**1. Explain the different types of data (qualitative and quantitative) and provide examples of each. Discuss nominal, ordinal, interval, and ratio scales.** Qualitative data describes categories (e.g., colors, car brands), while quantitative data represents numerical measurements (e.g., height, temperature). Nominal scales categorize without order (e.g., gender). Ordinal scales categorize with order but unequal intervals (e.g., education levels). Interval scales have ordered categories with equal intervals but no true zero (e.g., temperature in Celsius). Ratio scales have ordered categories with equal intervals and a true zero (e.g., height).

**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.** Measures of central tendency describe the center of a dataset. The mean (average) is suitable for symmetrically distributed numerical data. The median (middle value) is best for skewed data or data with outliers. The mode (most frequent value) is appropriate for categorical or discrete data, and for identifying peaks in distributions.

**3. Explain the concept of dispersion. How do variance and standard deviation measure the spread of data?** Dispersion measures the spread or variability of data. Variance is the average of squared deviations from the mean. Standard deviation is the square root of the variance. Both indicate how much individual data points deviate from the mean; higher values mean greater spread.

**4. What is a box plot, and what can it tell you about the distribution of data?** A box plot, also known as a box-and-whisker plot, visually displays the distribution of numerical data through its quartiles. It shows the median, interquartile range (IQR), and potential outliers, indicating symmetry or skewness and spread.

**5. Discuss the role of random sampling in making inferences about populations.** Random sampling is crucial for making valid inferences about a larger population from a smaller sample. It ensures every member of the population has an equal chance of selection, reducing bias and making the sample representative.

**6. Explain the concept of skewness and its types. How does skewness affect the interpretation of data?** Skewness measures the asymmetry of a probability distribution. Positive skewness (right-skewed) means the tail is longer on the right; negative skewness (left-skewed) means it's longer on the left. Skewness affects interpretation by indicating where data is concentrated and the direction of extreme values.

**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 the third quartile (Q3). It represents the middle 50% of the data. Outliers are often detected as values falling below Q1 - 1.5 \* IQR or above Q3 + 1.5 \* IQR.

**8. Discuss the conditions under which the binomial distribution is used.** The binomial distribution is used when there's a fixed number of independent trials, each with only two possible outcomes (success/failure), and the probability of success remains constant for every trial.

**9. Explain the properties of the normal distribution and the empirical rule (68-95-99.7 rule).** The normal distribution is a symmetric, bell-shaped distribution. The empirical rule states that for a normal distribution, approximately 68% of data falls within one standard deviation of the mean, 95% within two, and 99.7% within three[cite: 11].

**10. Provide a real-life example of a Poisson process and calculate the probability for a specific event.** A Poisson process describes events occurring independently at a constant average rate over a fixed interval. For example, the number of calls received by a call center per hour. If the average rate is 5 calls per hour, the probability of receiving exactly 3 calls can be calculated using the Poisson probability mass function.

**11. Explain what a random variable is and differentiate between discrete and continuous random variables.** A random variable is a variable whose value is a numerical outcome of a random phenomenon. Discrete random variables can take on a finite or countably infinite number of values (e.g., number of heads in coin flips). Continuous random variables can take on any value within a given range (e.g., height of a person).

**12. Provide an example dataset, calculate both covariance and correlation, and interpret the results.** Dataset: X = [1, 2, 3], Y = [2, 4, 5]. Covariance measures the joint variability of two variables. A positive covariance indicates they tend to increase/decrease together. Correlation measures the strength and direction of a linear relationship, normalized between -1 and 1. A correlation close to 1 implies a strong positive linear relationship, while 0 means no linear relationship