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. True
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) Central Limit Theorem
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) Modeling bounded count data
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)All of the mentioned
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   Answer. **c) poisson**
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
   Answer. 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
   Answer. b) Hypothesis
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. a) **0**
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) Outliers cannot conform to the regression relationship

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

10. What do you understand by the term Normal Distribution?

Answer. normal distribution is a stastical phenomenon representing a symmetric bell shaped curve. Most value located near he mean only few of them appear right and left tail.

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

Answer. Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you. Your application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea.

12. What is A/B testing?

Answer. A/B testing is shorthand for a **simple randomized controlled experiment**, in which two samples (A and B) of a single vector-variable are compared. These values are similar except for one variation which might affect a user's behavior. A/B tests are widely considered the simplest form of controlled experiment.

13. Is mean imputation of missing data acceptable practice ?

Answer. Bad practice in general

If just estimating means: 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?

Answer. Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

15. What are the various branches of statistics?

Answer. the two main branches of branches are discriptive statistics and inferential statistics.both of these employed in scientific analysis of data .