B. Sc. Part I • Semester I

MATHEMATICAL STATISTICS

PAPER I

G. V. KUMBHOJKAR

PHADKE PRAKASHAN

Paper - I

Classification & tabulation of data

* **Def^n of Statistics :-**
Statistics is the subject which deals with
1) Collection of data
2) Presentation of data
3) Analysis of data (सूक्ष्म अभ्यास करणे)
4) Interpretation of data (स्पष्टीकरण)

⊙ **Data** - The numerical information with reference to any problem is called as statastical data.

eg. Data collected on height of student, weight of students etc.

* **Types of Data :-**
In general there are two types of data,
(1) Qualitative data (गुणात्मक)
(2) Quantitative data (परिमाणात्मक)

(1) **Qualitative data** -
The data collected on qualitative character is called qualitative data.

IMP The qualitative character is non-measurable, is called attribute. (गुणधर्म)

eg. cleverness of student, honesty of student, colour of flower etc.

Scanned by CamScanner

(2) Quantitative data -

The data collected on quantitative character, changes called quantitative data.

The characters which are measurable changes its value from time to time is called variable.

e.g. Height of students,
weight of students etc.

There are two types of quantitative data.
1) Descrete data
2) Continuous data

1) Descrete Data :-

If the data contains only integer value within small range, then data is called descrete data.

e.g. The data collected on no. of students present in the class.

2) Continuous data :-

If the data contains integer as well as fration values the data is called continuous data.

e.g. The data collected on height of students, age of students etc.

* Methods of Collection of data:-
There are two methods suggested to collect the data.
① Primary data.
② Secondary data

① Primary data - The data collected by person personally for his own purpose is called primary data.

② Secondary data - The data collected by some other person, if it is used second time for different purpose is called Secondary data.
In comparison the primary data is the recent information and hence is more reliable than secondary data.
objective

* ~~Types of~~

* Classification of Data :-
Classification of Data -
Classification is the process of arranging the raw data into different groups according to their similarities.

* Types of Classification :-
According to nature of data their are two types of classification

① Qualitative classification
② Quantitative classification

① Qualitative classification :-
The classification made by using qualitative data (attribute) is called Qualitative classification
The Qualitative character which is non-measurable is called attribute.
e.g. Colour of flower, Honesty of student, claverness of student etc.

There are two types of Qualitative classification :-
① Simple Qualitative classification
② Manifold Qualitative classification

① Simple Qualitative classification :-
"If the classification is made by using single ~~attitude~~ attribute classification is called simple Qualitative classification."

② Manifold Qualitative classification :-
If the classification is made by using two or more attributes, classification is called Manifold Qualitative classification.

Q. Consider the following classification of different persons according to different Qualitative

Persons → male, female

male → married, unmarried; female → married, unmarried

manifold classification { simple

② Quantitative classification :-

If the classification is made by using Quantitative data (variable), the classification is called Quantitative classification.

This classification is always represented by using frequency table. Hence it is mostly knows as frequency destribution.

According to nature of Quantitative data there are two types -

① Discrete frequency destribution
② continuous frequency destribution

① Discrete frequency Distribution -

If the frequency distribution is made by using discrete data the distribution is called Discrete frequency distribution.

The Method of forming discrete freq^n distribution is an be

Que

e.g. ① The data below gives the no. of childrens below age 13 from 10 different families.

2, 4, 6, 0, 3, 4, 5, 3, 8, 4
4, 5, 3, 5, 4, 6, 7, 4, 3, 7
4, 5, 3, 6, 4, 3, 5, 3, 2, 5
3, 5, 1, 3, 1, 5, 2, 3, 1, 5
6, 2, 7, 3, 5, 1, 7, 0, 3, 5
4, 3, 6, 5, 4, 3, 1, 3, 5, 0
4, 3, 5, 3, 7, 6, 4, 5, 3, 4

Relative fr

No. of childrens	Tally marks	freq^n marks	l.c.f	M.c.f	R.f
0	|||	3	3	Total: 70	3/70
1	𝍸	5	8	67	5/70
2	||||	4	12	62	4/70
3	𝍸 𝍸 𝍸 |||	18	30	58	18/70
4	𝍸 𝍸 |||	13	43	40	13/70
5	𝍸 𝍸	15	58	27	15/70
6	𝍸 |	6	64	12	6/70
7	𝍸	5	69	6	5/70
8	|	1	70	1	1/70
		Total = 70		Total = 1	

② Continuous frequency Distribution OR Grouped frequency Distribution :-

If the data is continuous the classification of data is made by ~~using~~ Dividing the data in Different groups the classification is called Continuous OR Grouped frequency distribution.

