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
   Ans:(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
   Ans:(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
   Ans:(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
   Ans:(d)
5. _____ random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
   Ans:(c)
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
   Ans:(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
   Ans:(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
   Ans:(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
   Ans:(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?
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 statistics?

Answer 10:

NORMAL DISTRIBUTION: The normal distribution describes the symmetrical plot of data around its mean value,where the width of curve defined by its standard deviation. It is visually depicted by Bell curve.

Answer 11:
Missing data is the data that is not captured for a variable for the observation in question.When dealing with missing data,we can use two primary methods to solve the error: imputation or data removal.
Data Imputation: The imputation method substitutes reasonable guesses for missing data. It's most useful when the percentage of missing data is low. If the portion of missing data is too high, the results lack natural variation that could result in an effective model.
Answer 12:
A/B testing is a form of statistical and two-sample hypothesis testing. Statistical hypothesis testing is a method in which a sample dataset is compared against the population data. Two-sample hypothesis testing is a method in determining whether the differences between the two samples are statistically significant or not.
Answer 13:
Mean imputation is not acceptable practice to take care of missing data because of these reasons:
i)It does not preserve the relationships among variables.
ii)It leads to underestimate of standard errors.
Answer 14:
Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.
Answer 15:
The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.
Descriptive Statistics:
Descriptive statistics is the part of statistics that deals with presenting the data we have. The basic aim of descriptive statistics is to 'present the data' in an understandable way. If we simply write down every piece of data, it means little to someone who sees it; it needs to be summarized. descriptive statistics can be categorized into:

i) Measure of Central Tendency: Mean,Mode,Median
ii) Measures of Variability:Quartile,Ranges,variance,Standard Deviation

Inferential Statistics:
Inference statistics are statistical techniques that allow us to utilize data from a sample to conclude, predict the behavior of a given population, and make judgments or decisions. Furthermore, a statistician uses these techniques mainly for data analysis, writing, and drawing conclusions from the limited data. This is accomplished by taking samples and determining their reliability. There are some different types of inferential statistics, which include the following, which are shown below:

- Regression analysis
- Analysis of variance (ANOVA)
- Analysis of covariance (ANCOVA)
- Statistical significance (t-test)
- Correlation analysis