STATISTICS WORKSHEET- 6

Answers:

1. d) All of the mentioned
2. a) Discrete
3. a) pdf (Probability Density Function)
4. c) mean
5. c) empirical mean
6. a) variance
7. c) 0 and 1
8. b) bootstrap
9. a) frequency

10. A boxplot displays the summary of a dataset's distribution while a histogram displays the distribution of a dataset by dividing it into equal intervals and plotting frequency density. Boxplots show median, quartiles, outliers, as histograms only show frequency.

11. To choose metrics, identify what you want to measure and what is important to the situation, then select metrics tht accurately reflect those things. Consider factors such as data availability, meaningfulness, and interpretability.

12. To determine if a finding is significant, you perform statistical tests to determine the possibility that the result was due to chance, rather than a real effect. If the probability of the result being due to chance is low, the result is considered significant.

13. Examples of data that do not have a Gaussian or log-normal distribution are:
    a) Categorical data: Data that only have a limited number of categories, such as color, gender, etc.
    b) Exponential data: Data that follow an exponential distribution, such as waiting time between events.
    c) Poisson data: Data that follows a Poisson distribution, such as the number of calls received by a callcenter in a given hour.
    d) Uniform data: Data that are evenly distributed between two values, such as exam scores between 0-100.

14. Suppose you have a data set with the following values: 1, 2, 5, 1000. The mean of this data set is 202.75, but it is heavily influenced by the large value of 1000. The median, however, is 2, which gives a better representation of the "typical" value in this data set. In this case, the median is a better measure than the mean.

15. The likelihood is a measure of how well a given set of parameters fits a particular dataset. It expresses the strength of evidence for the parameter values being true, given the data. It is often used to compare different models and choose the best one.