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)
   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. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
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
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
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

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the basic branches of statistics?

Ans1: a
Ans 2: a
Ans3: b
Ans4: c
Ans5: c
Ans6: a
Ans7: b
Ans8: a
Ans9: c
Ans10: Normal distribution is a most common distribution which is refer to a probability distribution where the random variables distributed symmetrically it has called by two name the other name is gaussians distribution. It is significant in statistics and it is used to get real-valued random variables, whose distribution are unknown.

Ans11: first, i will check the columns and rows if it is not necessary for dataset then I'll remove the data rows or columns from the dataset. second step I can take to fill null values and copy the values and then fill the missing data or take alternative values to fill missing data.

Ans12: An A/B is a relational hypothesis is test between two data sets where a person could found statistical relation between two datasets. After testing they compared against each other to see if they having a similar relation between both datasets or if there's any statistical relationship left or no.

Ans14: Linear regression is used to predict the value of variables. Linear regression in statistics is an approach for develop the relationship between scaler response and one more explanatory variables which is called as dependent and independent variables and the one explanatory variable is called simple linear regression.

Ans15: The basic branches of statistics are:

a) Data collection
b) Descriptive statistics
c) Inferential statistics