The groups or classes are generally taken of equal size.

① e.g. Consider the following data of marks obtain in sum subject out of 100 in the class of 60 students.

42	59	37	62	79	41	64	37	22	48
72	65	39	67	49	29	33	67	45	53
62	59	37	28	54	53	24	69	53	74
37	77	43	57	41	60	72	52	69	71
42	49	23	18	43	59	52	47	38	39
42	51	7	47	53	32	47	67	36	42

Marks	Tally marks	frequency	l.c.f.	m.c.f.
5-14	|	1	1	60
15-24	||||	4	5	59
25-34	~~卌~~ ~~卌~~ ||||	4	9	55
35-44	卌 卌 卌 |	16	25	51
45-54	卌 卌 卌	15	40	35
55-64	卌 |||	8	48	20
65-74	卌 卌	10	58	12
75-84	||	2	60	2

② The data below gives hight in cm of ~~line~~ the students in the class. Classifide the data using class width of 10 units, also calculate l.c.f. & m.c.f.

109.2	122.6	99.7	67.9	82.5	109.6	119.2	87.7	111.2	121.6
69.3	72.9	112	123.8	92.4	89.3	63.5	98.3	86.2	92.[illegible]
102.4	78.9	119.9	112.4	132.2	79.3	88.5	105.7	110	97.2
66	88.7	110	124.7	117.2	87.3	83.9	70	128.2	69.[illegible]
112.4	92.8	97.2	82.4	90	110.2	87.6	92.9	128.3	99.[illegible]

class	Tally	freq.	l.c.f.	m.c.f.
60-70	卌	5	5	50
70-80	||||	4	9	45
80-90	卌 卌	10	19	41
90-100	卌 卌	10	29	31
100-110	||||	4	33	21
110-120	卌 卌	10	43	17
120-130	卌 |	6	49	7
130-140	|	1	50	1
		Total = 50		

① Classes :- The non-overlaping groups which are made to classify the given data are called as classes.

e.g. 60-70, 70-80, 80-90 etc.

② Classes limit :- The two numbers indicating boundary of the class are called as class limits of the class.

The number to the left hand side or which is minimum bet^n two limits is called as lower limits of the class. Similarly the number the right hand side all which is maximum bet^n two limits is called upper limits of the class.

③ Mid value OR class Mark :- The average (mine) of two class limits is called as class marks OR Mid value.

ie. $M.V = \frac{\text{Upper limits} + \text{lower limit}}{2}$

e.g. ① 60-70 $= \frac{60+70}{2} = \frac{130}{2} = 65$

④ Class Interval OR class width :- The difference bet^n two successive upper limits OR lower limits mid value is called class width of that class.

Scanned by CamScanner

e.g. $10 \left\{ \begin{array}{l} 60-70 \\ 70-80 \\ 80-90 \end{array} \right. \Big\} 10$ M.V. $\left. \begin{array}{l} 65 \\ 75 \\ 85 \end{array} \right\}$ 10 diff^n

class width - 10

⑤ Class frequency (f) :- The no. of observations belonging particular class gives the class freq^n of that class.

⑥ Frequency Distribution :- The table showing distribution of all the observations in the data over different group of observations represents the freq^n dist^n or classification of the data.

