NOTE: To understand these concepts I highly recommend reading chapter 14 from class 10 NCERT book.

Mean of Grouped Data

The mean (or average) of observations, as we know, is the sum of the values of all the observations divided by the total number of observations.

Now, the sum of the values of all the observations $= f_1x_1 + f_2x_2 + \ldots + f_nx_n$, and the number of observations $= f_1 + f_2 + \ldots + f_n$.

So, the mean $\overline{x}$ of the data is given by

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}$$

Recall that we can write this in short form by using the Greek letter Σ (capital sigma) which means summation. That is,

$$\overline{x} = \frac{\sum\limits_{i=1}^{n} f_ix_i}{\sum\limits_{i=1}^{n} f_i}$$

which, more briefly, is written as $\overline{x} = \dfrac{\Sigma f_ix_i}{\Sigma f_i}$, if it is understood that i varies from 1 to n.

Frequency Polygon

A frequency polygon is almost identical to a histogram, which is used to compare sets of data or to display a cumulative frequency distribution. It uses a line graph to represent quantitative data.

Steps to Draw Frequency Polygon

Step 1- Choose the class interval and mark the values on the horizontal axes
Step 2- Mark the mid value of each interval on the horizontal axes.
Step 3- Mark the frequency of the class on the vertical axes.
Step 4- Corresponding to the frequency of each class interval,
mark a point at the height in the middle of the class interval
Step 5- Connect these points using the line segment.
Step 6- The obtained representation is a frequency polygon.

Now, for each class-interval, we require a point which would serve as the representative of the whole class. *It is assumed that the frequency of each class-interval is centred around its mid-point.* So the *mid-point* (or *class mark*) of each class can be chosen to represent the observations falling in the class. Recall that we find the mid-point of a class (or its class mark) by finding the average of its upper and lower limits. That is,

$$\boxed{\textbf{Class mark} =} \frac{\textbf{Upper class limit + Lower class limit}}{\textbf{2}}$$

Cumulative Frequency Distribution

DEFINITION: The cumulative frequency is the total of frequencies, in which
the frequency of the first class interval is added to frequency of the second
class interval and then the sum is added to the frequency of the third
class interval and so on. Hence, the table that represents the cumulative
frequencies that are divided over different classes is called the
cumulative frequency table or cumulative frequency distribution.

Steps to Construct Less than Cumulative Frequency Curve

1. Mark the upper limit on the horizontal axis or x-axis.
2. Mark the cumulative frequency on the vertical axis or y-axis.
3. Plot the points (x, y) in the coordinate plane where x represents the upper
limit value and y represents the cumulative frequency.
4. Finally, join the points and draw the smooth curve.
5. The curve so obtained gives a cumulative frequency distribution graph of
less than type.

MODE of Grouped Data

A mode is that value among the observations which occurs most often, that is, the value of the observation having the maximum frequency.

In a grouped frequency distribution, it is not possible to determine the mode by looking at the frequencies. Here, we can only locate a class with the maximum frequency, called the modal class. The mode is a value inside the modal class, and is given by the formula:

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

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.

Median of Grouped Data

The median is a measure of central tendency which gives the value of the middle-most observation in the data. Recall that for finding the median of ungrouped data, we first arrange the data values of the observations in ascending order. Then, if n is odd, the median is (n+1)/2 th observation. And, if n is even, then the median will be the average of the (n/2) th and the (n/2)+1 th observations.

we find the cumulative frequencies of all the classes and n/2. We now locate the class whose cumulative frequency is greater than (and nearest to) n/2. This is called the median class.

After finding the median class, we use the following formula for calculating the median.

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

where $l =$ lower limit of median class,

                  $n =$ number of observations,

            cf $=$ cumulative frequency of class preceding the median class,

                 $f =$ frequency of median class,

                 $h =$ class size (assuming class size to be equal).

REMARKS:

1. There is a empirical relationship between the three measures of central tendency :

$$3 \text{ Median} = \text{Mode} + 2 \text{ Mean}$$

2. The median of grouped data with unequal class sizes can also be calculated. However, we shall not discuss it here.

3. Read tips and tricks as well as concepts pdf files.

4. Practice a lot.