Introduction

Page No.
Date

Data:

An information of facts and figures collected for a definite purpose in any manner is termed as data. Simply collection of raw facts and figures is called data.

Statistics:

Statistics is the science of collecting, organizing presenting, analysing and interpreating the numerical data.

Types of statistics:

a) Mathematical statistics.
b) Applied statistics.

a) Mathematical Statistics:

It is an application of mathematics to statistics. It involves collection of facts using mathematical techniques. It can be sub-divided into two parts:

1) Descriptive statistics.
2) Inferential statistics.

1) Descriptive statistics:

Statistics which deals with collection of data, presentation of data using table, diagram, graph etc and summarize data using measure of central tendency, dispersion, skewness, kurtosis etc is called descriptive statistics.

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

2) Inferential Statistics:

Statistics which deals with sample selection from population and statistical technique used to draw conclusion about population on the basis of statistical measures obtained from sample. It is used in estimation of parameters and statistical testing of hypothesis.

b) Applied Statistics:

Statistics which deals with application of statistical methods to specific problem is called applied statistics.

* Functions of Statistics:

- Condensation
- Comparision
- Forecasting
- Estimation
- Testing of hypothesis.

* Limitation of Statistics:

Despite it's power, essential usefulness and universal appliability statistics has it's own limitation and imperfection which are as under:

a) Statistics doesn't study individual.
b) It doesn't study qualitative phenomena.
c) Statistical data don't sever the story.
d) Statistical results are not always unquestionable.
e) Statistical laws are true on the average or the long run.
f) Statistical data are liable to be missed easily.

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

Terms used in Statistics:

a) **Data:**

On the basis of collection procedure data is divided into two types.

→ Primary data.
→ Secondary data.

Primary data:

The data which are originally collected by investigator or researcher for the first time for the purpose of statistical inquiry is called primary data.

Different methods used for collecting primary data

→ Direct personal interview.
→ Indirect oral interview
→ mailed questionnaire.
→ Information through telephone
→ Schedule sent through enumerator.
→ Information through corresponding.

Secondary data:

The data that has been already collected for a particular purpose and uses for next purpose is called secondary data.

Sources of secondary data:

→ Published source.
→ Unpublished source.

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

Population:

It is totality of units or items under study belonging to a particular class or group. Population can be divided into following ways:

Finite population.

Infinite population.

Sample:

A Sample is a random selection of members of a population. It is a small group drawn from the population that has the characteristics of the entire population.

Parameter:

A parameter is a data based on an entire population. Any numerical characteristics of a pop^n is a parameter. It may be averages and standard deviations.

Statistics:

The pop^n mean and pop^n standard deviation ~~and~~ are represented by the greek alphabet μ and σ, respectively. In same way, the numerical characteristics of a sample is referred as statistics.

Variable:

A variable is any characteristics, number or quantity that can be measured or counted. A variable may also be called a data item.

Scanned with CamScanner

Page No.
Date

Application of Statistics in the field of CSE:

Are given below:

Statistics is used for data mining, speech recognition vision and image analysis, data compression, artificial intelligence, networking and traffic modeling.

A statistical background is used for understanding algorithm and statistical properties that form the backbone of computer science.

Scanned with CamScanner

Chapter : 2

Descriptive statistics

Page No.
Date

Measure of Central tendency:

The measure of central single representative value of entire data is called measure of central tendency.

Objectives

To get a single value that represents the characteristics of the entire data.
To facilitate comparison

Requisites of a good average

It should be rigidly (clearly) defined.
It should be easy to understand and simple to calculate.
It should be based on all observation.
It should be suitable for further mathematical treatment
It should be be affected by extreme values.
It should be least affected by fluctations of sampling

Measures of central tendency

They are as follows:
Arithmetic mean
Geometric mean
Harmonic mean
Median (Partition values)
Mode.

Scanned with CamScanner

Arithmetic mean:

It is the sum of all observation divided by no. of observation. It can be divided into:

a) Simple arithmetic mean
b) Weighted arithmetic mean

Calculation of Simple mean:

By direct method:

Let, $x_1, x_2, \ldots x_n$ be the no. of observation then mean of all observation is denoted by $(\bar{x})$ and is given by,

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

$$\boxed{\therefore \bar{x} = \frac{\Sigma x}{n}}$$

By short-cut method:

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

where,

A = assumed value

$d = x - A$

n = no. of observation

Scanned with CamScanner

Page No.
Date

III) Step-deviation method: (Coading method)

$$\bar{x} = A + \frac{\Sigma d'}{n} \times h$$

where,

$$d' = \frac{x - A}{h}$$

h = Common factor

For individual series

eg:

1) Find mean by direct, short-cut and step deviation method of:

x: 60 80 90 30 100 150

Sol^n

By direct method,

$n = 6$

$\Sigma x = 60 + 80 + 90 + 30 + 100 + 150$

$= 510$

we have,

$$\bar{x} = \frac{\Sigma x}{n} = \frac{510}{6} = 85$$

$$\boxed{\therefore \bar{x} = 85}$$

By short-cut method,

x	$d = x - A$
60	-30
80	-10
90 A	0
30	-60
100	10
150	60
	$\Sigma d = -30$

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

We have

$$\bar{x} = A + \frac{\Sigma d}{n} \Rightarrow 90 + \left(\frac{-30}{6}\right) \Rightarrow 85$$

By step-deviation method,

x	$d' = \frac{x-A}{h} \Rightarrow \frac{x-90}{10}$
60	-3
80	-1
90	0
30	-6
100	1
150	6
	$\Sigma d' = -3$

We have

$$\bar{x} = A + \frac{\Sigma d'}{n} \times h$$

$$= 90 + \left(\frac{-3}{6}\right) \times 10$$

$$\boxed{\therefore \bar{x} = 85}$$

For discrete series:

Let observation x_1 occurs f_1 times, observation x_2 occurs f_2 times then x_n occurs f_n times then A.M is denoted by $\bar{x}$ and is given by,

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n}$$

$$\boxed{\therefore \bar{x} = \frac{\Sigma fx}{N}} \leftarrow \text{Direct method}$$

Scanned with CamScanner

Page No.
Date

By short-Cut method,

Let, $d = x - A$

$$\therefore \bar{x} = A + \frac{\Sigma d}{N}$$

By step-deviation method,

$$\bar{x} = A + \frac{\Sigma d'}{N} \times h$$

where,

$$d' = \frac{x - A}{h}$$

A = Assume mean

h = Common Factor

N = no. of observation

For eg:

Calculate mean by (i) Direct (ii) step deviation and (iii) short-cut method of:

x:	1780	1760	1680	1750	1830	1940	1100	1300	1060	19
f:	10	6	4	8	7	9	11	10	7	8

Soln

Calculation table

x	f	fx	$d = x - A$ ($A = 1750$)	$d' = \frac{x-A}{h}$ ($h = 10$)	fd	fd'
1780	10	17800	30	3	300	30
1760	6	10560	10	1	60	6
1680	4	6720	-70	-7	-280	-28
1750	8	14000	0	0	0	0
1830	7	16470	80	8	720	72
1940	9	13580	190	19	1330	133
1100	11	12100	-650	-65	-7150	-715
1300	10	18000	50	5	500	50
1060	7	7420	-690	-69	-4830	-483
1950	8	15600	200	2	1600	160
	$N = 80$	$\Sigma fx = 132250$			$\Sigma fd = -7750$	$\Sigma fd' = -775$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

By direct method

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{132250}{80} = 1653.125$$

By short cut method

$$\bar{x} = A + \frac{\Sigma fd}{N} \Rightarrow 1750 + \left(\frac{-7750}{80}\right) \Rightarrow 1653.125$$

By step-deviation method

$$\bar{x} = A + \frac{\Sigma fd'}{N} \times h \Rightarrow 1750 + \left(\frac{-775}{80}\right) \times 10 \Rightarrow 1653.125$$

Mean for Continuous Series:

Class :	0-10	10-20	20-30
f :	5	7	9

Exclusive type Continuous series

Class:	0-9	10-19	20-29	30-39
f :	7	5	6	9

Inclusive type Continuous Series

Changing Inclusive into Exclusive

$$\text{Correction Factor (K)} = \frac{\text{Lower limit of } 2^{nd} - \text{Upper limit of } 1^{st}}{2}$$

$$\therefore K = \frac{10-9}{2} \Rightarrow \frac{1}{2} \Rightarrow 0.5$$

Now

Exclusive class = (lower − K) − (upper + K)

∴ Exclusive class is

Class :	−0.5 − 9.5	9.5 − 19.5	19.5 − 29.5	29.5 − 39.5
f :	7	5	6	9

Scanned with CamScanner

Page No.
Date

(a) Direct- method:

$$\text{mean } (\bar{x}) = \frac{\Sigma fx}{N}$$

where,

x = mid-value

(b) By short- cut method:

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

$$\because d = x - A$$

(c) Step- deviation method: (coding method)

$$\bar{x} = A + \frac{\Sigma fd'}{N} \times h$$

where,

$$d' = \frac{x - A}{h}$$

A = assume value

h = common factor

For eg:

Find mean by (i) direct method (ii) short cut method (iii) step- deviation method.

Class:	0-10	10-20	20-30	30-40	40-50	50-60
F:	5	6	10	8	9	7

⇒ Sol^n

Calculation table,

Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

Page No.
Date

Class	f	x	fx	$d = x - A$	$d' = \frac{x-A}{h}$	fd	fd'
0-10	5	5	25	-20	-2	-100	-10
10-20	6	15	90	-10	-1	-60	-6
20-30	10 h	25 A	250	0	0	0	0
30-40	8	35	280	10	1	80	8
40-50	9	45	405	20	2	180	18
50-60	7	55	385	30	3	210	21
	N=45		$\Sigma fx = 1435$			$\Sigma fd = 310$	$\Sigma fd' = 31$

By direct method,

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{1435}{45} = 31.89$$

By short-cut method,

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

$$= 25 + \frac{310}{45}$$

$$= 31.89$$

By step-deviation method,

$$\bar{x} = A + \frac{\Sigma fd'}{N} \times h$$

$$= 25 + \frac{31}{45} \times 10$$

$$= 31.89$$

Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

Page No.
Date

Weighted mean:

If importance is given that is weighted mean and importance is not given that is unweighted mean. The term weight stands for the relative importance of the different items.

The formula for calculating weighted arithmetic mean, denoted by $\bar{x}_w$ is given by,

$$\bar{x}_w = \frac{\Sigma W x}{\Sigma W}$$

where,

x = value of variable ie, $x_1, x_2, \ldots x_n$

W = weight attached to variable values ie. $w_1, w_2, \ldots w_n$ respectively.

Eg:

x :	50	60	40	30
W :	10	7	12	4

weighted mean = ?

Sol^n:

$$\bar{x}_w = \frac{\Sigma w x}{\Sigma W}$$

$$= \frac{50 + 60 + 40 + 30}{10 + 7 + 12 + 4}$$

$$= \frac{230}{33}$$

$$= 6.97$$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

Median:

The middle most value which divides total observation into two equal parts is called median. It is also called positional value because it divides total observation into two equal parts. It is denoted by M_d.

