Basic Tools of Economic Analysis

1 CHAPTER

INTRODUCTION

In Book One, the mean, median and mode as measures of location were discussed. Notably, they are also measures of averages. However, to obtain these averages, we need other measures that show how other values deviate from averages. These are called the measures of dispersion.
This chapter will also treat linear equations as one of the basic tools of economic analysis.

OBJECTIVES

At the end of this chapter, students should be able to show simple economic relationships between, as well as calculate each of the following measures:

- Range;
- Quartile;
- Percentile;
- Variance;
- Mean deviation;
- Standard deviation;

1.1 Measures of Dispersion

This is the degree to which numerical data is spread around the average. These measures of dispersion include:

a. Range
b. Quartile
c. Variance
d. Mean deviation
e. Standard deviation

1.1.1 The Range

The range (R) of a distribution is simply the difference between two extreme items or observations. It is the difference between the lowest figure and the highest figure in a series of figures or values. For example, the range (R) of the set of numbers 30, 60, 60, 85 and 90, will be (90 − 30) = 60 marks. The formula is given as:

$R = X_h - X_1$,

where X_h is the highest value; $X_1$is the lowest value. When $X_h = 90$ and $X_1 = 30$, then

$R = 90 - 30 = 60$

1.1.1.1 Advantages of the Use of Range

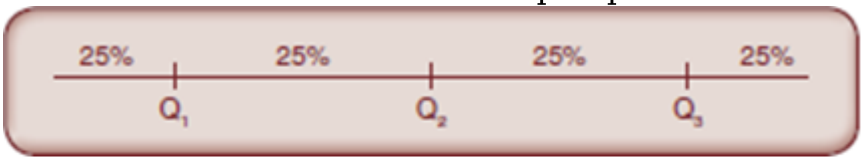
a. It is the simplest measure of dispersion.
b. It is a good indication of spread.
c. It is not difficult to calculate.
d. It is easy to understand.
e. It is useful for further statistical studies.
f. It gives rough estimate of the difference between two values.

1.1.1.2 Disadvantages of the Use of Range

a. The range does not take all the values into account.
b. It cannot be used to calculate open-ended discussions.
c. It is very difficult to interpret.

1.1.2 The Quartile

This is a means of identifying the range within which most values in a distribution occurs. It is similar to the median in concept, but the quartile divides a distribution into four equal parts



The lower quartile is the value below which 25% of the observation falls, and the upper quartile is the value above which 25% of the observation falls. Therefore, 50% of the observation falls between the lower and the upper quartiles, and this is known as the median.

Q_1 = lower quartile = 25% quartile
Q_2 = middle quartile = 50% quartile
Q_3 = upper quartile = 75% quartile

Example: Calculate the quartile of the following distribution:

Observation X	Frequency F	Cumulative Frequency CF
3	2	2
4	3	5
6	6	11 (Q_1)
7	8	19
10	9	28 (Q_2)
11	8	36 (Q_3)
15	5	41
20	3	44
	44	

Solution $Q_1 = \frac{25}{100} \times 44 = 11\text{th}$ item which is 6.

$Q_2 = \frac{50}{100} \times 44 = 22\text{nd}$ item which is 10.

$Q_3 = \frac{75}{100} \times 44 = 33\text{rd}$ item which is 11.

1.1.3 The Percentile

The percentile is a value on a scale of one hundred that indicates the percent of a distribution that is equal to or below it. It is calculated in the same manner as the quartile. The same formula is used but there are 99 percentiles instead of three quartiles. The only difference is that instead of applying the formula to the class in which the first, second or third quartile belong, it would now be applied to the class within which the particular percentage values lie.

1.1.4 Variance

The variance is the square of the standard deviation. By calculation, it is derived through the summation of the squares of the difference between each observation and the mean divided by the number of observations.

1.1.4.1 Advantages of the Use of Variance
a. It gives more information for further computation about a given data.
b. It is very important and useful in data analysis.
1.1.4.2 Disadvantages of the Use of Variance
a. It is not a sound mathematical index.
b. It exaggerates the dispersion of data by squaring their variations.

1.1.5 The Mean Deviation

This measures the dispersion around the arithmetic mean and it is defined as the sum of differences of all values from the arithmetic mean divided by the frequency of the number.