IMP ⑦ Inclusive freq^n Dist^n :- While classified the data if the upper limit of the class is included in the same class, the classification is called Inclusive freq^n dist^n. In this classification generally there is gap bet^n upper limit of the class & lower limit of next class. ie there is clear distingction upper limit & lower limit of next class.

e.g. 10-19
20-29
30-39

IMP ⑧ Exclusive freq^n dist^n :- While classified the data if the upper limit of the class is Excluded from the class & in included in the next

class the classification is called Exclusive classification.
Here the class starts from the end of previous class. ie upper limit of the class & lower ~~eta~~ limit of the next class

e.g. 10 - 20
20 - 30
30 - 40

⑨ Class boundaries :- When classification is Inclusive to convert it two exclusive class boundaries are use these class boundaries are obtained by extending the class limits equally from the group bet^{n} upper limit and lower limit of next class & then classification is made.

e.g.

Inclusive class	Exclusive class
10 - 19	9.5 - 19.5
20 - 29	19.5 - 29.5
30 - 39	29.5 - 39.5

⑩ Open end classes :- Some times we get some abnormal observations in the data because of these abnormal observations we get some blank classes in the in the distribution To avoid blank classes we may open 1st & last class class in the distribution & this classes are called as open end class.

Date
Page

e.g. below 10 - open end class
10-20
20-30
above 30 - open end class

(11) Cummulative Frequencies :-
The frequencies obtain the by add^n of class frequencies are called as cummulative frequencies.
There are two types:
(a) Less than cummulative frequencies (l.c.f)
(b) More than cummulative frequencies (m.c.f.)

(a) Less than Cumulative Frequencies (l.c.f.) :-
If we go on adding class frequencies from top of the table to the bottom of the table we will get less than cummulative frequencies.
This frequencies are useful the find the no. of observations below upper limit of perticular class.

(b) More than Cummulative Frequencies (m.c.f.) :-
If we go on adding class frequencies from bottom of the table to the top of the table we will get more than cummulative frequencies.
This freq^n are useful find the no. of observations above the lower limit of perticular class.
The table showing cummulative frequency is called as cummulative freq^n table.

(12) Relative frequency :- The proportion (प्रमाण) of class freq^n to the total freq^n is called as relative frequency.

ie. Relative freq^n = $\frac{\text{class freq}^n}{\text{Total freq}^n} = \frac{f}{N}$

V. IMP

Requirement of good classifications –

(1) The classes or the groups should be of uniform width & may be five or multiple of five.

(2) As far as possible the classes should starts from five or multiple of five.

(3) The class width should not be too large or too small but keep according requirement of no. of classes.

(6) The no. of classes should be bet^n 6 & 20.

classmate
Date 2-8-16
Page

Chapter :- Measures of Central Tendency

* **Introduction :-**

Most of the times we have to deal with huge data but the problem may arrives that we are not able to remember on the values gives data. The shortcut used at this time is we can find one central representative figure. for the complete data. Such a fig. is generally called as average. And is obtained by using methods of measures of central tendency or measures of location.

In most of the data we observe that most of the values occurs at centre of the data. The property of concentration of data at the central part is called as property of central tendency. And the average representative value is form this central part.

Here we apply diff. st methods to find out representative average according to our requirement.

IMP * **Requirements of good measure :**

1) It should be rigidly define.
2) It should be simple to understand & easy to calculate.
3) It should be depend upon all the observations in the data.
4) It should be capable of further algebric. treatment.
5) It should not affected by extreme values.
6) It should have sampling stability.

Page

Arthmatic Mean (A.M) [Mean] :-

The Arthmatic mean denoted by '$\bar{x}$' is defined as, the sum of all the observation in data divided by the no. of observation in the data.

Case - I

Ungrouped data (raw data) :-

In case of Ungrouped data of 'n' observation say $x_1, x_2, x_3 \dots x_n$ on variable 'X',

$$A.M. = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

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

Case - II

For descrete data :-

In case of descrete freq^n distribution where the values x_i, $i = 1, 2, \dots n$ occurs in freq^n f_i, $i = 1, 2, \dots n$. The arthmatin mean is given by

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

Case - III

For continuous data :-

IF data is in the form of continuous or grouped freq^n distribution. The A.M is calculated by using mid values of classes as observations & formula difference.

$$\boxed{\bar{x} = \frac{\sum f_i M_i}{\sum f_i}}$$

where M_i - mid value of i^{th} class.

