

Chapter 6 - Measures Of Dispersion

Question 1:

A measure of dispersion is a good supplement to the central value in understanding a frequency distribution. Comment.

Answer:

The study of the averages is only one sided distribution story. In order to understand the frequency distribution fully, it is essential to study the variability of the observations. The average measures center of the data whereas the quantum of the variation is measured by the measures of dispersion like range, quartile deviation, mean deviation and Standard Deviation. For example, if a country has very high income group people and very low income group people, then we can say that the country has large income disparity.

Question 2:

Which measure of dispersion is the best and how?

Answer:

Standard Deviation is the best measure of dispersion as it satisfies the most essentials of the good measure of dispersion. The following points make Standard Deviation the best measure of dispersion:

1. Most of the statistical theory is based on Standard Deviation. It helps to make comparison between variability of two or more sets of data. Also, Standard Deviation helps in testing the significance of random samples and in regression and correlation analysis.
2. It is based on the values of all the observations. In other words, Standard Deviation makes use of every item in a particular distribution.
3. Standard Deviation has a precise value and is a well-defined and definite measure of dispersion. That is, it is rigidly defined.
4. It is independent of the origin.
5. It is widely used measure of dispersion as all data distribution is nearer to the normal distribution.
6. It enables algebraic treatment. It has correct mathematical processes in comparison to range, quartile deviation and mean deviation.

Question 3:

Some measures of dispersion depend upon the spread of values whereas some calculate the variation of values from a central value. Do you agree?

Answer:

Yes, it is true that some measures of dispersion depend upon the spread of values, whereas some calculate the variation of values from the central value. The spread of values is determined by the absolute measures of dispersion like Range, Quartile Mean Deviation, and Standard Deviation. These measures express dispersion in terms of original unit of the series and it cannot be used for the comparison of statistical data having different units. On the other hand, the relative measures of the dispersion calculate the variability of the values from a central value. The relative measure includes coefficient of Range, Mean Deviation and Variation. It is used when the comparison has to be made between two statistical sets. These measures are free from any units.

Question 4:

In a town, 25% of the persons earned more than Rs 45,000 whereas 75% earned more than 18,000. Calculate the absolute and relative values of dispersion.

Answer:

Absolute Value of Dispersion

$$\begin{aligned}\text{Range} &= L - S \\ &= 45,000 - 18,000 = \text{Rs } 27,000\end{aligned}$$

Relative Value of Dispersion

$$\begin{aligned}\text{Coefficient of Range} &= \frac{L - S}{L + S} \\ &= \frac{45000 - 18000}{45000 + 18000} \\ &= 0.428\end{aligned}$$

Question 5:

The yield of wheat and rice per acre for 10 districts of a state is as under:

District	1	2	3	4	5	6	7	8
Wheat	12	10	15	19	21	16	18	9
Rice	22	29	12	23	18	15	12	34

Calculate for each crop,

- (i) Range
- (ii) Q.D.
- (iii) Mean Deviation about Mean
- (iv) Mean Deviation about Median
- (v) Standard Deviation
- (vi) Which crop has greater variation?
- (vii) Compare the value of different measures for each crop.

Answer:

(i) Range

a. Wheat

Highest value of distribution (H) = 25

Lowest value of distribution (L) = 9

$$\begin{aligned}
 \text{Range} &= H - L \\
 &= 25 - 9 \\
 &= 16
 \end{aligned}$$

b. Rice

Highest value of distribution (H) = 34

Lowest value of distribution (L) = 12

$$\begin{aligned}
 \text{Range} &= H - L \\
 &= 34 - 12 \\
 &= 22
 \end{aligned}$$

(ii) Quartile Deviation

a. Wheat

Arranging the production of wheat in increasing order

9, 10, 10, 12, 15, 16, 18, 19, 21, 25

$$\begin{aligned}
 Q_1 &= \frac{N+1}{4} \text{th item} \\
 &= \frac{10+1}{4} \text{th item} \\
 &= \frac{11}{4} \text{th item} \\
 &= 2.75^{\text{th}} \text{ item} \\
 &= \text{Size of 2th item} + 0.75 (\text{size of 3rd item} - \text{size of 2nd item}) \\
 &= 10 + 0.75 (10 - 10) \\
 &= 10 + 0.75 \times 0 \\
 &= 10
 \end{aligned}$$

