Q1. Bernoulli random variables take (only) the values 1 and 0.

a) True b) False

Ans1. A) True.

Q2. 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

Ans2. A) Central Mean Theorem

Q3. 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

Ans3. B) Modeling bounded count data

Q4. 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

Ans4. D) All of the mentioned

Q5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Ans5. C) Poisson.

Q6. Usually replacing the standard error by its estimated value does change the CLT.

a) True
b) False

Ans6. b) False.

Q7. Which of the following testing is concerned with making decisions using data?
a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned

Ans7. b) Hypothesis.

Q8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0
b) 5
c) 1
d) 10

Ans8. 0

Q9. 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

Ans9. C

Q10. What do you understand by the term Normal Distribution?

Ans10. In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape. The mean, median and mode are exactly the same. The distribution is symmetric about the mean—half the values fall below the mean and half above the mean. The distribution can be described by two values: the mean and the standard deviation.

Q11. How do you handle missing data? What imputation techniques do you recommend?

Ans11. There are several ways to handle missing data issue in a dataset, to name a few would be.

1. **Drop the missing Values-**
   - 1.1 **Drop the Variable data-** We need to evaluate if we can drop the (particular column) data which has missing value and if it doesn't impact much on the analysis results then we can decide to drop the variable.
   - 1.2 **Drop the Data entry -** if the data doesn't have much observations we can decide to drop the data entry in data set.

2. **Replace the missing values-**
   - **2.1 Replace it with an average value- we can replace the missing values with average value of similar datapoints.**
   - **2.2 Replace it by Frequency- For data type where average method might not work then we replace the data by most common value.**

The replacement of data in place of missing data would be my recommendation.

Q12. What is A/B testing?

Ans.12 A/B testing in its simplest sense is an experiment on two variants to see which performs better based on a given metric. Typically, two consumer groups are exposed to two different versions of the same thing to see if there is a significant difference in metrics like sessions, click-through rate, and/or conversions.

Q13. Is mean imputation of missing data acceptable practice?

Ans.153The process of replacing null values in a data collection with the data's mean is known as mean imputation.

Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate, and the confidence interval is narrower.

Q14. What is linear regression in statistics?

Ans.14 Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable

Q15. What are the various branches of statistics?

Ans.15 The two main branches of statistics are descriptive statistics and inferential statistics.

Descriptive Statistics - The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.

Inferential Statistics- The branch of statistics that analyzes sample data to draw conclusions about a population.