* Properties of Arthmatic Mean :-

① The sum of observations in the data is given by product of no. of observations in the data and mean of the data

i.e. $\boxed{\sum x_i = n \times \bar{x}}$

Proof :

we have by def^n

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

$$\Rightarrow \sum x_i = n \times \bar{x}$$

Similarly,

$\boxed{\sum x_i = \sum f_i \times \bar{x}}$ $\boxed{\sum x_i = N \cdot \bar{x}}$, $N = \sum f_i$

Imp ② The sum of deviations of observations taken from A.M. is always zero.

i.e. $\sum (x_i - \bar{x}) = 0$

In case of discrete $freq^n$ $dist^n$

$$\sum f_i (x_i - \bar{x}) = 0$$

consider $\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$

$$= \sum_{i=1}^{n} x_i = n\bar{x}$$

from property (1)

$= n\bar{x} - n\bar{x}$

$= 0$

Simillarly,

for discrete frequency distribution

Consider, $\Sigma f_i(x_i - \bar{x}) = \Sigma(f_i x_i - f_i \bar{x})$

$= \Sigma f_i x_i - \cancel{\Sigma f_i} \bar{x} \Sigma f_i$

$= \Sigma f_i x_i - N\bar{x}$

from property I

$= N\bar{x} - N\bar{x}$

$= 0$

③ Effect of change of origin and scale.

Let us consider the data of N values say, (x_i, f_i) $i = 1, 2, 3, \ldots n$ of the variable 'x'.

Let $\bar{x}$ be the mean of these values then,

$$\bar{x} = \frac{\Sigma x_i f_i}{N}, \quad N = \Sigma f_i$$

If we at (substract) some const. adding from each observations it is called change of origin.

If we multiply (Divide) each observation by some concents value it is called change of scale.

Proof : let us transform the variable x to the new variable 'U' by the relⁿ

$$U = \frac{x - a}{h}, \quad a, h - \text{constants}$$

$$\Rightarrow u_i = \frac{x_i - a}{h} \quad i = 1, 2, \ldots n$$

$$\Rightarrow x_i = a + hu_i$$

$$\Rightarrow f_i x_i = f_i(a + hu_i)$$

$\Rightarrow f_i x_i = a f_i + h f_i u_i$

Take sum & divide it by N

$\frac{1}{N} \cdot \Sigma f_i x_i = \frac{1}{N} \Sigma (a f_i + h f_i u_i)$

$= \frac{1}{N} a \Sigma f_i + \frac{1}{N} h \cdot \Sigma f_i u_i$

$= \frac{1}{N} aN + h \frac{1}{N} \Sigma f_i u_i$

$$\boxed{\bar{x} = a + h\bar{u}}$$

Therefore mean is affected by both change of origin as well as change of scale.

(4) The sum of squares of deviation of x_i is minimum if it is taken from the A.M.

i.e. $\Sigma (x_i - A)^2$ is minimum if $A = \bar{x}$

Proof: Consider the descrete freq^n distribution (x_i, f_i) $i = 1, 2, 3, \ldots n$

Now by def^n,

$\bar{x} = \frac{\Sigma f_i x_i}{N}$

Cosider,

$\Sigma f_i (x_i - A)^2 = Z$

from the conditions of the maxima & minima 'z' will be minimum if $\frac{dz}{dA} = 0$ & $\frac{d^2 z}{d^2 A} > 0$

Now, consider

$\frac{dz}{dA} = 2 \Sigma f_i (x_i - A)(-1) = 0$ —(i)

$\Rightarrow \Sigma f_i (x_i - A) = 0$ — (ii)

$\Rightarrow \Sigma (f_i x_i - f_i A) = 0$

Scanned by CamScanner

$$\Rightarrow \sum f_i x_i - A \sum f_i = 0$$
$$\Rightarrow AN = \sum f_i x_i$$
$$A = \frac{\sum f_i x_i}{N}$$
$$\boxed{\therefore A = \bar{x}}$$

Now, Differentiate eqⁿ ① w.r.to 'A' we get

$$\frac{d^2 z}{dA^2} = -2 \sum f_i (0-1)$$
$$\frac{d^2 z}{dA^2} = 2 \sum f_i \geq 0$$

ie z will be minimum if $A = \bar{x}$

ie. sum of squares on deviation is minimum when it is taken from A.M.

