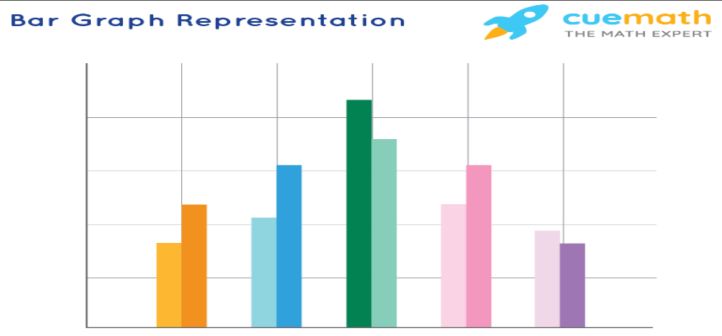
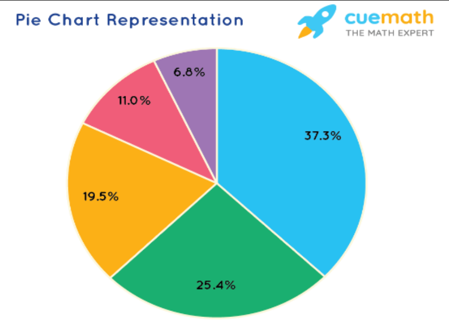
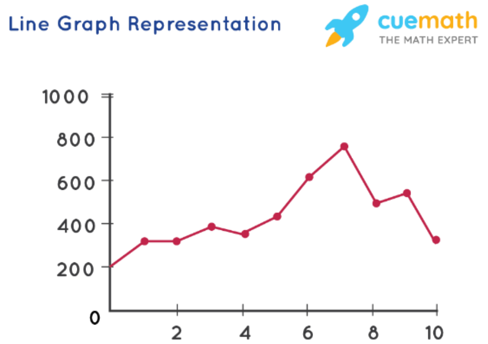
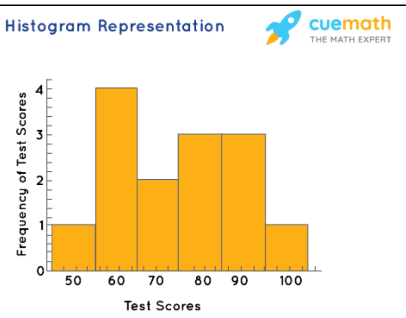
Statistics

AUC-CET-23-24-DAWAH

Statistics

Statistics is a branch of science dealing with the collection of data, organizing summarizing, presenting and analyzing data and drawing valid conclusions and thereafter making reasonable decisions on the basis of such analysis

<table>
<tr><th>Data Representation</th><th>Description</th></tr>
<tr><td>
</td><td>Bar Graph A group of data represented with rectangular bars with lengths proportional to the values is a .bar graph
The bars can either be vertically or horizontally plotted</td></tr>
<tr><td>
</td><td>Pie Chart
The pie chart is a type of graph in which a circle is divided into Sectors where each sector represents a proportion of .the whole</td></tr>
<tr><td>
</td><td>Line graph
The line graph represents the data in a form of series that is connected with a straight line. These series are called .markers</td></tr>
<tr><td>
</td><td>Histogram The histogram is a type of graph where the diagram consists of rectangles, the area is proportional to the frequency of a variable and the width is equal to the class interval. Here is an example of a histogram</td></tr>
</table>

Frequency Distribution Table — cuemath THE MATH EXPERT

Height	Frequency
139	1
145	2
150	2
136	1
152	1
144	1
138	2

Frequency Distribution

The frequency distribution table in statistics showcases the data in ascending order along with their .corresponding frequencies

The frequency of the data is often .represented by f

in every statistical study usually consists of:

1. To classify, group and sort the data of the sample.

2. To tabulate and plot data according to their frequencies.

3. To calculate numerical measures that summarize the information contained in the sample (*sample statistics*).

Frequency distribution is the arranged data, summarized by distributing it into classes or categories with their frequencies.

