1. INTRODUCTION

Suppose we want to compare the income distribution of workers in two factories and determine which factory pays more to its workers. If we compare on the basis of individual workers, we cannot conclude anything. However, if for the given data, we get a representative value that signifies the characteristics of the data, the comparison becomes easy.

It is generally observed that observations or data on a variable tend to gather around some central value. This gathering of data towards a central value is called **central tendency** or the middle value of the distribution, also known as middle of the data set.

A certain value representative of the whole data and signifying its characteristics is called an **average** of the data.

Three types of averages are useful for analyzing data.

They are : (i) **Mean**, (ii) **Median**, (iii) **Mode**.

2. MEAN FOR A GROUPED FREQUENCY DISTRIUBTION

2.1 DIRECT METHOD

Step 1: For each class, find the class mark x_i, as

$x_i = \frac{1}{2}$ (lower limit + upper limit)

Step 2: Calculate $f_i x_i$ for each i.

Step 3: Use the formula : $= \frac{\sum(f_i x_i)}{\sum f_i}$; $\bar{x} = \frac{f_1 \times 1 + f_2 \times 2}{f_1 + f_2 + f_3 + ... f_n}$

3. ASSUMED-MEAN METHOD

Following steps are taken to solve cases by assumed-mean method.

Step 1: For each class interval, calculate the class mark x_i by using the formula: $x_i = \frac{1}{2}$ (lower limit + upper limit).

Step 2: Choose a value of x_i in the middle as the assumed mean and denote it by A.

Step 3: Calculate the deviations $d_i = (x_i - A)$ for each i.

Step 4: Calculate the $(f_i d_i)$ for each i.

Step 5: Find $n = \sum f_i$.

Step 6: Calculate the mean, $\bar{x}$, by using the formula: $\bar{x} = A + \frac{\sum f_i d_i}{n}$.

4. STEP-DEVIATION METHOD

Following steps are taken to solve cases by step-deviation method.

Step 1: For each class interval, calculate the class mark x_i by using the formula: $x_i = \frac{1}{2}$ (lower limit + upper limit).

Step 2: Choose a value of x_i in the middle of the x_i column as the assumed mean and denote it by A.

Step 3: Calculate h = [(upper limit) – (lower limit)].

Step 4: Calculate $u_i = \frac{(x_i - A)}{h}$ for each class.

Step 5: Calculate f_iu_i for each class and find $\sum(f_iu_i)$.

Step 6: Calculate the mean, by using the formula: $\bar{x} = A + \left[h\frac{\sum(f_iu_i)}{\sum f_i}\right]$.

5. MEAN FOR AN INCLUSIVE SERIES

6. MODE

It is that value of a variate which occurs most often. More precisely, mode is that value of the variable at which the concentration of the data is maximum.

Modal Class : In a frequency distribution, the class having maximum frequency is called the modal class.

Formula for Calculating Mode:

We have:

Mode, $M_0 = \ell + h.\left[\frac{(f_1 - f_0)}{(2f_1 - f_0 - f_2)}\right]$,

where

ℓ = lower limit of the modal class,

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;

h = width of the class interval (assuming all class width to be equal)

7. METHOD FOR FINDING THE MEDIAN FOR GROUPED DATA

Median of a distribution is the value of the variable which divides it into two equal parts. It is the value which exceeds and is exceeded by the same number of observation i.e., it is the value such that the number of observation above it is equal to the number of observation below it.

In case of grouped frequency distribution, the class corresponding to the cumulative (c.f) just greater than $\frac{N}{2}$ is called the **median class**.

Following steps are involved in finding the median of the given frequency distribution.

Step 1: For the given frequency distribution, prepare the cumulative frequency table and obtain $N = \Sigma f_i$.

Step 2: Find $(N / 2)$.

Step 3: Find the cumulative frequency just greater than $(N / 2)$ and find the corresponding class, known as median class.

Step 4: Use the formula:

Median, $$Me = \ell + \left[h \times \frac{\left(\frac{N}{2} - c\right)}{f} \right],$$

Where,

ℓ = lower limit of median class,
h = width of median class, (class size)
f = frequency of median class,
c = cumulative frequency of the class preceding the median class,
$N = \Sigma f_i$. (number of observations)

8. RELATIONSHIP AMONG MEAN, MEDIAN AND MODE

We have, Mode = 3(Median) −2(Mean)

OR

Median = Mode $+ \frac{2}{3}$ (Mean − Mode)

OR

Mean = Mode $+ \frac{3}{2}$ (Median − Mode)

9. GRAPHICAL PRESENTATION OF CUMULATIVE FREQUENCY DISTRIBUTION

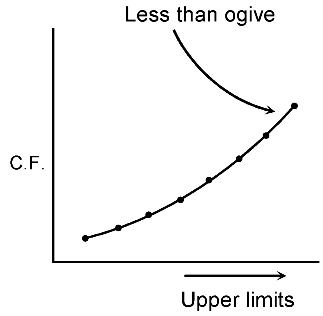
Let a grouped frequency distribution be given to us.

9.1 FOR A 'LESS THAN' SERIES

On a graph paper, mark the upper class limits along the *x*-axis and the corresponding cumulative frequencies along the *y*-axis.

(i) On joining these points successively by line segments, we get a polygon, called cumulative frequency polygon.

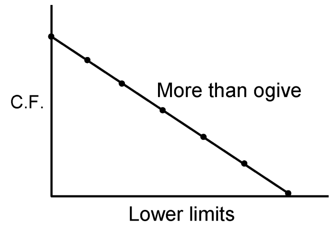
(ii) On joining these points successively by smooth curves, we get a curve, known as cumulative frequency curve or an ogive.



9.2 FOR A 'GREATER (MORE) THAN' SERIES

On a graph paper, mark the lower class limits along the *x*-axis and the corresponding cumulative frequencies along the *y*-axis.

(i) On joining these points successively by line segments, we get a polygon, called cumulative frequency polygon.



(ii) On joining these points successively by smooth curves, we get a curve, known as cumulative frequency curve or an ogive.