Department of biology

((Biostatistics))

1^{st} stage

chapter 3

By

Mm. Ali Alawadi

Second: Measures of Dispersion and Variation

Measures of central tendency are insufficient to compare the nature of statistical data, so there are measures of dispersion, which are:

First: Measures of Absolute Dispersion

1. **Range**

   is the difference between the largest value and the smallest value in the dataset

   Range = Largest Value – Smallest Value

$$\boldsymbol{R = U - L}$$

R = Range

U = highest value

L = smallest value

Example (1): Calculate the range (R) for the following values: 90,110,50,100,80

$$R = U - L$$

$$R = 110 - 50 = 60$$

Example (2): The following data represents the pH of 6 surface wells used for irrigation in a given area:

PH: 6.9, 7.6 , 8.3 , 7.6 , 7.7 , 7.4

Since pH vary, the range R is calculated according to the following equation:

$$\boldsymbol{R = U - L}$$

$$R = 8.3 - 6.9 = 1.4$$

Range features:

1. It's easy to calculate.
2. It gives a quick idea of the nature of the data.
3. It is frequently used to monitor production and describe weather conditions.

Disadvantages of the range:

1. It depends on only two values and neglected the rest of the data.
2. It is influenced by outliers and is an approximate measure that is not reliable.

Important Note

It is possible to have two groups with different values or degrees, some of which are similar but have the same range as follows:

A: 6.9 - 7.6 - 8.3 - 7.6 - 7.7 - 7.4 (n=6)

B: 7.4 – 8.3 – 7.4 – 7.4 – 6.9 – 7.4 – 7.4 (n= 7)

Here in both groups, we find that the range is (1.4)

Therefore, the range is used to give a quick initial idea (that does not need complex calculations) of the difference or similarity between the values of the group in question. In cases that require accuracy in determining the degrees of variation or dispersion, we use other dispersion scales that take all values without exception, especially variance and mean deviation.

2. (Average deviation- AD)

The mean deviation is defined as the average of the sum of the deviations of the absolute values from their mean

If we want to measure the degree of dispersion between the values given in the table below, which represents the number of visits made by 6 supervisors to 6 schools.

6	5	4	3	2	1	**Admin hierarchy**
4	6	1	7	4	2	**Number of visits**

Obviously, the number of visits varies according to the supervisor (dispersed). It is possible to calculate the difference of each of them from a specific basis (a specific value), here we first calculate the arithmetic average of the number of visits:

$$\bar{x} \frac{\Sigma x}{n}$$

$$\bar{x} \frac{2+4+7+1+6+4}{6} = \frac{24}{4} = 4$$

The amount of difference of each value (number of visits) of the six values from the arithmetic mean as in the following table (through the difference from the arithmetic mean, which is 4

6	5	4	3	2	1	Admin hierarchy
4	6	1	7	4	2	Number of visits
0	+2	-3	+3	0	-2	Arithmetic mean

This means that the number one admin traffic differs from the arithmetic mean by two visits, while the number of visits of the second admin is no different from the arithmetic average and so on. Through this, we notice that there are differences in the amount of deviation from the arithmetic mean, but the sum of the deviations is equal to zero, and this is the law of the sum of the differences from the mean.

But the conclusion is how it is built to determine the concept of differences or dispersion from the arithmetic mean (i.e. the magnitude of the differences and whether they are negative or positive) so we must get rid of the signal first through absolute value or squared. If we take the absolute value, we get that the differences will be.

0, +2, -3, +3 ,0, -2

Since these absolute differences vary, it is possible to calculate their arithmetic mean, which will represent a measure of dispersion or difference from the arithmetic mean and is called (mean deviation) and symbolized by AD , as in the following equation:

fi|yi-y|	|yi-y|	fiyi	yi	fi	الفئات
32.25	6.45	305	61	5	60-62
62.10	3.45	1152	64	18	63-65
18.90	0.45	2814	67	42	66-68
68.85	2.55	1890	70	27	69-71
44.40	5.55	584	73	8	72-74
226.50		6745		100	

$$AD = \frac{\Sigma|Xi - X|}{n}$$

If we apply this equation to the previous example, we find the value of the mean deviation:

$$AD = \frac{|2-4| + |4-4| + |7-4| + |1-4| + 6-4| + |4-4|}{6}$$

$$AD = \frac{2+0+3+3+2+0}{6} = \frac{10}{6} = 1.7$$

This means that the average deviation of one value from the six values (visits) is 1.7 visits.

Example: Find the mean deviation of the following frequency distribution table

$$y = \frac{\Sigma\, fi\ yi}{\Sigma\ \mathrm{fi}} = \frac{6745}{100} = 67.45$$

$$AD = \frac{\Sigma\ \mathrm{fi}|\mathrm{yi} - \mathrm{y}|}{\Sigma\ \mathrm{fi}} = \frac{226.50}{100} = 2.265$$

3. Variance

Variance is defined in principle as the average square of the differences of values from their arithmetic mean. Population and sample variations vary in terms of symbols and formulas.

