Measures of Dispersion

Variation

The deviation of an entry x in a population data set is the difference between the entry and the mean μ of the data set.

$\therefore$ Deviation of $x = x - \mu$

Variance:

The variance is the average of the squared differences of the given values from their arithmetic mean.

As a formula, the variance of population observations $x_1, x_2, \dots, x_N$, commonly denoted by σ^2 is

$$\sigma^2 = \frac{\Sigma(x_i - \bar{x})^2}{N} \quad \dots (1)$$

where $\bar{x}$ is the mean of all the observations and N is the total number of observations.

Computing variance for frequency distribution:

$$\sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N} \quad \text{or} \quad \sigma^2 = \left[\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2\right] \times h^2$$

For grouped data x_i will be the mid value of the i-th class.

-10, 0, 10, 20, 30   8, 9, 10, 11, 12

Mean = 10   Mean = 10

Standard deviation: The positive square root of the variance is the standard deviation.

That is standard deviation is the positive square root of the mean-square deviations of the observations from their arithmetic mean.

If $x_1, x_2, \ldots, x_N$ be N observations of a variable, then the standard deviation is defined as

$$\sigma = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{N}}$$

For frequency distribution. Standard ~~defn~~ deviation is defined as

$$\sigma = \sqrt{\frac{\Sigma f_i(x_i-\bar{x})^2}{N}} \quad \text{or} \quad \sigma = \left[\sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2}\,\right] \times h$$

Coefficient of variance:

A coefficient of variance is computed as a percentage of the standard deviation of the distribution of the mean of the same distribution. Symbolically

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

where σ = Standard deviation

$\bar{x}$ = mean

Problem: Calculate the variance, Standard deviation and co-efficient of variance for the age distribution given below:

Age	24.5-29.5	29.5-34.5	34.5-39.5	39.5-44.5	44.5-49.5	49.5-54.5
Frequency	3	9	15	12	7	4

Soln:

Age	Frequency f_i	Mid value x_i	$d_i = \frac{x_i - A}{c}$	$f_i d_i$	$f_i d_i^2$
24.5-29.5	3	27	-2	-6	12
29.5-34.5	9	32	-1	-9	9
34.5-39.5	15	37 = A	0	0	0
39.5-44.5	12	42	1	12	12
44.5-49.5	7	47	2	14	28
49.5-54.5	4	52	3	12	36
	$\Sigma f_i = 50$			$\Sigma f_i d_i = 23$	$\Sigma f_i d_i^2 = 97$

Variance,

$$\sigma^2 = \left[\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2\right] \times h^2$$

$$= \left[\frac{97}{50} - \left(\frac{23}{50}\right)^2\right] \times 5^2$$

$$= 43.21$$

Standard deviation,

$$\sigma = \left[\sqrt{\frac{\sum f_i d_i^2}{N} - \left(\frac{\sum f_i d_i}{N}\right)^2}\,\right] \times h$$

$$= \sqrt{43.21} = 6.573$$

Co-efficient of variance:

We know, co-efficient of variance,

$$\frac{\sigma}{\bar{x}} \times 100$$

Now, mean, $\bar{x} = A + \frac{\sum f_i d_i}{N} \times h$

$$= 37 + \frac{23}{50} \times 5$$

$$= 39 + 2.3$$

$$= 39.3$$

$\therefore$ $C.V = \frac{6.573}{39.3} \times 100$

$$= 16.73\%$$

(i) Calculate variance, standard deviation and co-efficient of variance from the following frequency distribution:

Profit	10-20	20-30	30-40	40-50	50-60
No. of companies	8	12	20	6	4

(ii)

Profit	0-10	10-20	20-30	30-40	40-50
No. of Companies	6	25	36	20	13

(iii)

Profit	0-10	10-20	20-30	30-40	40-50	50-60
No. of Companies	8	12	20	30	20	10

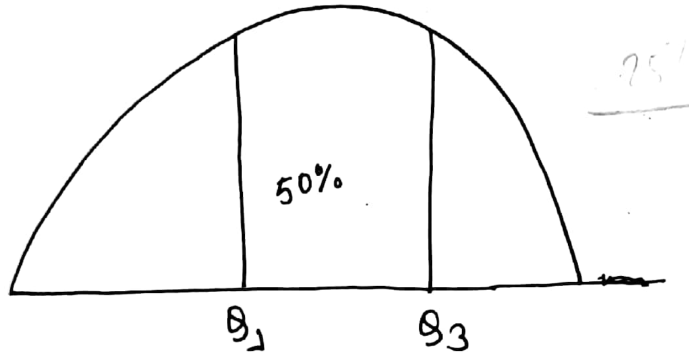
(iv)

Tax	5-10	10-15	15-20	20-25	25-30	30-35	35-40
No. of Managers	18	30	46	28	20	12	6

nterquartile range or quartile deviation:

Interquartile range represents the difference between the third quartile Q_3 and the first quartile Q_1. Symbolically $Q_3 - Q_1$ is interquartile range. The semi-inter quartile range is called the quartile deviation.

$$\therefore Q.D = \frac{Q_3 - Q_1}{2}$$



problem: The profit earned by 100 companies are given below:

profits	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of companies	4	8	18	30	15	10	8	7

(i) calculate the range within which middle 50% companies fall

(ii) calculate quartile deviation.

