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) 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 – a) 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 - a) 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 – c) 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 – a) 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 – a) 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 fordifficult statistics?
   a) baggyer
   b) bootstrap
   c) jackknife
   d) none of the mentioned

Ans – b) 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 – b) summarized

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

Q-10.What is the difference between a boxplot and histogram?

Ans - Histograms and box plots are very similar in that they both help to visualize and describe numeric data. Although histograms are better in determining the underlying distribution of the data, box plots allow you to compare multiple data sets better than histograms as they are less detailed and take up less space.

Q-11.How to select metrics?

Ans – To select metrics, we need to understand what the problem is we are trying to solve, what are the questions we need to answer, and what are the goals we want to achieve. Then we need to identify the metrics that are directly related to solving the problem and answering the questions. After that we will need to evaluate the metrics based on their relevance, accuracy, and ease of interpretation.

Q-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.

Q-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.

Q-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. The median indicates that half of all incomes fall below 27581, and half are above it. For these data, the mean overestimates where most household incomes fall.

Q-15 . What is the Likelihood?

Ans- Likelihood, being the outcome of a likelihood function thus defined, describes the plausibility, under a certain statistical model (the null hypothesis in hypothesis testing), of a certain parameter value after observing a particular outcome.