Elementary Statistics

By Prof. Mirza Manzoor Ahmad

Measures of Central tendency

In statistics, a **central tendency** (or, more commonly, a **measure of central tendency**) is a central or typical value for a probability distribution.It may also be called a **center** or **location** of the distribution. Colloquially, measures of central tendency are often called *averages*. The term *central tendency* dates from the late 1920s.The most common measures of central tendency are the arithmetic mean, the median and the mode.

Properties of a Good Average or Measure of central tendency.

According to Prof.Yule, the following are the properties that an ideal average or measure of central tendency should possess.

(i) It should be rigidly defined.

(ii) It should be easy to understand and calculate.

(iii) It should be based on all the observations.

(iv) It should be suitable for further mathematical treatment.

(v) It should be affected as little as possible by fluctuations of sampling.

(vi) It should not be affected much by extreme observations.

Arithmetic Mean Arithmetic mean is a number which is obtained by adding the values of all the items of a series and dividing the total by number of items.

Example:

There are six kindergarten classrooms in a small school district in Florida. The class sizes of each of these kindergartens are 26, 20, 25, 18, 20 and 23. A researcher writing a report about schools in her town wants to come up with a figure to describe the typical kindergarten class size in this town. She asks a friend for help and her friend suggests her to calculate the average of these class sizes.

To do this, the researcher finds out that she needs to add the kindergarten class sizes together and then divide this sum by six, which is the total number of schools in the district. Adding the six kindergarten class sizes together gives the researcher a total of 132. If she then divides 132 by six, she gets 22. Therefore, the average kindergarten class size in this school district is 22.

$$Average = \frac{26+20+25+18+20+23}{6} = \frac{132}{6} = 22$$

This average is also known as the arithmetic mean of a set of values.

Merits of Arithmetic Mean

- It is easy to understand and calculate.
- It is rigidly defined.
- It is based on all observations of the series.
- It is used for further algebraic manipulations.

Demerits of Arithmetic Mean

- It is too much affected by extreme values.
- Its graphical presentation is not possible.
- It cannot be used in qualitative information.

- Sometimes it gives bias results because it assigns more weight to bigger items.

Computation of Mean

We generally use two methods in calculation of the mean (i) Direct Method and (ii) Short cut Method.

Direct Method to Calculate Mean

Individual Series. In this method, we calculate mean by adding all the values of the items and then dividing the aggregate by the total number of items.

Thus, if there are a total of *n* numbers in a data set whose values are given by a group of *x*-values, then the arithmetic mean of these values, represented by 'm', can be found using this formula:

$$m = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

To be precise, the formula can be symbolically put as,

$m = \frac{\sum X}{N}$ where,

m means arithmetic mean

$\sum$ X stands for sum of size of items.

N means number of items.

In our kindergarten class size example, *n* is 6, or the number of kindergarten classrooms, while the *x*-values are given by the class sizes in each of the kindergartens within the school district. If you recall adding the total number of students in the six classrooms gave us 132. We can plug these values into our formula, dividing 132 by six, and find once again that the average class size is 22.

Short cut Method

In this method, we assume an arbitrary figure as mean and then the deviations of the items of the series from the assumed average are taken. Then we divide the sum of these deviations by the number of items and by adding the assumed average to it, the actual average is obtained.

Symbolically,

m = A + $\frac{\sum dx}{N}$ where

m = Arithmetic Mean

A = Assumed Mean

$\sum d_x$ = Sum of deviations

N = Number of items.

Problem.

Calculate the mean using the both Direct Method and Short cut Method.

5,10,15,20,25.

Sol.

(Direct Method)

X: 5,10,15,20,25.

$\sum X: 5+10+15+20+25$

$= 75$

N = 5

$m = \frac{\sum X}{N}$

$m = \frac{75}{5}$

$= 15.$

Sol.
(Short cut Method)
Let A=10

X	d_x /(X-A)
5	-5
10	0
15	5
20	10
25	15
N=5	25

Now A=10,$\sum d_x$= 25 and n =5
Substituting these values in the formula, we get

$$m = A + \frac{\sum dx}{N}$$
$$= 10+\frac{25}{5}$$
$$= 10+5$$
$$= 15$$

Discrete Series. In the discrete series, to get the arithmetic average we multiply the frequencies by their respective size of the items and summing up the products we divide by the sum of the frequencies.

