ClassRoom

Deciphering the Median Formula - Part 2: from the Ogives

MATHEMATICS CO-DEVELOPMENT GROUP

In Part 1 of this series, we discussed how to obtain the median formula $M = l + \frac{\frac{N}{2} - m}{f} \times c$ for grouped data from the corresponding histogram. In this article, we will discuss how the same formula can be derived from the corresponding ogive, the graph mapping the cumulative frequencies against the upper (or lower) class-limits. Since there are two kinds of ogives viz. (i) less than and (ii) more than, we will show how either can be used for this purpose. In addition, we will discuss why the ogives intersect at the median.

Ogives are a type of statistical graphs introduced at the secondary level. They plot the 'growth' (or 'decay') of the data and are linked with the cumulative frequencies. Let us consider a grouped data of ages of people living in an island (Table 1) and the (less than) cumulative frequencies (Table 2). Similarly, we can also calculate the 'more than' cumulative frequencies (Table 3).

Table 1

Age (in years)	No. of people
0-10	20
10-20	21
20-30	23
30-40	16
40-50	11
50-60	10
60-70	7
70-80	3
80-90	1

Table 2

Age (in years)	Less than Cumulative frequency
≤ 10	20
≤ 20	41
≤ 30	64
≤ 40	80
≤ 50	91
≤ 60	101
≤ 70	108
≤ 80	111
≤ 90	112

Table 3

Age (in years)	More than Cumulative frequency
≥ 0	112
≥ 10	92
≥ 20	71
≥ 30	48
≥ 40	32
≥ 50	21
≥ 60	11
≥ 70	4
≥ 80	1

Keywords: Statistics, median, ogive, reasoning, visualisation, inference

Figure 1 represents the ‘Less than’ ogive that plots the ‘less than’ cumulative frequencies against the data values. In other words, it is the graph that plots the points (10, 20), (20, 41), . . . (90, 112) from Table 2 and connects the consecutive points by line segments. Similarly, Figure 3 shows the ‘More than’ ogive based on the ‘more than’ cumulative frequencies in Table 3.

Let us recall that for ungrouped quantitative data, the median splits the entire data set in two parts – each with the same number of data points. Now, if we extend that to the grouped data, there should be $112 \div 2 = 56$ data points less than the median and 56 greater than it. Or in other words, the cumulative frequency for the median should be 56. Since data values are plotted along the x-axis, median is going to be an x-coordinate. Similarly, since cumulative frequencies are plotted along the y-axis, the y-coordinate, i.e., the cumulative frequency corresponding to the median should be 56. So, if M is the median, the point $(M, 56)$ should be on the ogive.

So, to find M, we draw the line $y = 56$, and find the x-coordinate of the point where this line intersects the ogive. This line intersects the ogive in the line segment connecting the points A (20, 41) and B (30, 64). Note that this is the line segment corresponding to the median class. Let the point of intersection between the ogive and the horizontal line be E. So, $E = (M, 56)$.

