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

   Correct option - d

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

   Correct option - a

3. Which of the following function is associated with a continuous random variable?
   a) pdf
   b) pmv
   c) pmf
   d) all of the mentioned

   Correct option - a

4. The expected value or_____of a random variable is the center of its distribution.
   a) mode
   b) median
   c) mean
   d) bayesian inference

   Correct option - c

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

   Correct option – c

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

   Correct option – a

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
   Correct option – c

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

   Correct option – b

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

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

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

    A histogram is a type of bar chart that graphically displays the frequencies of a data set.

    A box plot, also called a box-and-whisker plot, is a chart that graphically represents the five most important descriptive values for a data set. These values include the minimum value, the first quartile, the median, the third quartile, and the maximum value.

11. How to select metrics?
    - For regression: MSPE,MSAE, R-square , adjusted R-square
    - For classification: Precision-recall, ROC-AUC, accuracy, Log-loss
    - Unsupervised : Rand index, Mutual information

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

    Statistical significance of an insight can be assess through the hypothesis testing.

13. Give examples of data that does not have a Gaussian distribution, nor log-normal.

    Poisson distribution, found with rare events such as number of accidents

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

    If data is skewed or it have outliners it’s better to use median than the mean.

15. What is the Likelihood?

    The likelihood measures the goodness of fit of a statistical model to a sample of data for given values of the unknown parameters