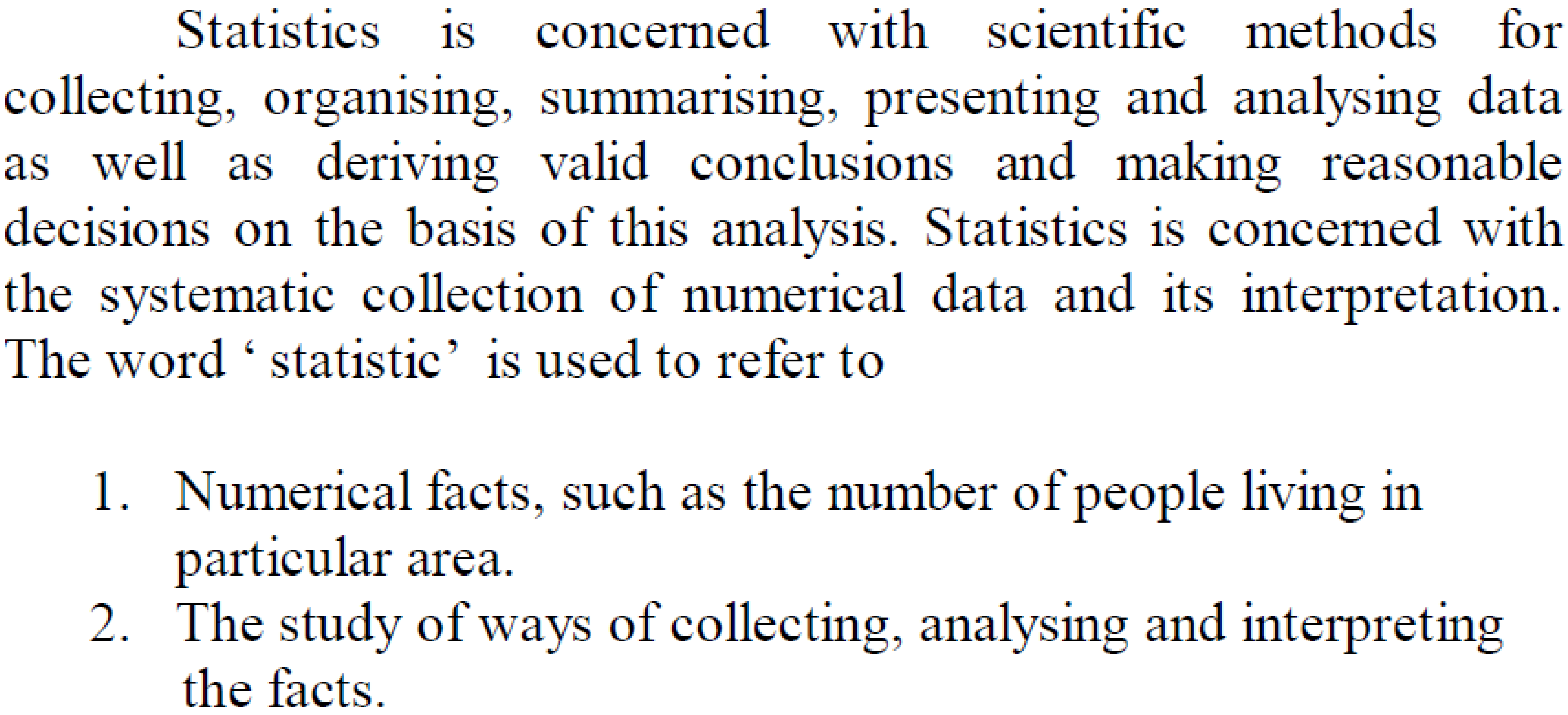
Statistics

Statistics



Statistics is concerned with scientific methods for collecting, organising, summarising, presenting and analysing data as well as deriving valid conclusions and making reasonable decisions on the basis of this analysis. Statistics is concerned with the systematic collection of numerical data and its interpretation. The word ‘ statistic’ is used to refer to

1. Numerical facts, such as the number of people living in particular area.
2. The study of ways of collecting, analysing and interpreting the facts.

Statistics

*“**Statistics** is the mathematical science involving the collection, analysis and interpretation of data”*

“Statistics are the classified facts representing the conditions of people in a state. In particular they are the facts, which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement”

“Statistics are measurements, enumerations or estimates of natural phenomenon usually systematically arranged, analysed and presented as to exhibit important interrelationships among them”

Collection of Data

1. **Collection of Data:** It is the first step and this is the foundation upon which the entire data set. Careful planning is essential before collecting the data. There are different methods of collection of data such as census, sampling, primary, secondary, etc., and the investigator should make use of correct method.

Presentation of data

The mass data collected should be presented in a suitable, concise form for further analysis. The collected data may be presented in the form of tabular or diagrammatic or graphic form.

Analysis of data

The data presented should be carefully analysed for making inference from the presented data such as measures of central tendencies, dispersion, correlation, regression etc.,

Interpretation of data

The final step is drawing conclusion from the data collected. A valid conclusion must be drawn on the basis of analysis. A high degree of skill and experience is necessary for the interpretation.

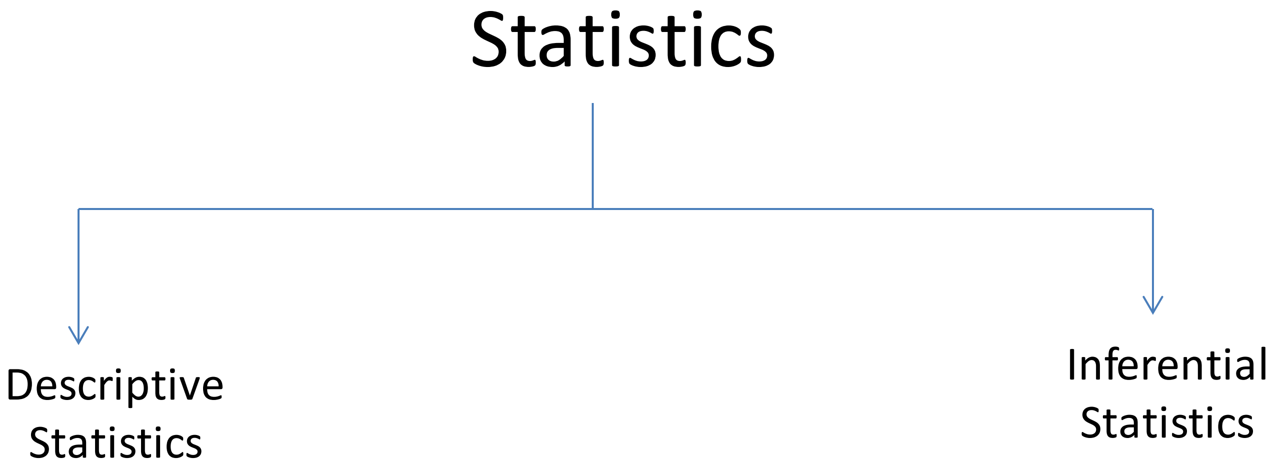
List of fields of application of statistics

- **Actuarial science** is the discipline that applies mathematical and statistical methods to assess risk in the insurance and finance industries.
- **Astrostatistics** is the discipline that applies statistical analysis to the understanding of astronomical data.
- **Biostatistics** is a branch of biology that studies biological phenomena and observations by means of statistical analysis, and includes medical statistics.
- **Business analytics** is a rapidly developing business process that applies statistical methods to data sets (often very large) to develop new insights and understanding of business performance & opportunities
- **Chemo-metrics** is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods.
- **Demography** is the statistical study of all populations. It can be a very general science that can be applied to any kind of dynamic population, that is, one that changes over time or space.
- **Econometrics** is a branch of economics that applies statistical methods to the empirical study of economic theories and relationships.
- **Environmental statistics** is the application of statistical methods to environmental science. Weather, climate, air and water quality are included, as are studies of plant and animal populations.
- **Epidemiology** is the study of factors affecting the health and illness of populations, and serves as the foundation and logic of interventions made in the interest of public health and preventive medicine.
- **Geostatistics** is a branch of geography that deals with the analysis of data from disciplines such as petroleum geology, hydrogeology, hydrology, meteorology, oceanography, geochemistry, geography.
- **Machine Learning**

List of fields of application of statistics

- **Operations research** (or Operational Research) is an interdisciplinary branch of applied mathematics and formal science that uses methods such as mathematical modeling, statistics, and algorithms to arrive at optimal or near optimal solutions to complex problems.
- **Population ecology** is a sub-field of ecology that deals with the dynamics of species populations and how these populations interact with the environment.
- **Psychometric** is the theory and technique of educational and psychological measurement of knowledge, abilities, attitudes, and personality traits.
- **Quality control** reviews the factors involved in manufacturing and production; it can make use of statistical sampling of product items to aid decisions in process control or in accepting deliveries.
- **Quantitative psychology** is the science of statistically explaining and changing mental processes and behaviors in humans.
- **Reliability Engineering** is the study of the ability of a system or component to perform its required functions under stated conditions for a specified period of time
- **Statistical finance**, an area of econophysics, is an empirical attempt to shift finance from its normative roots to a positivist framework using exemplars from statistical physics with an emphasis on emergent or collective properties of financial markets.
- **Statistical mechanics** is the application of probability theory, which includes mathematical tools for dealing with large populations, to the field of mechanics, which is concerned with the motion of particles or objects when subjected to a force.
- **Statistical physics** is one of the fundamental theories of physics, and uses methods of probability theory in solving physical problems.
- **Statistical Signal Processing**
- **Statistical thermodynamics** is the study of the microscopic behaviors of thermodynamic systems using probability theory and provides a molecular level interpretation of thermodynamic quantities such as work, heat, free energy, and entropy.

Statistics



Statistics

Descriptive Statistics

Inferential Statistics

Statistics

- **Descriptive Statistics:** Collection, Organization, summarization and presentation of data.

- **Inferential Statistics:** Generalizing from sample to population, performing estimations and hypothesis testing, and making predictions.

Data Collection

For statistical analysis, whether it is business, economics, social sciences, science, or other fields, the basic problem is to collect facts and figures relating to particular phenomenon under study.

Data Collection

- ❑ Objectives and scope of the enquiry.
- ❑ Statistical units to be used.
- ❑ Sources of information.
- ❑ Methods of data collection.

Data Collection

Primary Data: Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study.

Such data is original in character and is generated by survey conducted by individuals or research institution or any organisation.

Secondary Data: Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency.

Data Classification

“Classification is the process of arranging the data into sequences and groups according to their common characteristics, or separating them into different but related parts” – Secrist.

“A classification is a scheme for breaking a category into a set of parts, called classes, according to some precisely defined differing characteristics possessed by all the elements of the category” – Tuttle A.M.

Data Classification

- Data is collected for the purpose of analysis.
- Data collected in any statistical investigation is known as raw data.
- **Having collected and edited the data, the next important step is to organise it.** i.e. to present in a readily comprehensible condensed form which will highlight the important characteristics of the data.

Data Classification

The collected data, also known as raw data or ungrouped data are always in an un organised form and need to be organised and presented in meaningful and readily comprehensible form in order to facilitate further statistical analysis.

Objects of Classification:

✓ It condenses the mass of data.

✓ It eliminates unnecessary details.

✓ It facilitates comparison and highlights the significant aspect of data.

✓ It enables one to get a mental picture of the information and helps in drawing inferences.

✓ It helps in the statistical treatment of the information collected.

Data Classification

Geographical [Area-wise or regional]

Country	Average Output
India	123.5
USA	345.6
Pakistan	201.45
China	856.89
Sudan	123.4
Russia	345.5
Bangladesh	456.43

Classification

Chronological Classification:

Population of India (In crores)

Year	Population
1901	23.8
1911	25.0
1921	25.2
1931	27.9
1941	31.9
1951	43.9
1961	68.2

Data Classification

Qualitative classification:

✓ Genius

✓ Highly Intelligent

✓ Average Intelligent

✓ Below average

✓ Dull

Data Classification

- Quantities Classification:

Daily earnings	Number of stores
1-100	23
101-200	25
201-300	25
301-400	27
401-500	31
501-600	43
601-700	68

Data Classification

Why Classification?

✓It condenses the data.

✓It facilitates comparison.

✓It helps to study the relationships.

Data Classification

The statistical data collected are generally raw data or ungrouped data.

Monthly salary (Rs. In ,000) of 30 employees in a Company -

80, 70, 55, 50, 60, 65, 40, 30, 80, 90, 75, 45, 35, 65, 70, 80, 82, 55, 65, 80, 60, 55, 38, 65, 75, 85, 90, 65, 45, 75

- ✓ *The above figures are nothing but raw or ungrouped data.*
- ✓ *This representation of data does not furnish any useful information.*

Data Classification

Ascending order -

30, 35, 38, 40, 45, 45, 50, 55, 55, 55, 60, 60, 65, 65, 65, 65, 65, 65, 70, 70, 75, 75, 75, 80, 80, 80, 80, 85, 90, 90.

Advantages-

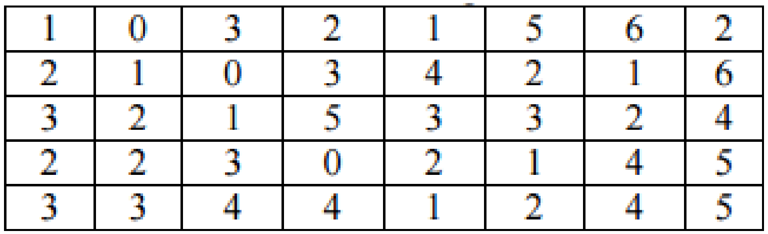
- ✓ Maximum and minimum values.
- ✓ It also gives a rough idea of the distribution of the items over the range .

Frequency table

Average salary	No. of Emp.
30-40	3
40-50	3
50-60	5
60-70	7
70-80	5
80-90	7
Total	**30**

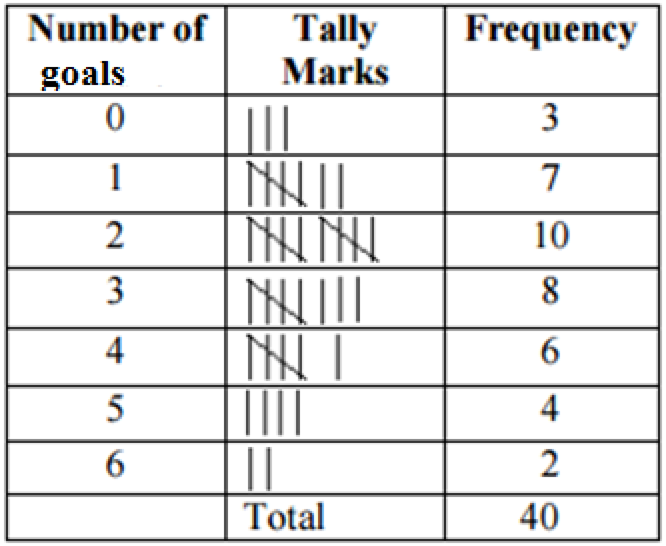
Discrete /Ungrouped frequency distribution

- ✓ A Survey of 40 football matched was conducted and following data obtained.



1	0	3	2	1	5	6	2
2	1	0	3	4	2	1	6
3	2	1	5	3	3	2	4
2	2	3	0	2	1	4	5
3	3	4	4	1	2	4	5

Frequency distribution of the number of goals



Number of goals	Tally Marks	Frequency
0	𝍷𝍷𝍷	3
1	𝍸 𝍷𝍷	7
2	𝍸 𝍸	10
3	𝍸 𝍷𝍷𝍷	8
4	𝍸 𝍷	6
5	𝍷𝍷𝍷𝍷	4
6	𝍷𝍷	2
	Total	40

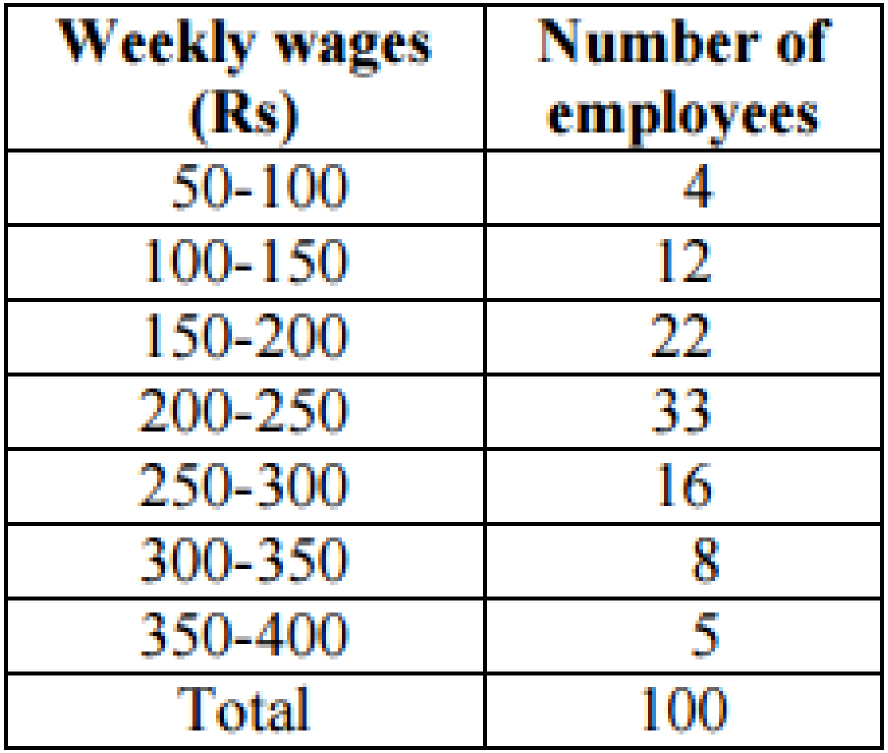
Discrete /Ungrouped frequency distribution

- Frequency refers to discrete value.
- Data are presented in a way that exact measurement of units are clearly indicated.
- Each class is distinct and separate from the other class.
- Non-continuity from one class to another class exist.

Continuous frequency distribution

- Continuous frequency distribution refers to groups of values.
- Advantage – Random variable can take any fractional value and the same can be presented in the form of contagious frequency distribution.

Continuous Frequency Distribution



Weekly wages (Rs)	Number of employees
50-100	4
100-150	12
150-200	22
200-250	33
250-300	16
300-350	8
350-400	5
Total	100

Definitions

- Class limits- The class limits are the lowest and the highest values that can be included in the class. For example, take the class 30-40. The lowest value of the class is 30 and highest class is 40.
- Class Interval: The class interval may be defined as the size of each grouping of data. For example, 50-75, 75-100, 100-125… are class intervals.
- Width or size of the class interval: The difference between the lower and upper class limits is called Width or size of class interval and is denoted by ‘ C’ .
- Range: The difference between largest and smallest value of the observation is called The Range and is denoted by ‘ R’ ie R = Largest value – Smallest value R = L - S

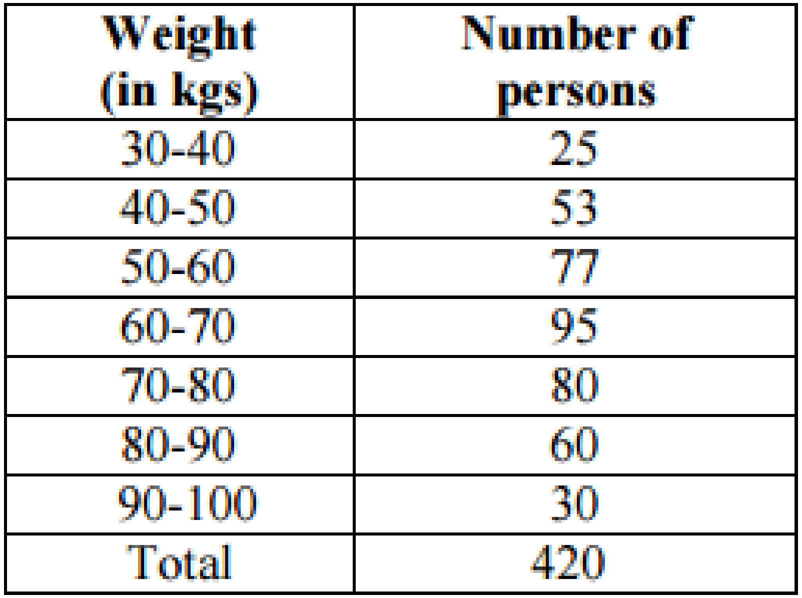
Definitions

Mid-value or mid-point:

- The central point of a class interval is called the mid value or mid-point. It is found out by adding the upper and lower limits of a class and dividing the sum by 2.
- Mid-Value = (L+ U)/ 2
- For example, if the class interval is 20-30 then the mid-value is (20 +30)/ 2 + = 25

Definitions

Frequency: Number of observations falling within a particular class interval is called ***frequency*** of that class.



Weight (in kgs)	Number of persons
30-40	25
40-50	53
50-60	77
60-70	95
70-80	80
80-90	60
90-100	30
Total	420

Frequency

Number of class intervals

- The number of class interval in a frequency is matter of importance.
- The number of class interval should not be too many.
- For an ideal frequency distribution, the number of class intervals can vary from 5 to 15.
- To decide the number of class intervals for the frequency distributive in the whole data, we choose the lowest and the highest of the values. The difference between them will enable us to decide the class intervals.

Number of class intervals:

- ✓ Decision the number of class groupings depends largely on the judgement of the individual investigator and/or the range that will be used to group the data.
- ✓ Following two rules are often used to decide approximate number of classes in a frequency distribution:

Number of class intervals:

(a) If k represents the number of classes and N the total number of observations, then the value of k will be the samllest exponent of the number 2, so that 2^k>= N.

e.g. If N = 30 observations.

2^5 = 32 (> 30) Thus we may choose k = 5

Number of class intervals

(b) Sturges’ Rule $K = 1 + 3. 322 \log10 N$

Where -

N = Total number of observations.

K = Number of class intervals.

