Unit - 3

DISPERSION

The measure of central tendancy gives us an idea of the concentration of the Observation about the central part of the distribution. if we know the average along. we ~~cano~~ cannot have a complete idea about the distribution for eg:- consider the following data sets.

I : 7, 8, 9, 10, 11

II : 3, 6, 9, 12, 15

III : 1, 5, 9, 13, 17.

in all these 3 data sets, the no of observations is 5 and the mean is 9. Thus we see that the measures of central tendancy alone are inadequate to give us a complete idea of the distribution. There must be some other measures. One such measure is dispersion.

the meaning of the word dispersion is Scatterdness

Measures of dispersion are statistical devices which reveal the extent to which individual items in a given Series are different from its average.

The important measures of dispersion are range, quartile deviation, mean deviation and standard deviation.

ABSOLUTE AND RELATIVE MEASURE OF DISPERSION.

A measure of dispersion can be expressed either in the absolute form or in the relative form. absolute measures are expressed in the same units in which the original data are given.

for eg:- if the income of sample items are measured in terms of rupees. an absolute measure of dispersion of income is also given in terms of rupees.

A measure of relative dispersion is the ratio of a measure of absolute dispersion to an appropriate average. it is a pure number without having any unit of measurement.

for eg:- the income of people in India is measured in terms of rupees while the income of people in USA is measured in terms of dollars. in this case variability of income in two country can be ~~comp~~ compared only by using relative measures of dispersi

Relative dispersion is usually called Coefficient of dispersion.

Coefficient of Range.

The relative measure of range is called coefficient of range. and it is given by

$$C \cdot R = \frac{L-S}{L+S}$$ L - largest value. S - Smallest value.

eg

① $= \frac{96-3}{96+3} = \frac{93}{99} = 0.9393$.

② $C \cdot R = \frac{59.5 - 9.5}{59.5 + 9.5} = \frac{50}{69} = 0.72$

Q) calculate range and CR for the data.

50 - 150	10
150 - 250	18
250 - 350	24
350 - 450	20
450 - 550.	16.

Range = 550 − 50 = 500

$$CR = \frac{550-50}{550+50} = \frac{500}{600} = \underline{0.8333}$$

D

Range

Range is a crude measure of dispersion. it is defined as the difference between the largest and smallest values in a series.

Range = L - S.

Where L is the largest value.

S is the smallest value

eg:- find range of the following data

8, 9, 36, 24, 18, 96, 7, 12, 3.

Range = 96-3

= 93

for a continious frequency table. Range is the difference between the true upper limit of the highest class and true lower limit of the lowest class.

Eg:- Calculate the Range for the following data.

Class	freq.	Continious class.
10 - 19.	3	9.5 - 19.5
20 - 29	7	19.5 - 29.5.
30 - 39	10	29.5 - 39.5.
40 - 49	6	39.5 - 49.5.
50 - 59.	4.	49.5 - 59.5.

Range = 59.5 - 9.5.

= 50

Merits And Demerits of Range.

Merits

- Range is easy to calculate
- Simple to understand.
- it can be effectively used in studying variations in exchange ~~changes~~ rates. Share prices.

Demerits

- Range is not based on all items in a given data.
- In the calculation of range. deviations from any average are not taken
- Value of range depends exclusively on extreme values. and thus it is very much effected by fluctuations in sampling.
- Range cannot be calculated from frequency tables with open end classes.

Quartile Deviation

Quartiles are the three points which divide the frequency distribution into 4 equal parts.

A measure of dispersion based on the quartiles is quartile deviation. it is also called the Semi inter Quartile range. The inter quartile range is the difference between the 3rd quartile and the 1st quartile ie $IQR = Q_3 - Q_1$

The quartile deviation [QD] is half of IQR.

$QD = \frac{1}{2}(Q_3 - Q_1)$

The Coefficient of Quartile deviation

it is obtained by dividing the quartile deviation with the arithmatic mean of Q_3 and Q_1.

ie Coefficient of $QD = \dfrac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

eg:- Calculate Quartile deviation and Coefficient of Quartile deviation if $Q_1 = 33.31$ and $Q_3 = 39.07$.

