MEASURES OF DISPERSION

In summarizing a set of data, it is generally not enough to indicate its average only but also to specify the degree to which numerical data tend to spread about the average. Measures that describe the spread or the dispersion of a data set are called measures of dispersion.

Consider the following series of data
12 8 1 3 16
6 7 8 9 10
The two series have the same arithmetic mean $\bar{X} = 8$. Are they similar?

THE RANGE
The range of a set of data is the difference between the largest and smallest data values in the set.

$$\mathbf{R = X_{max} - X_{min}}$$

THE SEMI-INTERQUARTILE RANGE
The semi-interquartile range, or quartile deviation is defined as half the difference between the third and first quartile. The semi-interquartile range is denoted by I_Q and is given by

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

The semi-interquartile range is common as a measure of dispersion, but the interquartile range ($Q_3 - Q_1$) is sometimes used.

Example: For the following frequency distribution, find the semi-interquartile range

x_i	[20 25[	[25 30[	[30 35[	[35 40[	[40 45[	[45 50[	[50 55[	[55 60[
n_i	4	10	24	34	14	8	4	2

Solution
1- First we find the cumulative frequency

Xi	n_i	$n_i \nearrow$
[20 25[	4	4
[25 30[	10	14
[30 35[	24	38
[35 40[	34	72
[40 45[	14	86
[45 50[	8	94
[50 55[	4	98
[55 60[	2	100
sum	100	/

2 - We calculate the first and the third quartile

$$Q_1 = 30 + \frac{\frac{100}{4} - 14}{24} * 5 = 32.29$$

$$Q_3 = 40 + \frac{\frac{3(100)}{4} - 72}{14} * 5 = 41.07$$

2 - We calculate the semi-interquartile range

$$I_Q = \frac{41.07 - 32.29}{2} = 4.39$$

THE MEAN DEVIATION

The mean deviation (MD), or average deviation, of a set of numbers is defined as the arithmetic mean absolute of deviations from the arithmetic mean $\bar{X}$.

The mean deviation of ungrouped data: The mean deviation is defined by

$$MD = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}$$

Example: find the mean deviation for the following set of numbers

3, 6, 11, 18, 7

solution

The mean $\bar{X} = \frac{45}{5} = 9$

$$MD = \frac{|3 - 9| + |6 - 9| + |7 - 9| + |11 - 9| + |18 - 9|}{5}$$

$$= \frac{6 + 3 + 2 + 2 + 9}{5}$$

$$= 4.4$$

The mean deviation of grouped data: The mean deviation is given by

$$MD = \frac{\sum_{i=1}^{n} n_i |X_i - \bar{X}|}{\sum n_i}$$

Example: Find the mean deviation for the following frequency distribution

x_i	[20 25[	[25 30[	[30 35[	[35 40[	[40 45[	[45 50[	[50 55[	[55 60[
n_i	4	10	24	34	14	8	4	2

Solution

xi	n_i	c_i	$n_i\ c_i$	$\lvert X_i - \bar{X}\rvert$	$n_i \lvert X_i - \bar{X}\rvert$
[20 25[	4	22.5	90	14,7	58,8
[25 30[	10	27.5	275	9,7	97
[30 35[	24	32.5	780	4,7	112,8
[35 40[	34	37.5	1275	0,3	10,2
[40 45[	14	42.5	595	5,3	74,2
[45 50[	8	47.5	380	10,3	82,4
[50 55[	4	52.5	210	15,3	61,2
[55 60[	2	57.5	115	20,3	40,6
SUM	**100**	/	**3720**	/	537,2

$$\bar{X} = \frac{3720}{100} = 37.2$$

$$MD = \frac{537.2}{100} = 5.372$$

THE VARIANCE

The variance of a set of data is defined as the mean square of the deviations from the mean.

