Chapter 4 – Measures of Dispersion

Objectives:

At the end of the topic, students are expected to achieve the following:

1. Define and identify the different measures of dispersion;
2. Determine whether a given set is a population or a sample;
3. Solve dispersion values from an ungrouped and grouped data sets.

4.1 DISPERSION

Dispersion is the state of getting dispersed or spread. **Statistical dispersion** means the extent to which numerical data is likely to vary about an average value. Dispersion helps to understand the distribution of the data.

In statistics, the measures of dispersion help to interpret the variability of data, i.e., to know how homogeneous or heterogeneous the data is. In simple terms, it shows how squeezed or how scattered the variable is.

Data dispersion measurement can be used to check how each element in the data set differs from one another. Dispersion (also known as variability, scatter, or spread) is also referred to as the extent in determining how stretched or squeezed a distribution is. Aside from the idea of how each element in the data set behaves, measures of dispersion can also be used to compare two or more different data sets.

For example, there are two different data sets, data set A and data set B. These data sets are shown below:

$$A = \{11, 13, 13, 16, 17, 20\}, \quad \bar{x} = 15$$

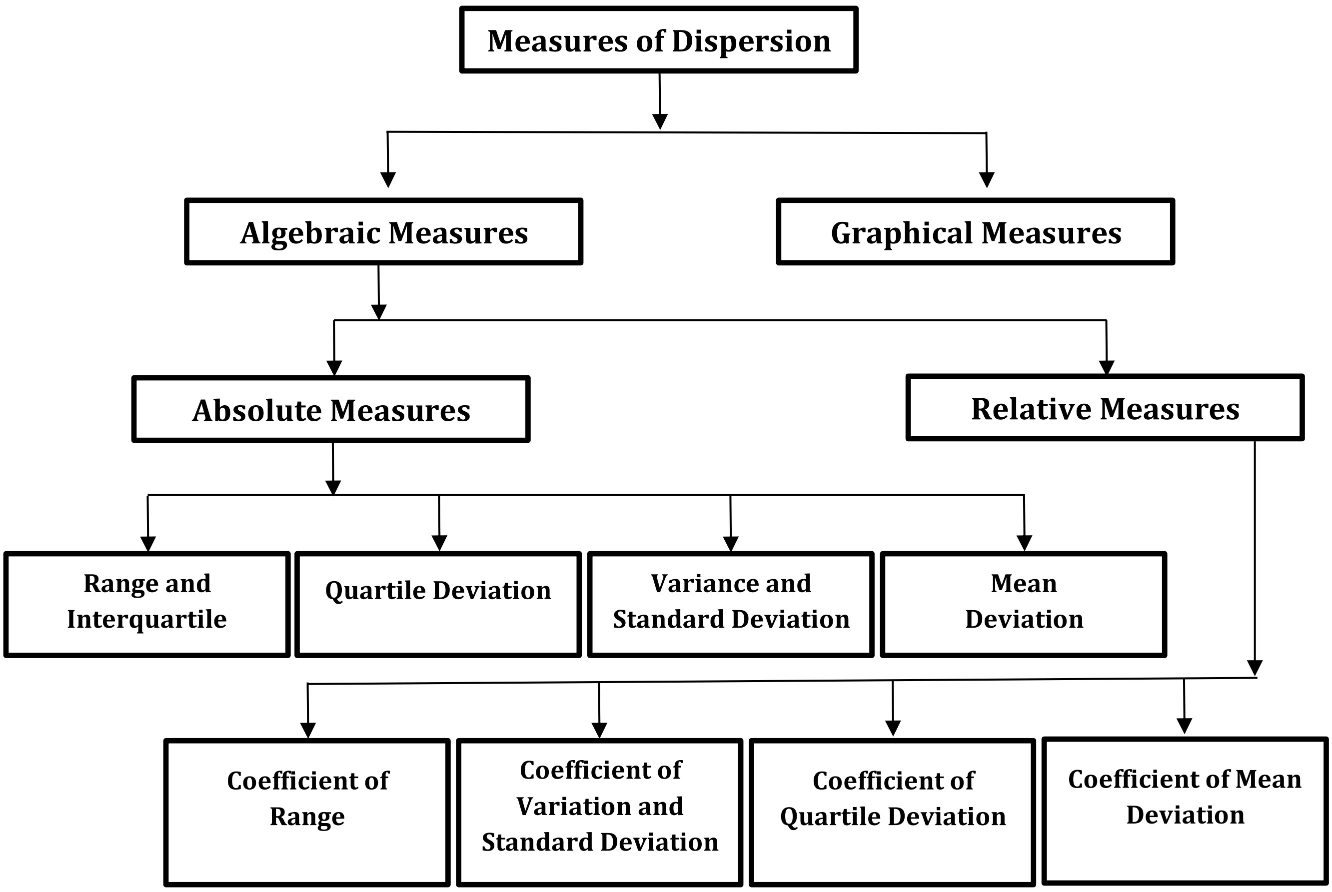
$$B = \{15, 15, 15, 15, 15, 15\}, \quad \bar{x} = 15$$

