F.Y.B.COM Module-III Dispersion

Prof. Anil O.Khadse
Dept. of Maths and Stats
Sheth NKTT College, Thane

LEARNING OBJECTIVE

- They will be able to:
  - Describe the homogeneity or heterogeneity of the distribution,
  - understand the reliability of the mean,
  - compare the distributions as regards the variability.

Dispersion (Definition)

- Dispersion is the spreadness / scattered ness of the data series around it average.

- Dispersion is the extent to which values in a distribution differ from the average of the distribution.

Why measures of dispersion? (Significance)

- Determine the reliability of an average
- Serve as a basis for the control of the variability
- To compare the variability of two or more series and
- Facilitate the use of other statistical measures.

Dispersion Example

- Number of minutes 20 clients waited to see a consulting doctor

Consultant Doctor

X		Y	
05	15	15	16
12	03	12	18
04	19	15	14
37	11	13	17
06	34	11	15

- X:Mean Time – 14.6 minutes
- Y:Mean waiting time 14.6 minutes
- What is the difference in the two series?

X: High variability, Less consistency.
Y: Low variability, More Consistency

Characteristics of an Ideal Measure of Dispersion

1. It should be rigidly defined.
2. It should be easy to understand and easy to calculate.
3. It should be based on all the observations of the data.
4. It should be easily subjected to further mathematical treatment.
5. It should be least affected by the sampling fluctuation
6. It should not be unduly affected by the extreme values.

Measure of Dispersion

- Absolute: Measure the dispersion in the original unit of the data.
- Variability in 2 or more distr^{n} can be compared provided they are given in the same unit and have the same average.
- Relative: Measure of dispersion is free from unit of measurement of data.
- It is the ratio of a measure of absolute dispersion to the average, from which absolute deviations are measured.
- It is called as co-efficient of dispersion.

Measures of Dispersion

Absolute

1.Range

2.Quartile Deviation

3. Mean Deviation

4. Standard Deviation

Relative

1. Coeff. of Range

2. Coeff of Q.D

3.Coeff of M.D

4. Coefficient of Variation

Range:

The difference between the values of the two extreme items of a series.

Range = Maximum - Minimum

Example:

42	28	28	61	31	23	50	34	32	37

Maximum= 61   Minimum = 23

Range = 61-23 = 38

Co-efficient of Range:

$$\text{Coeff.of Range} = \frac{Max - Min}{Max + Min}$$

$$= \frac{61-23}{61+23}$$

$$= \frac{38}{84} = 0.452$$

Example:

The temperature recorded at the end of every 4 hours during a day at the two cities A and B

A: 28, 30, 34, 31, 28, 27

B: 25, 28, 34, 40, 32, 26

Decide in which city there is variation in temperature.

$$Range = Max - Min$$

For City A,

$\text{Max} = 34, \text{Min} = 27$

$$Range = 34 - 27 = 7$$

$$Coeff.of\ Range = \frac{Max-Min}{Max+Min} = \frac{34-27}{34+27} = \frac{7}{61} = 0.11$$

For City B,

$\text{Max} = 40, \text{Min} = 25$

$$Range = 40 - 25 = 15$$

$$Coeff.of\ Range = \frac{Max-Min}{Max+Min} = \frac{40-25}{40+25} = \frac{15}{65} = 0.23$$

City B

Quartiles Deviation

Q. D or Semi Interquartile Range:

$$= (Q_3 - Q_1)/ 2$$

Coefficient of quartile deviation:

$$Coeff.of\ Q.D = \frac{Q3 - Q1}{Q3 + Q1}$$

Ex: Calculate Q.D and coefficient of Q.D for the following data

Height	58	54	60	61	62	63	64
No. of students	8	12	20	13	7	5	3

Height	No. of students	<c.f
58	8	8
54	12	20
60	20	40
61	13	53
62	7	60
63	5	65
64	3	68
	N=68	