Calculation of Median:

(i) For individual series:

Steps

(1) Arrange the data either in ascending or descending order.

(2) Find number of observation (n)

(3) Find position of median using formula,

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

Eg:

Find median value of:

x: 50 60 90 30 80 40 70 100 20

Soln

Arranging the data in ascending order,

20, 30, 40, 50, 60, 70, 80, 90, 100

$\therefore n = 9$

Now,

Position of median $(M_d) = \left(\frac{n+1}{2}\right)^{th}$ item

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

$= 5^{th}$ item

$\therefore$ median = value of 5th item

= 60

REDMI NOTE 5 PRO

Scanned with CamScanner

Page No.
Date

Eg:

Find median value of:

x:	50	20	80	60	90	100
f:	7	9	8	4	6	9

Soln

Calculation of median,

x	f	C.F
20	9	9
50	7	16
60	4	20
80	8	28
90	6	34
100	9	43
	N = 43	

$\therefore$ Position of median $= \left(\frac{N+1}{2}\right)^{th}$ item

$= \left(\frac{43+1}{2}\right)^{th}$ item

$= 22^{th}$ item

$\boxed{\therefore M_d = 80}$

$\therefore$ The CF just greater than 22 is 28 so, the median = 80

Scanned with CamScanner

Page No.
Date

Eg: (2)

Find median of;

x : 20 15 30 18 35 60

Sol^n:

Arranging data in ascending order

15, 18, 20, 30, 35, 60

$\therefore n = 6$

Now,

Position of median $(M_d) = \left(\frac{n+1}{2}\right)^{th}$ item

$= \left(\frac{6+1}{2}\right)^{th}$ item

$= 3.5^{th}$ item

$\therefore$ median $= \frac{\text{value of } 3^{rd} \text{ item} + \text{value of } 4^{th} \text{ item}}{2}$

$= \frac{20+30}{2}$

$= 25$

For discrete series:

Steps

(1) Arrange the data in ascending or descending order

(2) Prepare Cumulative Frequency (CF) table

(3) Find $N = (\Sigma F)$

(4) Find position of median using formula,

$M_d = \left(\frac{N+1}{2}\right)^{th}$ item

5) The value corresponding to $\left(\frac{N+1}{2}\right)^{th}$ item in CF

~~or~~ otherwise just greater than $\left(\frac{N+1}{2}\right)^{th}$ item in CF is median.

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

For Continuous Series:

Steps

(i) Arrange the class in ascending order.
(ii) Prepare the table.
(iii) Find N.
(iv) Find position of median class using formula,

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

(v) The class corresponding to or just greater than $(N/2)$ in CF is median class.
(vi) After finding median class use formula of median as,

$$M_d = L + \frac{N/2 - CF}{F} \times h$$

where,

L = lower limit of median class
h = ~~common~~ median class size (class interval)
F = median class frequency
~~CF = CF of just before (preceeding) median class CF.~~
CF = Cumulative frequency of the class preceeding the median class.

Note: Class should be in exclusive form

Eg:
Find median value of:

Class :	0-10	10-20	20-30	30-40	40-50
F :	5	7	9	8	4

Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

Soln,

Calculation table

Class	f	CF
0-10	5	5
10-20	7	12
20-30	9	21
30-40	8	29
40-50	4	33
	N=33	

Now,

Position of median class $= \left(\frac{N}{2}\right) = \frac{33}{2} = 16.5$

The CF just greater than 16.5 is 21. So,
median class = (20-30)

Here,

$L = 20$, $f = 9$, $CF = 12$, $h = 10$, $N/2 = 16.5$

Using formula,

$$M_d = L + \frac{N/2 - CF}{f} \times h$$

$$= 20 + \frac{16.5 - 12}{9} \times 10$$

$$= 20 + \frac{4.5 \times 10}{9}$$

$$= 25$$

$$\boxed{\therefore M_d = 25}$$

//

Scanned with CamScanner

REDMI NOTE 5 PRO
MI DUAL CAMERA

Page No.
Date

Eg(2):

Find median class of:

Class :	0-9	10-19	20-29	30-39	40-49
F :	7	8	10	6	5

Soln

The given class is inclusive type class so first converting into exclusive type class.

Calculation table

Class	F	CF
-0.5 - 9.5	7	7
9.5 - 19.5	8	15
19.5 - 29.5	10	25
29.5 - 39.5	6	31
39.5 - 49.5	5	36
	N=36	

∴ Position of median class = $(N/2) = 36/2 = 18$

∴ The CF just greater than 18 is 25. So the corresponding median class = (19.5 - 29.5)

Here,

$L = 19.5,\ h = 10,\ F = 10,\ CF = 15,\ N/2 = 18$

Using formula

$$M_d = L + \frac{N/2 - CF}{F} \times h$$

$$= 19.5 + \frac{18-15}{10} \times 10$$

$$= 19.5 + 3$$

$$\therefore M_d = 22.5$$

Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

Page No.
Date

Partition values:

Those variate values which divide the total observation into equal numbers of parts are called partition values. If total observation is divided into four equal parts then it is called quartiles (Q_i) ie. Q_1, Q_2, Q_3. If total observation is divided into 10 equal parts then it is called deciles (D_j) ie $D_1, D_2, \ldots D_9$. Similarly if total observation is divided into 100 equal parts then it is called percentiles (P_K) ie. $P_1, P_2, P_3 \ldots P_{99}$.

Calculation of partition values:

(1) For individual series:

Steps

(a) Arrange the data in either ascending or descending order.
(b) Find no. of item (n).
(c) Apply the formula for calculation of these value

(i) Position of Quartiles $(Q_i) = i\left(\frac{n+1}{4}\right)^{th}$ item

(ii) Position of deciles $(D_j) = j\left(\frac{n+1}{10}\right)^{th}$ item

(iii) Position of Percentiles $(P_K) = K\left(\frac{n+1}{100}\right)^{th}$ item

Eg:

Find 3rd quartile, 7th decile and 48th percentile

x: 10, 25, 80, 60, 39, 75, 60, 78, 99, 120, 150

Soln

Arranging the data in ascending order,

x: 10, 25, 39, 60, 60, 75, 78, 80, 99, 120, 150

Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

Page No.
Date

$\therefore n = 11$

Now,

For 3rd quartile,

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

$= 3\left(\frac{11+1}{4}\right)^{th}$ item

$= 9^{th}$ item

$\therefore Q_3 =$ value of 9th item $= 99$

For 7th deciles,

Position of 7th deciles $= 7\left(\frac{11+1}{10}\right)^{th}$ item

$= 8.4^{th}$ item

$\therefore D_7 =$ value of 8th item $+ 0.4(9^{th} - 8^{th})$

$= 80 + 0.4(99 - 80)$

$\therefore D_7 = 87.6$

For 48th Percentile:

Position of 48th Percentile $= 48\left(\frac{11+1}{100}\right)^{th}$ item

$= 5.76^{th}$ item

$\therefore P_{48} =$ value of 5th item $+ 0.76(6^{th} - 5^{th})$

$= 60 + 0.76(75 - 60)$

$= 71.4$

Scanned with CamScanner

Page No.
Date

(2) For discrete series:

Step:

(i) Arrange the data in ascending order of magnitude.

(ii) Find out Cumulative frequency (CF)

(iii) Find $N = (\Sigma f)$

(iv) Apply the formula

(a) Size of $Q_i = i\left(\frac{N+1}{4}\right)^{th}$ item where $i = 1, 2, 3$

(b) size of $D_j = j\left(\frac{N+1}{10}\right)^{th}$ item where $j = 1, 2, 3, \ldots, 9$

(c) size of $P_K = K\left(\frac{N+1}{100}\right)^{th}$ item where $K = 1, 2, 3, \ldots, 99$

(v) Now look at the CF column and find that the CF which is either equal to or next higher to that size (size determined in step (iv)) and determine the value of variable corresponding to it. That gives the values of quartiles (or deciles or percentiles).

Eg

Find 3rd quartile, 7th deciles and 58th percentile of:

x:	50	30	80	90	150	70	100
f:	6	9	8	10	4	7	3

Sol^n Calculation table,

x	f	CF
30	9	9
50	6	15
~~80~~ 70	7	22
80	8	30
90	10	40
100	3	43
150	4	47

$N = 47$

Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

Page No.
Date

For 3rd quartile

Position of $Q_3 = 3\left(\frac{N+1}{4}\right)^{th}$ item

$= 3\left(\frac{47+1}{4}\right)^{th}$ item

$= 36^{th}$ item

∴ The CF just greater than 36 is 40. So the Q_3 value = 90

For 7th deciles

Position of $D_7 = 7\left(\frac{N+1}{10}\right)^{th}$ item

$= 7\left(\frac{47+1}{10}\right)^{th}$ item

$= 33.6^{th}$ item

∴ The CF just greater than 33.6 is 40. So value of $D_7 = 90$

For 58th Percentile

~~For~~ Position of $P_{58} = 58\left(\frac{47+1}{100}\right)^{th}$ item

$= 27.84^{th}$ item

∴ The CF just greater than 27.84 is 30. So value of $P_{58} = 80$

//

Scanned with CamScanner

Page No.
Date

(3) For Continuous Series:

Steps

(i) Arrange the data in ascending order.

