Measure of Dispersion

Measure of dispersion: The degree to which the numerical data tends to spread about an average value is called dispersion of data.

Importance of measure of dispersion: The central value like mean is generally used to convey the general behaviour of a data set. In order to understand the frequency distribution fully. It is essential to study the variable of the observations. The average measures ~~of~~ center of the data whereas the quantum of the variation is measured by the measure of dispersion.

Types of measures disperpsion:

There are two types of dispersions.

1. Absolute measure of dispersion
2. Relative measure of dispersion

Absolute measure of dispersion: These measures give us an idea about the amount of dispersion in a set of observations.

Different types of absolute measures are:

(i) Range

(ii) Semi-interquartile or Quartile deviation (Q.D)

(iii) Mean deviation (M.D)

(iv) Variance and Standard deviation

Relative Measure of dispersion: Different types of Relative measure are: (i) coefficient of Range

(ii) Coefficient of quartile deviation

(iii) Coefficient of mean deviation

(iv) coefficient of variation (CV)

Range: Range is the difference ~~between~~ of the highest and the lowest observations of the distributions.

∴ Range = Maximum Values − Minimum Values

⇒ For grouped data, the difference between the highest observations in the last class interval and the lowest observation in the first class interval.

⇒ Range can never be zero.

Example: Find the range of the following data set:
{1, 4, 5, 8, 6, 7, 5, 6, 7, 4, 10, 9, 10}

Maximum Value = 10

Minimum Value = 1

∴ Range = 10 − 1

= 9.

Advantages of Ranges:

① It is easy to understand and to calculate.

② It gives a quick idea about the variability of a set of data.

③ It is the simplest of all measures of dispersion.

Disadvantages of Ranges:

① It is very much affected by extreme values.

② It provides us with an idea of only two extreme values in a set of data.

③ It cannot be computed for data set having open ended class interval.

Mean deviation: ~~The arith~~

The average deviation is the arithmetic mean of the absolutes values ~~is~~ of the deviations from the mean / median / mode.

For ungroup data : $M.D(\bar{x}) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$

For group data : $M.D(\bar{x}) = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{n}$

Mean M.D$(\bar{x})$

About Median:

For ungroup data, $M.D.(Me) = \frac{\sum_{i=1}^{n} |x_i - Me|}{n}$

For group data, $M.D.(Me) = \frac{\sum_{i=1}^{n} f_i |x_i - Me|}{n}$

About mode:

For ungroup data, $M.D.(M_0) = \frac{\sum_{i=1}^{n} |x_i - M_0|}{n}$

For group data, $M.D.(M_0) = \frac{\sum_{i=1}^{n} f_i |x_i - M_0|}{n}$

Example: Mean deviation of 3, 6, 6, 7, 8, 11, 15, 16 (ungroup data)

x_i	$\lvert x_i - \bar{x} \rvert$
3	$\lvert 3-9 \rvert = \lvert -6 \rvert = 6$
6	$\lvert 6-9 \rvert = \lvert -3 \rvert = 3$
6	$\lvert 6-9 \rvert = \lvert -3 \rvert = 3$
7	$\lvert 7-9 \rvert = \lvert -2 \rvert = 2$
8	$\lvert 8-9 \rvert = \lvert -1 \rvert = 1$
11	$\lvert 11-9 \rvert = \lvert 2 \rvert = 2$
15	$\lvert 15-9 \rvert = 6$
16	$\lvert 16-9 \rvert = 7$
	$\sum \lvert x_i - \bar{x} \rvert = 30$

$$\bar{x} = \frac{3+6+6+7+8+11+15+16}{8} = 9$$

Mean deviation, $M.D.(\bar{x}) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$

$$= \frac{30}{8}$$

$$= 3.75.$$

Variance: The variance is the arithmetic mean of the squared deviations from the mean. It is denoted by σ^2.

Ungroup data:

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

Group data:

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

Standard deviation: The positive square root of the variance is called standard deviation. It is denoted by σ.

Ungroup data:

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

Group data:

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

Example-1 Find Variance and standard deviation for the following data set. (Ungroup data)

5, 10, 8, 12, 20, 24, 25, 15, 16, 22.

Solution: We know that,

For ungroup data,

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \Big|\ n = 10$$

$$= \frac{5+10+8+12+20+24+25+15+16+22}{10}$$

$= 15.7.$

x_i	$(x_i-\bar{x})$, $\bar{x}=15.7$	$(x_i-\bar{x})^2$
5	-10.7	114.49
10	-5.7	32.49
8	-7.7	59.29
12	-3.7	13.69
20	4.3	18.49
24	8.3	68.89
25	9.3	86.49
15	-0.7	0.49
16	0.3	0.09
22	0.6	36.69
$\sum_{i=1}^{n} x_i = 157$		$\sum_{i=1}^{n}(x_i-\bar{x})^2 = 434.1$

$\therefore$ Variance, $\sigma^2 = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \dfrac{434.1}{10}$