For Q1, N/4= 68/ 4 = 17

Q1= 54

For Q3,
3N/4 = 3x68/4= 51

Q3 = 61

$$Q.D = \frac{Q3-Q1}{2}$$

$$= \frac{61-54}{2} = 3.5$$

$$\boldsymbol{Coeff.of\ Q.D = \frac{Q3-Q1}{Q3+Q1}}$$

$$\boldsymbol{= \frac{61-54}{61+54}}$$

$$\boldsymbol{= \frac{7}{115}}$$

$$\boldsymbol{Coeff.of\ Q.D} = 0.061$$

Ex: Calculate Q.D and coefficient of Q.D for the following data

Marks	0-20	20-40	40-60	60-80	80-100
No. of students	23	37	50	24	16

Marks	f	< c.f
0-20	23	23
20-40	37	60
40-60	50	110
60-80	24	134
80-100	16	150

For Q1

N/4 = 150/4 = 37.5

20-40 is Q1 class

For Q3

3N/4 = 3x150/4 = 112.5

60-80 is Q3 class

$$Q1 = l1 + \frac{(\frac{N}{4} - c.f)}{f}(l2 - l1)$$

$$= 20 + \frac{(37.5 - 23)}{37}(40 - 20)$$

$$= 20 + \frac{(14.5)}{37}(20)$$

$$Q1 = 20 + 7.84 = 27.84$$

$$Q3 = l1 + \frac{(\frac{3N}{4} - c.f)}{f}(l2 - l1)$$

$$= 60 + \frac{(112.5 - 110)}{24}(80 - 60)$$

$$= 60 + \frac{(2.5)}{24}(20)$$

$$= 60 + 2.08 = 62.08$$

$$\boldsymbol{Coeff.of\ Q.D = \frac{Q3 - Q1}{Q3 + Q1}}$$

$$\boldsymbol{Coeff.of\ Q.D = \frac{62.08 - 27.84}{62.08 + 27.84}} = \frac{\mathbf{34.24}}{\mathbf{89.92}} = 0.38$$

Mean Deviation (MD)

- The average of difference of the values of items from some average of the series (ignoring negative sign), i.e. the arithmetic mean of the absolute differences of the values from their average .

Mean Deviation

For individual series: $X_{1,}$ X_2, ……… X_n

$$M.D\ from\ A = \frac{\sum\ |x - A\ |}{n}$$

Where A =Mean/ Median /Mode

For Frequency distribution: If $X_{1,}$ X_2, ……… $X_{n\ \&}$ with corresponding frequency $f_{1,}$ f_2, ……… f_n

$$M.D\ from\ A = \frac{\sum f\ |x - A\ |}{N}$$

Mean Deviation

1. $M.D\ from\ Mean = \frac{\sum f\ |x - \bar{x}\ |}{N}$

2. $M.D\ from\ Median = \frac{\sum f\ |x - M\ |}{N}$

3. $M.D\ from\ Mode = \frac{\sum f\ |x{-}Mode\ |}{N}$

Coefficient od M.D=$\frac{M.D}{A}$

Example:

Find mean deviation from mean from the following

15, 22, 27, 11, 9, 21, 14

X	$x-\bar{x}$	$\lvert x-\bar{x}\rvert$
15	-2	2
22	5	5
27	10	10
11	-6	6
9	-8	8
21	4	4
14	-3	3
$\Sigma x =$ 119		$\Sigma\lvert x-\bar{x}\rvert =$ 38

$$\bar{x} = \frac{\sum x}{n} = \frac{119}{7} = 17$$

$M.D\ from\ Mean$

$$= \frac{\sum \lvert x-\bar{x}\rvert}{n} = \frac{38}{7}$$

Example:

Find Mean Deviation from mean for the following series.

X	f	fx	$\|x-\bar{x}\|$	$f\|x-\bar{x}\|$
6	2	12	2	4
7	5	35	1	5
8	4	32	0	0
9	3	27	1	3
10	2	20	2	4
Total	**16**	**126**		$\sum f\|x-\bar{x}\|$ **=16**

$$\bar{x} = \frac{\sum fx}{N} = \frac{126}{16} = 7.875 = 8$$

$$M.D\ from\ Mean = \frac{\sum f\ |x-\bar{x}\ |}{N} = \frac{16}{16}=1$$

Ex: Find mean deviation from median from the following data

Efficiency Index	Employees	
18-22	20	
22-26	30	
26-30	11	
30-34	3	
34-38	1	

Ex: Find mean deviation from median from the following data

Efficiency Index	f	< c.f
18-22	20	20
22-26	30	50
26-30	11	61
30-34	3	64
34-38	1	65
	N=65	

For Med, N/2=65/2= 32.5

22-26 is median class

$$\text{Med}= l1 + \frac{(\frac{N}{2} - c.f)}{f}(l2 - l1)$$

$$= 22 + \frac{(32.5-20)}{30}(26-22)$$

$$= 22 + 1.66 = 23.66$$

C.I	f	<c.f	M.P x	|X - M|	f|X - M|
18-22	20	20	20	3.66	73.2
22-26	30	50	24	0.34	10.20
26-30	11	61	28	4.34	47.74
30-34	3	64	32	8.34	25.02
34-38	1	65	36	12.34	12.34
	N=65				$\sum f\ \lvert x-M \rvert$ **168.5**

$$M.D\ from\ Med = \frac{\sum f\ \lvert x-M \rvert}{N} = \frac{168.5}{65} = 2.59$$

Standard Deviation (σ)

It is the positive square root of the average of squares of deviations of the observations from the mean. This is also called root mean squared deviation (σ) .

