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
   Answer: (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
   Answer: (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
   Answer: (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
   Answer: (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
   Answer: (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
   Answer: (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
   Answer: (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
   Answer: (b)

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

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

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

    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?
    Vision for the product: How do you envision the product being used, and how will that change over time? Do you anticipate weekly product usage? If so, WAU would be the best metric to drive success. But if you also anticipate that people will transition from desktop usage to mobile usage, you might want to consider selecting DAU as your goal metric since mobile platforms tend to increase usage frequency.

    Current product usage: You might have a vision for your product, but how are people actually using your product right now? If people are using your product multiple times a day, then it makes sense to select DAU or an intraday metric (number of sessions or hourly active users) your most important metric.

    The competition: How are competitors' products being used? Do people use competitors' products on a daily basis but your product is used on a weekly basis? If so, use DAU as the metric because your product will not succeed in the competitive environment as people are engaging it with less. Setting a daily usage goal may force product decisions that you otherwise may not have done.

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

15. What is the Likelihood?
    Likelihood that any parameter (or a of a set of parameters) should have any assigned value (or set of values) is proportional to the probability that if this were so, the totality of observations should be that observed


FLIP ROBO