Symbolically, the mean deviation is given as:

MD = $\frac{1}{n}\Sigma|x-\bar{x}|$
where *n* is the frequency or number of times the value occurs; Σ is the sum; *x* is the variable or values; $\bar{x}$ is the arithmetic mean; the two bars "||" represent the absolute values.
Example: Consider the following marks:
30, 60, 60, 85, 90
Solution:

$$\frac{30 + 60 + 85 + 90}{5} = 65$$

The MD is determined as follows:

Marks X	Mean $\bar{X}$	Deviation from the Mean $\lvert x-\bar{x}\rvert$	Absolute Deviation $\lvert d\rvert$ or $\lvert x-\bar{x}\rvert$
30	65	−35	35
60	65	−5	5
60	65	−5	5
85	65	20	20
90	65	25	25
$\Sigma\lvert X-\bar{X}\rvert = \Sigma = 90$		$\Sigma\lvert x-\bar{x}\rvert = \Sigma = 90$	

The summation of the absolute value is 90.
Divide the summation by *n*, that is:

$$\frac{1}{n}\Sigma\,|x-\bar{x}|$$

We have:

MD = $\frac{1}{5}\times 90 = 18$
The average deviation from the mean 65 marks is 18.

1.1.6 The Standard Deviation

This is a measure that shows the extent to which data deviates from the average. It is, by definition, the square root of the variance. In other words, it is also called a measure of the degree of scattering of a frequency distribution about its arithmetic mean and is given as standard deviation

$= \sqrt{\text{variance}}.$

To calculate the standard deviation, the following procedures apply:

a) Calculate the arithmetic mean $\bar{x}$ from the x variables.
b) Subtract the arithmetic mean from each of the variables $x_1, x_2, x_3, \ldots, x_n$ to obtain the deviation $(x - \bar{x})$..
c) Square each of the deviations from the mean, that is, $(x_1 - \bar{x})^2$, $(x_2 - \bar{x})^2 \ldots (x_n - \bar{x})^2$.
d) Find the sum of the squares of the deviations and divide the sum by the number of variables.
e) The square root of the result from step d gives the standard deviation.

Example: Find the standard deviation of the following distribution 2, 3, 4, 5, 6, 6, 7, 8, 9, 10.

Solution

$\Sigma x = 60$

$N = 10$

Therefore, $x = 60/10 = 6$

X	$d = (x - \bar{x})$	$d^2 = (x - \bar{x})^2$
2	−4	16
3	−3	9
4	−2	4
5	−1	1
6	0	0
6	0	0
7	1	1
8	2	4
9	3	9
10	4	16

SD= $\sqrt{\Sigma d^2}$ or $\sqrt{\Sigma (x - \bar{x})^2} = 60 = 7.1.$

The calculation of the standard deviation from a grouped data distribution follows the same step as for an ungrouped series.

In mathematical symbols, the standard deviation for group data is S, where $S = \sqrt{\Sigma f (x - \bar{x})^2 / (n - 1)}$ where x are the mid-values of the class ranges and $\Sigma f = n$.

Example: Find the standard deviation of the following distribution of spelling mistakes in a number of pages of a document, where X is equal to the number of mistakes and F is the frequency:

X	F
1 to < 3	1
3 – 5	3
5 – 7	7
7 – 9	13
9 – 11	9
11 – 13	5
13 – 15	1
15 – 17	1
Total (Σf)	40

Solution

Mid-value (X)	f	X^2	fx	fx^2
2	1	4	2	4
4	3	16	12	48
6	7	36	42	252
8	13	64	104	832
10	9	100	90	900
12	5	144	60	720
14	1	196	14	196
16	1	256	16	256
Total	40	816	340	3, 208

Variance = $S^2 = (\Sigma fx^2 - (\Sigma fx)^2/\Sigma f)/(\Sigma f - 1)$
$S^2 = (3208 - 340^2/40)\,/(40 - 1) = (3208 - 2890)/39$
= 318/39
= 8.15
Therefore, the standard deviation is calculated thus:
SD = $S = \sqrt{8.15} = 2.9$

1.1.6.2 Advantages of Standard Deviation

a) It is a good measure of dispersion, since all the values are used in its computation.
b) It is most useful mathematically, especially for further statistical analysis.
c) Standard deviation is used in samples to interpret the characteristics of the population from which they are drawn.
d) With the help of standard deviation, industries can set up confidence intervals or tolerance limits to assess the products.
e) Those whose specifications fall outside the limits are rejected.

