

CHAPTER - 14

STATISTICS

Introduction:

Statistics is the branch of mathematics which deals with the collection, analysis and interpretation of numerical data. In our day – to- day life, we come across a wide variety of information in the form of facts, numerical figures, tables, graphs etc. This information is provided by newspapers, televisions, magazines and other means of communications. You can relate to cricket batting or bowling averages, company profits, recorded city temperatures, expenditures in five year plan of the government, polling results etc. *The facts or figures, which are numerical or otherwise collected with a definite purpose are called data.* In “Latin” the singular form of “data” is “datum”.

Nowadays, we are becoming more and more information oriented and we utilize data in every part of our life in one or the other form. Therefore, it is very important for us to know how useful information can be extracted from such data. This extraction of meaningful information is studied in Statistics a branch of Mathematics.

The word “*Statistics*” appears to have been derived from the Latin word “*Status*” meaning “*a political state*”. In its origin Statistics was simply the collection of data on different aspects of people’s life, useful to the State. But over a period of time, however, its scope broadened. *Thus Statistics deals with the collection, organization, analysis and interpretation of data.*

Mean: The mean of a set of data values is the sum of all the data values divided by the number of data values.

$$\text{Mean} = \frac{\text{Sum of data values}}{\text{No. of data values}}$$

$$\bar{x} = \frac{\sum x}{n}$$

Arithmetic Mean:

$$\bar{x} = \frac{\text{Sum of terms}}{\text{Number of terms}} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

For two numbers “a” and “b”

$$\text{A.M.} = \frac{a + b}{2}$$

$\sum (x_i - \bar{x}) = 0$, where \bar{x} is the Arithmetic mean.

If \bar{x}_1 and \bar{x}_2 are the respective Arithmetic means of two different sets. If data having “ a_1 ” and “ a_2 ” elements respectively. Then the mean of the total set is

$$\bar{x} = \frac{a_1\bar{x}_1 + a_2\bar{x}_2}{a_1 + a_2} = (a_1 a_2 \dots a_n)^{1/n}$$

Geometric Mean:

For two numbers “a” and “b”

$$\text{G.M.} = \sqrt{ab} = (ab)^{1/2}$$

Harmonic Mean:

$$\text{H.M} = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_n}}$$

For two numbers “a” and “b”

$$\text{H.M} = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2ab}{a + b}$$

Note AM \geq GM \geq HM

Mean of Ungrouped Data:

The information collected systematically regarding a population or a sample survey is called *ungrouped data*.

$$\text{Mean} = \frac{\text{Sum of Observations}}{\text{Number of observations}}$$

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

Mean of grouped Data:

If x_1, x_2, \dots, x_n are n -observations with respective frequencies $f_1, f_2, f_3, \dots, f_n$. then the mean.

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \dots + f_nx_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$$

Example: Mean of the data shown below will be

Marks (f_i)	20	30	35	40
Number of students	10	6	6	2

Solution:

$$\begin{aligned}\bar{x} &= \frac{20 \times 10 + 30 \times 6 + 35 \times 6 + 40 \times 2}{10 + 6 + 6 + 2} \\ &= \frac{200 + 180 + 210 + 80}{24} = 27.91\end{aligned}$$

Assumed mean method of calculating mean

$$\begin{aligned}\bar{x} &= a + \frac{\sum f_i d_i}{\sum f_i} = a + \bar{d} \\ \bar{d} &= \bar{x} - a = \bar{x} - a = \frac{\sum (f_i x_i)}{\sum f_i} = \frac{\sum (f_i x_i)}{\sum f_i} - \frac{\sum f_i a}{\sum f_i} \\ &= \frac{\sum f_i (x_i - a)}{\sum f_i} = \frac{\sum f_i d_i}{\sum f_i}\end{aligned}$$

Example: The table below shows the number of people within different age group who visited the mall on week end.