$= 43.41$

$\therefore$ Standard deviation, $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$

$= \sqrt{43.41}$

$= 6.59$

(Ans)

Example-2 Find Variance and Standard deviation for the following ~~set~~ data set. (Group data)

Class interval	5-10	10-15	15-20	20-25
Frequency	2	5	8	3

Solution:

[f_ix_i এর জন্য একটা ঘর লাগাব]

Class interval	Frequency (f_i)	Mid Value (x_i)	$\cancel{f_ix_i}$ $\bar{x}$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
5-10	2	7.5		-8.33	69.38	138.76
10-15	5	12.5	$\bar{x} = \frac{\sum_{i=1}^{n} f_ix_i}{n}$	-3.33	11.08	55.4
15-20	8	17.5		1.67	2.78	22.24
20-25	3	22.5	$= \frac{285}{18}$	6.67	44.48	133.44
	$n=18$		$=15.83$			$\sum f_i(x_i-\bar{x})^2 = 349.84$

$$\therefore \text{Variance, } \sigma^2 = \frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{n}$$

$$= \frac{349.84}{18}$$

$$= 19.41$$

$$\therefore \text{Standard deviation, } \sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{n}}$$

$$= \sqrt{19.41}$$

$$= 4.40.$$

Properties of Standard deviations:

① The standard deviation is zero if all the observations under study are same.

② Standard Deviation is independent of change of origin but not of scale.

③ For two observations, standard deviation is the half of the range $\left|\frac{x_1-x_2}{2}\right|$

④ It is suitable for further algebraic treatment

⑤ Standard deviation of n natural numbers is $\sqrt{\frac{n^2-1}{12}}$

Coefficient of Variation: The coefficient of variation is the ratio between the standard deviation of a sample and its mean.

i.e, $C.V = \frac{\sigma}{\bar{x}}$

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

Uses of coefficient of variation: Without an understanding of the relative size of the standard deviation compared to the original data, the standard deviation is somewhat meaningless for use with the comparison of data sets.

To address this problem the coefficient of variation is used.

Example-1 A distribution is $\bar{x} = 140$ and $\sigma = 28.28$ and the other is $\bar{x} = 150$ and $\sigma = 24$ which of the two has a greater dispersion?

Solution: Here, $\bar{x} = 140$, $\sigma = 28.28$

$$\therefore C.V._1 = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{28.28}{140} \times 100$$

$$= 20.2\%$$

Again, $\bar{x} = 150$, $\sigma = 24$

$$\therefore C.V._2 = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{24}{150} \times 100$$

$$= 16\%$$

$\therefore$ The 1st distribution has a higher dispersion.

Standard Error of Mean (SEM):

$$SEM = \frac{SD}{\sqrt{n}}$$

Compute Mean deviation, Variance, Standard Deviation (SD), Coefficient Variance and Standard Error of Mean (SEM) from the following data.

Class-Intervals	Frequency
5-10	2
10-15	5
15-20	8
20-25	3

Solution:

Class Intervals	Frequency (f_i)	Mid point (x_i)	$f_i x_i$	$\bar{x}$	$\lvert x_i-\bar{x}\rvert$	$f_i\lvert x_i-\bar{x}\rvert$	$\lvert x_i-\bar{x}\rvert^2$	$f_i(x_i-\bar{x})^2$
5-10	2	7.5	15		8.33	16.66	69.388	138.776
10-15	5	12.5	62.5	$\bar{x}=$	3.33	16.65	11.089	55.45
15-20	8	17.5	140	$\frac{\sum f_i x_i}{n}$	1.67	13.36	2.789	22.32
20-25	3	22.5	67.5	$=\frac{285}{18}$ $=15.83$	6.67	20.01	44.489	133.47
	$n=18$		$\sum_{i=1}^{n} f_i x_i = 285$			$\sum f_i\lvert x_i-\bar{x}\rvert = 66.68$		$\sum f_i(x_i-\bar{x})^2 = 350.016$

$\therefore$ Mean deviation, $M.D.(\bar{x}) = \dfrac{\sum_{i=1}^{n} f_i\lvert x_i-\bar{x}\rvert}{n}$

$$= \frac{66.68}{18}$$

$$= 3.70.$$

$\therefore$ Variance, $\sigma^2 = \dfrac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{n}$

$= \dfrac{350.016}{18}$

$= 19.445$

$\therefore$ Standard deviation (SD), $\sigma = \sqrt{\dfrac{\sum f_i (x_i-\bar{x})^2}{n}}$

$= \sqrt{19.445}$

$= 4.40$

$\therefore$ Coefficient of Variance, $C.V = \dfrac{\sigma}{\bar{x}} \times 100$

$= \dfrac{4.40}{15.83} \times 100\%$

$= 27.79\%$

$\therefore$ Standard error of Mean (SEM) $= \dfrac{SD}{\sqrt{n}}$

$= \dfrac{4.40}{\sqrt{18}}$

$= 1.03.$