⑤ Combined Mean

Let $\bar{x}_i$ $i = 1, 2, \cdots k$ $\bar{x}_i$ are the means of k component series of sizes 'n_i' $i = 1, 2, \ldots k$.

Then mean $\bar{x}$ of combined observations is given by

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \ldots n_k \bar{x}_k}{n_1 + n_2 + \ldots n_k}$$

Let $x_{11}, x_{12}, \cdots x_{1n_1}$, be n_1 observations of sets, $x_{21}, x_{22}, \ldots x_{2n_2}$ be n_2 " of 2nd series

similarly

$x_{k1}, x_{k2}, \ldots x_{kn_k}$ are n_k observation of k series

Then by defⁿ of A.M. we have

$$\bar{x}_1 = \frac{x_{11} + x_{12} + \cdots x_{1n_1}}{n_1}$$

$$\Rightarrow x_{11} + x_{12} + \cdots + x_{1n_1} = n_1\bar{x}_1$$

simillarly

$$x_{21} + x_{22} + \cdots + x_{2n_2} = n_2\bar{x}_2$$

&

$$x_{k1} + x_{k2} + \cdots x_{kn_k} = n_k\bar{x}_k$$

Now consider combined mean of all the observations

$$\bar{x} = \frac{(x_{11} + x_{12} + \cdots x_{1n_1}) + (x_{21} + x_{22} + \cdots x_{2n_2}) + \cdots + (x_{k1} + x_{k2} + \cdots x_{kn_k})}{n_1 + n_2 + \cdots + n_k}$$

$$= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots n_k\bar{x}_k}{n_1 + n_2 + \cdots n_k}$$

Merits & Demerits of A.M.

Merits – 1) It is rigidly define.

2) It is easy to calculate & simple to understand

3) It is based upon all the values in the data.

4) It is capable of further algebric treatment.

5) It has sampling stability.

Demerits –

① It is affected by extreme values.

② It is not suitable for open end classes.

③ Sometimes it gives absurd result
④ Equality of two means does not mean that equality of distribution law.
⑤ Some times it is not actual value in the data.

- **Median :**

If the given values of 'x' are arrange in increasing or decreasing order of magnitude then the middle most value in this arrangement is called the median of x.

In other words the (value which divides the data into two equal parts is called median)

Case - I — for ungrouped data

When the no. of obs^n 'n' is odd then $\left(\frac{n+1}{2}\right)^{th}$ obs. in the arrangement is median.

If n is given then median is the A.M of $\frac{n}{2}^{th}$ & $\left(\frac{n+1}{2}\right)^{th}$ obs^n in the arrangment.

Case II - for descrete data

For descrete freq^n distribution the median is obtained by considering less than cummulative freq^n (L.C.F).

The steps are as follows :

i) find $N/2$, where $N = \Sigma f_i$

ii) find L.C.F. just $\geq N/2$

iii) Corresponding value of x is median

Scanned by CamScanner

Case III - For continuous

For continuous freq^n distribution obtained by considering l.c.f.

The steps are as follows :

(i) find $N/2$, where $N = \Sigma f_i$

(ii) find l.c.f. just $\geq N/2$

(iii) Corresponding class is median class.

$$\text{Median} = L + \frac{N/2 - cf}{f} \times h$$

where

L - lower limit of median class

c.f. - l.c.f. of premedian class

f. - freq^n of median class

h. - class width

Ungrouped

e.g. Find the median for following data sets :

a) 3, 8, 5, 6, 13, 16, 17, 19, 21

b) 121, 123, 151, 191, 147, 100, 103, 104

Sol^n :

a) data in assending order

3, 5, 6, 8, 13, 16, 17, 19, 21

Here $n = 9$ is odd

$\therefore$ Median = Value of $\left(\frac{n+1}{2}\right)^{th}$ obs^n

= value of 5^{th} obs^n

= 13

b) data in assending order

100, 103, 104, 121, 123, 147, 151, 191

Here $n = 8$ is even

$\therefore$ Median = A.M. of $\left\{ \frac{n}{2}^{th} \text{ \& } \frac{n}{2}+1^{th} \text{ obs}^n \right\}$

