STATISTICS WORKSHEET- 6

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following can be considered as random variable?
   a) The outcome from the roll of a die
   b) The outcome of flip of a coin
   c) The outcome of exam
   d) All of the mentioned
2. Which of the following random variable that take on only a countable number of possibilities?
   a) Discrete
   b) Non Discrete
   c) Continuous
   d) All of the mentioned
3. Which of the following function is associated with a continuous random variable?
   a) pdf
   b) pmv
   c) pmf
   d) all of the mentioned
4. The expected value or_____of a random variable is the center of its distribution.
   a) mode
   b) median
   c) mean
   d) bayesian inference
5. Which of the following of a random variable is not a measure of spread?
   a) variance
   b) standard deviation
   c) empirical mean
   d) all of the mentioned
6. The_____of the Chi-squared distribution is twice the degrees of freedom.
   a) variance
   b) standard deviation
   c) mode
   d) none of the mentioned
7. The beta distribution is the default prior for parameters between _____
   a) 0 and 10
   b) 1 and 2
   c) 0 and 1
   d) None of the mentioned
8. Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?
   a) baggyer
   b) bootstrap
   c) jacknife
   d) none of the mentioned

9. Data that summarize all observations in a category are called_____data.
   a) frequency
   b) summarized
   c) raw
   d) none of the mentioned

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What is the difference between a boxplot and histogram?

Box plot and Histogram can be used use Seaborn . where Box plot is Used using .boxplot Histogram is used using .countplot.

Box plot - gives the quartiles and indicate the median data to compare easily

Histogram - gives only the count

11. How to select metrics?

The model may give satisfying results when evaluated using a metric say accuracy_score but may give poor results when evaluated against other metrics such as logarithmic_loss or any other such metric. Hence, it is very much important to choose the right metric to evaluate the Machine Learning model.

Choice of metrics influences how the performance of machine learning algorithms is measured and compared. They influence how we weight the importance of different characteristics in the results.

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.
Adjusted R Square.

12. How do you assess the statistical significance of an insight?

Steps in Testing for Statistical Significance
1) State the Research Hypothesis
2) State the Null Hypothesis
3) Select a probability of error level (alpha level)
4) Select and compute the test for statistical significance
5) Interpret the results

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

    Exponential distributions do not have a log-normal distribution or a Gaussian distribution. In fact, any ty of data that is categorical will not have these distributions as well.

    Example: Duration of a phone car, time until the next earthquake, etc.

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

    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 27581, and half are above it. For these data, the mean overestimates where most household incomes fall.

15. What is the Likelihood?

    Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample. Let P(X; T) be the distribution of a random vector X, where T is the vector of parameters of the distribution.