Both data sets have the same arithmetic mean $(\bar{x})$. Thus, the arithmetic mean cannot distinguish the difference or similarity between these two sets. This is where the measures of dispersion come in. The diagram below shows the different categories in measuring the dispersion of data.

There are two types of measures of dispersion, and each type has its measures. These types are the ***absolute measures*** and the ***relative measures***. However, we will only talk about the absolute measures in this chapter.

Types of measures of dispersion:

1. **Absolute Measure of Dispersion** – expresses the variations in terms of the average of deviations of observations like the standard or the means deviations. It contains the same unit as the original data. The commonly used measures of dispersion under this type are as follows:
   a. **Range** – it is simply the difference between the maximum value and the minimum value given in the data set.
   b. **Variance** – it is calculated by deducting the mean from each data in the set then squaring them and adding each square and finally dividing by the total number of values in the data set.
   c. **Standard Deviation** – this is the square root of the variance.
   d. **Interquartile Range** – this is the distance between the first and the third quartile marks. It tells us the range of the middle half of the data.
   e. **Quartile Deviation** – this is the half of the distance between the first and the third quartile marks, or simply, half of the inter-quartile range.
   f. **Mean Deviation** – this tells us how far, on average, the distance of the individual data to the measure of central tendency used.
2. **Relative Measure of Dispersion** – used to compare the distribution of two or data sets. This measure compares values without units. The most common measures under this type are:
   a. Coefficient of range
   b. Coefficient of variation
   c. Coefficient of standard deviation
   d. Coefficient of quartile deviation
   e. Coefficient of mean deviation



Two Algebraic Data Dispersion Measures:

1. **Absolute Variations/Measures** – are the dispersion of data in the same set, which will determine how much they are different from each other. These measures use the original units of data and are most useful in understanding the dispersion within the context of your experiment and measurements.
2. **Relative Variations/Measures** – are calculated as ratios or percentages; for example, one relative measure of dispersion is the ratio of the standard deviation to the mean. These measures of dispersion are always dimensionless, and they are particularly useful for making comparisons between separate data sets or different experiments that might use different units. They are sometimes called **coefficients of dispersion**.

The easiest way to differentiate relative dispersion and absolute dispersion is to check whether your statistic involves units. Absolute measures always have units, while relative measures don't have. However, our lesson will only focus on the absolute dispersion measures.

4.2 RANGE, INTERQUARTILE RANGE, AND QUARTILE DEVIATION

In statistical analysis, comparing two sets of data using the measures of central tendency is often not adequate. Not adequate in the sense that when two different data sets have the same arithmetic means, it doesn't mean that they are the same, thus, they can be entirely different. In addition to this measure used, one needs to know the extent of variability between these two different sets. This is where the measures of dispersion come in.

