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 is (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 is (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 is (d)

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

   Answer is (c)

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

   Answer is (b)

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 is (b)

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 is (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 is (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?

Normal distribution can be understood to mean the tendency for the probability of data points or cluster to occupy the middle in the range of data distribution in comparison to others that occupy the space away from the Centre.

11. How do you handle missing data? What imputation techniques do you recommend? What is A/B testing?

The problem of missing data can be solved in several ways:

- Drop the entire variable with the missing value (if the action is estimated to cause minimum impact)
- Drop the data entry.
- Data trace. contact the source of the dataset to see if the missing data can be recovered.
- Replace missing values by either comparing data in nearby variables or calculating the of the variable.

12. Is mean imputation of missing data acceptable practice?

Yes on condition that its not going to cause extensive variation in the final analysis

13. What is linear regression in statistics?

Linear regression is when a given variable determine change in the value of another variable.

14. What are the various branches of statistics?

There are at least two branches of statistics namely:

- Descriptive Statistics – gathering and organizing data into exploration form for future analysis.
- Inferential statistics - for testing hypotheses to formulate sensible conclusions.