**Statistics**

* **Measures of Dispersion:**

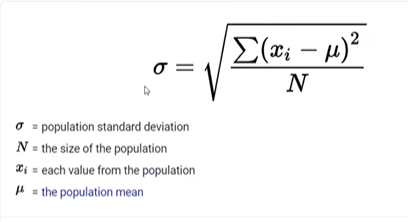
It means how much spread your data is in the bell curve. To measure the spread of the data we have to use the **Variance** **and Standard Deviation**.

**Variance** – Measure the spread of the data. Value of variance is high spread of data is also high.

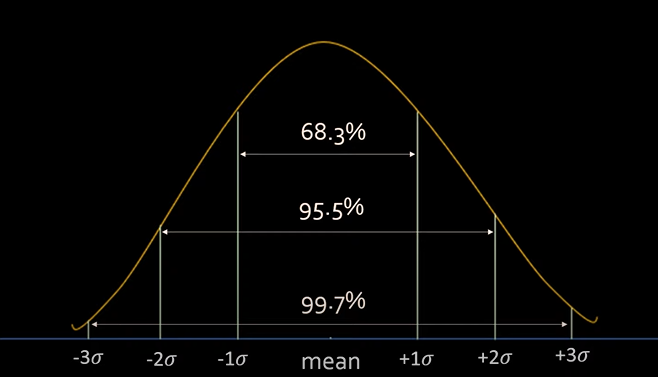
**Population Variance = (Xi-u) ^2 / N**

**Sample Variance = (Xi-u) ^2 / n-1**

* **Standard Deviation** – It actually helps to find how much data point is away from the mean.



* Formula:+ (Mean)\*(S.D.), 2S.D. =+ (Mean)\*2(SD), 3S.D. =+ (Mean)\*3(S.D.)
* 1 time standard deviation covers data up to 68.3%. 2 time standard deviation covers data up to 95.5%, and 3 time standard deviation covers data up to 99.7%.
* Below Gaussian distribution diagram shows how much data can cover by 1 SD, 2 SD and 3 SD. Data point which fall outside of 3 SD is going to be outlier.



* **Z-score:**

It is used to perform standardization process.

**Formula = (Xi – mean) / S.D.**

**IMP 🡪 In Z-Score mean is always 0 and standard deviation is 1.**

**Use case of Z-Score –** Dataset may have features with different scales and unit such as weight in kg, height in cm, age in years. User cannot perform operation on the data as all features have different unit. Here Z-score use to perform standardization process which convert all the features value at same scale.

* **Histogram:**

Histogram is one of the type of graph which is basically use to visualize the data to find out frequency of the data points and outlier.

* **Probability:**

**Sample Space** – It is define as how many different types of outcome you will get from experiment such as tossing coin, rolling a dice.

E.g. – Tossing a coin

Probability of getting head/Tail =1/2

Whereas 1 is the expected outcomes

2 is the total sample space outcomes

**Mutual Exclusive and Non-Mutual Exclusive –**

**Mutual Exclusive:** Two events are mutual exclusive if they both cannot occur at same time.

E.g. Rolling a Dice: 1/6 …. Here you cannot get 1 and 2 at same time.

E.g. Tossing coin

Additive Probability rule**:** rolling a dice what is the probability of getting 1 or 2.

P (1) or p (2) = p (1) + p (2) = (1/6 + 1/6) =1/3

**Non-Mutual Exclusive:** Two events are mutual exclusive if they both can occur at same time.

E.g. taking a card from a deck. What is the probability of getting K or heart.

P(A or B) = P(A) + P(B) – P(A intersection B)

**Note: Naive Bayes’ is algorithms we use probability.**

**Independent and Dependent events:**

**Independent** **event-** In independent events probability of one element is not effects on probability of other elements.

E.g. If you roll a dice then probability of getting any no. from 1 to 6 is 1/6. The probability of getting 1 is not going to effect on probability of 2 or 3, it is still remains as 1/6. This is nothing but independent event.

Pr(A and B) = Pr(A) \* Pr(B) \_\_\_\_ Probability of getting A and B.

**Dependent events-** In dependent events probability of one element is effects on probability of other elements.

E.g. If you have bag contains 5 balls each of different color (red, white, green, orange, black). The probability of getting any color ball is now 1/5. If you pick random ball from bag now the probability of getting others ball is ¼, because you have already picked one ball so it has affected on the probability of other balls. This is nothing but dependent event.

Pr(A and B) = Pr(A) \* P(B/A) \_\_\_\_\_\_\_\_\_\_\_\_ This is also called as conditional probability.

**Use case of probability in Machine Learning:**

Probability can be used in classification as well as regression problems.

Naive Bayes algorithms – to predict the values.

**Permutation and Combination:**

**Permutation nPr** = n! / (n-r)!

**Combination nCr** = n!/r!(n-r)!