(ii) Find out Cumulative frequency (C.F)

(iii) Find N (ΣF)

(iv) Apply the formula

(a) Size of $Q_i = \frac{N}{4} i$, where $i = 1, 2, 3$

(b) Size of $D_i = \frac{N}{10} i$, where $i = 1, 2, 3, \ldots 9$

(c) Size of $P_i = \frac{N}{100} i$, where $i = 1, 2, 3, \ldots 99$

(v) Now look at the C.F column and find that total which is either equal to or next higher to that size (size determined in step (iv)) and determine the class of variable corresponding to it. For exact value of the quartiles, deciles and percentiles used the following formula

(a) i^{th} Quartile, $Q_i = L + \frac{\frac{N}{4} i - CF}{F} \times h$

(b) i^{th} deciles, $D_i = \frac{L + \frac{N}{10} i - CF}{F} \times h$

(c) i^{th} Percentile $P_i = \frac{L + \frac{N}{100} i - CF}{F} \times h$

Eg:

Find 3rd quartile, 6th decile and 70th percentile of:

Class	0-10	10-20	20-30	30-40	40-50	50-60
F :	10	7	9	6	8	7

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

Soln

Calculation table

Class	F	C.F
0-10	10	10
10-20	7	17
20-30	9	26
~~30~~-30-40	6	32
40-50	8	40
50-60	7	47
	$N = 47$	

For 3rd quartile:

Position of $Q_3 = 3\left(\frac{N}{4}\right)^{th}$ item

$= 3\left(\frac{47}{4}\right)^{th}$ item

$= 35.25^{th}$ item

$\therefore$ The CF just greater than 35.25 is 40. So,
Q_3 class = (40-50)

Here,

$L = 40$, $F = 8$, $CF = 32$, $h = 10$, $3N/4 = 35.25$

$\therefore Q_3 = L + \frac{\frac{3N}{4} - CF}{F} \times h$

$= 40 + \frac{35.25 - 32}{8} \times 10$

$\therefore Q_3 = 44.0625$

For 6th decile:

Position of $D_6 = 6\left(\frac{N}{10}\right)^{th}$ item

Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

$$= \left(\frac{6 \times 47}{10}\right)^{th} \text{ item}$$

$$= 28.2^{th} \text{ item}$$

$\therefore$ The CF just greater than 28.2 is 32 so D_6 class = (30–40)

Here

$L = 30$, $F = 6$, $CF = 26$, $h = 10$, $\frac{6N}{10} = 28.2$

$$\therefore D_6 = L + \frac{\frac{6N}{100} - C.F}{F} \times h$$

$$= 30 + \frac{28.2 - 26}{6} \times 10$$

$$= 35.33$$

<u>For 70th Percentile</u>

Position of $P_{70} = \left(\frac{70 \times 47}{100}\right)^{th}$ item

$$= 32.9^{th} \text{ item}$$

The CF just greater than 32.9 is 40, so P_{70} class = (40–50)

Here

$L = 40$, $F = 8$, $CF = 32$, $h = 10$, $\frac{70N}{100} = 32.9$

$$\therefore P_{70} = L + \frac{\frac{70N}{100} - CF}{F} \times h$$

$$= 40 + \frac{32.9 - 32}{8} \times 10$$

$$= 41.125$$

//

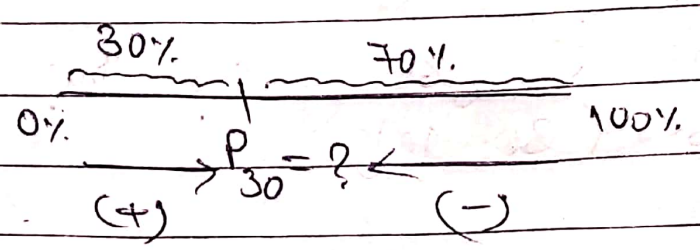
Scanned with CamScanner

REDMI NOTE 5 PRO
MI DUAL CAMERA

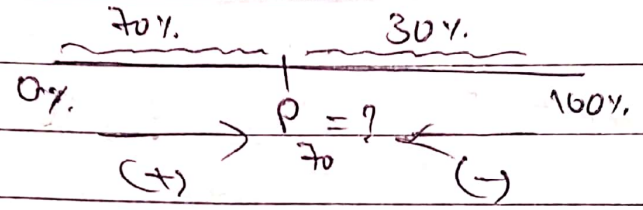
Date

SOME NOTES

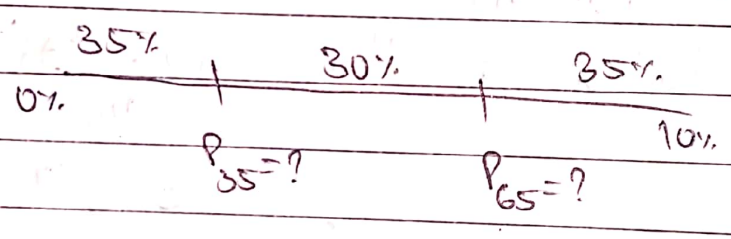
Find the maximum of minimum 30%



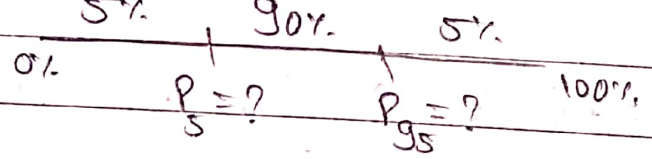
Minimum of maximum 30%



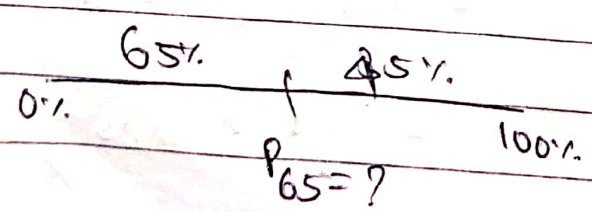
Limit of middle 30%



Middle 90%

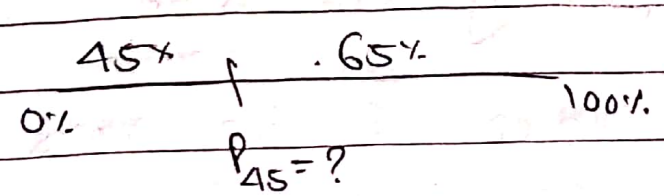


Largest of smallest 65%



Scanned with CamScanner
REDMI NOTE 5 PRO
MI DUAL CAMERA

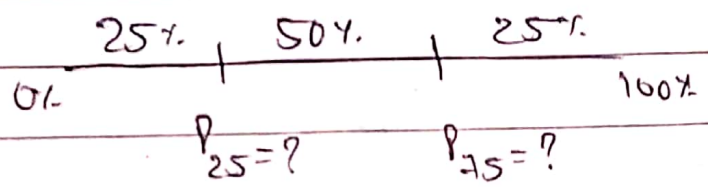
smallest of largest 65%



smallest of largest 80% $\Rightarrow P_{20} = ?$

largest of smallest 10% $\Rightarrow P_{10} = ?$

Range of middle 50% (R)



$$\boxed{R = P_{75} - P_{25}}$$

The following table give the frequency distribution of the marks of 800 candidates in a class test.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of students	10	40	80	140	170	130	100	70	40	20

If 30% of the students fails, find the highest marks obtained by fail students.

If the minimum marks required for passing are 35, what percentage of the students pass the test.

Find the limits of marks of central 30% students and range marks of central 30% students.

Find the percentage of the students who get the marks

between 75 and 95.

(v) Find the percentage of students who get marks less than 55.

(vi) Find the median of the marks.

⇒ Sol^n

Calculation table

Marks	No. of students (F)	C.F
0–10	10	10
10–20	40	50
20–30	80	130
30–40	140	270
40–50	170	440
50–60	130	570
60–70	100	670
70–80	70	740
80–90	40	780
90–100	20	800

$N = 800$

(i) The highest mark obtained by fail 30% of the students is given by P_{30}.

$\therefore$ size of $P_{30} = \frac{30N}{100} = \frac{30 \times 800}{100} = 240^{th}$ item

Thus, P_{30} lies in class (30–40)

$$\therefore P_{30} = L + \frac{\frac{30N}{100} - CF}{F} \times h$$

$$= 30 + \frac{240 - 130}{140} \times 10$$

$$= 37.86$$

Frequency $(f_N) = \left(\frac{\text{New class}}{\text{Old class}} \times F\right)$

(ii) No. of students who secured more than 35 marks

$$= \frac{40-35}{40-30} \times (140 + 170 + 130 + 100 + 70 + 40 + 20)$$

$$= 600$$

$$\text{Percentage} = \frac{600}{800} \times 100 \Rightarrow 75\%$$

(iii) The limit of marks of central 30% students is given by P_{35} to P_{65}

$$\text{Size of } P_{35} = \frac{35N}{100} = \frac{35 \times 800}{100} = 280^{th} \text{ item}$$

Thus P_{35} lies in class (40-50)

$$\therefore P_{35} = L + \frac{\frac{35N}{100} - CF}{F} \times h$$

$$= 40 + \frac{280 - 270}{170} \times 10$$

$$= 40.58$$

$$\text{Size of } P_{65} = \frac{65N}{100} = \frac{65 \times 800}{100} = 520^{th} \text{ item}$$

Thus P_{65} lies in class (50-60)

$$\therefore P_{65} = L + \frac{\frac{65N}{100} - CF}{F} \times h$$

$$= 50 + \frac{520 - 440}{130} \times 10$$

$$= 56.15$$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

$\therefore$ The limit of marks of central 30% student is P_{35} to P_{65} is 40.58 to 56.15.

$\therefore$ The range marks of central 30% student $= P_{65} - P_{35}$

$= 56.15 - 40.58$

$= 15.57$

(iv) Number of students who got the marks greater than 75 and less than 95.

$= \frac{80-75}{80-70} \times 70 + \frac{95-90}{100-90} \times 20 + 40$

$= 85$

$\therefore$ Percentage of no. of students $= \frac{85}{800} \times 100\%$

$= \cancel{63.12\%}$ 10.63%

(v) No. of students who get the marks less than 55

$= \frac{55-50}{60-50} \times 130 + 170 + 140 + 80 + 40 + 10$

$= 505$

$\therefore$ Percentage of no. of students $= \frac{505}{800} \times 100\%$

$= 63.13\%$

(vi) Size of $Md = \frac{N}{2} = \frac{800}{2} = 400^{th}$ item

$\therefore$ median lies in class (40-50)

$\therefore Md = L + \frac{N/2 - CF}{f} \times h = 40 + \frac{400 - 270}{170} \times 10$

$= 47.65$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Mode (M_o):

The mode value is that value in a series of observation which occurs maximum number of times.

Calculation of mode

(i) For individual series:

mode = maximum repeated value

(ii) For discrete series:

mode = maximum frequency value

(iii) For continuous series:

Steps

(a) Arrange the class in ascending order

(b) make class size equal and exclusive.

(c) See maximum frequency class called model class.

(d) After finding model class we use formula of mode as,

$$M_o = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

where,

L = lower limit of model class

f_1 = frequency of model class

f_0 = preceding (before) frequency of model class

f_2 = succeeding (later) frequency of model class

h = class interval

Eg:

Find mode value of:

x: 50 60 90 80 100 80 40 50 80 150

Sol/ Here

80 ~~pea~~ repeats maximum number of times so,

mode = 80

Date

(2) Find mode value of:

x:	50	80	60	30	100	150
F:	7	9	10	8	4	6

Soln:

Here,

maximum frequency value is 10, so mode

(3) Find the mode value of:

Class:	0-10	10-20	20-30	30-40	40-50	50-60
F:	7	9	12	8	6	4

Soln:

Calculation table

Class	Frequency	
0-10	7	
10-20	9	f_0
20-30	12	f_1
30-40	8	f_2
40-50	6	
50-60	4	

Here,

maximum frequency value is 12 so mode class = (20-30)

So,

$L = 20, \; h = 10, \; f_1 = 12, \; f_0 = 9, \; f_2 = 8$

we have,

$$M_0 = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

$$= 20 + \frac{12 - 9}{2 \times 12 - 9 - 8} \times 10$$

$$\therefore M_0 = 24.29$$

//

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

39.5 - 49.5 20

Find the mode value of:

x :	0-9	10-29	30-39	40-49	50-69
f :	6	16	12	10	20

Soln

Since the classes are unequal so adjustment is to be made before calculating mode by making class intervals equal.

x :	0-9	10-19	20-29	30-39	40-49	50-59	60-69
F :	6	8	8	12	10	10	10

Since the given data is inexclusive type data, so, first converting it into exclusive type data.

