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: 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
   Ans: 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
   Ans: 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
   Ans: 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
   Ans: a

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: 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
   Ans: 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
   Ans: b

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

10. What is the difference between a boxplot and histogram?
    Ans: Boxplots may also depict values that are far outside of the normal range of responses (referred to as outliers). A histogram is a graphical representation of the spread of data points.

11. How to select metrics?
    Ans: Metrics like accuracy, precision, recall are good ways to evaluate classification models for balanced datasets, but if the data is imbalanced then other methods like ROC/AUC perform better in evaluating the model performance.

12. How do you assess the statistical significance of an insight?
    Ans: Researchers use a measurement known as the p-value to determine statistical significance: if the p-value falls below the significance level, then the result is statistically significant. The p-value is a function of the means and standard deviations of the data samples.

13. Give examples of data that doesnot 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: The likelihood is the probability that a particular outcome is observed when the true value of the parameter is , equivalent to the probability mass on ; it is not a probability density over the parameter . The likelihood, , should not be confused with , which is the posterior probability of given the data .