Measures of Dispersion (مقاييس التشتت)

When the values of a set of observations are close to their mean, their dispersion is less than when they are widely dispersed. There are several measures of dispersion, including the following:

1. The Mean Deviation (الانحراف المتوسط)

It is defined as the sum of the absolute deviations of values from their arithmetic mean divided by their number and is symbolized by the symbol $M.D$.

If any data set consisting of n values $x_1, x_2, \cdots, x_n$, then the mean deviation of these values $M.D$ is defined as:

$$M.D = \frac{\sum|x_i - \overline{x}|}{n}$$

Example 1: Here are the ages of 14 cancer patients.

37 48 53 46 42 49 44 38 32 33 52 51 47 41. Determine the mean deviation.

Solution: To find the mean deviation $M.D$, we must know the arithmetic mean $\overline{x}$.

$$\overline{x} = \sum_{i=1}^{14} x_i \Big/ 14 = \frac{616}{14} = 44$$

Now, we create the following table:

x_i	32	33	37	38	41	42	44	46	47	48	49	51	53	55
$\lvert x_i - \overline{x}\rvert$	12	11	7	6	3	2	0	2	3	4	5	7	9	11

$$M.D = \frac{\sum|x_i - \overline{x}|}{n} = \frac{82}{14} = 5.857$$

If we have a frequency table of class with their frequencies, the mean deviation is equal to the sum of the absolute deviations of the values from their arithmetic mean multiplied by their frequencies divided by the sum of the frequencies.

$$M.D = \frac{\sum f_i|x_i - \overline{x}|}{\sum f_i}$$

Example 2: The following table shows the age distribution of accident deaths.

Class	5-14	15-24	25-34	35-44	45-54	55-64	65-74
Frequency	393	514	460	341	415	299	278

Determine the mean deviation.

Solution: First, you need to find the midpoints x_i of each class, and then the following table is formed:

Class	5-14	15-24	25-34	35-44	45-54	55-64	65-74
x_i	9.5	19.5	29.5	39.5	49.5	59.5	69.5
f_i	393	514	460	341	415	299	278
$x_i . f_i$	3733.5	10023	13570	13469.5	20542.5	17790.5	19321
$\|x_i - \overline{x}\|$	27	17	7	3	13	23	33
$f_i\|x_i - \overline{x}\|$	10611	8738	3220	1023	5395	6877	9174

$$\overline{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{98450}{2700} = 36.5$$

$$M.D = \frac{\sum f_i |x_i - \overline{x}|}{\sum f_i} = \frac{45038}{2700} = 16.7$$

2. The Variance (التباين)

To calculate the variance of a sample of values, we subtract the mean from each value, square the resulting differences, then add the squared differences. This squared sum of the deviations of the values from their mean is divided by the sample size minus 1. If any data set consisting of n values $x_1, x_2, \cdots, x_n$, then the variance of these values S^2 is defined as:

$$S^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

Example 3: Determine the variance of a sample in example 1.

Solution:

x_i	32	33	37	38	41	42	44	46	47	48	49	51	53	55
$(x_i - \overline{x})$	-12	-11	-7	-6	-3	-2	0	2	3	4	5	7	9	11
$(x_i - \overline{x})^2$	144	121	49	36	9	4	0	4	9	16	25	49	81	121

$$S^2 = \frac{\sum(x_i - \overline{x})^2}{n-1} = \frac{668}{14} = 47.7$$

The variance can be calculated using the following formula:

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

Example 4: The concentrations of antibodies in the blood serum of a sample of fourth-year students in the Medical Physics Department were as follows:

2.05 , 0.94 ,1.83 ,1.17 , 2.16 , 1.25. Determine the variance.

Solution:

x_i	2.05	0.94	1.83	1.17	2.16	1.25	$\sum x_i = 9.4$
${x_i}^2$	4.2025	0.8836	3.3489	1.3689	4.6656	1.5625	$\sum {x_i}^2 = 16.032$

$$S^2 = \frac{\sum {x_i}^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{16.032 - \frac{(9.4)^2}{6}}{5} = \frac{16.032 - 14.737}{5} = 0.259$$

For calculating the variance when the frequency of the observations is given, such that $x_1, x_2, \cdots, x_n$ is the recorded observations, and $f_1, f_2, \cdots, f_n$ is the respective frequencies of the observations then:

$$S^2 = \frac{\sum f_i(x_i - \overline{x})^2}{\sum f_i - 1}. \quad \text{Or} \quad S^2 = \frac{\sum f_i {x_i}^2 - \frac{(\sum f_i x_i)^2}{\sum f_i}}{\sum f_i - 1}$$

Example 5: Determine the variance of a sample in example 2.

Solution:

Class	5-14	15-24	25-34	35-44	45-54	55-64	65-74
x_i	9.5	19.5	29.5	39.5	49.5	59.5	69.5
f_i	393	514	460	341	415	299	278
$x_i.f_i$	3733.5	10023	13570	13469.5	20542.5	17790.5	19321
$(x_i - \overline{x})$	-27	-17	-7	3	13	23	33
$(x_i - \overline{x})^2$	729	289	49	9	169	529	1089
$f_i(x_i - \overline{x})^2$	286497	148546	22540	3069	70135	158171	302742

$$S^2 = \frac{991700}{2700-1} = 334.356$$

Example 6: In 2014, 50 students from the College of Science at the University of Babylon donated blood to the wounded of the Ali al-Akbar Brigade who were wounded during the operations to liberate Talafar from Daeish gangs. The level of hemoglobin in their blood was summarized in the following table. Calculate the variance.

