- The Variance and standard deviation

The variance is dividing the sum of the squares deviations on the degrees of freedom (n - 1) and it is symbolizes for the society variance is symbol (σ^2) and for sample is symbol (S^2) .

$$S^2 = \frac{\sum yi^2 - (\sum yi)^2 / n}{n - 1}$$

$$S^2 = \sum(y_i - \bar{y})^2 / n - 1 = \frac{\sum yi^2 - (\sum yi)^2 / n}{n - 1}$$

Standard deviation :It is the square root of the variance that sample and is symbolized by the S.

$$S = \sqrt{[\sum(yi - \bar{y})^2] / n - 1} = \sqrt{[\sum yi^2 - (\sum yi)^2 / n] / n - 1} .$$

Example 1. The following data represents the amount of cotton crop in kilograms per acre in five farms . find the standard deviation? Yi = 9, 8, 6, 5, 7.

solution :

Yi	Yi^2
9	81
8	64
6	36
5	25
7	49
$\sum y_i = 35$	$\sum y_i^2 = 255$

$S^2 = \sum y_i^2 - (\sum y_i)^2 / n / n - 1.$

$S^2 = 255 - (35)^2 / 5 / 5 - 1 =$

$S^2 = 255 - 1225/5/4 =$

$S^2 = (255 - 245)/4 =$

$S^2 = 10/4 =$

$S^2 = 2.50$.

$$S = \sqrt{\sum yi^2 - (\sum yi)^2 / n / n - 1}$$

$\mathbf{S} = \sqrt{255 - (35)2/5/5 - 1} = \sqrt{255 - 1225/5/4} = \sqrt{255 - 245/4} =$

$\mathbf{S} = \sqrt{2.50} = 1.581$ kg .

B/ The classified data: If there were y_1, y_2, y_3 ...... y_n are representing categories centers in the frequency distribution table with replicates f_1, f_2, f_3, ... f_n, the standard deviation of these observations is :

$$\mathbf{S} = \sqrt{[\sum \mathbf{fi}\,(\mathbf{yi} - \bar{y})^2]/\sum fi - 1} = \sqrt{[\sum \mathbf{fiyi^2} - (\sum \mathbf{fi}yi)^2/\sum fi]/\sum fi - 1} .$$

Example 2. Calculate the standard deviation and variance of the frequency distribution table :

Factions	f_i	y_i	$f_i\,y_i$	Y_i^2	$f_i\,y_i^2$
60 – 62	5	61	305	3721	18605
63 - 65	18	64	1152	4096	73728
66 - 68	42	67	2814	4489	188538
69 - 71	27	70	1890	4900	132300
72 - 74	8	73	584	5329	42632
Total	100	335	6745		455803

$\mathbf{S^2 = [\sum f_i y_i^2 - (\sum f_i y_i)^2/\sum f_i]/\sum f_i - 1}$ **.**

$\mathbf{S^2} = [455803 - (6745)^2/100]/100 - 1=$

$\mathbf{S^2} = [455803 - (22515025)/100]/99=$

$\mathbf{S^2} = [455803 - 225150.25]/99=$

$\mathbf{S^2} = [852.75]/99 =$

$\mathbf{S^2} = 8.613$ **.**

$$\mathbf{S} = \sqrt{[\sum \mathbf{fiyi^2} - (\sum \mathbf{fiyi})^2/\sum \mathbf{fi}]/\sum \mathbf{fi} - \mathbf{1}} \ \mathbf{.}$$

$S = \sqrt{[455803 - (6745)^2/100]/100 - 1} =$

$S = \sqrt{[455803 - (22515025)/100]/99} =$

$S = \sqrt{[455803 - (454950.25)]/99} =$

$S = \sqrt{(852.75)/99} = \sqrt{8.613}$.

$S = \sqrt{8.613}$.

$S = 2.934$.

- Characteristics of the variation and standard deviation :

(1). When you add or subtract a fixed number (k) to each value of the observations values. The variation and standard deviation are not changing :

The new value of variance = The original value of variance .

The new value of standard deviation = The original value of standard deviation .

(2). If you multiplying every value of the observations values by a fixed number (k) is :

The new value of variance = The original value of variance × square of fixed number (k)
The new value of standard deviation = The original value of standard deviation × fixed number (k) .

(3). If were both X and Y are independent variables, the Z variable was equal to sum of both of them : $Z_i = X_i + Y_i$:

variation Z = variation X + variation Y :

$$S_z^2 = S^2_x + S^2_y \ .$$

(4). If there were two values groups that composed of n_1 and n_2 of observations with have S_1^2 and S_2^2 variation respectively, the pooled variance of all observations are :

Pooled variance : $S^2p = [(n_1 - 1) S^2_1 + (n_2 - 1) S^2_2] / [(n_1 - 1) + (n_2 - 1)]$.

Weighting variance : $S^2p = (SS_1 + SS_2) / (n_1 + n_2 - 2)$. The variation of weighted arithmetic mesn :

Example 1. Calculate the variance and standard deviation of these values and then add each of them 3 and calculate the variance and standard deviation of the new values.

The original values: y_i = 8, 3, 2, 12, 10 .