For individual series: If $x_1, x_2, \ldots\ldots\ldots x_n$ are the values of variable x, S.D is defined as

$$S.D, \sigma = \frac{\sqrt{\sum (x - \bar{x})^2}}{n}$$

$$S.D, \sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

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

Standard Deviation (σ)

For discrete series:

If $X_{1,}\ X_{2,}\ \ldots\ldots\ldots\ X_n$ *& with corresponding frequency* $f_{1,} f_{2,}\ \ldots\ldots\ldots\ f_n$ *then s.d is defined as*

$$S.D,\sigma = \frac{\sqrt{\sum f(x-\bar{x})^2}}{N}$$

$$= \sqrt{\frac{\sum f x^2}{N} - \bar{x}^2}$$

Where $\bar{x} = \frac{\sum f x}{N}$

Standard Deviation (σ) Contd.

Variance: It is the square of the s.d

Variance = $(s.d)^2$

Coefficient of Variation (CV): Corresponding Relative measure of dispersion.

$$C.V = \frac{s.d}{Mean} \times 100$$

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

Ex. Find standard deviation 25, 28, 32, 20, 35, 40

x	$x - \bar{x}$	$(x - \bar{x})^2$
25	-5	25
28	-2	4
32	2	4
20	-10	100
35	5	25
40	10	100
Σ x=180		$\Sigma(x - \bar{x})^2$ **258**

n = 6

$$x = \frac{\Sigma x}{n} = \frac{180}{6} = 30$$

$$S.D, \sigma = \frac{\sqrt{\Sigma(x - x)^2}}{n}$$

$$S.D, \sigma = \frac{\sqrt{258}}{6}$$

$$\sigma = \sqrt{43} = 6.557$$

Ex. Find standard deviation 8,12, 9, 15, 10, 7, 13

x	x^2
8	64
12	144
9	81
15	225
10	100
7	49
13	169
$\sum x=74$	$\sum x^2$ **832**

$$\bar{x} = \frac{\sum x}{n} = \frac{74}{7} = 10.571$$

$$S.D, \sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{832}{7} - (10.571)^2}$$

$$= \sqrt{118.857 - 111.755}$$

$$= \sqrt{7.102} = 2.665$$

Ex: Find mean and standard deviation. Also find Coefficient of Variation

X	9	8	7	10	11
f	20	10	5	10	5

x	f	fx	fx^2
9	20	180	1620
8	10	80	640
7	5	35	245
10	10	100	1000
11	5	55	605
Total	*N=50*	$\sum fx$= *450*	$\sum fx^2$ = *4110*

Mean, $x = \frac{\sum f x}{N} = \frac{450}{50} = 9$

$$\text{s.d} = \sqrt{\frac{\sum f x^2}{N} - \bar{x}^2}$$

$$\text{s.d} = \sqrt{\frac{4110}{50} - 9^2}$$

$$\text{s.d} = \sqrt{82.200 - 81}$$

$$\mathrm{s.d} = \sqrt{1.200} = 1.095$$

$$C.V = \frac{s.d}{Mean} \times 100$$

$$C.V = \frac{1.095}{9} \times 100$$

$$C.V = 12.17\%$$

Ex: Calculate mean and s.d for the following data

Sales '000Rs'	No. of Shops
0-10	3
10-20	5
20-30	8
30-40	3
40-50	1

Solution

Sales	f	$M.P\ (x)$	fx	fx^2
0-10	3	5		
10-20	5	15		
20-30	8	25		
30-40	3	35		
40-50	1	45		
	N=20		$\sum fx$	$\sum fx^2$

Solution

Sales	f	$M.P\ (x)$	fx	fx^2
0-10	3	5	15	75
10-20	5	15	75	1125
20-30	8	25	200	5000
30-40	3	35	105	3675
40-50	1	45	45	2025
	N=20		$\sum fx=440$	$\sum fx^2=11900$

Mean $\bar{x} = \frac{\sum f x}{N} = \frac{440}{20}$ =22

$$\mathrm{s.d} = \sqrt{\frac{\sum f x^2}{N} - x^2}$$

$$\mathrm{s.d} = \sqrt{\frac{11900}{20} - (22)^2}$$

$$\mathrm{s.d} = \sqrt{595 - 484}$$

$$\mathrm{s.d} = \sqrt{111}$$

$$\mathrm{s.d} = \mathit{10.54}$$

Ex: The following distribution gives the weight of forty children's . Calculate coefficient of variation

