11 Statistics

Frequency table and mean

We have seen the use of such numbers as the mean and median in diverse instances—to understand the educational standard of children in a class or to understand the economic standard of people in a locality.

- The table below gives the number of workers doing various jobs in a factory and their daily wages:

Daily Wages (Rupees)	Number of Workers
210	2
225	4
250	6
270	2
300	1

What is the mean wage?

Here, the mean is the total wages divided by the total number of workers. See how the total wages is computed in the table below:

Daily Wages (Rupees)	Number of Workers	Total Wages (Rupees)
210	2	420
225	4	900
250	6	1500
270	2	540
300	1	300
Total	**15**	**3660**

Repeated addition

To find the mean of a collection of numbers, we have to divide their sum by their number. If some numbers in the collection are repeated, we can find their sum by multiplication. For example, suppose that the ages of 10 children are recorded as below:

13	13	14	15	13
15	14	15	13	15

The sum is easily computed as

$$(4 \times 13) + (2 \times 14) + (4 \times 15) = 140$$

From this, we can compute the mean age as

$$\frac{140}{10} = 14$$

Small, large and mean

The mean of two numbers is the number exactly at their middle. In algebraic language, the mean of the numbers a and b is $\frac{1}{2}(a+b)$.

What about three numbers? Let them be a, b, c in ascending order of magnitude. The mean is $\frac{1}{3}(a+b+c)$. Here, b and c are greater than or equal to a and so the number $\frac{1}{3}(a+b+c)$ is greater than or equal to $\frac{1}{3}(a+a+a)=a$. On the other hand, since a and b are less than or equal to c, the number $\frac{1}{3}(a+b+c)$ is less than or equal to $\frac{1}{3}(c+c+c)=c$.

Thus, the mean is between the smallest number a and the largest number c.

This is true for four numbers also, isn't it? Check it out. What if we take more numbers?

So, the mean is

$$3660 \div 15 = 244$$

Now look at this problem:

- The table below shows the classification of 50 persons in a locality according to their daily income:

Daily Income (Rupees)	Number of People
145 - 155	7
155 - 165	9
165 - 175	14
175 - 185	11
185 - 195	7
195 - 205	2

What is the mean daily income?

How do we find the total daily income of these 50 persons from this?

What is the difference between this table and the earlier one?

For example, look at the first line of this table: from this we know only that there are 7 persons whose daily income is between 145 rupees and 155 rupees. We don't know how many persons earn exactly 145 rupees or how many earn exactly 155 rupees. So, how do we compute the total daily income of these 7 persons?

To get the total, we don't need the individual incomes of these 7 persons; it is enough if we have their mean income. Here, the mean must be between 145 and 155 (see the box **Small, large and mean**). Moreover, this mean would be a number around 150. So, we assume the mean to be 150 and proceed with the computation.

Likewise, the mean daily income of the 9 persons earning between 155 rupees and 165 rupees is assumed to be the middle value 160 of 155 and 165.

So, we enlarge the first table as below:

Daily Income (Rupees)	Number of People	Class Average	Total Income
145 - 155	7	150	1050
155 - 165	9	160	1440
165 - 175	14	170	2380
175 - 185	11	180	1980
185 - 195	7	190	1330
195 - 205	2	200	400
Total	**50**		**8580**

Now can't we compute the mean?

$$8580 \div 50 = 171.6$$

Thus we can take the mean daily income as 172 rupees.

Now try these problems:

- The table below classifies the number of days of a month according to the amount of rainfall received in a certain locality.

  Find the mean daily rainfall during this month.

Rainfall (mm)	Number of Days
54	3
56	5
58	6
55	3
50	2
47	4
44	5
41	2

Distribution and mean

Suppose 7 numbers between 145 and 155 are as below:

145, 147, 147, 150, 152, 152, 155

We can compute the mean and it is 149.71.

These numbers are distributed more or less equally on either side of the middle number, 150. And the mean 149.71 does not differ much from 150.

Now what if the numbers are as below?

145, 145, 145, 146, 146, 148, 155

Most of these are near 145. And what about the mean? It is 147.14. What if most of the numbers are near 155?