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Calculation table

x	f	
-0.5 - 9.5	6	
9.5 - 19.5	8	
19.5 - 29.5	8	f_0
29.5 - 39.5	12	f_1
39.5 - 49.5	10	f_2
49.5 - 59.5	10	
59.5 - 69.5	10	

Here,

maximum frequency value is 12. So, mode class = (29.5 - 39.5)

$\therefore L = 29.5,\ f_1 = 12,\ f_0 = 8,\ f_2 = 10,\ h = 10$

we have,

$$M_0 = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

$$= 29.5 + \frac{12 - 8}{2 \times 12 - 8 - 10} \times 10$$

$$\therefore M_0 = 36.17$$

//

(4) find mode value of:

x:	0-10	10-20	20-30	30-40	40-50	50-60
f:	7	9	10	10	6	4

Soln//

This is bi-modal case, we find mode by using formula,

$$\text{Mode} = 3\,\text{median} - 2\,\text{mean}$$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

Calculation table

Class	F	Mid-value (x)	CF	fx
0-10	7	5	7	35
10-20	9	15	16	125
20-30	10	25	26	250
30-40	10	35	36	350
40-50	6	45	42	270
50-60	4	55	46	220
	N = 46			$\Sigma fx = 1260$

$\therefore$ mean $(\bar{x}) = \frac{\Sigma fx}{N} = \frac{1260}{46} = 27.39$

For median:

Position of median class $= \left(\frac{N}{2}\right)^{th}$ item

$= \left(\frac{46}{2}\right)^{th}$ item

$= 23^{th}$ item

Thus, media class lies in (20-30)

Here,

$l = 20$, $h = 10$, $f = 10$, $CF = 16$, $N/2 = 23$

$\therefore$ median $= l + \frac{N/2 - CF}{f} \times h$

$= 20 + \frac{23-16}{10} \times 10$

$= 27$

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

Now,

$$\text{Mode } (M_o) = 3\text{median} - 2\text{mean}$$
$$= 3 \times 27 - 2 \times 27.39$$
$$= 26.22$$

Calculation of mode by Graphically:

In a frequency distribution the value of mode can also be determined graphically. The steps calculations are:

Draw a histogram of the given data.

Draw two lines diagonally in the inside of the modal class box, starting from each upper corner of the box to the upper corner of the adjacent box.

Draw a 1st line from the intersection of the two diagonal lines to the x-axis which gives us the modal value.

Draw a histogram for the following distribution and find mode and check by the value by direct calculation.

x :	10-15	15-20	20-25	25-30	30-35	35-40	40-45
f :	60	140	110	150	120	100	50

Soln,

The histogram of this data is given below locating the mode graphically.

Mode

mode = 27 (By inspection)

Direct calculation,

By inscepion, mode lies in class (25-30)

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

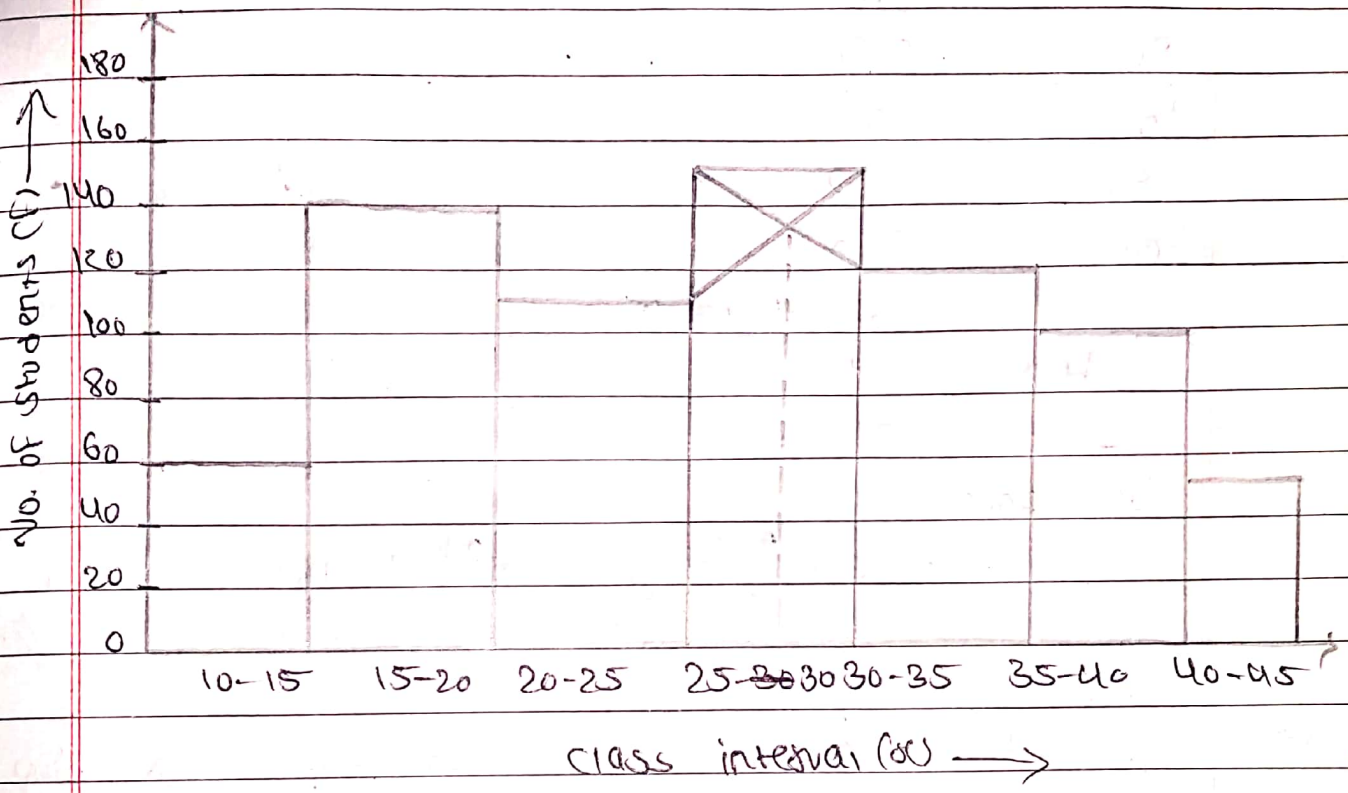
Page No.
Date

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

$$= 25 + \frac{150 - 110}{2 \times 150 - 110 - 120} \times 5$$

$$= 27.86$$

$\therefore$ Mode $= 27.86$



Geometric mean:

To find the average of ratios and percentage we use geometric mean. It is denoted by GM. and is given by,

$$GM = \text{Antilog}\left(\frac{\Sigma \log x}{n}\right) \quad \text{(for individual series)}$$

$$GM = \text{Antilog}\left(\frac{\Sigma f \log x}{n}\right) \quad \text{(for discrete \& continuous series)}$$

Scanned with CamScanner

$\sum_{i=1} w_i x_i$

Page No.

Date $\sum_i w_i$

Eg:

Find GM of:-

x: 50 60 40 80 100 150 200

Soln

Calculation table

x	$\log x$
50	1.70
60	1.78
40	1.60
80	1.90
100	2
150	2.18
200	2.30

$\Sigma \log x = 13.46$

Here

$n = 7$

we have

$$GM = \text{Anti}\log\left(\frac{\Sigma \log x}{n}\right) = \text{Anti}\log\left(\frac{13.46}{7}\right) = 83.73$$

(*) Find the average profit percentage per shop from the selling of Dell computer from following data.

Profit (%)	4-8	8-12	12-16	16-20	20-24	24-28
No. of shop	6	10	18	30	15	12

28-32	32-36	36-40
10	6	2

Soln

Page No.
Date

Calculation of G.M

Profit (₹)	No. of shop (f)	mid point (m)	$\log m$	$f \log m$
4-8	6	6	0.78	4.68
8-12	10	10	1	10
12-16	18	14	1.15	20.70
16-20	30	18	1.26	37.80
20-24	15	22	1.34	20.10
24-28	12	26	1.41	16.92
28-32	10	30	1.48	14.8
32-36	6	34	1.53	9.18
36-40	2	38	1.58	3.16
	$N = 109$			$\Sigma f \log m = 137.34$

we have

$$GM = \text{Antilog}\left[\frac{1}{n} \cancel{\Sigma f \log} \Sigma f \log m\right]$$

$$= \text{Antilog}\left[\frac{137.34}{109}\right]$$

$$= 18.20$$

Weighted Geometric mean

In situation in which all items are not of equally importance than we attach weight. The term weight refers the relative importance given to various items.

Let, observation x_1 takes w_1 weight, observation x_2 takes w_2 weight ..... observation x_n takes w_n weight then the weighted geometric mean denoted by GM_w is given by

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

Scanned with CamScanner
REDMI NOTE 5 PRO

Page No.
Date

Eg:

Following data represents interest rate of bank in fixed deposit according to amount deposited.

Amount (million Rs)	5	10	15	20
Interest rate (x)	6	6.5	7	8

Find average rate of interest

Sol^n

Amount (w)	Interest rate (x)	$\log x$	$w \log x$
5	6	0.778	3.89
10	6.5	0.812	8.12
15	7	0.845	4.225
20	8	0.903	18.06
$\Sigma w = 40$			$\Sigma w \log x = 34.295$

Average rate of interest

$$GM_w = \text{Antilog} \frac{\Sigma w \log x}{\Sigma w}$$

$$= \text{Antilog} \frac{34.295}{40}$$

$$= 7.2$$

Hence

Average rate of interest = 7.2%.

Combined geometric mean:

If sizes and geometric mean of two component series are known then we can find the combined geometric mean of these series

If n_1 and n_2 are sizes and GM_1 and GM_2 are their respective geometric mean of two series then combined geometric mean is

Page No.
Date

$$GM_{12} = \text{Antilog}\left(\frac{n_1 \log GM_1 + n_2 \log GM_2}{n_1 + n_2}\right)$$

Eg:

Two series of sizes 10 and 12 have G.M of 7.89 and 6.47 respectively, find combined G.M.