There are two ways of classifying data:

Non-grouping: Sorting values from lowest to highest value (if there is an order). Used with qualitative variables and discrete variables with few distinct values.

1, 2, 4, 2, 2, 2, 3, 2, 1, 1, 0, 2, 2, 0, 2, 2, 1, 2, 2, 3, 1, 2, 2, 1, 2

Grouping: Grouping values into intervals (classes) and sort them from lowest to highest intervals. Used with continuous variables and discrete variables with many distinct values.

185 ,111 ,111 ,172 ,171 ,158 ,171 ,181 ,173 ,171 ,171 ,111 ,188 ,151 ,115 ,178 ,177 ,118 ,187 ,112 ,175, 182, 167, 169, 172, 186, 172, 176, 168, 187.

Sample classification :

It consists in grouping the values that are the same and sorting them if there is an order among them. Example. X = Height

Statistics

Frequency count :

It consists in counting the number of times that every value appears in the sample. Example. X=Height

Sample frequencies : Definition Sample frequencies. Given a sample of **(n)** values of a variable **(x)**, for every value of the variable **(xi)** we define Absolute Frequency **ni** : The number of times that value **xi** appears **xi** in the sample. Relative Frequency **fi** : The proportion of times that value appears in the sample.

$$f_i = n_i / n$$

Cumulative Absolute Frequency N_i: The number of values in the sample less than or equal to x_i.

$$N_i = n_1 + \dots + n_i = N_{i-1} + n_i$$

EXAMPLE/ teacher gave a test in statistics to his student there marks were

3	8	6	5	6	4	7	6
5	3	5	6	3	5	4	4
3	6	7	8	1	10	7	6
4	5	0	7	6	5	6	7
1	7	5	4	5	8	5	7

The marks could be tabulated as follows:

Frequency Distribution Table

Mark (x)	Tally mark	Frequency (f)
0	/	1
1	//	2
2		0
3	////	4
4	~~////~~	5
5	~~////~~ ////	9
6	~~////~~ ///	8
7	~~////~~ //	7
8	///	3
9		0
10	/	1
		$\sum f = 40$

Now it is easier to gather the following types of information from the frequency table than the raw data.

(a) The highest mark is 10 and the lowest is 0.

(b) 4 students scored more than 7.

(c) 12 students scored less than 5.

(d) 8 students scored 6.

(e) Noone scored 9.

(f) 40 students did the test.

$$\text{Relative frquency of a score} = \frac{\text{Frequency of the score}}{\text{Total frequency}} = \frac{f}{\sum f}$$

Relative Frequency Distribution

Marks (x)	Frequency (f)	Relative frequency
0	1	1/40=.025
1	2	2/40=.050
2	0	0/40=.000
3	4	=.100
4	5	=.125
5	9	=.225
6	8	=.200
7	7	=.175
8	3	=.075
9	0	=.000
10	1	=.025
	$\sum f = 40$	

Frequency table :

The set of values of a variable with their respective frequencies is called frequency distribution of the variable in the sample, and it is usually represented as a frequency table.

X values	Absolute frequency	Relative frequency	Cumulative absolute frequency	Cumulative relative frequency
x_1	n_1	f_1	N_1	F_1
⋮	⋮	⋮	⋮	⋮
x_i	n_i	f_i	N_i	F_i
⋮	⋮	⋮	⋮	⋮
x_k	n_k	f_{kf}	N_k	F_k

Example - Quantitative variable and non-grouped data.