$$\begin{aligned}
 Q_3 &= \frac{3(N+1)}{4} \text{th item} \\
 &= \frac{3(10+1)}{4} \text{th item} \\
 &= \frac{33}{4} \text{th item} \\
 &= 8.25^{\text{th}} \\
 &= \text{Size of 8th item} + 0.25 (\text{size of 9th item} - \text{size of 8th item}) \\
 &= 19 + 0.25 (21 - 19) \\
 &= 19 + 0.25 \times 2 \\
 &= 19 + 0.50 = 19.50
 \end{aligned}$$

$$\begin{aligned}
 \text{Quartile Deviation} &= \frac{Q_3 - Q_1}{2} \\
 &= \frac{19.50 - 10}{2} \\
 &= \frac{9.50}{2} \\
 &= 4.75
 \end{aligned}$$

b. Rice

Arranging the data of production of rice

12, 12, 12, 15, 18, 18, 22, 23, 29, 34

$$Q_1 = \frac{N+1}{4} \text{th item}$$

$$= \left(\frac{10+1}{4} \right) \text{th item}$$

$$= 2.75 \text{th item}$$

$$= \text{size of } 2^{\text{nd}} \text{ item} + 0.75 (\text{size of } 3^{\text{rd}} \text{ item} - \text{size of } 2^{\text{nd}} \text{ item})$$

$$= 12 + 0.75 (12 - 12)$$

$$= 12 + 0.75 \times 0$$

$$= 12$$

$$Q_3 = \frac{3(N+1)}{4} \text{th item}$$

$$= \frac{3(10+1)}{4} \text{th item}$$

$$= \frac{33}{4} \text{th item}$$

$$= 8.25 \text{th item}$$

$$= \text{Size of } 8^{\text{th}} \text{ item} + 0.25 (\text{size of } 9^{\text{th}} \text{ item} - \text{size of } 8^{\text{th}} \text{ item})$$

$$= 23 + 0.25 (29 - 23)$$

$$= 23 + 0.25 \times 6$$

$$= 23 + 1.5$$

$$= 24.5$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{24.5 - 12}{2}$$

$$= \frac{12.50}{2}$$

$$= 6.25$$

(iii) Mean Deviation about Mean

a. Wheat

Wheat Production x	$ d = x - \bar{A\bar{x}} $ $\bar{A\bar{x}} = 15$
9	6
10	5
10	5
12	3
15	0
16	1
18	3
19	4

21	6
25	10
$\sum x = 155$	$\sum d = 43$

$$\text{Mean} = \frac{\sum x}{N} = \frac{155}{10} = 15.5$$

Mean Deviation from Mean

$$\begin{aligned} MD(\bar{X}) &= \frac{\sum |d| + (\bar{x} - A\bar{x})(\sum f_B - \sum f_A)}{n} \\ &= \frac{43 + (15.5 - 15)(5 - 5)}{10} \\ &= \frac{43}{10} = 4.3 \end{aligned}$$

b. Rice

Rice Production x	$ d = x - A\bar{x} $ $A\bar{x} = 18$
12	6
12	6
12	6
15	3
18	0
18	0
22	4
23	5
29	11
34	16
$\sum x = 195$	$\sum d = 57$

(iv) Mean Deviation about Median

a. Wheat

Production of Wheat	$ d = x - 15 $
------------------------	------------------

x	
09	6
10	5
10	5
12	3
15	0
16	1
18	3
19	4
21	6
25	10
	$\sum d = 43$

$$\text{Median} = \text{size of } \left(\frac{N+1}{2} \right) \text{th item}$$

$$= \text{size of } \left(\frac{10+1}{2} \right) \text{th item}$$

$$= \text{size of } (5.5) \text{th item} = \frac{\text{size of 5}^{\text{th}} \text{ item} + \text{size of 6}^{\text{th}} \text{ item}}{2} = \frac{15+16}{2} = 15.5$$

$$\text{MD}_{\text{Median}} = \frac{\sum |d| + (\text{Median} - A)(\sum f_B - \sum f_A)}{n}$$

$$= \frac{57 + (18-18)(6-4)}{10}$$

$$= \frac{57}{10} = 5.7$$

b. Rice

Production of rice	$d = X-18$
12	6
12	6
12	6
15	3
18	0

18	0
22	4
23	5
29	11
34	16
	$\sum d = 57$

Since n is even

Therefore, Median = size of $\left(\frac{N+1}{2}\right)$ th item

= size of $\left(\frac{10+1}{2}\right)$ th item

\Rightarrow size of (5.5) th item = $\frac{\text{size of 5th item} + \text{size of 6th item}}{2} = \frac{36}{2} = 18$

$$MD_{\text{Median}} = \frac{\sum |d| + (\text{Median} - A)(\sum f_B - \sum f_A)}{n}$$

$$= \frac{57 + (18 - 18)(6 - 4)}{10}$$

$$= \frac{57}{10} = 5.7$$

(v) Standard Deviation

a. Wheat

Reduction of Wheat x	$Ax = 15$ $d = x - Ax$	d^2
9	-6	36
10	-5	25
10	-5	25
12	-3	9
15	0	0
16	1	1
18	3	9
19	4	16
21	6	36
25	10	100

