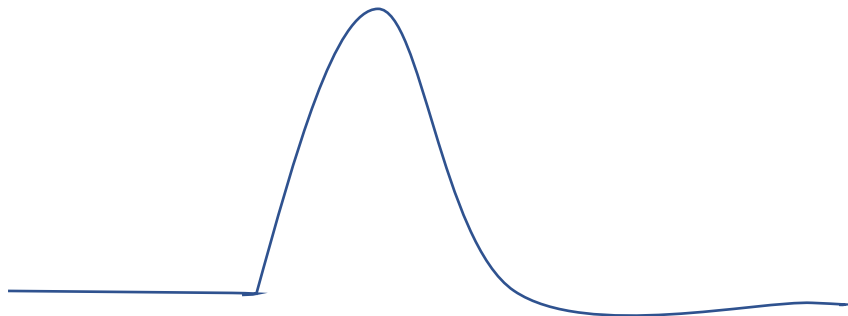


STATISTICS WORKSHEET-1

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 Theorem
3. Which of the following is incorrect with respect to use of Poisson distribution?
c) Modelling contingency tables
4. Point out the correct statement.
d) All of the mentioned
5. _____ random variables are used to model rates.
d) All of the mentioned
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 _____ 0 _____ 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. Normal distribution is a type of continuous probability distribution. It is the most widely known and used of all distributions. Since it approximates many natural phenomena so well, it is used as a standard of reference for many probability problems. It is also known as bell-curve.



11. One of the most accepted methods for handling missing data is deletion. When the number of observations with missing data is greater than the number of observations without missing values, we apply deletion.

Imputation- another general method for dealing with missing data is to fill in the missing value with a substituted value with `.fillna()` method.

Simple Imputation- In this method, we need to evaluate variables' distribution to determine which central tendency summary (mean, median, mode) to apply.

KNN Imputation- KNN imputation is a recommended or a fairer approach to the simple imputation method. It operates by replacing missing values with the average mean of the neighbours nearest to it.

12. A/B testing allows decision makers to choose the best design for a website by looking at the analytics results obtained with two possible alternatives A and B. some examples of A/B testing are: hypotheses test, chi-squared test, z-test, student t test

It is a way to compare the two versions of a variable to find out which perform better in a controlled environment.

13. the process of replacing null values in a data collection with the data's mean is known as mean imputation. It is considered terrible practice as it ignores feature correlation. Also, it decreases the variance of our data while increasing bias. So the model is less accurate.

14. Linear Regression is commonly used type of predictive analysis. Its estimates are used to explain the relationship between one dependent variable and one or more independent variables.

The simplest form of linear regression equation with one dependent and one independent variable is given by,

$$Y=mX+c$$

Where, Y= estimated dependent variable

m= regression coefficient

X= independent variable

c= constant

15. Various branches of Statistics are:

i. Descriptive Statistics-It is the part of statistics that deals with presenting the data we have either visually or numerically.

ii. Inferential Statistics- It is the aspect that deals with making conclusions about the data. E.g., hypotheses testing