is 5 arrivals/hour, the probability of 7 arrivals: $P(X = 7) = \frac{\lambda^7 e^{-\lambda}}{7!}, , \lambda = 5$

---

11. **Random Variables**

- **Definition**: Variables whose outcomes are determined by chance.
    - **Discrete**: Countable outcomes (e.g., number of heads in 10 coin flips).
    - **Continuous**: Infinite outcomes (e.g., time taken to run a race).

---

12. **Covariance and Correlation**

- **Example Dataset**: [X = [2, 4, 6], , Y = [3, 6, 9]]
- **Covariance**: [\text{Cov(X, Y)} = \frac{\sum{(X_i - \bar{X})(Y_i - \bar{Y})}}{n-1}]
- **Correlation**: Standardized measure ((-1) to (1)) indicating strength and direction of relationship. [\text{Correlation} = \frac{\text{Cov(X, Y)}}{\sigma_X \cdot \sigma_Y}]
- **Interpretation**: A positive value indicates a direct relationship, and a negative value indicates an inverse relationship.

Double-click (or enter) to edit