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)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 ANSWER: **a)**Central Limit 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 **Answer : b)modeling bounded count data**
   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 **Answer : d)**
   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 **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 **Answer : 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 **Answer: a) 0**
   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 **Answer:c)**
   c) Outliers cannot conform to the regression relationship
   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?

ANS: The normal distribution is a continuous probability that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions.

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

ANS: Depending on the type of the imputed variable (i.e. continuous, ordinal, nominal) and missing data pattern (i.e. monotone, non-monotone), below are a few commonly used models. If you plan to do it in SAS, there are SAS codes that you can write to identify the missing data pattern.

- Logistic Regression
- Discriminant Regression
- Markov Chain Monte Carlo (MCMC

12. What is A/B testing?
ANS:A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. Here, either you can use random experiments, or you can apply scientific and statistical methods
13. Is mean imputation of missing data acceptable practice?
ANS: True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the **mean remains unbiased**. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.
14. What is linear regression in statistics?
ANS: In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).
15. What are the various branches of statistics?
ANS: There are three real branches of statistics: data collection, descriptive statistics and inferential statistics