STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer.

1. Bernoulli random variables take (only) the values 1 and 0.
   Ans. 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?
   Ans. a) Central Limit Theorem

3. Which of the following is incorrect with respect to use of Poisson distribution?
   Ans. b) Modeling bounded count data

4. Point out the correct statement.
   Ans. d) All of the mentioned

5. _____ random variables are used to model rates.
   Ans. c) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.
   Ans. b) False

7. Which of the following testing is concerned with making decisions using data?
   Ans. b) Hypothesis

8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.
   Ans. a) 0

9. Which of the following statement is incorrect with respect to outliers?
   Ans. 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?
    Ans. In a normal distribution, data is symmetrically distribution with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region

and tapering off as they go further away from the centre. Normal distribution is also called Gaussian distributions or bell curves because of their shape.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans. Data can have missing values for a number of reasons such as observations that were not recorded and data corruption.

- Delete the record missing values
- Create a separate model to handle missing values
- Statistic method mean median and mode

12. What is A/B testing?
Ans. A/B testing is a way to compare two versions of a 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. A/B testing includes an application of statistical hypothesis testing**.**

13. Is mean imputation of missing data acceptable practice?
Ans. It is acceptable when the missing value proportion is not large enough. But, when the missing values are large enough and you impute them with the mean, the standard errors will be lesser than what they actually would have been**.**

14. What is linear regression in statistics?
Ans. Linear Regression is a statistical supervised learning technique to predict the quantitative variable by forming a linear relationship with one or more independent features.
It helps to determine:

- If   independent variable does a good job in predicting the dependent variable.
- Which independent variable plays a significant role in predicting the dependent variable.

15. What are the various branches of statistics?
Ans. There are two different branches of statistics.

1. **Descriptive Statistics:** – These statistics helps one to describe the data. It is like learning the behaviour of the data. These methods help collect, summarize, analyse and finally present the data. Every time a data scientist starts a new project give or take, he goes through descriptive statistics to understand and compare data. Most of the analysis here is univariate in nature i.e., it involves exploring one variable at a time.

   Example of Descriptive Statistics The average score of the college students in the math test. The average age of the people who voted for the winning candidate in the last election. The average length of the statistics book.

2. **Inferential statistics:** - Inference statistics often speak in terms of probability by using descriptive statistics. Besides, a statistician uses these techniques for data analysis, drafting, and making conclusions from limited information. That is obtained by taking samples and testing how reliable they are.

Example of Inferential Statistics

Suppose you want to get an idea about the percentage of the people who love shopping at FILA. We take the sample of the population and find the proportion of individuals who love the FILA brand. With the assistance of probability, this sample proportion allows us to make a few assumptions about the population proportion. This study belongs to inferential statistics.