Find (fi, Ni, Fi) for the following of number of children in 25 families are:

1, 2, 4, 2, 2, 2, 3, 2, 1, 1, 0, 2, 2, 0, 2, 2, 1, 2, 2, 3, 1, 2, 2, 1, 2

Solution :The frequency table for the number of children in this sample is

Relative Frequency f_i: $f_i = n_i / n$

Cumulative Absolute Frequency N_i: $N_i = n_1 + \cdots + n_i$

Cumulative Relative Frequency F_i: $F_i = N_i / n$

x_i	n_i	f_i	N_i	F_i
0	2	0.08	2	0.08
1	6	0.24	8	0.32
2	14	0.56	22	0.88
3	2	0.08	24	0.96
4	1	0.04	25	1
$\sum$	25	1		

$$f_i = \frac{n_i}{\sum ni}$$

$$Fi = \frac{Ni}{\sum ni}$$

Example - Quantitative variable and grouped data. The heights (in cm) of 30 students are:

179, 173, 181, 170, 158, 174, 172, 166, 194, 185,
162, 187, 198, 177, 178, 165, 154, 188, 166, 171, 175, 182, 167, 169, 172, 186, 172, 176, 168, 187.

Solution: The frequency table for the height in this sample is :

Relative Frequency f_i: $f_i = n_i / n$

Cumulative Absolute Frequency N_i: $N_i = n_1 + \cdots + n_i$

Cumulative Relative Frequency F_i: $F_i = N_i / n$

x_i	n_i	f_i	N_i	F_i
(150, 160]	2	0.07	2	0.07
(160, 170]	8	0.27	10	0.34
(170, 180]	11	0.36	21	0.70
(180, 190]	7	0.23	28	0.93
(190, 200]	2	0.07	30	1
$\sum$	30	1		

EXAMPLE/Wages of 100 workers

Wages in Rs.	0-10	10-20	20-30	30-40	40-50
Numbers of workers	12	23	35	20	10

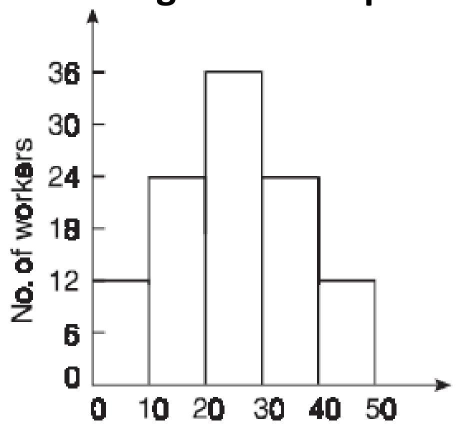
Graphical representation. It is often useful to represent frequency distribution by: means of a diagram. The different types of diagrams are

1-Histogram.

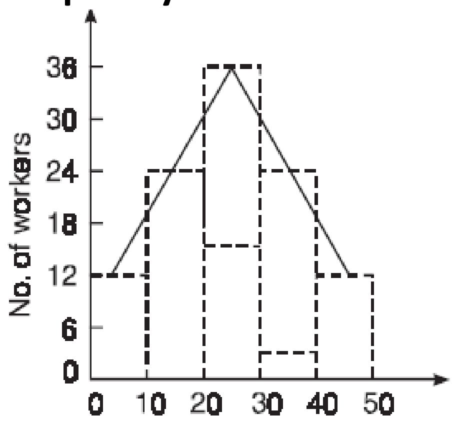
2-Frequency polygon .

3-Frequency curve .

1-Histogram consists of a set of rectangles having their heights proportional to the class frequencies for equal class-intervals. For unequal class-interval, the areas of rectangles are Proportional to the frequency .



Histogram



Frequency polygon

2-Frequency Polygon is a line graph of class-frequency plotted against class-mark.it can be obtained by connecting mid-points on the tops of the rectangles in the histogram

AVERAGE OR MEASURES OF CENTRAL TENDENCY

An average is a value which is representative of a set of data. Average value may also be termed as measures of central tendency. There are *five* types of averages in common

(i) Arithmetic average or mean

(ii) Median

(iii) Mode

(iv) Geometric Mean

(v) Harmonic Mean

In general, when all the central tendency statistics can be calculated, is advisable to use them as representative values in the following order:

1. The mean. Mean takes more information from the sample than the others, as it takes into account the magnitude of data.