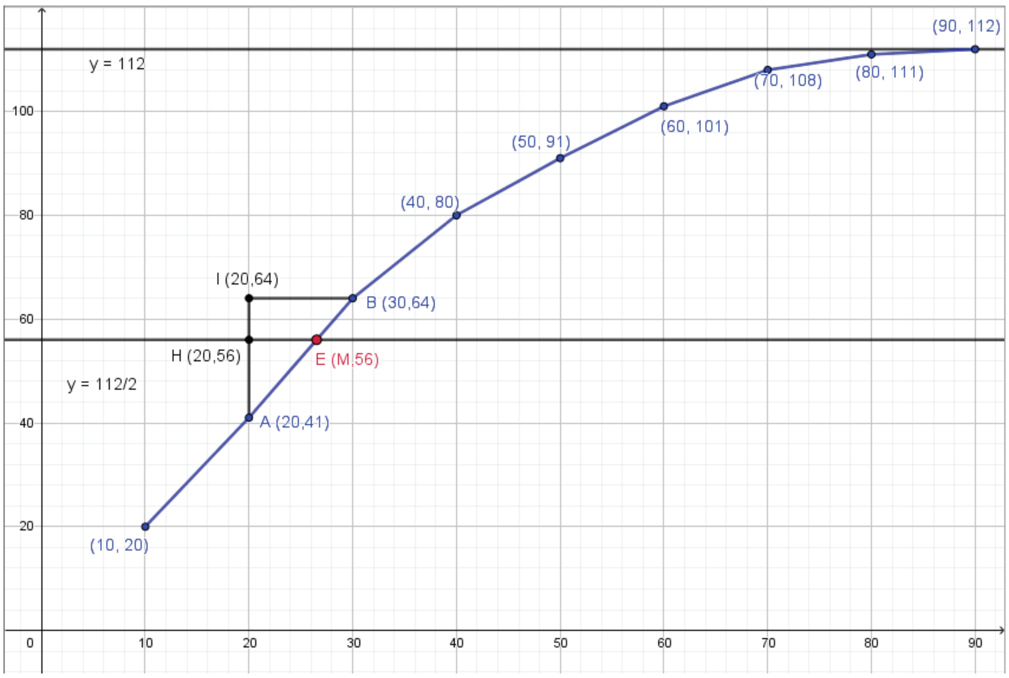

Figure 1. Less than ogive

Now, let us complete the similar right triangles ΔAEH and ΔABI with AE and AB as hypotenuses respectively by drawing their horizontal and vertical legs as shown in Figure 1 and Figure 2. Therefore, $I = (20, 64)$ and $H = (20, 56)$. So, $HE = M - 20$ and $IB = 30 - 20 = 10$, the class-width; while $HA = 56 - 41$ and $IA = 64 - 41 = 23$, the frequency of the median class.

$$\text{Now, } HE : IB = HA : IA \ldots.. \quad (1)$$

$$\Rightarrow (M-20) : 10 = (56-41) : 23 \Rightarrow (M-20) \times 23 = (56-41) \times 10$$

$$\Rightarrow M-20 = (56-41) \times 10/23 \Rightarrow M = 20 + (56-41) \times 10/23$$

$$\Rightarrow M = 20 + (112/2 - 41) \times 10/23 \ldots.. \quad (2)$$

which is very similar to the median formula!

Let us now generalize by using algebra (Figure 2):

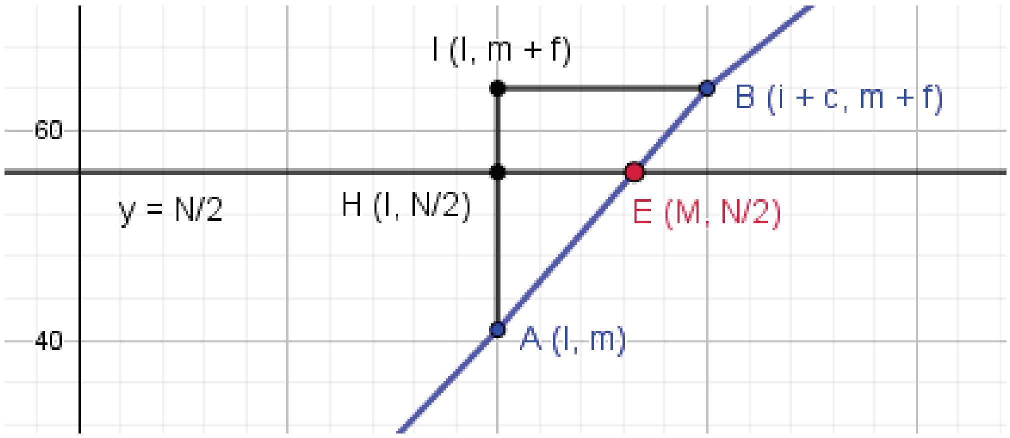

Figure 2.

$$\text{So (1)} \Rightarrow \frac{M-l}{c} = \frac{\frac{N}{2}-m}{f}$$

$$\Rightarrow M - l = \left(\frac{\frac{N}{2}-m}{f}\right) \times c$$

$$\Rightarrow M = l + \left(\frac{\frac{N}{2}-m}{f}\right) \times c \ldots. \quad (3)$$

