ASSIGNMENT-4

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

   Ans:- 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.

   Ans:- 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

   Ans:- pdf

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

   Ans:- 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

   Ans:- Variance

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

   Ans:- 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

Ans:- 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

   Ans:- bootstrap

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

   Ans:- 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?

    Ans:- Histograms are a special kind of bar graph that shows a bar for a range of data values instead of a single value. A box plot is a data display that draws a box over a number line to show the interquartile range of the data. The 'whiskers' of a box plot show the least and greatest values in the data set.

11. How to select metrics?

    Ans:- Metrics are used to monitor and measure the performance of a model(during training and testing),and don.t need to be differentiable

    Regression Metrics:- (1) Mean Absolute Error(MAE)
    (2) Mean squared Error(MSE)
    (3) Root Mean Squared Error (RMSE)
    (4) R-Squared
    Classification Metrics:- (1) Accuracy
    (2) Confusion Metrics (not a metric but fundamental to others)
    (3) Precision and recall
    (4) F-1 score
    (5) AU-ROC

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

    Ans :- To assess statistical significance, you would use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Second, you'd calculate the p-value, which is the likelihood of getting the test's observed findings if the null hypothesis is true. Finally, you would select the threshold of significance (alpha) and reject the null hypothesis if the p-value is smaller than the alpha — in other words, the result is statistically significant.

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

    Ans:- Exponential distributions do not have a log-normal distribution or a Gaussian distribution. In

fact, any type 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.

    Ans:- Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

15. What is the Likelihood?

    Ans:- Likelihood is not a probability, but is proportional to a probability; the two terms can't be used interchangeably. In this post, we will be dissecting likelihood as a concept and understand it's importance in machine learning.