Table of Contents

Measures of central Tendency

Mathematics & Statistics for Business
(Statistics)
Lesson: Measures of central Tendency
Lesson Developer: Madhu Gupta
College/Department: Janki Devi Memorial College of Commerce, University of Delhi

2.1 Introduction

Fig. 2.1 Balance depicting measure of central tendency
Source: http://images.google.co.in/images?hl=en&source=hp&q=balance&btnG=Search+Images&gbv=2&aq=f&aqi=&aql=&oq=&gs_rfai=

After collection, classification and presentation of data we need to analyze and interpret the same. There are various numerical measures which describe the inherent characteristics of collected data.(see web link 1). The first such measure is "measure of central tendency" or **'average'** which gives us a single value which describes the characteristics of the entire group and facilitates easy expression, reference and interpretation. It reduces the complex mass of data into a single figure, which is "gist" if not the substance of the series.

Value addition: Did you Know?

Descriptive statistics

A frequency distribution or data set can be fully described by four main characteristics (called descriptive statistics). These are :

1. Measure of central tendency – measures a single value around which all or most of the values in the data set clusters.

2. Measure of dispersion – measures the extent to which individual values are spread.

http://www.techshout.com/software/2006/16/microsoft-excel-hit-by-new-vulnerability/

3. Measure of skewness – measures the extent of departure of individual values from symmetrical distribution. It describes the shape of the distribution. A distribution can be positively skewed, symmetrical or negatively skewed.

4. Measure of kurtosis– measures the height or peakedness of the distribution as compared to normal distribution.

2.1.1 Meaning

Measure of central tendency, also known as **average**, is one of the most widely used statistical device. It is that value of the distribution where most of the items of the series tend to cluster. Average is computed to obtain a single value which can represent a whole series of observations. A measure of central tendency, by condensing a mass of data in one single value, enables us to get a bird's eye view of the entire data and make comparisons. For example if the marks obtained by 1000 students of college A and 1000 students of college B are before us, it may not be possible for us to draw any inference from them or to compare the two series relating to these marks readily. On the other hand, comparison would become very easy if we get a single value to represent the marks of all the 1000 students of college A and another single value to represent the marks of all the 1000 students of college B. An average is used to represent a whole series and as such, its value always lies between the minimum and maximum value and generally it is **located in the centre** or middle of the distribution.

Value addition: Did you Know?

The word average is very commonly used in day to day conversations, for example, we often talk about pass percentage of a university or college, average height, average income of employees in an industry, life expectancy of an average male or female etc. In the world of business and finance, BSE SENSEX, NIFTY, wholesale and consumer price index, per capita income, percentage of growth in GDP, inflation, population, & rate of return on investment etc. are some of the commonly used averages.

Averages also appears daily on T.V., in the newspaper and other journals. Here are some interesting facts gathered from various published media:

1. Average daily footfalls in Ambience Mall in Gurgaon are 59356.

2. Average loan size of urban poor has increased from Rs.5000 in 2008 to Rs.10000 in 2009.

3. Average annual package offered by international banks to B-School Indian graduates is Rs. 15 lakhs.

4. Average net worth of top 25 Indian billionaires is Rs.8152.8 millions Dollar.

5. Life expectancy of an average Indian is only 66.09 years as compared to 77.5 years of an American.

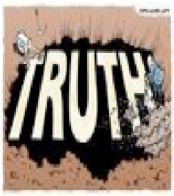
6 In 2009, fertility rate in India was 2.68 children born/ woman.

7 Sensex was 17933.14 on 9th April 2010.

8 In March 2010, food article price index rose to 17.7%.

Value addition: Interesting Facts!

Some interesting facts about average from everyday life:



Cats average 16 hours of sleep a day, more than any other mammal.

Every square inch of the human body has an average of 32 million bacteria on it.

In 1900 the average age at death in the US was 47.

On an average a woman speaks 7,000 words per day; Men manage with just over 2000.

On an average, 42,000 balls are used and 650 matches are played at the annual Wimbledon

tennis tournament.

