

STATISTICS WORKSHEET 1

1. A) True
2. A) central limit theorem
3. D) All of the above
- 4.
5. C) poisson
6. B) false
7. B) hypothesis
8. A) 0
9. C) outliers cannot conform to the regression relationship
10. Normal distribution is also called as bell curve which represents the probability distribution line, in which 68% of data will lie between the +1 standard deviation to -1 standard deviation, 95% of data will lie between the +2 standard deviation to -2 standard deviation and 99.7% of data lies between the +3 standard deviation to -3 standard deviation and outside + (or) - 3 standard deviation data points are considered as the outliers
11. There are three techniques to handle the missing data they are
 - * delete the record missing value – this type can be used only when we have huge data value, deleting the missing data values does not make any changes to predicted output value by the model we go for delete the record missing value
 - * create a separate model to handle missing values – in some cases the input values are missing in the particular column or rows in such cases we create separate model and particular missing columns are given for the model test terms for predicting the value and the columns and rows which are having the data is given for model training purpose by this model will train thoroughly and predict the missing value
 - * statistical methods mean, median and mode – in this the missing value are filled with the mean, median and mode value, where the mean is the sum of the data divided by number of data values, median is the centre value of the data when it is arranged in the ascending order and mode is the most repeated value in that data value
12. A to B test basically used to compare two different products and the comparison done by using the user input and at the last concluding which is the better performing model
13. If the data are missing completely at random the estimate of the mean remains unbiased

14. linear regression analysis is used to predict the value of 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 value is called as independent variable

The linear regression equation is

$$Y = a+bx+e$$

Where , y-dependent variable

x-independent variable

b-co efficient value

e- error

a- intercept value

15. There are two main branches of statistics the are

* Descriptive statistics – deals with the presentation and collection of data , this is usually the first part of a statistical analysis . where summarizing data using number and graph , data summary done by bar graph , histograms, pie chart etc

Descriptive further classified into central tendency and description of data

*inferential statistics- using sample data to make an inference or draw the conclusion of the population ,

inferential