Thus if the number of observation is 10, then the number of class intervals is $K = 1 + 3. 322 \log 10 = 4.322$

If 100 observations are being studied, the number of class interval is $K = 1 + 3. 322 \log 100 = 7.644$ @ 8

Let us consider the weights in kg of 50 college students.

42	62	46	54	41	37	54	44	32	45
47	50	58	49	51	42	46	37	42	39
54	39	51	58	47	64	43	48	49	48
49	61	41	40	58	49	59	57	57	34
56	38	45	52	46	40	63	41	51	41

Here the size of the class interval as per sturges rule is obtained as follows

$$\text{Size of class interval} = C = \frac{\text{Range}}{1+3.322\log N}$$

$$= \frac{64 - 32}{1+3.322\log(50)} = \frac{32}{6.64} \qquad 5$$

Thus the number of class interval is 7 and size of each class is 5. The required size of each class is 5. The required frequency distribution is prepared using tally marks as given below:

Class Interval	**Tally marks**	**Frequency**
30-35	𝍷𝍷	2
35-40	𝍸 𝍷	6
40-45	𝍸 𝍸 𝍷𝍷	12
45-50	𝍸 𝍸 𝍷𝍷𝍷𝍷	14
50-55	𝍸 𝍷	6
55-60	𝍸 𝍷	6
60-65	𝍷𝍷𝍷𝍷	4
Total		50

Size of the class interval

- Since the size of the class interval is inversely proportional to the number of class interval in a given distribution.

Size of class interval = C

C = Range/ Number of class intervals

= Range /1+3.322*$\log_{10}$N

*where Range = Largest Value – smallest value in the distribution

Types of class intervals

- Exclusive
- Inclusive
- Open-end

Exclusive method

✓When the data are classified in such a way that the upper limit of a class interval is the lower limit of the succeeding class interval (i.e. no data point falls into more than one class interval), then it is said be the exclusive method of classifying data.

Exclusive method

Dividents declared in %	No. of companies
0-10	5
10-20	6
20-30	10
30-40	5
40-50	3

- ✓ 5 companies declared dividends ranging from 0 to 10 percent.

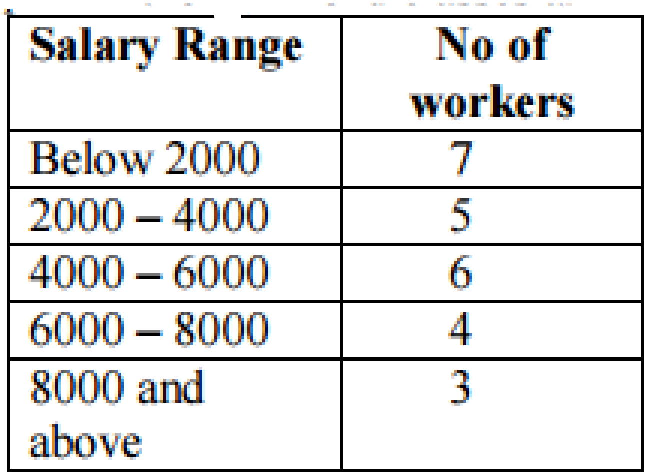
Company which declared exactly 10 % dividend would not be included in the class 0-10 but would be included in the next class 10-20.

Inclusive Method

✓ When the data are classified in such a way that both lower and upper limits of a class interval are included in the interval itself, then it is said to be the inclusive method of classifying data.

Class intervals	Frequency
0-4	5
5-9	22
10-14	13
15-19	8
20-24	2

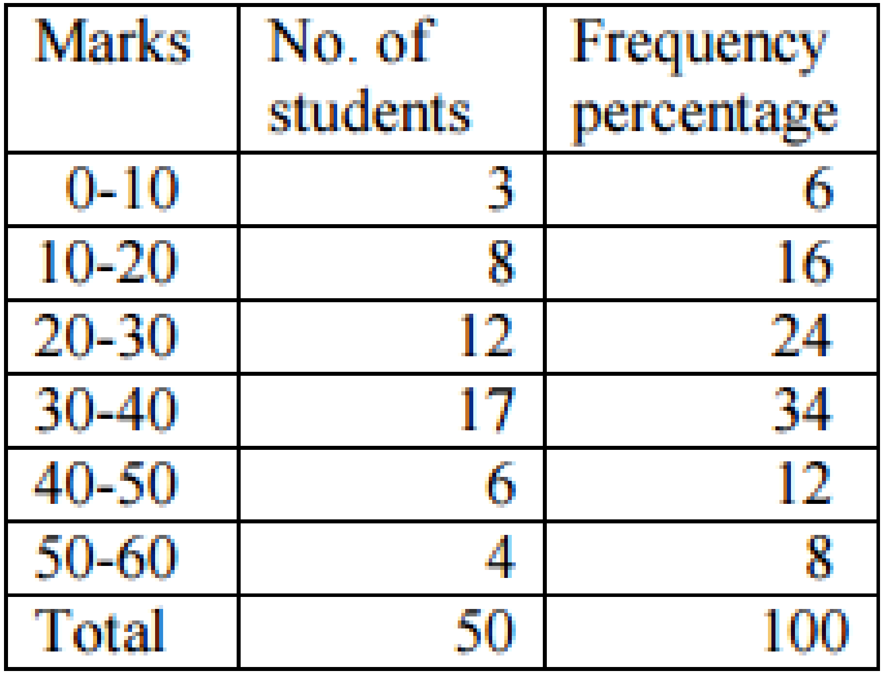
Open end classes



Salary Range	No of workers
Below 2000	7
2000 – 4000	5
4000 – 6000	6
6000 – 8000	4
8000 and above	3

Percentage frequency table

$$\text{Frequency percentage} = \frac{\text{Actual Frequency}}{\text{Total Frequency}} \times 100$$

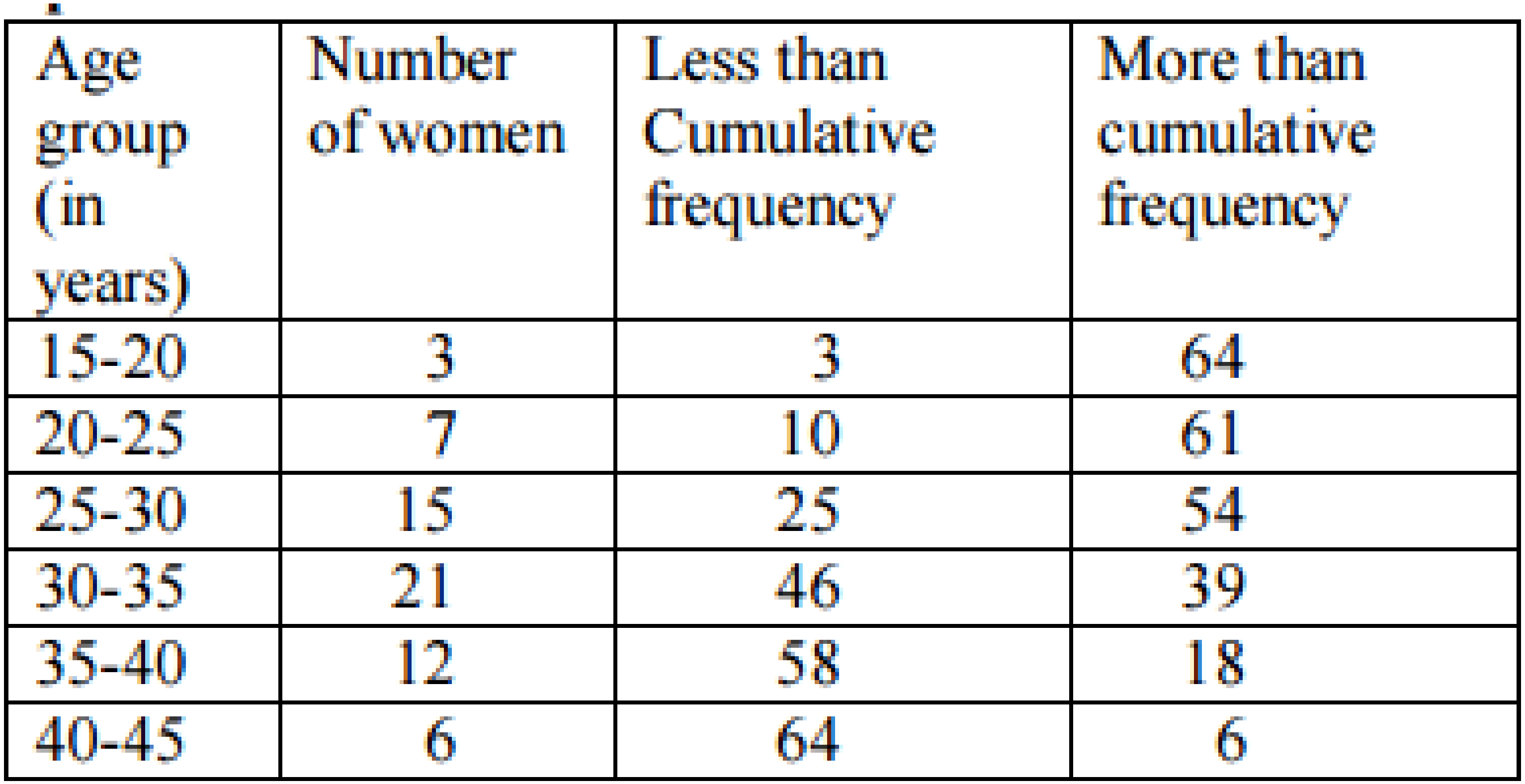


Marks	No. of students	Frequency percentage
0-10	3	6
10-20	8	16
20-30	12	24
30-40	17	34
40-50	6	12
50-60	4	8
Total	50	100

Cumulative frequency table:

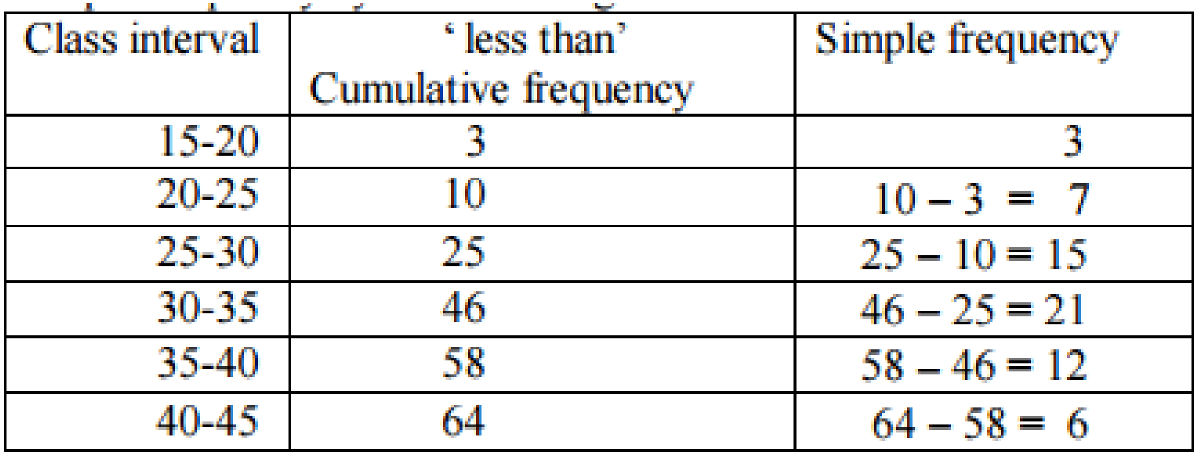
Cumulative frequency distribution has a running total of the values. It is constructed by adding the frequency of the first class interval to the frequency of the second class interval. Again add that total to the frequency in the third class interval continuing until the final total appearing opposite to the last class interval will be the total of all frequencies. The cumulative frequency may be downward or upward. A downward cumulation results in a list presenting the number of frequencies "less than" any given amount as revealed by the lower limit of succeeding class interval and the upward cumulative results in a list presenting the number of frequencies "more than" and given amount is revealed by the upper limit of a preceding class interval.

Cumulative frequency table

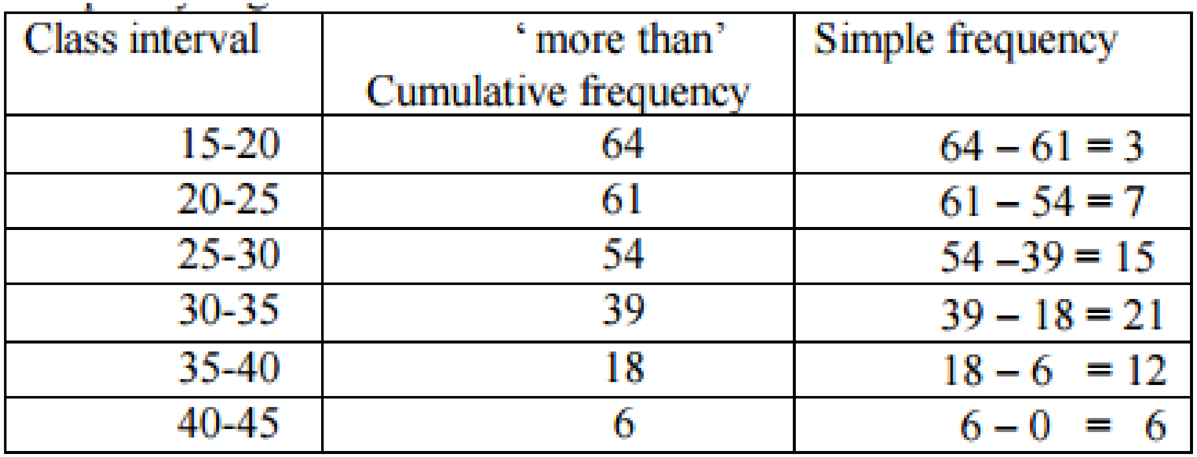


Age group (in years)	Number of women	Less than Cumulative frequency	More than cumulative frequency
15-20	3	3	64
20-25	7	10	61
25-30	15	25	54
30-35	21	46	39
35-40	12	58	18
40-45	6	64	6

Conversion of cumulative frequency to simple Frequency -

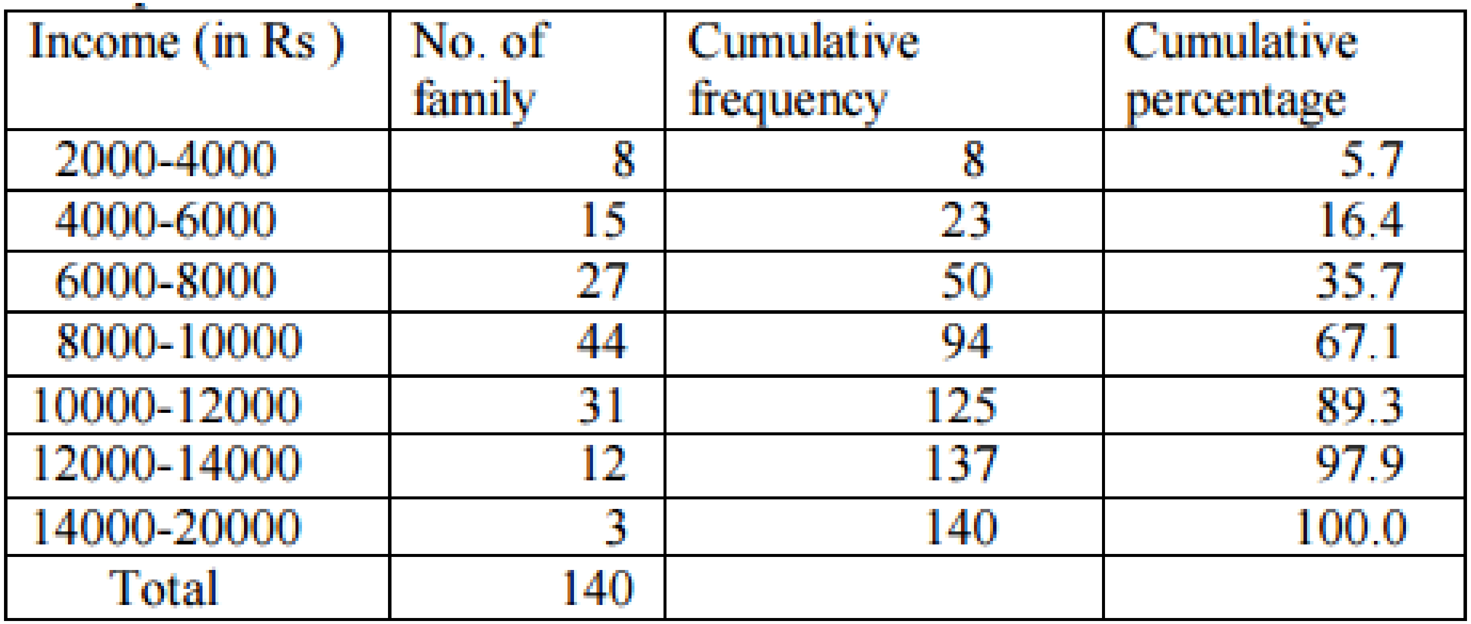


Class interval	‘ less than’ Cumulative frequency	Simple frequency
15-20	3	3
20-25	10	10 – 3 = 7
25-30	25	25 – 10 = 15
30-35	46	46 – 25 = 21
35-40	58	58 – 46 = 12
40-45	64	64 – 58 = 6



Class interval	‘ more than’ Cumulative frequency	Simple frequency
15-20	64	64 – 61 = 3
20-25	61	61 – 54 = 7
25-30	54	54 –39 = 15
30-35	39	39 – 18 = 21
35-40	18	18 – 6 = 12
40-45	6	6 – 0 = 6

Cumulative percentage Frequency table



Income (in Rs)	No. of family	Cumulative frequency	Cumulative percentage
2000-4000	8	8	5.7
4000-6000	15	23	16.4
6000-8000	27	50	35.7
8000-10000	44	94	67.1
10000-12000	31	125	89.3
12000-14000	12	137	97.9
14000-20000	3	140	100.0
Total	140		

Following is the distribution of persons according to different income groups. Calculate arithmetic mean.

Income Rs(100)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of persons	6	8	10	12	7	4	3

Income C.I	Number of Persons (f)	Mid X	$d = \frac{x - A}{c}$	Fd
0-10	6	5	-3	-18
10-20	8	15	-2	-16
20-30	10	25	-1	-10
30-40	12	A $\boxed{35}$	0	0
40-50	7	45	1	7
50-60	4	55	2	8
60-70	3	65	3	9
	50			-20

Mean = $\bar{x} = A + \frac{\sum fd}{N}$

$= 35 - \frac{20}{50} \times 10$

$= 35 - 4$

$= 31$

Measures of Central Tendency

- ✓ Frequency distributions and corresponding graphical representations make raw data more meaningful, yet they fail to identify three major properties that describe a set of quantitative data.
- ✓ The numerical value of an observation (also called as central value) around which most numerical values of other observations in the data set show a tendency to cluster or group, which is called central tendency.
- ✓ The extent to which numerical values are dispersed around the central value, which is called variation (measure of dispersion)
- ✓ The extent of the departure of numerical values from symmetrical(normal) distribution around the central value, which is called skewness.

Measures of Central Tendency

✓ **Mathematical Averages**

(a) Arithmetic Mean

(b) Geometric Mean

(c) Harmonic Mean

✓ **Averages of Position**

(a) Median

(b) Quartiles

(c) Deciles

(d) Percentiles

(e) Mode

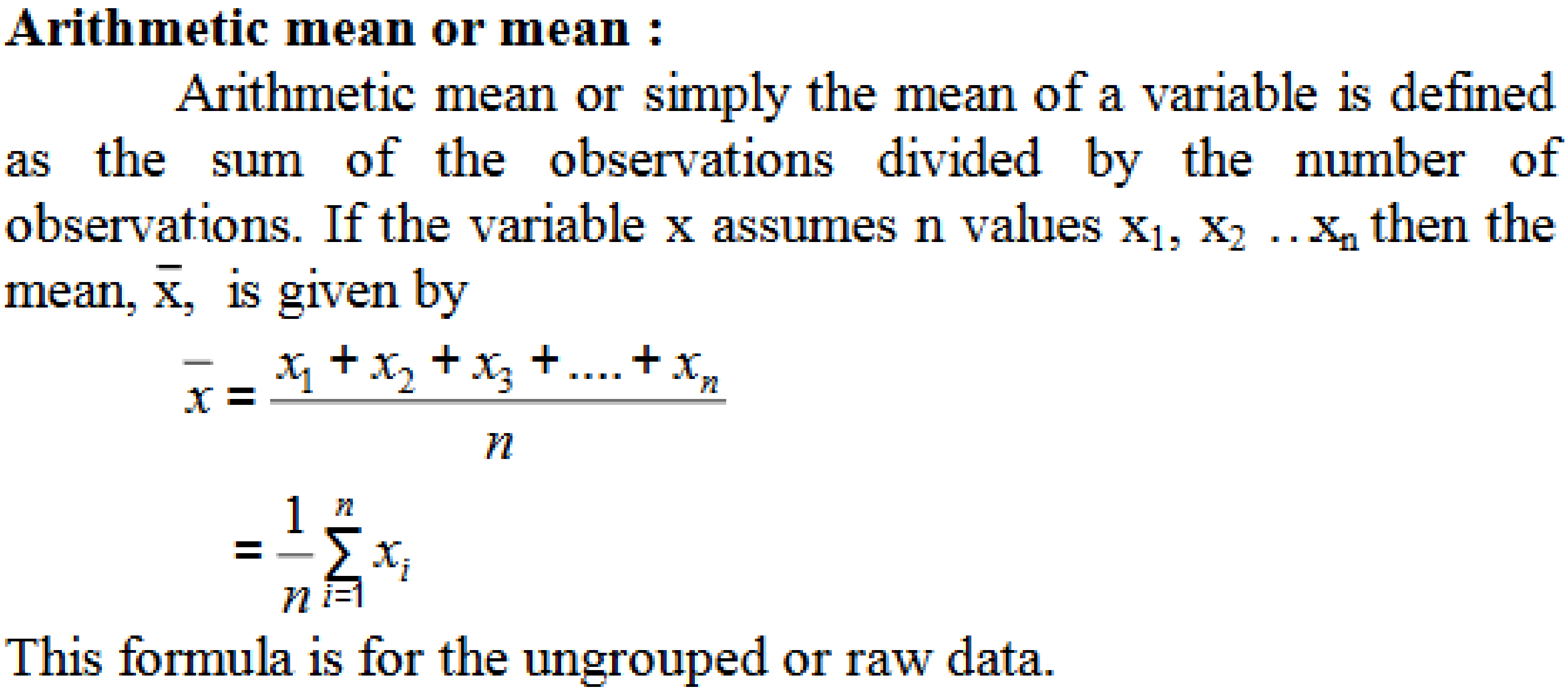
Measures of Central Tendency

“A measure of central tendency is a typical value around which other figures congregate.”

