**Worksheet\_Set\_4 3**

**Statistics**

. Which of the following can be considered as random variable?

Ans:- All of the Mentioned

. Which of the following random variable that take on only a countable number of possibilities?

Ans:-Discrete

Which of the following function is associated with a continuous random variable?

Ans:- Probability Density Function(PDF)

The expected value or \_\_\_\_\_\_\_ of a random variable is the center of its distribution

Ans:- Mean

Which of the following of a random variable is not a measure of spread?

Ans:-Variance

. The \_\_\_\_\_\_\_\_\_ of the Chi-squared distribution is twice the degrees of freedom.

Ans:-None of the mentioned

The beta distribution is the default prior for parameters between \_\_\_\_\_\_\_\_\_\_\_\_

Ans:-0 and 1

Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?

Ans:-Bootstrap

Data that summarize all observations in a category are called \_\_\_\_\_\_\_\_\_\_ data

Ans:-summarized

What is the difference between a boxplot and histogram?

Ans:- Boxplot- A boxplot is a graphical representation of statistical data based on the minimum, first quartile, median, third quartile, and maximum. It is used to display the distribution of a set of continuous or numerical data values.

Histogram- A histogram is a graphical representation of the distribution of a set of continuous or numerical data values. It displays the data values as bars, where the height of each bar represents the frequency of the data points falling within a certain range or bin.

How to select metrics?

Ans:- Selection metrics refer to the evaluation criteria used to choose the best model or solution from a set of candidate models or solutions in a machine learning or data analysis task. The selection metrics are chosen based on the specific requirements and objectives of the task, and they help to determine the most accurate, reliable, and efficient model among the candidates. Common selection metrics include accuracy, precision, recall, F1-score, AUC, etc.

How do you assess the statistical significance of an insight?

Ans:-hypothesis test can be performed. This involves formulating a null hypothesis and alternative hypothesis calculating a test statistic and its corresponding p-value, and comparing the p-value to a significance level. If the p-value is less than the significance level, you reject the null hypothesis and conclude that the insight is statistically significant. Otherwise, you fail to reject the null hypothesis and conclude that the insight may be due to chance.

Give examples of data that doesnot have a Gaussian distribution, nor log-normal.

Ans:- Categorical data,Poisson Distribution,Exponential Distribution,pareto distribution,skewed distribution.

Give an example where the median is a better measure than the mean.

Ans:- The median is often a better measure than the mean in the presence of outliers or extreme values.

consider the income distribution of a country. If a few people have extremely high incomes, the mean income of the country will be significantly higher than the median income, which is the middle value in the distribution. The mean will be influenced by the outliers, but the median will not.

What is the Likelihood?

Ans:- Likelihood is a statistical concept used to quantify the agreement between the observed data and a model. It is defined as the probability of observing a particular set of data given the parameters of a model.