<https://chatgpt.com/share/b58fb359-9b15-44f5-8d31-8c6f888b6ecd>

 50 questions covering various topics in descriptive statistics and probability distributions, which are fundamental to data science:

Mean, Median, Mode

1. What is the mean of the dataset: [10, 15, 12, 18, 20]?
2. Calculate the median of the dataset: [5, 7, 12, 13, 15, 16, 20].
3. Determine the mode of the dataset: [6, 8, 8, 10, 10, 10, 12].
4. How do you interpret the mean of a dataset in terms of central tendency?
5. If a dataset is [4, 9, 11, 15, 17, 20], what is the median if an additional value 10 is added to the dataset?

Variance and Standard Deviation

1. Calculate the variance of the dataset: [2, 4, 4, 4, 5, 5, 7, 9].
2. What is the standard deviation of the dataset: [1, 3, 5, 7, 9]?
3. Explain the relationship between variance and standard deviation.
4. Why is the standard deviation preferred over variance when comparing variability?
5. Compute the variance and standard deviation of the dataset: [12, 15, 14, 16, 18].

Correlation and Covariance

1. Define the correlation coefficient and its range.
2. Calculate the correlation coefficient between [X: 1, 2, 3, 4, 5] and [Y: 2, 4, 5, 4, 5].
3. What is covariance and how is it different from correlation?
4. Compute the covariance of the datasets [X: 1, 2, 3, 4, 5] and [Y: 3, 6, 7, 8, 9].
5. Why is correlation a better measure of the relationship between two variables compared to covariance?

Skewness and Kurtosis

1. Define skewness and describe the difference between positive and negative skewness.
2. Given the dataset [3, 5, 6, 8, 9, 15, 21], determine if it is positively or negatively skewed.
3. What is kurtosis and what does it tell us about the distribution of data?
4. How does high kurtosis affect the tails of a distribution?
5. Calculate the skewness of the dataset: [2, 4, 6, 8, 10, 10, 12].

Probability Distributions

1. Define a probability distribution and its properties.
2. What is the difference between a discrete and a continuous probability distribution?
3. Provide an example of a real-world scenario that can be modeled using a normal distribution.
4. Explain the binomial distribution and give an example of its application.
5. Describe the Poisson distribution and its use cases.

Probability Density Function (PDF) and Cumulative Distribution Function (CDF)

1. What is a probability density function (PDF) and how is it used?
2. Explain the cumulative distribution function (CDF) and its significance.
3. How is the CDF related to the PDF?
4. For a continuous random variable, describe how you would use the PDF to find probabilities.
5. Given a normal distribution with mean 0 and standard deviation 1, calculate the CDF at (z = 1.96).

Probability Mass Function (PMF)

1. Define the probability mass function (PMF) and its properties.
2. How is the PMF used in the context of discrete random variables?
3. Provide an example of a PMF for a fair six-sided die.
4. Explain the relationship between the PMF and the cumulative distribution function (CDF) for discrete random variables.
5. Calculate the PMF for the number of heads in two coin tosses.

Practical Applications

1. Describe a scenario in data science where calculating the mean would be crucial.
2. Explain why understanding the variance is important when analyzing a dataset.
3. How can correlation analysis be used in predicting sales based on advertising spend?
4. Describe how skewness can affect the choice of statistical methods in data analysis.
5. Why is it important to consider kurtosis when analyzing financial returns?

Advanced Concepts

1. What is the significance of the central limit theorem in probability distributions?
2. How does the law of large numbers apply to empirical data analysis?
3. Explain the difference between a population and a sample in statistics.
4. What is the role of standard error in inferential statistics?
5. How would you interpret a z-score in the context of a normal distribution?

Visualization

1. Create a histogram for the dataset: [10, 15, 15, 20, 25, 30, 35, 40].
2. Draw a box plot for the dataset: [5, 6, 7, 8, 9, 10, 15, 20] and identify any outliers.
3. Explain how a scatter plot can be used to identify correlation between two variables.
4. What information can be obtained from a QQ plot?
5. Describe the use of a heatmap in visualizing correlation matrices.