THE VARIANCE FOR UNGROUPED DATA The variance of a sample or a population is given by the following formulas

The variance of a sample $s^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$

The variance of a population $\sigma^2 = \frac{\sum(X_i - \mu)^2}{N}$

THE VARIANCE FOR GROUPED DATA The variance of a sample or a population is given by the following formulas

The variance of a sample $s^2 = \frac{\sum n_i(X_i - \bar{X})^2}{n-1}$

The variance of a population $\sigma^2 = \frac{\sum n_i(X_i - \mu)^2}{N}$

THE STANDARD DEVIATION

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

THE STANDARD DEVIATION FOR UNGROUPED DATA The standard deviation of a sample or a population is given by the following formulas

the standard deviation of a sample $s = \sqrt{s^2} = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}$

the standard deviation of a population $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(X_i - \mu)^2}{N}}$

THE STANDARD DEVIATION FOR GROUPED DATA The standard deviation of a sample or a population is given by the following formulas

the standard deviation of a sample $s = \sqrt{s^2} = \sqrt{\frac{\sum n_i(X_i - \bar{X})^2}{n-1}}$

the standard deviation of a population $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum n_i(X_i - \mu)^2}{N}}$

Example: find the variance and the standard deviation for the following set of numbers

3, 6, 11, 18, 7

solution

The mean $\bar{X} = \frac{45}{5} = 9$

the variance

$$S^2 = \frac{|3-9|^2 + |6-9|^2 + |7-9|^2 + |11-9|^2 + |18-9|^2}{5}$$

$$= \frac{36 + 9 + 4 + 4 + 81}{4}$$

$$= 33.5$$

the standard deviation

$s = \sqrt{s^2} = 5.79$

Example: Find the variance and the standard deviation for the following frequency distribution

x_i	[20 25[	[25 30[	[30 35[	[35 40[	[40 45[	[45 50[	[50 55[	[55 60[
n_i	4	10	24	34	14	8	4	2

Solution

xi	n_i	c_i	$n_i\ c_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$n_i(X_i - \bar{X})^2$
[20 25[	4	22.5	90	-14,7	216,09	864,36
[25 30[	10	27.5	275	-9,7	94,09	940,9
[30 35[	24	32.5	780	-4,7	22,09	530,16
[35 40[	34	37.5	1275	0,3	0,09	3,06
[40 45[	14	42.5	595	5,3	28,09	393,26
[45 50[	8	47.5	380	10,3	106,09	848,72
[50 55[	4	52.5	210	15,3	234,09	936,36
[55 60[	2	57.5	115	20,3	412,09	824,18
SUM	**100**	/	**3720**	/		5341

$$\bar{X} = \frac{3720}{100} = 37.2$$

The variance: $s^2 = \frac{\sum n_i(X_i - \bar{X})^2}{n-1}$

$$s^2 = \frac{5341}{99} = 53.95$$

the standard deviation of a sample $s = \sqrt{s^2} = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}$

$$s = \sqrt{53.95} = 7.34$$

SHORT METHODS FOR COMPUTING THE VARIANCE
for ungrouped data

$$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right] = \frac{1}{n-1}\left[\sum x_i^2 - n\bar{x}^2\right]$$

for grouped data

$$s^2 = \frac{1}{n-1}\left[\sum n_i\, x_i^2 - \frac{\left(\sum n_i\, x_i\right)^2}{n}\right] = \frac{1}{n-1}\left[\sum n_i x_i^2 - n\bar{x}^2\right]$$

If the absolute dispersion is the standard deviation **S** and if the average is the mean $\bar{X}$, then the relative dispersion is called the coefficient of variation and is generally expressed as a percentage.

$$CV = \frac{S}{\bar{X}}$$

EMPIRICAL RELATIONS BETWEEN MEASURES OF DISPERSION

For moderately skewed distributions, we have the empirical formulas

Mean deviation = $\frac{4}{5}$ (standard deviation)

Semi-interquartile range = $\frac{2}{3}$ (standard deviation)

These are consequences of the fact that for the normal distribution we find that the mean deviation and semi-interquartile range are equal, respectively, to 0.7979 and 0.6745 times the standard deviation.