“An average stands for the whole group of which it forms a part yet represents the whole.”

$x = \frac{}{n}$

Arithmetic Mean



Arithmetic mean or mean :

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. If the variable x assumes n values x_1, x_2 . . x_n then the mean, $\bar{x}$, is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots. + x_n}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i$$

This formula is for the ungrouped or raw data.

Arithmetic Mean

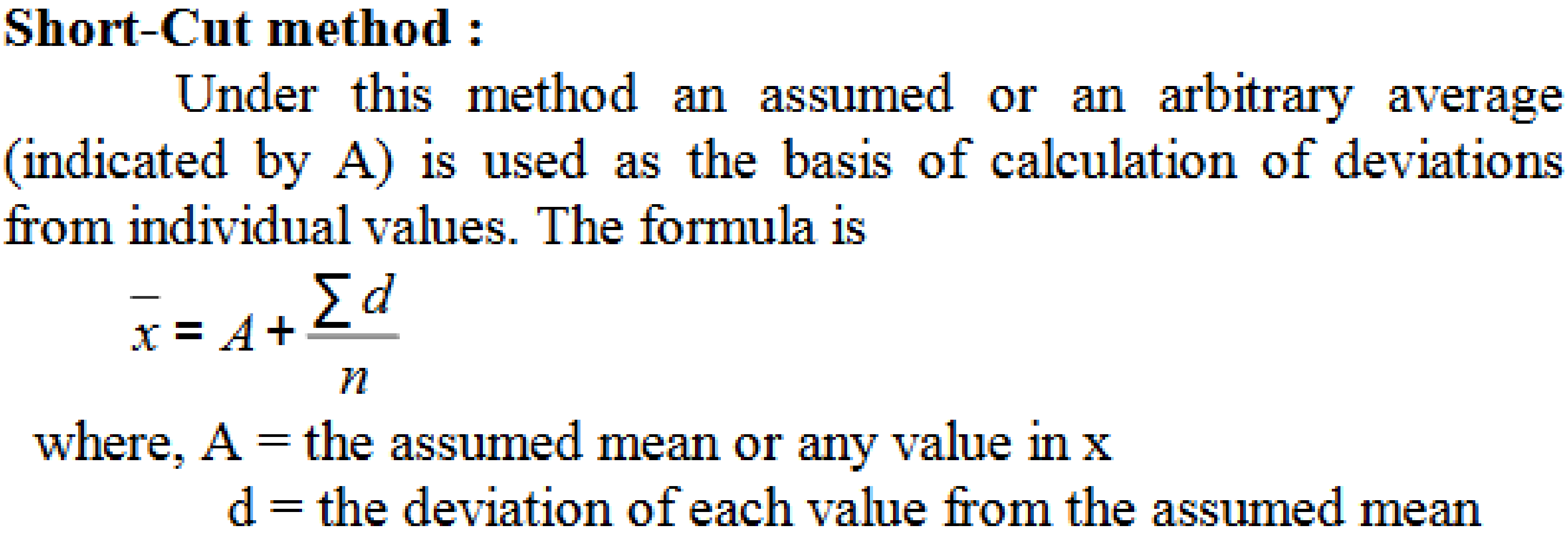
- Find the arithmetic mean (average) of TCS share price.

Price- 2600,2533, 2631, 2628,2740,2644

$$\bar{x} = \frac{2600+2533+ 2631+2628+2740+2644}{6}$$

$$\bar{x} = 2629$$

Arithmetic Mean



Short-Cut method :

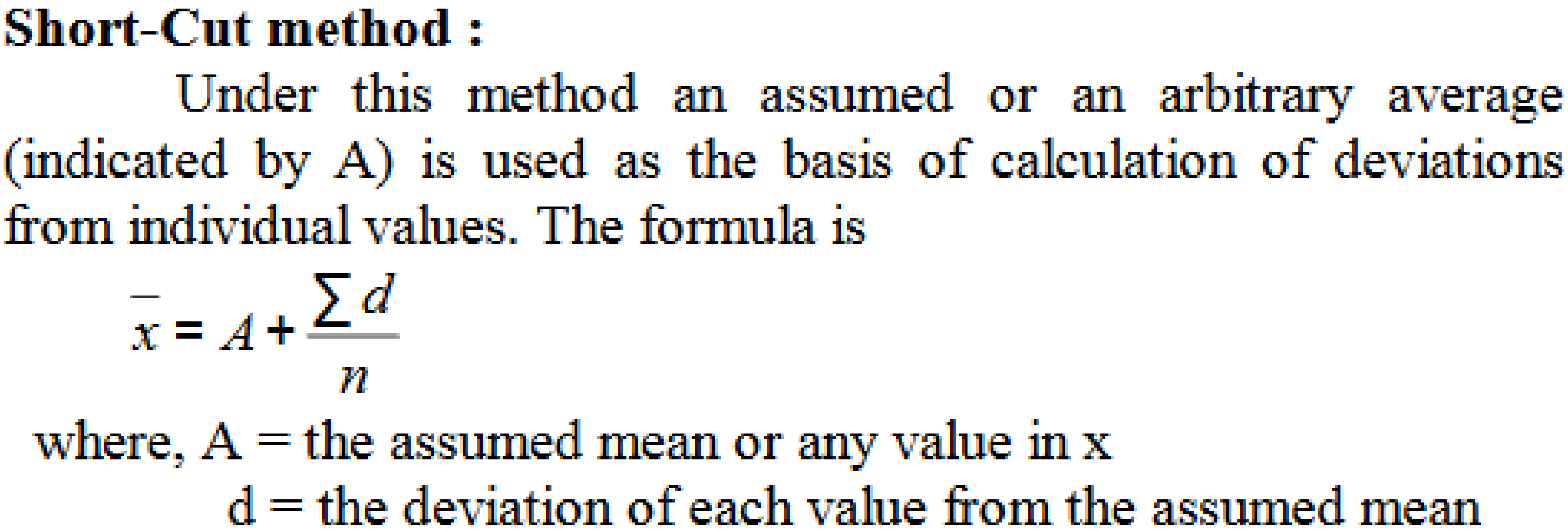
Under this method an assumed or an arbitrary average (indicated by A) is used as the basis of calculation of deviations from individual values. The formula is

$$\bar{x} = A + \frac{\sum d}{n}$$

where, A = the assumed mean or any value in x

d = the deviation of each value from the assumed mean

Arithmetic Mean



Short-Cut method :

Under this method an assumed or an arbitrary average (indicated by A) is used as the basis of calculation of deviations from individual values. The formula is

$$\bar{x} = A + \frac{\sum d}{n}$$

where, A = the assumed mean or any value in x

d = the deviation of each value from the assumed mean

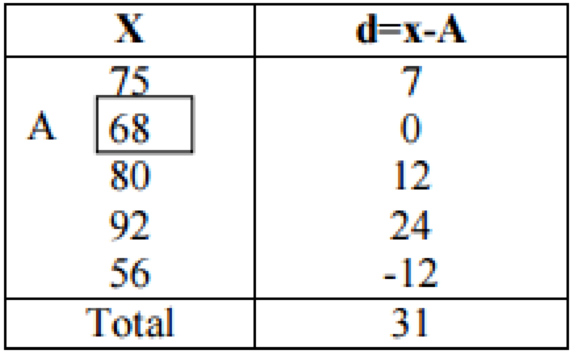
Arithmetic Mean-Step deviation method

$$\bar{x} = A + \frac{\sum d}{n}$$

where, A = the assumed mean or any value in x

d = the deviation of each value from the assumed mean

Share price of a company for the 5 days is – 75 , 68, 80, 92, 56. Find average share price



X	d=x-A
75	7
A $\boxed{68}$	0
80	12
92	24
56	-12
Total	31

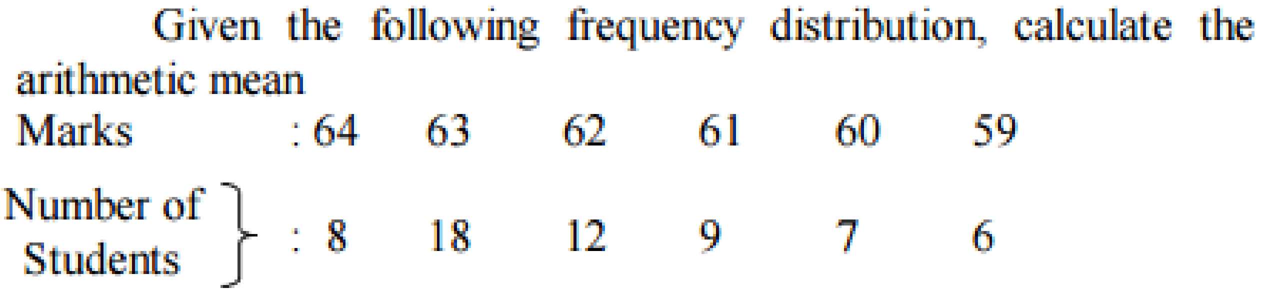
$$\bar{x} = A + \frac{\sum d}{n}$$

$$= 68 + \frac{31}{5}$$

$$= 68 + 6.2$$

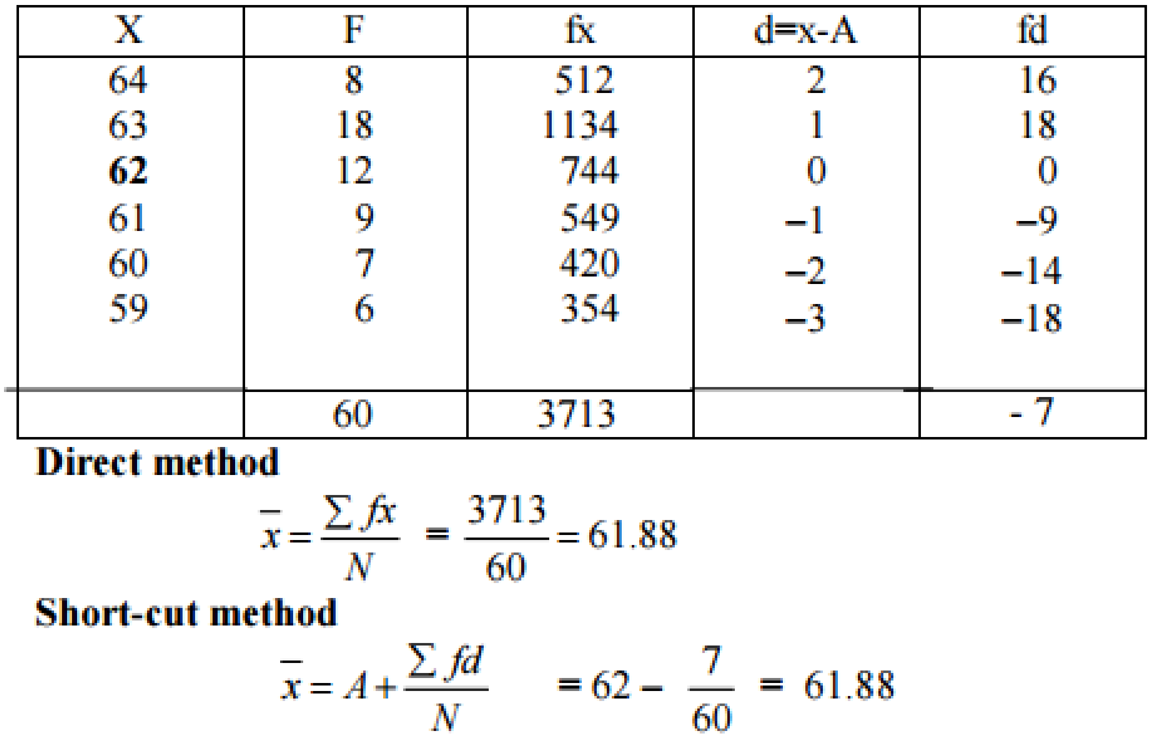
$$= 74.2$$

Arithmetic Mean-



Given the following frequency distribution, calculate the arithmetic mean

Marks	: 64	63	62	61	60	59
Number of Students	: 8	18	12	9	7	6



X	F	fx	d=x-A	fd
64	8	512	2	16
63	18	1134	1	18
62	12	744	0	0
61	9	549	–1	–9
60	7	420	–2	–14
59	6	354	–3	–18
	60	3713		- 7

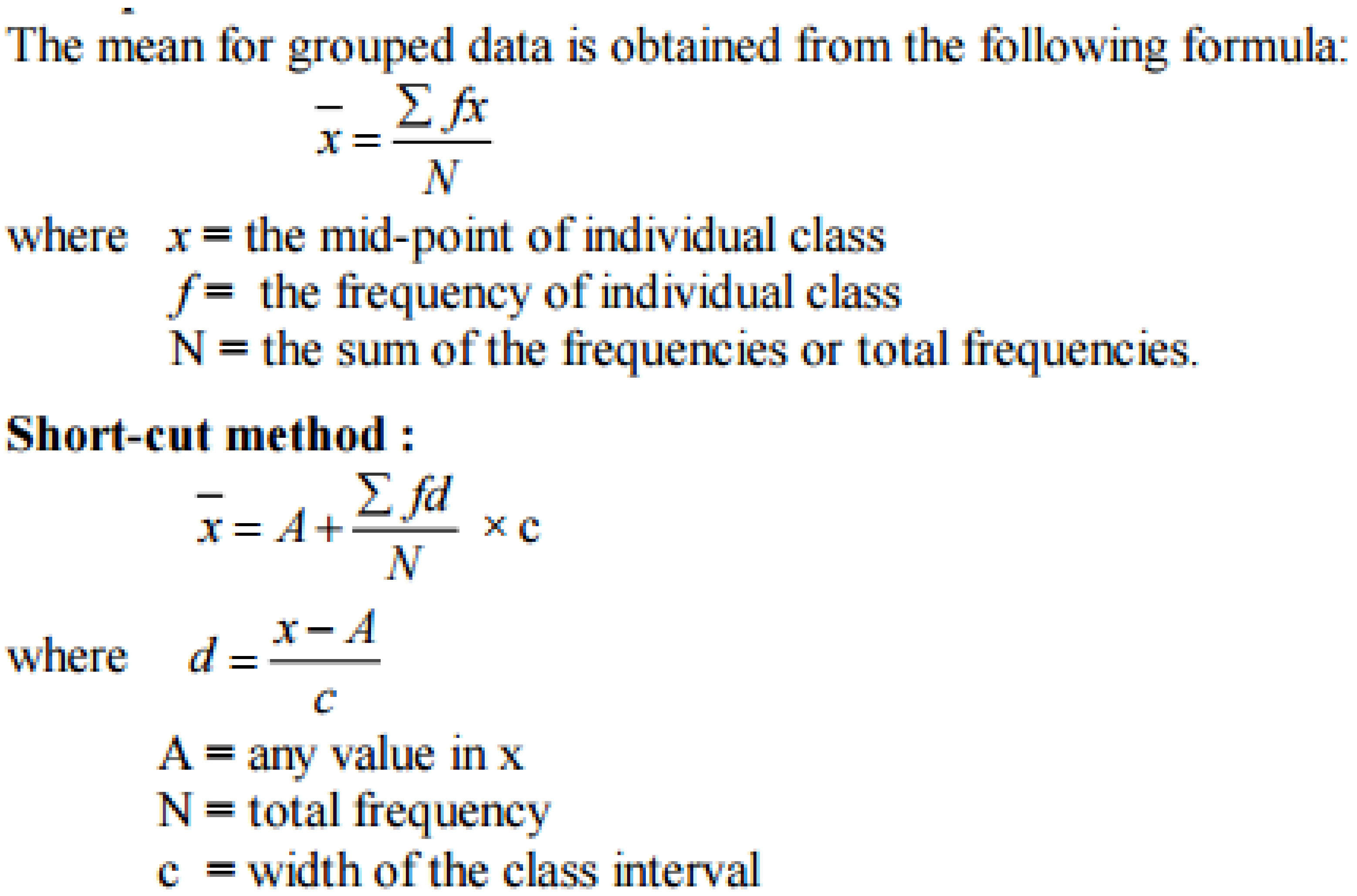
Direct method

$$\bar{x} = \frac{\sum fx}{N} = \frac{3713}{60} = 61.88$$

Short-cut method

$$\bar{x} = A + \frac{\sum fd}{N} = 62 - \frac{7}{60} = 61.88$$

Arithmetic Mean- [Grouped data]



The mean for grouped data is obtained from the following formula:

$$\bar{x} = \frac{\sum fx}{N}$$

where x = the mid-point of individual class

f = the frequency of individual class

N = the sum of the frequencies or total frequencies.

Short-cut method :

$$\bar{x} = A + \frac{\sum fd}{N} \times \mathrm{c}$$

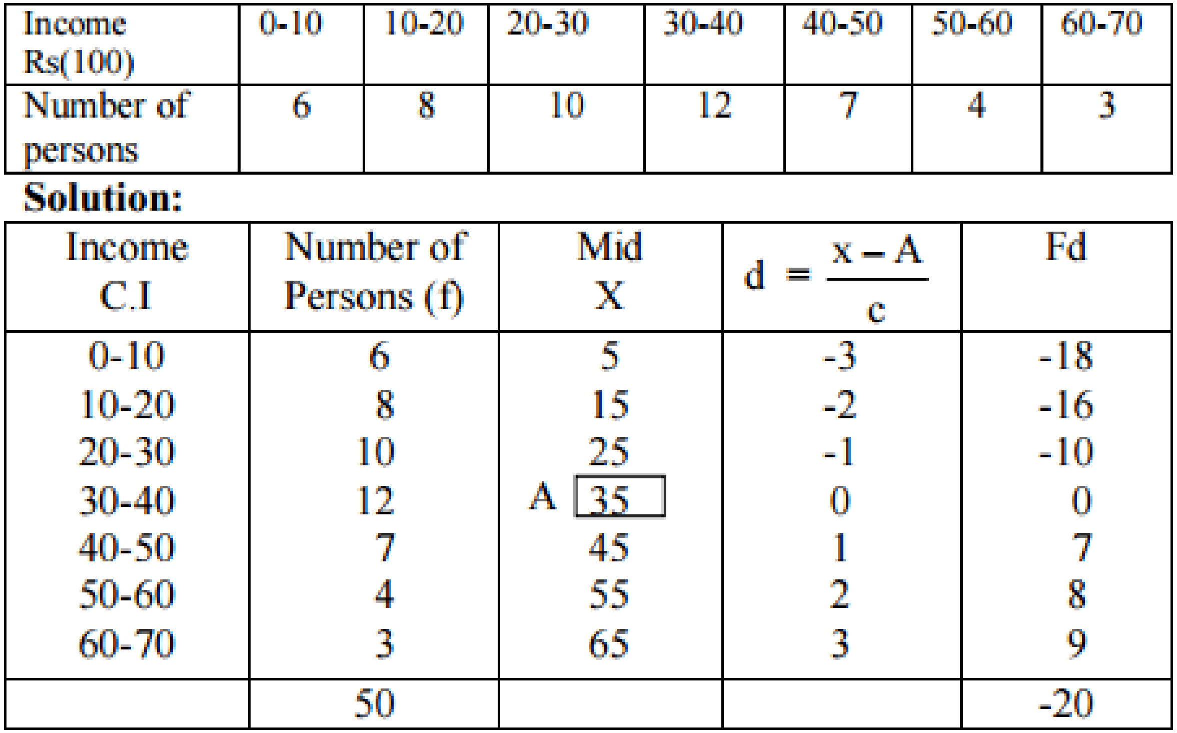
where $d = \dfrac{x - A}{c}$

A = any value in x

N = total frequency

c = width of the class interval

Arithmetic Mean- [Grouped data]



Income Rs(100)	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Number of persons	6	8	10	12	7	4	3

Solution:

Income C.I	Number of Persons (f)	Mid X	$d = \dfrac{x - A}{c}$	Fd
0-10	6	5	-3	-18
10-20	8	15	-2	-16
20-30	10	25	-1	-10
30-40	12	A $\boxed{35}$	0	0
40-50	7	45	1	7
50-60	4	55	2	8
60-70	3	65	3	9
	50			-20

$$\bar{x} = A + \frac{\sum fd}{N} \times \text{c}$$

$$= 35 - \frac{20}{50} \times 10$$

$$= 35 - 4$$

$$= 31$$

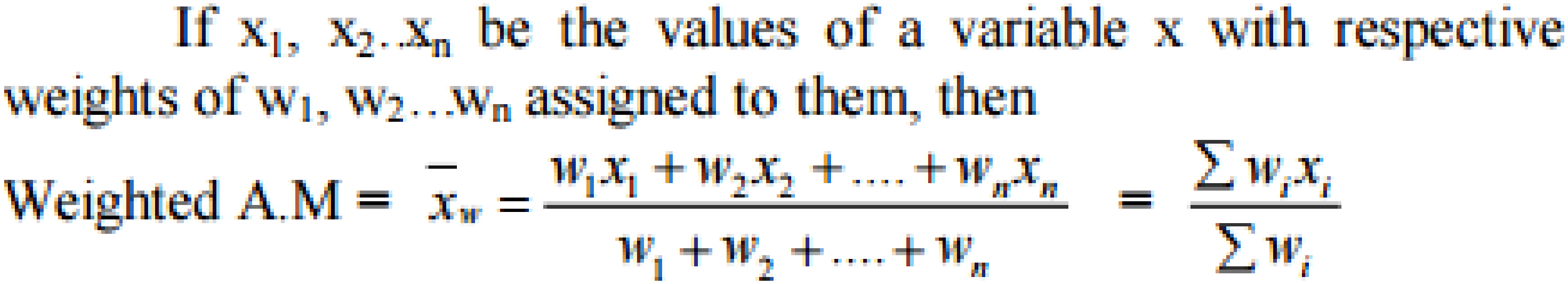
$$\bar{x} = \frac{\Sigma f X}{N} = \frac{1550}{50} = 31$$

Weighted AM

- ✓ AM gives equal importance (or weight) to each observation in the data set.
- ✓ However there are situations in which values of individual observations in the data set are not of equal importance.
- ✓ If the values occur with different frequencies, then computing AM of values may not be a rue representative of the data set characteristic and thus may be misleading.
- ✓ Under these circumstances, we may attach to each observation value “weight” w1, w2--- wn as the indicator of their importance and calculate weighted AM.

