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

   A) 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

   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

   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

5. **Poisson** 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

   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

   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

   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

C)<u>Outliers cannot confirm to the regression relationship</u>

FLIP ROBO

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 various branches of statics?

10) A normal distribution is an arrangement of the data set in which most value cluster in the middle of the Range and the rest taper off symmetrically toward either extreme. The precise shape can vary According to the distribution of the population but the peak is always in the middle and true curve is Always symmetrical.

11) Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets Where participants have missing fields. Mean imputation. Simply calculate the mean of the observed Values for the variable for all individuals who are non-missing.

12) A/B testing (also known as bucket or split-run testing) is a user experience research methodology. A/B Testing is a way to compare two version of single variable, typically by testing a subject's response to Variant A against variant B and determining which of the two variants is more effective.

13) True, imputing the mean preserve the mean of the observed data. So, if the data are missing completely At random, the estimate of the mean remains unbiased.That's good things. Since most research studies Are interested in the relationship among variables, mean imputation is not good solution.

14) In statistics, linear regression is a linear approach for modelling the relationship between a scalar Response and one or more explanatory variables. The case of one explanatory variable is called Simple linear regression, for more then one, the process is called multiple linear regression.

15) The two main branches on statistics are descripted statistics and inferential statistics. Both of these are Are employed n scientific analysis of data both are equally important for the student of statistics.