Middle and mean

For a collection numbers in an arithmetic sequence, we have seen that their sum is the product of half the sum of the first and the last, with their number. So for such numbers, the mean is half the sum of the first and the last; that is, the number right in the middle of the first and the last numbers.

In an arithmetic sequence the numbers are distributed in the same fashion on either side of the middle number of the first and the last, isn't it so?

- The table below classifies the members of a committee according to their ages:

Age	Number of Members
25 - 30	6
30 - 35	14
35 - 40	16
40 - 45	22
45 - 50	5
50 - 55	4
55 - 60	3

Calculate the mean age of the members of this committee.

- The table below shows the number of students in Class 10 of a school, classified according to their heights:

Height(cm)	Number of Students
120 - 125	19
125 - 130	36
130 - 135	23
135 - 140	23
140 - 145	43
145 - 150	21
150 - 155	23
155 - 160	12

Calculate the mean height.

Frequency table and median

We have seen examples of instances where the mean does not give a correct picture of the data. See this table, showing the classification of 25 families in a locality, on the basis of their monthly incomes:

Monthly Income (Rupees)	Number of Families
4000	2
5000	6
6000	7
7000	3
8000	3
9000	2
10000	2
Total	**25**

We get the mean monthly income as Rs. 6520 (Try!) But the table shows that about sixty percent of the families have only a monthly income less than or equal to six thousand rupees. So, the mean is not a very good indicator of this distribution.

How do we compute the median here? Median occurs at the middle, we know. More precisely, here 12 families should have income less than the median income and 12 families should have income more than the median income.

To find this, we need only arrange the incomes in ascending order and find the income of the thirteenth family. From the table, we see that the first two families have an income of 4000 rupees, and the next 6 have an income of 5000 rupees; in other words, when we take the first 8 families, the top income reaches 5000 rupees. What we need is the income of the 13th family in this order. So, we must take the next 5 families. We see that the next 7 families have an income of 6000 rupees. That is, the income of the family numbers 9th to the 15th is 6000 rupees. In particular, this is the income of the 13th family. Thus the median monthly income is Rs. 6000.

Mean and median

We use numbers such as mean and median to get a quick appraisal of data given by a large mass of numbers. (See the section **Ways of statistics** of the lesson, **Statistics** in the Class 9 textbook).

In instances where the distribution has a relatively high frequency at the middle, with frequencies decreasing in more or less the same way on either side, the mean gives a somewhat correct picture of the distribution.

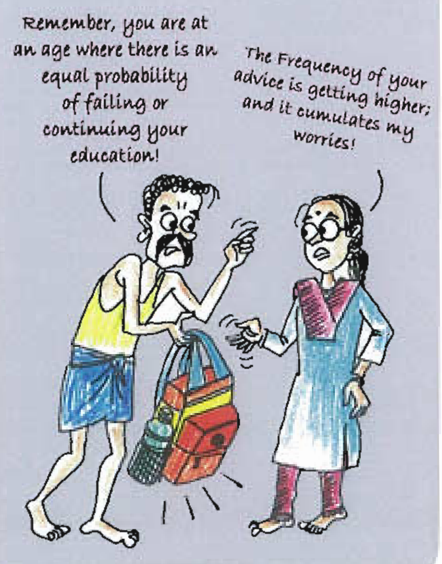
But when the frequency at an end is very high, the mean tends to move towards that part. And it does not give a true picture of the data. In such cases, the median may represent the data better.

Cumulative frequency

We have seen how a table giving classes and their frequencies can be changed to a table giving the frequency upto the upper limit of each class, by adding up the frequencies. These sums of frequencies are called *cumulative frequencies*.

In such instances, we compute the median based on the assumption that at every stage, the change in cumulative frequency is proportional to the change in the numbers given in classes; the median is *defined* as the number corresponding to the cumulative frequency equal to half the total frequency.

Such a notion first arose in connection with probability theory. The problem was to find the age at which one has an equal probability of living or dying, based on a mortality table. For this, the number of people above and below that age must be the same.