Scanned with CamScanner

Class Interval	Frequency	Cumulative Frequency
20-30	4	4
30-40	8	12
40-50	18	30
50-60	30	60
60-70	15	75
70-80	10	85
80-90	8	93
90-100	7	100

1st Quartile

Here, $N=100$, $\frac{1\times100}{4}=25$

(40-50) is the 1st quartile class because 25th observation lies (40-50)

$$Q_1 = 40 + \frac{25-12}{18} \times 10$$

$$= 47.22$$

Again 3rd quartile

$$\frac{3\times100}{4} = 75$$

75th observation Lies on the class (60-70).
so quartile class is (60-70)

$$\therefore Q_3 = 60 + \frac{75-60}{15} \times 10$$

$$= 70$$

(i) Range within which middle 50% companies fall

$$= Q_3 - Q_1$$

$$= 70 - 47.22$$

$$= 22.78 \quad \text{(Ans)}$$

(ii) Quartile deviation,

$$\frac{Q_3 - Q_1}{2} = \frac{70 - 47.22}{2}$$

$$= 11.39$$

H.W

Age	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of members	6	5	8	15	7	6	3

(i) calculate the range within which middle 50% members fall.

(ii) calculate quartile deviation.

Mean Deviation: Mean deviation is an average of absolute deviation of each observation from the mean. Symbolically,

$$M.D = \frac{\Sigma f_i |x_i - \bar{x}|}{N}, \quad \text{where } \bar{x} = \text{mean}$$

Problem: Calculate the mean deviation for the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	6	5	8	15	7	6	3

Soln:

Marks	f_i	Mid Value x_i	d_i	$f_i d_i$	$\bar{x}$	$\|x_i-\bar{x}\|$	$f_i\|x_i-\bar{x}\|$
0-10	6	5	-3	-18		28.4	170.4
10-20	5	15	-2	-10		18.4	92
20-30	8	25	-1	-8	$\bar{x} = A + \frac{\Sigma f_i d_i}{N} \times h$	8.4	67.2
30-40	15	35	0	0	$= 35 + \frac{-8}{50} \times 10$	1.6	24
40-50	7	45	1	7	$= 33.4$	11.6	81.2
50-60	6	55	2	12		21.6	129.6
60-70	3	65	3	9		31.6	94.6

$\Sigma f_i = 50$ $\quad \Sigma f_i d_i = -8$ $\quad \Sigma f_i |x_i - \bar{x}| = 658.4$

$$\text{Mean Deviation} = \frac{\sum f_i |x_i - \bar{x}|}{N}$$
$$= \frac{658.4}{50}$$
$$= 13.168 \text{ (Ans)}$$

Empirical Relationship

Quartile Deviation (Q.D) = $\frac{2}{3}$ Standard deviation (6)

Mean deviation (M.D) = $\frac{4}{5}$ Standard deviation (6)

Quartile Deviation (Q.D) = $\frac{5}{6}$ Mean deviation.

Problem: Calculate ~~Standard~~ deviation and then calculate mean deviation using empirical relation for the following data;

Profit (Lakhs)	10-20	20-30	30-40	40-50	50-60
No. of companies	8	12	20	6	4

Profit	No. of Companies f_i	Mid value x_i	d_i	$f_i d_i$	$f_i d_i^2$
10-20	8	15	-2	-16	32
20-30	12	25	-1	-12	12
30-40	20	35	0	0	0
40-50	6	45	1	6	6
50-60	4	55	2	8	16
				$\Sigma f_i d_i = -14$	$\Sigma f_i d_i^2 = 66$

Standard deviation.

$$\sigma = \sqrt{\frac{\Sigma f_i d_i^2}{N} - \left(\frac{\Sigma f_i d_i}{N}\right)^2} \times h$$

$$= \sqrt{\frac{66}{50} - \left(\frac{-14}{50}\right)^2} \times 10$$

$\simeq 11.14$ Lakh

Mean deviation $= \frac{4}{5} \times \sigma$

$= \frac{4}{5} \times 11.14$

$= 8.912$ Lakh

* problem: Calculate mean deviation and then calculate quartile deviation using the following empirical relation for the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	6	5	8	15	7	6	3