Measures of Variability

Variability

Variability means "Scatter" or "Spread". Thus measures of variability refer to the scatter or spread of scores around their central tendency. The measures of variability indicate how the distribution scatter above and below the central tender.

Definitions of Variability:

Dictionary of Education- C.V. Good, **"The scatter or variability of the observations of a distribution about some measure of central tendency."**

Collins Dictionary of Statistics: **"Dispersion is the spread of a distribution"**

Need of Variability:

1. Helps to as-certain the measures of deviation:
The measures of variability help us to measure the degree of deviation, which exist in the data. By that can determine the limits within which the data will navy in some measureable variety or quality.

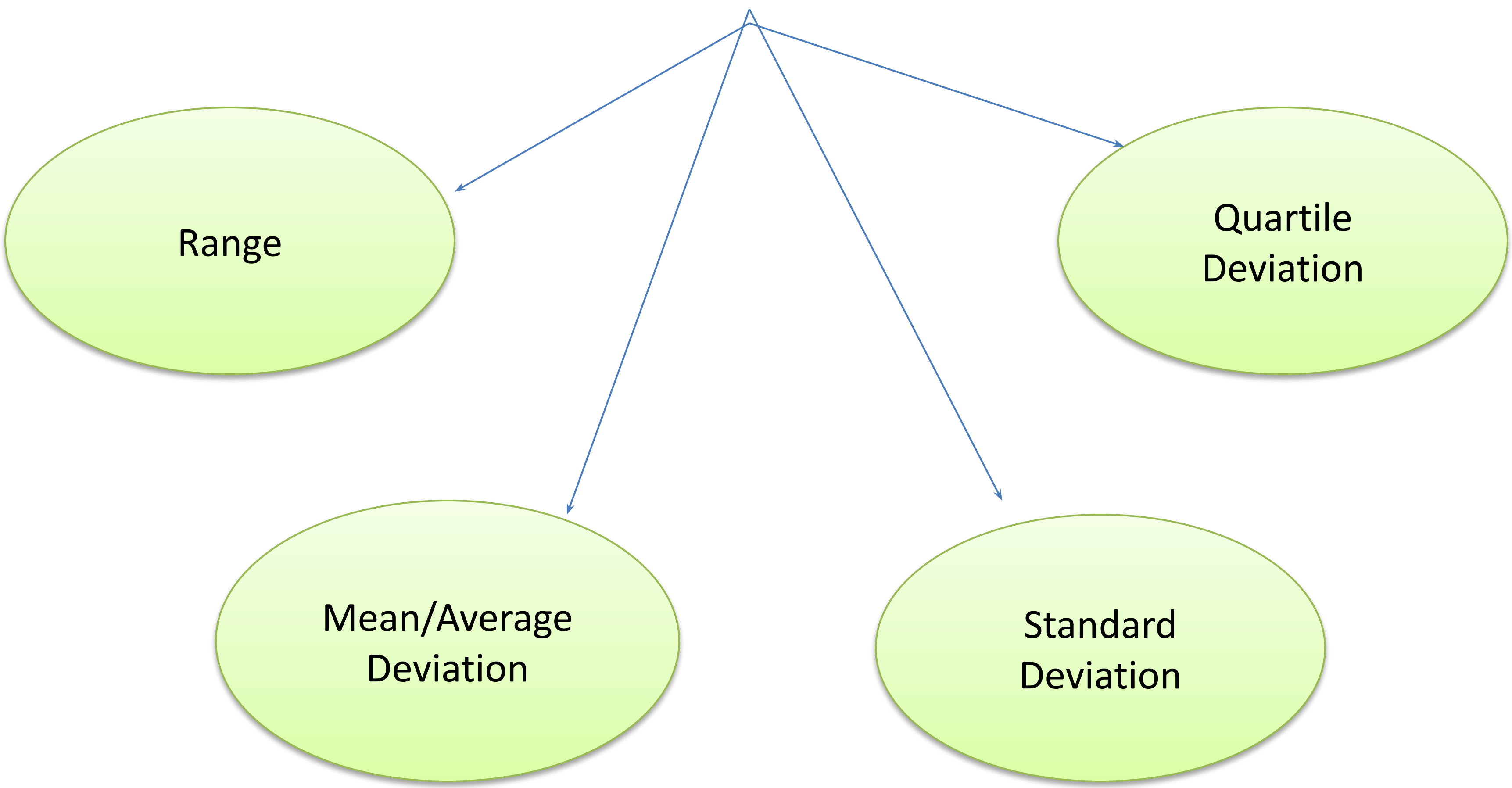
2. It helps to compare different group:
With the help of measures of validity we can compare the original data expressed in different units.

