STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
   a) True
   b) False
2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?
   a) Central Limit Theorem
   b) Central Mean Theorem
   c) Centroid Limit Theorem
   d) All of the mentioned
3. Which of the following is incorrect with respect to use of Poisson distribution?
   a) Modeling event/time data
   b) Modeling bounded count data
   c) Modeling contingency tables
   d) All of the mentioned
4. Point out the correct statement.
   a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
   b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   c) The square of a standard normal random variable follows what is called chi-squared distribution
   d) All of the mentioned
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
6. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
7. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned
8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   a) 0
   b) 5
   c) 1
   d) 10
9. Which of the following statement is incorrect with respect to outliers?
   a) Outliers can have varying degrees of influence
   b) Outliers can be the result of spurious or real processes
   c) Outliers cannot conform to the regression relationship
   d) None of the mentioned

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

10. What do you understand by the term Normal Distribution?

- Ans: Generally all the processes (like weight, height, rainfall, temperature,…) which are governed by nature follow normal distribution.
- Many distributions can be approximated to normal distribution.
- When Exploratory data Analysis is used then using the distribution graph one can guess whether data follows normality or not by observing skewness and kurtosis
- For regression analysis normality if observed is required otherwise nonparametric tests are to be used (Specially when data is small)

11. How do you handle missing data? What imputation techniques do you recommend?
- If missing values are quantitative then those values can be replaced by arithmetic mean using fillna() or impute()
- If missing values are qualitative then those values can be replaced by median using fillna() or impute()
- In advanced stage one can replace such values by estimated values.
- Sometimes depending on situation if we get intuition that missing data is not much contributive then one can discard those rows/ records.

12. What is A/B testing?

13. Is mean imputation of missing data acceptable practice?
- It depends on situation

14. What is linear regression in statistics?
- For bivariate data if we observe that there is the correlation between the variables then if one wants to estimate the variable on the basis of other variables then regression techniques is used.
- If model used for regression (Prediction) is the linear model then it is called linear regression

15. What are the various branches of statistics?
- Inferential Statistics
- Hypothesis testing
- Design and Analysis of experiment
- Time series analysis
- Bayesian inference