	$\sum d = 5$	$\sum d^2 = 257$
--	--------------	------------------

$$\begin{aligned}
 \sigma &= \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} \\
 &= \sqrt{\frac{257}{10} - \left(\frac{5}{10}\right)^2} \\
 &= \sqrt{\frac{257}{10} - \frac{25}{100}} \\
 &= \sqrt{\frac{2570 - 25}{100}} \\
 &= 5.04
 \end{aligned}$$

b. Rice

Production of rice x	$Ax = 18$ $d = x - Ax$	d^2
12	-6	36
12	-6	36
12	-6	36
15	-3	9
18	0	0
22	4	16
23	5	25
29	11	121
34	16	256
	$\sum d = 15$	$\sum d^2 = 535$

$$\begin{aligned}
 \sigma &= \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} \\
 &= \sqrt{\frac{535}{10} - \left(\frac{15}{10}\right)^2} \\
 &= \sqrt{\frac{535}{10} - \frac{225}{100}} \\
 &= \sqrt{51.25} \\
 &= 7.16
 \end{aligned}$$

(vi) Coefficient of Variation

a. Wheat

$$\begin{aligned} CV &= \frac{\sigma}{\bar{X}} \times 100 \\ &= \frac{5.04}{15.5} \times 100 \\ &= 32.51 \end{aligned}$$

b. Rice

$$\begin{aligned} CV &= \frac{\sigma}{\bar{X}} \times 100 \\ &= \frac{7.16}{19.5} \times 100 \\ &= 36.71 \end{aligned}$$

(vi) Rice crop has greater variation as the coefficient of variation is higher for rice as compared to that of wheat.

(vii) Rice crop has higher Range, Quartile Deviation, Mean Deviation about mean, Mean Deviation about median, Standard Deviation and Coefficient of Variation.

Question 6:

In the previous question, calculate the relative measures of variation and indicate the value which, in your opinion, is more reliable.

Answer:

(i) Coefficient of Range

a) Wheat

$$\begin{aligned} \text{Coefficient of Range} &= \frac{H - L}{H + L} \\ &= \frac{25 - 9}{25 + 9} \\ &= \frac{16}{34} = 0.47 \end{aligned}$$

b) Rice

$$\begin{aligned} \text{Coefficient of Range} &= \frac{H - L}{H + L} \\ &= \frac{34 - 12}{34 + 12} \\ &= \frac{22}{46} = 0.48 \end{aligned}$$

(ii) Coefficient of Quartile Deviation

a) Wheat

$$\begin{aligned}\text{Coefficient of Quartile Deviation} &= \frac{Q_3 - Q_1}{Q_3 + Q_1} \\ &= \frac{19.50 - 10}{19.50 + 10} \\ &= 0.32\end{aligned}$$

b) Rice

$$Q_1 = 12, Q_3 = 24.5$$

$$\begin{aligned}\text{Coefficient of Quartile Deviation} &= \frac{Q_3 - Q_1}{Q_3 + Q_1} \\ &= \frac{24.5 - 12}{24.5 + 12} \\ &= 0.34\end{aligned}$$

(iii) Coefficient of Mean Deviation from mean

a) Wheat

$$\begin{aligned}\text{Coefficient of } MD_x &= \frac{MD_x}{\bar{X}} \\ &= \frac{4.3}{15.5} \\ &= 0.28\end{aligned}$$

$$\begin{aligned}\text{Coefficient of } MD_M &= \frac{MD_M}{M} \\ &= \frac{4.3}{15} = 0.29\end{aligned}$$

b) Rice

$$\begin{aligned}\text{Coefficient of } MD_x &= \frac{MD_x}{\bar{X}} \\ &= \frac{6}{19.5} \\ &= 0.31\end{aligned}$$

$$\begin{aligned}\text{Coefficient of } MD_M &= \frac{MD_M}{M} \\ &= \frac{5.7}{18} = 0.32\end{aligned}$$

(iv) Coefficient of Variation

a) Wheat

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{5.04}{15.5} \times 100 = 32.51$$

b) Rice

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{7.16}{19.5} \times 100 = 36.71$$

The coefficient of variation is more reliable than all other measures.