Soln,

Let $n_1 = 10$, $n_2 = 12$, $GM_1 = 7.89$ & $GM_2 = 6.47$

then,

$$GM_{12} = \text{Antilog}\left(\frac{n_1 \log GM_1 + n_2 \log GM_2}{n_1 + n_2}\right)$$

$$= \text{Antilog}\left(\frac{10 \times \log 7.89 + 12 \times \log 6.47}{10 + 12}\right)$$

$$= \text{Antilog}(0.849)$$

$$= 7.07$$

Hence

Combined GM = 7.07 //

Harmonic mean:

To find the average mean of speed velocity we use harmonic mean. It is denoted by H.M and is given by

$$H.M = \frac{n}{\Sigma 1/x} \rightarrow \text{For individual series}$$

$$= \frac{N}{\Sigma F. 1/x} \rightarrow \text{For discrete ~~and continuous~~ series}$$

$$= \frac{N}{\Sigma F. 1/m} \rightarrow \text{For Continuous series}$$

Eg:

Calculate ~~mean~~ H.M of:

x: 10 15 20 40 60 80 100

Soln,

Calculation table

x	$1/x$
10	0.1
15	0.067
20	0.05
40	0.025
60	0.0017
80	0.0125
100	0.010

$\Sigma 1/x = 0.282$

$\therefore n = 7$

Now,

$$H.M = \frac{n}{\Sigma 1/x} = \frac{7}{0.282} = 24.823$$

Eg:

Speed	60-62	63-65	66-68	69-71	72-74
number	15	118	142	127	18

Soln,

Calculation table

Speed	number (F)	mid-value (m)	$1/m$	$F.1/m$
60-62	15	61	$1/61$	0.24590
63-65	118	64	$1/64$	1.84375
66-68	142	67	$1/67$	2.12940
69-71	127	70	$1/70$	1.81429
72-74	18	73	$1/73$	0.24657

$\Sigma N = 420$ $\Sigma F.1/m = 6.26991$

we have,

$$H.M = \frac{N}{\Sigma F.1/m} = \frac{420}{6.26991}$$

$$= 66.987$$

Hence, average speed is 66.987

$A.M \geq G.M \geq H.M$

Page No.
Date

Weighted Harmonic mean:

Let, observation x_1 takes w_1 weight, observation x_2 takes w_2 weight ..... observation x_n takes w_n weight then the weighted harmonic mean denoted by HM_w is given by,

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

:

A person travels a distance of 10 km by a bus at speed of 25 km/hr, after then travels a distance of 5 km by taxi at speed of 38 km/hr and a distance of 8 km by motorbike at speed of 45 km/hr. Find average speed of entire trip

Soln,

vehicle	Speed in km/hr(x)	Distance in km(w)	w/x
Bus	25	10	0.4
Taxi	38	5	0.1315
motorbike	45	8	0.1777
		$\Sigma w = 23$	$\Sigma w/x = 0.709$

we have,

$$\text{Average speed } (HM_w) = \frac{\Sigma w}{\Sigma w/x}$$

$$= \frac{23}{0.709}$$

$$= 32.44$$

Hence,

Average speed of entire trip is 32.44 km/hr.

Scanned with CamScanner

Page No.
Date

Measure of dispersion:

Dispersion means scatterness of the data from central value. It provides an idea of the homogenity (compactness) or heterogenity (scatter) of the distribution.

The objective of measuring skewness are

i) To determine the reliability of an average.
ii) To serve as a basis for the control of variability
iii) To compare two or more series with regard to their variability.
iv) To facilitate the use of other statistical measures.

Properties (Requisties) of Good measure of Dispersion

i) It should be simple to understand.
ii) It should be easy to compute.
iii) It should be rigidly defined
iv) It should be based on all the observations.
v) It should be capable of further algebraic treatment

Absolute and Relative measure of Dispersion

An absolute measure of dispersion express the variation in the same units of measurement as the original data. It is used to compare the variability of two distributions having the same units & almost the same average size. If the two distributions are expressed in different units (eg: Kilogram of sugar verses tones of sugar) or in different average size (eg: manager's salary verses workers salary), we use relative measure of dispersion to compare the

Page No.
Date

Variability of two distributions.

A relative measure of dispersion is defined by,

$$\text{Relative dispersion} = \frac{\text{Absolute dispersion}}{\text{Suitable average}}$$

Here suitable average means the average used in computing absolute dispersion. Relative dispersion is also called coefficient of dispersion because it is a pure no. independent of original units of measurement.

Methods of studying Dispersion:

Range
Quartile deviation
Mean deviation
(d) Standard deviation
(e) Coefficient of variation

Range:

A simple but crude measure of dispersion is the range which is defined as the difference between maximum (or largest) and minimum (or smallest) values in a given data.

ie. Range = L − S

where

L = largest item
S = smallest item

The range is an absolute measure of dispersion & it has the same units of measurement as that of the given data. The relative measure corresponding to the range called the coefficient of range, is defined by

$$\text{Coefficient of Range} = \frac{L-S}{L+S}$$

Scanned with CamScanner

Page No.
Date

Eg:

(Q) Find range and coefficient of range of:

x:	50	90	80	150	60	40	100

Soln, Here

Maximum value (L) = 150

Minimum value (S) = 40

we have

Range = L − S = 150 − 40 ⇒ 110

& Coefficient of Range = $\frac{L-S}{L+S}$ = $\frac{150-40}{150+40}$ ⇒ 1

(Q) Find range and coefficient of range of:

x:	50	60	100	120	150
F:	10	9	10	8	9

Soln,

Largest value (L) = 150

Smallest value (S) = 50

we have

Range = L − S = 150 − 50 ⇒ 100

& Coefficient of Range = $\frac{L-S}{L+S}$ = $\frac{150-50}{150+50}$ ⇒ 0.5

(Q) Calculate range & coefficient of range of:

x:	10-20	20-30	30-40	40-50	50-60
F:	7	9	12	8	4

Soln, Here

Largest value (L) = upper limit of highest class ⇒ 60

Smallest value (S) = lower limit of smallest class ⇒ 10

we have

Range = L − S = 60 − 10 ⇒ 50

& Coefficient of Range = $\frac{60-10}{60+10}$ ⇒ 0.75

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

Quartile deviation (semi Inter Quartile Range):

The difference between third quartile (Q_3) and first quartile (Q_1) is called inter-quartile range & half of inter quartile range is called semi inter quartile range or Quartile deviation. It is denoted by Q.D.

$$\therefore \quad QD = \frac{Q_3 - Q_1}{2}$$

And,

$$\text{Coefficient of quartile deviation (Q.D)} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Eg:

(Q) Find QD and it's coefficient of:

x: 40 30 60 80 90 150 120.

Sol^n,

Arranging the data in ascending order,

x: 30 40 60 80 90 120 150

$n = 7$

for Q_1

$$\text{Position of } Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{item} = \frac{7+1}{4} = 2^{th} \text{ item}$$

$$\therefore Q_1 = 40$$

for Q_3

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

$$= 3\left(\frac{7+1}{4}\right)^{th} \text{item}$$

$$= 6^{th} \text{ item}$$

$$\therefore Q_3 = 120$$

Using formula,

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Date

$$QD = \frac{Q_3 - Q_1}{2} = \frac{120-40}{2} = 40$$

& Coefficient of $Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{120-40}{120+40} = 0.25$

find Q.D and it's coefficient of:

x :	30	40	60	80	90	120	150
f :	7	5	9	8	6	9	5

Sol:

Calculation table

x	f	c.f
30	7	7
40	5	12
60	9	21
80	8	29
90	6	35
120	9	44
150	5	49

$\therefore N = 49$

for Q_1

Position of $Q_1 = \left(\frac{N+1}{4}\right)^{th}$ item $= \left(\frac{49+1}{4}\right)^{th}$ item

$= 12.5^{th}$ item

$\therefore Q_1 = 60$ (ie. size corresponding to c.f 21)

for Q_3

Position of $Q_3 = 3\left(\frac{N+1}{4}\right)^{th} = (3 \times 12.5)^{th}$ item

$= 37.5^{th}$ item

$\therefore Q_3 = 120$ (ie size corresponding to c.f 44)

Scanned with CamScanner

Position of Q_3 Class $= 3(N/4)^{th}$ item

$= (3 \times 10)^{th}$ item

$= \cancel{30^{th}}$ item

$= 30^{th}$ item.

$\therefore$ Q_3 lies in the class $(30-40)$

$\therefore$ $Q_3 = 30 + \frac{30-26}{8} \times 10$

$= 35$

we have

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{35 - 13.33}{2} = 10.835$$

& Coeff. of $Q.D = \frac{35 - 13.33}{35 + 13.33} = 0.448$ //

<u>Percentile Range:</u>

Like semi-inter quartile range the percentile range and semi-percentile range are also used as a measure of dispersion. They are defined as follows:

Percentile range $= P_{90} - P_{10}$

Semi-percentile range $= \frac{P_{90} - P_{10}}{2}$

where

$P_{10} = 10^{th}$ Percentile & $P_{90} = 90^{th}$ percentile.

<u>Mean Deviation (MD) or Average Deviation</u>

It is also known as absolute deviation & can be calculated either from mean or median or mode.

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

we have

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{120-60}{2} = 30$$

& Coefficient of $Q.D = \frac{120-60}{120+60} = 0.167$

Class:	0-10	10-20	20-30	30-40	40-50
F:	7	9	10	8	6

Soln

Calculation table

Class	F	CF
0-10	7	7
10-20	9	16
20-30	10	26
30-40	8	34
40-50	6	40

$N = 40$

for Q_1

Position of Q_1 Class $= \left(\frac{N}{4}\right)^{th}$ item

$= (40/4)^{th}$ item

$= 10^{th}$ item

$\therefore$ Q_1 lies in the class (10-20)

$$\therefore Q_1 = L + \frac{N/4 - CF}{F} \times h$$

$$= 10 + \frac{10-7}{9} \times 10$$

$$= 13.33$$

for Q_3

Scanned with CamScanner

Page No.
Date

Calculation of mean deviation

(1) For individual series:

Mean deviation (MD) from mean ($\bar{x}$) = $\frac{\Sigma |x-\bar{x}|}{n}$

M.D from median (m_d) = $\frac{\Sigma |x-m_d|}{n}$

M.D from mode (m_o) = $\frac{\Sigma |x-m_o|}{n}$

(2) For discrete ~~or~~ or continuous series:

M.D from mean = $\frac{\Sigma f |x-\bar{x}|}{N}$

M.D from median = $\frac{\Sigma f |x-m_d|}{N}$

M.D from mode = $\frac{\Sigma f |x-m_o|}{N}$

where x is mid-value in case of continuous series

~~The vertical line~~

Co.eff of M.D from mean = $\frac{\text{M.D from mean}}{\text{mean}}$

Co.eff of M.D from median = $\frac{\text{M.D from median}}{\text{median}}$

Coeff. of M.D from mode = $\frac{\text{M.D from mode}}{\text{mode}}$

Eg:

Calculate mean deviation from mean, median & mode of:

x:	40	30	20	80	60	100	120	80

Soln

Arranging the data in ascending order,

x: 20 30 40 60 80 80 100 120

$\therefore n = 8$

Now,

$$\bar{x} = \frac{\Sigma x}{n} = \frac{20+30+40+60+80+80+100+120}{8}$$

$$= 66.25$$

For median

Position of median $= \left(\frac{n+1}{2}\right)^{th}$ item

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

$$= 4.5^{th} \text{ item}$$

$$\therefore m_d = \frac{4^{th} \text{ item} + 5^{th} \text{ item}}{2}$$

$$= \frac{60+80}{2}$$

$$= 70$$

For mode

m_o = maximum repeated value = 80

Calculation table

x	$\lvert x-\bar{x}\rvert$	$\lvert x-m_d\rvert$	$\lvert x-m_o\rvert$
20	46.25	50	60
30	36.25	40	50
40	26.25	30	40
60	6.25	10	0
80	13.75	10	20
80	13.75	10	20
100	33.75	30	20
120	53.75	50	40
	$\Sigma\lvert x-\bar{x}\rvert = 230$	$230 = \Sigma\lvert x-m_d\rvert$	$\Sigma\lvert x-m_o\rvert = 230$