Direct Method

$$m= \frac{\sum fx}{N}$$

Short-cut-Method

$$m=A+\frac{\sum fdx}{N}$$

Problem.
Compute the mean from the following data.
X: 5 10 15 20 25
f : 3 1 2 5 4
Sol.
(Direct Method)

X (1)	**f (2)**	**fX (3)**
5	3	15
10	1	10
15	2	30
20	5	100
25	4	100
	15	**255**

$$m = \frac{\sum fX}{N}$$
$$= \frac{255}{15}$$
$$= 17$$

(Short -cut- Method)
Let A be 15

X (1)	**f (2)**	**d_x (3)**	**fd_x (5)**
5	3	-10	-30
10	1	- 5	- 5
15	2	0	0
20	5	5	25
25	4	10	40
	15		**30**

$$m = A+\frac{\sum f dx}{N}$$
$$= 15+\frac{30}{15}$$
$$= 15+2$$
$$= 17$$

Steps for the computation of Arithmetic Mean in case of Frequency Distribution.

- Locate the mid value of the class (Grouped Data) by summing up the lower limit and the upper limit of the class and divide the result by 2.
- Multiply each value of X or the mid value of the class (in case of grouped data/continuous frequency distribution) by its corresponding frequency f.
- Obtain the sum of the products as obtained in step second above to get ∑fx.
- Divide the sum obtained in step third by N=∑f, the total frequency.
- The resulting value gives the arithmetic mean.

Note: To take deviations of the size from the assumed mean and proceed with other steps as desired by the concerned formula's, follow the same procedure as is employed in earlier calculations.

Problem.
Calculate the Arithmetic Mean from the following data.
X : 11 - 13, 13 – 15, 15 – 17, 17 – 19, 19 -21, 21- 23, 23 - 25.
f : 3 4 5 6 5 4 3
Sol.
(Direct Method)

X (1)	f (2)	x **(mid value)** (3)	f x (4)	d_x (5)	fd_x (6)
11-13	3	12	36	-6	-18
13-15	4	14	56	-4	-16
15-17	5	16	80	-2	-10
17-19	6	18	108	0	0
19-21	5	20	100	2	10
21-23	4	22	88	4	16
23-25	3	24	72	6	18
	30		**540**		**0**

$$m = \frac{\sum f x}{N}$$
$$= \frac{540}{30}$$
$$= 18$$

(Short-cut- Method)

$m = A+\frac{\sum fdx}{N}$

$= 18+\frac{0}{30}$

$= 18+0$

$= 18$

Note: Represent the desired columns only in a tabular form.

Step Deviation Method.

To simplify the calculations, deviation from assumed mean are divided by a common factor (i.e., the class interval) if that is of the same magnitude throughout all the sizes of the series. The total of the products of deviations from the assumed mean and frequencies are multiplied by this common factor and divided by the sum of the frequencies and added to the assumed average.

$m= A+\frac{\sum fdx}{N} \times i$ where

i stands for class interval

Problem.

Calculate the Arithmetic Mean from the following data using Step Deviation Method.

X : 11 - 13, 13 – 15, 15 – 17, 17 – 19, 19 -21, 21- 23, 23 - 25.

f : 3 4 5 6 5 4 3

Sol.

X (1)	f (2)	x (mid value) (3)	dx $(x-A)/i$ (4)	fdx (5)
11-13	3	12	-3	-9
13-15	4	14	-2	-8
15-17	5	16	-1	-5
17-19	6	18	0	0
19-21	5	20	1	5
21-23	4	22	2	8
23-25	3	24	3	**9**
	30			**0**

$m= A+\frac{\sum fdx}{N} \times i$ where i=2

$=18+\frac{o}{30}\times 2$

$= 18+0$

$= 18$

<u>MEDIAN</u>

Median is the value of the middle item of a series arranged in an ascending or a descending order of magnitude.

Merits:

- It is very easy to calculate.
- Its value is not much affected by extreme items.
- It satisfies most of the conditions of an ideal average.
- It can be determined graphically.

Demerits:

- It is not suitable for further mathematical treatment.
- It involves additional work of arranging data in ascending or descending order.
- It gives very little importance to extreme values.

Computation of Median

Case (i) Ungrouped Data.

To locate median in the case of ungrouped data, first array the data in ascending or descending order. If the data set contains an odd number of items, the middle item of the array is the median. If there are even numbers of items, the median is the average of the two middle items.

Median = Size of $\left(\frac{N+1}{2}\right)$th item.

