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. 10. 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 : c

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.
10. What do you understand by the term Normal Distribution?
Explanation : Let's suppose there is an class. I am analyzing the grade of each student's secured in examination .Let grades are A,B,C,D,E. So in class most students secure grade 'C', smaller students secured 'B or D' and rest smaller secure 'A or E'. With the help of mean and deviation when we plot the distribution plot, it will be look like Normal Distributed.

11. How do you handle missing data? What imputation techniques do you recommend?
Explanation : There are different technique and algorithm that deal automatically on missing data.

12. What is A/B testing?

13. Is mean imputation of missing data acceptable practice?
Explanation : No.
Suppose we have data of human like human age and fitness score. Let's suppose age are between 18-60, according to their age they have fitness score. Let suppose data is missing for age 18. When we use mean imputation(mean of age 18-60) then the data will be wrong, because the value for fitness score greater than the actual fitness score.
14. What is linear regression in statistics?
Explanation : It's an predictive analysis. We have a set of independent variable based on the independent Variables we are trying to predict the value or label or target variable.

15. What are the various branches of statistics?
Explanation : Data Collection : How actual data is collected
Descriptive Statics : Deal with presentation of data we have
Inferential Statistics : Deal with making conclusion about data**.**