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

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

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

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

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned
Answer: c

6. Usually replacing the standard error by its estimated value does change the CLT.
a) True **b) False**
Answer: b

7. Which of the following testing is concerned with making decisions using data?
a) Probability
b**) Hypothesis**
c) Causal
d) None of the mentioned
Answer: b

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

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:

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. Normal distribution is a bell shape curve with continuous probability distribution which is symmetric about its mean. It has zero mean and skew.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans. Missing data can be handled in many ways:

a. Predicting missing values
b. Delete the row or column of missing values if not value to dataset much
c. Substituting the missing values (by mean or mode)
d. By imputation techniques

According to recommendation, by calculating statistical value of columns(ie mean or mode) and replace all missing values for that column with statistic.

12. What is A/B testing?
Ans. A process in which a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not. A/B resting is the example of statistical hypothesis testing.

13. Is mean imputation of missing data acceptable practice?
Ans. No,it not fit in case of correlation and many more.(eg. If we have to take admission on the basis of university rank in that case missing value filled with mean does not work)

14. What is linear regression in statistics?
Ans. Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. A linear regression line has an equation of the form y=mx+c, where x is the explanatory variable, y is the dependent variable, m is coefficient and c is intercept.

15. What are the various branches of statistics?
Ans. There are two branches of statistics

1. Descriptive statistics
2. Inferential statistics