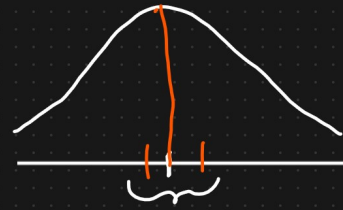① Measure of Central Tendency

① Mean or Average

② Median

③ Mode

① Mean

Population (N)   Sample (n)

$$X = \{1, 1, 2, 2, 3, 3, 4, 5, 5, 6\}$$

Population mean $(\mu) = \sum\limits_{i=1}^{n} \dfrac{x_i}{N}$ $(\bar{x})$. Sample mean $= \sum\limits_{i=1}^{n} \dfrac{x_i}{n}$

$$= \left[\dfrac{1+1+2+2+3+3+4+5+5+6}{10}\right] \qquad =$$

$$= \dfrac{32}{10} = 3.2 \text{ //.}$$

② Median

Steps $X = \{4, 5, 2, 3, 2, 1\}$.

① Sort the Random Variable $\{1, 2, 2, 3, 4, 5\}$.

② No. of elements   Count = 6

③ if Count == even

$\{1,2,\boxed{2,3},4,5\}$

$\Downarrow$

Median = $\dfrac{2+3}{2}$ = 2.5

④ if Count is Odd

$\{1,2,2,\boxed{3},4,5,6\}$

Median = 3

Why Median?

outlier
$\downarrow$

$x = \{1,2,3,4,5\}$

$X = \{1,2,3,4,5,100\}$

$\bar{x} = \dfrac{1+2+3+4+5}{5}$

$\bar{x} = \dfrac{1+2+3+4+5+100}{6}$

$\bar{x} = \dfrac{15}{5} = 3$

$= \dfrac{115}{6} \sim 19.$

$X = \{1,2,\boxed{3,4},5,100\}$

Median = $\dfrac{3+4}{2} = 3.5$

Median is used to find the central Tendency when outliers is present.

③ Mode : Frequency Maximum

$\{2,1,1,1,4,5,7,8,9,9,10\}$

Mode = 1

EDA AND Feature Engineering

Age	Weight	Salary	Gender (Mode)	Degree (Mode)
24	70	40K	M	BE
25	80	70K	F	–
27	95	45K	F	–
→24	–	50K	M	PHD
→32	–	60K	–	B.E
→–	60	–	–	Mashn
→–	65	55K	–	BSC
→40	72	–	M	B.E

DATA IS MISSING