The following given solved problems illustrate the method.

Case (ii) Frequency Distribution.

In case the variable takes the value with respective frequencies, median is the size of the (N+1)/2th item. In this case use of cumulative frequency (c.f.) distribution facilitates the calculations. The steps involved are:

- Prepare the <u>less than</u> cumulative frequency distribution.
- Find N/2
- See the c.f. just greater than N/2
- The corresponding value of the variable gives median.

The following given solved problems illustrate the method.

Case (iii) Continuous frequency Distribution.

In case of continuous series,the steps involved are as follows:

- Prepare the <u>less than</u> cumulative frequency distribution.
- Find N/2
- See the c.f. just greater than N/2
- The corresponding class contains the median value and is called the median class.

The value of median is now obtained by using the interpolation formula :

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

Where

l is the lower limit of the median class.

f is frequency of the median class.

h is the magnitude of the median class.

N is summation of the frequencies.

c is c.f. of the class preceding the median class.

Problem.

Find median for the following data:

4,1,6, 3,5.

Sol.

S.No.	Ascending order	Descending order
1.	1	6
2.	3	5
3.	4	4
4.	5	3
5.	6	1

Median = Size of $\left(\frac{N+1}{2}\right)$th item.

Median = Size of $\left(\frac{5+1}{2}\right)$th item.

Median = Size of $\left(\frac{6}{2}\right)$th item.

Median = Size of 3^{rd} item. Since the size of 3^{rd} item is 4, therefore median is 4.

Problem.

Find median for the following data:

9, 3, 8,2,5,1.

Sol.

S.No.	Ascending order
1.	1
2.	2
3.	3
4.	5
5.	8
6.	9

Median = Size of $\left(\frac{N+1}{2}\right)$th item.

Median = Size of $\left(\frac{6+1}{2}\right)$th item.

Median = Size of $\left(\frac{7}{2}\right)$th item.

Median = Size of 3.5th item.

Since Median is the size of 3.5th item in the array, so we are required to determine the average of 3rd and 4th item values.

Size of 3rd item is 3 and that of 4th item is 5, therefore the average of the two values works out to be:

$$(3+5)/2=8/2 =4$$

Therefore, median is 4.

Problem.

Calculate Median for the following distribution:

X: 2, 3, 4, 5, 6, 7.

f : 2, 3, 9, 21, 11, 5.

Sol.

X	f	c.f.
2	2	2
3	3	5
4	9	14
5	21	35
6	11	46
7	5	51

Median = Size of $\left(\frac{N+1}{2}\right)$th item

Median = Size of $\left(\frac{51+1}{2}\right)$th item

Median = Size of $\left(\frac{52}{2}\right)$th item

Median = Size of 26thitem

c.f.next higher to 26 is 35

Therefore median=5.

Problem.

Find median from the data given below:

Marks : 0-10,10-20,20-30,30-40,40-50,50-60.

No. of Students : 12 18 27 20 17 6

Sol.

X	x	f	c.f.
0-10	5	12	12
10-20	15	18	30
20-30	25	27	57
30-40	35	20	77
40-50	45	17	94
50-60	55	6	100

Median = size of N/2th item.
= size of 100/2th item.
= size of 50^{th} item.

c.f. just greater than 50 is 57 and it represents class interval 20-30

Median = l+ $\frac{h}{f}\left(\frac{N}{2}-c\right)$

Median = 20+ $\frac{10}{27}\left(\frac{100}{2}-30\right)$

Median = 20+ $\frac{10}{27}(50-30)$

Median = 20+ $\frac{10}{27}(20)$

Median = 20+7.407 =27.41

Mode

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster densely.

Merits

- It is easily understood.
- It is not affected by extreme observations.
- It can be easily calculated simply by inspection.

Demerits

- It is not based on all the observations of a series.
- It is ill defined.
- It is affected to a great extent by sampling fluctuations in comparison with mean.

Methods of Estimating Mode

Generally the following methods are used estimating mode of a series.

- Locating the most frequently repeated value in the array.
- Estimating the mode by interpolation.
- Estimating the mode from the median and the mean.

Computation of Mode

Problem.

Find out mode:

2, 5, 6, 5, 9, 3.

Sol. (By Inspection)

Since the size 5 occurs maximum number of times (i.e.twice), hence mode is 5.

Problem

Find out mode:

2, 5, 6, 5, 9, 3, 2.

