Statical worksheet = 6

1. (d) All of the mentioned

2. (a)Discrete

3. (a)pdf (Probability density function)

4. (c)Chi-squared distribution

5. (c)Mean

6. (a)Varience

7. (c)0 and 1

8. (b)Bootstrap

9. (b)Summarized

10. Histograms are bar charts that show the frequency of a numerical variable's values and are used to approximate the probability distribution of the given variable. Boxplot gather other information like the quartiles, the range, and outliers. Boxplots are especially useful when you want to compare multiple charts at the same time because they take up less space than histograms.

11. The metric(s) chosen to evaluate a machine learning model depends on various factors: a) To find a regression or classification task. b) To find the business objective like precision vs recall c) To know distribution of the target variable. There are a number of metrics that can be used, including adjusted r-squared, MAE, MSE, accuracy, recall, precision, f1 score, and the list goes on.

12. First, you would state the null hypothesis and alternative hypothesis. Second, you would calculate the p-value, the probability of obtaining the observed results of a test assuming that the null hypothesis is true. Last, you would set the level of the significance (alpha) and if the p-value is less than the alpha, you would reject the null — in other words, the result is statistically significant.
13. Any type of categorical data won't have a gaussian distribution or lognormal distribution. Exponential distributions — eg. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

14. When there are a number of outliers that positively or negatively skew the data.

15. The probability of some of the observed outcomes under specific parameter values is regarded as the likelihood of the set of parameter values under certain observed outcomes