On an average, right-handed people live 9 years longer than their left-handed counterparts.

Smokers are likely to die on an average six and a half years earlier than non-smokers.

The average person drinks about 16, 000 gallons of water in a lifetime.

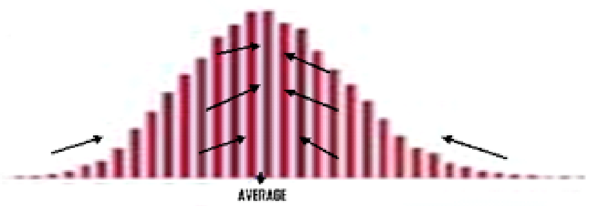
The average person walks the equivalent of twice around the world in a lifetime.

Source ;http://www.corsinet.com/trivia/average.html

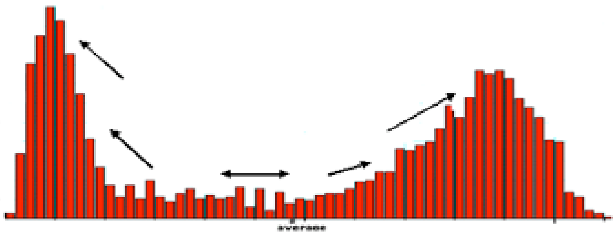
Picture source: http://images.google.co.in/images?hl=en&gbv=2&tbs=isch:1&q=cartoons+pictures&sa=N&start=40&ndsp=20

Average value should be such where most of the items of the series tend to **cluster**. In bell shaped distributions, this is true and frequencies tend to cluster around the average. But in case of J-shaped (skewed) and U- shaped distributions this is not true and frequencies do not cluster around average. In these cases proper care must be taken while interpreting its value.

Bell – Shaped distribution



U - Shaped distribution-



J - Shaped distribution

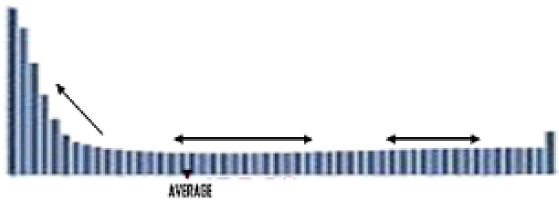


Fig. 2.2 Clustering of frequencies in different types of distribution

The averages are also called **measures of location** because they determine the position of the distribution on the axis of the variable.

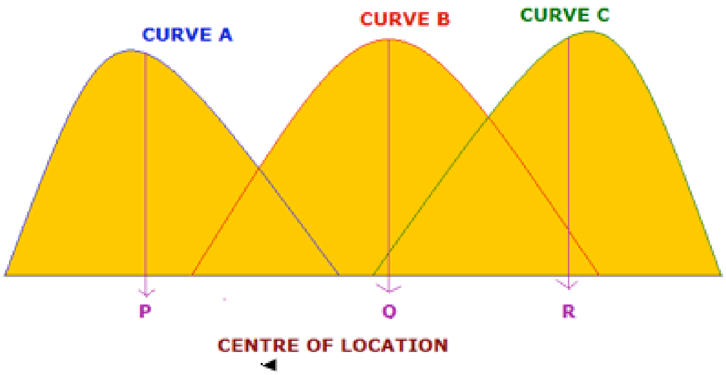


fig. 2.3   Centre of location of three distributions

2.1.2 Objectives Of Averaging

There are two main objectives of averaging**:**

1. **To get a single values to represent all the individual values in the series**

Average gives a bird's eye view of the huge mass of data, which are ordinarily not intelligible. It sets aside the unnecessary details of the data and aid the human mind in grasping the true significance of large aggregates of facts and measurements. Thus, the questions like 'how cold was Delhi in January this year', or 'what was the electricity consumption in June in a city' may be answered by a single figure of average temperature in January and average electricity consumption of the city in June. It is not necessary to give huge unwieldy set of numerical data of temperatures and electricity consumptions of different days.