Table 4		
Symbol	Meaning	In the example
N	Total frequency	112
c	(uniform) class-width	10
l	Lower limit of median class	20
f	Frequency of median class	23
m	(Less than) cumulative frequency for l	41
m'	(More than) cumulative frequency for l	71

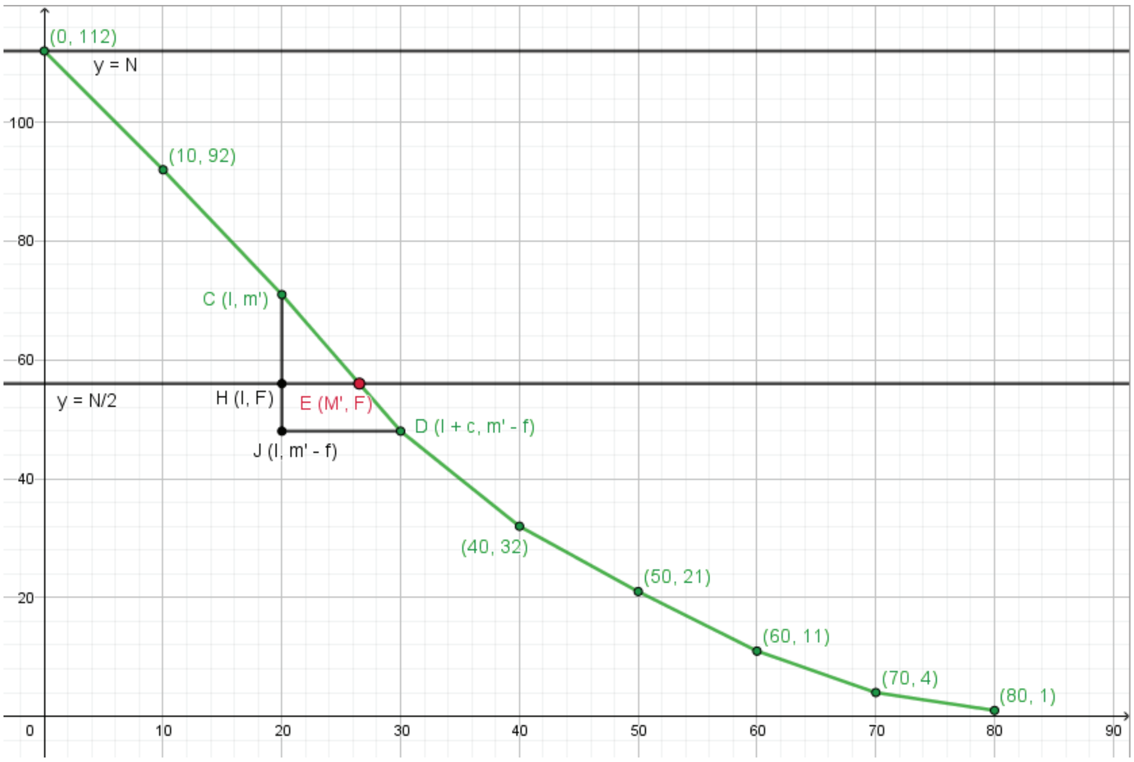

Figure 3. More than ogive

Similarly, we can get the same from 'more than' ogive. Note that this graph (Figure 3) is made by joining the points (0, 112), (10, 92), . . . (80, 1) from Table 3. The line $y = 56$ intersects the line segment CD corresponding to the median class at E (Figure 3 and Figure 4). So, $C = (20, 71)$, $D = (30, 48)$ and $E = (M, 56)$ as before. We complete the similar right triangles ΔCHE and ΔCJD on the hypotenuses CE and CD respectively. So, $H = (20, 56)$ as before and $J = (20, 48)$. Therefore, $HE = M - 20$ (as before), $CH = 71 - 56$, $JD = 30 - 20 = 10$ (class-width), and $CJ = 71 - 48 = 23$ (frequency of the median class).

$$\text{Since } HE : JD = CH : CJ, \text{ we get } (M - 20) : 10 = (71 - 56) : 23$$

$$\Rightarrow M = 20 + (71 - 56) \times 10/23$$

$$\Rightarrow M = 20 + (71 - 112/2) \times 10/23 \ldots . \quad (4)$$

Note that (4) is identical to (2) except 112/2 − 41 being replaced by 71 − 112/2.

