Name : Ankit Dadarwala **ID :** 30

Batch No: 1832
Internship19

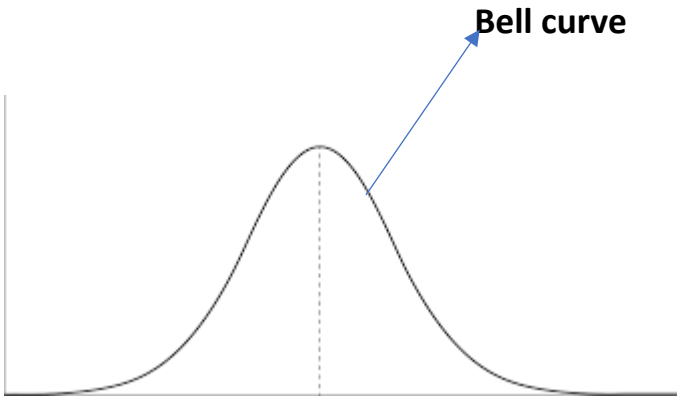
Subject : Statistics Worksheet 1

1. **A.**
2. **A.**
3. **B.**
4. **D.**
5. **C.**
6. **B.**
7. **B.**
8. **A.**
9. **C.**

10. **Normal distribution is an import concept in statistics and machine learning. It work with linear regression, central limit theorem, EDA. It also call Gaussian distribution. It have bell shape curve that symmetrical from mean point to the both halves of curve.**

It useful in dataset to check data distribution are normal or not by histogram, KDE plot.

For standard distribution mean is 0 and std is 1.



11. **Missing data can be handle by imputers like Mean, Median, Mode.**

Python has library scikit learn in which imputer method used
For numerical values we use mean for that particular columns
For categorical data we used most frequent(mode) fill null values.
Python pandas also has method df.fillna(data,method use)
In deep learning datawig library use for all null values

12. **A/B testing means show your current experience to half of your customer and alternat experience to to remaining half of the customer and compare both of them and find the difference of the performance, then either continue with the old one or switch all with the new one.**

13. **No, mean imputation of missing data is not acceptable practice. Because using mean can significantly reduce the model's accuracy and bias the result.**

14. **Linear Regression is used for predictive analysis show the relation between the continues variable it shows the relationship between the independent variable and Dependent variable. If There is one single input variable its simple linear regression and there are more than one input variable than it is called multiple linear regression. It simply uses traditional slope intercept form.**

    Y= mx+b Where, Y= Dependent variable B=intercept of line M=Linear Regression Co-efficient X= independent variable