Scanned with CamScanner

Page No.
Date

We have

M.D from mean $= \frac{\Sigma|x-\bar{x}|}{n} = \frac{230}{8} = 28.75$

M.D from median $= \frac{\Sigma|x-M_d|}{n} = \frac{230}{8} = 28.75$

M.D from Mode $= \frac{\Sigma|x-M_o|}{n} = \frac{230}{8} = 28.75$ //

(Q) Calculate mean deviation from mean, median and Mode ~~of~~ : and also it's coefficient of :

x :	20	80	60	30	90	100	120
F :	7	6	4	6	6	5	4

Soln //

Calculation table

x	F	C.F	Fx	$\lvert x-\bar{x}\rvert$	$F.\lvert x-\bar{x}\rvert$
20	7	7	140	48	336
~~80~~ 30	6	13	180	38	228
~~60~~ 60	4	17	240	8	32
80	6	25	600	12	72
90	6	31	540	22	132
100	5	36	500	32	160
120	4	40	400	52	208
	$N=40$		$\Sigma Fx = 2720$	~~$\Sigma\lvert x-\bar{x}\rvert$~~	$\Sigma F.\lvert x-\bar{x}\rvert = 1168$

Now,

$(\bar{x}) = \frac{\Sigma Fx}{N} = \frac{2720}{40} = 68$

$\therefore$ M.D from mean $= \frac{\Sigma F.|x-\bar{x}|}{N}$

$= \frac{1168}{40} = 29.2$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

Arranging the data in ascending order,

x: 20 30 40 60 80 80 100 120

$\therefore n = 8$

Now,

$$\bar{x} = \frac{\Sigma x}{n} = \frac{20+30+40+60+80+80+100+120}{8}$$

$$= 66.25$$

For median

Position of median $= \left(\frac{n+1}{2}\right)^{th}$ item

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

$$= 4.5^{th} \text{ item}$$

$$\therefore M_d = \frac{4^{th} \text{ item} + 5^{th} \text{ item}}{2}$$

$$= \frac{60+80}{2}$$

$$= 70$$

For mode

M_o = maximum repeated value = 80

Calculation table

x	$\lvert x-\bar{x}\rvert$	$\lvert x-M_d\rvert$	$\lvert x-M_o\rvert$
20	46.25	50	60
30	36.25	40	50
40	26.25	30	40
60	6.25	10	0
80	13.75	10	20
80	13.75	10	20
100	33.75	30	20
120	53.75	50	40
	$\Sigma\lvert x-\bar{x}\rvert = 230$	$230 = \Sigma\lvert x-M_d\rvert$	$\Sigma\lvert x-M_o\rvert = 230$

Page No.
Date

We have,

M.D from mean $= \frac{\Sigma|x-\bar{x}|}{n} = \frac{230}{8} = 28.75$

M.D from median $= \frac{\Sigma|x-M_d|}{n} = \frac{230}{8} = 28.75$

M.D from mode $= \frac{\Sigma|x-M_o|}{n} = \frac{230}{8} = 28.75$ //

(Q) Calculate mean deviation from mean, median and mode ~~of~~: and also it's coefficient of:

x:	20	80	60	30	90	100	120
F:	7	6	4	6	6	5	4

Soln //

Calculation table

x	F	C.F	Fx	$\lvert x-\bar{x}\rvert$	$F.\lvert x-\bar{x}\rvert$
20	7	7	140	48	336
~~80~~ 30	6	13	180	38	228
~~60~~ 60	4	17	240	8	32
80	6	25	600	12	72
90	6	31	540	22	132
100	5	36	500	32	160
120	4	40	400	52	208
	$N=40$		$\Sigma Fx = 2720$	~~$\Sigma\lvert x-\bar{x}\rvert$~~	$\Sigma F.\lvert x-\bar{x}\rvert = 1168$

Now,

$(\bar{x}) = \frac{\Sigma Fx}{N} = \frac{2720}{40} = 68$

$\therefore$ M.D from mean $= \frac{\Sigma F.\lvert x-\bar{x}\rvert}{N}$

$= \frac{1168}{40} = 29.2$

REDMI NOTE 5 PRO
Scanned with CamScanner

REDMI NOTE 5 PRO
MI DUAL CAMERA

For median

Position of median $= \left(\frac{N+1}{2}\right)^{th}$ item

$= \left(\frac{40+1}{2}\right)^{th}$ item

$= (20.5)^{th}$ item

The CF just greater than 20.5 is 25 so,
median = 80

$\therefore$ M.D from median $= \frac{\Sigma f.|x-M_d|}{N}$

$= \frac{1120}{40}$

$= 28$

$\lvert x-M_d \rvert$	$f.\lvert x-M_d \rvert$
60	420
50	300
20	80
0	0
10	60
20	100
40	~~80~~ 160

$\Sigma f.|x-M_d| = 1120$

For mode

~~$|x-M_o|$~~

~~Here mode =~~

maximum frequency value is 7 so, mode=2

$\therefore$ M.D from $M_o = \frac{\Sigma f.|x-M_o|}{N}$

$= \frac{1800}{40}$

$= 45$

$\lvert x-M_o \rvert$	$f.\lvert x-M_o \rvert$
0	0
10	60
40	160
60	360
70	420
80	400
100	400

$\Sigma f.|x-M_o| = 1800$

Scanned with CamScanner

Page No.
Date

Again,

Co-efficient of MD from mean $= \frac{\text{M.D from } \bar{x}}{\bar{x}}$

$= \frac{29.2}{68}$

$= 0.43$

Coeff. of MD from median $= \frac{\text{MD from median}}{\text{median}}$

$= \frac{28}{80}$

$= 0.35$

Coeff. of MD from mode $= \frac{\text{MD from mode}}{\text{Mode}}$

$= \frac{45}{20}$

$= 2.25$

Standard deviation (S.D):

The standard deviation is the square root of the mean of the squared deviation from the arithmetic mean and is denoted by S.D or 'σ'. It is also called "root-mean-squared-deviation" (RMSD).

For individual series

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

REDMI NOTE 5 PRO

Scanned with CamScanner

Coefficient of standard deviation $= \frac{\sigma}{\bar{x}}$

Page No.
Date

For discrete or continuous series

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}} = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

Computing SD for Individual observations (ungrouped data)

(a) Direct method or Actual mean method:

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

(b) Deviation method:

$$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

where

$d = x - A$ and $A =$ Assumed mean

Computing SD for discrete or continuous distribution (grouped data):

(a) Direct method:

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}} = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

(b) Deviation method:

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \quad ; \quad d = x - A$$

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date

(c) Step-deviation method:

$$\sigma = \sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2} \times h \; ; \quad d' = \frac{d}{h} = \frac{x - A}{h}$$

(For continuous series x denoted mid point of class)

Combined standard deviation:

Let, n_1 and n_2 be the no. of observation with mean and standard deviation $\bar{x}_1$ and $\bar{x}_2$ and σ_1 and σ_2 respectively. Then, the combined S.D is denoted by σ_{12} and is given by,

$$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

where

$$d_1 = \bar{x}_1 - \bar{x}_{12}, \quad d_2 = \bar{x}_2 - \bar{x}_{12}$$

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \text{Combined mean.}$$

Note:

M.D = 0.7979 (σ)
Q.D = 0.6745 (σ)

Eg:

The mean of two samples of sizes 50 and 100 respectively are 54.1 and 50.3 and the SD's are 8 and 7. Find the mean and SD of the sample size 150 obtained by combining the two samples.

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Here,

$n_1 = 50$, $n_2 = 100$, $\bar{x}_1 = 54.1$, $\bar{x}_2 = 50.3$, $\sigma_1 = 8$, $\sigma_2 = 4$

Hence,

the combined mean is,

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} = \frac{50 \times 54.1 + 100 \times 50.3}{54.1 + 50.3} = 51.57$$

Now,

$$d_1 = \bar{x}_1 - \bar{x}_{12} = 54.1 - 51.57 = 2.53$$

$$d_2 = \bar{x}_2 - \bar{x}_{12} = 50.3 - 51.57 = -1.27$$

∴ The combined S.D is,

$$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1+n_2}}$$

$$= \sqrt{\frac{50(64 + 6.4) + 100(79 + 1.6)}{50 + 100}}$$

$$= 7.56$$

//

Varience:

The variance of the values of a variable 'x' is defined as the square of the standard deviation.

Hence,

$$Var(x) = (\sigma)^2$$

$$\therefore \sigma = \sqrt{Var(x)}.$$

Sample standard deviation:

A ~~sample~~ S.D of a sample which is used to estimate the S.D of a population based on random sample is said to be sample S.D.

Scanned with CamScanner

Page No.
Date

Let, $x_1, x_2, x_3, \ldots x_n$ be n sample observation then sample S.D denoted by 's' is given by positive square root of sum of square of deviation of sample observation from their mean divided by sample size minus one.

ie,

$$s = \sqrt{\frac{1}{n-1} \Sigma(x-\bar{x})^2} = \sqrt{\frac{1}{n-1} (\Sigma x^2 - n(\bar{x})^2)}$$

Sample variance:

It is square of sample S.D and is given by,

$$s^2 = \frac{1}{n-1} (\Sigma x^2 - n\bar{x}^2) = \frac{1}{n-1} \Sigma(x-\bar{x})^2.$$

Eg:

Given the following set of data from a sample of size 7, 12, 7, 4, 9, 0, 7, 3. Find S.D and variance.

