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
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?

    Answer: 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?

    Answer: Step 1 Why is the measurement required?
    Step 2 What needs to be measured?
    Step 3 What is the precision of measurement required?
    Step 4 How will it be measured?
    Step 5 What use will the measurement be put to? By whom?

12. How do you assess the statistical significance of an insight?
    Answer: 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.
    Answer: Any distribution of money or value will be non-Gaussian. For example: distributions of income; distributions of house prices; distributions of bets placed on a sporting event. These distributions cannot have negative values and will usually have extended right hand tails.
    Non-normal distributions may lack symmetry, may have extreme values, or may have a flatter or steeper “dome” than a typical bell. There is nothing inherently wrong with non-normal data; some traits simply do not follow a bell curve. For example, data about coffee and alcohol consumption are rarely bell shaped.

14. Give an example where the median is a better measure than the mean.
    Answer: The median better represents the central tendency for the skewed distribution. These data are based on the U.S. household income for 2006. 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?
    Answer: The term Likelihood refers to the process of determining the best data distribution given a specific situation in the data. When calculating the probability of a given outcome, you assume the model's parameters are reliable.