Age group (Class interval)	10 – 25	25 – 40	40 – 55	55 – 70	70 – 85	85 - 100
Number of people	3	11	10	8	6	2

Solution:

Class Interval	Number of people (f_i)	Class mark (x_i)	$d_i = x_i - 47.5$	$f_i d_i$
10 – 25	3	17.5	- 30	-90
25 – 40	11	32.5	- 15	-165
40 – 55	10	47.5	0	0

Class Interval	Number of people (f_i)	Class mark (x_i)	$d_i = x_i - 47.5$	$f_i d_i$
55 – 70	8	62.5	15	120
70 – 85	6	77.5	30	180
85 – 100	2	92.5	45	90
	$\Sigma f_i = 40$			$\Sigma f_i d_i = 135$

Thus
$$\bar{x} = a + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

$$= 47.5 + \frac{135}{40} = 50.875$$

Step deviation method:

$$\bar{x} = a + h \left(\frac{\Sigma f_i u_i}{\Sigma f_i} \right)$$

$$u_i = \frac{x_i - a}{h}$$

Where “a” is the assumed mean and “h” is the class size.

Example:

Number of wickets	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65	65 – 75	75 – 85
Number of bowlers	6	6	7	4	4	2	1

Solution:

Class Interval	Number of bowlers (f_i)	(x_i)	$d_i = x_i - 50$	$u_i = \frac{x_i - 50}{10}$	$f_i u_i$
15 – 25	6	20	-30	-3	-18
25 – 35	6	30	-20	-2	-12
35 – 45	7	40	-10	-1	-7

Class Interval	Number of bowlers (f_i)	(x_i)	$di = x_i - 50$	$u_i = \frac{x_i - 50}{10}$	$f_i u_i$
45 – 55	4	50	0	0	0
55 – 65	4	60	10	1	4
65 – 75	2	70	20	2	4
75 – 85	1	80	30	3	3
	$\Sigma f_i = 30$				$= -26$

$$\bar{x} = a + \left(\frac{\Sigma f_i u_i}{\Sigma f_i} \right) h$$

$$= 50 + \left(\frac{-26}{30} \right) \times 10 = 41.33$$

Median: Arrange the numbers in ascending or descending order, and why they are arranged as such, median is

- The middle term when number of terms is odd.
- Then average of middle two terms when the number of terms is even.
- Divide the distribution in two equal parts

Median of grouped Data:

Data which have been arranged in groups or classes rather than showing all the original figures.

$$\text{Median} = l + \left(\frac{\frac{n}{2} - c_f}{f} \right) \times h$$

Where

- l = lower limit of median class.
- n = number of observations
- c_f = cumulative frequency of class preceding the median class
- f = frequency of median class
- h = class size

Note: Find cumulative frequencies of all the classes and $\frac{n}{2}$ locate the class where cumulative frequency is greater than $\frac{n}{2}$. That is called median class.

Mode: The number which has the highest frequency in the mode.

Mode of grouped Data:

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

Where

- l = lower limit of median class. (Class with maximum frequency)
- h = size of class interval
- f_0 = frequency of class preceding the modal class
- f_1 = frequency of modal class
- f_2 = frequency of succeeding to the modal class.

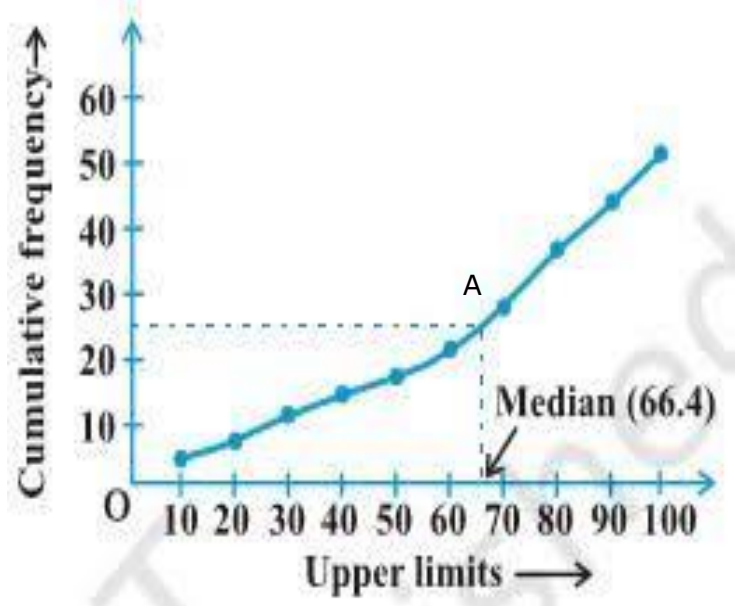
Note: *The empirical formula says Mode = 3 (median) – 2 (mean)*

There are three methods of drawing Ogive.

