Statistics

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

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

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.

- **MODE**
  - **Mode of ungrouped data**
- The mode or modal value of a distribution is the observation for which the frequency is the maximum.
  - **Mode of grouped data**

Mode of a grouped data is given by:

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

where, l = Lower limit of the modal class
h = Size of the class interval (assuming all class sizes to be equal)
f_1 = Frequency of the modal class
f_0 = Frequency of the class preceding the modal class
f_2= Frequency of the class succeeding the modal class

Example: Find the mode of the following distribution.

Class interval	Frequency
0 – 5	4
5 – 10	9
10 – 15	7
15 – 20	10
20 – 25	5
25 – 30	6

Solution: The maximum class frequency is 10.
Modal class is 15 – 20
l = 15, h = 5
f_1 = 10, f_0 = 7, f_2 = 5

$$\therefore \text{Mode} = 15 + \left(\frac{10-7}{2\times10-7-5}\right)\times5$$
$$= 15 + \frac{15}{8}$$
$$= 15 + 1.875 = 16.875$$

- **Median of grouped data**

Median of a grouped data is given by:

$$= l + \left(\frac{\frac{n}{2} - cf}{f}\right)\times h$$

Median
where l = Lower limit of median class
n = Number of observations
cf = Cumulative frequency of the class preceding the median class
f = Frequency of the median class
h = Class size (assuming class size to be equal)

Example: Find the median of the following distribution.

Class interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120

Frequency	7	8	6	8	6	5

Solution: The cumulative frequency for the given data can be written as:

Class interval	Frequency	Cumulative frequency
0 – 20	7	7
20 – 40	8	7 + 8= 15
40 – 60	6	15 + 6 = 21
60 – 80	8	21 + 8 = 29
80 – 100	6	29 + 6 = 35
100 – 120	5	35 + 5 = 40

Here, $n = 40$

$$\therefore \frac{n}{2} = \frac{40}{2} = 20$$

lies in the class 40 – 60

Median class is 40 – 60

$$\text{Median} = l + \left(\frac{\frac{n}{2} - cf}{f} \right) \times h$$

$$l = 40, cf = 15, f = 6, h = 20$$

$$\therefore \text{Median} = 40 + \left(\frac{20 - 15}{6} \right) \times 20$$

$$= 40 + \frac{5}{6} \times 20$$

$$= 40 + 16.66 (\text{approx})$$

$$= 55.66 (\text{approx})$$

- **Graphical representation of cumulative frequency distribution Ogive**
-
  - **OGIVE (of the less- than type)**

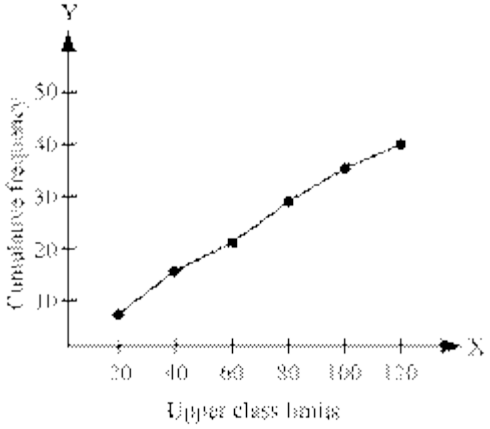
Example 1: Draw ogive of the less-than type for the given distribution.

Class interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120
Frequency	7	8	6	8	6	5

Solution: The cumulative frequency distribution for the given data can be found as:

Class interval	Upper class limit	Frequency	Cumulative frequency
0 – 20	20	7	7
20 – 40	40	8	15
40 – 60	60	6	21
60 – 80	80	8	29
80 – 100	100	6	35
100 – 120	120	5	40

By taking the horizontal axis as the upper class limit and the vertical axis as the corresponding cumulative frequency, we can plot the cumulative frequency for each upper class limit.
Then, the required ogive (of the less-than type) is obtained as:



-
  - **OGIVE (of the more-than type)**

Example 2: Draw ogive of the more-than type for the following distribution.

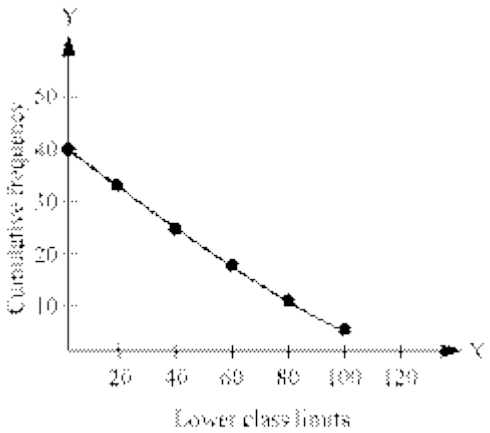
Class interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120
Frequency	7	8	6	8	6	5

Solution: The cumulative frequency for the given data can be found as:

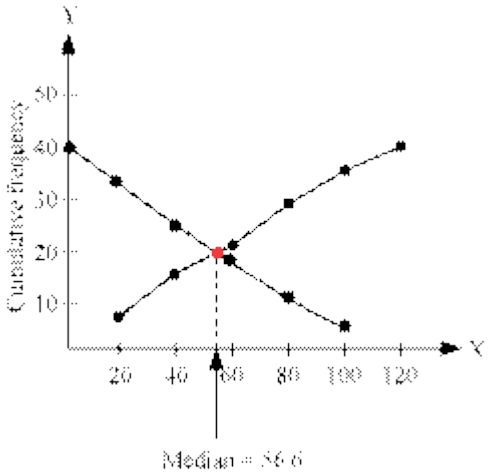
Class interval	Lower class limit	Frequency	Cumulative frequency
0 – 20	0	7	40
20 – 40	20	8	33

40 – 60	40	6	25
60 – 80	60	8	19
80 – 100	80	6	11
100 – 120	100	5	5

By taking the horizontal axis as the lower class limit and the vertical axis as the corresponding cumulative frequency, we can plot the cumulative frequency for each lower class limit.
Then, the required ogive (of the more-than type) is obtained as:



Note:
The *x*-coordinate of the point of intersection of the "more-than ogive" and "less-than ogive" of a given grouped data gives its median.



- **Empirical relationship between the three measures of central tendency**
  3 Median = Mode + 2 Mean