The variance of the population of limited size (known size) is calculated according to the following equation:

$$S^2 = \frac{\Sigma Xi^2 - \frac{(\Sigma \mathrm{Xi})^2}{n}}{n - 1}$$

Example: Calculate the following sample variance:

4,6,1,7,4,2

We first calculate the arithmetic mean of the sample (it represents the sum of the values over its number)

$$X = \frac{4 + 6 + 1 + 7 + 4 + 2}{6} = \frac{24}{6} = 4$$

$\Sigma Xi^2 = 4^2 + 6^2 + 1^2 + 7^2 + 4^2 + 2^2$

$\Sigma Xi^2 = 16 + 36 + 1 + 49 + 16 + 4 = 122$

$(\Sigma \mathrm{Xi})^2 = 24^2 = 576$

$\frac{(\Sigma \mathrm{Xi})^2}{n} = \frac{576}{6} = 96$

$$S^2 = \frac{\Sigma Xi^2 \frac{(\Sigma \mathrm{Xi})^2}{n}}{n-1}$$

$$S^2 = \frac{122 - 96}{6 - 1} = \frac{26}{5} = 5.2$$

<u>i.e. the amount of sample variation 5.2 visits</u>

In the case of classified data:

Example: The following is the distribution of grades of 15 students in the subject of life statistics, very variation.

Iteration (fi)	Categories
1	10 -14
3	15 -19
5	20 -24
4	25 -29
2	30 -34

$$S^2 = \frac{\Sigma fiyi^2 \frac{(\Sigma \mathrm{fiyi})^2}{\Sigma fi}}{\Sigma fi - 1}$$

$Fiyi^2$	fiyi	yi^2	yi	fi	Categories
144	**12**	**144**	**12**	**1**	10 -14
867	**51**	**289**	**17**	**3**	**15 -19**
2420	**110**	**484**	**22**	**5**	**20 -24**
2916	**108**	**729**	**27**	**4**	**25 -29**
2048	**64**	**1024**	**32**	**2**	**30 -34**
8395	**345**			**15**	

$$S^2 = \frac{8395 \frac{(345)^2}{15}}{15 - 1} = 32.857$$

4. Standard deviation- SD

Standard deviation or one of the most widely used measures of dispersion and its standard units are the same units by which the values of the studied elements are measured. Standard deviation is defined as the square root of variance and is denoted by the population standard deviation (σ) and the standard deviation of the sample with the symbol (S) or (SD) and therefore:

$$\sigma = \sqrt{\sigma^2}$$
$$S = \sqrt{S^2}$$

In the previous example, the magnitude of the variance was 5.2

That is, the average deviation of one value (number of visits) from the arithmetic mean is 2.28 visits. It takes the units of the original sample so it is frequently used.

The importance of standard deviation is to judge the degree of dispersion of the values of a particular group, so if the value of the standard deviation is relatively small, it indicates that there is a large convergence (or little dispersion) between the values. (The deviation may be higher or lower than the mean so write ±).

Sometimes the value of the standard deviation is not enough on its own, especially if we have several groups and perhaps different units of measurement, so we resort to looking at the proportion of what the standard deviation constitutes from the arithmetic mean, and this leads us to a new scale called (coefficient of variation).

Standard error -SE

It is one of the commonly used measures of dispersion, and represents the standard or standard deviation divided by the root number of observations (n), or it results from dividing the variance by the number of observations under the root, and it is calculated by the following equation:

$$SE \frac{S}{\sqrt{n}}$$

As in the previous example, the variance is 5.2

$$SE \frac{5.2}{\sqrt{6}} = \frac{\mathbf{5.2}}{\mathbf{2.44}} = \mathbf{2.13}$$

Second: Measures of Relative Dispersion

1. Coefficient of variation - CV

The coefficient of variation is defined as the percentage formed by the standard deviation (S) or denoted by (SD) from the arithmetic mean (X) and denoted by (CV). It is calculated according to the following equation:

$$CV = \frac{S}{X} \times 100\%$$

Example: If we have a sample of students' grades for mathematics with an average of (58) and the variance of the sample is (64), that is, the standard deviation is (8), and another sample for genetics with an average of (70) and its variance is (74), that is, the standard deviation is (8.61), find the coefficient of difference for each of them, explaining the results:

Solution:

The coefficient of variation of mathematics is:

$$CV = \frac{8}{58} \times 100\% = 13.79\%$$

The coefficient of variation of the heredity material is:

$$CV = \frac{8.61}{70} \times 100\% = 12.30\%$$

This means that the degree of dispersion of genetics is lower than the degree of dispersion of mathematics.

The coefficient of variation the lower the better, and when it increases significantly (unacceptable) can resort to increasing the sample size (the number of observations studied) to reduce it, as used to compare between traits or methods of measurement or methods of teaching.