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
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) conform to the regression relationship Outliers cannot
   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?

A) Normal distribution is a continuous probability distribution wherein values lie in a symmetrical fashion mostly situated around the mean. We use various functions in numpy library to mathematically calculate the values for a normal distribution.

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

A) Missing data can be handled by various methods:

1) Deleting the record of missing values: But it is not suitable for less number of records, when there is huge flow of record we can use it.
2) Create a separate model to handle the missing data.
3) Statistical methods such as mean , median and mode methods . Also known as **reduced feature models, distribution-based imputation, prediction value imputation**.

12. What is A/B testing?

A) A/B testing is also called split testing. The testing can be done to the same product or website to know the Acceptance rate or willing ness of that particular product or website.

13. Is mean imputation of missing data acceptable practice?

A) Bad practice in general. mean imputation preserves the mean of the observed data. leads to an underestimate of the standard deviation. Distorts relationships between variables by "pulling" estimates of the correlation toward zero.

14. What is linear regression in statistics?

A) Establish the relationship between two variables if exists.
Example: Income and spends, student heights and exam results.
Forecasting new observations: To know about where the relationship should be developed for unobserved values , it means predicting the future business or sales.

I n this model we use two variables that is dependent variables and independent variables.
Dependent variables: its values depends on something else. Denoted by Y Variable.
Independent variables: its values explain the other values and is independent and is denoted by X Variable.

15. What are the various branches of statistics?

A) The two main branches of statistics are descriptive statistics and inferential statistics.
Descriptive Statistics: Descriptive statistics is the first part of statistics that deals with the collection of data.
Inferential statistics: Inferential statics include techniques can be used by the statisticians to collect the data or information from sample to conclusion.