$QD = \frac{1}{2}(Q_3 - Q_1)$

$= \frac{1}{2}(39.07 - 33.31)$

$= \underline{\underline{2.88}}$

Coeff of $QD = \dfrac{Q_3 - Q_1}{Q_3 + Q_1} = \dfrac{39.07 - 33.31}{39.07 + 33.31} = \dfrac{5.76}{72.38}$

$= 0.07957$

$= \underline{\underline{0.08}}$

eg:- calculate Q.D for the following data.

18, 4, 9, 6, 5, 16, 20.

Ascending Order 4, 5, 6, 9, 16, 18, 20 7

$\frac{7+1}{4}^{th} = 2^{m} = \underline{\underline{5}}$. $Q_1 = 5$

$$\frac{3(N+1)}{4} = \frac{3 \times 8^2}{4_1} = 6 \quad Q_3 = 18.$$

$Q_1 = 5$

$Q_3 = 18$

$$QD = \frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}(18-5) = \frac{13}{2} = \underline{\underline{6.5}}$$

$$\text{Coefficient} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{13}{23} = \underline{\underline{0.5652}}$$

Q) Calculate coefficient of QD for following data

Class	
32-34	14
35 - 37	62
38 - 40	99
41 - 43	18
44 - 46	7

Ans

		cf
31.5 - 34.5	14	14
34.5 - 37.5	62	76
37.5 - 40.5	99	175
41.5 - 43.5	18	193
44.5 - 46.5	7	200
	200	

$Q_1 = \frac{N}{4}$ th $= \frac{200}{4} = 50$th. $\quad$ 34.5 - 37.5

$$Q_1 = L_1 + \frac{\left(\frac{N}{4} - m_1\right)}{f_1} \times C$$

$$Q_1 = 34.5 + \frac{\left(\frac{200}{4} - 14\right)}{62} \times 3$$

$$= 34.5 + \frac{(50 - 14)}{62} \times 3$$

$$= 34.5 + \frac{36}{62} \times 3$$

$$= 34.5 + 0.5806 \times 3$$

$$= 34.5 + 1.741$$

$$= \underline{36.24}$$

$l \Rightarrow$ lower limit of class.

$m \rightarrow$ c.f. just before the class.

$f \Rightarrow$ frequency of class.

$c \Rightarrow$ class width.

$$Q_3 = L_3 + \frac{\left(\frac{3N}{4} - m_3\right)}{f_3} \times c$$

$=$

$$Q_3 = \frac{3N}{4} = 3 \times 50 = 150$$

$$= 37.5 + \frac{\left(3 \times \frac{200}{4} - 76\right)}{99} \times 3$$

$$= \frac{150 - 76}{\cancel{99}\ 33} \times \cancel{3}$$

$$= 37.5 + 2.24$$

$$= \underline{39.74}$$

accordingly there are three types of mean deviations
they are (1) Mean deviation from mean.
(2) Mean deviation from median
(3) Mean deviation from mode.

Mean deviation from Mean

Mean deviation from mean is defined as the arithematic mean of the absolute deviations from arithematic mean.

Calculation of MD. from Mean

Let $x_1, x_2, x_3 \dots x_n$ are the observations. and let $\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$. be the arithematic ~~and~~ mean. Then.

Mean deviation about Mean

$$MD(\bar{x}) = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{N}$$

$$= \frac{1}{N} \sum_{i=1}^{n} |x_i - \bar{x}|.$$

Q) Calculate MD from AM.

X: 26, 32, 16, 49, 7

$\bar{x} = \frac{26 + 32 + 16 + 49 + 7}{5}$

$= \frac{130}{5}$

$= \underline{\underline{26}}$

X	$\lvert x_i - \bar{x} \rvert$
26	26-26 = 0
32	32-26 = 6
16	16-26 = 10
49	49-26 = 23
7	7-26 = 19
	58

$MD(\bar{x}) = \frac{1}{N} 58 = \frac{58}{5} = \underline{\underline{11.6}}$.

$$QD = \frac{1}{2}(Q_3 - Q_1)$$

$$= \frac{1}{2}(39.74 - 36.24)$$

$$= \left(\frac{3.5}{2}\right)$$

$$= 1.75$$

$$C.SD = \frac{3.5}{75.98} = \underline{0.04}$$

Merits And Demerits Of Quartile Deviation

- Quartile deviation is rigidly defined.
- it is easy to calculate & Simple to understand.
- It is not affected by extreme values.
- it can be computed from frequency tables with open end classes.