2. The median. Median takes less information than mean but more than mode, as it takes into account the order of data.

3. The mode. Mode is the measure that fewer information takes from the sample, as it only takes into account the absolute frequency of values.

But, *be careful with outliers*, as the mean can be distorted by them. In that case it is better to use the median as the value most representative.

Sample arithmetic mean *X*. The *sample arithmetic mean* of a variable *X* is the

(I)- //sum of observed values in the sample divided by the sample size:

ARITHMETIC MEAN

(a)- If x1, x2 ,x3, ... xn are n numbers then their arithmetic mean A.M is defined by

$$A.M. = \frac{x_1 + x_2 + x_3 + \ldots\ldots + x_n}{n} = \frac{\sum x}{n}$$

If the number x1 occurs f1times, *x2* occurs *f2* times and so on, then

$$A.M. = \frac{f_1 x_1 + f_2 x_2 + \ldots\ldots + f_n x_n}{f_1 + f_2 + \ldots\ldots + f_n} = \frac{\sum f x}{\sum f}$$

Formula to Find Arithmetic Mean

cuemath THE MATH EXPERT

Formula:

Ungrouped Data: $\bar{x} = \dfrac{x_1 + x_2 + \cdots x_n}{n}$

Grouped Data: $\bar{x} = \dfrac{\Sigma fx}{n}$

Where: f = frequency in each class
x = midpoint of each class
n = total number of scores

This is known as *direct* method

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

Also , it can be calculated from the ***frequency table*** with the formula

$$\boxed{\bar{x} = \frac{\sum x_i n_i}{n} = \sum x_i f_i}$$

Example. Find the Arithmetic mean of 20, 22, 25, 28, 30

Solution.

$$A.M. = \frac{20+22+25+28+30}{5} = \frac{125}{5} = 25$$

. **Example**: Compute the arithmetic mean of the first 6 odd natural numbers

Solution: The first 6 odd natural numbers: 1, 3, 5, 7, 9, 11

.$\bar{x} = (1+3+5+7+9+11) / 6 = 36/6 = 6$

.Thus, the arithmetic mean is 6

Example. Find the Arithmetic Mean of the following

Numbers	8	10	15	20
Frequency	5	8	8	4

Solution. $\Sigma fx = 8\times5+10\times8+15\times8+20\times4 = 40+80+120+80 = 320$

$$\Sigma f = 5+8+8+4 = 25$$

$$A.M. = \frac{\Sigma fx}{\Sigma f} = \frac{320}{25} = 12.8.$$

(b) Short cut method

Let **a** be the assumed mean, **d** the deviation of the variety **x** from **a**. Then

$$\frac{\Sigma fd}{\Sigma f} = \frac{\Sigma f(x-a)}{\Sigma f} = \frac{\Sigma fx}{\Sigma f} - \frac{\Sigma fa}{\Sigma f} = \text{A.M.} - \frac{a\Sigma f}{\Sigma f} = A.M. - a$$

$$A.M. = a + \frac{\Sigma fd}{\Sigma f}$$

Example. Find the Arithmetic Mean for the following distribution

Class	0–10	10–20	20–30	30–40	40–50
Frequency	7	8	20	10	5

Solution. Let assumed mean (a) = 25

Class	Mid-value x	Frequency f	$x - 25 = d$	$f.d$
0–10	5	7	−20	−140
10–20	15	8	−10	−80
20–30	25	20	0	0
30–40	35	10	+10	+100
40–50	45	5	+20	+100
Total		50		−20

$$A.M. = a + \frac{\Sigma fd}{\Sigma f} = 25 + \frac{-20}{50} = 24.6 \qquad \textbf{Ans.}$$

(c) ***Step deviation method***

Let a be the assumed mean, i the width of the class interval and

$$D = \frac{x-a}{i}, \quad A.M. = a + \frac{\sum fD}{\sum f}\, i$$

Example . Find the Arithmetic Mean of the data given in example 3 by step deviation method

