FLIP ROBO

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:- (C)

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:- (A)

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

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

    A boxplot and histogram are two different ways to represent and visualize the distribution of data.

    A boxplot provides a summary of a set of data values by showing the median, quartiles (25th and 75th percentiles), and outliers (values outside of 1.5 times the interquartile range), all in a single diagram.

    A histogram, on the other hand, is a graphical representation of the distribution of data using bars of different heights. The height of each bar represents the frequency of the data values falling within a certain range or bin.

    So, the main difference between a boxplot and histogram is that a boxplot summarizes the data by showing its central tendency and variability, while a histogram shows the distribution of the data through frequency of data values falling into different bins.

11. How to select metrics?
    To select metrics, the following steps can be considered:

    Define the goals and objectives of your analysis.
    Identify the relevant data sources.
    Choose the most appropriate measurement scale (e.g. nominal, ordinal, interval, ratio).
    Decide which statistical methods will best measure the desired metric(s).
    Decide which visualizations best represent the data and the desired metric(s).
    Validate the metrics by checking the data for outliers, errors and missing values.
    Evaluate the results and refine the metrics as necessary.

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

    To assess the statistical significance of an insight, you need to perform a statistical test that measures the likelihood of observing a result given a certain assumption. You can use hypothesis tests such as t-tests, ANOVA, chi-squared tests, etc. to determine if the result is significant or if it could have occurred by chance. Another approach is to calculate a confidence interval for a metric, which gives an estimate of the range of values that the metric is likely to take given a certain level of confidence. If the confidence interval does not include zero or a null value, then it can be considered statistically significant

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.
    Exponential Distribution
    Poisson Distribution
    Uniform Distribution
    Cauchy Distribution
    Bernoulli Distribution
    Skewed Distributions (e.g. gamma, beta, etc.)
    Multinomial Distribution
    Heavy-tailed Distributions (e.g. Pareto, Lévy, etc.)
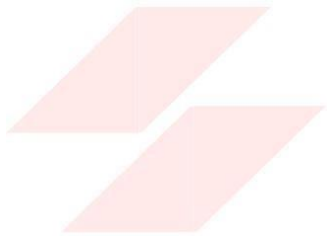   Mixture Distributions (e.g. Gaussian Mixture Model, etc.)

14. Give an example where the median is a better measure than the mean.
    Data with Outliers: If a dataset contains extreme values, called outliers, the mean can be heavily influenced by them, while the median remains robust and less sensitive to outliers.
    For example, consider salary data of a company where most employees earn a moderate salary, but a few executives have very high salaries. The mean salary would be skewed high due to the high salaries of executives, while the median salary would give a better representation of the typical salary earned by employees.

15. What is the Likelihood?

    Likelihood is a statistical concept that refers to the probability of observing a particular set of data given a set of parameters and a specific model. It provides a way to quantify how well a given model fits the observed data. The likelihood function is often used in estimation, hypothesis testing, and model selection, among other statistical applications. The likelihood is a scalar value, with higher values indicating a better fit between the model and the data.