Demerits.

- quartile deviation is not based on all observations
- QD is not capable to further algebraic treatment.
- it is affected by Sampling fluctuations.

Mean Deviation

Mean deviation is defined as the arithmetic mean of the absolute deviations of observations from any average. the Commonly used averages from which deviations are taken are mean, median and mode.

Calculation of mean deviation from frequency table.

Let $x_1, x_2, \ldots x_n$ be the values of the variable and $f_1, f_2, \ldots f_n$ be the respective frequency.

The $M.D(\bar{x}) = \dfrac{f_1|x_1-\bar{x}| + f_2|x_2-\bar{x}| + f_3|x_3-\bar{x}| + \cdots + f_n|x_n-\bar{x}|}{f_1+f_2+\cdots+f_n}$

$= \dfrac{1}{N} \sum_{i=1}^{n} f_i |x_i - \bar{x}|$.

Q) Calculate Mean deviation about A.M from the following data.

X	f	Xf	$\lvert x_1-\bar{x}\rvert$	$f\lvert x_1-\bar{x}\rvert$
12	3	36	2.9	8.7
13	7	91	1.9	13.3
14	10	140	0.9	9.0
15	15	225	0.1	1.5
16	6	96	1.1	6.6
17	5	85	2.1	10.5
18	4	72	3.1	12.4
	50	745	~~12.1~~	62

$\bar{x} = 14.9$.

$MD = \frac{1}{N} \sum_{i=1} f_1 |x_1 - \bar{x}|$

$MD = \frac{1}{N} = \frac{62}{50}$ 1.24.

$= \underline{\underline{1.24}}$

Mean Deviation about Median

Let $x_1, x_2, \ldots x_n$ be observation and M be the median.

Then $MD(M) = \dfrac{|x_1 - M| + |x_2 - M| + \cdots + |x_n - x|}{n}$.

$$= \frac{1}{n} \sum_{i=1}^{n} |x_i - M|.$$

Q) Calculate mean deviation about the Median for the following data.

$x : 6, 7, 15, 16, 4, 2, 1.$

$1, 2, 4, \underline{6}, 7, 15, 16$

Median = 6.

x	$\lvert x_1 - M\rvert$
1	5
2	4
4	2
6	0
7	1
15	9
16	10
	31

$$MD(M) = \frac{1}{N} \sum_{i=1}^{n} |x_1 - M|$$

$$= \frac{1}{7} \times 31$$

$$= \underline{4.428}$$

Calculation of MD(x) from continious frequency table.

Let $x_1, x_2, x_3 \cdots x_n$ are the mid values of a continious frequency table with respective frequency. table with respective frequency $f_1, f_2 \cdots f_3$.

then the mean deviation from mean.

$$MD.(\bar{x}) = \frac{f_1|x_1-\bar{x}|+f_2|x_2-\bar{x}|+\cdots+f_n|x_n-\bar{x}|}{f_1+f_2+f \cdots +f_n}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i|x_i-\bar{x}|.$$

Q) calculate $MD(\bar{x})$. from following data.

Class.	Freq.	x.	xf	$(x_i-\bar{x})$	$f(x_i-\bar{x})$.
0-4	6	2.	12	8.8	52.8
4-8	14.	6	84	4.8	67.2
8-12	18	10	180	0.8	14.4
12-16.	10	14	140	3.2	32.0
16-20.	8	18	144	7.2.	57.6.
20-24.	4.	22.	88	11.2.	44.8.
			648		268.8

$$\bar{x} = \frac{648}{60}$$

$$= \underline{10.8}$$

$$MD = \frac{268.8}{60}$$

$$= \underline{4.48}.$$

Mean deviation about median for frequency data.
Let $x_1, x_2, \cdots, x_n$ be observations with respective frequency $f_1, f_2, \ldots f_n$ then

$$MD(M) = \frac{f_1|x_1 - M| + f|x_2 - M| + \cdots + f_n|x_n - M|}{f_1 + f_2 + \cdots + f_n}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - M|.$$

