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

a) True

b) 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

8. 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

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

Ans:

Normal distribution means we are having good quality data with low variance symmetrical distributed Mean is 0 ,no outliers ,frequent data points ,data is not much deviated from mean

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

Ans:

It depends on data and domain knowledge of business ,

If data is continuous and not that much sensitive means wrong value not going to affect then as per data, I will take one of the following ways

Replace with mean

Replace with prediction by algorithms as assigning missing data as label

If data is sensitive

Like car data is there and brand is missing ,consider actual brand is Honda and by any technic we fill as tata so it will not be going to make any sense cause tata is not having the model Honda have or other characteristics

Instead of writing any similar or suitable I will prefer to delete missing data records

12. What is A/B testing?

Ans:

A/B testing is basically performed to make hypothesis on basis of two data set is there any statistical relationship between them.

13. Is mean imputation of missing data acceptable practice?

Ans:

No, its not a good practise . It reduces the accuracy of model. In some case it looks too much weird value not making any sense.

14. What is linear regression in statistics?

Ans:

It is a linear predictor function which predict the target value based on other values

y=mx+c

where

y=predicted value

m=coefficient or slope

x=data points

c=intercept

15. What are the various branches of statistics

Ans:

Descriptive Statistics

Inferential Statistics

Prescriptive Statistics

Predictive Statistics