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

ans:- a) 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

ans:- a) Central Limit Theorem

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

ans:- b) Modeling bounded count data

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

ans:- d) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

ans:- c) Poisson

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

a) True

b) False

ans:- 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

ans:- b) Hypothesis

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

ans:- a) 0

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

ans:- c) Outliers cannot conform to the regression relationship

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:-A normal distribution is a probability distribution that shows the data points are centered or spread out at around the mean or central value. This means that the distribution has more data around the mean. The data distribution decreases as you move away from the center. The resulting curve is symmetrical about the mean and forms a bell-shaped distribution.

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

Ans:- there are many ways to handle the missing values present in any data.some ways are like predicting the values formissing values by already provided values with the help of mean , median or sime of by extreme value method also some by arbitrary method and also by regression line etc. this means we just predict the value for missing with the help of data values already provided or present in the data.

The imputation techniques are:

1-mean,median,mode replacement

2-random sample imputation

3-capturing nan value with feature

4-end of distribution imputation

5-arbitrary imputation

6-frequent categories imputation

12. What is A/B testing?

Ans:- A/B testing also called split testing or bucket testing which compares the performance of two versions of content like app or webpage to see which one appeals more to visitors/viewers and to identify the better performance. It tests a control (A) version against a variant (B) version to measure which one is most successful based on your key metrics.also it's a method to helps you to make decision based on the real data rather than just guessing.

13. Is mean imputation of missing data acceptable practice?

Ans:-mean imputation have certain good sides like very easy to use to fill the missing values present in data by taking out the mean value of value present in data and fill those mean value in the missing value column.but there are also some drawbacks of it. Mean is not a good option always.it may lost the realation between the variables.

14. What is linear regression in statistics?

Ans:- linear regression in statistics is a statistical analysis technique used to predict the value of unknown varable using known variable with the helps of linear equation is known as linear regression.

Equation- Y=aX+b, where y is dependent variable ,a is the slope , X is independent variable and b is the y-intercept.

15. What are the various branches of statistics?

Ans:- the two main branches of statistics are

1-inferential statistics

2-descriptive statistics-a)measure of central tendancy b)measure of variability

Also there are more branches id decriptive stat.