Q) Calculate Mean deviation about median from the following data

x	f	c.f.	$\lvert x_i - M \rvert$	$f\lvert x_i - M \rvert$
12	3	3	3	9
13	7	10	2	14
14	10	20	1	10
15	15	35	0	0
16	6	41	1	6
17	5	46	2	10
18	4	50	3	12
				61

Median = $\frac{50}{2} = 25$

Median = 15

$$MD(M) = \frac{1}{N}\sum_{i=1}^{n} |x_i - M|$$

$$= \frac{61}{50}$$

$$= 1.22$$

$= \cancel{20 + \frac{12}{4} \times 10}$

$= 20 + \cancel{3 \times 30}$

$=$

$= 20 + \left(\frac{12}{15} \times 10\right)$

$= \underline{\underline{28}}$

$$M.D = \frac{1}{N} \sum_{i=1}^{n} f_i |x_i - M|$$

$= \frac{1}{50} \times 508$

$= \underline{\underline{10.16}}$

Q) Calculate Mean Deviation for the following data

Class		C.f
20-30	3	3
30-40	61	64
40-50	132	192
50-60	153	349
60-70	140	489
70-80	51	540
80-90	2	542

Mean deviation about median for continous frequency data

Let $x_1, x_2, x_3 \dots x_n$ be the midvalues of continious frequency take with respective frequencies $f_1, f_2, \dots f_n$, then the MD about the Median is defined as

$$M.D(M) = \frac{f_1|x_1-M| + f_2|x_2-M| + \dots + f_n|x_n-M|}{f_1+f_2+f_3+\dots+f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i|x_i-M|}{\sum f_i}$$

Q) Calculate mean deviation about median for following data.

Class	frequency	cf	Middle value x	$x_1 - M$	
0-10	6	6	5	23	138
10-20	7	13	15	13	91
20-30	15	28	25	3	45
30-40	16	44	35	7	112
40-50	4	48	45	17	68
50-60	2	50	55	27	54
					508

Median class $= \frac{50}{2} = 25$ m.

$L_1 = 20$

Median $= L_1 + \frac{\left(\frac{N}{2} - m_1\right)}{f} \times C$

$= 20 + \frac{(25-13)}{15} \times 10$

Mean Deviation from Mode.

Mean deviation from mode is defined as the arithmetic mean of the absolute deviation from the mode.

For a row data, mean deviation from MD (Mo). can be calculated as follows

Let $x_1, x_2, \ldots x_n$ be the values and. M_0 be the mode of the data then.

$$MD(M_0) = \frac{|x_1 - M_0| + |x_2 - M_0| + \cdots + |x_n - M_0|}{n}$$

$$= \frac{\sum_{i=1}^{n} |x_i - M_0|}{n}$$

	$\lvert x - M_o \rvert$
5	1
4	0
8	4
3	1
4	0
6	2
4	0
5	1

$$MD = \frac{\sum_{i=1}^{n} \lvert x_i - M_o \rvert}{N} = \frac{1+4+1+2+1}{8} = \frac{9}{8} = \underline{1.125}$$

eg ②:- Calculate M.D from mode of the following data

Class	Freq
0-5	3
5-10	8
10-15	10
15-20	14
20-25	6
25-30	5

$$\text{Mode} = L_1 + \frac{(f_1 - f_0) \times C}{2f_1 - f_0 - f_2}$$

L_1 - lower limit of the class
f_1 - frequency of modal class
f_0 - frequency of previous class
f_2 - frequency after modal class
C - class width

$$\text{Mode} = 15 + \frac{(14-10) \times 5}{(2 \times 14) - 10 - 6}$$

$$\begin{array}{r} 28 - \\ 16 \\ \hline \underline{12} \end{array}$$

$$= 15 + \frac{4 \times 5}{12}$$

$$= 15 + 1.66$$

$$= \underline{16.66}$$

for frequency data, MD(Mo) can be calculated as follows.
let $x_1, x_2, \ldots x_n$ be the values with respective frequencies, $f_1, f_2, f_3, \ldots f_n$.

