Statistic

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 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

Answer : 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

Answer : All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical b) Binomial c) Poisson d) All of the mentioned

Answer : Poisson

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

a) True b) False

Answer : True

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

Answer : 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

Answer : 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

Answer : Outliers cannot conform to the regression relationship

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

Answer : Normal distribution is also known as Gaussian distribution, it is a probability distribution that is symmetric about the mean, Showing that data near mean are more frequent in occurrence than data far from the mean. In graphical form it appear like a bell curve. In a normal distribution the mesn is 0 and the standard devition is 1.

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

Answer : Missing data appear when no value is available in one or more variables of an individual. We can handle missing data by deletions and basic imputation techniques. K-Nearest Neighbour Imputation is a best imputation techniques.

12. What is A/B testing?

Answer : A/B testing is basically Statistical hypothesis testing, we can call it Statistical inference. It is an analytical method for making decision that estimates population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

Answer : NO, Mean imputation is typically considered terrible practice since it ignores feature correlation. 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.

14. What is linear regression in statistics?

Answer : Linear regression is a kind of statistical analysis that attempts to show a relationship between two variables. Linear regression looks at various data points and plots a tend line.

15. What are the various branches of statistics?

Answer : The two main branches of statistics are Descriptive statistics and Inferential statistics.