To generalize with algebra (Figure 4), we need to add only the 'more than' cumulative frequency m' for the lower limit of the median class l and the rest remain the same.

$$\text{So, (3) generalizes to } M = l + \frac{m' - \frac{N}{2}}{f} \times c \ldots . \quad (5)$$

$\therefore$ we can combine (3) and (5) by $M = l + \frac{|\frac{N}{2} - m|}{f} \times c$ where m is the cumulative frequency corresponding to l

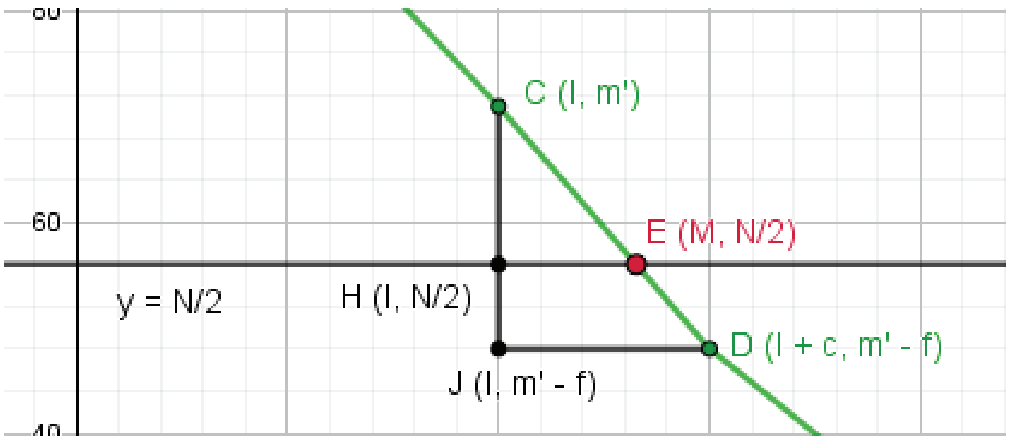

Figure 4.

The last part of the article explores where these two ogives intersect (Figure 5).

Consider the combined Table 5, which is basically Table 1-3 put together. Note that the shaded parts refer to the median class and the corresponding line segments (Figure 5).

Table 5					
Class-Interval	Frequency	Upper Class Limit	Less than CF	Lower Class Limit	More than CF
0-10	20	≤ 10	20	≥ 0	112
10-20	21	≤ 20	41	≥ 10	92
20-30	23	≤ 30	64	≥ 20	71
30-40	16	≤ 40	80	≥ 30	48
40-50	11	≤ 50	91	≥ 40	32
50-60	10	≤ 60	101	≥ 50	21
60-70	7	≤ 70	108	≥ 60	11
70-80	3	≤ 80	111	≥ 70	4
80-90	1	≤ 90	112	≥ 80	1

