STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.

   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 Theoram

3. Which of the following is incorrect with respect to use of Poisson distribution?

   B) Modeling bounded count data

4. Point out the correct statement.

   D) All of the mentioned

5. _____ random variables are used to model rates.

   C) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.

   B) False

7. Which of the following testing is concerned with making decisions using data?

   B) Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

   A) 0

9. Which of the following statement is incorrect with respect to outliers?

   C) Outliers cannot conform to the regression relationship

Q10 What do you understand by the term Normal Distribution?

Normal Distribution is important because of the central limit theorem. Normal Distribution is a continuous probability distribution that possesses a bell-shaped curve. Normal distribution curves are sometimes designed with a histogram inside the curve. The graphs are commonly used in mathematics, statistics and corporate data analytics

Q11. How do you handle missing data? What imputation techniques do you recommend?

The answer depends a lot on the data itself. You need to know the type of missingness, the circumstances under which the data were generated and a whole lot of other factors.

Deletion: List wise and Pair wise, You can delete the the missing values using these two methods. In list wise method you need to erase complete row which contains missing value. And pair wise method will do analysis by neglecting the missing value and considering only the available values.

Imputation: You can also choose imputation method to handle missing data in case you want to replace them. There are different types of imputation methods, you can select them based upon your data.

Prediction model: You can build a model just to predict the missing values.

The imputation techniques I recommend is replacing missing values with the mean or median values of the dataset.

Q12. What is A/B testing?

A/B testing is a basic approach to compare two versions of something to figure out which performs better. Even though A/B testing can be used for any experiments it is often used in the digital marketing world (websites, apps, email campaigns…).A/B testing is a digital marketing technique that involves comparing two versions of a web page or application to see which performs better. These two versions, A and B are randomly shown to different visitors at the same time. Then a statistical analysis of the results determines which version performs better according to certain predefined indicators such as conversion rate.

Q13. Is mean imputation of missing data acceptable practice?

The process of replacing null values in a data collection with the data's mean is known as mean imputation.Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

Q14. What is linear regression in statistics?

Linear regression is a statistical method that tries to show a relationship between variables. It looks at different data points and plots a trend line. A simple example of linear regression is finding that the cost of repairing a piece of machinery increases with time.

$$y = mx + b$$

- y is the estimated dependant variable
- m is the regression coefficient
- x is the independent variable
- b is the constant

Q15. What are the various branches of statistics?

Statistics is a study of presentation, analysis, collection, interpretation and organization of data

There are two main branches of statistics –

Inferential Statistic.

Descriptive Statistic.

Inferential Statistics: Inferential statistics used to make inference and describe about the population. These stats are more useful when its not easy or possible to examine each member of the population.

Descriptive Statistics: Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphical form.