1.1.6.3 Disadvantage of Standard Deviation

Its calculation may pose a problem to mathematically-uninclined minds.

Linear Equation

A linear equation is one in a straight line. It takes the form of $y = ax + c$ where a and c are constants and x and y are variables. Consider the figure below:

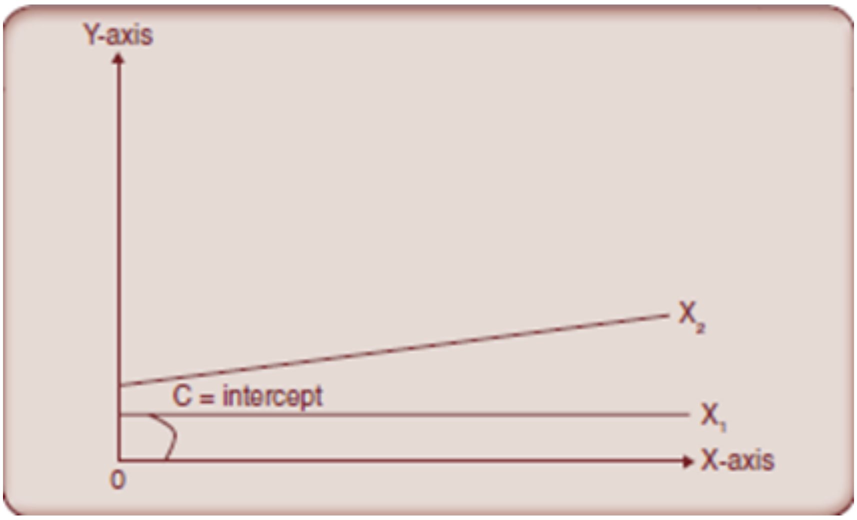


FIG. 1.1 Ay = ax + c line

C is the intercept on the y axis and "a" is the gradient of the line $y = ax + c$. It is a variable whose value depends on the value of x. If the values of ax and c are known (a and c are constants) you can find the value of y so far the straight line is established.

Linear equation establishes a trend line form in which a forecast can be made.

Summary

This chapter has discussed:

§ The five measures of dispersion, namely;
a. Range **b.** Quartile **c.** Mean deviation
d. Standard deviation **e.** Variance

§ The quartile which divides a distribution into four equal parts.

§ The standard deviation which is obtained by using the formula

$$SD = \sqrt{(\Sigma fx^2 - (\Sigma fx)^2/\Sigma f)/(\Sigma f - 1)}$$

§

§ The variance which is the square of the standard deviation.

§ A linear equation which is one in a straight line. It takes the form of $y = ax + c$ where a and c are constants and x and y are variables. Consider the figure below:

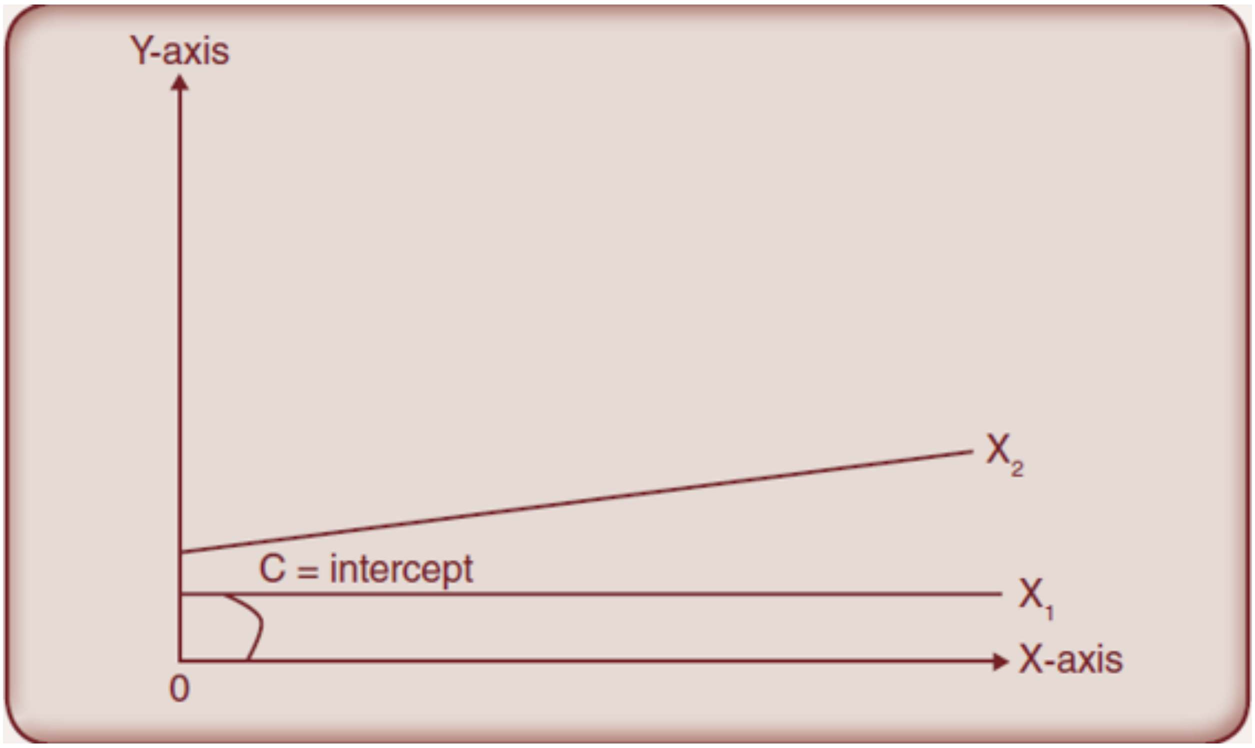


FIG. 1.1 Ay = ax + c line

C is the intercept on the *y* axis and “a” is the gradient of the line $y = ax + c$. *Y* is a variable whose values depend on the value of *x*. If the values of *ax* and *c* are known (*a* and *c* are constant) you can find the value of *y*, ones the straight line is established, the value can be determined.

§ Linear equation which establishes a trend line from which a forecast can be made.

Class Activity

Students should practice the drawing of the graph and do some calculations as the teacher deems fit.

Revision Questions

Objective Questions

1. The graph of the function $x = a + by$ is:
- **(a)** Linear
- **(b)** Quadratic
- **(c)** Cubical
- **(d)** Experimental

2. The difference between the highest and the lowest number in a set of data is:
- **(a)** Range
- **(b)** Median
- **(c)** Variance
- **(d)** Mode **(SSCE 2001)**

3. Which of the following is **not** an advantage of tabular presentation of data?
- **(a)** It enables easy location of required figures
- **(b)** It makes for easy comparison of figures
- **(c)** It occupies more space than mathematical equations
- **(d)** It shows whether the figures are increasing or decreasing **(SSCE 2003)**

4. Find the mean deviation of the following distribution: 3, 5, 5, 6, 8, 9
- **(a)** 7.0
- **(b)** 1.6
- **(c)** 8.0

(d) 12.0

5. Find the range of the following scores: 25, 37, 49, 59, 89, and 98

(a) 64
(b) 10
(c) 73
(d) 44

Essay Questions

1. What are the measures of dispersion?

2. The quantity of fertilizer in bags used by 20 farmers in one year is shown in the following distribution: 2, 4, 5, 5, 4, 8, 6, 7, 6, 2, 4, 5, 6, 6, 7, 10, 12, 10, 5, 6. From the given data:

(a) Calculate the range
(b) Mean deviation
(c) Standard deviation
(d) The variance of the distribution

3. The quantity of milk demanded by 8 consumers in a week is shown in the following distribution: 2, 4, 6, 8, 4, 6, 7, 10.
From the information above, calculate:

(a) The mean deviation
(b) Variance
(c) Standard deviation

4. Define the variance and the standard deviation of given set of observation. Use the following table in which F is the frequency of an observation x, to calculate the mean deviation and standard deviation of the distribution

X	0	1	2	3	4	5	6
F	2	13	16	20	45	38	4

Glossary

Linear equation: This is a straight line equation
Range: It is the difference between the lowest and the highest Value
Mean: Average
Frequency: Number of occurrence
Variable: This is an item that can take different values