FLIP ROBO

STATISTICS WORKSHEET-1

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

Answers – 1- a), 2- a), 3- b), 4- d), 5- c), 6- b), 7- b), 8- a) 9- c)

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. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False

7. 1. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned

8. 4. 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 subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?
    Ans- **Normal distribution**, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

11. How do you handle missing data? What imputation techniques do you recommend?
    Ans- Real-world data is messy and usually holds a lot of missing values. Missing data can skew anything for data scientists and, A data scientist doesn't want to design biased estimates that point to invalid results. Behind, any analysis is only as great as the data. Missing data appear when no value is available in one or more variables of an individual. Due to Missing data, the statistical power of the analysis can reduce, which can impact the validity of the results.
    Basic Imputation Techniques:
    - Imputation with a constant value
    - Imputation using the statistics (mean, median, mode)
    - K_Nearest Neighbor Imputation

12. What is A/B testing?
    Ans- A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

13. Is mean imputation of missing data acceptable practice?
    Ans- Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?
    Ans- **Linear regression** quantifies the relationship between one or more predictor variable(s) and one outcome variable. Linear regression is commonly used for predictive analysis and modeling.

15. What are the various branches of statistics? **Ans**- There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.

Ans-