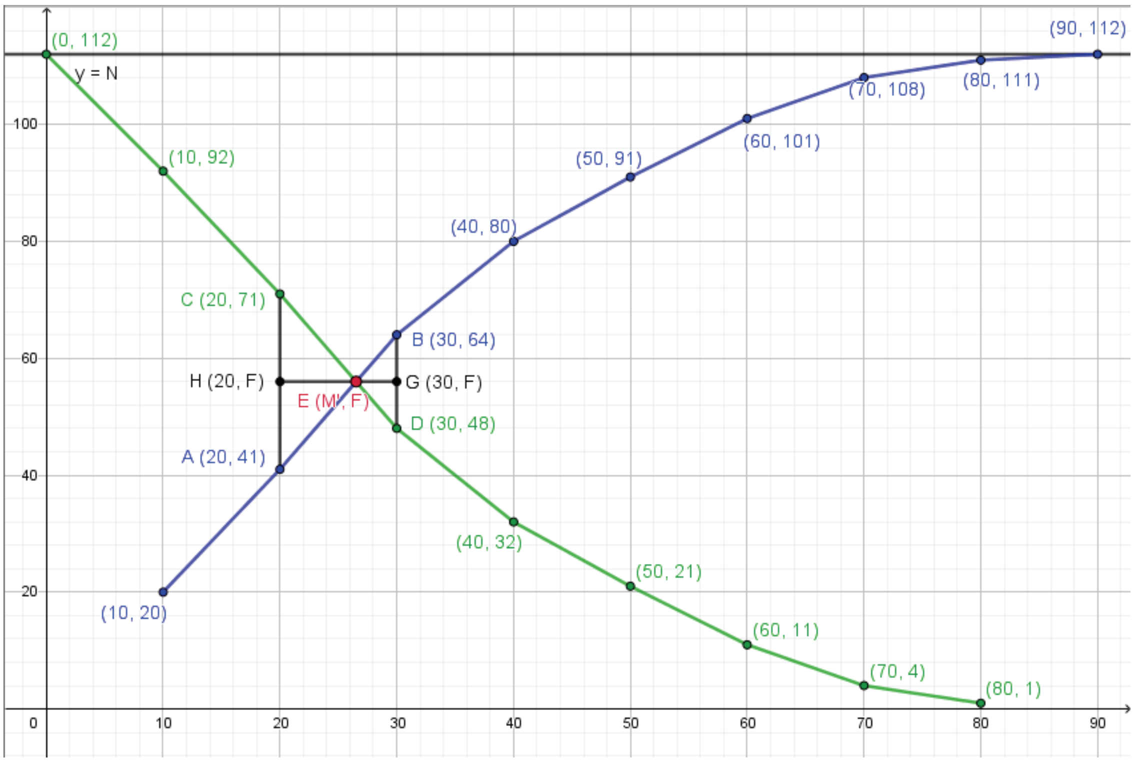

Figure 5.

A, *B*, *C* and *D* are defined as before. So, *AB* and *CD* are the line segments corresponding to the median class in the less than and the more than ogives respectively. *E* (*M'*, *F*) is the point of intersection of *AB* and *CD*. We have to show that E is on the line $y = 56$ i.e. $F = 56$. Then *M'* becomes the median. Join *AC* and *BD*, and let them intersect the line $y = F$ at *H* and *G* respectively.

Since $\Delta AHE \sim \Delta BGE$ and $\Delta CHE \sim \Delta DGE$, we get $HE : GE = AH : BG$ and $HE : GE = CH : DG$

$$\Rightarrow AH : BG = CH : DG \Rightarrow AH \times DG = CH \times BG \ldots . \quad (6)$$

$$\Rightarrow (F - 41) \times (F - 48) = (71 - F) \times (64 - F)$$

$$\Rightarrow F^2 - (41 + 48)F + 41 \times 48 = F^2 - (71 + 64)F + 71 \times 64$$

$$\Rightarrow (71 + 64 - 41 - 48)F = 71 \times 64 - 41 \times 48 \ldots . \quad (7)$$

To generalize with algebra, Table 5 becomes Table 6. Let $a_{k-1} - a_k$ be the median class

$$\text{So } a_{k-1} = l, a_k - a_{k-1} = c, f_k = f, \sum_{i=1}^{k-1} f_i = m, N - \sum_{i=1}^{k-1} f_i = m'$$

$$\text{So, } m' = N - \sum_{i=1}^{k-1} f_i = N - m \ldots \quad (8)$$

Table 6					
Class-Interval	Frequency	Upper Class Limit	Less than CF	Lower Class Limit	More than CF
$a_0 - a_1$	f_1	$< a_1$	f_1	$> a_0$	N
$a_1 - a_2$	f_2	$< a_2$	$f_1 + f_2$	$> a_1$	$N - f_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_{k-2} - a_{k-1}$	f_{k-1}	$< a_{k-1}$	$\sum_{i=1}^{k-1} f_i$	$> a_{k-2}$	$N - \sum_{i=1}^{k-2} f_i$
$a_{k-1} - a_k$	f_k	$< a_k$	$\sum_{i=1}^{k} f_i$	$> a_{k-1}$	$N - \sum_{i=1}^{k-1} f_i$
$a_k - a_{k+1}$	f_{k+1}	$< a_{k+1}$	$\sum_{i=1}^{k+1} f_i$	$> a_k$	$N - \sum_{i=1}^{k} f_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_{n-1} - a_n$	f_n	$< a_n$	$\sum_{i=1}^{n} f_i = N$	$> a_{n-1}$	$N - \sum_{i=1}^{n-1} f_i$