For example, three friends, Jean, Carol, and Liza, compare the six months incomes of their family's business. The incomes are as follows:

Month	Jean	Carol	Liza
January	0.00	9,000.00	5,000.00
February	11,000.00	10,000.00	7,000.00
March	20,000.00	14,000.00	8,000.00
April	19,000.00	17,000.00	0.00
May	15,000.00	18,000.00	20,000.00
June	25,000.00	22,000.00	50,000.00
Total Income	**฿ 90,000.00**	**฿ 90,000.00**	**฿ 90,000.00**
Average Income	**฿ 15,000.00**	**฿ 15,000.00**	**฿ 15,000.00**

Notice that the averages of their family's business income are the same. When we try to look closely at the incomes, there is a real difference in it. This is quite obvious that the average only tells us one aspect of the data, that is, a representative size of the values. This is where the measures of dispersion come in.

Range (r) and Midrange (r_{mid})

The **range** is the difference between the highest value and the lowest value of the data. A higher value of range means higher dispersion and a lower value means lower dispersion. The **midrange** is another form of the range, where its purpose is to determine the value that sits in the middle of the two extreme values.

Range of Ungrouped Data:

$$r = x_{max} - x_{min}$$

Range of Grouped Data:

$$r = U_{highest\ class} - L_{lowest\ class}$$

Midrange of Ungrouped Data:

$$r_{mid} = \frac{x_{max} + x_{min}}{2}$$

Midrange of Grouped Data:

$$r_{mid} = \frac{\left(U_{highest\ class} + L_{lowest\ class}\right)}{2}$$

The range is unduly affected by extreme values. It is not based on all the values. As long as the minimum and maximum values remain unaltered, any change in other values does not affect the range. It cannot be calculated for open-ended frequency distribution.

Open-ended distributions are those in which either the lower limit of the lowest class or the upper limit of the highest class or both are not specified.

Interquartile Range (IQR)

The **interquartile range** is a measure of where the "middle fifty" is in a data set. While the range is the measure of where the beginning and end are in a set, an interquartile range is a measure of where the bulk of the values lie. That's why it's preferred over many other measures of dispersion when reporting things like school performance. This measure doesn't include the lower 25% and the upper 25% of the data. This makes the IQR not affected by extreme values.

The formula for the computation of this measure is:

Interquartile Range:

$$IQR = Q_3 - Q_1$$

Quartile Deviation (QD)

The quartile deviation is just half of the interquartile range. This measure is sometimes referred to as the semi-interquartile range. Since this measure uses the interquartile range, the bases of its computations are the middle 50% of the data set, thus, excluding the lower and upper 25% of the data.

Quartile Deviation:

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

Example 1. Calculate the range, the interquartile range, and the quartile deviation of the following data:

20 25 29 30 35 39 41 48 51 60 70

Example 2. A set of data has a mid-range of 40 and a range of 20. What are the minimum and the maximum values of this set of data?

Example 3. In a midterm exam, three students have an arithmetic mean of 67, a median of 65, and a range of 16. What is the highest score of these three students?

Example 4. Isaac recorded the arithmetic mean of temperatures in ${}^{0}C$ of ten fridges in a factory. The range of the temperature for these ten fridges is 6.2. He later misplaces one value, the lowest temperature value. The recorded nine values are:

3.0, 4.2, -1.1, 0.3, 0.1, -0.6, 2.1, -0.2, 0.0

What is the arithmetic mean of these ten fridges' temperatures?

Example 5. The distribution below presents the scores of forty students on a certain test. Calculate the range, the interquartile range, and the quartile deviation of this distribution.

Test Scores	f	F
1 – 10	5	
11 – 20	8	
21 – 30	16	
31 – 40	7	
41 – 50	4	

Example 6. The ages of the residents of a certain community are tabulated below. Calculate the range, the interquartile range, and the quartile deviation of this distribution. (***This is an example of uneven distribution.)

Test Scores	***f***	***F***
0 - 10	42	
11 – 20	65	
21 – 40	80	
41 – 55	50	
56 – 75	25	

Exercise 4.1

1. The monthly salaries of eight employees who work for a small company in Bangkok are listed below. What are the range, the interquartile range, and the quartile deviation of their salary?

฿ 12,000	฿ 11,500	฿ 10,500	฿ 18,000
฿ 15,000	฿ 13,000	฿ 21,500	฿ 16,000

2. Find the range, the interquartile range, and the quartile deviation of the data set: 15, 14, 10, 8, 12, 8, 16, 12, 12, 13.