3. It is useful to supplement the information provided by the measures of central tendency.

4. It is useful to calculate further advance statistics based on the measures of dispersion.

Measures of Variability:

There are four measures of variability:



Range:

Range is the difference between in a series. It is the most general measure of spread or scatter. It is a measure of variability of the varieties or observation among themselves and does not given an idea about the spread of the observations around some central value.

Range = H—L

Here H = Highest score

L = Lowest score

Example:

In a class, 20 students have secured the marks as following:

22, 48, 43, 60, 55, 25, 15, 45, 35, 68, 50, 70, 35, 40, 42, 48, 53, 44, 55, 52

Here—The Highest score is 70

The Lowest score is 15

Range = H — L = 70 –15 = 55

Merits of Range:

1. Range is easily calculated and readily understood.
2. It is the simplest measure of variability.
3. It provides a quick estimate of the measure of variability.

Demerits of Range:

1. Range is greatly affected by fluctuation of scores.
2. It is not based on all the observations of the series. It only takes the highest and the lowest scores in to account.
3. In case of open ended distributions range cannot be used.
4. It is affected greatly by fluctuations in sampling.
5. It is affected greatly by extreme scores.
6. The series is not truly represented by range. A symmetrical and A symmetrical distribution may have same range but not the same dispersion.

Uses of Range:

- Range is used as a measure of dispersion when variations in the value of the variable are not much.
- Range is the best measure of variability when the data are too scattered or too scant.
- Range is used when the knowledge of extreme score or total spread is wanted.
- When a quick estimate of variability is wanted range is used.

The Average/Mean Deviation (A.D.):

"Average deviation is the arithmetic mean of all the deviations of different scores from the mean value of the scores without the regard for sign of the deviation."

Definitions:

Collins Dictionary of Statistics:

"Average deviation is the mean of the absolute values of the differences between the values of a variable and the mean of its distribution."

Dictionary of Education, C.V. Good:

"A measure expressing the average amount by which the individual items in a distribution deviate from a measure of central tendency such as mean of median."

Computation of Average Deviation:

There are two situations for computing average deviation:

(a) When data are ungrouped.

(b) When data are grouped.

Computation of AD from ungrouped data.

Example: 10, 15, 10, 20, 25, 15, 25, 20, 17, 23.

प्राप्तांक X	विचलन $(X - M)$	विचलन 'd'
10	10 – 18	= –8
15	15 – 18	= –3
10	10 – 18	= –8
20	20 – 18	= +2
25	25 – 18	= +7
15	15 – 18	= –3
25	25 – 18	= +7
20	20 – 18	= +2
17	17 – 18	= +1
23	23 – 18	= +5
$\Sigma X = 180$ $M = 18$		Σ 'd' = 46 (= +23 – 23)

$$\text{M.D.} = \frac{\sum 'd'}{N}$$

$$\text{M.D.} = \frac{46}{10} = 4.60$$

Calculate AD by Long Method from Grouped Data:
Example:

