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?

    when there is moderate variation among the observed frequencies, which causes the histogram to look ragged and non-symmetrical due to the way the data is grouped. This may lead one to assume the data is slightly skewed. However, when a box plot is used to graph the same data points, the chart indicates a perfect normal distribution

11. How to select metrics?
    Metrics will be selected based on which model we are using like for classification model there are different metrics available(and for regression model metrics will be different.

12. How do you assess the statistical significance of an insight?
    Statistical significance is often calculated with statistical hypothesis testing, which tests the validity of a hypothesis by figuring out the probability that your results have happened by chance.

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.
    Any distribution of money or value will be non--Gaussian. For example: distributions of income; distributions of house prices; distributions of bets placed on a sporting event. These distributions cannot have negative values and will usually have extended right hand tails

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

    Median will be better where distribution tends to be skewed,like income.

15. What is the Likelihood?
    The term Likelihood refers to the process of determining the best data distribution given a specific   situation in the data.