STATISTICS WORKSHEET- 4

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?

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.

11. How to select metrics?

Ans:- KEY STEPS TO SELECTING EVALUATION METRICS

1. **Classification.**
   This algorithm will predict data type from defined data arrays. For example, it may respond with yes/no/not sure.
2. **Regression.**
   The algorithm will predict some values. For example, weather forecast for tomorrow.
3. **Ranking.**
   The model will predict an order of items.

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

Ans: - 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 does not have a Gaussian distribution, nor log-normal.

Ans: - Any type of categorical data won't have a gaussian distribution or lognormal distribution. Exponential distributions - eg. the amount of time that a car battery lasts or the amount of time until an earthquake occurs.

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 a strange concept in that it is not a probability but is proportional to a probability.