STATISTICS

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: a) The outcome from the roll of a die, b) The outcome of flip of a coin are the examples of random variable.

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) 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: c) 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: 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: c) empirical mean

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) 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: 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 above

Answer: 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: b) summarized

Q10 to Q15 are subjective answer type questions, Answer them briefly.

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

Answer: A boxplot and histogram are both graphical representations of data. A boxplot is a summary of a set of data containing the minimum, first quartile, median, third quartile, and maximum. A histogram is a graphical representation of the distribution of numerical data. It shows the number of observations within each given interval or bin

11. How to select metrics?

Answer: 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.

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

Answer: To assess the statistical significance of an insight, one can use statistical hypothesis testing. This involves formulating a null

hypothesis and an alternative hypothesis and then using a test statistic and p-value to determine whether the insight is statistically significant or not.

13. Give examples of data that does not have a Gaussian distribution, nor log-normal?

Answer: Examples of data that does not have a Gaussian distribution, nor log-normal are:

- **Exponential Distribution**
- **Poisson Distribution**
- **Bernoulli Distribution**
- **Uniform Distribution**

14. Give an example where the median is a better measure than the mean?

Answer: An example where median is a better measure than mean is when the data has outliers. In such cases, the mean can be skewed by the outliers, while the median is not affected by them.

15. What is the Likelihood?

Answer: Likelihood is a way to measure how well a set of parameters fits a given set of observations. It is a function of the parameters of the model and is used to estimate the parameters that are most likely to have generated the observed data. The higher the likelihood, the more likely the observed data is to have been generated by the given set of parameters.

ASSIGNMENT - 6