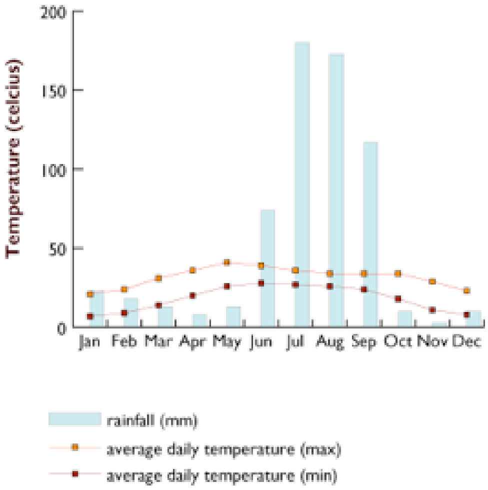


Fig 2.4- Average temperature of New-Delhi,2009

Source- http://www.bbc.co.uk/weather/world/city_guides/results.shtml?tt=TT002240

Value addition: Misconceptions

Average is a single value which represents a whole series of observations**.** However It does not mean all values of the distribution are uniform and are equal to average. Average always lies somewhere between the lowest and the highest values and proper care should be taken in interpreting its value

In this regard there is an interesting **old story** of a person who, along with his family of ten members, was going to a neighboring village. In between they had to cross a river whose average depth was understood to be 4 feet. He calculated the average height (mean) of the members of his family and found it to be 4 feet 8 inches. Since average height was more than the average depth of the river, he ordered his family to cross the river. Four members of his family drowned while crossing. He was very surprised and checked his calculations again which came out to be correct. He could not understand where he made the mistake and finally remarked-

'Ausat jyon ka tyon, kunba dooba kyon?'

(the average is the same , why has the family drowned)

In fact, average depth and average height does not mean uniform depth and uniform height. Individual values may be less or more than the average value.

Fig.Source-http://www.1st-art-gallery.com/(after)-Le-Moyne,-Jacques-(de-Morgues)/A-Family-Crossing-A-River.html

2. To facilitate comparison

Averages are very helpful for making comparative studies. For example, if the average of the marks of the students of college A is 72 % and that of college B is 78 %, we may say performance of the students of college B is better than that of college A. This would not be possible if we had two full series of marks of individual students of the two colleges. But this certainly doesn't mean that all students of college B are better than that of college A In both the colleges there may be certain students who must have scored less than the average marks and similarly there will be certain students who must have scored more than the average marks. We have to be very careful while making such comparisons and must consider other factors such as dispersion before drawing any inferences.

2.1.3. Characteristics Of A Good Average

There are different types of averages, and, the question is, which one should be used. Since it is to be used to represent all the values in a series it should satisfy certain properties.

First is should be easy to understand and calculate so that a person of ordinary intelligence easily understands it. Others wise its use will be confined to a limited number of persons. However convenience of computation should not be sought at the expense of accuracy.

Fig 2.5 A confused man

Source-http://images.google.co.in/images?hl=en&source=hp&q=statistics&btnG=Search+Images&gbv=2&aq=f&aqi=&aql=&oq=&gs_rfai

Second, it should be rigidly defined so that it is subject to one and only one interpretation. An average should be a precise and definite value. The user may not introduce any bias i.e. it is not left to the mere estimation of the observer, because in that case it will not be a true representative of the series. It will be preferable if it can be defined by an algebraic formula so that if different people compute the average from the same figures, they all get the same value.

Third, it should be based on all the item of the series. If some of values of the series are not taken into account in computing an average, it cannot be said to be a true representative of the whole series.