Sol. (By Inspection)

Since the two sizes (i.e., 2 and 5) repeat maximum but equal number of times (i.e.twice), hence mode is ill defined.

Problem

Find out mode:

X : 1, 2, 3, 4, 5, 6, 7, 8.

F :2, 9, 3, 4, 8, 7, 8, 3.

Sol. (Grouping Method)

X	(i)	(ii)	(iii)	(iv)	(v)	(vi)
1	2					
		11				
2	9			14		
			12			
3	3				16	
		7				
4	4					15
			12			
5	8			19		
		15				
6	7				23	
			15			
7	8					18
		11				
8	3					

Method : 1. The frequency of each items is written in col. (i)
2. They are added in two's at a time in col. (ii) and col.(iii)
3. They are added in three's in columns (iv),(v) and (vi).
4. The frequency which does not fall in the group is left free.

ANALYSIS TABLE

Column	Size of item having max.frequency					
(i)	2					
(ii)			5	6		
(iii)				6	7	
(iv)		4	5	6		
(v)			5	6	7	
(vi)				6	7	8
Total	**1**	**1**	**3**	**5**	**3**	**1**

It follows from the table that the size 6 occurs large number of times. Therefore, mode is 6.

Problem.

Find out mode using interpolation method.

X : 0-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70,70-80.

f : 2 9 3 4 8 7 8 3

Sol. (Interpolation Method)

X	(i)	(ii)	(iii)	(iv)	(v)	(vi)
0-10	2					
		11				
10-20	9			14		
			12			
20-30	3				16	
		7				
30-40	4					15
			12			
40-50	8			19		
		15				
50-60	7				23	
			15			
60-70	8					18
		11				
70-80	3					

ANALYSIS TABLE

Column Size of item having max.frequency

X	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80.
(i)		10-20						
(ii)					40-50	50-60		
(iii)						50-60	60-70	
(iv)				30-40	40-50	50-60		
(v)					40-50	50-60	60-70	
(vi)						50-60	60-70	70-80
Total	**0**	**1**	**0**	**1**	**3**	**5**	**3**	**1**

It follows from the table that the size 50-60 occurs large number of times. Therefore, 50-60 is modal class.

By interpolation,

$Z=l+\left(\frac{f-t}{2f-t-u}\times i\right)$

where,

Z stand for mode, l for the lower limit of modal class, f for frequency of modal class, t for frequency of the class preceding modal class, u frequency of the class succeeding modal class and i for class interval.

$Z = 50+\left(\frac{7-8}{2(7)-8-8}\times 10\right)$

$Z = 50+\left(\frac{-1}{14-16}\times 10\right)$

$Z = 50+\left(\frac{-1}{-2}\times 10\right)$

$Z = 50+5$

$Z = 55$

Note: This method is used in case of continuous frequency distributions. First of all a modal class is determined. A modal class is the class in which mode of series lies. Having determined the modal class, the next issue will be to interpolate the value of the mode within this modal class as illustrated in the above example.

Locating mode from mean and median
Problem
Given, Mean = 15 and median =16, find out the mode?
Sol.
Mode =3median-2Mean
Mode= 3(16)-2(15)
Mode= 48-30
Mode= 18.

MEASURES OF DISPERSION

Dispersion

Dispersion is defined as the extent of scatteredness of items around a measure of central tendency. The objective of measuring scatteredness is to obtain a single summary figure which exhibits the extent of the scatteredness of the values.

Absolute and relative dispersions

Dispersion is said to be in absolute form when it states the actual amount by which the value of an item on an average deviates from a measure of central tendency. Absolute measures are expressed in concrete units i.e. the units in terms of which the data has been expressed.

A relative measure of dispersion is obtained by dividing the absolute measure by a quantity in respect of which absolute deviation has been computed. It is usually expressed in a percentage form. It is used for making comparisons between two or more distributions.

Range, Quartile Deviation, Mean Deviation and Standard Deviation come under the classification of absolute measures while as coefficients of Range, Quartile Deviation, Mean Deviation and variation represent relative measures of dispersion.

Mean Deviation

Mean Deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency.

Merits

- It is rigidly defined.
- It is easy to calculate.
- It is based on all the observation of a series.
- It is less affected by the presence of extreme items.

Demerits

- It ignores signs which are seriously objectionable.
- It is not capable of further algebraic treatment.

Computation of Mean Deviation

Problem

Calculate Mean Deviation and its co-efficient about mean for the following data:

90,160,200,360,400,500,600,650.

