STATISTIC 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 Theorem

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?
a) Probability
b) Hypothesis

8.Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0

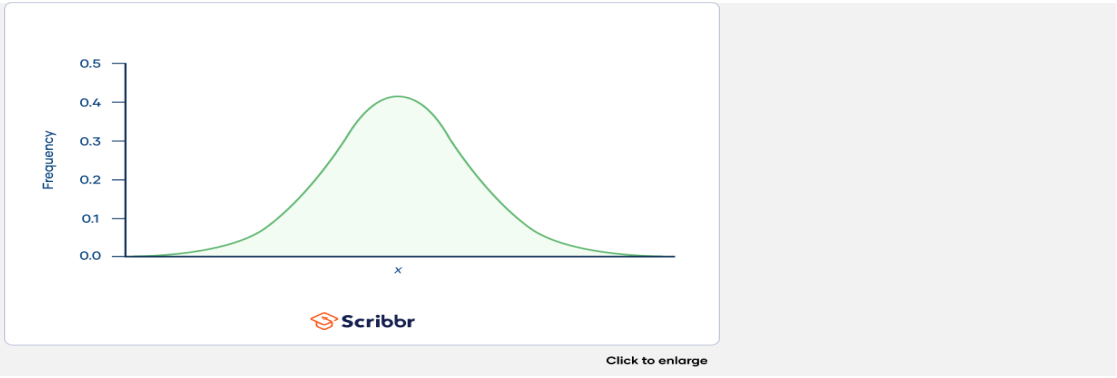
9. Which of the following statement is incorrect with respect to outliers?
c) Outliers cannot conform to the regression relationship

10) What do you understand by the term normal distribution?

Ans) In a normal distribution, data is symmetrically distributed 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 center.

Normal distributions are also called Gaussian distributions or bell curves because of their shape.

Understanding the properties of normal distributions means you can use inferential statistics to compare different groups and make estimates about populations using samples.



11) How do you handle missing data? what imputation techniques do you recommend?

Ans) Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you.

Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

IMPUTATION TECHNIQUES :

Mean imputation:

Calculate the mean of the observed values for that variable for all non-missing people. It has the advantage of maintaining the same mean and sample size, but it also has a slew of drawbacks.

Substitution:

Assume the value from a new person who was not included in the sample.To put it another way, pick a new subject and employ their worth instead.

Hot deck imputation

A value picked at random from a sample member who has comparable values on other variables.

Cold deck imputation

A value picked deliberately from an individual with similar values on other variables.In most aspects, this is comparable to Hot Deck, but without the random variance.

Regression imputation

The result of regressing the missing variable on other factors to get a predicted value. As a result, instead of utilising the mean, you're relying on the anticipated value, which is influenced by other factors.

12) What is A/B testing?

Ans) A/B testing is a shorthand for a simple randomized controlled experiment, in which two samples (A and B) of a single vector variable are compared.

These values are similar except for one variation which might affect a user's behavior. A/B tests are widely considered the simplest form of controlled experiment.

However, by adding more variants to the test, its complexity grows.

A/B tests are useful for understanding user engagement and satisfaction of online features like a new feature or product.

Large social media sites like LinkedIn, Facebook, and Instagram use A/B testing to make user experiences more successful and as a way to streamline their services

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 basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things:

(1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable?

(2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable?

These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula

y = c + b*x, where:

y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

There are many names for a regression's dependent variable. It may be called an outcome variable, criterion variable, endogenous variable, or regressan. The

independent variables can be called exogenous variables, predictor variables, or regressors.

Three major uses for regression analysis are :

(1) determining the strength of predictors,

(2) forecasting an effect (3) trend forecasting.

15)What are the various branches of statistics?

Ans) 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 deals with the presentation and collection of data. This is usually the first part of a statistical analysis.
It is usually not as simple as it sounds, and the statistician needs to be aware of designing experiments, choosing the right focus group and avoid biases that are so easy to creep into the experiment.
Different areas of study require different kinds of analysis using descriptive statistics.

For example, a physicist studying turbulence in the laboratory needs the average quantities that vary over small intervals of time.

Inferential Statistics:

Inferential statistics, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.
In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.

Most predictions of the future and generalizations about a population by studying a smaller sample come under the purview of inferential statistics.
Most social sciences experiments deal with studying a small sample population that helps determine how the population in general behaves. By designing the right experiment, the researcher is able to draw conclusions relevant to his study.

Both descriptive and inferential statistics go hand in hand and one cannot exist without the other. Good scientific methodology needs to be followed in both these steps of statistical analysis and both these branches of statistics are equally important for a researcher.