$$M.D. = \frac{\Sigma fd'}{N} \text{ या } \frac{\Sigma |fx|}{N}$$

प्राप्तांक	f	X	fX	'd' $\lvert X - M \rvert$	f 'd' ($f \times$ 'd')
40 – 45	3	42	126	14.3	42.9
35 – 40	4	37	148	9.3	37.2
30 – 35	4	32	128	4.3	17.2
25 – 30	7	27	189	0.7	4.9
20 – 25	5	22	110	5.7	28.5
15 – 20	3	17	51	10.7	32.1
10 – 15	2	12	24	15.7	31.4
	N = 28		$\Sigma fX = 776$		$\Sigma f'd' = 194.2$

$$M = \frac{\Sigma fX}{N} = \frac{776}{28} = 27.7$$

$$M.D. = \frac{\Sigma f'd'}{N} = \frac{194.2}{28} = 6.93$$

Calculate M.D. by Short Method from Grouped Data:

Example:

प्राप्तांक	f	x'	fx'	X	d' $(X-M)$	fd' $(f \times d')$
40 – 45	3	+3	+9	42	14.3	42.9
35 – 40	4	+2	+8	37	9.3	37.2
30 – 35	4	+1	+4	32	4.3	17.2
....	....	....	= (+21)			
25 – 30	7	0	0	27	0.7	4.9
....	....	....				
20 – 25	5	–1	–5	22	5.7	28.5
15 – 20	3	–2	–6	17	10.7	32.1
10 – 15	2	–3	–6	12	15.7	31.4
			= (–17)			
	N = 28		$\Sigma fx' = 4$			$\Sigma fd' = 194.2$

$$\text{Mean (M)} = \text{A.M.} + \frac{\sum fx'}{N} \times \text{C.I.}$$

$$= 27 + \frac{4}{28} \times 5 = 27.7$$

$$\text{M.D.} = \frac{\sum fd'}{N} = \frac{194.2}{28} = 6.93$$

Merits of A.D.:

- Average deviation is rigidly defined and its value is precise and definite.
- It is easy to calculate.
- It is easy to understand. Because it is the average of the deviations from a measure of central tendency.
- It is based on all the observations.
- It is less affected by the value of extreme scores.

Demerits of A.D.:

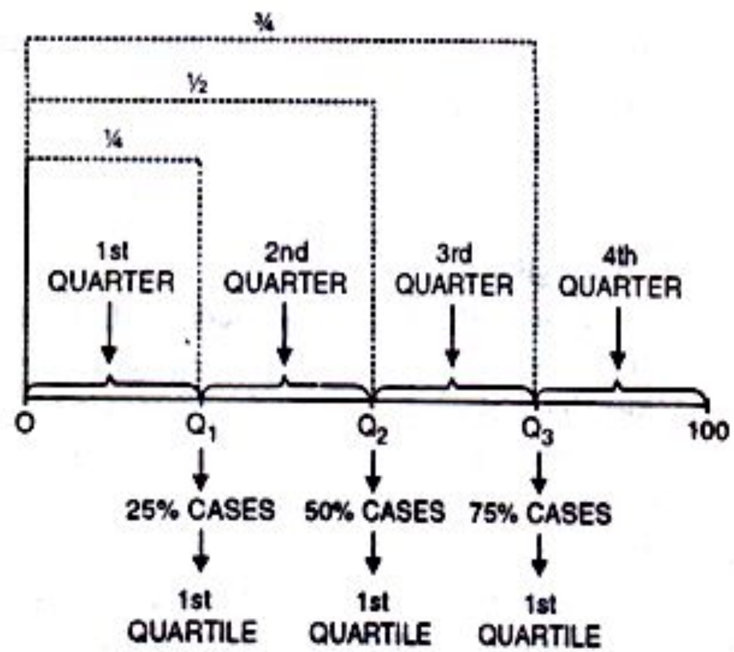
- The most serious drawback with average deviation is that it ignores the algebraic signs of the deviations which is against the fundamental rules of mathematics.
- Further algebraic treatment is not possible in case of AD.
- It is very rarely used. Because of standard deviation is generally used as a measure of dispersion.
- When calculated from mode AD does not give accurate measure of dispersion.

