Measures of Central tendency

For the purpose of understanding the salient features of a data set, it is essential to extract from the data just a few numbers which summarise the essential information. The summary numbers include Measures of central tendency, Measures of variation, Measures of skewness and Measures of kurtosis. These summarising numbers are called (i) Parameters if the data is from the whole population and (ii) Statistics (plural of a statistic) if the data is from a sample.

A measure of central tendency is a single value that serves as a representative of a set of observations. It is called a measure of central tendency because other values tend to cluster around it. The purpose of a measure of central tendency is to pinpoint the centre of a set of data.

Measures of central tendency are also called measures of average and include the mean, median, mode and the mid-range.

Mean

The mean can be classified into the following categories:

1. Arithmetic mean
2. Weighted arithmetic mean
3. Harmonic mean
4. Geometric mean

Arithmetic mean

Is the most commonly used measure of central tendency.

Arithmetic mean is the amount that each observation in a data set would assume or get if the total amount were shared/divided equally among all observations in the data set.

There are two methods for calculating the mean.

1. **Direct method**

   The arithmetic mean is obtained by adding all observations and dividing the total by the number of observations.

2. **Indirect method or deviation method**

   In this method an arbitrary assumed mean is used as a basis for calculating deviations from individual values in the data set.

3 **Step deviation method**

   It is used when the deviations, $d = X - A$, have a common factor.

Illustration 1
Mr. Otieno has 5 children who go to school. The daily allowances to these children are 80, 150, 270, 285, and 200 respectively. What is the mean allowance for the children?

Arithmetic mean is the amount each individual would get if the total amount were divided equally among all the individuals in the distribution.

General Representation
If a variable, X, takes the values x_1, x_2, x_3..................x_n then the arithmetic means is given by:

$$\text{Arithmetic mean} = \frac{x_1+x_2+x_3+ \qquad +x_n}{n} = \frac{\sum_{i=1}^{n} xi}{n}$$

Where $\sum$ is the summation sign (Greek letter sigma), and n is the number of observations
$\sum xi$ means sum of x_i where i = 1, 2, 3, ---n.

Population arithmetic mean is given by

$$\mu = \frac{\sum_{i=1}^{n} Xi}{N}$$ where N is the number of elements in the population.

Sample arithmetic mean is given by

$$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n}$$ where n is the number of elements in the sample.

$$\overline{X} \rightarrow bar, \mu \rightarrow mew$$

Mean is the sum of all the observations divided by the total number of observations.
Illustration 2
The monthly salaries paid to employees in a firm are 200, 100, 300, 100, 400, 600, 550, 100 and 350. Determine the arithmetic mean.
Illustration 3
The monthly salaries paid to female employees in a company are: 400, 100, 300, and 600; while the monthly salaries paid to male employees are: 100, 200, 550, 100, and 350.

(i) Find the arithmetic mean of salaries paid to female employees.

(ii) Find the arithmetic mean of salaries paid to male employees.
(iii) Find the combined arithmetic mean of salaries paid to all employees.

Illustration 4

Find out the combined arithmetic mean from the following information: $\overline{X}_1 = 50, n_1 = 10, \overline{X}_2 = 30, n_2 = 5, \overline{X}_3 = 80, n_3 = 8.$

Illustration 5

The daily earnings of employees working on a daily basis in a firm are:

Daily earning	100	120	140	160	180	200	220
Number of employees	3	6	10	5	4	2	7

Determine the arithmetic mean.

Illustration 6

The human resource manager at a city hospital began a study of the overtime hours of the registered nurses. Twenty five nurses were selected at random and their overtime hours during a month were recorded:
13,13,12,15,7,15,5,12,6,7,12,10,9,13,12,9,6,10,5,6,9,6,9,12,5

Calculate the arithmetic mean of overtime hours during the month.

Illustration 7

A company is planning to improve plant safety. For this, accident data for the last 50 weeks was compiled. These data are grouped into the frequency distribution as shown below.

Number of accidents	0 - 4	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29
Number of weeks	5	10	12	13	8	2

Calculate the arithmetic mean of accidents per week.

Median

Is the value that divides data into two equal parts if such data has been arranged in ascending or descending order.

Median of ungrouped data

Find the median of 5, 6, 4, 7, 8,0, 7.
Find the median of 10, 7, 12, 10, 16, 15, 18, 17

Median for a simple frequency distribution.