Then $$MD(Mo) = \frac{f_1|x_1 - Mo| + f_2|x_2 - Mo| + \cdots + f_n|x_n - Mo|}{f_1 + f_2 + \cdots + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i |x_i - Mo|}{\sum_{i=1}^{n} f_i}$$

for continuous frequency data, MD (Mo) can be calculated as follows.
let $x_1, x_2, \ldots x_n$ be the mid values with respective frequencies $f_1, f_2, f_3, \ldots f_n$. Then

$$MD(Mo) = \frac{f_1|x_1 - Mo| + f_2|x_2 - Mo| + \cdots + f_n|x_n - Mo|}{f_1 + f_2 + \cdots + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i |x_i - Mo|}{\sum_{i=1}^{n} f_i}$$

eg:-
Calculate mean deviation from mode for the following data

x: 5, 4, 8, 3, 4, 6, 4, 5
Mode = 4.

Class	Freq	x	$\lvert x - M_o \rvert$	× f
0-5	3	2.5	14.1667	42.498
5-10	8	7.5	9.1667	73.328
10-15	10	12.5	4.1667	41.66
15-20	14	17.5	0.8333	11.662
20-25	6	22.5	5.8333	34.998
25-30	5	27.5	10.8333	54.165
				258.311

$$MD = \frac{\sum_{i=1}^{n} \lvert x_i - M_o \rvert}{N} = \frac{258.311}{46} = \underline{\underline{5.615}} = \underline{\underline{5.62}}$$

Coefficient of Mean Deviation

Coefficient of Mean Deviation can be calculated by dividing the absolute measure of Mean deviation with the particular average from which deviations were taken

Coefficient of M.D from AM = $\frac{MD(\bar{x})}{\bar{x}}$

Coefficient of M.D from Median = $\frac{MD(M)}{M}$

Coefficient of M.D from Mode = $\frac{MD(M_o)}{M_o}$

Q) Coefficient of MD. of . eg 2

~~MoD~~ = Coefficient of MD = $\frac{5.62}{16.66} = \underline{\underline{0.337}}$

Calculation of S.D

(i) Raw data:

Let $x_1, x_2, \ldots, x_n$, are the values and $\bar{x}$ be its A.M.

Then the SD –

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_1-\bar{x})2}{n}}$$

Note that $\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}2)$.

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}$$

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\, n\bar{x} + n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum_{i=1}^{n} x^2 - n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n\left(\frac{\sum_{i=1}^{n} x_1}{n}\right)^2$$

$$= \sum_{i=1}^{n} x_1^2 - \frac{(\Sigma x_i)^2}{n}$$

$$\therefore \sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i^2\right)^2}{n}}{n}}$$

Merits And Demerits Of Mean Deviation

- Mean deviation is rigidly defined
- it is easy to calculate and simple to understand
- it is based on all values in a given data.
- Mean deviation is less affected by extreme values

Demerits

- Mean deviation cannot be used for further alge manipulations
- Mean deviation gives best results when deviations are taken from median .. However if the variabl in a data is high then median is not a representative average.

Standard Deviation

Standard deviation is defined as th positive square root of AM. of the square deviations from A.M. it is usually denoted the symbol of σ (small sigma).

The square of standard deviation called variance σ^2.

Q

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_1^2}{n} - \left(\frac{\sum_{i=1}^{n} x_1}{n^2}\right)^2}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_1}{n}\right)^2} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x})^2}$$

Q) Calculate S.D for the following data

X : 9, 5, 7, 3, 11.