Solution. Let **a = 25**

Class	Mid-value x	frequency f	$D = \frac{x-a}{i}$	$f.D$
0–10	5	7	−2	−14
10–20	15	8	−1	−8
20–30	25	20	0	0
30–40	35	10	+1	+10
40–50	45	5	+2	+10
Total		50		−2

$$A.M. = a + \frac{\sum fD}{\sum f}\,.\,i = 25 + \frac{-2}{50} \times 10 = 24.6 \qquad \text{Ans.}$$

(ii)- MEDIAN

Median is defined as the measure of the central item when they are arranged in ascending or descending order of magnitude When the total number of the items is odd and equal to say(n) then the value

OF MEDIAN$=\frac{1}{2}(n+1)th$

When the total number of the frequencies is *even*, say n, then there are two middle item and so the mean of the values of

Statistics

$\frac{1}{2}n$ and ($\frac{1}{2}$ n+1)th

$$\text{Med}(X) = \begin{cases} X[\frac{n}{2}] & \text{if n is even} \\ \frac{(X[\frac{n-1}{2}]+X[\frac{n+1}{2}])}{2} & \text{if n is odd} \end{cases}$$

Example . Find the median of 6, 8, 9, 10, 11, 12, 13

$$\text{The middle item} = \frac{1}{2}(7+1)^{th} = 4th$$

$$\text{Median} = \text{Value of the 4thitem} = 10$$

For grouped data, $$\text{Median} = l + \frac{\frac{1}{2}N - F}{f} \cdot i$$

Where(L) is the lower limit of the median class,(*f*)is the frequency of the class, i is the width of the class-interval(*f*) is the total of all the preceding frequencies of the median-class And(N) is total frequency of the data

Example . Find the value of Median from the following data

No. of days for which absent (less than)	5	10	15	20	25	30	35	40	45
No. of students	29	224	465	582	634	644	650	653	655

Solution. The given cumulative frequency distribution will first be converted into ordinary frequency as under

Class-Interval	*Cumulative frequency*	*Ordinary frequency*
0–5	29	29 = 29
5–10	$\boxed{224}$	224 – 29 = 195
$\boxed{10–15}$	465	465 – 224 = $\boxed{241}$
15–20	582	582 – 465 = 117
20–25	634	634 – 582 = 52
25–30	644	644 –634 = 10
30–35	650	650 – 644 = 6
35–40	653	653 – 650 = 3
40–45	655	655 – 653 = 2

$$\text{Median} = \text{size of } \frac{655}{2} \text{ or } 327.5\text{th item}$$

327.5th item lies in 10–15 which is the median class.

$$M = l + \frac{\frac{N}{2} - C}{f} i$$

Where(L) stands for lower limit of median class

(N) stands for the total frequency

(C) stands for the cumulative frequency just preceding the median class

(i) stands for class interval

f stands for frequency for the median class

$$\text{Median} = 10 + \frac{\frac{655}{2} - 224}{241} \times 5$$

$$= 10 + \frac{103.5 \times 5}{241} = 10 + 2.15 = 12.15$$

EXAMPLE//Calculate the median from the following data: Marks
0 – 10 ,10 – 30 ,30 – 60, 60 – 80 ,80 – 90
Number of students 5 ,15 ,30, 8 ,2.
Answer: The median for the given data is 40.

Let's understand how to find the median for a grouped data.
Explanation:

We need to calculate the cumulative frequencies to find the median.

