# 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
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
3. Which of the following function is associated with a continuous random variable?
   a) pdf
   b) pmv
   c) pmf
   d) all of the mentioned
4. The expected value or_____of a random variable is the center of its distribution.
   a) mode
   b) median
   c) mean
   d) bayesian inference
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
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
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
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

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

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

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

Ans: A boxplot and a histogram are both used to visually represent data, but they have different purposes. A histogram is a graphical representation of the distribution of a dataset, showing the number of observations that fall within each given range of values (bins). It is used to show the distribution of continuous data. On the other hand, a boxplot is a standardized way of displaying the distribution of data based on five number summary ("minimum", first quartile (Q1), median, third quartile (Q3), and "maximum"). It is used to show the distribution of continuous or discrete data.

11. How to select metrics?

Ans: To select metrics, it is important to consider the goals and objectives of the analysis, as well as the type and structure of the data. One should start by identifying the key variables and factors of interest, and then determining which metrics will best measure and describe these variables. Additionally, it is important to consider the level of measurement (nominal, ordinal, interval, ratio) and the type of data (continuous, categorical, count) when selecting metrics.

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

Ans: To assess the statistical significance of an insight, one can use statistical hypothesis testing, specifically p-value. The p-value is a measure of the evidence against a null hypothesis. A small p-value (typically $\leq 0.05$) indicates strong evidence against the null hypothesis, so you reject the null hypothesis and conclude that your sample provides enough evidence that the population parameter differs from the value specified in the null hypothesis.

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

Ans: Examples of data that doesn't have a Gaussian distribution, nor log-normal are: the data following a logistic distribution, the data following a Poisson distribution, the data following a binomial distribution.

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

Ans: An example where median is a better measure than mean is when there are outliers present in the data. Outliers are extreme values that deviate significantly from the rest of the observations, and they can greatly affect the mean, but not the median. Therefore, when outliers are present, the median gives a better representation of the central tendency of the data.

15. What is the Likelihood?

Ans: Likelihood is a function that describes the probability of obtaining a certain set of data given a set of parameters. It is used in statistical modeling to estimate the probability of a model given a set of observations. The likelihood function is used to update the estimates of the parameters in the model through an optimization process, such as maximum likelihood estimation (MLE).


FLIP ROBO