$\sum_{i=1}^{n}$ 9+5+7+3+11

24
11
35

$$= \sqrt{\cancel{\frac{\sum_{i=1}^{n} x_i^2}{n}} - \cancel{\left(\frac{\sum_{i=1}^{n} x_1}{n}\right)^2}}$$

$= \sqrt{}$

$\bar{x} = \frac{35}{5} = 7$

x^2	x	$x-\bar{x}$	$(x-\bar{x})^2$
81	9	2	4
25	5	2	4
49	7	0	0
9	3	4	16
121	11	4	16
285			40

$$\sigma = \sqrt{\frac{\sum(x_1-\bar{x})^2}{n}}$$

$$= \sqrt{\frac{40}{5}}$$

$$= \sqrt{8}$$

$$= \underline{2\sqrt{2}}$$

$$\sigma = \sqrt{\frac{\sum x_1^2}{n} - \left(\frac{\sum x_1}{n}\right)^2} = \sqrt{\frac{285}{5} - \left(\frac{35}{5}\right)^2}$$

(3) Calculation of S.D for frequency data

Let $x_1, x_2, \dots x_n$ are the values i... the respective frequencies $f_1, f_2, \dots f_n$ then the S.D is defined as

$$SD = \sqrt{\frac{f_1(x_1-\bar{x})^2 + f_2(x_2-\bar{x})^2 + f_n(x_n-\bar{x})^2}{f_1+f_2+f_3+\dots+f_n}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} f_1 x_1^2}{N} - \left(\frac{\sum f_1 x_1}{N}\right)^2}$$

Q)

x	f	$x_i f_i$	$x_1^2 f_1$
12	4	48	576
13	11	143	1859
14	32	448	6272
15	21	315	4725
16	15	240	3840
17	8	136	2312
18	5	90	1620
20	4	80	1600
	100	1500	22,804

$$\frac{\sum_{i=1}^{n} f_1 x_1^2}{N}$$

$$\frac{1500}{100}$$

$$\frac{22804}{100} = \underline{228.04} - 225 \quad \sqrt{3.04} = 1.7$$

$= \sqrt{57 - \frac{49}{25}}$

$= \sqrt{57 - 1.9}$

$= \sqrt{\frac{285}{5} - \bar{x}^2}$.

$= \sqrt{\frac{285}{5} - 49}$.

$= \sqrt{57 - 49}$

$= \sqrt{8}$

$= \underline{\underline{2\sqrt{2}}}$.

Q) find standard deviation of

65, 66, 67, 68, 69, 67, 71, 70.

x	$d = x - 69$	d^2
65	-4	16
66	-3	9
67	-2	4
68	-1	1
69	0	0
67	-2	4
71	2	4
70	1	1
		39

$S.D(x) = SD(d)$.

$= \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n} - \bar{d}^2}$.

$= \sqrt{\frac{39}{8} - \left(\frac{-9}{8}\right)^2}$

$= \sqrt{4.87 - \frac{81}{64}}$

$= \sqrt{4.87 - 1.65}$

$= \sqrt{3.22} = 1.79$

Q). Calculate S.D.

x	f	M.O value	$x_i f_1$	$x_i^2 f_1$
25-30	4	27.5	110	3025
30-35	20	32.5	650	21125
35-40	38	37.5	1425	53437
40-45	24	42.5	1020	43350
45-50	10	47.5	475	22562
50-55	4	52.5	210	11025
	100		3890	154524

$$S.D = \sqrt{\frac{\sum_{i=1}^{n} f_1 x_1^2}{N} - \left(\frac{\sum f_1 x_1}{N}\right)^2}$$

$$= \sqrt{\frac{154524}{100} - \left(\frac{3890}{100}\right)^2}$$

$$= \sqrt{1545.24 - (38.9)^2}$$

$$= \sqrt{1545.24 - 1,513.21}$$

$$= \sqrt{32.03}$$

$$= \underline{\underline{5.65}}$$

Combined Standard Deviation

if two sample are combined, standard deviation of the combined group can be estimated using following formula.

$$\text{Combined S.D} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2}}, \text{ Where}$$

n_1 - no of elements in 1st group.
n_2 - no of elements in 2nd group.
σ_1 - SD of Ist group
σ_2 - SD of 2nd group.

Coefficient of Variation (CV)

The relative measure of standard deviation is called Coefficient of Variation. It is used to compare the variability of two or more series of observation. usually coefficient of variation is given as percentage.

$$\text{Coefficient of Variation (CV)} = \frac{S.D.}{\bar{x}} \times 100.$$

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

The series with lesser ~~C.D.~~ C.V is more consistant and the series with higher C.V is more variable,

eg:- The monthly wages paid to the workers in two factories ~~on~~ A and B are given below.