Weighted Arithmetic mean

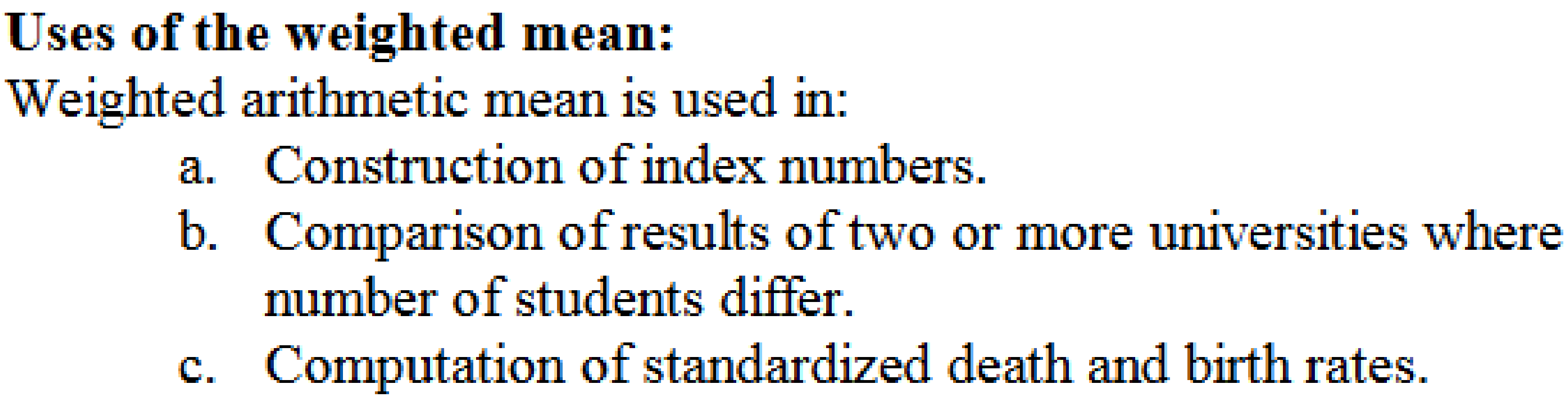
The average whose component items are being multiplied by certain values known as "weights" and the aggregate of the multiplied results are being divided by the total sum of their "weight" is called weiahted AM.



If x_1, $x_2..x_n$ be the values of a variable x with respective weights of w_1, $w_2...w_n$ assigned to them, then

$$\text{Weighted A.M} = \bar{x}_w = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum w_i x_i}{\sum w_i}$$

Weighted Arithmetic mean

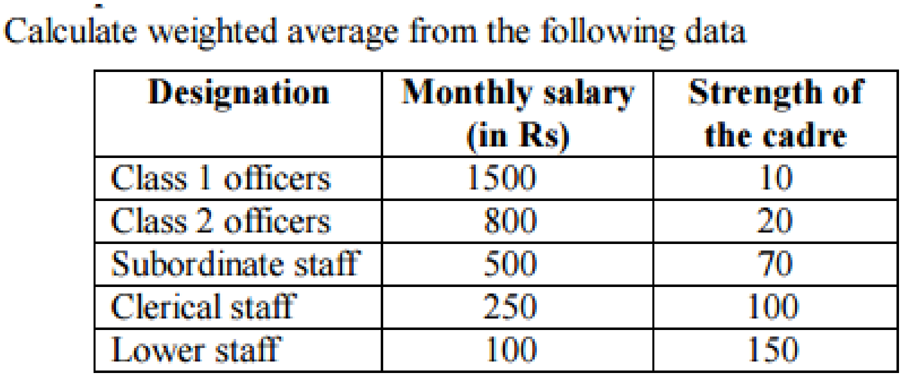


Uses of the weighted mean:

Weighted arithmetic mean is used in:

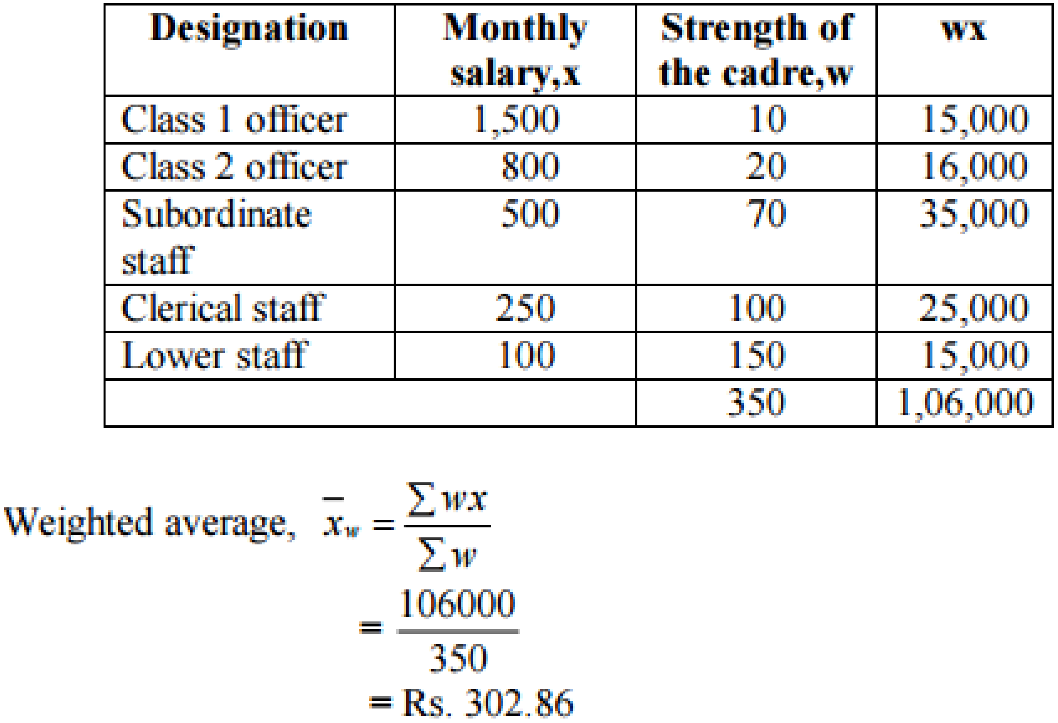
a. Construction of index numbers.
b. Comparison of results of two or more universities where number of students differ.
c. Computation of standardized death and birth rates.

Weighted Arithmetic mean



Calculate weighted average from the following data

Designation	Monthly salary (in Rs)	Strength of the cadre
Class 1 officers	1500	10
Class 2 officers	800	20
Subordinate staff	500	70
Clerical staff	250	100
Lower staff	100	150



Designation	Monthly salary,x	Strength of the cadre,w	wx
Class 1 officer	1,500	10	15,000
Class 2 officer	800	20	16,000
Subordinate staff	500	70	35,000
Clerical staff	250	100	25,000
Lower staff	100	150	15,000
		350	1,06,000

$$\text{Weighted average, } \bar{x}_w = \frac{\sum wx}{\sum w}$$

$$= \frac{106000}{350}$$

$$= \text{Rs. } 302.86$$

Weighted Arithmetic mean

A candidate obtained the following % of marks in examination: English 60; Hindi 75; Maths 63, Physics 59, Chemistry 55. Find the candidate weighted AM if weights 1,2,1,3,3 respectively are allocated to subjects

Subject	Marks (%)	Weight(W)	WX
English	60	1	60
Hindi	75	2	150
Maths	63	1	63
Physics	59	3	177
Chemistry	55	3	165
		10	615

Weighted AM (in%) = 615/10 = 61.5

Properties - AM

- ✓ The algebraic sum of the deviations of given set of observations from their arithmetic mean is zero. $\sum_{i=1}^{n}(xi - mean) = 0$
- ✓ If we know the sizes and means of two component series, then we can find the mean of the resultant series obtained on combination of given series. If n1 and n2 are the sizes and $\bar{X}1$ and $\bar{X}2\ are\ the\ respective\ means\ of\ the\ two\ groups\ then\ \ \bar{X}\ of\ the\ combined\ group\ of\ size\ n1 + n2\ is\ given\ by$

$$\bar{X} = \frac{n1\bar{X}1 + n2\bar{X}2}{n1 + n2}$$

- ✓ The sum of squares of deviations of the given set of observations is minimum when taken from arithmetic mean.

Arithmetic Mean

Merits –

- ✓ Arithmetic mean rigidly defined by Algebraic Formula.
- ✓ It is easy to calculate and simple to understand.
- ✓ It is based on all observations of the given data.
- ✓ It is capable of being treated mathematically hence it is widely used in statistical analysis.

Arithmetic Mean

Demerits –

- ✓ As the data becomes skewed the mean loses its ability to provide the best central location for the data. i.e. It is affected very much by extreme values.
- ✓ It can neither be determined by inspection or by graphical location
- ✓ Arithmetic mean cannot be computed for non parametric data.
- ✓ It cannot be calculated for open-end classes

Arithmetic Mean

- Demerits –

Emp.No	1	2	3	4	5	6	7	8	9	10
Salary	15k	18k	16k	14k	15k	15k	12k	17k	90k	95k

μ = 30.7

Arithmetic Mean

- Trimmed mean: A trimmed mean is computed by trimming away a certain percent of both the largest and smallest set of values.

Geometric Mean

Geometric Mean of set of n observations is the nth root of their products.

If *x1, x2 . . . Xn are the observations* then

$$\text{G.M} = \sqrt[n]{x_1 . x_2 \ldots x_n}$$

$$= (x_1 . x_2 \ldots x_n)^{1/n}$$

$$\log \text{GM} = \frac{1}{n} \log(x_1 . x_2 \ldots x_n)$$

$$= \frac{1}{n} (\log x_1 + \log x_2 + .. + \log x_n$$

$$= \frac{\sum \log x_i}{n}$$

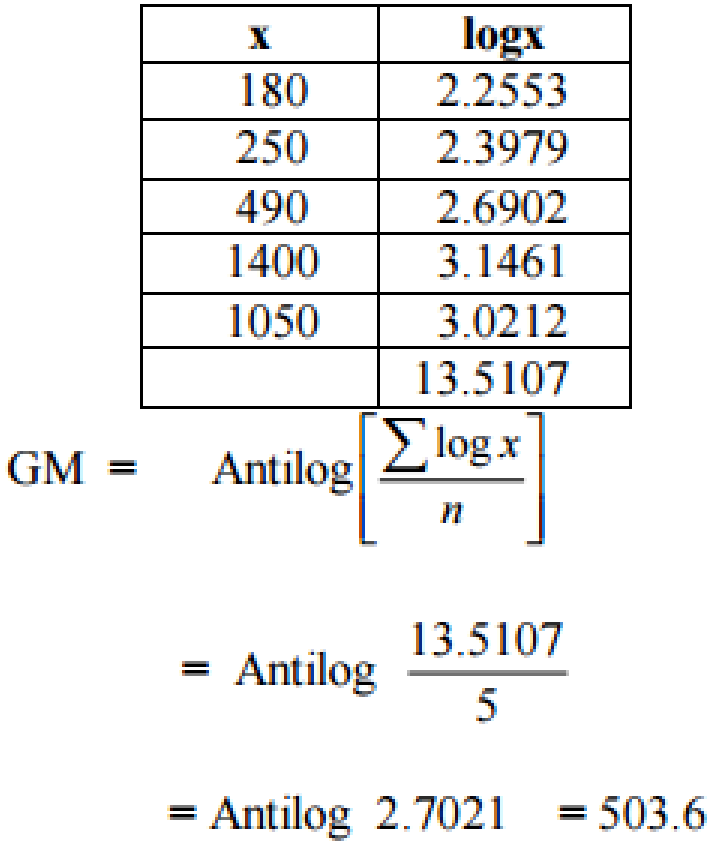
$$\text{GM} = \text{Antilog } \frac{\sum \log x_i}{n}$$

For grouped data

$$\text{GM} = \text{Antilog} \left[\frac{\sum f \log x_i}{N} \right]$$

Geometric Mean

Calculate the Geometric average monthly income of 5 families.



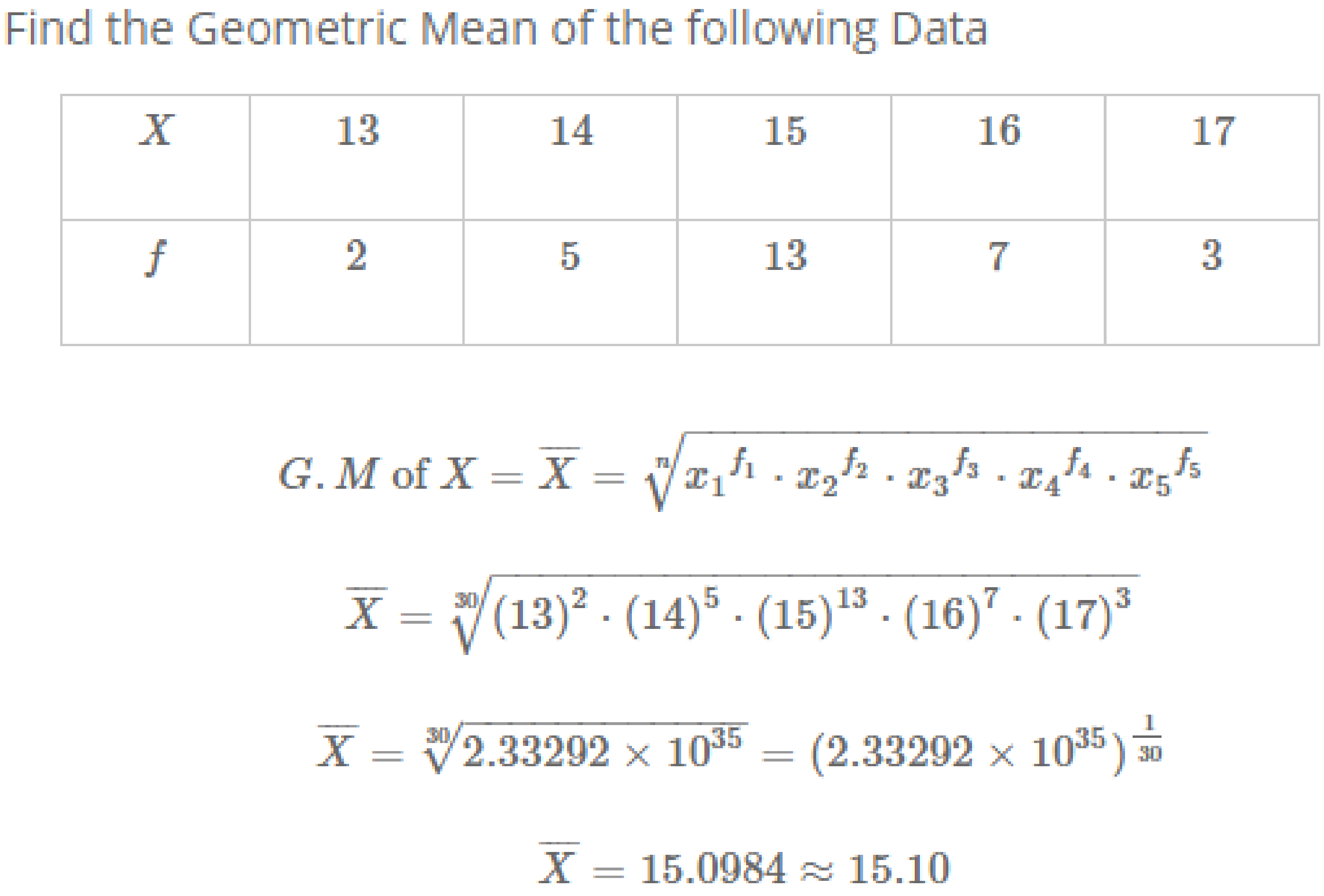
x	logx
180	2.2553
250	2.3979
490	2.6902
1400	3.1461
1050	3.0212
	13.5107

$$GM = \text{Antilog}\left[\frac{\sum \log x}{n}\right]$$

$$= \text{Antilog}\ \frac{13.5107}{5}$$

$$= \text{Antilog}\ 2.7021 \quad = 503.6$$

Geometric Mean



Find the Geometric Mean of the following Data

X	13	14	15	16	17
f	2	5	13	7	3

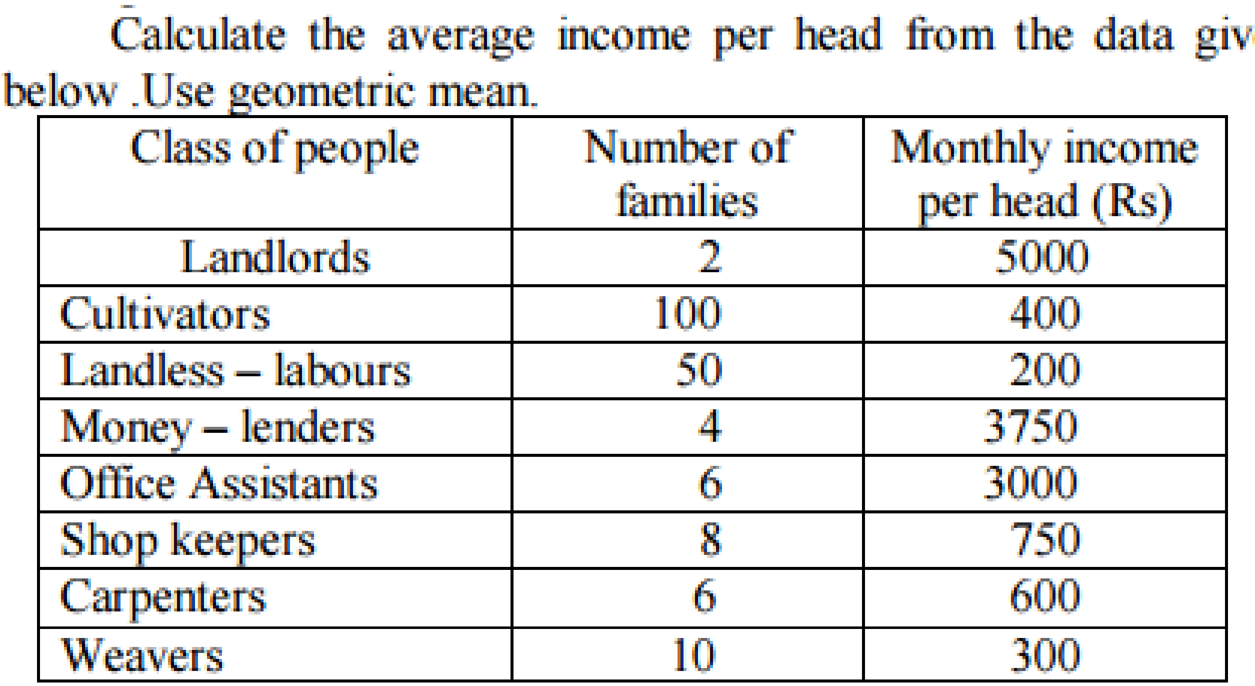
$$G.M \text{ of } X = \overline{X} = \sqrt[n]{{x_1}^{f_1} \cdot {x_2}^{f_2} \cdot {x_3}^{f_3} \cdot {x_4}^{f_4} \cdot {x_5}^{f_5}}$$

$$\overline{X} = \sqrt[30]{(13)^2 \cdot (14)^5 \cdot (15)^{13} \cdot (16)^7 \cdot (17)^3}$$

$$\overline{X} = \sqrt[30]{2.33292 \times 10^{35}} = (2.33292 \times 10^{35})^{\frac{1}{30}}$$

$$\overline{X} = 15.0984 \approx 15.10$$

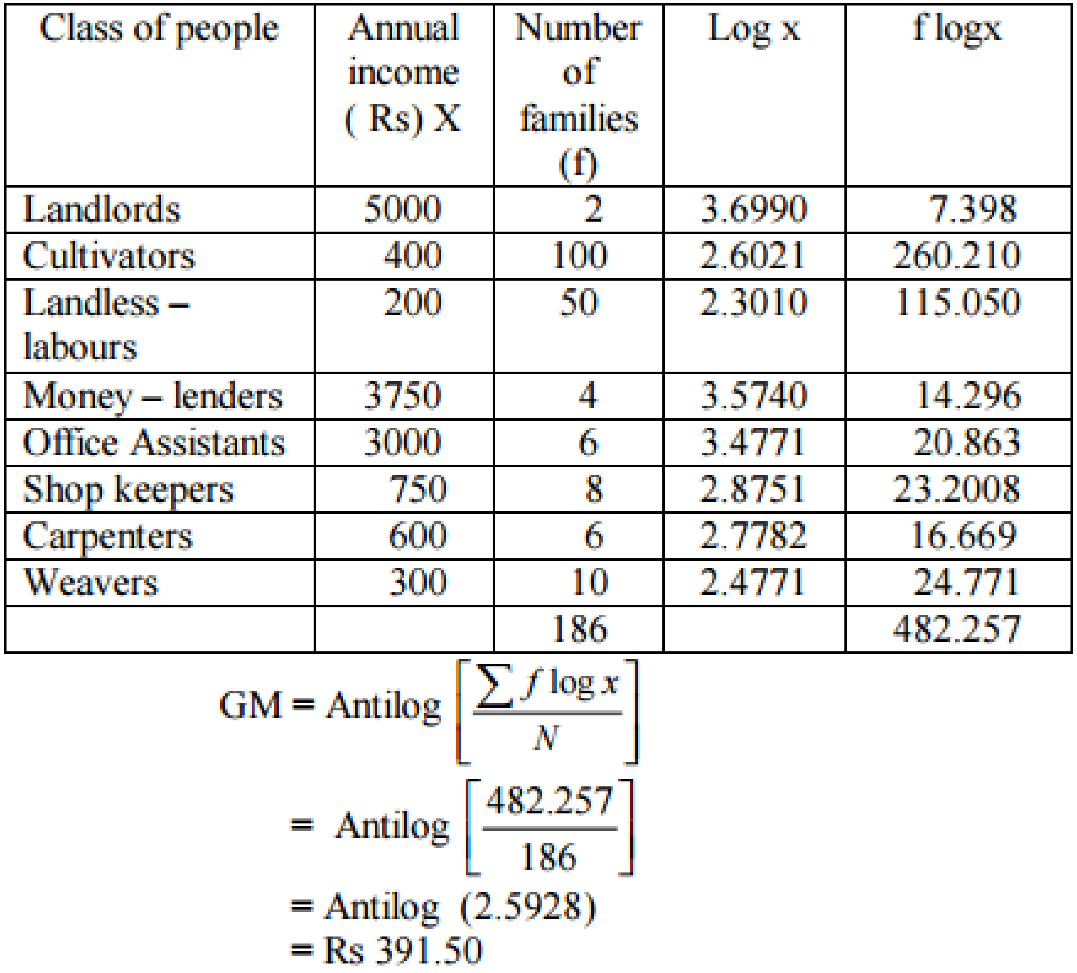
Geometric Mean



Calculate the average income per head from the data given below .Use geometric mean.

