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
   Answer- true
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
   Answer- central limit theorm
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
   Answer-B

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
   Answer - D

5. _____ random variables are used to model rates. a) Empirical b) Binomial c) Poisson d) All of the mentioned
   Answer- poisson
6. Usually replacing the standard error by its estimated value does change the CLT. a) True b) False
   Answer- 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
   Answer- B
8. Normalized data are cantered at_____and have units equal to standard deviations of the original data. a) 0 b) 5 c) 1 d) 10
   Answer- A
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
   Answer- D

WORKSHEET

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

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

Answer- A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution.

11. How do you handle missing data? What imputation techniques do you recommend?

Answer- Handling missing data is an important step in data analysis to ensure accurate and meaningful results. There are several approaches to handle missing data, and the choice of technique depends on the nature and extent of missingness, as well as the specific analysis objectives. Here are some commonly used techniques for handling missing data:

1. Complete Case Analysis (CCA): This approach involves excluding any observations with missing data from the analysis. It is the simplest method but can lead to biased results if the missingness is not random.
2. Pairwise Deletion: In this approach, missing values are ignored on a pairwise basis, allowing different observations to contribute to different parts of the analysis. However, this method can also lead to biased results and loss of information if the missingness is not random.
3. Mean/Mode Imputation: In this technique, missing values in a variable are replaced with the mean (for numeric variables) or mode (for categorical variables) of the available data. It is a straightforward approach but can lead to underestimation of variances and distortions in relationships.
4. Multiple Imputation: Multiple imputation generates multiple plausible values for missing data based on the observed data and the assumption of missingness. It creates multiple complete datasets with imputed values, and the analysis is performed on each dataset. The results are then combined to obtain the final estimates and appropriate standard errors.
5. Model-Based Imputation: Model-based imputation involves creating a statistical model to estimate missing values based on observed data. The model can be based on regression, propensity scores, or other statistical techniques. The missing values are then imputed using the estimated model.
6. Hot-Deck Imputation: Hot-deck imputation involves randomly assigning missing values from observed data values in similar cases. It can be used when there is reason to believe that missing values are related to other observed values in the dataset.

The choice of imputation technique depends on factors such as the missing data pattern, the underlying assumptions, and the specific analysis objectives. It is recommended to carefully assess the missing data mechanism, evaluate the potential biases introduced by different imputation methods, and conduct sensitivity analyses to understand the impact of missingness on the results.

Additionally, it is important to note that imputation should be used judiciously and the limitations and potential biases associated with imputed data should be acknowledged and reported in any data analysis or research study

12. What is A/B testing?

Answer- also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics. Essentially, A/B testing eliminates all the

guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to 'control' or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable.

The version that moves your business metric(s) in the positive direction is known as the 'winner.' Implementing the changes of this winning variation on your tested page(s) / element(s) can help optimize your website and increase business ROI.

The metrics for conversion are unique to each website. For instance, in the case of eCommerce, it may be the sale of the products. Meanwhile, for B2B, it may be the generation of qualified leads.

A/B testing is one of the components of the overarching process of Conversion Rate Optimization (CRO), using which you can gather both qualitative and quantitative user insights. You can further use this collected data to understand user behaviour, engagement rate, pain points, and even satisfaction with website features, including new features, revamped page sections, etc. If you're not A/B testing your website, you're surely losing out on a lot of potential business revenue.

13. Is mean imputation of missing data acceptable practice?

    Answer- Mean imputation is typically considered terrible practice since it ignores feature correlation

14. What is linear regression in statistics?

    Answer- Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?
    Answer- Two branches, descriptive statistics and inferential statistics, comprise the field of statistics