Factory	No of Workers	Average Monthly income	Variance
A	586	5250	4900
B	648	4750	5625

which firm has greater variability in wages.

$C.V_A = \frac{\sigma_A}{\bar{x}} \times 100$ $\quad \sigma_A = \sqrt{4900} = 70$

$\bar{x} = \frac{5250}{\cancel{5088}} =$

$= \cancel{\frac{4900}{}} \frac{70 \times 100}{5250}$

$= \underline{1.333}$

$C.V_B = \frac{\sigma_B}{\bar{x}} \times 100$ $\quad \sigma_B = \sqrt{5625} = 75$

$\bar{x} = 4750$

$= \frac{75 \times 100}{4750}$

$= \underline{1.578}$

factory B ~~C.V_B~~ has greater variability in wages.

eg 2) estimate S.D and C.V. for the following data

Class	x_1	freq	$x-A/20$	d	$f.d$	$f.d^2$
80 - 100	90	6	$-60/20=$	-3	-18	54
100 - 120	110	10	$-40/20$	-2	-20	40
120 - 140	130	24	$-20/20$	-1	-24	24
140 - 160	150	30	$0/20$	0	0	0
160 - 180	170	20	$20/20$	1	20	20
180 - 200	190	6	$40/20$	2	12	24
200 - 220	210	4	$60/20$	3	12	36
					-18	198

$$S.D = \left(\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}\right) \times 20$$

$$\sqrt{\frac{198}{100} - \left(\frac{-18}{100}\right)^2}$$

$$\sqrt{\quad - (0.18)^2}$$

$$\sqrt{1.98 - 0.0324}$$

$$\sqrt{1.947}$$

$$1.395 \times 20$$

$$\underline{\underline{27.90}}$$

Note that

$$d_i = \frac{x_i - A}{c} = \frac{x_i - 150}{20}$$

$$20d_i + 150$$

$$\Sigma x_i = \Sigma 20d_i + \ldots$$

$$C.V = \frac{S.D}{\bar{x}}$$

$$\bar{x} = \frac{-18}{100}$$

$$= \frac{27.90}{\cancel{0.18}} \times 100$$

$$\bar{x} = \frac{-18}{100} \times 20 + 150$$

$$= \frac{27.90}{146.4} \times 100$$

$$= \frac{-360 + 15}{100}$$

$$= 150 - 3.6$$

$$= \underline{146.4}$$

$$= \underline{\underline{19.05}}$$

Merits & Demerits

Merit

- Standard deviation is rigidly defined.
- Standard deviation is based on all observations
- S.D is capable of further algebraic treatment
- it is possible to calculate combined S.D.

of two or more series.

Demerits

- SD is more difficult to calculate compared to other measures of dispersion.
- S.D is much affected by extreme values.

Properties

1. S.D of a constant series. ie. c, c, c, c. is zero.

Proof. Let c, c, ... c, be the observation

$$SD = \sqrt{\frac{\Sigma (x_1 - \bar{x})^2}{n}} \qquad \text{here } \bar{x} = \frac{c+c+c\cdots c}{n} = \frac{nc}{n} = \underline{\underline{c}}$$

$$= \sqrt{\frac{(c-c)^2 + (c-c)^2 + \cdots + (c-c)^2}{n}} = \sqrt{\frac{0}{n}} = \sqrt{0} = \underline{\underline{0}}$$

S.D is unchanged if a constant is added or substracted from all observations.

(a) if $x_1, x_2 \ldots x_n$ are observation and $x_1 + A, x_2 + A, \ldots x_n + A$. or $x_1 - A, x_2 - A, \ldots x_n - A$,

then $S.D(x) = SD(x+A) = SD(x-A)$.

Proof.

$$SD(x) = \sqrt{\frac{\sum_{i=1}^{n} (x_1 - \bar{x})^2}{n}} \qquad (\overline{x+A}) = \bar{x} + A.$$

$$SD(x+A) = \sqrt{\frac{\left(\sum_{i=1}^{n} (x_1 + A - (\overline{x+A})\right)^2}{n}}$$