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?

Ans. d) All of the mentioned

2. Which of the following random variable that take on only a countable number of possibilities?

Ans. a) Discrete

3. Which of the following function is associated with a continuous random variable?

Ans. a) pdf

4. The expected value or _____ of a random variable is the center of its distribution.

Ans. C) mean

5. Which of the following of a random variable is not a measure of spread?

Ans. A) variance

6. The _____ of the Chi-squared distribution is twice the degrees of freedom.

Ans. b) standard deviation

7. The beta distribution is the default prior for parameters between _____

Ans. c) 0 and 1

8. Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?

Ans. b) bootstrap

9. Data that summarize all observations in a category are called _____ data.

Ans. 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?

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. KEY STEPS TO SELECTING EVALUATION METRICS

Classification. This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.

Regression. The algorithm will predict some values. For example, weather forecast for tomorrow.

Ranking. The model will predict an order of items.

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

Ans. Steps in Testing for Statistical Significance

State the Research Hypothesis.
State the Null Hypothesis.
Select a probability of error level (alpha level)
Select and compute the test for statistical significance.
Interpret the results.

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. The mean is used for normal distributions. The median is generally used for skewed distributions. The mean is not a robust tool since it is largely influenced by outliers. The median is better suited for skewed

distributions to derive at central tendency since it is much more robust and sensible.

15. What is the Likelihood?

Ans. **The likelihood function (often simply called the likelihood) represents the probability of random variable realizations conditional on particular values of the statistical parameters. Thus, when evaluated on a given sample, the likelihood function indicates which parameter values are more likely than others, in the sense that they would have made the observed data more probable**