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

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

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

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 – b

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Ans – c

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

a) True

b) False

Ans – b

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

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

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

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 random variable (x) is said to have a normal distribution with mean (μ) and variance (σ^2). If the random variable follows the properties of normal distribution it has three properties, that properties is called empirical formulas.

1. The probability of variable, that falls within the range of ($\mu - \sigma$) and ($\mu + \sigma$) that is the range of 1st standard deviation will approximately 68%. It means 68% of data points belonging to the random variable x, falls within the range of 1st standard deviation.
2. 95% of data points belonging to the random variable (x) falls within the range of 2nd standard deviation.
3. The 99.7% of data points belonging to the random variable (x) lies within the range of 3rd standard deviation.

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

Ans – As we know that, the machine learning models are not good with missing data. So, for this we handle the missing data in a such way that, our model will give good result on testing data. If you have missing data then we have two way,

1. Simply drop that column or row whose have a missing data.
2. Fill some appropriate value in to missing data.

For filling the data, we used "Imputation Technique". To impute or fill them, we have two ways

a. Univariate   b. Multivariate

Univariate: - In univariate we focus on one column at a time. For this, we used **"Simple Imputer Technique"**. We fill the data depending upon data type whether the data is continuous or categorical. For numeric we used mean, median method and for categorical mode method used.

Multivariate: - For multivariate we have two technique,

1. KNN Imputer
2. Iterative Imputer.

12. What is A/B testing?

Ans – It is also called as split testing. A to B testing is basically used to compare two product and those comparison is basically made user input. A to B testing is heavily used in e-commerce sector where you have different product, that are dynamically popping Infront of you.

Example – Netflix movies recommendation, in which you will get different recommendation of movies. One thing that you have notice there, there thumbnail also different.

A to B is used to comparing the version of web page, email, marketing sites.

Process of A to B testing: -

1. Collecting the data.
2. Identifying our goals
3. Create variations
4. Run the Experiments
5. Analysis results

13. Is mean imputation of missing data acceptable practice?

Ans –No, Bad Practice: In mean imputation we take a mean of data and assign this mean to the missing data. But practice of this, it is only feasible to in some condition. By this our model performance will poor because mean imputation is **ignoring the "feature correlation".**

Mean imputation decreases the variance of our data while increasing bias. As result of the reduced variance, model less accurate.

Mean imputation reduced the standard error, which invalidates most hypothesis tests and the calculation of confidence interval.

14. What is linear regression in statistics?

Ans – You have to want to establish relationship between independent variable and dependent variable best thing to do regression and most prominently you will used this where depending variable is continuous in nature.

Linear means, if you increase the values of "x" then value of "y" also increases that relationship is called linear relationship and we solve such problems via linear regression.

Linear regression is the process of finding a line that best fit the data points available, so that we can use it to predict output. For this we have two types of linear regression.

1. Simple Linear Regression

2. Multiple Linear Regression

15. What are the various branches of statistics?

Ans – Statistics is the science of collecting, organizing, analysis data and doing so for better decision making. There is a two type of statistic

    1.Descriptive statistics: - You take the population or sample data and explore the particular data, visualizing data in the form of numbers.

Example – Bar plot, histogram, PDF, CDF, normal distribution, measure of central tendency.

2.Inferential Statistics: - We take sample from population and try to do some test, then we come up with inference and conclusion for that population.

It also called as confident interval.

Example: - Z-test, T-test, Hypothesis testing