The new values after adding 3 for each value. x_i = 11 , 6 , 5 , 15 , 13 .

y_i	y_i^2	x_i	x_i^2
8	64	11	121
3	9	6	36
2	4	5	25
12	144	15	225
10	100	13	169
35	321	50	576

$\mathbf{S^2_y = [\sum y_i^2 - (\sum y_i)^2 / n] / n - 1}$ **.**

$S^2_y = [321 - (35)^2 / 5] / 5 - 1 =$

$S^2_y = [321 - 1225/5]/4 =$

$S^2_y = [(321 – 245)]/4 =$

$S^2_y = 76/4 = 19$.

$S^2_y = 19$.

$\mathbf{S_y} = \sqrt{[\sum \mathbf{yi}^2 - (\sum \mathbf{yi})^2 / \mathbf{n}]/\mathbf{n} - \mathbf{1}} =$

$S_y = \sqrt{[321 - (35)^2 /5] /5 - 1} =$

$S_y = \sqrt{19} =$

$S_y = 4.358$.

$\mathbf{S^2_x = [\sum x_i^2 – (\sum x_i)^2 /n] /n -1}$ **.**

$S^2_x = [576 – (50)^2 /5]/5 -1 =$

$S^2_x = [576 - 2500/5]/4 =$

$S^2_x = [(576 – 500)]/4 =$

$S^2_x = 76/4 =$

$S^2_x = 19$.

$\mathbf{S_x} = \sqrt{[\sum \mathbf{xi}^2 - (\sum \mathbf{xi})^2 / \mathbf{n}]/\mathbf{n} - \mathbf{1}} =$

$\mathbf{S_x} = \sqrt{[576 - (50)^2 /5]/5 - 1} =$

$S_x = \sqrt{[576 - 2500/5]/4} =$

$S_x = \sqrt{[(576 - 500)]/4} =$

$S_x = \sqrt{[76/4]} =$

$\mathbf{S_x} = \sqrt{\mathbf{19}} =$

$\mathbf{S_x} = 4.358$.

- The relationship between the standard deviation and the mean deviation :

If the distribution is symmetric (sprain simple quickly) is :

The mean deviation = 4/5 of standard deviation that is,

M.D. = 4/5 × SD .

NOTE: When you measure the dispersion of the samples averages that belongs to the community as it uses the standard error (SE) or standard deviation (SD) or has the symbol is $S\bar{y}$.

$$S_{\bar{y}} = \frac{\sqrt{S^2}}{\sqrt{n}}$$

$$S_{\bar{y}} = \frac{\sqrt{S^2}}{\sqrt{n}}$$

$S_{\bar{y}} = S/\sqrt{n}$.

Standardized scores :

In many cases, we need to compare two items from two different groups. In this case, it must be converted the observations into standard units so that the comparison is meaningful using the arithmetic mean and standard deviation for each group.The standard scores have no units were used in the measurement. As we convert all the values to standard scores, the arithmetic mean of these standard scores is zero and the variance is equal to one. The Z is naturally distributed by arithmetic mean is equal to zero and a standard deviation is to one .

The Z_i value is called a standard score . $Z_i = (Y_i - \bar{Y}) / S$.

Example 1. A student was received 84 score in the final exam in mathematics. The mathematical mean in the mathmatic exam for all the students was 76 and with a standard deviation is to 10. In the physics exam, the same student was received 90 score. The mathematical mean in the mathmatic exam for all the students was 82 and with a standard deviation is to 16. In which subjects was the student's viability higher ?

solution:

When converting these two degrees to standard scores we find that $Z_i = (Y_i - \bar{Y}) / S$.

For mathematics: $\bar{Z} = (84 - 76) / 10 = 0.8$.

For physics: $\bar{Z} = (90 - 82) / 16 = 0.5$.

Thus, the student's ability to pass the mathematics exam is higher than in physics, which is the opposite of previous comparisons .

Example 2. Convert the following values to standard scores : yi = 6, 2, 8, 7, 5 .

$\bar{Y} = \sum Y_i / n$

$\bar{Y} = 6 + 2 + 8 + 7 + 5 /5 =$

$\bar{Y} = 28/5 = 5.6$.

$$\mathbf{S} = \sqrt{[\sum \mathbf{yi^2} - (\sum yi)^2 /n] /n - 1} \;.$$

$S = \sqrt{[(6^2 + 2^2 + 8^2 + 7^2 + 5^2) - (28)^2/5]/5 - 1}$ =

$S = \sqrt{[(36 + 4 + 64 + 49 + 25) - (784)/5]/4}$ =

$S = \sqrt{[178 - 156.8]/4}$ =

$S = \sqrt{[21.2]/4}$ =

$S = \sqrt{5.3}$ =

S = 2.30 .

$Z = (Y_i - \bar{Y}) / S$

$Z_1 = (6 - 5.6) / 2.30 = 0.4/ 2.30 = 0.174$.

$Z_2 = (2 - 5.6) / 2.30 = - 3.6/ 2.30 = - 1.565$.

$Z_3 = (8 - 5.6) / 2.30 = 2.4/ 2.30 = 1.043$.

$Z_4 = (7 - 5.6) / 2.30 = 1.4/ 2.30 = 0.609$.

$Z_5 = (5 - 5.6) / 2.30 =- 0.6/ 2.30 =- 0.261$.

Classes (Z scores) الدرجات القياسية	**fi التكرار**	**Percentage (%)**
High scores : 1.043	1	20%
Medium scores : 0.609 , 0.174	2	40%
Low scores : -1.565 , -0.261.	2	40%
Sum	5	100%