Class	12.8-13.8	13.9-14.9	15.0-16.0	16.1-17.1	17.2-18.2	18.3-19.3
f_i	3	5	15	16	10	1

Solution:

Class	12.8-13.8	13.9-14.9	15.0-16.0	16.1-17.1	17.2-18.2	18.3-19.3
x_i	13.3	14.4	15.5	16.6	17.7	18.8
f_i	3	5	15	16	10	1
$f_i x_i$	39.9	72	232.5	265.6	177	18.8
${x_i}^2$	176.89	207.36	240.25	275.56	313.29	353.44
$f_i {x_i}^2$	530.67	1036.8	3603.75	4408.96	3132.9	353.44

$$S^2 = \frac{\sum f_i {x_i}^2 - \frac{(\sum f_i x_i)^2}{\sum f_i}}{\sum f_i - 1} = \frac{13066.52 - \frac{(805.8)^2}{50}}{50-1} = 1.64$$

3. Standard Deviation (الانحراف المعياري)

Standard deviation is defined as the square root of the variance. It is symbolized by the symbol S.

Example 7: A student in the Chemistry Department of the College of Science at Babylon University determined chloride in tap water using Moore's method. He repeated the experiment four times and obtained the following results: 12.69 , 12.58 ,13.02 ,12.63. Calculate the standard deviation.

Solution:

x_i	12.69	12.58	13.02	12.63	$\sum x_i = 50.92$
${x_i}^2$	161.0361	158.2564	169.5204	159.5169	$\sum {x_i}^2 = 648.3298$

$$S = \sqrt{S^2} = \sqrt{\frac{\sum {x_i}^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{648.3298 - \frac{(50.92)^2}{4}}{3}}$$

$$= \sqrt{0.0394} = 0.1985$$

Algebraic Properties of Dispersion Measures

1. Let $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ be two set of data such that $y_i = x_i \mp a$, $\forall i = 1,2,\cdots,n$, where a any constant. Then:

$i)\ M.D_x = M.D_y.$ $\quad ii) S_x^2 = S_y^2.$ $\quad iii)\ S_x = S_y.$

In other words, when a certain value is added or subtracted from all the data, the values of all the dispersion measures are not affected.

2. Let $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ be two set of data such that $y_i = bx_i$, $\forall i = 1,2,\cdots,n$, where b any constant. Then:

$i)\ M.D_x = bM.D_y.$ $\quad ii)\ S_x^2 = b^2 S_y^2.$ $\quad iii)\ S_x = bS_y.$

3. Let $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ be two set of data such that $y_i = a \mp bx_i$, $\forall i = 1,2,\cdots,n$, where a, b any constants. Then:

$i)\ M.D_x = bM.D_y.$ $\quad ii)\ S_x^2 = b^2 S_y^2.$ $\quad iii)\ S_x = bS_y.$

The above algebraic properties can be summarized in the following table:

Data	**Mean Deviation**	**Variance**	**Standard Deviation**
$x_1, x_2, \dots, x_n$	$M.D$	S^2	S
$x_1 \mp a, x_2 \mp a, \dots, x_n \mp a$	$M.D$	S^2	S
$bx_1, bx_2, \dots, bx_n$	$b\ M.D$	b^2S^2	bS
$a \mp bx_1, a \mp bx_2, \dots, a \mp bx_n$	$b\ M.D$	b^2S^2	bS

Example 8: Complete the following table:

Data	$M.D$	S^2	S
x_i: 2,3,5,7,8			
y_i: 0,1,3,5,6			
z_i: 6,9,15,21,24			

Solution:

$$\bar{y} = \frac{15}{5} = 3$$

y_i	0	1	3	5	6	Sum
$\lvert y_i - \bar{y} \rvert$	3	2	0	2	3	10
$(y_i - \bar{y})^2$	9	4	0	4	9	26

Data	$M.D$	S^2	S
x_i: 2,3,5,7,8	2	6.5	2.55
$y_i = x_i - 2$: 0,1,3,5,6	2	6.5	2.55
$z_i = 3x_i$: 6,9,15,21,24	6	$9 \times 6.5 = 58.5$	$3 \times 2.55 = 7.65$

H.W.

1. Hb% per 100 cc blood of 7 individuals is as follows: 14.7, 13.1, 11.5, 12.9, 13.8, 14.5 and 12.5. Calculate the mean deviation $M.D$, variance S^2 and standard deviation S.

2. Calculate the mean deviation $M.D$, variance S^2 and standard deviation S from the following frequency distribution table

Class	20-24	25-29	30-34	35-39	40-44	45-49
f_i	8	11	21	28	17	15

3. Complete the following table:

Data	$M.D$	S^2	S
x_i: 5, 7 , 11 , 12 , 15			
y_i: 15,17,21,22,25			
z_i: 10,14,22,24,30			