Soln,

x	x^2	$(x-\bar{x})$
0	0	
3	9	
4	16	
7	49	
7	49	
7	49	
9	81	
12	144	
$\Sigma x = 49$	$\Sigma x^2 = 397$	

$$\bar{x} = \frac{\Sigma x}{n} = \frac{49}{8} = 6.13$$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

Sample S.D $(s) = \sqrt{\frac{1}{n-1} \Sigma(x^2 - n\bar{x}^2)}$

$= \sqrt{\frac{1}{8-1} (397 - 8 \times 6.13^2)}$

$= 3.71$

Sample variance $(s^2) = (3.71)^2 = 13.76$ //

Coefficient of variation:

The Coefficient of variation (C.V) is the relative measure based on the standard deviation & defined as the ratio of standard deviation to the mean expressed in percent.

i.e.

$$(CV) = \frac{\sigma}{\bar{x}} \times 100\%$$

when samples SD (s) are given,

$$(CV) = \frac{s}{\bar{x}} \times 100\%$$

Calculate S.D, variance and Coefficient of variation (CV)

x: 10 60 50 40 20 30 70

Calculation table

x	x^2
10	100
20	400
30	900
40	1600
50	2500
60	3600
70	4900
$\Sigma x = 280$	$\Sigma x^2 = 14000$

Scanned with CamScanner

Page No.
Date

we have,

Standard deviation $(\sigma) = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

$= \sqrt{\frac{14000}{7} - \left(\frac{280}{7}\right)^2}$

$\boxed{\therefore \sigma = 20}$

variance $= (\sigma)^2 = (20)^2 = 400$

Coeff of variation $(C.V) = \frac{\sigma}{\bar{x}} \times 100\%$

$= \frac{20}{\frac{\Sigma x}{n}} \times 100\%$

$= \frac{20}{280/7} \times 100\%$

$= 50\%$

(Q) Calculate S.D variance and CV of:

x	10	60	50	40	20	30	70
f	5	6	8	10	8	4	9

Sol^n Calculation table

x	f	fx	fx^2
10	5	50	500
60	6	360	21600
50	8	400	7500
40	10	400	16000
20	8	160	3200
30	4	120	3600
70	9	630	44100
	$N = 45$	$\Sigma fx = 1870$	$\Sigma fx^2 = 96500$

Scanned with CamScanner

Page No.
Date

we have

$$(\bar{x}) = A + \frac{\Sigma fd'}{N} \times h$$

$$= 24.5 + \frac{17}{39} \times 10$$

$$= 28.86$$

$$SD(\sigma) = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{123}{39} - \left(\frac{17}{39}\right)^2} \times 10$$

$$= 1.72 \times 10$$

$$= 17.2$$

variance $= \sigma^2 = (17.2)^2 = 296.38$

$$C.V = \frac{\sigma}{\bar{x}} \times 100\%$$

$$= \frac{17.2}{28.86} \times 100\%$$

$$= 59.60\%$$

(Q) find combined S.D, combined C.V & variance of:

mean	S.D	No. of observation
60	7	150
80	9	120

Sol:

Let, $n_1 = 150$ $\quad n_2 = 120$

$\bar{x}_1 = 60$ $\quad \bar{x}_2 = 80$

$\sigma_1 = 7$ $\quad \sigma_2 = 9$

Combined S.D $(\sigma_{12}) = ?$

Combined C.V $= ?$

variance $= ?$

Scanned with CamScanner

Page No.
Date

We have,

Combined mean $(\bar{x}_{12}) = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$

$= \frac{150\times60 + 120\times80}{150+120}$

$= 68.89$

then,

$d_1 = \bar{x}_1 - \bar{x}_{12} = \cancel{68.89}\ 60 - 68.89 = -8.89$

$d_2 = \bar{x}_2 - \bar{x}_{12} = 80 - 68.89 = 11.11$

Now,

Combined SD $(\sigma_{12}) = \sqrt{\frac{n_1(d_1^2+\sigma_1^2) + n_2(d_2^2+\sigma_2^2)}{n_1+n_2}}$

$= \sqrt{\frac{150(79.03+491) + 120(123.43+81)}{150+120}}$

$= 12.73$

Variance $= (\sigma_{12})^2 = (12.73)^2 = 161.99$

Combined CV $= \frac{\sigma_{12}}{\bar{x}_{12}} \times 100\%$

$= \frac{12.73}{68.89} \times 100\%$

$= 18.48\%$

Skewness:

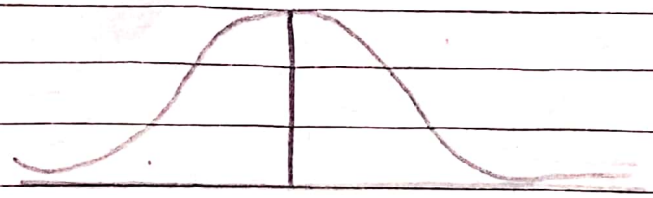
Skewness means 'lack of symmetry'. We study skewness to have an idea about the shape of the curve which we can draw with the help of the given data. So skewness shows the lack of symmetry in a distribution. It measures the degree of departure from symmetry and elongation of curve.

Scanned with CamScanner

Page No.
Date

Types of skewness

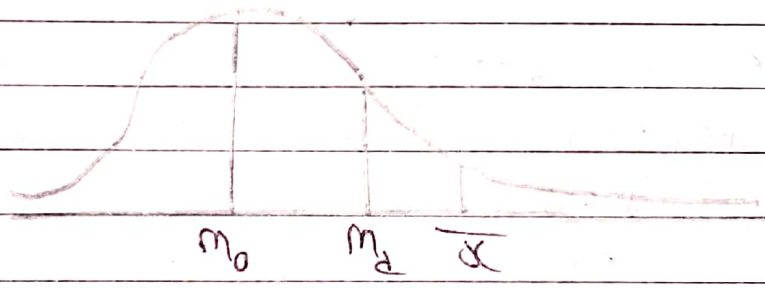
(i) Symmetrical distribution (no skewness)



mean = median = mode

It is clear from the above diagram that in a symmetrical distribution the values of mean median and mode coincide. The spread of the frequencies is the same on both side of the center point of the curve

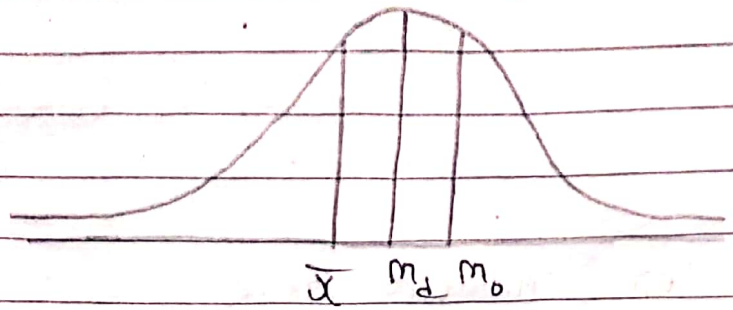
(ii) Positively skewed distribution (Right skewness)



It is clear from the above diagram in a positively skewed distribution, the value of the mean is maximum and that of the mode is least, the median lies in between the two. In positively skewed distribution the frequencies are spread out over a greater range of values on the right hand side than they are on the left hand side.

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

(iii) Negatively skewed distribution (Left skewness):



It is clear from above diagram in negatively skewed distribution the value of mode is maximum and mean of mode is least. The median lies in between the two. In the negatively skewed distribution the frequencies are spread out over a greater range of values on the left hand side than they are on right hand side.

(methods) Measure of skewness:

(i) Karl-Pearason's (coeff) measure of skewness. (S_{kp}).
(ii) Bowley's measure of skewness.
(iii) Kelly's measure of skewness.
(iv) Measure of skewness by using moments.

Karl-Pearason's Coefficient of skewness (skewness based on mean, median, mode & standard deviation):

$$\text{Skewness} = \text{mean} - \text{mode} \quad \Big\} \text{ absolute measure}$$
$$= \text{mean} - \text{median}$$

Relative measure:

It is the difference of mean and

Page No.
Date

mode divided by standard deviation. It is denoted by S_{KP} and is given by,

$$S_{KP} = \frac{\text{mean} - \text{mode}}{S.D} = \frac{\bar{x} - m_o}{\sigma} \quad \text{--- (i)}$$

If mode is ill defined then,

mode = ~~3mean~~ 3median − 2 mean

Hence,

$$S_{KP} = \frac{3(\bar{x} - m_d)}{\sigma} \quad \text{--- (ii)}$$

If we use formula (i) S_{KP} lies between −1 to +1.
If we use formula (ii) S_{KP} lies between −3 to +3.

Interpretation of S_{KP}:

i. If $S_{KP} = 0$ ie, $\bar{x} = m_d = m_o$, the distribution is ~~no~~ symmetrical.

ii. If $S_{KP} > 0$ ie $\bar{x} > m_d > m_o$, the distribution is positively skewed.

iii. If $S_{KP} < 0$ ie $\bar{x} < m_d < m_o$, the distribution is negatively skewed.

Eg:

Q) Find S_{KP} of:

x : 50 10 80 60 30 90 40 20 80

Soln

Calculation table

x	x^2
50	2500
10	100
80	6400
60	3600
30	900
90	8100
40	1600
20	400
80	~~1600~~ 6400
$\Sigma x = 460$	$\Sigma x^2 = 30,000$

we have

$$\bar{x} = \frac{\Sigma x}{n} = \frac{460}{9} = 51.1$$

m_0 = maximum repeated value
$= 80$

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{30000}{9} - \left(\frac{460}{9}\right)^2}$$

$$= 26.85$$

then

$$S_{KP} = \frac{\bar{x} - m_0}{\sigma}$$

$$= \frac{51.11 - 80}{26.85}$$

$$\boxed{\therefore S_{KP} = -1.08}$$

which shows that the distribu
is negatively skewed.

Bowley's measure of skewness (skewness based on Quartile

$$\text{Skewness} = Q_3 + Q_1 - 2m_d$$

Bowley's Coefficient of skewness is denoted by S_{KB} and is given by,

$$S_{KB} = \frac{Q_3 + Q_1 - 2m_d}{Q_3 - Q_1}$$

Note: S_{KB} lies between -1 to $+1$.

Scanned with CamScanner

Compute Bowley's Coefficient of skewness from:

7, 3, 14, 11, 9, 18, 20, 15

Arranging given data in ascending order

3, 7, 9, 11, 14, 18, 15, 18, 20

Here,

$n = 8$

$\therefore Q_1 = \left(\frac{n+1}{4}\right)^{th}$ item $= \left(\frac{8+1}{4}\right)^{th}$ item $= 2.25^{th}$ item

$= 2^{nd}$ item $+ 0.25$ (3rd item $- 2^{nd}$ item)

$= 7 + 0.25(9-7)$

$= 7.5$

$(Q_2)\ M_d = 2\left(\frac{n+1}{4}\right)^{th}$ item $= 2 \times 2.25 = 4.5^{th}$ item

$= 4^{th}$ item $+ 0.5$ (5th item $- 4^{th}$ item)

$= 11 + 0.5(14 - 11)$

$= 12.5$

$Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ item $= 3 \times 2.25 = 6.75^{th}$ item

$= 6^{th}$ item $+ 0.75$ (7th item $- 6^{th}$ item)

$= 15 + 0.75(18 - 15)$

$= 17.25$

Now,

$$S_{KB} = \frac{Q_3 + Q_1 - 2M_d}{Q_3 - Q_1}$$

$$= \frac{17.25 + 7.5 - 2 \times 12.5}{17.25 - 7.5}$$

$\therefore S_{KB} = -0.025 < 0$

Hence, the distribution is negatively skewed.

