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 is 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

Answer is 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

Answer is 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

Answer is 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

Answer is 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

Answer is 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

Answer is c) 0 and 1

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 is 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

Answer is b) summarized data

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 and box plots are graphical representations for the frequency of numeric data values. They aim to describe the data and explore the central tendency and variability before using advanced statistical analysis techniques.

Both histograms and box plots allow to visually assess the central tendency, the amount of variation in the data as well as the presence of gaps, outliers or unusual data points.

Both histograms and box plots are used to explore and present the data in an easy and understandable manner. Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.

11. How to select metrics?

These are the main way to select metrics to improve performance.

Define primary objective
Determine cause and effect
Create relevant activities
Evaluate periodically

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

To assess statistical significance, We need to use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Second, we need calculate the p-value, which is the likelihood of getting the test's observed findings if the null hypothesis is true. Finally, we need to select the threshold of significance (alpha) and reject the null hypothesis if the p-value is smaller than the alpha — in other words, the result is statistically significant.

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.

Non Gaussian distributed time series data arise when the mean or noise statistics vary with time.

If the mean varies with time, the variable could be non-stationary / time-varying (its trend changes with time), auto- or cross-correlated (it changes depending on its previous value or the values of other variables), or its value is computed from the values of other Gaussian variables but in a nonlinear way.
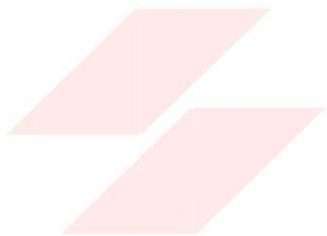
14. Give an example where the median is a better measure than the mean.

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.
The median is less affected by outliers and skewed data than the mean, and is usually the preferred measure of central tendency when the distribution is not symmetrical.

15. What is the Likelihood?

Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample.