**Statistics Worksheet 6**

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

10. A histogram is a graphical representation of the distribution of numerical data while a boxplot displays the distribution of the data through their quartiles. The histogram shows the data distribution by dividing the data into bins, while the boxplot shows the median, quartiles, and outliers of the data.

11. Metrics selection depends on the specific problem and the goal of the analysis. A good metric should be relevant, measurable, specific, and timely. One can select metrics by considering the business objective, customer needs, available data, and the impact of the metric on the decision-making process.

12. Statistical significance is a measure of the probability that the observed results occurred by chance. To assess the statistical significance of an insight, one can perform a hypothesis test, which involves comparing the observed result to a null hypothesis, assuming that there is no difference between the groups being compared.

13. Some examples of data that do not have a Gaussian distribution or a log-normal distribution include power-law distributed data, Pareto distributed data, and exponential distribution.

14. The median is a better measure than the mean in data with extreme outliers or a skewed distribution. For example, in income data, where a few individuals have extremely high or low incomes, the median provides a better measure of central tendency than the mean.

15. Likelihood is a statistical concept that measures the probability of observing the data given a specific model or hypothesis. In other words, it measures how well the data fits the model or hypothesis. Likelihood is used in maximum likelihood estimation, a method used to estimate the parameters of a statistical model by finding the parameter values that maximize the likelihood of the observed data.