1. Less than Method:

Steps involved in calculating median using less than Ogive approach:

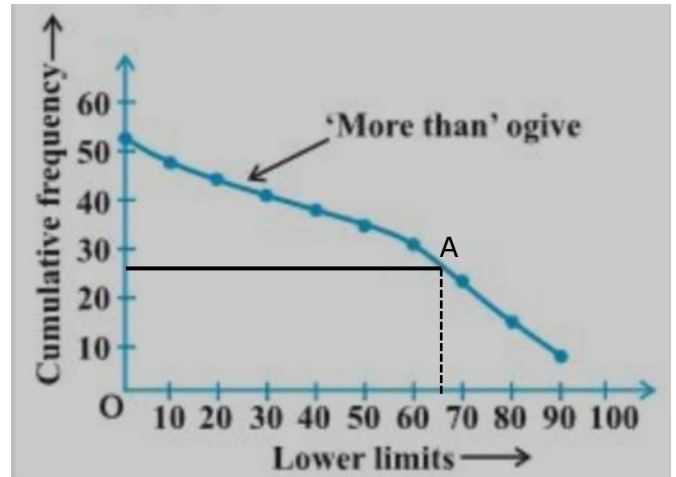
- ❖ Convert the series into a “less than” cumulative frequency distribution.
- ❖ Let “ N ” be the total number of students whose data is given. “ N ” will also be the cumulative frequency of the last interval. Find the $(\frac{N}{2})^{\text{th}}$ item and mark it on the Y-axis.
- ❖ Draw a perpendicular from that point to the right to cut the Ogive curve at the point “A”.
- ❖ From point “A” where the Ogive curve is cut, draw a perpendicular on the X-axis. The point at which it touches the X-axis will be median value of the series as shown in the graph:



2. More than Method:

Steps involved in calculating median using more than Ogive approach:

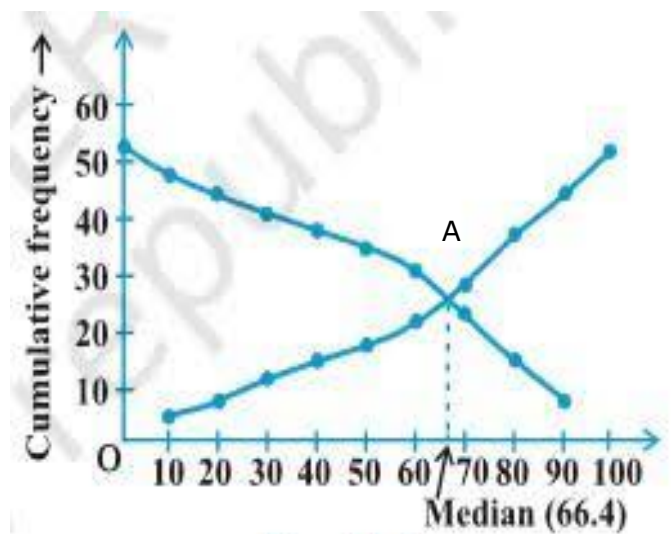
- ❖ Convert the series into a “**more than**” cumulative frequency distribution.
- ❖ Let “ N ” be the total number of students whose data is given. “ N ” will also be the cumulative frequency of the last interval. Find the $(\frac{N}{2})^{\text{th}}$ item and mark it on the Y-axis.
- ❖ Draw a perpendicular from that point to the right to cut the Ogive curve at the point “ A ”.
- ❖ From point “ A ” where the Ogive curve is cut, draw a perpendicular on the X-axis. The point at which it touches the X-axis will be median value of the series as shown in the graph:



3. Less than and More than Ogive Method:

Another way of graphical determination of median is through simultaneous graphic presentation of both the “less than” and “more than” Ogives:

- ❖ Mark the point “ A ” where the Ogive curves cut each other.
- ❖ Draw a perpendicular from “ A ” on the X-axis. The corresponding value on the X-axis would be the median value.
- ❖ The median of grouped data can be obtained graphically as the X-coordinate of the point of intersection of the two ogives of the data.