3. The distribution below is the summary of the test scores of students in a certain class. Find the range, the interquartile ranges, and the quartile deviation of this distribution.

Test Scores	*f*	*F*
46 – 55	3	
56 – 65	4	
66 – 75	8	
76 – 85	9	
86 – 95	4	
96 – 105	2	

4. A certain group of students was asked to give random real numbers from 11 to 100. The distribution below shows their responses. What are the range, the interquartile range, and the quartile deviation of this data set?

Number Responses	**Number of Students**
11 – 25	12
26 – 45	9
46 – 56	17
57 – 70	16
71 – 90	20
91 – 100	16

4.3 VARIANCE, STANDARD DEVIATION, AND MEAN DEVIATION

The range, the interquartile range, and the quartile deviation are known measures of dispersion from the spread of values. Those three measures consider the lower-end and the upper-end values in measuring the dispersion.

However, when trying to get the dispersion of certain data using the mean, then we need measures that can be categorized as the measures of dispersion from the mean. These measures are the variance, the standard deviation, and the mean deviation. Unlike the r, the IQR, and the QD, these measures make use of all the data in the data set.

Variance (σ^2 or s^2)

Variance measures how far a data set is spread out. It is the average of the squared differences of each data from the mean.

Variance of a Population (Ungrouped):

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$$

or

$$\sigma^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}$$

Variance of a Sample (Ungrouped):

$$s^2 = \frac{\sum(X-\bar{x})^2}{n-1}$$

or

$$s^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}$$

where:

X - the value of individual data
μ - population mean
$\bar{x}$ - sample mean
N - population size
n - sample size

Variance of a Population (Grouped):

$$\sigma^2 = \frac{\sum f(X-\mu)^2}{N}$$

or

$$\sigma^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N}$$

Variance of a Sample (Grouped):

$$s^2 = \frac{\sum f(X-\bar{x})^2}{n-1}$$

or

$$s^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}$$

where:

X - interval class mark or center
f - interval frequencies
μ - population mean
$\bar{x}$ - sample mean
N - population size
n - sample size

Notice that the formulas stated above are separated into grouped and ungrouped variance, as well as the variance of a population or a sample. The main reason why the formulas for the sample have an "$n-1$" in the denominator has something to do with the computation of an unbiased estimator. An unbiased estimator is an estimator whose expected value is the parameter being estimated. (***Hopefully more discussions on this in additional math.)

Standard Deviation (σ or s)

The **standard deviation** is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. The **standard deviation** is calculated as the square root of variance by determining each data point's deviation relative to the mean. If the data points are further from the mean, there is a higher deviation within the data set; thus, the more spread out the data, the higher the standard deviation.

The following statements are key takeaways in understanding standard deviation:

1. Standard deviation measures the dispersion of a data set relative to its mean.
2. A high standard deviation means that values are generally far from the mean, while a low standard deviation indicates that the values are clustered to the mean.
3. A volatile stock (in business and market) has a high standard deviation, while the deviation of a stable blue-chip stock is usually rather low.
4. As a downside, standard deviation calculates all uncertainty as risk, even when it's in the investors' favor – such as the above-average returns.
5. Standard deviation assumes that the distribution of the data is normal, thus, it can easily be affected by the outliers and extreme values.

Standard Deviation in Business:

Standard deviation is an especially useful tool in investing and trading strategies as it helps measure market and security volatility – and predict performance trends. As it relates to investing, for example, an index fund is likely to have a low standard deviation versus its benchmark index, as the fund's goal is to replicate the index.

On the other hand, one can expect aggressive growth funds to have a high standard deviation from relative stock indices, as their portfolio managers make aggressive bets to generate higher-than-average returns.

A lower standard deviation isn't necessarily preferable. It all depends on the investments and the investor's willingness to assume risk. When dealing with the amount of deviation in their portfolios, investors should consider their tolerance for volatility and their overall investment objectives. More aggressive investors may be comfortable with an investment strategy that opts for vehicles with higher-than-average volatility, while more conservative investors may not.