So, $A = \left(a_{k-1}, \sum_{i=1}^{k-1} f_i\right) = (l, m)$ and $B = \left(a_k, \sum_{i=1}^{k} f_i\right) = (l + c, m + f)$ in the less than ogive while $C = \left(a_{k-1}, N - \sum_{i=1}^{k-1} f_i\right) = (l, m')$ and $D = \left(a_k, N - \sum_{i=1}^{k} f_i\right) = (l + c, m' - f)$ in the more than ogive.

AB and CD intersect at $E\,(M', F)$. We need to show that $F = \frac{N}{2}$ since that will imply that $M' =$ median.

AC and BD intersect $y = F$ at $H\,(l, F)$ and $G\,(l + c, F)$ respectively (Figure 6).

So, (6): $AH \times DG = CH \times BG \Rightarrow (F - m) \times (F - m' + f) = (m' - F) \times (m + f - F)$

$$\Rightarrow F^2 - \left(m + m' - f\right) F + m\left(m' - f\right) = F^2 - \left(m' + m + f\right) F + m'(m + f)$$

$$\Rightarrow \left(m' + m + f - m - m' + f\right) F = m'(m + f) - m\left(m' - f\right)$$

$$\Rightarrow 2fF = \left(m' + m\right) f = Nf \ldots \text{by (8)} \Rightarrow F = \frac{N}{2}$$

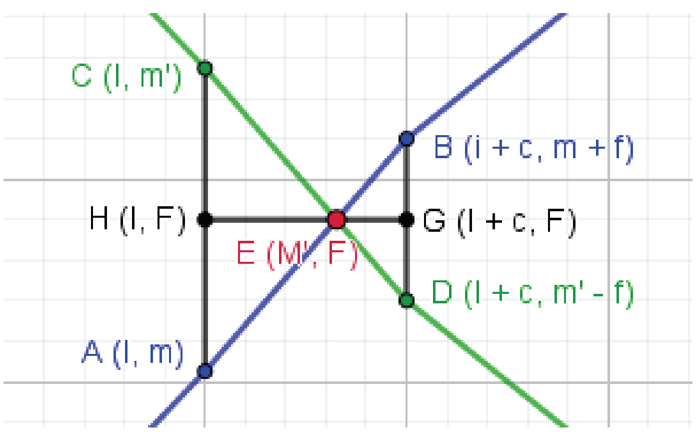

Figure 6.

We would like to observe how the ogives are symmetric about the line $y = \frac{N}{2}$ (Figure 5). This can be explained as follows:

- For any ogive, each line segment represents a particular class-interval
- So, except for the first and last line segments, the remaining ones in less than and more than ogives are paired
- Now, each such pair of line segments have the exact same run $= a_i - a_{i-1}$, upper class limit − lower class limit of the i^{th} class
- And each such pair has opposite rises, f_i i.e. the frequency of the i^{th} class for the less than ogive and $-f_i$ for the more than ogive

Even though the median formula is quite complicated and ogive is a new graph at the secondary level, note that none of the derivations require any sophisticated techniques. They were based on basic coordinate geometry viz. coordinates of points, lengths of horizontal and vertical line segments (which are essentially difference of x- or y-coordinates respectively), and similar triangles − all of which are part of secondary mathematics syllabus. If students can be taught to use the basic principle, then they can easily draw the relevant part of the less than ogive, as in Figure 2, for any given data and compute the median without having to remember the formula.

In the next article, we shall look into another such complicated formula − that of the mode.

Math Co-dev Group or more elaborately **MATHEMATICS CO-DEVELOPMENT GROUP** is an internal initiative of Azim Premji Foundation where math resource persons across states put their heads together to prepare simple materials for teachers to develop their understanding on different content areas and how to transact the same in their classrooms. It is a collaborative learning space where resources are collected from multiple sources, critiqued and explored in detail. Math Co-dev Group may be reached through yashvendra@azimpremjifoundation.org.