Sol.

X	Deviation from Mean d_x or /D/ (X-Mean)
90	280
160	210
200	170
360	10
400	30
500	130
600	230
650	280
2960	1340

Mean = $\frac{\sum X}{n} = \frac{2960}{8} = 370$

Mean Deviation about Mean $= \frac{\sum /D/}{N}$

Mean Deviation about Mean $= \frac{1340}{8} = 167.5$

Coefficient of mean deviation about mean $= \frac{mean\ deviation\ about\ mean}{mean}$

$= \frac{167.5}{370} = 0.453$

Problem

Calculate Mean Deviation about mean and its coefficient for the following data:

Marks	No. of students
0-10	5
10-20	8
20-30	15
30-40	16
40-50	6

Sol.

X	f	x	step devia- -tions from A=25 (d_x)	fd_x	/D/	f/D/
0-10	5	5	-2	-10	22	110
10-20	8	15	-1	-8	12	96
20-30	15	25	0	0	2	30
30-40	16	35	1	16	8	128
40-50	6	45	2	12	18	108
	50			10		472

Mean = $A + \frac{\sum fdx}{n} \times i$

$= 25 + \frac{10}{50} \times 10$

$= 25 + 2$

$= 27$

Mean Deviation about mean $= \frac{\sum f/D/}{N}$

$= \frac{472}{50} = 9.44$

Coefficient of mean deviation about mean $= \frac{mean\ deviation\ about\ mean}{mean}$

$= \frac{9.44}{27} = 0.349$

Steps : Case (i) Individual Series

- Calculate any measure of central tendency for the data or as is desired.
- Follow any of the three methods i.e. Direct, Short cut or Step Deviation method for the calculation of average.
- Take the deviations of the size from the computed measure of central tendency.

- Sum up the deviations and divide the result by the number of items. The figure, thus, obtained is mean deviation. However ± signs are ignored in this measure of dispersion.

Case (ii) Frequency Distribution

Calculate any measure of central tendency for the data or as is desired. Follow any of the three methods i.e. Direct, Short cut or Step Deviation method for the calculation of average.

Take the deviations of the size from the computed measure of central tendency.

Multiply the deviations taken with their respective frequencies.

Sum up the product of deviations and frequencies .Divide the result by the number of items. The figure, thus, obtained is mean deviation.

Standard Deviation

Standard Deviation is the square root of the arithmetic mean of the square of deviations of the items.

Merits

- It is rigidly defined.
- It is based on all observation of the series.
- It is suitable for further algebraic treatment.
- It is not readily comprehended.
- It pays more weightage to extreme values.

Computation of S.D.

Direct Method

In the calculation of S.D. mean is calculated and the deviations are taken from the mean and squared deviations are summed up and the total is divided by the number of items and the square root of resulting figure gives us the S.D.of the series. The formulas used are as follows:

In case of individual Series:

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum d^2_x}{N}}$$

In case of discrete and continuous Series:

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum fd^2_x}{N}}$$

Where,

$\boldsymbol{\sigma}$ = Standard deviation.

$\sum d^2_x$ = Sum of squares of deviation taken from mean.

$\sum fd^2_x$= Sum of products of squared deviations and freq.

N = Sum of frequencies.

Short cut method

In this method, deviations are taken from the assumed mean and then squared and divided by the number of items. From this figure, we subtract the square of the mean of the deviations from the assumed mean. The square root of the resulting figure would give us the standard deviation.

Individual Series

$$\boldsymbol{\sigma} = \sqrt{\frac{\sum d^2_x}{N} - \left[\frac{\sum d_x}{N}\right]^2}$$

Discrete and Continuous Series

$$\sigma = \sqrt{\frac{\sum fd^2_x}{N} - \left[\frac{\sum fd_x}{N}\right]^2}$$

***Step Deviation Met*hod.** In continuous series, we can also use the step deviation method. The formula is as follows:

$$\sigma = \sqrt{\frac{\sum fd^2_x}{N} - \left[\frac{\sum fd_x}{N}\right]^2} \times i$$

Variance

The square of standard deviation is called variance and is denoted by $\sigma 2$

Merits

- It is based on all the observations.
- It is not much affected by the fluctuations of sampling and is therefore useful in sampling theory test of significance.

Demerits

- It gives more importance to extreme observations.
- Since it depends upon the units of measurement of the observations, it cannot be used for comparing the dispersion of the distributions expressed in different units.
- It is difficult to understand and calculate.