Question 7:

A batsman is to be selected for a cricket team. The choice is between X and Y on the basis of their scores in five previous tests which are:

X	25	85	40	80	120
Y	50	70	65	45	80

Which batsman should be selected if we want,

- (i) a higher run getter, or
- (ii) a more reliable batsman in the team?

Answer:

Batsman X

X	$X - \bar{X} =$ x $X - 70$	x^2
25	- 45	2025
85	+ 15	225
40	- 30	900
80	10	100
120	50	2500
$\sum X = 350$		$\sum x^2 = 5750$

$$\bar{X} = \frac{\sum X}{N} = \frac{350}{5} = 70$$

$$\sigma = \sqrt{\frac{\sum x^2}{N}} = \sqrt{\frac{5750}{5}}$$

$$= 33.91$$

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{33.91}{70} \times 100 = 48.44$$

Batsman Y

Y	$Y - \bar{Y} = y$	y^2
50	-12	144
70	8	64
65	3	9
45	-17	289
80	18	324
$\sum Y = 310$		$\sum y^2 = 830$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{310}{5} = 62$$

$$\sigma = \sqrt{\frac{\sum y^2}{N}} = \sqrt{\frac{830}{5}}$$

$$= 12.88$$

$$CV = \frac{\sigma}{\bar{Y}} \times 100$$

$$= \frac{12.88}{62} \times 100$$

$$= 20.78$$

(i) Average of Batsman X is higher than that of Batsman Y , so he should be selected if we want to score higher run.

(ii) The Batsman X is more reliable than Batsman Y . This is because the coefficient of variation of Batsman X is higher than that of Batsman Y .

Question 8:

To check the quality of two brands of light bulbs, their life in burning hours was estimated as under for 100 bulbs of each brand.

Life (in hrs)	No. of bulbs	
	Brand A	Brand B
0 – 50	15	2
50 – 100	20	8
100 – 150	18	60
150 – 200	25	25
200 – 250	22	5
	100	100

(i) Which brand gives higher life?

(ii) Which brand is more dependable?

Answer:

For Brand A

Life (in hours)	No. of bulbs	M	$A = 125$	$d = \frac{x-A}{i}$	d^2	fd	fd^2
	f	x	$X - A = d$				
0 – 50	15	25	– 100	– 2	4	– 30	60
50 – 100	20	75	– 50	– 1	1	– 20	20
100 – 150	18	125	0	0	0	0	0
150 – 200	25	175	50	1	1	25	25
200 –	22	225	100	2	4	44	88

250							
	$\sum f = 100$					$\sum fd' = 19$	$\sum fd'^2 = 193$

$$\bar{X} = A + \frac{\sum fd'}{\sum f} \times i$$

$$= 125 + \frac{19}{100} \times 50$$

$$= 134.5$$

$$\sigma_x = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f} \right)^2 \times i}$$

$$= \sqrt{\frac{193}{100} - \left(\frac{19}{100} \right)^2 \times 50}$$

$$= \sqrt{\frac{193}{100} - \frac{361}{10000}} \times 50$$

$$= \sqrt{\frac{18939}{10000}} \times 50$$

$$= 68.5$$

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{68.5}{134.5} \times 100$$

$$= 50.93$$

For Brand B

Life (in hrs)	No. of bulbs (<i>f</i>)	Mid valuem	<i>A</i> = 125	$d' = \frac{X - A}{c}$	d'^2	fd'	fd'^2
0 – 50	2	25	– 100	– 2	4	– 4	8
50 – 100	8	75	– 50	– 1	1	– 8	8
100 – 150	60	125	0	0	0	0	0
150 –	25	175	50	1	1	25	25

200							
200	5	225	100	2	4	10	20
–							
250							
	$\sum f = 100$					$\sum fd' = 23$	$\sum fd'^2 = 61$

$$\bar{X} = A + \frac{\sum fd'}{\sum f} \times i$$

$$= 125 + \frac{23}{100} \times 50$$

$$= 136.5$$

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times i$$

$$= \sqrt{\frac{61}{100} - \left(\frac{23}{100}\right)^2} \times 50$$

$$= \sqrt{\frac{61}{100} - \frac{529}{10000}} \times 50$$

$$= \sqrt{\frac{5571}{10000}} \times 50$$

$$= 37.5$$

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{37.5}{136.5} \times 100$$

$$= 27.47$$

- (i) The average life of bulb of Brand B has comparatively higher life than the bulb of Brand A.
(ii) The bulbs of Brand B is more dependable as CV of Brand B is lesser than CV of Brand A.

Question 9:

Average daily wage of 50 workers of a factory was Rs 200 with a Standard Deviation of Rs 40. Each worker is given a raise of Rs 20. What is the new average daily wage and Standard Deviation? Have the wages become more or less uniform?

