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
   ANSWER 1 = 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
   ANSWER 2 = A) Discrete

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

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

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
   ASNWER 5 = C) empirical mean

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
   ANSWER 6 = A) variance

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
   ASNWER 7 = C) 0 and 1

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
   ANSWER 8 = B) bootstrap

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

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

10. What is the difference between a boxplot and histogram?
    ANSWER 10 = Histogram are used to check the distribution of the data; this is formed using the probability density functions. Where box plot also shows the distribution of data however it is mostly used to detect the outliers, it shows the quantile range of the data. For data distribution histogram is more desirable. Box plots allow you to compare multiple data sets better than histograms as they are less detailed and take up less space.

11. How to select metrics?
    ANSWER 11 = Depending upon the business problem or model that we use we select the metrics, For example for classification model we use confusion matrix, accuracy score, precision, f1, recall, auc as performance metrics, while for regression we use R2, MAE, MSE, RMSE for measuring performance.

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

ANSWER 12 = By using AB testing we can access the statistical significance of an insight.

1. Create a null hypothesis.
2. Create an alternative hypothesis.
3. Determine the significance level.
4. Decide on the type of test you'll use.
5. Perform a power analysis to find out your sample size.
6. Calculate the standard deviation.
7. Use the standard error formula.
8. Determine the t-score.
9. Find the degrees of freedom.
10. Use a t-table.

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.
    ANSWER 13 = Distribution like Pareto and Possion can be different from normal and log normal distribution.
    Examples of pareto distributions are -The sizes of human settlements (few cities, many hamlets/villages), The length distribution in jobs assigned to supercomputers (a few large ones, many small ones), Amount of time a user on Steam will spend playing different games.

    Example of poisson distributions are - telephone calls arriving in a system, Internet traffic, number of losses or claims occurring in a given period of time.

14. Give an example where the median is a better measure than the mean.
    ANSWER 14 = When there are many outliers in the data set than median is better than mean as mean gets highly affected by outliers.

15. What is the Likelihood?
    ANSWER 15 = Likelihood refers to the probability of something happening given some constraints, It is widely used in conditional probability (Bayes Theorem)