Board –CBSE	Class –10th	Topic – Statistics

Data: It is a collection of facts such as numbers, words, measurements, alphabets, symbols, observations, or even just descriptions of things.

Example: Data include marks of students, present-absent report of students, name of students, runs made by the batsman, etc.

Data Organization: The data available in an unorganized form is called raw data. The extraction of the information from these raw data to give meaning to these data is known as data organization.

Frequency of data: The number of times a particular quantity repeats itself in the given data is known as its frequency.

Example: The table below represents the number of cars possessed by different families in a society.

Number of cars	No of families
0	4
1	8
2	2

Here, the frequency of families who have one car is 8.

Frequency Distribution Tables The table which represents the number of times a particular quantity is repeated is known as the frequency distribution table.

Example: Table below represents the number of cars possessed by different families in a society.

Number of cars	Frequency
0	4
1	8
2	2
3	3
4	2
5	1

Mean of Grouped Data: The mean value of a variable is defined as the sum of all the values of the variable divided by the number of values. Suppose, if $x_1, x_2 ... x_n$ are observations with respective frequencies $f_1, f_2 ... f_n$, then this means observation x_1 occurs f_1 times, x_2 occurs f_2 times, and so on.

Now, the sum of the values of all the observations $= f_1x_1 + f_2x_2 + ... + f_nx_a$

and the number of observations $= f_1 + f_2 + ... + f_n$

Hence, the mean of the data is given by,

$$x^{-} = \frac{f_1x_1+f_2x_2+L+f_nx_n}{f_1+f_2+L+f_n} \qquad x^{-} = \frac{\Sigma f_ix_i}{\Sigma f_i}$$

Data Grouping: When the amount of data is huge, then the frequency distribution table for individual observation will result in a large table. In such a case, we form a group of data and then prepare a table. This type of table is called a grouped frequency distribution.

For Example: Suppose, we need to prepare a table for Science marks obtained by 60 students in a class. Then preparing a table for individual marks will result in a big table, so we will group the data as shown in the table below:

Range of Marks	No of students
$0 - 10$	2
$10 - 20$	9
$20 - 30$	22
$30 - 40$	20
$40 - 50$	6
$50 - 60$	1
Total	60

Class Interval or Class: It represents the range in which the data are grouped. For the above example, groups 0-10, 10-20, 20-30, etc. represent class intervals.

Lower class limit: The lowest number occurring in a particular class interval is known as its lower class limit. For the above example, if we consider the class interval 10-20 then 10 is called the lower class limit of that interval.

Upper-class limit: The highest number occurring in a particular class interval is known as its upper-class limit. For the above example, if we consider the class interval $10 - 20$ then 20 is called the upper-class limit of that interval.

Width or size of class interval: The difference between the upper-class limit and the lower class limit is called the width or size of class interval. For the above example, if we consider the class interval 10-20, then the width or size of this class interval will be 10.

Classmark: The frequency of each class interval is centered around its midpoint Classmark = (Upper-class limit + lower class limit)/2. For the above example, if we consider the class interval 10-20, then the classmark will be 15.

Methods to find mean:

Direct Method:

Example: A survey was conducted by a group of students as a part of their environment awareness program in which they collected the following data regarding the number of plants in 20 houses in a locality. Find the mean number of plants per house.

No 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

$$We\ know\ that,\ Classmark\ \left(x_i\right) = \frac{Upper-class\ limit + lower\ class\ limit}{2}.$$

No of plants	No of houses $\left(f_i\right)$	x_i	$f_i x_i$

$0 - 2$	1	1	1
$2 - 4$	2	3	6
$4 - 6$	1	5	5
$6 - 8$	5	7	35
$8 - 10$	6	9	54
$10 - 12$	2	11	22
$12 - 14$	3	13	39
Total	20		162

$\sum f_i = 20$

$\sum f_i x_i = 162$

$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{162}{20} = 8.1$

Therefore, the mean number of plants per house is 8. 1

Assumed Mean Method:

Example: The following distribution shows the daily pocket allowance of children of a locality. The mean pocket allowance is Rs 18. Find the missing frequency f.

Daily pocket allowance (in Rs)	$11 - 13$	$13 - 15$	$15 - 17$	$17 - 19$	$19 - 21$	$21 - 23$	$23 - 25$
No of children	7	6	9	13	f	5	4

$We\ know\ that,\ Class\ mark\ \left(x_i\right) = \frac{(Upper\ class\ limit + lower\ class\ limit)}{2}$

Given, mean pocket allowance, $x^{-} = 18\ Rs.$

Daily Pocket allowance (in Rs)	No of Children f_i	Classmark x_i	$d_i = x_i$ - 18	$f_i d_i$
11-13	7	12	-6	-42
13-15	6	14	-4	-24
15-17	9	16	-2	-18
17-19	13	18	0	0
19-21	F	20	2	2f
21-23	5	22	4	20
23-25	4	24	6	24
Total	$\sum f_{i=44} + f$			2f-40

From the table, we get,

$\Sigma f_{i=44+f}$

$\Sigma f_i d_{i=2f-40}$

$x^{-} = a + \sum f_i d_i \Sigma f_i$

$18 = 18 + (2f - 40)/(44 + f)$

$0 = (2f - 40)/(44 + f)$

$2f - 40 = 0$

$f = 20$

Therefore, the missing frequency is 20.