To make the computations easier, let's re-write our table as below:

Monthly Income (Rupees)	Number of Families
upto 4000	2
upto 5000	8
upto 6000	15
upto 7000	18
upto 8000	21
upto 9000	23
upto 10000	25

What if the tabulation is done as classes?

See this table, which gives the distribution of heights of children of a certain class in a school:

Height (cm)	Number of Children
135 - 140	4
140 - 145	7
145 - 150	18
150 - 155	11
155 - 160	6
160 - 165	5
Total	**51**

Here also, we first add up the frequencies and make a new table showing the number of children below specific heights:

Height (cm)	Number of Children
below 140	4
below 145	11
below 150	29
below 155	40
below 160	46
below 165	51

In instances such as this, the very definition of median is purely mathematical. For example, in the above example, let's first tabulate the numbers 140, 145, 150, . . . in the first column and 4, 11, 29,... in the second column as below:

x	140	145	150	155	160	165
y	4	11	29	40	46	51

Between the numbers we have taken as x, there are other numbers; and we don't know the numbers y corresponding to these. To compute them, we assume that at every stage, *the change in y is proportional to the corresponding change in x.*

For example, as the variable x changes from 140 to 145, the variable y changes from 4 to 11. So, to compute the y corresponding to $x = 141$, we use our proportionality assumption to get

$$\frac{y-4}{141-140} = \frac{11-4}{145-140}$$

From this, we get

$$y - 4 = \frac{7}{5}$$

and from this

$$y = \frac{27}{5} = 5.4$$

On the other hand, we can use the same technique to find what x must be for a specified value of y. For example, for $y = 41.5$, we must take the x satisfying the equation,

$$\frac{x-155}{160-155} = \frac{41.5-40}{46-40}$$

That is,

$$x = 155 + 5 \times \frac{1.5}{6} = 156.25$$

Now the median. Here the median is, by definition, that x for which

$$y = \frac{51}{2} = 25.5$$

Polynomial of proportionality

If the relation between two quantities is a first degree polynomial of the form $y = ax + b$, then the difference of two numbers taken as x and the difference in the corresponding numbers y would be proportional. For if $y_1 = ax_1 + b$ and $y_2 = ax_2 + b$, then

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{a(x_1 - x_2)}{x_1 - x_2} = a$$

On the other hand, suppose that two quantities, denoted x and y, are related such that the difference of numbers taken as x and the difference of the corresponding numbers y are proportional. Then the algebraic expression giving this relation would be a first degree polynomial. To prove this, let's take the constant of proportionality of this relation as a. Let the number x_1 be related to the number y_1 by this relation. Then for any pair (x, y) of related numbers, we have

$$\frac{y - y_1}{x - x_1} = a$$

That is,

$$y = ax + (y_1 - ax_1)$$

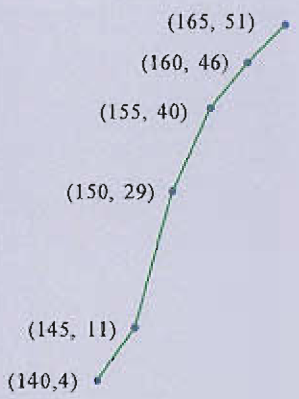
Writing $y_1 - ax_1$ as b, we get

$$y = ax + b$$

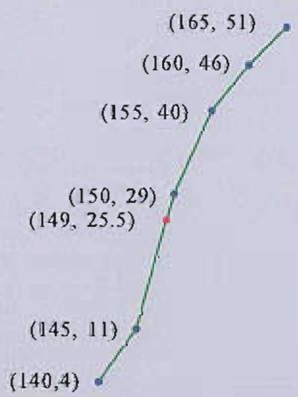
So,we can say that the assumption underlying the computation of the median is that the relation between the quantities and the cumulative frequencies at every stage is a first degree polynomial.

Median picture

If we plot the pairs (x, y) of the height problem as points, and join these by line segments, we get a picture like this:



The median is then the x- coordinate of the point on this with y-coordinate 25.5



The number $y = 25.5$ is between the tabulated numbers $y = 11$ and $y = 29$. And the corresponding numbers are $x = 145$ and $x = 150$. So, as seen in the examples above, corresponding to $y = 25.5$, we must have x such that

$$\frac{x - 145}{150 - 145} = \frac{25.5 - 11}{29 - 11}$$

That is,

$$x = 145 + 5 \times \frac{14.5}{18} \approx 149.03$$

So, in our problem, the median height of the children is 149 centimetres.