Coefficient of variation

Standard deviation is the only absolute measure of dispersion, depending upon the units of measurement. The relative measure of dispersion based on standard deviation is called the coefficient of standard deviation. According to Prof. Karl Pearson, "coefficient of variation is the percentage variation in mean, standard deviation being considered as the total variation in the mean."

Coefficient of Variation $= \frac{standard\ Deviation}{Mean} \times 100$

Practical Problems

Calculate S.D. for the following data:

X : 30,40 ,42,44,46,48,58.

Sol. Direct Method

X	d_x (X-mean)	d^2_x
30	-14	196
40	- 4	16
42	- 2	4
44	0	0
46	2	4
48	4	6
58	14	196
308		432

$m = \frac{\sum X}{n}$

$m = \frac{308}{7}$ = 44.

$$\sigma = \sqrt{\frac{\sum d^2_x}{N}}$$

$$\sigma = \sqrt{\frac{432}{7}} = \sqrt{61.7143} = 7.856$$

Sol.(Short cut Method)

X	d_x (X-A)	d^2_x
30	-14	196
40	- 4	16
42	- 2	4
44	0	0
46	2	4
48	4	16
58	14	196
308	0	432

Let A= 44

$$\sigma = \sqrt{\frac{\sum d^2_x}{N} - \left[\frac{\sum d_x}{N}\right]^2}$$

$$\sigma = \sqrt{\frac{432}{7} - \left[\frac{0}{7}\right]^2}$$

$$\sigma = \sqrt{61.714 - 0} = \sqrt{61.7143} = 7.856$$

PROBLEM

Calculate S.D. for the following data:

X = 12 13 14 15 16 17 18 20

f = 4 11 32 21 15 8 5 4

Sol. Direct Method

X	f	fX	d_x (X-m)	d^2_x	fd^2_x
12	4	48	-3	9	36
13	11	143	-2	4	44
14	32	448	-1	1	32
15	21	315	0	0	0
16	15	240	1	1	15
17	8	136	2	4	32
18	5	90	3	9	45
20	4	80	5	25	100
	100	1500			304

$$m = \frac{\sum Fx}{N}$$

$$m = \frac{1500}{100} = 15$$

$$\sigma = \sqrt{\frac{\sum fd^2_x}{N}}$$

$$\sigma = \sqrt{\frac{304}{100}} = \sqrt{3.04} = 1.74$$

Sol. (Short cut Method)

X	f	d_x (X-A)	fd_x	d^2_x	fd^2_x
12	4	-3	-12	9	36
13	11	-2	-22	4	44
14	32	-1	-32	1	32
15	21	0	0	0	0
16	15	1	15	1	15
17	8	2	16	4	32
18	5	3	15	9	45
20	4	5	20	25	100
	100		0		304

$$\sigma = \sqrt{\frac{\sum fd^2_x}{N} - \left[\frac{\sum fd_x}{N}\right]^2}$$

$$\sigma = \sqrt{\frac{304}{100} - \left[\frac{0}{100}\right]^2} = \sqrt{3.04} = 1.74 \text{ Ans.}$$

Problem

Calculate standard Deviation, Variance and coefficient of variation for the following data using Step Deviation and Short cut Method:

X : 0-10, 10-20, 20-30, 30-40, 40-50

f : 10 12 17 14 5

Sol.(Short cut Method)

X	x	f	dx	fd_x	d^2_x	fd^2_x
0-10	5	10	-20	-200	400	4000
10-20	15	12	-10	-120	100	1200
20-30	25	17	0	0	0	0
30-40	35	14	10	140	100	1400
40-50	45	5	20	100	400	2000
		58	0	-80		8600

Mean

Let A = 25

$m = A + \frac{\sum f dx}{n}$

$m = 25 + \frac{-80}{58} = 25 + (-1.4) = 25 - 1.4 = 23.6$

Standard deviation

$\sigma = \sqrt{\frac{\sum fd^2_x}{N} - \left[\frac{\sum fd_x}{N}\right]^2}$

$\sigma = \sqrt{\frac{8600}{58} - \left[\frac{-80}{58}\right]^2} = \sqrt{148.276 - 1.9} = \sqrt{146.38} = 12.099$ Ans.

Variance

V = σ^2

= 146.39 Ans.