Uses of Average Deviation:

- Average deviation is used when it is desired to weight all the deviations from the mean according to their size.
- When extreme scores influence standard deviation at that time AD is the best measure of dispersion.
- AD is used when we want to know the extent to which the measures are spread out either side of the mean.

The Quartile Deviation (QD):

"The quartile deviation or Q is the one-half the scale distance between the 75th and 25th percentiles in a frequency distribution."



Calculate the QD of following data:

प्राप्तांक (Scores)	आवृत्ति (f)	सं. आ. (Cf)
40 – 44	3	28
35 – 39	3	25
30 – 34	5	22
25 – 29	7	17
20 – 24	5	10
15 – 19	3	5
10 – 14	2	2
	N = 28	

हल : $Q = \frac{Q_3 - Q_1}{2}$

अतः यहाँ Q_3 एवं Q_1 का अलग-अलग मान ज्ञात करेंगे—

$$Q_3 = lm + \left(\frac{3N/4 - Cf}{fm}\right) \times i$$

यहाँ, $lm = 29.5$, $N = 28$, $Cf = 17$, $fm = 5$, $i = 5$

$$Q_3 = 29.5 + \left(\frac{3 \times 28/4 - 17}{5}\right) \times 5$$

$$= 29.5 + 4 = 33.5$$

$$Q_1 = lm + \left(\frac{N/4 - Cf}{fm}\right) \times i$$

यहाँ, $lm = 19.5, N = 28, Cf = 5, fm = 5, i = 5$

$$Q_1 = 19.5 + \left(\frac{28/4 - 5}{5}\right) \times 5$$

$$= 19.5 + 2 = 21.5$$

$Q_1 = 21.5$

$$Q = \frac{Q_3 - Q_1}{2}$$

यहाँ, $Q_3 = 33.5$ एवं $Q_1 = 21.5$

$$\therefore \quad Q = \frac{33.5 - 21.5}{2} = \frac{12}{2} = 6.0$$

उत्तर : $Q = 6.0$

Merits of Quartile Deviation:

- Quartile deviation is simple to calculate and easy to understand.
- It is more representative and trust worthy than range. In case of open ended class intervals it is used in studying measures of dispersion.
- In case of open ended class intervals it is used in studying measures of dispersion.
- It is a good index of score density at the middle of the distribution.
- When we take Median as the measure of central tendency at that time Q is preferred as the measure of dispersion.
- Like range it is not affected by extreme scores.

Demerits of Quartile Deviation:

- It is not based on all the observations of data. It ignores the first 25% and the last 25% of the scores.
- Further algebraic treatment is not possible in case of Q. It is only a positional average. It does not study variation of the values of a variable from any average. It merely indicates a distance on a scale.
- It is affected by fluctuation of scores. Its value is affected in any case, by a change in the value of a single score.
- Q is not a suitable measure of dispersion, when in a series there is a considerable variation in the values of various scores.

Uses of Quartile Deviation:

- When Median is the measure of central tendency at that time Q is used is used as the measure of dispersion.
- When extreme scores affect S.D. or the scores are scattered at that time Q is used as measure of variability.
- When our primary interest is to know the concentration around the median-the middle 50% of cases, at that time Q is used.
- When the class intervals are open ended, Q is used as measure of dispersion.

The Standard Deviation (SD):

"Standard deviation is the square root of the average value of the squared deviations of the scores from their arithmetical mean."

Definitions:

Collin's Dictionary of Statistics.

"Standard deviation is a measure of spread or dispersion. It is root mean squared deviation."

Dictionary of Education—C.V. Good.

"A widely used measure of variability, consisting of the square root of the mean of the squared deviations of scores from the mean of the distribution."

Computation of SD from Ungrouped Data:

Example: 10, 15, 10, 20, 25, 15, 25, 20, 17, 23.