Fourth, through an average should be based on all the items of the series, it should not be unduly affected by extreme observations. If one or two very small or very large items are able to influence very much the value of the average then such average will not be the true representative of the series

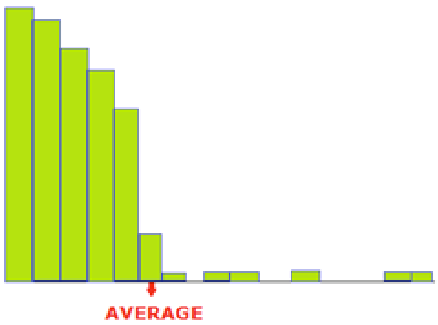


Fig 2.6 A few bigger items influencing the average

Last, it should have sampling stability and should be capable of further algebraic treatment. For this average should have a well built mathematical model so that different persons calculating the same average from the different samples of the same population should get approximately the same value. No doubt, averages of different samples are rarely the same, but one form of average may show much greater differences in the values of two samples than another. Of the two, that one in better in which this difference, technically called fluctuations' of sampling is less, otherwise it will have limited application and utility.

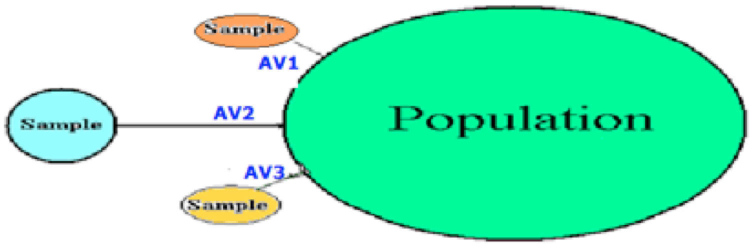


Fig 2.7 Population and its sample's average

It should be noted that no single average satisfies all the above properties and is sailable for all practical purposes. A suitable selection of an average **depends on** the nature of data and purpose of enquiry.

2.1.4. Types Of Averages

The different averages which we will study here are arithmetic mean, median, mode, geometric mean and harmonic mean. These averages are often categorized as **mathematical averages or positional averages**. The averages like arithmetic mean geometric mean and harmonic mean are called mathematical averages since they are computed with the help of a well built mathematical formula and are based on each and every value of the variable. Median and mode are called positional averages because their values are determined on the basis of their position in the distribution and not by the size of the items.

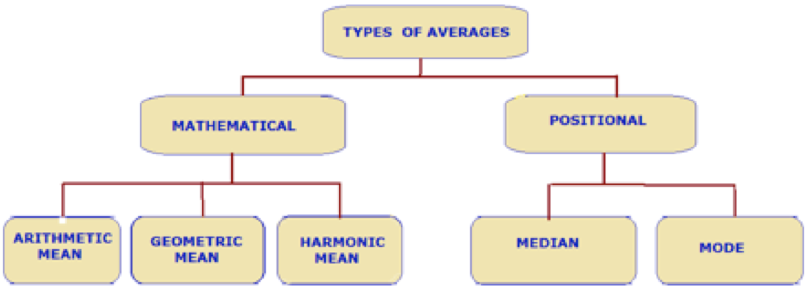


Fig 2.8 Mathematical & positional averages

2.2. Arithmetic mean

Arithmetic mean is the most important and widely used average. Most of the time, when we refer to the average of something, we are talking about Arithmetic mean. This is true in cases such as average income, average marks, and average temperature of a city. It is equal to the total of all the values in a series divided by the number of items in that series.

2.2.1. Calculation Of Arithmetic Mean

Direct method.

If X_1, X_2 ----- X_n are the given n observations in a sample, then their arithmetic mean denoted by $\overline{X}$ is given by

$$\overline{X} = \frac{X_1 + X_2 + ---- X_n}{n} = \frac{\Sigma X}{n}$$

Where ΣX is the sum of the observations.

Short cut method

When the values are large and the number of items is high, a short – cut method may be used to reduce the calculations. Under this method, one value is taken as an arbitrary origin or assumed mean, deviations of individual values are taken from that assumed meaan and then arithmetic mean is calculated by using the following formula:

$$\overline{X} = A + \frac{\Sigma d}{n}$$

Where, A is assumed mean,

n the number of items and

Σd is the sum of the deviations of individual values from assumed mean $[\Sigma \quad (X - A)]$.