Coefficient of Variation

$CV = \frac{standard\ Deviation}{Mean} \times 100$

$= \frac{\sigma}{m} \times 100 = \frac{12.099}{23.6} \times 100$= o.51 × 100 = 51 Ans.

Sol. (Step Deviation Method)

X	x	f	d_x (x-A)/i	fd_x	d^2_x	fd^2_x
0-10	5	10	-2	-20	4	40
10-20	15	12	-1	-12	1	12
20-30	25	17	0	0	0	0
30-40	35	14	1	14	1	14
40-50	45	5	2	10	4	20
		58	0	-8		86

Let A=25

Mean

$m = A + \frac{\sum f dx}{n} \times 10$

$m = 25 + \frac{-8}{58} \times 10 = 25 + (-0.1379) \times 10 = 25 - 0.1379 \times 10 = 25 - 1.379 = 23.6$

Standard Deviation

$\sigma = \sqrt{\frac{\sum fd^2_x}{N} - \left[\frac{\sum fd_x}{N}\right]^2} \times i$

$\sigma = \sqrt{\frac{86}{58} - \left[\frac{-8}{58}\right]^2} \times 10 = \sqrt{1.48 - 0.019} \times 10 = \sqrt{1.461} \times 10$

$\sigma = 1.2087 \times 10 = 12.09$ Ans.

Variance

V = σ^2

V = $(12.09)^2$ = 146.168Ans.

Coefficient of Variation

$$CV = \frac{standard\ Deviation}{Mean} \times 100$$

$$CV = \frac{\sigma}{m} \times 100$$

$$= \frac{12.099}{23.6} \times 100 = o.51 \times 100 = 51 \text{ Ans.}$$

Measures of Skewness

Skewness

Skewness refers to the asymmetry or lack of symmetry in the shape of frequency distribution. In symmetrical distribution the values of mean, median and mode coincide. A distribution which is not symmetrical is called a skewed or asymmetrical distribution.

Measures of skewness

The following are the measures of skewness:

Karl Pearson's Method

The method is based on the relationship among the three measures of central tendency and is popularly known as Karl Pearson's measure of skewness. The formula used in its computation is as follows:

1. Mean-Mode
2. Mean-Median
3. Median –Mode

These measures of skewness are absolute. The relative measures of skewness can be had by dividing the absolute measures by any measure of dispersion. The relative measures of skewness are also called coefficients of skewness.The coefficient of skewness is computed by using the following formula:

$$\text{Coefficient of skewness} = \frac{\text{mean-mode}}{\text{Standard deviation}}$$

Sometimes mode is ill-defined. In such a situation, the following formula is used:

$$\text{Coefficient of skewness} = \frac{3(\text{mean – median})}{\text{Standard Deviation}}$$

Bowley's Method

Bowley's measure of skewness is based on the quartiles and is given by:

$$Sk = Q_3+Q_1-2Md.$$

$$\text{Coefficient of SK} = \frac{Q_3+Q_1-2Md}{Q_3-Q_1}$$

The value of this coefficient of skewness varies between the limits ± 1. But the result obtained through this method should be taken with a grain of salt. It is just possible that the value of the coefficient may be zero and yet the series may not be symmetrical. The answer to this lies in the fact that quartiles are not based on all the observations of the series.

Practical Problems

Given that mean = 22, mode = 20 and SD = 3.06, calculate skewness and its coefficient using Karl Pearson's method. ?

Sol.

Given mean =22, mode = 20 and Standard Deviation = 3.06

Sk =Mean-Mode

22 – 20 = 2

$$\text{Coefficient of Sk} = \frac{\text{mean-mode}}{\text{Standard deviation}}$$

$$= \frac{22-20}{3.06}$$

$$= \frac{2}{3.06} = 0.65\text{Ans.}$$

Problem

Given mean = 22, Md. = 20 and SD = 3.06, Find out Karl Pearson's coefficient of Sk. ?

Sol.

Given mean = 22, Md. = 20 and SD = 3.06

$$\text{Coefficient of skewness} = \frac{3(\text{mean} - \text{median})}{\text{Standard Deviation}}$$

$$\text{Coefficient of skewness} = \frac{3(22-20)}{3.06}$$

$$= \frac{3(2)}{3.06}$$

$$= \frac{6}{3.06} = 1.96 \text{ Ans}$$

Problem

Given that Q_3 = 195.94, Q_1 = 138.0 and Md. =167.9,find Bowley's coefficient of Sk. ?

Sol.