$= \frac{121 + 123}{2}$

$= 122$

descrete

Que. Find the median for the following data sets:

x	f	l.c.f
10	35	35
15	160	195
20	90	285
25	55	340
30	25	365 = N = Σfi
	365	

Sol^n - i) $\frac{N}{2} = \frac{365}{2} = 182.5$

ii) l.c.f just ≥ 182.5 is 195

iii) Corresponding value of x is 15

∴ Median = 15

Continuous

Que. Obtain the value of median for following data:

Class	freq^n	exclusive classes	l.c.f.
0-9	12	-0.5 - 9.5	12
10-19	15	9.5 - 19.5	27 ... c.f
20-29	30 = fi	$\boxed{19.5 - 29.5}$	57
30-39	23	29.5 - 39.5	80
40-49	15	39.5 - 49.5	95
50 - above	5	49.5 - above	100

Sol^n i) $N/2 = 100/2 = 50$

ii) l.c.f just ≥ 50 is 57

iii) Corresponding value of x is 19.5 - 29.5

Median class = 19.5 - 29.5

$\therefore$ Median ~~class~~ = $L + \frac{N/2 - c.f}{f} \times h$

$= 19.5 + \frac{50-27}{30} \times 10$

$= 19.5 + \frac{23}{30} \times 10$

$= 19.5 + 7.66$

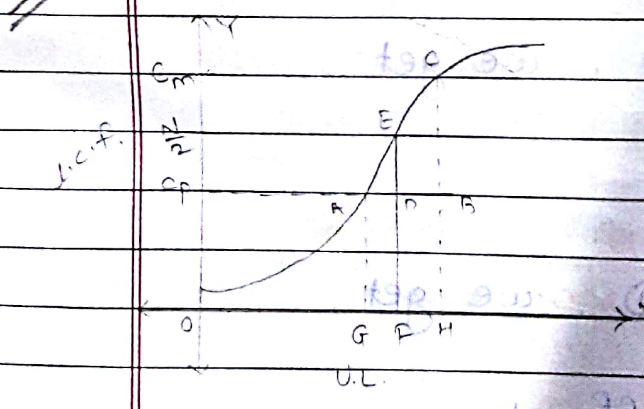
$\therefore$ Median = 27.16

* Graphical Method to obtain Value of Median :-

The median of freq^n distribution can be obtain graphically by using less than ogive curve. This is the graph of l.c.f & upper limits.

* Deviation for median in case of Grouped freq^n Dist^n :-

IMP



GH - Median class

Cp - l.c.f of premedian class

Cm - l.c.f of median class

Median is obs^n which devide the data into two equal parts. ie. it is the obs^n corresponding l.c.f $N/2$.

Hence from above fig. we have, median = OF

= OG + GF

Now, consider AC & straight line. Hence from the properties of similarity of triangle we have,

$$\Delta ABC \sim \Delta ADE$$

Hence we have,

$$\frac{AD}{AB} = \frac{DE}{BC}$$

$$\Rightarrow AD = \frac{DE}{BC} \times AB = GF \quad \text{——— ②}$$

Now we have,

$$DE = N/2 - C_p$$
$$= N/2 - cf \quad (\because \text{c.f of Premedian class})$$

$$AB = GH$$
$$= h = \text{class width}$$

$$BC = C_m - C_p$$
$$= f = \text{freq}^n \text{ of median class}$$

Put this values eq^n ②, we get

$$GF = \frac{N/2 - cf}{f} \times h$$

Put this value eq^n ①, we get,

$$\text{Median} = OG + GF$$
$$= OG + \frac{N/2 - cf}{f} \times h$$

$$\boxed{\text{Median} = OG + \frac{N/2 - c.f}{f} \times h.}$$

where, l = lower limit of median class
c.f = l.c.f of premedian class
f = freq^n of median class
h = class width.

