**Measure of central tendency:**

* Mean
* Median and
* Mode

**Mean:** We already know, it’s an average.

Mean = Sum of all elements / Number of elements.

Ex: x = {1, 2, 3, 4, 5}

Mean = (1+2+3+4+5)/5 = 3

Let’s say, Entire population is **‘N’**, and sample is **‘n’,** here **N ≥ n.**

We can calculate the population mean using the below formula.

Population Mean **(µ) =, i=1 to N**

Sample mean (xֿ) =**, i=1 to n**

Let us calculate using one more example**:**

**Population Age = {24, 23, 2, 1, 28, 27}**

* **N= 6**
* Population Mean **(µ) = (24+ 23+ 2+ 1+ 28+ 27)/6 = 17.5**

**Sample age = {24, 1, 2, 27}**

* **Here, n = 4**
* Sample mean **(xֿ)** = **(24+1+2+27)/4 = 13.5**

**Here µ≥ xֿ**

Or **xֿ ≤ µ**

**Now let us observe below example, how do we do in practical.**

|  |  |
| --- | --- |
| **Age** | **Salary** |
| 24 | 45 |
| 28 | 50 |
| 29 | NAN |
| NAN | 60 |
| 31 | 75 |
| 36 | 80 |
| NAN | NAN |

NAN = Not A Number

Now, in the above age, salary dataset, we have some NAN values; let us fill those NAN values with mean.

Lets find average of Age column. i.e (24+28+29+31+36)/7 = 29.6

Lets find average of Salary column i.e (45+50+60+75+80)/5 = 62

Updated dataset after handling the NAN values is as follows:

|  |  |
| --- | --- |
| **Age** | **Salary** |
| 24 | 45 |
| 28 | 50 |
| 29 | **62** |
| **29.6** | 60 |
| 31 | 75 |
| 36 | 80 |
| **29.6** | **62** |

Hope you got the clear idea now, how to fill the missing values using mean.

**Median:**

**Steps to find out the median:**

1. **Sort the given number**
2. **Find the central number**
   1. If the number of elements are **odd,** we find the central element as median
   2. If the number of are **even**, take the **middle two** number and **find the mean**.

Let’s say we have the sorted numbers as follows,

X = (1, 2, 3, 4, 5, 6, 7, 8, 100, 120).

Since given example consists of even number,

let’s find out the mean of central numbers. I;e (5+6)/2 = 5.5

So Median = 5.5, Mean = 25.6

I hope you got clear understanding between, mean and median.

Finally,

**Mode**: Most frequent occurring element.

This is majorly used in categorical variables.

Let’s consider the below example.

**Types of flowers:**

|  |  |
| --- | --- |
| **Before Mode operation** | **After Mode operation** |
| Rose | Rose |
| Lilly | Lilly |
| Sunflower | Sunflower |
| NAN | **Rose** |
| Rose | Rose |
| Hibiscus | Hibiscus |
| Rose | Rose |
| NAN | **Rose** |

Here, if we observe all NAN values are filled by **ROSE**, since its occurring majority of times when compare with other type of flowers.