Find the median of the following distribution:

Marks	No. of Students
3	3
5	4
6	1
7	5
9	2

Median for a grouped frequency distribution

Procedure

1. Sum all the frequencies (number of observations)
2. Obtain the position of the median by use of:

   $\frac{n}{2}$ if there is even number of observations.

   (i) $\frac{n+1}{2}$ if there is an odd number of observations.
3. Obtain less than cumulative frequencies.
4. Identify the class in which the median is located.
5. Determine the median using the formula

$$\text{Median} = \text{Lo} + \left(\frac{P - fo}{fm}\right) xC$$

Where

Lo is the lower class boundary of median class.

n is the total frequency

p is the position of the median

fm is the frequency of the median class

c is the class width of the median class.

fo cumulative frequency achieved just before the median class.

<u>Example:</u>

Obtain the median for the following grouped data.

Marks	0-19	20-39	40-59	60-69	70-79	80-89	90-99
No. of students	5	12	20	5	22	9	6

Mode

The mode of a distribution is the value that occurs most frequently i.e it is the most frequent/popular observation.

A distribution can be:

(i) Unimodal – if there is only one mode.

(ii) Bimodal – if there are two modes

(iii) Multimodal – if there are more than two modes.

Examples 1:

The marks of 6 students in a test are 25, 23, 47, 43, 23, 23. What is the mode ?

Example 2:

What is the mode for the following distribution?

Marks	5	20	25	30
Frequency	7	6	8	2

Mode of a grouped distribution

We can determine the mode of a grouped distribution either by graphical method or by Formula Method.

Formula Method

The mode is not rigidly defined.

1. $$\text{Mode} = Lo + \frac{\Delta 1}{\Delta 1 + \Delta 2} \times C$$

   Where Lo - Lower class boundary of modal class

   $\Delta 1$ is the difference between the frequency of the modal class and the frequency of the class that precedes the modal class.

   $\Delta 2$ is the difference between the frequency of the modal class and the frequency of the class that follows the modal class.

2. $$\text{Mode} = Lo + \frac{f_2}{f_1 + f_2} \times C$$

   where f1 – frequency of the class preceding the modal class.

   f2 – frequency of the class following the modal class

   fm – frequency of the modal class.

3. Mode = 3 median – 2 Arithmetic mean.

Example:

The ages of employees in a factory are as follows:

Age	11-20	21-30	31-40	41-50	51-60
No. of employees	3	35	40	25	2

<u>Position Measures: Quartiles, Deciles and Percentiles</u>

The median divides a data set into 2 equal parts if the data has been arranged in ascending, Median is a position measure.

<u>Quartiles</u>

Quartiles are position measures that divide a distribution (a set of data) into four equal parts. There are three quartiles, Q_1, Q_2 and Q_3.

Q1 is called the First Quartile or the Lower Quartile.

Q2 is the value which is halfway through the data and is therefore the median.Q2 is called the Second Quartile.

Q3 is the Third Quartile or Upper Quartile

Deciles

Are values that divide a distribution into ten equal parts. There are 9 deciles, D1, D2, D3---- D9.

D_1 is the first Decile

D_2 is the second Decile.

<u>Percentiles</u>

Are values that divide a distribution into a hundred equal parts. There are 99 percentiles; P_1, P_2, P_3--- P_{99}.

<u>Illustration 1</u>

Calculate the values of Q_1, Q_3, D_1, D_6, P_{30} and P_{75} from the following data.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No of students	5	15	33	65	76	69	49	35

<u>Illustration 2</u>

Calculate the values of Q_1, Q_3, D_1, D_6, P_{30} and P_{75} from the following data.

Marks	98-106	107-115	116-124	125-133	134-142	143-151	152-160	161-169
No of students	3	5	9	12	5	4	2	4

<u>Characteristics of a Good Measure of Central tendency.</u>

1. It should be rigidly defined. An average should be rigidly defined so that it has a defined and fixed value.
2. It should be easy to calculate and comprehend. For its wide applicability, an average should be simple to compute and easy to understand.
3. Its calculation should be based on all the observations. Since an average is a representative value of the whole mass of data, its calculations must involve all the observations.
4. It should not be much affected by a few extreme values of the observations. A few very small or large observations should not unduly affect an average. Otherwise, these extreme observations may distort the average value and it cannot be regarded as a representative value of the entire series. .
5. It should be capable of further algebraic treatment. This makes it useful in further statistical analysis i.e it should be amendable to mathematical treatment.