1 MARK QUESTIONS

- Q.1** The median for the data 2,4,6,8,10,12,14 is
a) 6 b) 8 c) 9.5 d) 10
- Q.2** The mean weekly pay for ten persons equals to Rs. 100, if one of the person gets a taxi of Rs. 10 per week, what is the new mean weekly pay
a) Rs. 99 b) Rs. 101 c) Rs. 200 d) Rs. 250
- Q.3** Each of the group formed from given data is called:
a) frequency b) raw data c) median d) class-interval
- Q.4** Mean of the 5 items of a data is 10. If each term is multiplied by 4, then the new mean will be
a) 40 b) 50 c) 30 d) 60
- Q.5** The Harmonic mean of 2, 4, 6 and 8 is
a) 3 b) 3.5 c) 3.84 d) 9
- Q.6** If the AM as well as GM of two +ve numbers is 8, what is their H.M.
a) 2 b) 4 c) 8 d) None of these
- Q.7** Geometric mean of 2, 6, 24 and 72 is
a) 12 b) $13\sqrt{3}$ c) $8\sqrt{3}$ d) none of these
- Q.8** Given the mean (30), the mode (35), what is the median
a) 30 b) 31.69 c) 40 d) 55.7
- Q.9** The range of the data 12, 16,17,23,29,13,2,5,19 is
a) 23 b) 24 c) 27 d) 29
- Q.10** The mean of 80 items was 42. Later it was found that the two items were misread as 87 and 6 instead of 187 and 66. Which of the following will be the correct mean.
a) 44 b) 45 c) 42.6 d) none of these

ANSWERS

- | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|
| Q.1 | (b) | Q.2 | (b) | Q.3 | (d) | Q.4 | (a) | Q.5 | (c) |
| Q.6 | (c) | Q.7 | (a) | Q.8 | (b) | Q.9 | (c) | Q.10 | (a) |

2 MARK QUESTIONS:

Q.1 Define mode.

Q.2 If the mean of the following data is 15

x: 5 10 15 20 25

f: 6 P 6 10 5

Find P.

Q.3 In a continuous frequency distribution, the median of the data is 21. If each observation is increased by 5, then find the new median.

Q.4 The mean of the following frequency table is 50, but the frequency f_1 and f_2 in class interval 20 – 40 and 60 – 80 respectively are not known. Find these frequencies, when sum of all the frequency is 120.

Class	Frequencies
0 – 20	17
20 – 40	f_1
40 – 60	32
60 – 80	f_2
80 – 100	19
Total	120

Q.5 Find the mode of the following frequency distribution.

Marks	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Number of students	12	35	45	25	13

Q.6 A class teacher has the following absentee record of 40 students of a class for the whole term. Find the mean number of days a student was absent.

Number of days	Number of students
0 – 6	11
6 – 10	10
10 – 14	7
14 – 20	4
20 – 28	4
28 – 38	3
38 – 40	1

Q.7 Give three examples of collecting data from day to day life.

Q.8 Find the mode of the following items.

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

Q.9 A student scored the following marks in 6 subject

30, 19, 25, 30, 27, 30.

Find his modal score

Q.10 Find mode using an empirical relation, when it is given that mean and median are 10.5 and 9.6 respectively.

Q.11 In a frequency distribution, if “a” assumed mean = 55. $\sum fi = 100$, $h = 10$ and $\sum fiui = - 30$, then find mean of the distribution.

ANSWER

Q.2 (8)	Q.3 (26)	Q.4 (1300 – 1450)	Q.5 (3)
Q.6 (5)	Q. 8 (5)	Q.9 (30)	Q.10 (7.8) Q.11 (52)

3 MARK QUESTIONS

Q.1 Find the value of “y” from the following observations, if they are already arranged in ascending order. The median is 63.

20, 24, 42, y, y+2, 73, 75, 80, 99

Q.2 While checking the value of 20 observations, it was noted that 125 was wrongly noted as 25 while calculating the mean and then mean was 60. Find the correct mean.

Q.3 The daily minimum steps climbed by a person during a week were as under:

Monday	35
Tuesday	30
Wednesday	27
Thursday	32
Friday	23
Saturday	28

Find the mean number of steps.

Q.4 From the following frequency distribution, find the median class.

Class Interval	Frequency
1000 – 1150	8
1150 – 1300	15
1300 – 1450	21
1450 – 1600	8

Q.5 Consider the following distribution, find the frequency of class 30 – 40.