* Partition Values :

The values which divides the data into diff. equal parts are called as partition values.

In practice we divide the data into either two or four or ten or hundred equal parts.

When we devide the data into two equal parts we have to find one observation at centre which is median.

A] Quartiles :-

If we divide the data into four equal parts, observations occuring at each quarter part is called as Quartile. Hence we get three quartiles Q_1, Q_2, & Q_3 but in general we calculate, Q_1 - 1st quartile & Q_3 - 3rd Quartile as 2nd quartile is median.

~~Calcul~~

* Calculation of Quartile -

Case (I) - for ungrouped data -

Arrange the data in assending order of magni-tude.

Q_1 = Value of $\left(\frac{n+1}{4}\right)^{th}$ observation in the arrangment

Q_3 = Value of $3\left(\frac{n+1}{4}\right)^{th}$ obs^n in the arrangment

Case (II) For descrete data :-

In case of descrete freq^n distribution quartiles are obtained by considering l.c.f.

The procedure is as follows :

For Q_1 -

i) find $N/4$

ii) find l.c.f. just $\geq N/4$

iii) Corresponding value of x is Q_1

For Q_3 -

i) find $3N/4$

ii) find l.c.f. just $\geq 3N/4$

iii) Corresponding value of x is Q_3

Case (III) For Continuous data :-

The procedure is as follows :

For Q_1 -

i) find $N/4$

ii) find l.c.f just $\geq N/4$

iii) Corresponding value is 1st quarter class

$$Q_1 = L + \frac{N/4 - cf}{f} \times h$$

For Q_3 -

i) find $3N/4$

ii) find l.c.f. just $\geq 3N/4$

iii) Corresponding class is 3rd quarter class

$$Q_3 = L + \frac{3N/4 - cf}{f} \times h$$

Remark -
$$Q_1 \leq Q_2 \leq Q_3$$

11-8-16

Deciles & Percentiles :-

1) **Deciles** - Deciles are the values which device the data into ten equal parts,

D_i = Value of $\left(\frac{i \times n}{10}\right)^{th}$ observation in the arranged data. $i = 1, 2, \dots 9$

Descrete freq^n distribution -

In case of descrete freq^n distribution deciles are obtained by considering l.c.f. the steps are as follows:

i) Find $\frac{i \times N}{10}$

ii) Find l.c.f. just $\geq \frac{i \times N}{10}$

* iii) Corresponding value of x is D_i, $i = 1, 2, \dots 9$

Continuous freq^n distribution -

In case of continuous freq^n distribution i^{th} decile D_i is obtained by considering l.c.f. The steps are as follows:

i) Find $\frac{i \times N}{10}$

ii) Find l.c.f. just $\geq \frac{i \times N}{10}$

* iii) Corresponding class is i^{th} Deciles class

$$\boxed{D_i = L + \frac{\frac{i \times N}{10} - c.f}{f} \times h} \quad i = 1, 2, \dots$$

Scanned by CamScanner

2) Percentiles - Percentiles are the values which divides the data into hundred equal parts. The ith percentile denoted by 'P_i', is defined as

$$P_i = \text{value of } \left(\frac{i \times n}{100}\right)^{th} \text{ obs}^n \text{ in the arranged data } i = 1, 2, \ldots 99.$$

Descrete freq^n distribution :-

In case of descrete freq^n distribution percentiles are obtained by condering l.c.f. The steps are as follows:

i) Find $\frac{i \times N}{100}$

ii) find l.c.f. just $\geq \frac{i \times N}{100}$

iii) Corresponding value of X is P_i

Continuous freq^n distribution :-

In case of continuous freq^n percentiles are obtained by considering l.c.f. The steps are as follows:

i) Find $\frac{i \times N}{100}$

ii) Find l.c.f. just $\geq \frac{i \times N}{100}$

iii) Corresponding class is ith percentile class

$$\boxed{P_i = L + \frac{\frac{i \times N}{100} - cf.}{f} \times h} \qquad i = 1, 2, \ldots 99$$

Que.
eg 1) Consider the following frequency distribution

x	f	l.c.f.
10	2	2
20	7	9
30	15	24
40	30	59
50	18	72
60	13	85
70	5	90
80	3	93

Find Quartiles, D_7, P_{80}.