Class of people	Number of families	Monthly income per head (Rs)
Landlords	2	5000
Cultivators	100	400
Landless – labours	50	200
Money – lenders	4	3750
Office Assistants	6	3000
Shop keepers	8	750
Carpenters	6	600
Weavers	10	300

Geometric Mean



Class of people	Annual income (Rs) X	Number of families (f)	Log x	f logx
Landlords	5000	2	3.6990	7.398
Cultivators	400	100	2.6021	260.210
Landless – labours	200	50	2.3010	115.050
Money – lenders	3750	4	3.5740	14.296
Office Assistants	3000	6	3.4771	20.863
Shop keepers	750	8	2.8751	23.2008
Carpenters	600	6	2.7782	16.669
Weavers	300	10	2.4771	24.771
		186		482.257

$$GM = \text{Antilog}\left[\frac{\sum f \log x}{N}\right]$$

$$= \text{Antilog}\left[\frac{482.257}{186}\right]$$

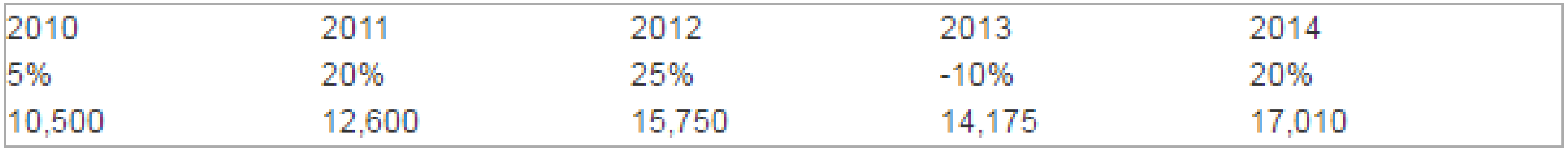
$$= \text{Antilog}\ (2.5928)$$

$$= \text{Rs } 391.50$$

Geometric Mean

- ✓ Investment return for Rs. 10,000 is as below.

Find the average rate of return.



2010	2011	2012	2013	2014
5%	20%	25%	-10%	20%
10,500	12,600	15,750	14,175	17,010

GM = $\sqrt[5]{1.05 \ast 1.2 \ast 1.25.0.9 \ast 1.2}$

GM = 11.20

Geometric Mean

- 1st year = 100% return
- 2nd year = -50% return

- Is your average return 25% or 0%

Geometric Mean

- Suppose your stock portfolio earned a 10% return in year 1 and a 20% return in year 2.

  What happened? If you invested \$100, then at the end of year 1 it grows to \$100 x (1 + 10%) = \$100 x (1.10) = \$110. So, we start year 2 with \$110, and that year goes very well, and we earn 20%, so our balance grows to \$110 x 1.20 = \$132.

  What is the geometric average return? It turns out that it's 1.1489%

  Now, let's look at what would have happened if we had earned that geometric average return every year: We start out investing \$100, and we get a return of 1.1489% in the first year, so we end the first year with a balance of \$100 x 1.1489 = \$114.89. Then, we start year 2 with \$114.89, and again we get a return of 14.89%, so we end year 2 with a balance of \$114.89 x 1.1489 = \$132.

Geometric distribution

- Find the average rate of increase in population which in the first decade had increased by 20%, in the next by 30% and in the third by 40%

Geometric Mean

- ✓ The geometric mean is used when dealing with average investment returns, Growth Rates, Portfolio return, etc.

Geometric Mean

Merits:

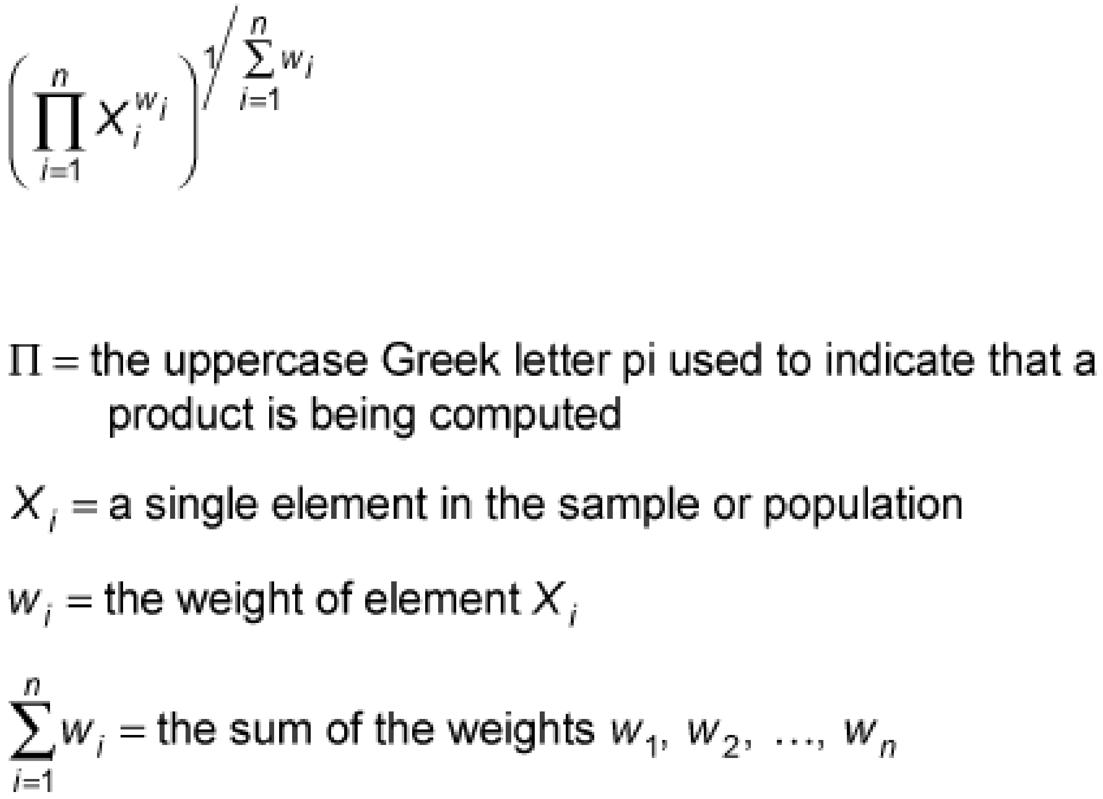
- ✓ It is rigidly defined.
- ✓ 2. It is based on all items.
- ✓ 3. It is very suitable for averaging ratios, rates and percentages.
- ✓ 4. It is capable of further mathematical treatment.
- ✓ 5. Unlike AM, it is not affected much by the presence of extreme values.

Geometric Mean

Demerits:

- ✓ It cannot be used when the values are negative or if any of the observations is zero
- ✓ It is difficult to calculate particularly when the items are very large or when there is a frequency distribution.

Weighted GM



$$\left(\prod_{i=1}^{n} X_i^{w_i}\right)^{1/\sum_{i=1}^{n} w_i}$$

Π = the uppercase Greek letter pi used to indicate that a product is being computed

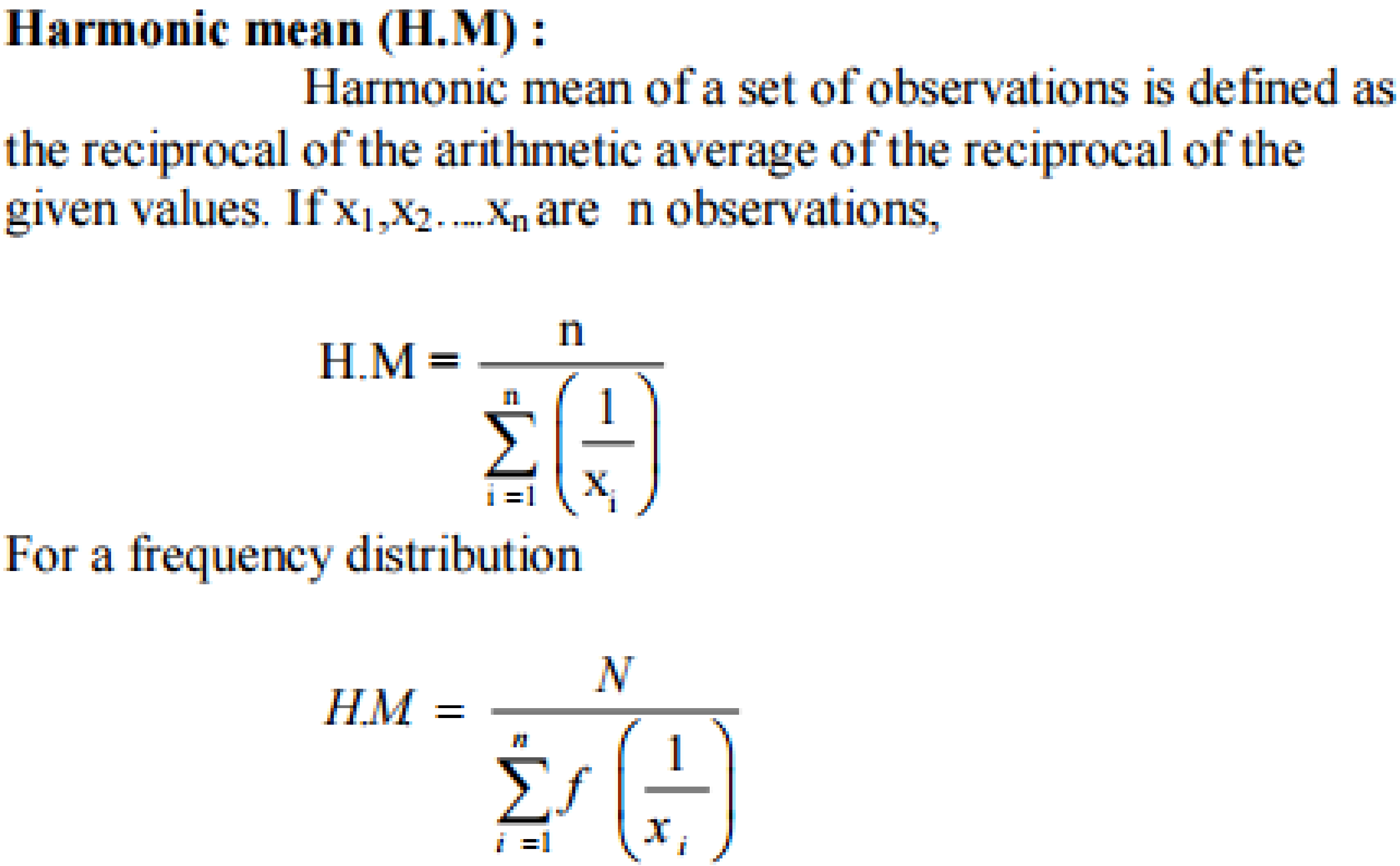
X_i = a single element in the sample or population

w_i = the weight of element X_i

$\sum_{i=1}^{n} w_i$ = the sum of the weights $w_1, w_2, \ldots, w_n$

GM(W) = Antilog(∑WlogX/ ∑W)

Harmonic Mean



Harmonic mean (H.M) :

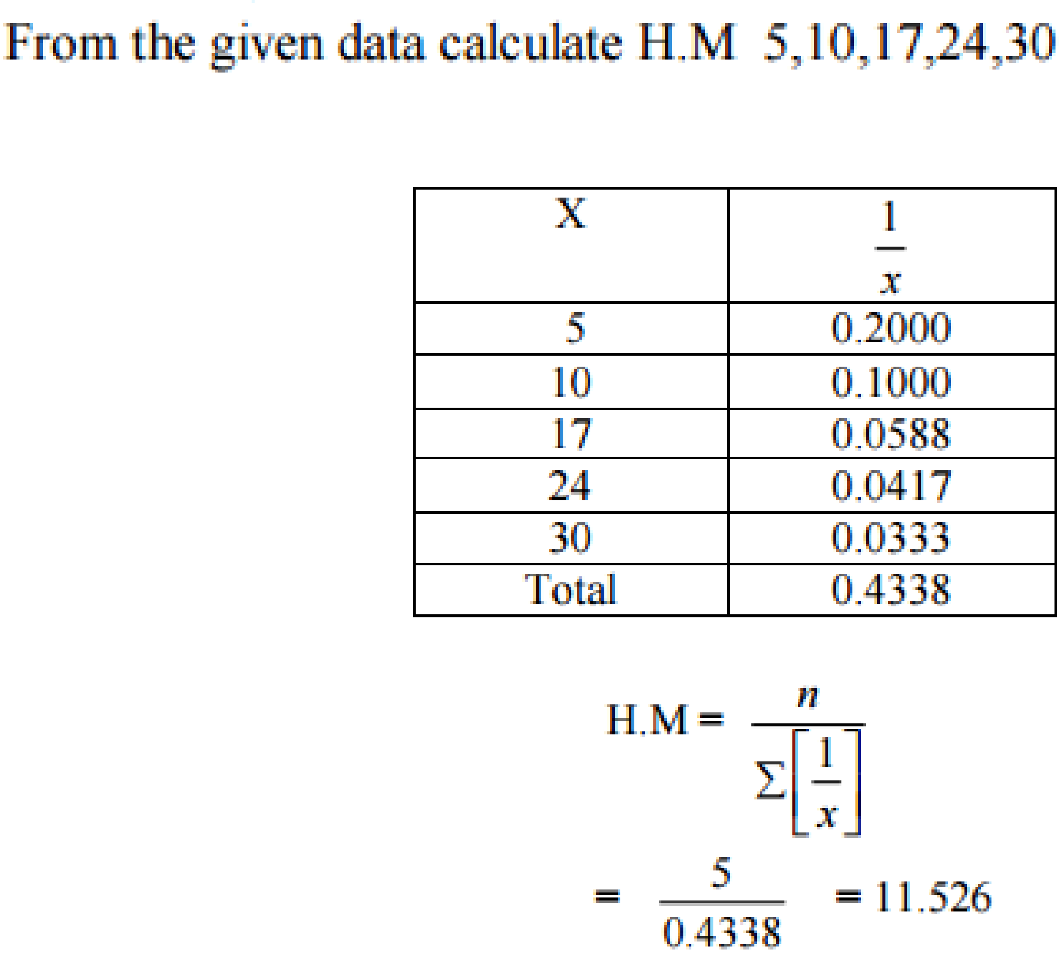
Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values. If $x_1, x_2 \ldots\ldots x_n$ are n observations,

$$H.M = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)}$$

For a frequency distribution

$$H.M = \frac{N}{\sum_{i=1}^{n} f\left(\frac{1}{x_i}\right)}$$

Harmonic Mean



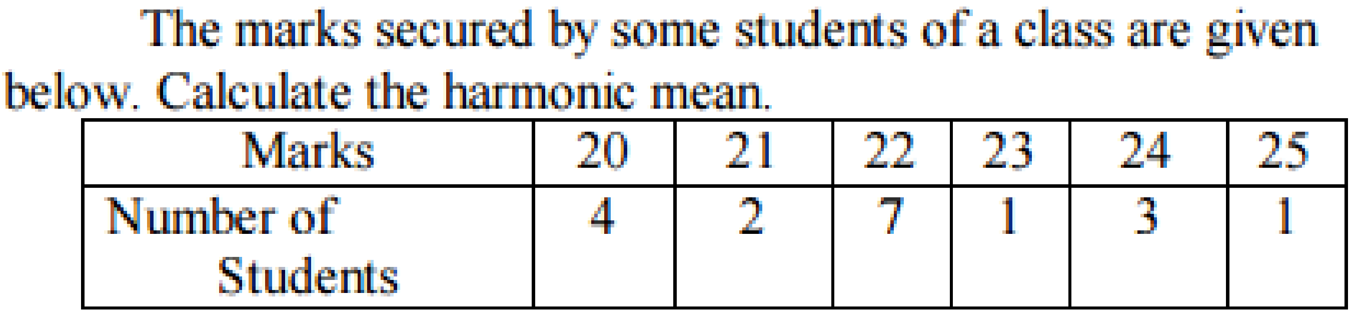
From the given data calculate H.M 5,10,17,24,30

X	$\frac{1}{x}$
5	0.2000
10	0.1000
17	0.0588
24	0.0417
30	0.0333
Total	0.4338

$$\text{H.M} = \frac{n}{\Sigma\left[\frac{1}{x}\right]}$$

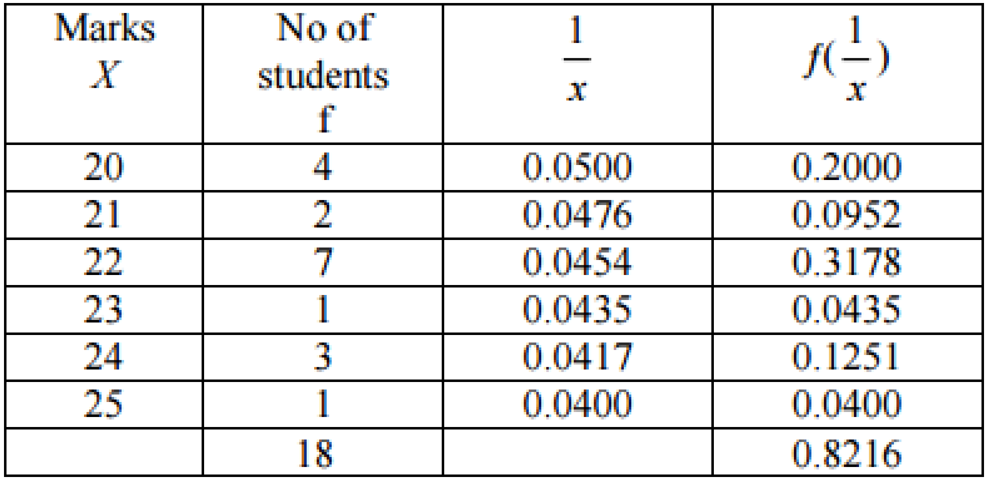
$$= \frac{5}{0.4338} = 11.526$$

Harmonic Mean

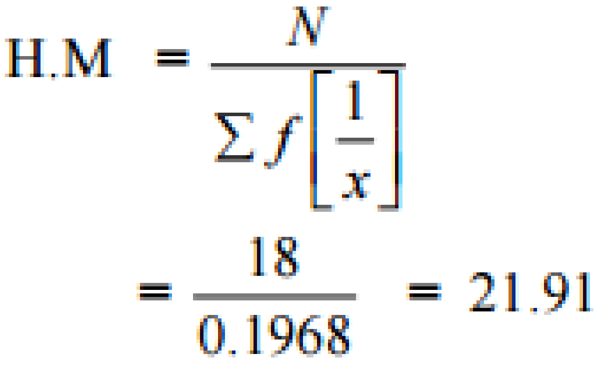


The marks secured by some students of a class are given below. Calculate the harmonic mean.

Marks	20	21	22	23	24	25
Number of Students	4	2	7	1	3	1



Marks X	No of students f	$\frac{1}{x}$	$f(\frac{1}{x})$
20	4	0.0500	0.2000
21	2	0.0476	0.0952
22	7	0.0454	0.3178
23	1	0.0435	0.0435
24	3	0.0417	0.1251
25	1	0.0400	0.0400
	18		0.8216



$$\text{H.M} = \frac{N}{\Sigma f\left[\frac{1}{x}\right]}$$

$$= \frac{18}{0.1968} = 21.91$$

Harmonic Mean

- ✓ A cyclist pedals from his house to his college at a speed of 10 k.m.p.h. and back from the college to his house at 15 k.m.p.h. Find the average speed.

Solution: Speed = Distance/ Time.

Let x be the distance from house to college.

Average speed = Total distance/Total time.

Average speed = 2x/[x/10 + x/15] = 12 kmph.

Harmonic Mean

- ✓ An investor buys Rs. 1,200 worth of shares in a company each month. During the first 5 months, he bought the shares at a price of Rs. 10, Rs. 12, Rs.15, Rs, 20 and Rs. 24 per share. After 5 months what is the average price paid for the shares by him?

Solution: Since the share value is chngine after a fixed unit of time (1 month), The required average price per share is harmonic mean of 10,12,15,20,24.

5/ [1/10 + 1/12 + 1/ 15 + 1/20 = 1/24] = 14.63.