Step-deviation method:

For Example: Consider the following distribution of daily wages of 50 workers of a factory. Find the mean daily wages of the workers of the factory.

Daily wages (in Rs)	$100 - 120$	$120 - 140$	$140 - 160$	$160 - 180$	$180 - 200$
No of workers	12	14	8	6	10

We know that, Class mark $\left(x_i\right) = \frac{(Upper-class\ limit + lower\ class\ limit)}{2}$; Here, Class size $(h) = 20$

.

Taking 150 as assured mean (a), d_i, u_i , and $f_i u_i$ can be calculated as follows:

Daily wages (in Rs)	No of workers f_1	x_4	$d_1 = x_4 - 150$	$u_1 = d_1/20$	f_u
$100 - 120$	12	110	-40	-2	-24
$120 - 140$	14	130	-20	-1	-14
$140 - 160$	8	150	0	0	0
$160 - 180$	6	170	20	1	6
$180 - 200$	10	190	40	2	20
Total					-12

From the table, we get,

$\Sigma f_{i=50}$

$\sum f_i u_{i=-12}$

$Mean\ x^{-} = a + \sum f_i u_i \Sigma f_i$

$= 150 + (-12/50)20$

$= 150 - 24/5$

$= 145.2$

Therefore, the mean daily wage of the workers of the factory is Rs145. 20

Mode of Grouped Data:

Modal class: The class interval having the highest frequency is called the modal class and Mode is obtained using the modal class.

$M_o = 1 + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)h$

Where

1 = lower limit of the modal class,

h = size of the class interval (assuming all class sizes to be equal),

f_1 = frequency of the modal class,

f_0 = frequency of the class preceding the modal class,

f_2 = frequency of the class succeeding the modal class.

Example: The following data gives the information on the observed lifetimes (in hours) of 225 electrical components. Determine the modal lifetimes of the components.

Lifetimes (in hours)	0 − 20	20 − 40	40 − 60	60 − 80	80 − 100	100 − 120
Frequency	10	35	52	61	38	29

For the given data, it can be observed that the maximum class frequency is 61 which belong to class interval 60 − 80.

Therefore, modal class = 60 − 80.

Lower class limit (I) of modal class = 60

Frequency (f_1) of modal class = 61

Frequency (f_0) of class preceding the modal class = 52

Frequency (f_2) of class succeeding the modal class = 38

Class size (h)

$M_0 = 1 + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right)h = 20 = 60 + \frac{(61-52)\times 20}{(2\times 61-52-38)}$

$= 60 + \frac{9\times 20}{(122-90)} = 60 + \frac{90}{16}$

$= 65.6259$

Therefore, the modal lifetime of electrical components is 65. 625 hours.

Median of Grouped Data: For the given data, we need to have a class interval, frequency distribution, and cumulative frequency distribution. Then, the median is calculated as

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

Where

l = lower limit of median class,

n = number of observations,

cf = cumulative frequency of class preceding the median class,

$f = frequency\ of\ median\ class,$

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

Example: The following frequency distribution gives the monthly consumption of electricity of 68 consumers of a locality. Find the median of the data.

Monthly consumption (in units)	No of consumers
$65 - 85$	4
$85 - 105$	5
$105 - 125$	13
$125 - 145$	20
$145 - 165$	14
$165 - 185$	8
$185 - 205$	4

To find the median of the given data, cumulative frequency is calculated as follows:

Monthly consumption (in units)	No of consumers	Cumulative frequency
$65 - 85$	4	4
$85 - 105$	5	$4 + 5 = 9$
$105 - 125$	13	$9 + 13 = 22$
$125 - 145$	20	$22 + 20 = 42$
$145 - 165$	14	$42 + 14 = 56$
$165 - 185$	8	$56 + 8 = 64$
$185 - 205$	4	64 +4 +68

From the table, we get $n = 68$.

The cumulative frequency (cf) is just greater than $\frac{n}{2}$ (i.e. $\frac{68}{2} = 34$) is 42, belonging to interval 125 −145.

Therefore, median class = $125 - 145$

Lower limit (I) of median class = 125

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

Class size $(h) = 20$

Frequency (f) of median class = 20

Cumulative frequency (cf) of class preceding median class = $125 + ((34 - 22)/20)(20)$

$= 125 + 12 = 137$

Therefore, the median of the given data is 137.

Graphical Representation of Cumulative Frequency Distribution:

Example: The following distribution gives the daily income of 50 workers of a factory. Convert the distribution above to a less than type cumulative frequency distribution, and draw its graph.

Daily income (in Rs)	No of workers
$100 - 120$	12
$120 - 140$	14
$140 - 160$	8
$160 - 180$	6
$180 - 200$	10

The less than type cumulative frequency distribution is given as follows:

Daily income (in Rs)	No of workers	Cumulative frequency
$100 - 120$	12	12
$120 - 140$	14	$12 + 14 = 26$

$140 - 160$	8	$26 + 8 = 34$
$160 - 180$	6	$34 + 6 = 40$
$180 - 200$	10	$40 + 10 = 50$

Now, we will draw the given curve by plotting points

(120, 12), (140, 26), (160, 34), (180, 40), (200, 50)