Measures of Central Tendency

- **Mean of data sets**

Mean or average of a data is given by the formula,

Mean = $\frac{\text{Sum of all observations}}{\text{Number of observations}}$

Note:

-
    - Mean always lies between the highest and lowest observations of the data.
    - It is not necessary that mean is any one of the observations of the data.

1. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then $x_1-\bar{x}+x_2-\bar{x}+x_3-\bar{x}+\ldots+x_n-\bar{x}=0$.
2. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean of x_1+p, x_2+p, x_3+p, ..., x_n+p is $(\bar{x}+p)$.
3. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean of x_1-p, x_2-p, x_3-p, ..., x_n-p is $(\bar{x}-p)$.
4. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean of px_1, px_2, px_3, ..., px_n is $p\bar{x}$.
5. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean of $\frac{x_1}{p}$, $\frac{x_2}{p}$, $\frac{x_3}{p}$, ..., $\frac{x_n}{p}$ is $\frac{\bar{x}}{p}$.

Example:

The runs scored by a batsman in 6 matches are as follows:

24, 126, 78, 43, 69, 86

What is the average run scored by the batsman?

Solution:

Total number of runs scored = 24 + 126 + 78 + 43 + 69 + 86

= 426

Number of matches = 6

∴ Average runs scored = $\frac{426}{6} = 71$

- **Mean of grouped data using direct method**

  Mean $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$, where f_i is the frequency corresponding to the class mark x_i.

  Example:

  Consider the following distribution of marks scored by the students of a class in a unit test.

Marks scored	10 – 20	20 – 30	30 – 40	40 – 50
Number of students	4	7	15	14

Find the mean marks obtained by the students

Solution:

Class interval	**Frequency (f_i)**	**Class mark(x_i)**	$f_i x_i$
10 – 20	4	15	60
20 – 30	7	25	175
30 – 40	15	35	525
40 – 50	14	45	630
Total	$\Sigma f_i = 40$		$\Sigma f_i x_i = 1390$

Mean $= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1390}{40} = 34.75$

Thus, the mean of the marks obtained by the students is 34.75.

- **Assumed-mean method**

$\bar{x} = a + \bar{d} = a + \frac{\sum f_i d_i}{\sum f_i}$, where '$a$' is the assumed mean, $d_i = x_i - a$, and f_i is the frequency corresponding to the class mark x_i

Example:

The table below shows the attendance of students for 30 working days in a particular school.

Attendance	300 – 320	320 – 340	340 – 360	360 – 380	380 – 400
Number of days	8	6	7	6	3

Find the average attendance in this school.

Solution:

$$\text{Class marks} = \frac{\text{Upper limit} + \text{Lower limit}}{2}$$

$$\therefore x_1 = \frac{300+320}{2} = 310$$

$$x_2 = \frac{320+340}{2} = 330$$

$$x_3 = \frac{340+360}{2} = 350$$

$$x_4 = \frac{360+380}{2} = 370$$

$$x_5 = \frac{380+400}{2} = 390$$

Let the assumed mean 'a' be 350.

Class interval	**Number of days (f_i)**	**Class mark(x_i)**	$d_i = x_i - a$	$f_i d_i$

300 – 320	8	310	–40	–320
320 – 340	6	330	–20	–120
340 – 360	7	350 = *a*	0	0
360 – 380	6	370	+20	+120
380 – 400	3	390	+40	+120
Total	$\sum f_i = 30$			$\sum f_i d_i = -200$

$$\therefore \bar{x} = a + \frac{\sum f_i d_i}{\sum f_i} = 350 + \frac{(-200)}{30} = 350 - 6.67 = 343.33 \approx 343$$

Thus, the required average attendance in the school is 343 students per day.

- **Step-deviation method**

$$\bar{x} = a + h\bar{u} = a + h\left(\frac{\sum f_i u_i}{\sum f_i}\right), \text{ where } u_i = \frac{x_i - a}{h},\ f_i$$

is the frequency corresponding to the class mark x_i , a is the assumed mean and h is the class size

Example: Find the mean of the following data.

Class interval	Frequency
600 – 800	4
800 – 1000	2
1000 – 1200	3
1200 – 1400	8
1400 – 1600	3

Solution:
Class size (h) = 200

Class interval	Frequency (f_i)	Class mark(x_i)	$d_i = x_i - a$	$u_i = \frac{x_i - a}{h}$	$f_i u_i$
600 – 800	4	700	–400	–2	–8
800 – 1000	2	900	–200	–1	–2
1000 – 1200	3	1100 = *a*	0	0	0
1200 – 1400	8	1300	200	1	8
1400 – 1600	3	1500	400	2	6
Total	20				4

$$\bar{x} = a + h\left(\frac{\sum f_i u_i}{\sum f_i}\right)$$

$$= 1100 + 200 \times \frac{4}{20}$$

$$= 1100 + 40$$

$$= 1140$$

Thus, the required mean is 1140.

1. The assumed-mean method and the step-deviation method are simplified forms of the direct method

2. The mean obtained by all the three methods is the same.
3. Step-deviation method is convenient to apply if all d_i's have a common factor.

Note: If the class sizes are unequal, and x_i are numerically large, then the step-deviation method is still applicable by taking h to be suitable divisor of all the d_i's.

- **Median**

Median is the value of the middlemost observation when the data is arranged in increasing or decreasing order.

To find the median, the observations are arranged in ascending or descending order and then, if the number of observations (n) is odd, the value of $\left(\frac{n+1}{2}\right)^{th}$ observation is the median.

If the number of observations (n) is even, then the mean of the values of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations is the median.

Example:

The weights of 7 students are as follows: 30, 35, 41, 29, 28, 32, 30. What is the median of this data?

Solution:

The observations in ascending order are 28, 29, 30, 30, 32, 35, 41.

Here, n = 7 (which is odd)

$$\therefore \text{Median} = \left(\frac{7+1}{2}\right)^{th} \text{observation}$$

$$= 4^{th} \text{observation}$$

$$= 30$$

Quartile Deviation

Quartile deviation is the half of the difference between third quartile, Q_3 and first quartile, Q_1 of the series.

∴ Quartile deviation = Q3-Q12

Quartile deviation gives half of the range of middle 50% observations. Quartile deviation is also known as semi-inter quartile range.

Calculation of Quartile Deviation

1. For an individual series, the first and third quartiles can be calculated using the following formula:

Q_1 = Value of n+14th ordered observation

Q_3 = Value of 3n+14th ordered observation

2. For a discrete series, the first and third quartiles can be calculated using the following formula:

If $N = \sum f$, then

Q_1 = Value of N+14th ordered observation

Q_3 = Value of 3N+14th ordered observation

3. For a continuos series, the first and third quartiles can be calculated using the following formula:

Q1=L+N4-c.f.f×h

Q3=L+3N4-c.f.f×h

Here, L = lower limit of the quartile class

f = frequency of the quartile class

h = class interval of quartile class

c.f. = total of all the frequencies below the quartile class

N = total frequency, ∑f

- **Mode**
  1. The mode of a set of observations is the observation that occurs most often.
  2. Mode of a large data can be calculated by forming a tally marks table.

Example: What is the mode of data: 247, 346, 335, 247, 335, 346, 247, 335, 346, 351, 351, 346, 247, 247, 346, and 247?

Solution:A tally marks table for the given data is as follows.

Data	Tally marks	Frequency
247	$\cancel{||||}\ |$	6
335	$|||$	3
346	$\cancel{||||}$	5
351	$||$	2

Therefore, mode of the data is 247.