Marks	**Number of students**	**Cumulative frequency**	
0 - 10	5	0 + 5	5
10 - 30	15	5 + 15	20

30 - 60	30	20 + 30	50
60 - 80	8	50 + 8	58
80 - 90	2	58 + 2	60

N = ∑fi = 60

N/2 = 60/2 = 30

Median Class is 30-60

l = 30, f = 30, c.f = 20, h = 30

Using Median formula:

$$\text{Median} = l + \left[\frac{\frac{n}{2} - c}{f}\right] \times h$$

$$=30+(30-20/30)\times 30=30+10/30\times 30=30+10 \text{ Median } =40$$

(iii)MODE

Mode is defined to be the size of the variable which occurs most frequently

Example_. Find the mode of the following items

0,6 ,5 ,0 ,6 ,2 ,6 ,6 ,7 ,3 ,2 ,7 ,6 ,1 ,0

Solution. 6 occurs 5 times and no other item occurs 5 or more than 5 times, hence the mode is 6

For grouped data, $$\text{Mode} = l + \frac{f - f_{-1}}{2f - f_{-1} - f_1} \cdot i$$

Where(L) is the lower limit of the modal class,(f) is the frequency of the modal class,(i) is the width of the class(f_{-1})is the frequency before the modal class and($f1$) is the frequency after the modal class

Emperical formula

$$\textbf{Mean} - \textbf{Mode} = 3\,[\textbf{Mean} - \textbf{Median}]$$

Example 8. *Find the mode from the following data:*

Age	0–6	6–12	12–18	18–24	24–30	30–36	36–42
Frequency	6	11	25	35	18	12	6

Solution.

Age	*Frequency*	*Cumulative frequency*
0–6	6	6
6–12	11	17
12–18	$25 = f_{-1}$	42
$\boxed{18–24}$	$35 = f$	77
24–30	$\boxed{18} = f_1$	95
30–36	12	107
36–42	6	113

$$\text{Mode} = l + \frac{f - f_{-1}}{2f - f_{-1} - f_1} \times i$$

$$= 18 + \frac{35 - 25}{70 - 25 - 18} \times 6$$

$$= 18 + \frac{60}{27} = 18 + 2.22 = 20.22$$

If x1, x2, x3, ….., xn be n values of varieties x, then the geometric mean G=$\sqrt[n]{x1 \times x2 \times x3 \ldots \ldots . \times xn}$

Example . *Find the geometric mean of 4, 8, 16.*

Solution. $G.M. = (4 \times 8 \times 16)^{1/3} = 8.$

HARMONIC MEAN

Harmonic mean of a series of values is defined as the reciprocal of the arithmetic mean of their reciprocals. Thus if H be the harmonic mean, then

$$\frac{1}{H} = \frac{1}{n}\left[\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots + \frac{1}{x_n}\right]$$

Example . *Calculate the harmonic mean of 4, 8, 16.*

Solution.

$$\frac{1}{H} = \frac{1}{3}\left[\frac{1}{4}+\frac{1}{8}+\frac{1}{16}\right] = \frac{7}{48}$$

$$H = \frac{48}{7} = 6.853$$

AVERAGE DEVIATION OR MEAN DEVIATION

It is the mean of the absolute values of the deviations of a given set of numbers from their arithmetic mean If x1, x2, x3, ..., xn be a set of numbers with frequencies *f*1, *f*2, ... *f*n respectively. Let x be the arithmetic mean of the numbers x1, x2, ..., xn, then

$$\text{Mean deviation} = \frac{\sum f_i \mid x_i - \bar{x} \mid}{\sum f_i}$$

Example . *Find the mean deviation of the following frequency distribution.*

Class	0-6	6-12	12-18	18-24	24-30
Frequency	8	10	12	9	5

Solution. $a = 15$

Class	Mid-value x	Frequency f	$d = x - a$	fd	$\lvert x - 14 \rvert$	$f\lvert x - 14 \rvert$
0–6	3	8	−12	−96	11	88
6–12	9	10	−6	−60	5	50
12–18	15	12	0	0	1	12
18–24	21	9	+6	54	7	63
24–30	27	5	+12	60	13	65
Total		44		−42		278

$$\text{Mean} = a + \frac{\Sigma fd}{\Sigma f} = 15 - \frac{42}{44} = 14 \text{ nearly}$$

$$\text{Average deviation} = \frac{\Sigma f\lvert x - \bar{x} \rvert}{\Sigma f} = \frac{278}{44} = 6.3$$ **Ans.**