Month	Price per share [x]	Total cost [fx]	Number of shares boutht [f]
1	10	1200	120
2	12	1200	100
3	15	1200	80
4	20	1200	60
5	24	1200	50
		Σ fx = 6000	Σ f= 410

Average price = $\Sigma fx / \Sigma f$

$= 6000 / 410$

$= 14.63$

Harmonic Mean

- ✓ **Weighted HM:** Instead of fixed (constant) distance being travelled with varying speeds, let us now suppose that different distances are travelled with corresponding different speeds. What is the average speed.

Speed = $\frac{Distance(S)}{Time(t)}$

t1 = $\frac{s1}{v1}$ t2 = $\frac{s2}{v2}$ …….. tn= $\frac{sn}{vn}$

Average speed = $\frac{\Sigma s}{\Sigma(\frac{s}{v})}$

Harmonic Mean

- If you make a trip which entails travelling 900 kms by train at an average speed of 60 kmph; 3000 kms by boad at an average speed of 25 kmph; 400 kms by plane at 350 kmph, and finally 15 ksm by taxi at 25 kmph. What is your average speed for the entire distance.

X (speed)	W(Distance)	W/X
60	900	15
25	3000	120
350	400	1.43
25	15	0.60
Total	4315	137.03

Average speed = $\frac{\Sigma W}{\Sigma(\frac{W}{X})}$

= 4315/ 134.03

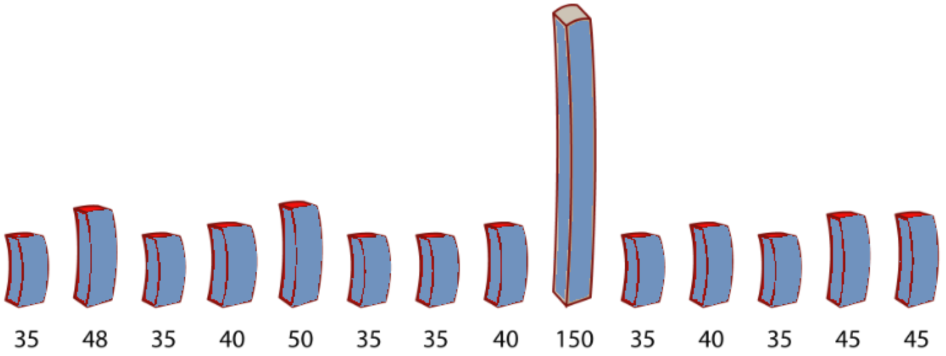
= 31.49 km. ph.

Harmonic Mean

Merits:

- ✓ It is rigidly defined.
- ✓ It is defined on all observations.
- ✓ It is the most suitable average when it is desired to give greater weight to smaller observations and less weight to the larger ones.
- ✓ It gives greater importance to small items and is therefore, useful only when small items have to be given greater weightage.

Harmonic Mean



If the population (or sample) has a few data points that are much higher than the rest (outliers), the harmonic mean is the appropriate average to use. Unlike the arithmetic mean, the harmonic mean gives less significance to high-value outliers–providing a truer picture of the average.

Median

✓"*The median is that value of the variable which divides the group into two equal parts, one part comprising all the values greater and the other, all the values less than median*" – L.R. Conner

✓Thus median of distribution may be defined as that value f the variable which exceeds and is exceeded by the same number of observations.

Median

Ungrouped Data :

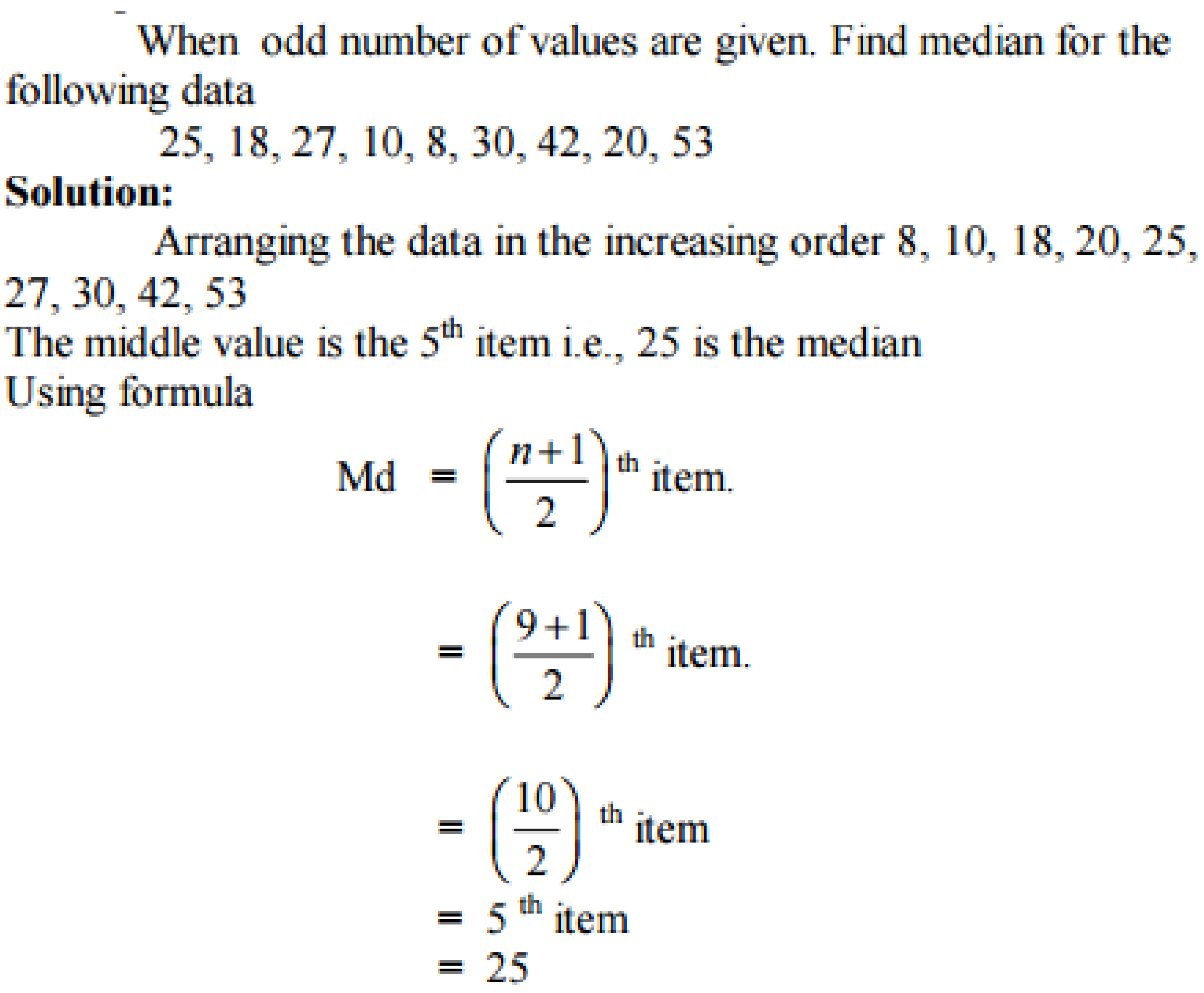
Arrange the given values in the increasing or decreasing order.

If the number of values are odd, median is the middle value.

If the number of values are even, median is the mean of middle two values. By formula Median = Md = $\frac{(n+1)}{2}$thitem.

Median

Ungrouped Data :



When odd number of values are given. Find median for the following data

25, 18, 27, 10, 8, 30, 42, 20, 53

Solution:

Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53

The middle value is the 5th item i.e., 25 is the median

Using formula

$$\text{Md} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item.}$$

$$= \left(\frac{9+1}{2}\right)^{\text{th}} \text{ item.}$$

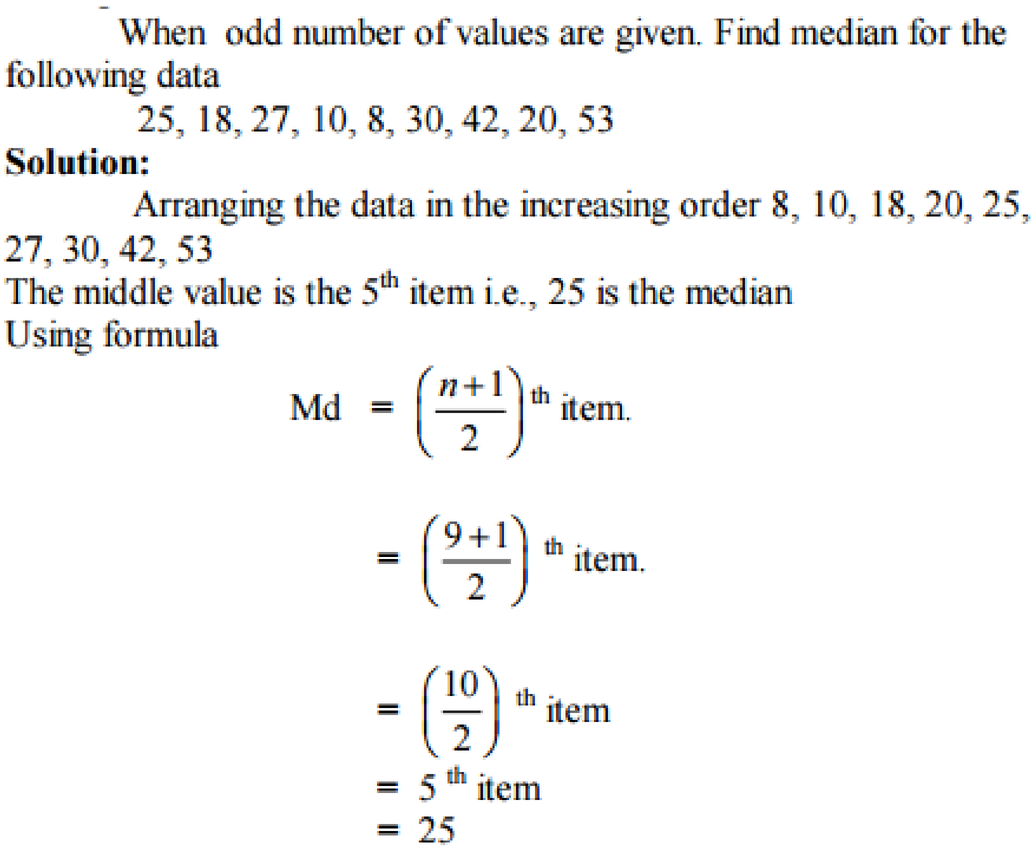
$$= \left(\frac{10}{2}\right)^{\text{th}} \text{ item}$$

$$= 5^{\text{th}} \text{ item}$$

$$= 25$$

Median

Ungrouped Data :



When odd number of values are given. Find median for the following data

25, 18, 27, 10, 8, 30, 42, 20, 53

Solution:

Arranging the data in the increasing order 8, 10, 18, 20, 25, 27, 30, 42, 53

The middle value is the 5th item i.e., 25 is the median

Using formula

$$\text{Md} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item.}$$

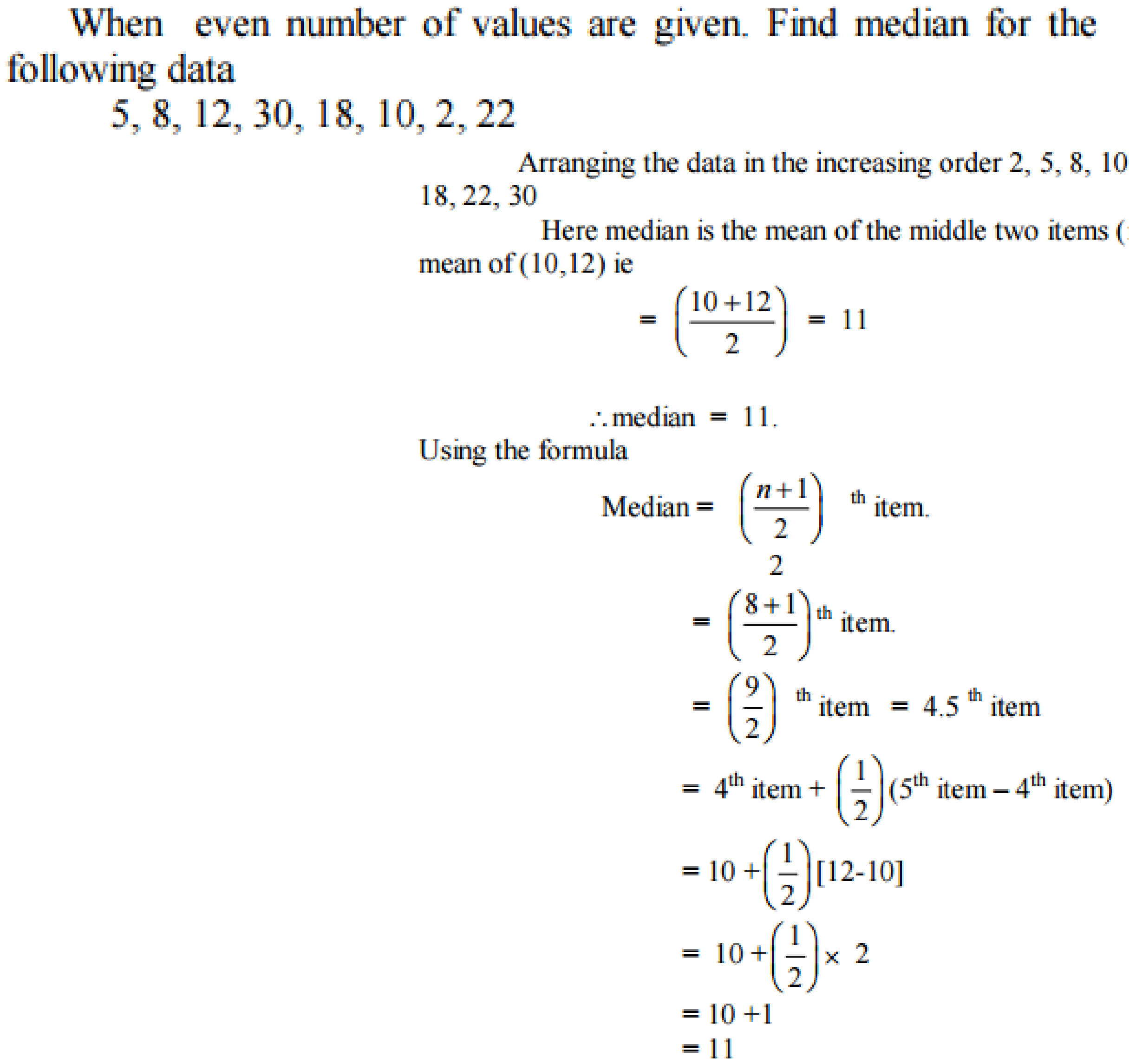
$$= \left(\frac{9+1}{2}\right)^{\text{th}} \text{item.}$$

$$= \left(\frac{10}{2}\right)^{\text{th}} \text{item}$$

$$= 5^{\text{th}} \text{item}$$

$$= 25$$

Median



When even number of values are given. Find median for the following data

5, 8, 12, 30, 18, 10, 2, 22

Arranging the data in the increasing order 2, 5, 8, 10, 12, 18, 22, 30

Here median is the mean of the middle two items (ie) mean of (10,12) ie

$$= \left(\frac{10+12}{2}\right) = 11$$

$$\therefore \text{median} = 11.$$

Using the formula

$$\text{Median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{ item.}$$

2

$$= \left(\frac{8+1}{2}\right)^{\text{th}} \text{ item.}$$

$$= \left(\frac{9}{2}\right)^{\text{th}} \text{ item} = 4.5^{\text{th}} \text{ item}$$

$$= 4^{\text{th}} \text{ item} + \left(\frac{1}{2}\right)(5^{\text{th}} \text{ item} - 4^{\text{th}} \text{ item})$$

$$= 10 + \left(\frac{1}{2}\right)[12\text{-}10]$$

$$= 10 + \left(\frac{1}{2}\right) \times 2$$

$$= 10 + 1$$

$$= 11$$

Median

The following table represents the marks obtained by a batch of 10 students in certain class tests in statistics and Accountancy.

Serial No	1	2	3	4	5	6	7	8	9	10
Marks (Statistics)	53	55	52	32	30	60	47	46	35	28
Marks (Accountancy)	57	45	24	31	25	84	43	80	32	72

Indicate in which subject is the level of knowledge higher ?

Serial No	1	2	3	4	5	6	7	8	9	10
Marks in Statistics	28	30	32	35	46	47	52	53	55	60
Marks in Accountancy	24	25	31	32	43	45	57	72	80	84

$$\text{Median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{item} = \left(\frac{10+1}{2}\right)^{\text{th}} \text{item} = 5.5^{\text{th}} \text{ item}$$

$$= \frac{\textit{Value of } 5^{th} \textit{ item} + \textit{value of } 6^{th} \textit{ item}}{2}$$

$$\text{Md (Statistics)} = \frac{46+47}{2} = 46.5$$

$$\text{Md (Accountancy)} = \frac{43+45}{2} = 44$$

There fore the level of knowledge in Statistics is higher than that in Accountancy.

Median

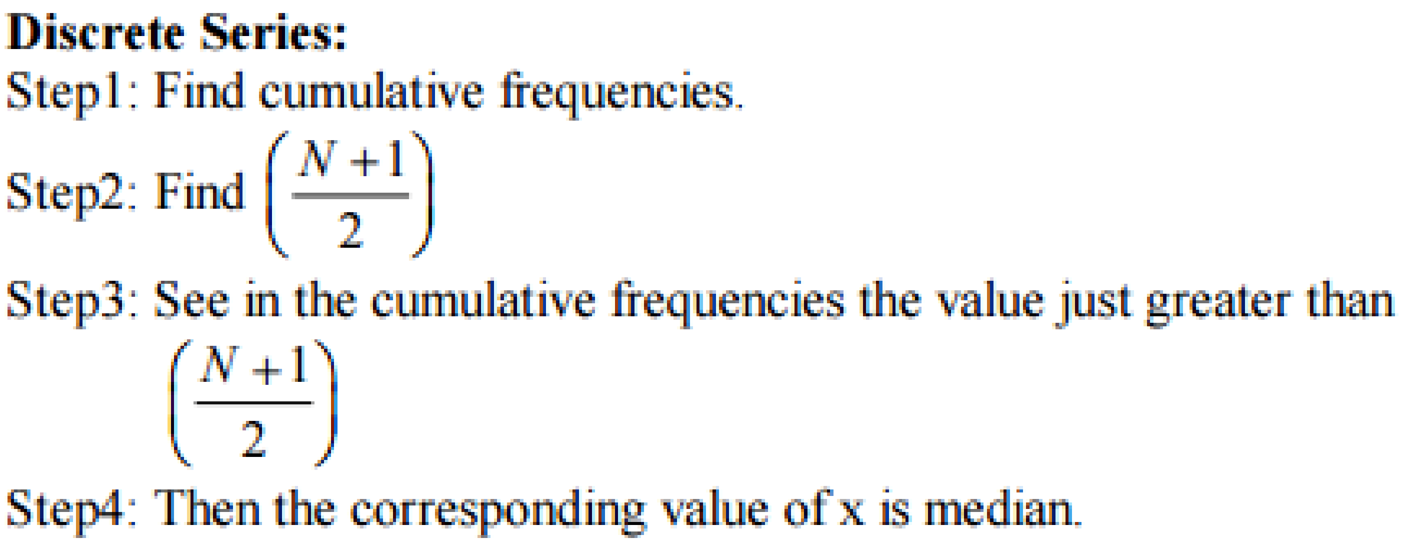
Grouped Data:

In a grouped distribution, values are associated with frequencies. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution.

Whatever may be the type of distribution , **cumulative frequencies** have to be calculated to know the total number of items.

Cumulative frequency: (C.F.) Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the pervious classes, ie adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

Median



Discrete Series:

Step1: Find cumulative frequencies.

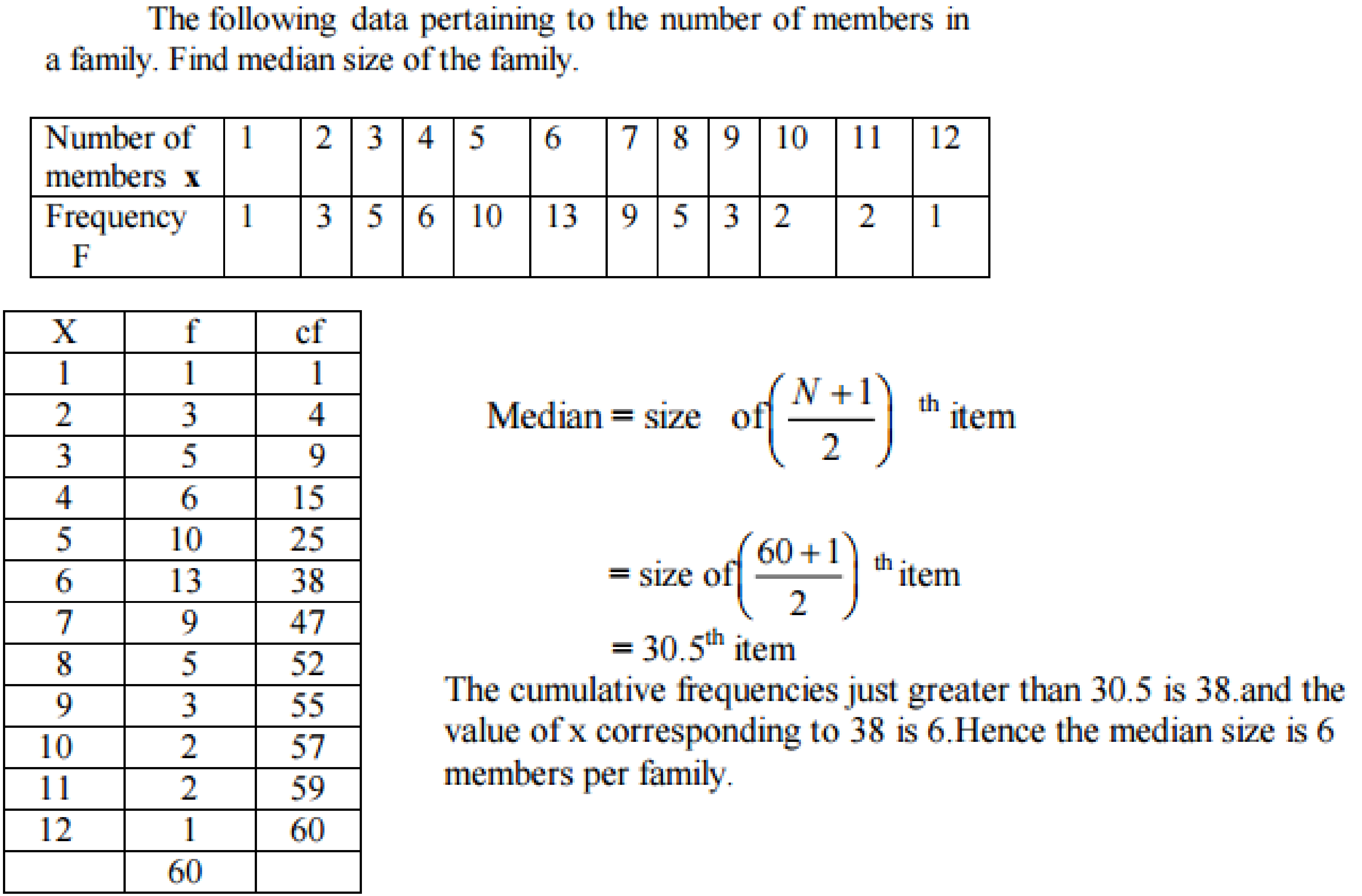
Step2: Find $\left(\frac{N+1}{2}\right)$

Step3: See in the cumulative frequencies the value just greater than

$\left(\frac{N+1}{2}\right)$

Step4: Then the corresponding value of x is median.