Scanned with CamScanner

Page No.
Date

Kelly's measure of skewness / (skewness based on deciles and percentiles) / ~~(Percentile~~

$$\text{skewness} = D_9 + D_1 - 2D_5$$
$$= P_{90} + P_{10} - 2P_{50}$$

Kelly's Coefficient of skewness (~~Percentile coefficient of~~) is denoted by S_{KK} and is given by

$$S_{KK} = \frac{D_9 + D_1 - 2D_5}{D_9 - D_1}$$

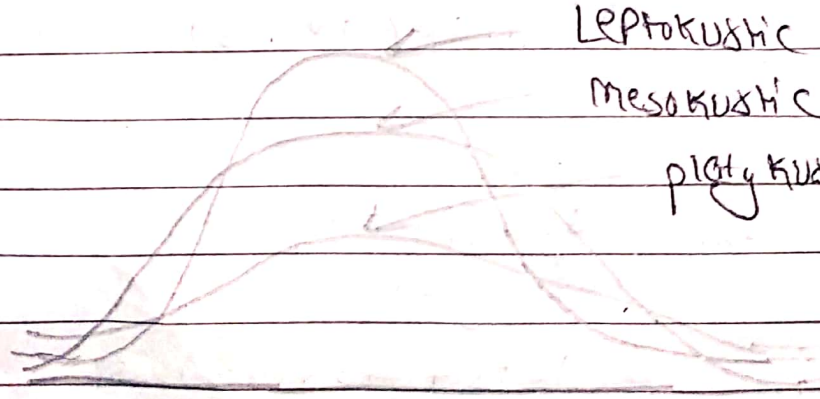
$$S_{KK} = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$$

Kurtosis:

It measures Peakedness or Flatness of the curve from the given frequency distribution.

Types of Kurtosis curve (Nature

Leptokurtic curve (Peaked)
Mesokurtic curve (Normal)
Platykurtic curve (flat)



Scanned with CamScanner

Page No.
Date

Measure (methods) of Kurtosis:
Percentile measure (Coefficient) of Kurtosis.
Measure of Kurtosis by using moments.

Percentile Coefficient of Kurtosis (Kelly's Coefficient of Kurtosis):

It is measure of Kurtosis based upon partition values (quartiles and percentiles). It is denoted by K & is given by,

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}$$

Interpretation
If $K = 0.263$, the distribution is mesokurtic.
If $K > 0.263$, the distribution is leptokurtic.
If $K < 0.263$, the distribution is platykurtic

Test the normality of the distribution

Class :	0-9	10-19	20-29	30-39	40-49	50-59
f :	7	8	10	6	8	5

Calculation table

Class	f	C.F
-0.5 - 9.5	7	7
9.5 - 19.5	8	15
19.5 - 29.5	10	25
29.5 - 39.5	6	31
39.5 - 49.5	8	39
49.5 - 59.5	5	44
	N = 44	

Scanned with CamScanner

Page No.
Date

For Q_3

Position of Q_3 class $= \left(\frac{3N}{4}\right)^{th}$ item

$= \left(\frac{3 \times 44}{4}\right)^{th}$ item

$= 33^{th}$ item

The CF just greater than 33 is 39. So, Q_3 class $Q_3 = 39.5 - 49.5$

$\therefore Q_3 = L + \frac{3N/4 - CF}{F} \times h$

$= 39.5 + \frac{33 - 31}{8} \times 10$

$\boxed{\therefore Q_3 = 42}$

For Q_1

Position of Q_1 class $= \left(\frac{N}{4}\right)^{th}$ item

$= (44/4)^{th}$ item

$= 11^{th}$ item

Thus, Q_1 ~~ties~~ class lies in $(9.5 - 19.5)$

$\therefore Q_1 = 9.5 + \frac{11 - 7}{8} \times 10$

$\boxed{\therefore Q_1 = 14.5}$

For P_{90}:

Position of $P_{90} = \frac{90N}{100} = \frac{90 \times 44}{100} = (39.6)^{th}$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

Page No.
Date

$\therefore$ P_{90} class lies in (49.5 – 59.5)

$$\therefore P_{90} = L + \frac{\frac{90N}{100} - CF}{f} \times h$$

$$= 49.5 + \frac{39.6 - 39}{5} \times 10$$

$$\boxed{\therefore P_{90} = 50.7}$$

For P_{10}

Position of $P_{10} = \frac{10N}{100} = \frac{10 \times 44}{100} = 4.4^{th}$ item

Thus, P_{10} class lies in (-0.5 – 9.5)

$$\therefore P_{10} = -0.5 + \frac{4.4 - 0}{7} \times 10$$

$$\boxed{\therefore P_{10} = 5.78}$$

we have

$$\text{Kurtosis } (K) = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

$$= \frac{42 - 14.5}{2(50.7 - 5.78)}$$

$$\boxed{\therefore K = 0.306}$$

Here,

$K = 0.306 > 0.263$ so, the curve is not normal ie the curve is leptokurtic curve.

//

Moments:

moments are ~~average~~ arithmetic average of various power of deviations of observation in which deviations are taken from mean (AM) or assumed

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Page No.
Date.

mean (A). They are used to determine characteristics of frequency distribution ie. central tendency, dispersion, skewness and kurtosis.

Types of moments:
Central moments.
Raw moments.

Central moments:

If deviation is taken from actual mean ($\bar{x}$) then it is called central moments. The rth central moment is denoted by μ_r and is given by

$$\mu_r = \frac{\Sigma(x-\bar{x})^r}{n}$$ → for individual series. ; $r = 1, 2, 3, 4$

$$\mu_r = \frac{\Sigma f(x-\bar{x})^r}{N}$$ → for discrete & continuous series

where x is mid-value in case of continuous series.

If $(x-\bar{x}) = x$ then,

$$\mu_r = \frac{\Sigma x^r}{n}$$

$$\mu_r = \frac{\Sigma f x^r}{N}$$

for $r = 1$,

$$\mu_1 = \frac{\Sigma x}{n}, \quad \mu_1 = \frac{\Sigma f x}{N}$$

for $r = 2$,

$$\mu_2 = \frac{\Sigma x^2}{n}, \quad \mu_2 = \frac{\Sigma f x^2}{N}$$

Scanned with CamScanner

for $r = 3$,

$$\mu_3 = \frac{\Sigma x^3}{n}, \quad \mu_3 = \frac{\Sigma f x^3}{N}$$

for $r = 4$,

$$\mu_4 = \frac{\Sigma x^4}{n}, \quad \mu_4 = \frac{\Sigma f x^4}{N}$$

Raw moments:

If deviation is taken from assumed mean (a) then it is called raw moment. The rth raw moment is denoted by μ_r' and is given by,

$$\mu_r' = \frac{\Sigma (x-a)^r}{n} \; ; \quad r = 1, 2, 3, 4$$

for individual series

$$\cancel{\mu_r' = \frac{\Sigma f (x-a)^r}{}}$$

$$\mu_r' = \frac{\Sigma f (x-a)^r}{N}$$ ← for discrete ~~& continuous series~~.

$$\mu_r' = \frac{\Sigma f (x-a)^r}{N}$$

$$= \frac{\Sigma f (hd')^r}{N}$$

$$= h^r \frac{\Sigma f (d')^r}{N}$$ ← for continuous series

If $(x-a) = d$ then,

$$\mu_r' = \frac{\Sigma d^r}{n}$$

$$\mu_r' = \frac{\Sigma f d^r}{N}$$

REDMI NOTE 5 PRO
MI DUAL CAMERA
Scanned with CamScanner

If $r = 1$,

$$\mu_1' = \frac{\Sigma d}{n}, \qquad \mu_1' = \frac{\Sigma f d}{N}$$

If $r = 2$,

$$\mu_2' = \frac{\Sigma d^2}{n}, \qquad \mu_2' = \frac{\Sigma f d^2}{N}$$

If $r = 3$,

$$\mu_3' = \frac{\Sigma d^3}{n}, \qquad \mu_3' = \frac{\Sigma f d^3}{N}$$

If $r = 4$,

$$\mu_4' = \frac{\Sigma d^4}{n}, \qquad \mu_4' = \frac{\Sigma f d^4}{N}$$

Relationship between Central Moments & Raw Moments

$$\mu_1 = \mu_1' - \mu_1' = 0$$
$$\mu_2 = \mu_2' - \mu_1'^2$$
$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$
$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1' - 3\mu_1'^4$$

Measure of skewness based upon moments:

$$\text{Skewness } (\beta_1) = \frac{\mu_3^2}{\mu_2^3}$$

If μ_3 is negative,

$$\text{Skewness } (\gamma_1) = \frac{\mu_3}{\mu_2 \sqrt{\mu_2}}$$

i.e. $\gamma_1 = \sqrt{\beta_1}$

Note:

measure of central tendency $(\bar{x}) = A + \mu_1'$

measure of dispersion $(\sigma) = \sqrt{\mu_2}$

variance $(\sigma^2)^2 = \mu_2$

Page No.
Date

Interpretation:

if $\mu_3 = 0$; $\beta_1 = 0$ & $\gamma_1 = 0$, the distribution is symmetrical.

if $\mu_3 > 0$; $\beta_1 > 0$ & $\gamma_1 > 0$, the distribution is positively skewed.

if $\mu_3 < 0$; $\gamma_1 < 0$, the distribution is negatively skewed.

Measure of Kurtosis based upon moments:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

& $\gamma_2 = \beta_2 - 3$

interpretation

if $\beta_2 = 3$; $\gamma_2 = 0$, the distribution is mesokurtic.

if ~~$\beta_2 > 3$~~ $\beta_2 > 3$; $\gamma_2 > 0$, the distribution is leptokurtic.

if $\beta_2 < 3$; $\gamma_2 < 0$, the distribution is platykurtic.

Given,

$A = 0$

First four raw moments,

$\mu_1' = -2$

$\mu_2' = 14$

$\mu_3' = -20$

$\mu_4' = 50$

First ~~four~~ four central moments are

$\mu_1 = \mu_1' - \mu_1' = (-2) - (-2) = 0$

Scanned with CamScanner

Date

$$u_2 = u_2' - (u_1')^2$$
$$= 14 - (-2)^2$$
$$= 10$$

$$u_3 = u_3' - 3u_2'u_1' + 2(u_1')^3$$
$$= -20 - 3 \times 14 \times (-2) + 2 \times (-2)^3$$
$$= -20 + 84 - 16$$
$$=$$

$$u_4 = u_4' - 4u_3'u_1' + 6u_2'u_1' - 3(u_1')^4$$
$$=$$

REDMI NOTE 5 PRO
MI DUAL CAMERA

Scanned with CamScanner

Steam Leaf Display:

Steam leaf consisting of stems as the main groups while the leaves are elements within each main group/stem.

Construct the steam and leaf plot for the following:
23 60 79 32 57 74 52 70 82 36 80 77 81 95
41 65 92

Sol:

Steam	Leaves	Frequency
2	3	1
3	2 6	2
4	1	1
5	7 2	2
6	0 5	2
7	9 4 0 7	4
8	2 1 0	3
9	5 2	2

Scanned with CamScanner