B. Discrete series

Direct method

Arithmetic mean of discrete series is obtained by multiplying each value of the variable by its corresponding frequency totaling them up and dividing that total by the total frequency. Formula used for arriving at the arithmetic mean is as follows;

$$\overline{X} = \frac{\Sigma fX}{\Sigma f}$$

Where f stands for frequency and Σf = N

Short cut method

Under this method, the deviations of items from an assumed mean are taken and then they are multiplied by their respective frequencies. Arithmetic mean is found out by using the following formula:

$$\overline{X} = A + \frac{\Sigma fd}{\Sigma f}$$

Where A = assumed mean

Σfd = total of the products of deviations from the assumed mean and the corresponding frequencies.

Illustration: From the following frequency distribution find out the mean salary of workers.

Salary (in 000' Rs):	60	61	62	63	64	65	66
No. of workers :	2	09	10	15	07	04	03

Solution: Calculation of mean

Salary (in 000Rs)	No. of Workers	f X	X-A (A=63)	fd

X	F		d	
60	2	120	-3	-6
61	9	549	-2	-18
62	10	620	-1	-10
63	15	945	0	0
64	7	448	1	7
65	4	260	2	8
66	3	198	3	9
	f = 50	fx = 3140		Σfd = -10

$$\overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{3140}{50} = 62.8 \quad (000\ Rs)$$

Alternatively,

$$\overline{X} = A + \frac{\Sigma fd}{\Sigma f}$$

$$\overline{X} = 63 + \frac{-10}{50}$$

$$= 63 - (.20) = 62.8 \quad (00\,0\ Rs)$$

Thus mean salary of the workers is Rs 62800.

C. Continuous series

Direct method

First, mid-point of each class is taken as the representative figure of that particular class. Then we multiply each mid-point by the corresponding frequency and apply the following formula to find out the arithmetic mean:

$\overline{X} = \Sigma f X / \Sigma f$ here X denotes mid points of various classes.

Short-cut method

In this method, deviations of mid-points from an arbitrary origin are taken. The formula applied to calculate mean is

$$\overline{X} = A + \frac{\Sigma fd}{\Sigma f}$$

Where A is assumed mean

$\Sigma fd = \Sigma f (X - A)$ and

X is the mid point of different classes.

Step-deviations method

This is an extension of the short cut method used for further simplification. Where figures of deviations are divisible by a common factor, we reduce them by dividing them by that common factor and then multiply their sum by the same common factor. Thus

$$\overline{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$$

Where $d' = \frac{X - A}{C}$ and

C is common factor

Illustration: Following is the data of overtime hours worked by 100 employees of GPL ltd.. You are required to calculate arithmetic mean from the following data by using direct, short-cut and step deviation method.

Overtime (hrs)	0-10	10-20	20-30	30-40	40-50
No. of employers	15	25	30	20	10

Fig 2.9 Employees working

Solution:

Solution:

Over Time (hrs)	No. of employers f	Mid points X	fx	d = X – A (A = 25)	fd	$d' = \frac{X-A}{C}$ (c = 10)	fd'
0-10	15	5	75	-20	-300	-2	-30
10-20	25	15	375	-10	-250	-1	-25
20-30	30	25	750	0	0	0	0
30-40	20	35	700	10	200	1	20
40-50	10	45	450	20	200	2	20
	Σf = 100		Σfx =2350		Σfd = -150		Σfd'=-15

$$Direct \quad method \quad \overline{X} = \frac{\Sigma fX}{\Sigma f} = \frac{2350}{100} = 23.5 \quad hours$$

$$Short-cut \quad method \quad \overline{X} = A + \frac{\Sigma fd}{\Sigma f} = 25 + \frac{-150}{100}$$

$$= 25 - 1.5 = 23.5 \quad hours$$

$$step-deviation \quad method \quad \overline{X} = A + \frac{\Sigma fd'}{\Sigma f} \times C$$

$$= 25 + \frac{-15}{100} \times 10$$

$$= 25 - 1.5 = 23.5 \quad hours$$

The result obtained in case of continuous series will be only approximate of the actual average. This is because we do not know every data point in the sample. We assumed that every value in a class was equal to its mid point.

