STATISTICS WORKSHEET- 6

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

10.

- histogram displays frequencies for a group of data , box plot show frequency of individual data.
- In histogram we display our information with bars.box plot clearly displays information about the first quartile, median, and third quartile
- Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data set

11.
we need to understand what kind of problems we are trying to solve before appying metrics.Here is a list of some common problems in machine learning:

- Classification. This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.

- Regression. The algorithm will predict some values. For example, weather forecast for tomorrow.

- Ranking. The model will predict an order of items. For example, we have a student group and need to rank all the students depending on their height from the tallest to the shortest.

Depending on the problem we can choose the appropriate metrics

Classification Metrics

- Accuracy.
- Logarithmic Loss.
- ROC, AUC.
- Confusion Matrix.
- Classification Report.

Regression Metrics

- Mean Absolute Error.
- Mean Squared Error.
- Root Mean Squared Error.
- Root Mean Squared Logarithmic Error.
- R Square.

Ranking metrics

- basic metric
- Kendall's tau coefficient

12.

1. First, you would state the null hypothesis andalternative hypothesis.
2. Second, you would calculate the p-value,the probability of obtaining the observed results of a test assumingthat the null hypothesis is true.
3. Last, you would set the level of thesignificance and if the p-value is less than the alpha, you would reject the null, in other words, the result is statisticallysignificant.

13.

The simplest example is the distribution of numbers that show up on the top of a fair die after a large number of throws. Each number from 1 to 6 will occur with approximately equal frequency. Increasing the number of throws will not tend to produce a bell-shaped histogram, in fact the fractional occurrence will approach a constant 1/6 over the possible numbers.

14.

Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed. The median indicates that half of all incomes fall below and half are above it. For these data, the mean overestimates where most household incomes fall.

15.

- Likelihood refers to the process of determining the best data distribution given a specific situation in the data.
- likelihood determines whether the parameters in a model can be trusted based on the sample data you have observed.
- Likelihood function gives us an idea of how well the data summarizes the probability distribution function