STATISTICS WORKSHEET-1

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
   Answer- 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
   Answer- 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
   Answer- 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
   Answer- D) All of the mentioned
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   Answer- C) Poisson
6. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
   Answer- B) False
7. Which of the following testing is concerned with making decisions using data?
   a) Probability
Hypothesis
   b) Causal
   c) None of the mentioned
   Answer- B) Hypothesis
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
   Answer- 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

Answer- 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?

Answer- The Mean, Median, Mode of normal distribution is equal. A normal distribution is the proper term for a probability bell curve. Normal distribution are symmetrical, but not all symmetrical distributions are normal. The normal distribution skewness is 0 and kurtosis is 3. Normal distribution is a limiting case of Poisson distribution as n--->infinity.

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

Answer - Firstly, We should check the data whether it is missing data. If it is a missing data we need to use some solution for data imputation depending on the kind of problem. For Imputation there are two methods they are :-
1)Time series problems
2)General problems

12. What is A/B testing?

Answer- A/B testing is also known as split testingor bucket testing. It is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is a basic randomized control experiment.

13. Is mean imputation of missing data acceptable practice?

Answer- Mean Imputation also called mean substitution. It's a popular solution to missing data, despite its drawbacks. But that doesn't make it a good solution , and it may not help you find relationships with strong parameters estimators .Even if they exist in the population.

14. What is linear regression in statistics?

Answer- Linear regression may be defined as the statistical model that analyse the linear relationship between a dependent variable with given set of independent variables . The case of one explanatory variable is called simple linear regression ,for more than one , the process is called multiple linear regression.

15. What are the various branches of statistics?

Answer- The two main branches of statistics are:
• Descriptive statistics
• Inferential statistics
Both of these are employed in scientific analysis of data.