Statistics – Standardization, Normalization   
and Z score

# Introduction:

* Standardization is a process used to center the data around the mean with a standard deviation of 1.
* Generally, dataset has different columns
* Different columns have different values
* Different values have different units
* Some columns have very less values
* Ex. 1 to 100 years
* Some columns have very high value
* Ex. K, Lakhs, Crs
* Different Columns have different scales
* When we apply ML models, comparing different scales is not recommended
* Interpretation [problem occurs
* Some math calculation also takes more time
* So, scale all the column values under one scale, it is easy to work on
* Foe Ex. Multiply two-colour images mean we are multiplying 255\*255
* If we change colour image to Gray, 255 becomes 1, then 1\*1

# Standardization:

* It is also called as Z – Standardization
* Standardization is a process used to center the data around the mean with a standard deviation of 1.
* From Empirical Rule (68-95-99.7),



* (value ranges from -3 to +3)
* **Application:** Standardization is widely used in algorithms like Support Vector Machines, Linear Regression, and PCA (Principal Component Analysis) where the distance between points is critical.

# Normalization:

* Normalization is the process of scaling the data to a range of [0, 1] (or sometimes [-1, 1]).
* Z- Standardization values vary from -3 to +3
* Normalization values vary from 0 to 1
* Formula:
* Normalization values vary from -1 to 1 (very rare)
* Normalization use is in image operations
* Colour image to Gray image:🡪 Normalization