In previous chapter, we learned how to summarize a raw data set using a single value that indicates the center of the distribution.
Obviously, this is an important property of data description.
But it gives only partial information of the data set.

For example, Record the examination results of two students
Student A: 50, 49, 51, 50
Student B: 100, 100, 0, 0
Where, $\bar{X}_A = 50$ and $\bar{X}_B = 50$. On average both students are equal in getting final exam score. Right???? But, actually are they same?
No, they are not same. Why? Because!!! The variation of each observation are not same for both students.

So, it is also important to describe "How much the observation vary from one to another. And, it is called "Dispersion or Variation".

So, what is dispersion and measures of dispersion?

Dispersion may describe how much the observation vary from one to another.

Descriptive measures that indicate the amount of variation in a data set are called measures of dispersion.

Why do we need?
To determine the reliability of an average
To compare the variability of two or more data sets.

There are two types of measures of dispersion,

1. Absolute measure
2. Relative measure

Absolute measures: Only measure the inherent variation of a data set.

Important measures are,

1. Range
2. Mean Deviation
3. Variance
4. Quartile deviation

Range: Difference between Highest value and Lowest value

For example, weights of 5 newly born babies (in pounds)

$$7.5,\ 4.5,\ 10.1,\ 9.6,\ and\ 5.5$$

$$Range = 10.1 - 4.5 = 5.6\ pounds$$

Mean Deviation:

Let $x_1, x_2, \ldots, x_n$ be n observations, where $\bar{x}$ is average/ or mean

$MD = \frac{\sum |x_i - \bar{x}|}{n}$ [for ungrouped data]

$MD = \frac{\sum f_i\, |x_i - \bar{x}|}{\sum f_i}$ [for grouped data]

For example, x=1,2,4,5. Calculate the mean deviation

<table>
<tr><th>$\boldsymbol{x_i}$</th><th>$\boldsymbol{\bar{x}}$</th><th>$\boldsymbol{|x_i - \bar{x}|}$</th></tr>
<tr><td>1</td><td rowspan="4">$\frac{\sum x_i}{n} = 3$</td><td>2</td></tr>
<tr><td>2</td><td>1</td></tr>
<tr><td>4</td><td>1</td></tr>
<tr><td>5</td><td>2</td></tr>
</table>

$$MD = \frac{\sum |x_i - \bar{x}|}{n} = \frac{6}{4} = 1.5$$

Variance:

For grouped, $\sigma^2 = \frac{\sum f_i (x_i - \mu)^2}{N}$

Let $x_1, x_2, \ldots, x_n$ be n observations

$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$ [Population variance (ungrouped)]

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

For grouped, $s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n-1}$

Standard Deviation:

Square root of variance

$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$ [Population standard deviation (ungrouped)]

$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$[Sample variance (ungrouped)]

For example: weights of 10 newly born babies (in pounds)

7.5, 4.5, 10.1, 9.6, 5.5, 6.6, 7.8, 5.9, 6.0, 5.5

a) Calculate the standard deviation

b) Interpretation

x_i	$\bar{x}$	$(x_i - \bar{x})^2$
7.5		0.36
4.5		0.57
10.1		10.24
9.6		7.29
5.5	$\frac{\sum x_i}{n} = 6.9$	1.96
6.6		0.09
7.8		0.81
5.9		1
6.0		0.81
5.5		1.96

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{30.28}{10-1}} = \sqrt{3.36} = 1.83\, pounds$$

The average variation of the weights of the newly born babies from the mean weights is 1.83 pounds.

Relative Measures: To compare the variation of two or more data sets having different or same units.

1. Coefficient of Range
2. Coefficient of MD
3. Coefficient of Variance
4. Coefficient of QD

Coefficient of Variance (CV):

$$\text{General formula of CV} = \frac{SD}{mean} \times 100$$

- It is a ratio
- Between standard deviation (SD) and mean
- Expressed as percent

$$CV = \frac{\sigma}{\mu} \times 100 \text{ ; } [For\ population]$$

$$CV = \frac{s}{\bar{x}} \times 100 \text{ ; } [For\ sample]$$

Higher CV means High Variation

Lower CV means Low Variation

[More consistent, More uniform, More homogeneous, More stable]

Section A: Mean = 30 and SD = 4

Section B: Mean = 25 and SD = 6

Which section is more consistent in getting final exam mark?

$$\boldsymbol{CV_A = \frac{4}{30} \times 100 = 13.3\%}$$

$$\boldsymbol{CV_B = \frac{6}{25} \times 100 = 24\%}$$

Which one is lower? CV of A!!! Right!!!!

Since, the CV of A is lower than the CV of B. Thus, section A is more consistent comparatively section B