Measures of Central Tendency

What is central tendency?

→ Central tendency is a statistic which best describes what the data (the distribution of data) is centered around.

Types:-

i) Mean: Average of all observations

$$\mu = \frac{\text{Sum of all observations}}{\text{Total no. of observations}}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Ex: marks of student: | 95 | 86 | 82 | 91 | 87 |

* want to know the performance of student?

$$\text{Avg marks} = \frac{95+86+82+91+87}{5}$$

$$= 90.2$$ ↑ mean / avg marks of student.

ii) median: The middle observation (sorted data)

n is odd: $\frac{n+1}{2}$th observation

| 1 | 2 | 3 | 4 | 5 | ⇒ 3

n is even: $\frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$ value

| 1 | 2 | 3 | 4 | 5 | 6 |

$\frac{3+4}{2} = 7/2 = 3.5$

iii) mode: most frequent observation

Ex: 1, 2, 2, 3, 3, 4, 4, 4, 5

mode = 4,,

Measures of dispersion

What are measures of dispersion?

→ measures of dispersion are used to understand the spread of data.
↳ measure of variability

Ex: Consider the marks of two students:-

S1: 10 20 30 40 50 $\quad \mu_1 = 150/5 = 30$ $\quad$ median $= \left(\frac{5+1}{2}\right)^{th}$ term $= 30$,,

S2: 20 29 30 31 32 $\quad \mu_2 = 150/5 = 30$ $\quad$ median $= 30$,,

→ mean & median are same.

* using central tendency, we conclude both students performs the same, but student1 have low marks (10 & 20) and student 2 is more consistent.

Types:

i) Range: The diff b/w maximum and minimum value.

In above example,

the range of student 1 = 50 - 10 = 40

the range of student 2 = 32 - 28 = 4 (10 times less compare to student 1)

ii) Variance: It measures the degree of dispersion of data around the mean.

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{n}$$

Ex1: | 20 | 22 | 25 | 28 | 30 |

$\mu = 25$

X	$x_i - \mu$	$(x_i - \mu)^2$
20	-5	25
22	-3	9
25	0	0
28	3	9
30	5	25
		$\Sigma = 68$

$\sigma^2 = 68/5$

$\boxed{\sigma^2 = 13.6}$

Ex2: | 20 | 30 | 40 | 50 | 60 |

↳ $\mu = 40$

$\sigma^2 = \frac{1000}{5}$

$\boxed{\sigma^2 = 200}$

iii) Standard Deviation: A measure of how dispersed the data is in relation to the mean.

$$\sigma = \sqrt{\frac{1}{n}\Sigma(x_i - \mu)^2}$$

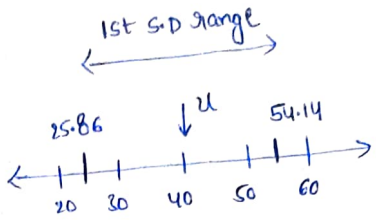
for above variance examples

Ex1 $\Rightarrow \sigma^2 = 13.6 \Rightarrow \boxed{\sigma = 3.69}$

Ex2 $\Rightarrow \sigma^2 = 200 \Rightarrow \sigma = 14.14$

Consider Ex2:

20	30	40	50	60



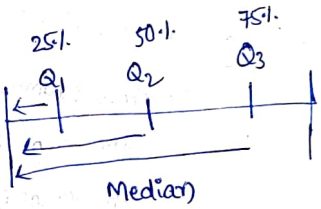
Left $= \mu - \sigma = 40 - 14.14$
$= 25.86$

Right $= \mu + \sigma = 40 + 14.14$
$= 54.14$

* In most cases, 66% of data will be covered in 1st S.D range and around 95% data around in 2nd S.D range.

iv) Quartiles:

Quartiles divide data into four equal parts.

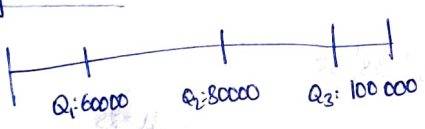


Q_1: value which below 25% of data lies

Q_2: " " " 50% " " "

Q_3: " " " 75% " " "

Ex: Salary in organization



the above graph represents,

75% employees have salary < 100000
50% " " " 80000
25% " " " < 60000

Interquartile Range:- Central 50% of data

$\boxed{IQR = Q_3 - Q_1}$ if IQR value is high → large spread / high dispersion and vice versa.