Weights (kg)	No. of children
5-10	4
10-15	8
15-20	12
20-25	10
25-30	6

Solution

Weights	f	$M.P\ (x)$	fx	fx^2
5-10	4	7.5	30	225
10-15	8	12.5	100	1250
15-20	12	17.5	210	3675
20-25	10	22.5	225	5062.5
25-30	6	27.5	165	4537.5
	N=40		$\sum fx = 730$	$\sum fx^2 = 14750$

Mean , $\bar{x} = \dfrac{\sum fx}{N} = \dfrac{730}{40}$ =18.25

$$\text{s.d} = \sqrt{\frac{\sum f x^2}{N} - \bar{x}^{\,2}}$$

$$\text{s.d} = \sqrt{\frac{14750}{40} - (18.25)^2}$$

$$\text{s.d} = \sqrt{368.75 - 333.0625}$$

$$\text{s.d} = \sqrt{35.6875} = 5.97$$

$$C.V = \frac{s.d}{Mean} \times 100$$

$$C.V = \frac{5.97}{18.25} \times 100 = 32.71\%$$

Combine S.D

If there are two groups containing n_1 and n_2 observations. Let $\overline{x_1}$, $\overline{x}_2$ be the averages of two groups and σ_1, σ_2 be the standard deviations of he two groups respectively.

	Group-I	Group-II
Number	n_1	n_2
Average/ Mean	$\overline{x}_1$	$\overline{x}_2$
S.D	σ_1	σ_2

Combine s.d by taking two groups together is given by

$$s.d\ \sigma = \frac{\sqrt{n_1({\sigma_1}^2 + d12\quad) + n_2({\sigma_2}^2 + d22)}}{n1 + n2}$$

Where

$d1 = \overline{x_1} - \overline{x}$

$d2 = \overline{x_2} - \overline{x}$

$\overline{x}$= combine Mean =$\frac{n_1\overline{x_1}+n_2\overline{x_2}}{n_1+n_2}$

Example:
Find combine mean and standard deviation from the following information

	Boys	Girls
Numbers	100	150
Mean weight	50 kg	40 kg
S.D	5 kg	6 kg

Given: $n_1 = 100,$ $n_2 = 150$ $\overline{x_1} = 50,$
$\overline{x_2} = 40$ $\sigma_1 = 5,$ $\sigma_2 = 6$
${\sigma_1}^2 = 25$ ${\sigma_2}^2 = 36$

$\overline{x}$= combine Mean $= \frac{n_1\overline{x_1}+n_2\overline{x_2}}{n_1+n_2}$

$$\overline{x}= \frac{100\times50+150\times40}{100+150} = \frac{5000+6000}{250} = \frac{11000}{250} = 44$$

$d1 = \overline{x_1} - \overline{x}$ =50 – 44 = 6 $\therefore d_1{}^2 = 36$

$d2 = \overline{x_2} - \overline{x}$ = 40 – 44 = -4 $\therefore d_2{}^2 = 16$

$$s.d\ \sigma = \frac{\sqrt{n_1(\sigma_1{}^2 + d12\ \ ) + n_2(\sigma_2{}^2 + d22)}}{n1 + n2}$$

$$= \frac{\sqrt{100(25+36)+150(36+16)}}{100+150}$$

$$= \frac{\sqrt{100(61)+150(52)}}{250}$$

$$= \frac{\sqrt{6100+7800}}{250}$$

$$= \frac{\sqrt{13900}}{250}$$

$$= \sqrt{55.6}$$

$$=7.46$$

Example:
Find combine mean and coefficient of variation for each group from the following information and decide which is more consistent

	Male	Female
Numbers	40	60
Mean Height	170 cm	160 cm
S.D	5 cm	2 cm

Given: $n_1 = 40,$ $n_2 = 60$ $\overline{x_1} = 170,$
$\overline{x_2} = 160$ $\sigma_1 = 5,$ $\sigma_2 = 2$

$\overline{x}$= combine Mean $=\frac{n_1\overline{x_1}+n_2\overline{x_2}}{n_1+n_2}$

$$\overline{x}= \frac{40\times170+60\times160}{40+60} \qquad \overline{x}= \frac{6800+9600}{100} = 164$$

$$C.V = \frac{s.d}{Mean}\times100$$

$$(C.V)M = \frac{5}{170} \times 100 = 2.94\%$$

$$(C.V)F = \frac{2}{160} \times 100 = 1.25\%$$

$$(C.V)F < (C.V)M$$

Female group is more consistent

Example:

The A.M of 5 observation is 57.2 and the sum of squares of x values is 17230. Find s.d

Sol:

Given: n= 5, A.M = 57.2 ,

sum of squares of x values , $\Sigma x^2 = 17230$, s.d = ?

$$s.d = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

$$s.d = \sqrt{\frac{17230}{5} - 57.2^2}$$

$$s.d = \sqrt{3446 - 3271.84}$$

$$= \sqrt{174.16}$$

$$= 13.1970$$

Example:

The coefficient of variation of group of observation is 23.0716% and the mean was 57.2. Find s.d

Sol:

Given: C.V = 23.0716, A.M = 57.2

s.d = ?

$$C.V = \frac{s.d}{Mean} \times 100$$

$$23.0716 = \frac{s.d}{57.2} \times 100$$

$$23.0716 \times 57.2 = s.d \times 100$$

$$1319.6955/100 = s.d$$

$$s.d = 13.1969$$

Example:

The coefficient of quartile deviation for a certain group of observation 0.1. The sum of two quartiles is 100. Find two quartile.

Sol:

Given: Coeff. Of Q.D = 0.1

Sum of two quartiles =100

Q3+Q1 = 100, Q1 =?, Q3 = ?

$$\boldsymbol{Coeff.of\ Q.D = \frac{Q3 - Q1}{Q3 + Q1}}$$

$$0.1 = \frac{Q3-Q1}{100}$$

$$Q3 - Q1 = 10$$

$$Q3 + Q1 = 100 \quad -----(1)$$
$$Q3 - Q1 = 10 \quad -----(2)$$

Solving equation 1 & 2 simultaneously

$$Q3 + \cancel{Q1} = 100$$
$$Q3 - \cancel{Q1} = 10$$

------------------------ *Adding 1& 2*

$$2Q3 = 110$$
$$Q3 = \frac{110}{2}$$
$$Q3 = 55$$

From eq. (1)

$$Q3 + Q1 = 100$$
$$55 + Q1 = 100$$
$$Q1 = 45$$

Merits & Demerits of Standard Deviation:

Merits:

- It is rigidly defined.
- It is based on all observations.
- It is capable of further mathematical treatment.
- It is not very much affected by sampling fluctuations.

Demerits:

- It is difficult to understand.
- It gives undue weightage for extreme values.
- It cannot be calculated for open end classes.

From the following data

	Group I	Group II
No. of Persons	150	200
Average Daily wages	Rs 200	Rs. 500
S.D	15	25

i) Which group has larger wage bill?
ii) Find combine mean and s.d
iii) Which group has more variability?

Example:

There are two groups of children's having 50 and 70 children's respectively. The arithmetic mean of weights of children's in two groups are 30 kgs and 40 kgs with the standard deviations 6 kgs and 5 kgs respectively. Find the combine mean and standard deviation of the entire group of 120 childrens.

	Group I	Group II
No. of Childrens	50	70
A.M	30 kgs	40 kgs
S.D	6	5

Given: $n_1 = 50,$ $n_2 = 70$ $\overline{x_1} = 30,$

$\overline{x_2} = 40$ $\sigma_1 = 16,$ $\sigma_2 = 5$

$\sigma_1{}^2 = 36$ $\sigma_2{}^2 = 25$

$\bar{x}$= combine Mean = $\frac{n_1\overline{x_1}+n_2\overline{x_2}}{n_1+n_2}$

$\bar{x}$=$\frac{50\times30+70\times40}{50+70}$=35.83

$d1 = \overline{x_1} - \bar{x}$ =30 – 35.83 = -5.83 $\therefore d_1{}^2 = 33.99$

$d2 = \overline{x_2} - \bar{x}$ = 40 – 35.83 =4.17 $\therefore d_2{}^2 = 17.39$

$$s.d\ \sigma = \frac{\sqrt{n_1(\sigma_1{}^2 + d12\ \ ) + n_2(\sigma_2{}^2 + d22)}}{n1 + n2}$$

$$s.d\ \sigma = \frac{\sqrt{50(36 + 33.99\ \ ) + 70(25 + 17.39)}}{50 + 70}$$

$$s.d\ \sigma = \frac{\sqrt{50(69.99\ \ ) + 70(42.39)}}{120}$$

$$s.d\ \sigma = \frac{\sqrt{3499.5 + 2967.3}}{120}$$

THANK YOU