Now look at this problem. The table shows the number of workers of a factory classified according to their ages:

Age	Number of Workers
25 - 30	6
30 - 35	8
35 - 40	12
40 - 45	20
45 - 50	16
50 - 55	6
Total	**68**

Let's find the median age. First we make a table showing the numbers of workers below each age:

Age	Number of Workers
below 30	6
below 35	14
below 40	26
below 45	46
below 50	62
below 55	68

Next we look at this as a relation between numbers:

x	30	35	40	45	50	55
y	6	14	26	46	62	68

Here the median is the number x which gives $y = \frac{68}{2} = 34$.

The number $y = 34$ occurs between $y = 26$ and $y = 46$ in the table.

Also, from the table we see that $y = 26$ corresponds to $x = 40$, and $y = 46$ corresponds to $x = 45$. So, as in the first problem, using our proportionality assumption,

$$\frac{x-40}{45-40} = \frac{34-26}{46-26}$$

and this gives

$$x = 40 + \left(5 \times \frac{8}{20}\right) = 42$$

This means, the median age is 42.

Now try these problems on your own:

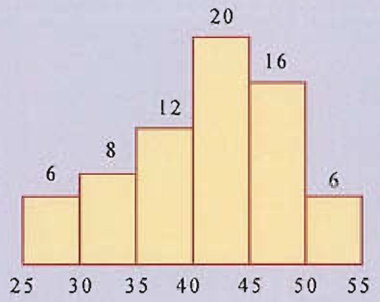
- The table below classifies according to weight, the infants born during a week in a hospital:

Weight (kg)	Number of Infants
2.500	4
2.600	6
2.750	8
2.800	10
3.000	12
3.150	10
3.250	8
3.300	7
3.500	5

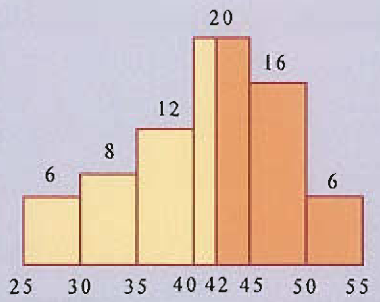
Find the median weight.

Median and area

Remember how we drew histograms of frequency distributions? Here's the histogram of the age problem:



The vertical line through the median splits the picture into two:

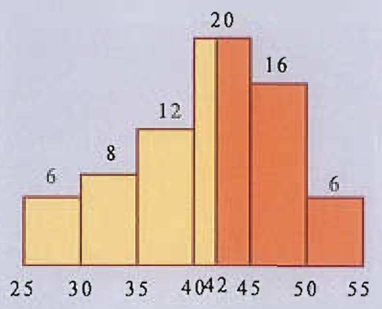


It's not difficult to see that the areas of these two parts are equal. (Try!)

Does the median have this property in all such distributions? Why?

Median and probability

We saw that the vertical line through the median splits the histogram into two parts of equal area.



So, if we mark a point on this picture, the probability of it falling on either part is the same (that is, the probability is $\frac{1}{2}$).

In other words, if we choose a worker of this factory, without any special consideration, the probability of his age to be less than 42, or to be more than 42, is the same.

- The table below shows the number of employees of an office, classified according to the income-tax paid by them:

Income Tax (Rupees)	Number of Employees
1000 - 2000	8
2000 - 3000	10
3000 - 4000	15
4000 - 5000	18
5000 - 6000	22
6000 - 7000	8
7000 - 8000	6
8000 - 9000	3

Compute the median income-tax.

- The table below classifies the candidates who took an examination, according to the marks scored by them:

Marks	Number of Candidates
0 - 10	44
10 - 20	40
20 - 30	35
30 - 40	20
40 - 50	12
50 - 60	10
60 - 70	8
70 - 80	6
80 - 90	4
90 - 100	1

Find the median mark.