For Q_1 :-

i) $\frac{N}{4} = \frac{93}{4} = 23.2$

ii) l.c.f just ≥ 23.2 is 24

iii) corresponding value of x is 30

$\boxed{\therefore \quad Q_1 = 30}$

for Q_3 :

i) $\frac{3N}{4} = 3 \times 23.2 = 69.6$

ii) l.c.f. just ≥ 69.6 is 72

iii) corresponding value of x is 50

$\therefore \quad Q_3 = 50$

for D_7 : 93

i) $\frac{7 \times 93}{10} = 65.1$

ii) l.c.f just ≥ 65.1 is 72

iii) corresponding value of x is 50

Scanned by CamScanner

for P_{80} :

i) $\frac{80 \times 93}{100} = 8 \times 9.3 = 74.4$

ii) l.c.f. just ≥ 74.4 is 85

iii) corresponding value of x is 60

$\therefore \; P_{80} = 60$

2 Consider following freq^n distribution :

class	freq^n	l.c.f.
5-10	3	3
10-15	5	8
15-20	8	16
20-25	15	31
25-30	18	49
30-35	13	55
35-40	18	73
40-45	7	80
45-50	4	84

Find Quartiles D_4 ; P_{74}

Sol^n for Q_1 :

i) $\frac{N}{4} = \frac{84}{4} = 21$

ii) l.c.f. just ≥ 21 is 31

iii) corresponding class is 20-25

$$Q_1 = L + \frac{N/4 - c.f}{f} \times h$$

classmate

Date

Page

$$Q_1 = L + \frac{N/4 - c.f}{f} \times h$$

$$= 20 + \frac{21-16}{15} \times 5$$

$$= 20 + \frac{5}{15} \times 5$$

$$= 20 + \frac{5}{3} = 20 + 1.67$$

$$\underline{\underline{Q_1 = 21.67}}$$

For Q_3

i) $\frac{3N}{4} = \frac{3 \times 84}{4} = 3 \times 21 = 63$

ii) c.f just ≥ 63 is 73

iii) corresponding class is 35-40

$$\therefore Q_3 = L + \frac{\frac{3N}{4} - c.f}{f} \times h$$

$$= 35 + \frac{3N/4 - c.f}{f} \times h$$

$$= 35 + \frac{63 - 62}{11} \times 5$$

$$= 35 + \frac{5}{11}$$

$$= 35 + 0.45$$

$$\underline{Q_3 = 35.45}$$

For D_4

i) $4N/10 = \frac{4 \times 84}{10} = \frac{336}{10} = 33.6$

ii) c.f. just ≥ 33.6 is 49

Scanned by CamScanner

iii) Corrosponding class is 25-30

$$\therefore D_4 = L + \frac{4N/10 - c.f}{f} \times h$$

$$= 25 + \frac{33.6 - 31}{18} \times 5$$

$$= 25 + \frac{2.6}{18} \times 5$$

$$= 25 + \frac{13}{18} = 25 + 0.72$$

$$\boxed{D_4 = 25.72}$$

For P_{74},

i) $\frac{74 \times N}{100} = \frac{74 \times 84}{100} = \frac{6216}{100} = 62.16$

ii) c.f just ≥ 62.16 is 73

iii) Corrosponding class is 35-40

$$\therefore P_{74} = L + \frac{\frac{74N}{100} - c.f}{f} \times h$$

$$= 35 + \frac{62.16 - 62}{11} \times 5$$

$$= 35 + \frac{0.16 \times 5}{11}$$

$$= 35 + \frac{0.80}{11}$$

$$= 35 + 0.072$$

$$\boxed{P_{74} = 35.072}$$