Marks	No. of Students
0 or more	63
10 or more	58
20 or more	55
30 or more	51
40 or more	48
50 ore more	42

Q.6 Following table shows sale of shoes in a store during the month

Shoe Size	Pairs sold
3	4
4	18
5	25
6	12
7	5
8	1

Find the modal size of the shoes sold.

Q.7 The following table shows the ages of the patients admitted in a hospital during a year.

Age (in years)	5 – 15	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65
Number of patients	6	11	21	23	14	5

Find the mode of the data.

Q.8 If the median of the distribution given below is 28.5. Find the values of “x” and “y”.

Class Interval	Frequency
0 – 10	5
10 – 20	x
20 – 30	20
30 – 40	15
40 – 50	y
50 – 60	5
Total	60

Q.9 A company manufactures car batteries of a particular type. The lives of 40 such batteries were recorded as below:

Life of batteries (in years)	2 – 2.5	2.5 – 3.0	3.0 – 3.5	3.5 – 4.0	4.0 – 4.5	4.5 – 5.0
No. of batteries	2	6	14	11	4	3

Find the modal life of a battery in years.

Q.10 Consider the following distribution:

Number of plants	0 – 2	2 – 4	4 – 6	6 – 8	8 – 10	10 - 12	12 – 14
No. of Houses	1	.2	1	5	6	2	3

ANSWERS

Q.1 (61) Q.2 (65) Q.3 (29.17) Q.4 ($f_1=28, f_2=24$)
 Q.5 (33.33) Q.6 (12.48 days) Q.7 (36.8 years) Q.8 ($x=8, y=7$)
 Q.10 (8.1)

4 MARK QUESTIONS

Q.1 The average score of boys in the examination of a school is 71 and that of the girls is 73. The average score of the school in the examination is 71.8. Find the ratio of the number of boys to the number of girls who appeared in the examination.

Q.2 Some students of a class donated for the welfare of old age persons. The contributions are as follows:

Amount in Rs.	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of students	5	8	12	11	4

Find median and mode for their contribution.

Q.3 In a retail market, fruit vendors were selling mangoes in the packing boxes. These boxes contained varying number of mangoes. The following was the distribution.

No. of mangoes	50 – 52	53 – 55	56 – 58	59 – 61	62 – 64
No. of boxes	15	110	135	115	25

Find the mean and median number of mangoes kept in a packing box.

Q.4 The length of 50 leaves of a plant are measured correct to the nearest millimeter and the data obtained is represented in the following table:

Length in (mm)	109 – 117	118 – 126	127 – 135	136 – 144	145 – 153	154 – 162	163 – 171
No. of leaves	4	6	14	13	6	4	3

Find the mean length of the leaves.

Q.5 In a hospital, during the month of October, number of patients admitted for dengue and their ages are as follows:

Age in years	0 – 8	8 – 16	16 – 24	24 – 32	32 – 40	40 – 48	48 – 56	56 – 64	64 – 72
No. of patients	10	12	8	25	15	11	21	30	22

Find the mean and median age of patients

Q.6 Find the missing frequencies (f_1 , f_2 and f_3) in the following frequency distribution when it is given that $f_1, f_2 = 4:3$ and mean = 50.

Class Interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	Total
Frequency	17	f_1	f_2	f_3	19	20

Q.7 A student noted the number of cars passing through a spot on a road for 100 periods, each of 3 minutes and summarized it in the table. Find the mode of the data.

Number of cars	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency	7	14	13	12	20	11	15	8

Q.8 Draw “less” than “Ogive” for the following frequency distribution.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
No. of students	5	3	4	3	3	4	7	9	7	8

Q.9 Draw “more than” Ogive for the following distribution. Hence find median.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
No. of students	5	3	4	3	3	4	7	9	7	8

Q.10 Draw “less than” Ogive and “more than” Ogive for the following distribution

Height	135 – 140	140 – 145	145 – 150	150 – 155	155 – 160	160 – 165
No of Plants	4	7	18	11	6	4

ANSWERS

Q.1 (3:2) Q.2 (Median = 51.66, Mode = 56) Q.3 (57.16)

Q.4 (137.30 mm) Q.5 (Mean = 41.92, Median = 45.09)

Q.6 ($f_1=28, f_2 = 32, f_3 = 24$) Q.7 (44.7) Q.9 (66)