Given that Q_3 = 195.94, Q_1 = 138.0 and Md. =167.9

$$Sk = Q_3+Q_1-2Md.$$

$$= 195.94+138.0-2(167.9)$$

$$= 333.94-335.8 = -1.86$$

$$\text{Coefficient of SK} = \frac{Q_3+Q_1-2Md}{Q_3-Q_1}$$

$$= \frac{195.94+138.0-2(167.9)}{195.94-138.0}$$

$$= \frac{333.94-335.8}{57.94}$$

$$= \frac{-1.86}{57.94}$$

$$= -0.03 \text{ Ans.}$$

Note:-In case the values of Q_3, Q_1 and median are not given. Compute the said values by employing the formula of median and quartile deviation from the given data.

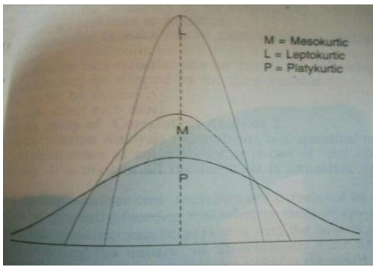
Moments and Kurtosis

Moment is a familiar mechanical term which refers to the measures of a force with respect to its tendency to provide rotation. The strength of the tendency depends on the amount of force and the distance from the origin of the point at which the force is exerted.

The concept of moment is of great significance in statistical work. With the help of moments we can measure the central tendency of a set of observations, their variability, their asymmetry and the height of the peak of the curves.

Kurtosis is a Greek term meaning bulginess. In statistics kurtosis refers to the degree of flatness or peakedness in a region about the mode of a frequency curve. The degree of kurtosis of a distribution is measured relative to the peakedness of normal curve. In other words, measures of kurtosis tell us the extent to which a distribution is more peaked or flat-topped than the normal curve. If a curve is more peaked than the normal curve, it is called **LEPTOKURTIC.** In such a case items are more closely bunched around the mode. On the other hand, if a curve is more flat-topped than the normal curve, it is called **PLATYKURTIC.**

The normal curve itself is known as **MESOKURTIC.**



Measures of Kurtosis

The most important measure of kurtosis is the value of the coefficient $ß_2$.It is defined as:

$$ß_2 = \frac{\mu_4}{\mu_2^2}$$

Where $\mu_4$is the 4thmoment and μ_2 is the 2nd moment.

The greater the value of $ß_2$, the more peaked is the distribution.

For a normal curve, the value of $ß_2=3$.When the value of $ß_2$is greater than3,the curve is more peaked than the normal curve i.e.leptokurtic.When the value of $ß_2$is less than 3, then the curve is less peaked than the normal curve, i.e. platykurtic. The normal curve and other curves with $ß_2= 3$ are called mesokurtic.

SometimesΥ_2, the derivative of $ß_2$ is used as a measure of kurtosis.Υ_2 is defined as

$\Upsilon_2= ß_2 - 3$

For a normal distribution $\Upsilon_2=0$. If Υ_2 is positive, the curve is leptokurtic and if Υ_2 is negative, the curve is platykurtic.

Practical Problem

Calculate first four moments and kurtosis for the following data:

X : 1,2, 8,9,10.

Sol.

X	$(X-\bar{X})$	$(X-\bar{X})^2$	$(X-\bar{X})^3$	$(X-\bar{X})^4$
1	-5	25	-125	625
2	-4	16	- 64	256
8	2	4	8	64
9	3	9	27	81
10	4	16	64	256
30	0	70	-90	1282

$$X = \frac{\sum X}{N} = \frac{30}{5} = 6$$

$$\mu_1 = \frac{\sum (X-\bar{X})}{N} = \frac{0}{5} = 0$$

$$\mu_2 = \frac{\sum (X-\bar{X})^2}{N} = \frac{70}{5} = 14$$

$$\mu_3 = \frac{\sum (X-\bar{X})^3}{N} = \frac{-90}{5} = -18$$

$$\mu_4 = \frac{\sum (X-\bar{X})^4}{N} = \frac{1282}{5} = 256.4$$

<u>Kurtosis</u>

Karl Pearson's kurtosis is given by the following formula,

$$ß_2 = \frac{\mu_4}{\mu_2^2}$$

$$ß_2 = \frac{256.4}{196} = \textbf{1.31Ans}$$

$$\Upsilon_2 = ß_2 - 3 = 1.31 - 3 = -1.69 \text{ Ans.}$$