प्राप्तांक X	मध्यमान से प्राप्तांकों का विचलन $(X-M) = d$ या x	विचलनों का वर्ग $(X-M)^2 = d^2$ या x^2
10	$10-18 = -8$	$(-8)^2 = 64$
15	$15-18 = -3$	$(-3)^2 = 9$
10	$10-18 = -8$	$(-8)^2 = 64$
20	$20-18 = +2$	$(+2)^2 = 4$
25	$25-18 = +7$	$(+7)^2 = 49$
15	$15-18 = -3$	$(-3)^2 = 9$
25	$25-18 = +7$	$(+7)^2 = 49$
20	$20-18 = +2$	$(+2)^2 = 4$
17	$17-18 = -1$	$(-1)^2 = 1$
23	$23-18 = +5$	$(+5)^2 = 25$
$\Sigma X = 180$		Σx^2 या $\Sigma d^2 = 278$

$$M = \frac{\Sigma X}{N} = \frac{180}{10} = 18$$

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

$$\Sigma x^2 = 278, \ N = 10$$

$$\sigma = \sqrt{\frac{278}{10}} = \sqrt{27.8}$$

$$\sigma = 5.27$$

In grouped data SD can be calculated in two methods:

1. Direct method or Long method
2. Short method or Assumed Mean method

Direct method or Long method: **Example:**

प्राप्तांक	f	म.वि. X	$f \times X$ fX	$X - M$ d	$f \times d$ fd	$fd \times d$ fd^2
40—45	3	42	126	14.3	42.9	613.47
35—40	4	37	148	9.3	37.2	345.96
30—35	4	32	128	4.3	17.2	73.96
25—30	7	27	189	– 0.7	– 4.9	3.43
20—25	5	22	110	– 5.7	–28.5	162.45
15—20	3	17	51	– 10.7	–32.1	343.47
10—15	2	12	24	– 15.7	–31.4	492.98
	N=27		$\Sigma fX = 776$			$\Sigma fd^2 = 2035.72$

$$M = \frac{\Sigma fX}{N} = \frac{776}{28}$$

$$M = 27.71$$

$$\text{S.D.}(\sigma) = \sqrt{\frac{\Sigma fd^2}{N}}$$

$$\Sigma fd^2 = 2035.72, \; N = 28$$

$$\text{S.D.}(\sigma) = \sqrt{\frac{2035.72}{28}}$$

$$= \sqrt{72.70}$$

$$= 8.52$$

Merits of S.D:

- Standard deviation is rigidly defined and its value is always definite.
- It is based on all the observations of data.
- It is capable of further algebraic treatment and possesses many mathematical properties.
- Unlike Q and AD it is less affected by fluctuations of scores.
- Unlike AD, it does not ignore the negative signs. By squaring of deviations it overcomes these difficulties.
- It is the reliable and most accurate measure of variability. It always goes with the mean which is the most stable measure of central tendency.
- S.D. gives a measure that is comparable meaning from one test to other. Above all the normal curve units are expressed in a unit.

Demerits of S.D:

- S.D. is difficult to understand and not easy to calculate.
- S.D. gives more weight to extreme scores and loss to those which are nearer to the mean. It is because the squares of the deviations, which are big in size, would be proportionately greater than the squares of those deviations which are comparatively small.

Uses of S.D:

- S.D. is used when our thrust is to measure the variability having greatest stability.
- When extreme deviations might affect the variability at that time S.D. is used.
- S.D. is used for calculating the further statistics like coefficient of correlation, standard scores, standard errors, Analysis of Variance, Analysis of Co-variance etc.
- When the interpretation of scores is made in terms of the NPC, S.D is used.
- When we want to determine the reliability and validity of test scores, S.D. is used.

Disclaimer: The content displayed in the PPT has been taken from variety of different websites and book sources. This study material has been created for the academic benefits of the students alone and I do not seek any personal advantage out of it.