Standard deviation is one of the key fundamental risk measures that analysts, portfolio managers, advisors use. Investment firms report the standard deviation of their mutual funds and other products. A large dispersion shows how much the return on the fund is deviating from the expected normal returns. Because it is easy to understand, this statistic is regularly reported to the end clients and investors.

Since standard deviation is just the square root of the variance, its formula is the same as that of the variance. The only difference is the removal of the square factor on the left side, and inserting the square root symbol on the right side. Note that, the units of measurement between these two measures of dispersion are different. The unit of the variance is in terms of square units (i.e., square meter m^2), while that of the standard deviation is the same unit used in the data (i.e., meter m).

Standard Deviation of a Population (Ungrouped):

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$$

or

$$\sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

Standard Deviation of a Sample (Ungrouped):

$$s = \sqrt{\frac{\sum(X-\bar{x})^2}{n-1}}$$

or

$$s = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}}$$

Standard Deviation of a Population (Grouped):

$$\sigma = \sqrt{\frac{\sum f(X-\mu)^2}{N}}$$

or

$$\sigma = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N}}$$

Standard Deviation of a Sample (Grouped):

$$s = \sqrt{\frac{\sum f(X-\bar{x})^2}{n-1}}$$

or

$$s = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1}}$$

Example 7. Calculate the variance, and the standard deviation of the following data, treating it first as a population, then treating it next as a sample. What can you observe about the results? (***Note: it will be easier to calculate the values if the data are organized as columns.)

20 25 29 30 35 39 41 48 51 60 70

Population	**Sample**
Variance:	*Variance:*
Standard Deviation:	*Standard Deviation:*

Example 8. The distribution below presents the scores of M.6 students on a certain test. Calculate the variance and the standard deviation of this distribution. (***Note: you need to extend the table by adding some more columns.)

Test Scores	f	F
1 – 10	5	
11 – 20	8	
21 – 30	16	
31 – 40	7	
41 – 50	4	

Mean Deviation (MD)

The **mean deviation** (sometimes called the **mean absolute deviation** or MAD) is defined as a statistical measure that is used to calculate the average deviation from the mean value of the given data set. The mean deviation of the data values can be easily calculated using the below procedure.

Step 1: Find the mean value for the given data values.

Step 2: Subtract the mean value from each of the data values given. (Note: Ignore the minus symbol)

Step 3: Now, find the mean of those values obtained in step 2.

These three steps in finding the mean deviation apply to both ungrouped and grouped data. However, for the grouped data, step 2 is about subtracting the value of the class marks X by the obtained arithmetic mean μ or $\bar{x}$.

Mean Deviation (Ungrouped):

$$MAD = \frac{\sum(x-\mu)}{N}$$

Mean Deviation (Grouped):

$$MAD = \frac{\sum[f(X-\mu)]}{N}$$

Example 9. Calculate the mean deviation of the following data. (Note: it will be easier to calculate the values if the data are organized as columns.)

20 25 29 30 35 39 41 48 51 60 70

Example 10. The distribution below presents the scores of M.6 students on a certain test. Calculate the mean deviation of this distribution.

Test Scores	f	X	$(X-\mu)$	$f(X-\mu)$
1 – 10	5			
11 – 20	8			
21 – 30	16			
31 – 40	7			
41 – 50	4			

Exercise 4.2

1. The monthly salaries of eight employees who work for a small company in Bangkok are listed below. What is the variance, the population, and the mean deviation of their salary?

฿ 12,000	฿ 11,500	฿ 10,500	฿ 18,000
฿ 15,000	฿ 13,000	฿ 21,500	฿ 16,000

2. Find the mean deviation of the data set: 15, 14, 10, 8, 12, 8, 16, 12, 12, 13.

3. The distribution below is the summary of the test scores of students in a certain class. Find the variance, the standard deviation, and the mean deviation of this distribution.

Test Scores	f	F
46 – 55	3	
56 – 65	4	
66 – 75	8	
76 – 85	9	
86 – 95	4	
96 – 105	2	