2.2.2 Mathematical Properties Of Arithmetic Mean

Arithmetic mean has some very important mathematical properties which are as follows:-

1. if the number of items and their aritmetic mean are both know the aggregate or the sum of items can be obtained by multiplying the average by the number of items.

$$\Sigma X = n.\overline{X}$$

this property of mean in known as "Total value" property and is not possessed by any other measure of central tendency.

Illustration: The average monthly sales for the first 11 months of the year on respect of a certain salesman were rs 24000 but due to his illness during the last month the average sales fr the whole year came down to Rs. 22,750. What was the value of his sales during the last month.

Solution: $\Sigma X = \overline{X}$

or total sales = Average sales x no. of months.
Now let the sales during the last month br K.

Then the total sales of 11 month = 24000 x 11 = 264000 Rs and total sales of 12 months (whole year) = 264000 + K.

Average of the whole year = 22,750 (given)

$$Average \quad sales = \frac{Total \quad sales}{No. \quad of \quad months}$$

Or $$22750 = \frac{264000 + k}{12}$$

Or 273000 = 264000 + K

Or k = 9000 Rs

sales of last month are Rs 9000

Illustration: The mean of 40 observations wae 160. It was detected on rechecking that two values 105 and 180 were wrongly copied as 125 and 120 for the computation of mean. Find the correct mean.

$$Correct \quad mean = \frac{Correct \quad Total}{Correct \quad no. \quad of \quad items}$$

$$= \frac{Incorrect \quad total - wrong \quad items + Correct \quad itmes}{Correct \quad no. \quad of \quad items}$$

Soultions:

Here,

Incorrect sum of 40 observations = 160 x 40 = 6400

Correct sum of 4 observations = 6400 - 125 - 120 + 105 +108 = 6440

Correct mean = $\frac{6400}{40} = 160$

2. The sum of the deviations of the individual items from their arithmetic mean is always zero. i. e.

$$\Sigma\,(X - \overline{X}) = 0$$

For example, the arithmetic mean of 15, 20, 25, 30 and 35 is 25. if the difference of each of these items is calculated it whould be -10, -5, 0, +5 and + 10. Their total is zero.

Due to this property the mean is also called point of balance i.e. at this point (mean) sum of the positive deviations from mean is equal to the sum of the negative deviations from mean.

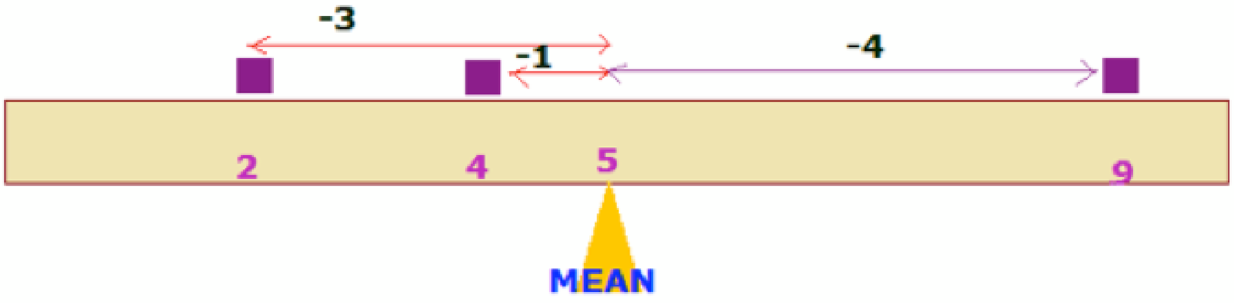


Fig 2.10 Arithmetic mean as a balance point

3. The sum the squared deviations of the items from arithmetic mean is minimum i.e. it is less that the sum of the squared deviations of the items from any other value.

$$\Sigma\,(X - \overline{X})^2 < \Sigma\,(X - A)^2$$

Where A is any other value except mean. Therefore, we call mean as the **least squares measure** of central tendency. This is property of mean is also known as the least square property.