Median



The following data pertaining to the number of members in a family. Find median size of the family.

Number of members **x**	1	2	3	4	5	6	7	8	9	10	11	12
Frequency F	1	3	5	6	10	13	9	5	3	2	2	1

X	f	cf
1	1	1
2	3	4
3	5	9
4	6	15
5	10	25
6	13	38
7	9	47
8	5	52
9	3	55
10	2	57
11	2	59
12	1	60
	60	

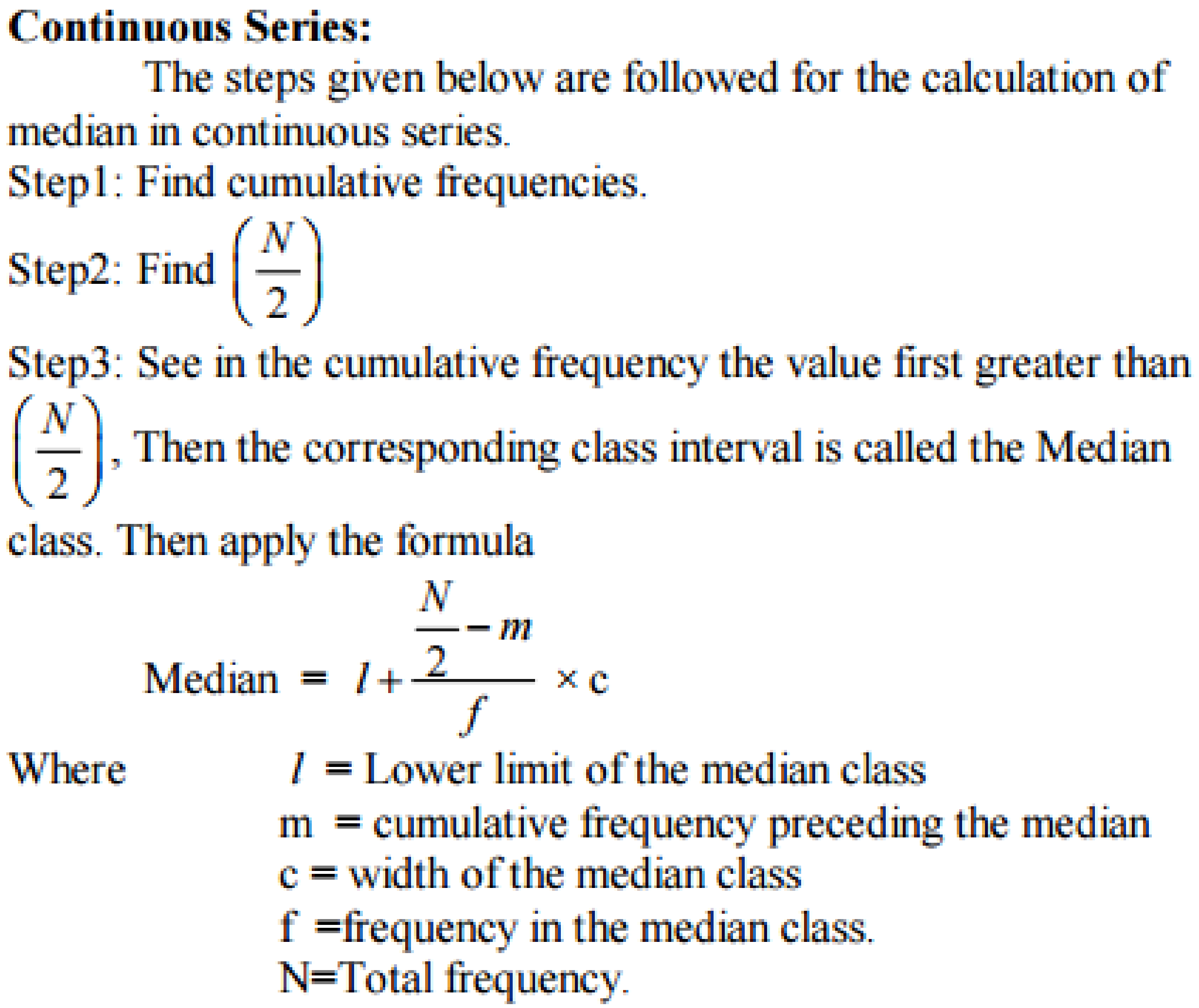
$$\text{Median} = \text{size of} \left(\frac{N+1}{2}\right)^{\text{th}} \text{item}$$

$$= \text{size of} \left(\frac{60+1}{2}\right)^{\text{th}} \text{item}$$

$$= 30.5^{\text{th}} \text{ item}$$

The cumulative frequencies just greater than 30.5 is 38.and the value of x corresponding to 38 is 6.Hence the median size is 6 members per family.

Median



Continuous Series:

The steps given below are followed for the calculation of median in continuous series.

Step1: Find cumulative frequencies.

Step2: Find $\left(\frac{N}{2}\right)$

Step3: See in the cumulative frequency the value first greater than $\left(\frac{N}{2}\right)$, Then the corresponding class interval is called the Median class. Then apply the formula

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where l = Lower limit of the median class

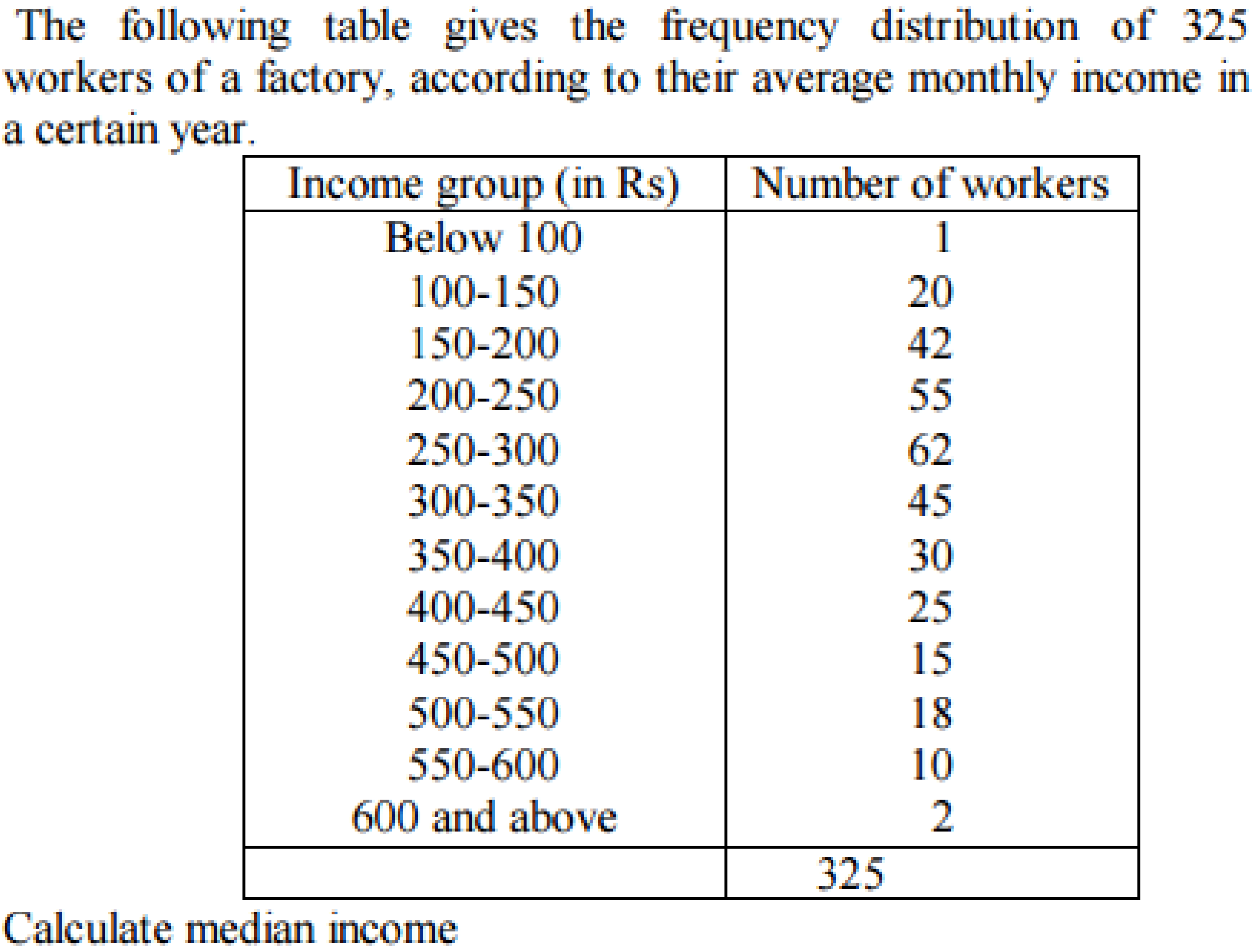
m = cumulative frequency preceding the median

c = width of the median class

f =frequency in the median class.

N=Total frequency.

Median

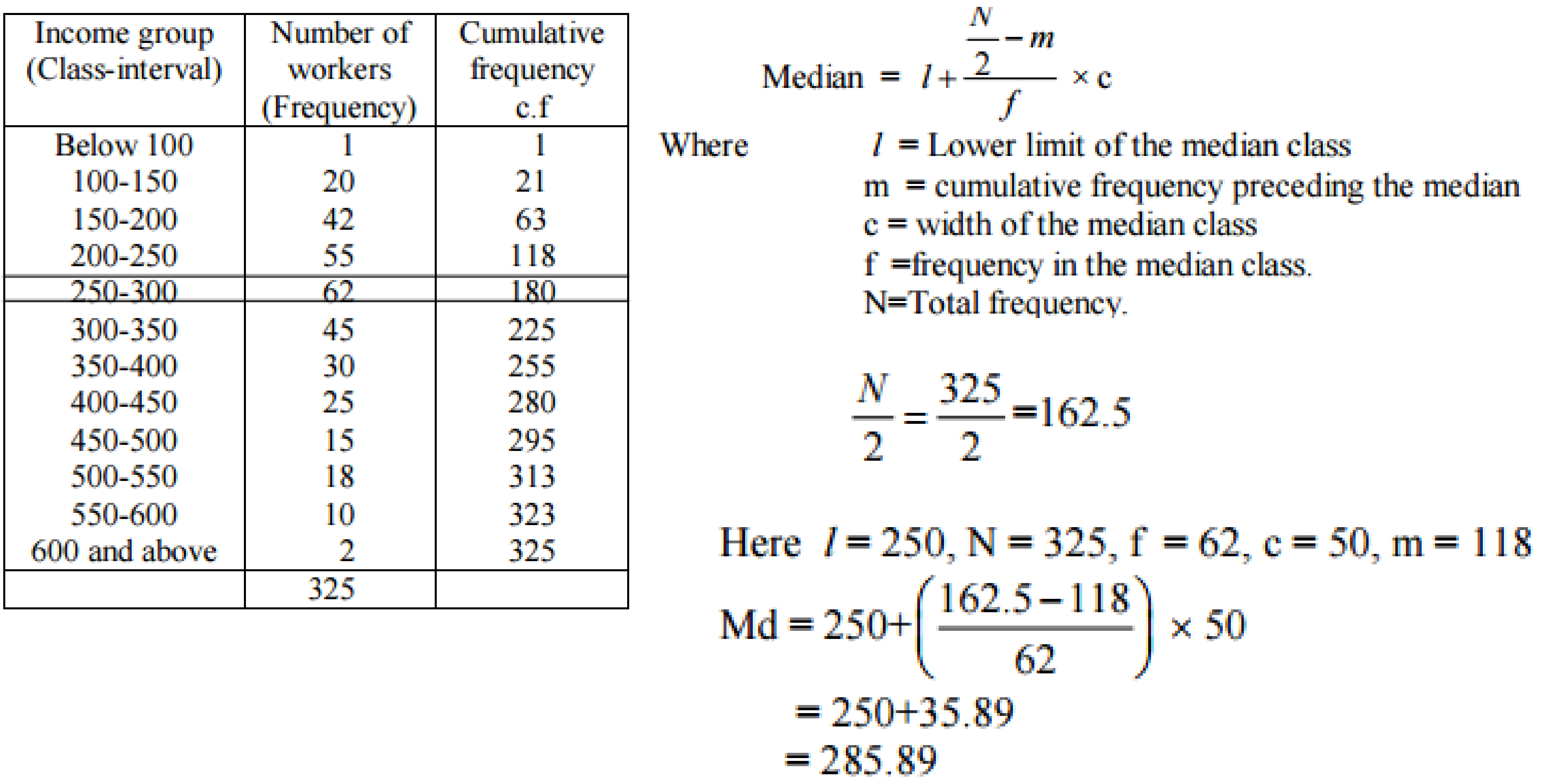


The following table gives the frequency distribution of 325 workers of a factory, according to their average monthly income in a certain year.

Income group (in Rs)	Number of workers
Below 100	1
100-150	20
150-200	42
200-250	55
250-300	62
300-350	45
350-400	30
400-450	25
450-500	15
500-550	18
550-600	10
600 and above	2
	325

Calculate median income

Median



Income group (Class-interval)	Number of workers (Frequency)	Cumulative frequency c.f
Below 100	1	1
100-150	20	21
150-200	42	63
200-250	55	118
250-300	62	180
300-350	45	225
350-400	30	255
400-450	25	280
450-500	15	295
500-550	18	313
550-600	10	323
600 and above	2	325
	325	

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times \text{c}$$

Where l = Lower limit of the median class

m = cumulative frequency preceding the median

c = width of the median class

f =frequency in the median class.

N=Total frequency.

$$\frac{N}{2} = \frac{325}{2} = 162.5$$

Here $l = 250$, N = 325, f = 62, c = 50, m = 118

$$\text{Md} = 250 + \left(\frac{162.5 - 118}{62}\right) \times 50$$

$$= 250 + 35.89$$

$$= 285.89$$

Median

Merits:

- It is rigidly defined.
- Since median is a positional average, it is not affected at all by extreme values. So it is very useful in the case of skewed distributions.
- Median can be computed while dealing with a distribution with open end classes.

Median

- ✓ In case of even number of items or continuous series, median is an estimated value other than any value in the series.
- ✓ It is not suitable for further mathematical treatment except its use in mean deviation.
- ✓ It is not taken into account all the observations.

Partition Values

- ✓ The values which divide the series into a number of equal parts are called the partition values.
- ✓ Median may be regarded as a particular partition value which divides the given data into two equal parts.

Partition Values

- ✓ **Quartiles:** The values which divide the given data into four equal parts are known as quartiles.
- ✓ There will be three such points *Q1, Q2, Q3, such that Q1 < Q2<Q3.*
- ✓ *Q1, known as the lower or first quartile is the value which has 25% of the items of the distribution below it and consequently 75% of the items are greater than it.*
- ✓ *Q2, the second quartile, coincides with the median and has an equal number of observations above and below it.*
- ✓ *Q3, known as the upper or third quartile, has 75% of the observations below it and consequently 25% of the observations above it.*

Quartile Deviation

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Where $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ item and $Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ item

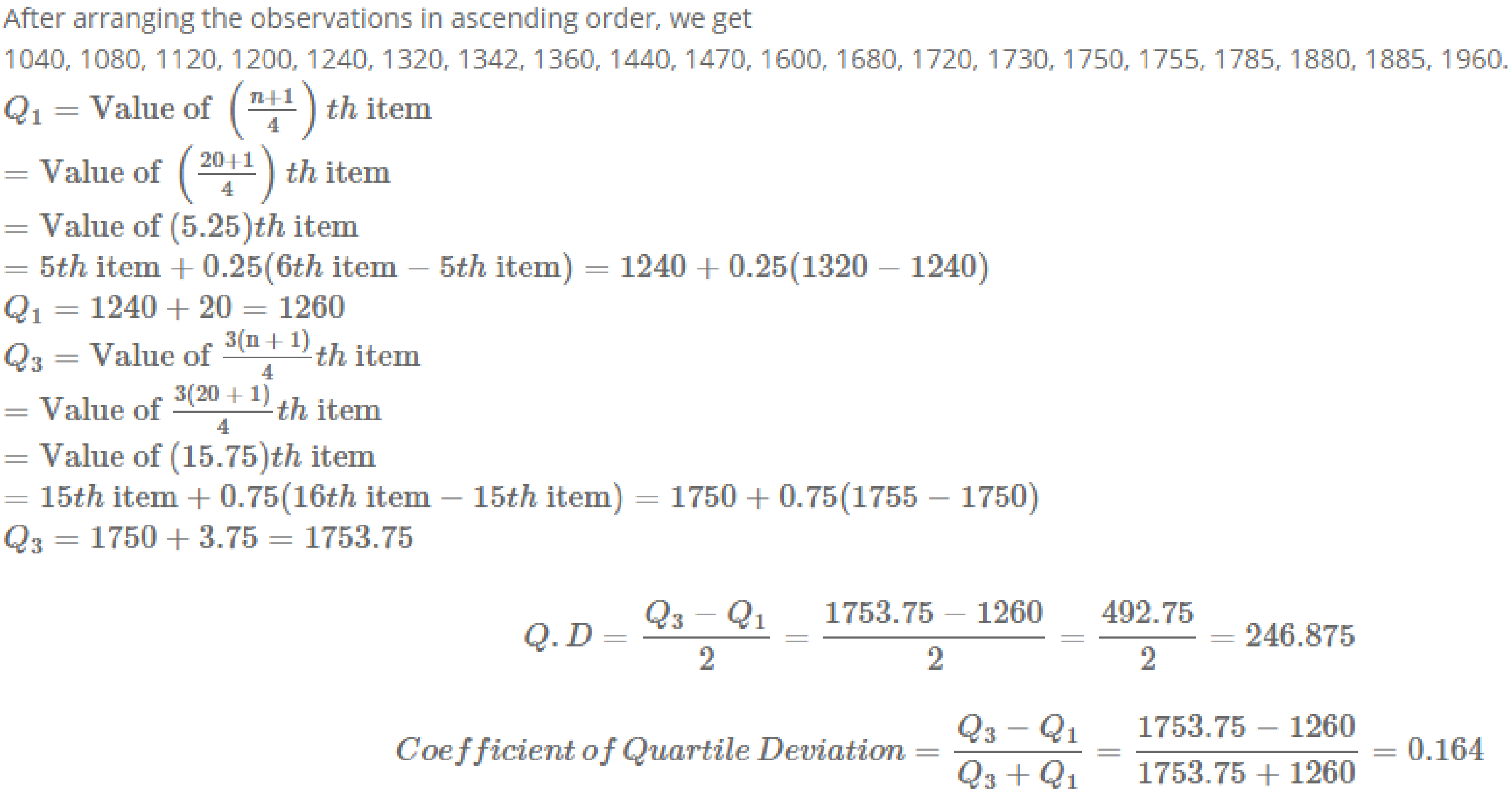
In a symmetric distribution, the two quartiles Q1, and Q3 are equidistant from median. Thus Median(±) quartile deviation coveres 50 % of the observations.

Coefficient of quartile deviation = **(Q3 – Q1)/ (Q3 + Q1)**

Quartile Deviation

The wheat production (in Kg) of 20 acres is given as: 1120, 1240, 1320, 1040, 1080, 1200, 1440, 1360, 1680, 1730, 1785, 1342, 1960, 1880, 1755, 1720, 1600, 1470, 1750, and 1885. Find the quartile deviation and coefficient of quartile deviation.

Quartile Deviation



After arranging the observations in ascending order, we get

1040, 1080, 1120, 1200, 1240, 1320, 1342, 1360, 1440, 1470, 1600, 1680, 1720, 1730, 1750, 1755, 1785, 1880, 1885, 1960.