Answer:

$$N = 50$$

$$\bar{X} = 200$$

$$s = 40$$

$$\text{Average Wage} = \frac{\text{Total Wages}}{\text{No. of Workers}}$$

$$200 = \frac{\text{Total Wages}}{50}$$

$$\text{So, Total Wages} = 200 \times 50$$

$$= \text{Rs } 10,000$$

$$\text{Now, increased wage rate} = \text{Rs } 20$$

$$\text{Total raise} = 50 \times 20 = \text{Rs } 1,000$$

$$\text{Total Wage after raise} = 10,000 + 1,000$$

$$= \text{Rs } 11,000$$

$$\text{New Average Wage} = \frac{\text{New Total Wages}}{\text{No. of Workers}}$$

$$= \frac{11,000}{50}$$

$$= \text{Rs } 220$$

$$\text{Initial Standard Deviation} = \text{Rs } 40$$

$$\text{So, New Standard Deviation} = \text{Rs } 40 + \text{Rs } 20$$

$$= \text{Rs } 60$$

Note: New Standard Deviation will rise by the same amount as the wage of each worker has increased.

Question 10:

If in the previous question, each worker is given a hike of 10% in wages, how are the Mean and Standard Deviation values affected?

Answer:

$$\text{Average wage} = \text{Rs } 200$$

$$\text{Hike in wages} = 10\% \text{ of Rs } 200$$

$$= \frac{10}{100} \times 200$$

$$= \text{Rs } 20$$

$$\text{Individual raise given to each worker} = \text{Rs } 20$$

$$\text{Total raise in wage} = 50 \times 20 = \text{Rs } 1,000$$

$$\text{New Total Wage} = \text{Rs } 10,000 + \text{Rs } 1,000$$

$$= \text{Rs } 11,000$$

$$\text{New Average Wage} = \frac{\text{New Total Wage}}{\text{No. of Workers}}$$

$$= \frac{11,000}{50} = \text{Rs } 220$$

$$\text{Initial Standard Deviation} = \text{Rs } 40$$

$$\text{So, New Standard Deviation} = \text{Rs } 40 + 20$$

$$= \text{Rs } 60$$

Question 11:

Calculate the Mean Deviation using mean and Standard Deviation for the following distribution.

Classes	Frequencies
20 – 40	3
40 – 80	6
80 – 100	20
100 – 120	12
120 – 140	9
Total	50

Answer:

Classe s	Frequenc y	m	$A = 90$ $d = X - A$	$ d $	$f d $	$d' = \frac{X - A}{i}$	fd'	d'^2	fd'^2
-------------	---------------	-----	-------------------------	-------	--------	------------------------	-------	--------	---------

20 – 40	3	30	– 60	60	180	– 6	– 18	36	108
40 – 80	6	60	– 30	30	180	– 3	– 18	9	54
80 – 100	20	90	0	0	0	0	0	0	0
100 – 120	12	110	20	20	240	2	24	4	48
120 – 140	9	130	40	40	360	4	36	16	144
	$\sum f = 50$				$\sum f d = 960$		$\sum fd' = 24$		$\sum fd'^2 = 354$

$$\bar{X} = A + \frac{\sum fd'}{\sum f} \times i$$

$$= 90 + \frac{24}{50} \times 10$$

$$= 94.8$$

$$MD_{\bar{X}} = \frac{\sum f|d|}{\sum f}$$

$$= \frac{960}{50} = 19.2$$

$$\sigma = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f} \right)^2} \times i$$

$$= \sqrt{\frac{354}{50} - \left(\frac{24}{50} \right)^2} \times 10$$

$$= \sqrt{\frac{354}{50} - \frac{576}{2500}} \times 10$$

$$= \sqrt{6.85} \times 10$$

$$= 26.17$$

Mean Deviation from mean

$$MD_{\sigma} = \frac{\sum f|D|}{N}$$

$$= \frac{998.4}{50} = 19.968$$

Question 12:

The sum of 10 values is 100 and the sum of their squares is 1090. Find out the Coefficient of Variation.

Answer:

$$\sum X = 100$$

$$N = 10$$

$$\text{So, } \bar{X} = \frac{\sum X}{N} = \frac{100}{10} = 10$$

$$\sum X^2 = 1090$$

$$\sigma = \sqrt{\frac{\sum X^2}{N} - (\bar{X})^2}$$

$$= \sqrt{\frac{1090}{10} - (10)^2}$$

$$= \sqrt{9}$$

$$= 3$$

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{3}{10} \times 100 = 30$$