Assignment Answers:

1. **Types of Data (Qualitative and Quantitative):**
    - **Qualitative Data** (Categorical): Describes qualities or characteristics.
        - *Nominal:* No natural order. Example: Gender (Male, Female), Blood Type (A, B, AB, O).
        - *Ordinal:* Ordered categories. Example: Satisfaction (Poor, Fair, Good, Excellent).
    - **Quantitative Data** (Numerical): Represents measurable quantities.
        - *Interval:* Numeric scale without true zero. Example: Temperature in Celsius.
        - *Ratio:* Numeric scale with true zero. Example: Height, Weight, Age.
2. **Measures of Central Tendency:**
    - **Mean (Average):** Useful when data is symmetric and has no outliers.
        - *Formula:* Mean = $\Sigma x / n$
    - **Median:** Middle value; best for skewed data or outliers.
    - **Mode:** Most frequent value; useful for categorical or repeated values.
3. **Dispersion (Spread of Data):**
    - **Range:** Difference between highest and lowest values.
    - **Variance:** Measures average squared deviation from mean.
        - *Population Variance:* $\sigma^2 = \Sigma(x - \mu)^2 / N$
        - *Sample Variance:* $s^2 = \Sigma(x - \bar{x})^2 / (n - 1)$
    - **Standard Deviation:** Square root of variance; shows average deviation.
4. **Box Plot and Data Distribution:**
    - Shows five-number summary: Minimum, Q1, Median (Q2), Q3, Maximum.
    - Detects skewness, spread, and outliers.
    - *Outlier Detection:* Q1 - 1.5×IQR and Q3 + 1.5×IQR.
5. **Importance of Random Sampling:**
    - Removes selection bias.
    - Ensures equal probability of selection.
    - Helps make reliable population inferences.
6. **Skewness and Its Interpretation:**
    - **Positive Skew:** Right tail longer; Mean > Median.
    - **Negative Skew:** Left tail longer; Mean < Median.
    - Affects mean and spread; important for selecting correct statistical tools.
7. **Interquartile Range (IQR) and Outliers:**
    - IQR = Q3 - Q1
    - Measures middle 50% spread of data.
    - Used to detect outliers via lower and upper fences.
8. **Binomial Distribution Conditions:**
    - Fixed number of trials.

- Two outcomes: Success or Failure.
- Trials are independent.
- Constant probability of success.
- *Formula:* P(X = k) = C(n, k) × p^k × (1 - p)^(n - k)

9. **Normal Distribution & Empirical Rule:**
    - Bell-shaped, symmetric curve.
    - Mean = Median = Mode.
    - *Empirical Rule:*
        - 68% within 1 standard deviation
        - 95% within 2 standard deviations
        - 99.7% within 3 standard deviations
10. **Poisson Process Example:**
    - Models random events over time (fixed average rate λ).
    - Example: 5 phone calls per hour
    - *Find P(X = 3):* P(X = 3) = (e^-5 × 5^3) / 3! ≈ 0.1404
11. **Random Variables:**
    - **Discrete:** Countable values (e.g., number of heads).
    - **Continuous:** Any value within a range (e.g., height).
12. **Covariance and Correlation Example:**
    - Dataset:

Student	Hours Studied (X)	Exam Score (Y)
A	2	65
B	4	70
C	6	80
D	8	90

    - Mean of X = 5, Mean of Y = 76.25
    - Covariance = 54.7
    - Correlation (r) = 0.98 → Strong Positive Relationship
    - *Interpretation:* As study time increases, exam scores also increase significantly.