$Q_1 = \text{Value of } \left(\frac{n+1}{4}\right) th \text{ item}$

$= \text{Value of } \left(\frac{20+1}{4}\right) th \text{ item}$

$= \text{Value of } (5.25) th \text{ item}$

$= 5th \text{ item} + 0.25(6th \text{ item} - 5th \text{ item}) = 1240 + 0.25(1320 - 1240)$

$Q_1 = 1240 + 20 = 1260$

$Q_3 = \text{Value of } \frac{3(\text{n}+1)}{4} th \text{ item}$

$= \text{Value of } \frac{3(20+1)}{4} th \text{ item}$

$= \text{Value of } (15.75) th \text{ item}$

$= 15th \text{ item} + 0.75(16th \text{ item} - 15th \text{ item}) = 1750 + 0.75(1755 - 1750)$

$Q_3 = 1750 + 3.75 = 1753.75$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{1753.75 - 1260}{2} = \frac{492.75}{2} = 246.875$$

$$Coefficient\, of\, Quartile\, Deviation = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1753.75 - 1260}{1753.75 + 1260} = 0.164$$

Quartile Deviation

Compute quartiles for the data given below 25,18,30, 8, 15, 5, 10, 35, 40, 45

Solution :

5, 8, 10, 15, 18,25, 30,35,40, 45

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= \left(\frac{10+1}{4}\right)^{th} \text{ item}$$

$$= (2.75)^{th} \text{ item}$$

$$= 2^{nd} \text{ item} + \left(\frac{3}{4}\right) (3^{rd} \text{ item-}2^{nd} \text{ item})$$

$$= 8 + \frac{3}{4} \quad (10\text{-}8)$$

$$= 8 + \frac{3}{4} \times 2$$

$$= 8 + 1.5$$

$$= 9.5$$

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= 3 \times (2.75)^{th} \text{ item}$$

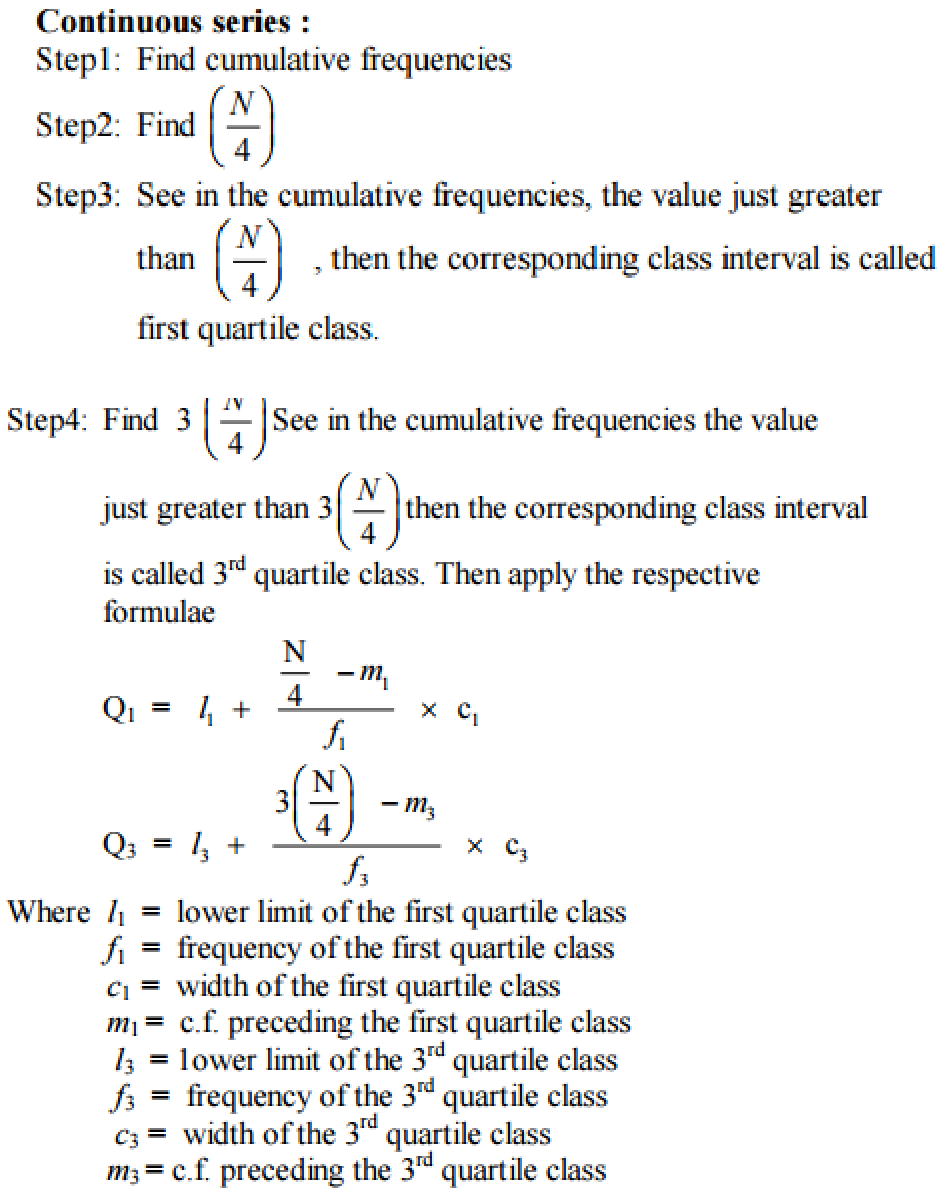
$$= (8.25)^{th} \text{ item}$$

$$= 8^{th} \text{ item} + \frac{1}{4} [9^{th} \text{ item-}8^{th} \text{ item}]$$

$$= 35 + \frac{1}{4} [40\text{-}35]$$

$$= 35+1.25=36.25$$

Quartile Deviation



Continuous series :

Step1: Find cumulative frequencies

Step2: Find $\left(\frac{N}{4}\right)$

Step3: See in the cumulative frequencies, the value just greater than $\left(\frac{N}{4}\right)$, then the corresponding class interval is called first quartile class.

Step4: Find 3 $\left(\frac{N}{4}\right)$ See in the cumulative frequencies the value just greater than $3\left(\frac{N}{4}\right)$ then the corresponding class interval is called 3rd quartile class. Then apply the respective formulae

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$$Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times c_3$$

Where l_1 = lower limit of the first quartile class
f_1 = frequency of the first quartile class
c_1 = width of the first quartile class
m_1 = c.f. preceding the first quartile class
l_3 = 1ower limit of the 3rd quartile class
f_3 = frequency of the 3rd quartile class
c_3 = width of the 3rd quartile class
m_3 = c.f. preceding the 3rd quartile class

C.I.	f	cf
0-10	11	11
10-20	18	29
20-30	25	54
30-40	28	82
40-50	30	112
50-60	33	145
60-70	22	167
70-80	15	182
80-90	12	194
90-100	10	204
	204	

$$\left(\frac{N}{4}\right) = \left(\frac{204}{4}\right) = 51 \qquad 3\left(\frac{N}{4}\right) = 153$$

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$$= 20 + \frac{51-29}{25} \times 10 = 20 + 8.8 = 28.8$$

$$Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times c_3$$

$$= 60 + \frac{153-145}{22} \times 12 = 60 + 4.36 = 64.36$$

Deciles

- Deciles are the values which divide the series into ten equal parts.
- There are 9 deciles. D1, D2 … D9.
- D5 coincides with the median.
- The method of computing the deciles Di(I = 1,2 ….9) is same as discussed for Q1 and Q2. To compute the ith decile, see c.f. just greater than $\frac{i*n}{10}$. The corresponding value of X is Di.

Calculate D_3 and D_7 for the data given below

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency :	5	7	12	16	10	8	4

C.I	f	c.f
0-10	5	5
10-20	7	12
20-30	12	24
30-40	16	40
40-50	10	50
50-60	8	58
60-70	4	62
	62	

$$D_3 \text{ item } = \left(\frac{3N}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{3\times 62}{10}\right)^{th} \text{ item}$$

$$= (18.6)^{th} \text{ item}$$

which lies in the interval 20-30

$$\therefore D_3 = l + \frac{3\left(\frac{N}{10}\right) - m}{f} \times c$$

$$= 20 + \frac{18.6 - 12}{12} \times 10$$

$$= 20 + 5.5 = 25.5$$

$$D_7 \text{ item} = \left(\frac{7\times N}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{7\times 62}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{434}{10}\right)^{th} \text{ item} = (43.4)^{th} \text{ item}$$

which lies in the interval(40-50)

$$D_7 = l + \frac{\left(\frac{7N}{10}\right) - m}{f} \times c$$

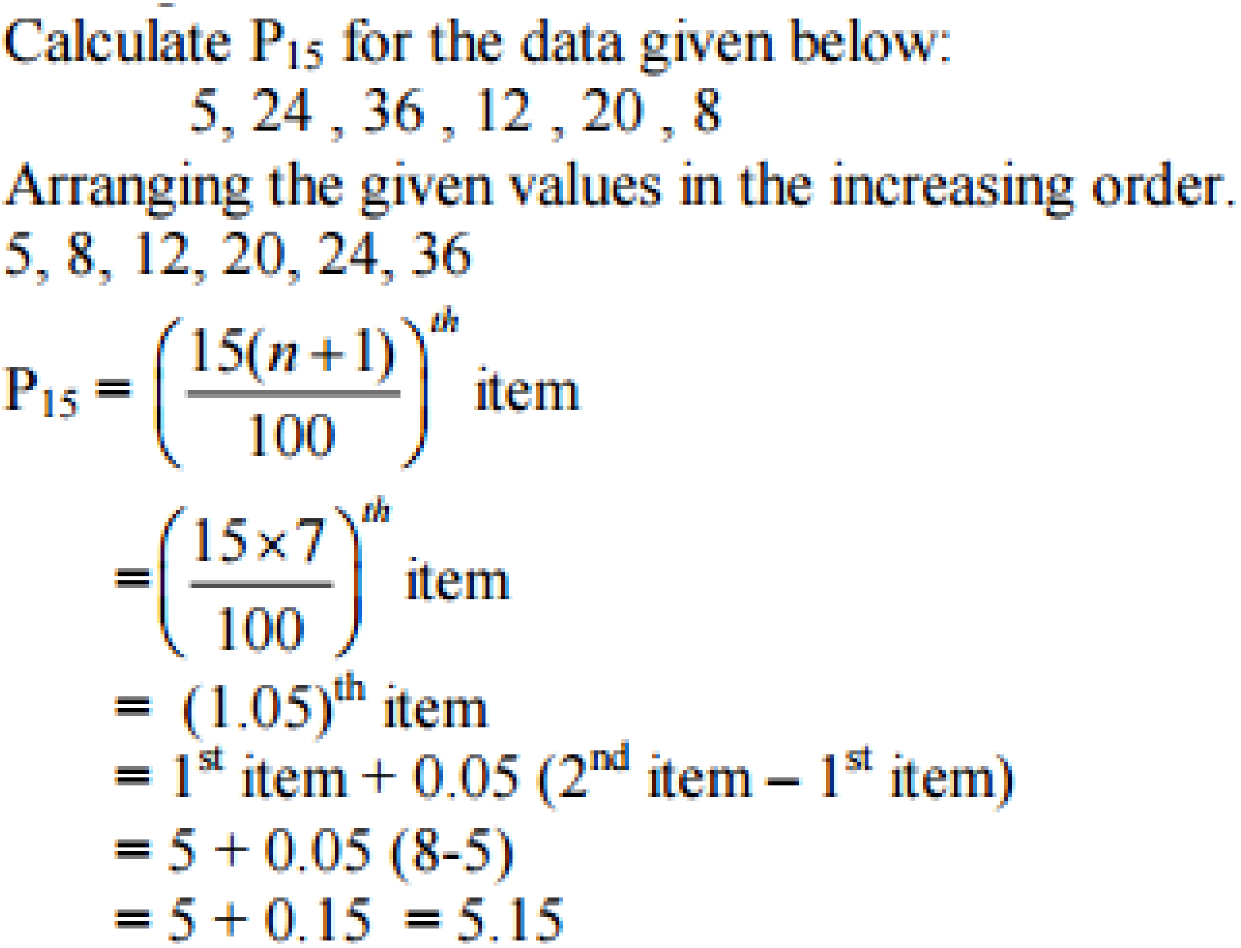
$$= 40 + \frac{43.4 - 40}{10} \times 10$$

$$= 40 + 3.4 = 43.4$$

Percentiles

- The percentile values divide the distribution into 100 parts each containing 1 percent of the cases. The percentile (Pk) is that value of the variable up to which lie exactly k% of the total number of observations.
- P25 = Q1 ;
- P50 = D5 = Q2 = Median
- P75 = Q3

Percentiles



Calculate P_{15} for the data given below:

5, 24 , 36 , 12 , 20 , 8

Arranging the given values in the increasing order.

5, 8, 12, 20, 24, 36

$$P_{15} = \left(\frac{15(n+1)}{100}\right)^{th} \text{ item}$$

$$= \left(\frac{15 \times 7}{100}\right)^{th} \text{ item}$$

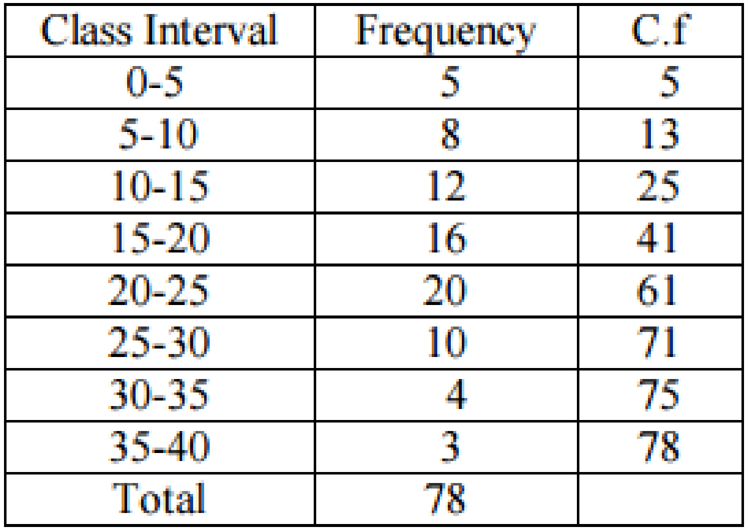
$$= (1.05)^{th} \text{ item}$$

$$= 1^{st} \text{ item} + 0.05\ (2^{nd} \text{ item} - 1^{st} \text{ item})$$

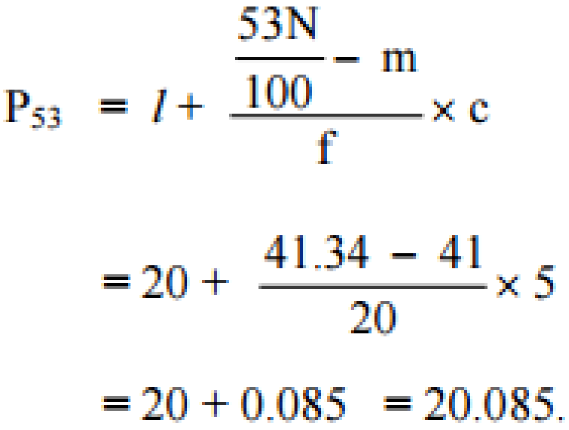
$$= 5 + 0.05\ (8\text{-}5)$$

$$= 5 + 0.15 \ = 5.15$$

Percentiles



Class Interval	Frequency	C.f
0-5	5	5
5-10	8	13
10-15	12	25
15-20	16	41
20-25	20	61
25-30	10	71
30-35	4	75
35-40	3	78
Total	78	



$$P_{53} = l + \frac{\frac{53N}{100} - m}{f} \times c$$

$$= 20 + \frac{41.34 - 41}{20} \times 5$$

$$= 20 + 0.085 \quad = 20.085.$$

Mode

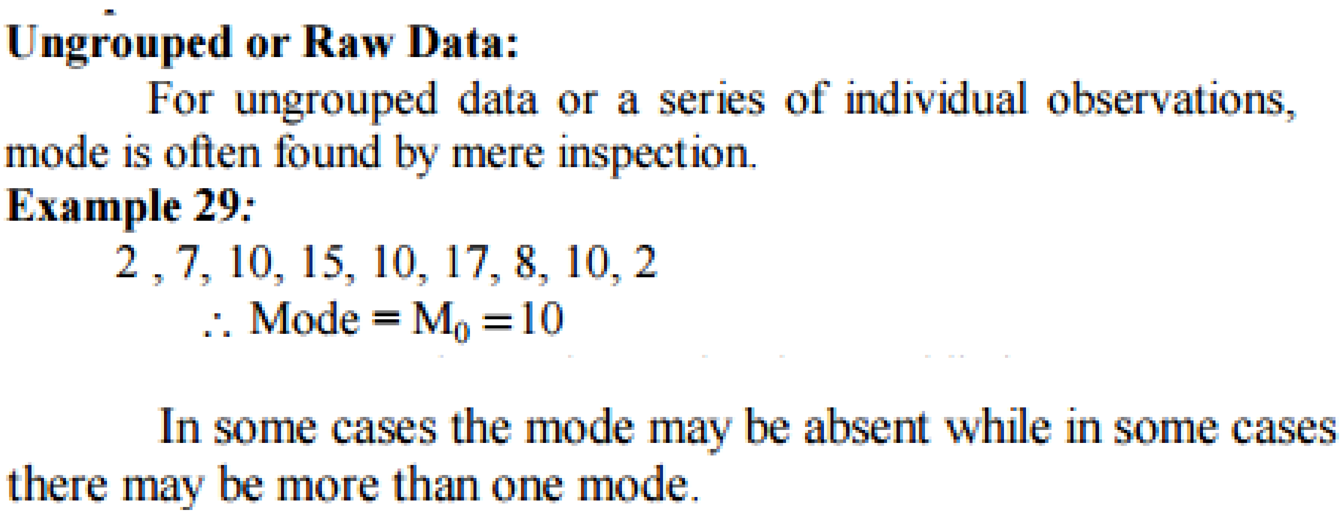
- ✓ Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely.
- ✓ Mode is the value of a series which is predominant in it.

Mode

- The average size of the shoe sold in a shop is 7.
- Average height of an Indian male is 1.66 meters,
- Average size of the shirt sold in a ready-made garment shop is 35 cm.

✓ The average referred to is neither mean nor median but mode. The most frequent value in the distribution.

Mode



Ungrouped or Raw Data:

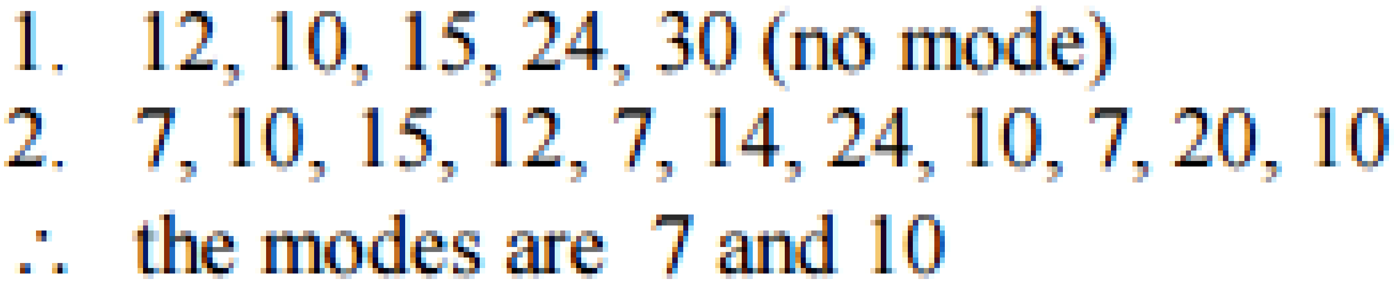
For ungrouped data or a series of individual observations, mode is often found by mere inspection.

Example 29:

2 , 7, 10, 15, 10, 17, 8, 10, 2

$\therefore$ Mode = $M_0 = 10$

In some cases the mode may be absent while in some cases there may be more than one mode.



1. 12, 10, 15, 24, 30 (no mode)
2. 7, 10, 15, 12, 7, 14, 24, 10, 7, 20, 10

$\therefore$ the modes are 7 and 10

Grouped Data:

For Discrete distribution, see the highest frequency and corresponding value of X is mode.

Continuous distribution :

See the highest frequency then the corresponding value of class interval is called the modal class. Then apply the formula.

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

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
The above formula can also be written as

1. If $(2f_1 - f_0 - f_2)$ comes out to be zero, then mode is obtained by the following formula taking absolute differences within vertical lines.
2. $M_0 = l + \dfrac{(f_1 - f_0)}{|f_1 - f_0| + |f_1 - f_2|} \times c$
3. If mode lies in the first class interval, then f_0 is taken as zero.

Calculate mode for the following :

C- I	f
0-50	5
50-100	14
100-150	40
150-200	91
200-250	150
250-300	87
300-350	60
350-400	38
400 and above	15

The highest frequency is 150 and corresponding class interval is 200 – 250, which is the modal class.
Here l=200,f_1=150,f_0=91, f_2=87, C=50

$$\text{Mode} = M_0 = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

$$= 200 + \frac{150\text{-}91}{2 \times 150 - 91 - 87} \times 50$$

$$= 200 + \frac{2950}{122}$$

$$= 200 + 24.18 = 224.18$$

Mode

Merits of Mode:

- It is easy to calculate and in some cases it can be located mere inspection
- Mode is not affected by extreme values.
- It can be calculated for open-end classes.
- It is usually an actual value of an important part of the series.
- In some circumstances it is the best representative of data.

Mode

Demerits of Mode:

- It is not based on all observations.
- It is not capable of further mathematical treatment.
- Mode is ill-defined generally, it is not possible to find mode in some cases.
- As compared with mean, mode is affected to a great extent, by sampling fluctuations.
- It is unsuitable in cases where relative importance of items has to be considered.

Relationship Between AM, GM, HM

- AM ≥ GM ≥ HM
- For two numbers. G^2 = A X H

Selection of Average

✓ Nature and availability of data and purpose plays important role in selecting the average

✓ AM is not recommended while dealing with frequency distribution with extreme observations or open end classes.

✓ Median and mode are averages to be used while dealing with open end classes.

✓ Mode is particularly used in business decisions.

✓ Harmonic mean is to be used in computing special types of average rates or ratios where time factor is variable and the act being performed like distance is constant.

✓ GM is used for calculating returns, diminishing value, etc.