The Economic Journal of Nepal
Vol. IV. No. 4, Oct.-Dec. 1981 EIC-TU

Mutual Deviation : A Measure of Dispersion

Shiva Gautam★

In the first part of this article a measure of dispersion; Mutual Deviation has been developed and its relationship with standard deviation (for univariate distribution) is shown. In the second part its practical application to measure the dispersion (two-dimensional data) is illustrated where the standard deviation fails. The location of data (bivariate only) is also calculated.

Part I

1. INTRODUCTION

Let $X = (x_1 . x_2 , \ldots\ldots . x_n)^1$ denote the vector whose elements are the n different values of a variate. $F_{nx1} = (f_1 , f_2 , \ldots\ldots , f_n)^1$ be another vector of frequencies such that $x_i = (i = 1, 2, \ldots , n)$ has a frequency of f_i $(i = 1, 2, \ldots , n)$ and $\Sigma f = N$ = total frequencies.

Mutual deviation tries to measure the spread of different values of the variate among themselves. There are various methods for the measurement of scatterness or dispersion of a system of data, each of which are developed on some logical thinking based on what actually one means by scatterness or dispersion. Mutual deviation measures the spread of a system of data by

★ Mr. Gautam is Currently associated with Nepal Jute Development Corporation, Biratnagar, Nepal.

measuring the average distance of a value from another different value. To avoid the ambiguity and to retain its mathematical characteristics not the average distance but the average of the squares of distances or differences is taken as the square of the Mutual deviation.

BUILDING OF THE MODEL

These were mentioned that in the above system data x_i is repeated f_i times. While measuring the scatterness Mutual deviation does not consider the dispersion among the same values for dispersion does not exists among the same values. Let us take two different values x_i and x_j with respective frequencies f_i and f_j. Then each x_i is located at a distance of $(x_i - x_j)$ from each x_j ; i = j.

This expression will be repeated $f_i \; f_j$ times as there are $f_i \; x_i^s$ and $f_2 \; x_2^s$ .Assuming similarly for each possible pair of $x_{i,} \; x_j$ (i = j) in the given system we obtain the square of Mutual deviation which is nothing but the arithmetic averages of the squares of all $(x_i - x_j)$.

If M be the Mutual deviation, then

$$M^2 = \frac{f_1 f_2 (x_1 - x_2)^2 + f_1 f_3 (x_1 - x_3)^2 + \dots + f_{n-1} f_n (x_{n-1} - x_n)^2}{f_1 f_2 + f_1 f_3 + \dots\dots + f_{n-1} f_n}$$

$$= \frac{f_1 f_2 (x_1^2 + x_2^2 - 2x_1 x_2) + f_1 f_3 (x_1^2 + x_3^2 - 2x_1 x_3) + \dots f_{n-1} f_n (x_{n-1}^2 + x_n^2 - 2x_{n-1}}{\tfrac{1}{2}(2f_1 f_2 + 2f_1 f_3 + \dots + 2f_{n-1} f_n)}$$

$$= \frac{f_1 x_1^2 (f_2 + f_3 + \dots + f_n) + f_2 x_2^2 (f_1 + f_3 + \dots + f_n) + \dots f_n x_n^2 (f_1 + f_2 + \dots + f_{n-1}) - (2f_1 f_2 x_1 x_2 + 2f_1 f_3 x_1 x_3 + \dots\dots 2f_{n-1} f_n x_{n-1} x_n)}{\tfrac{1}{2}(2f_1 f_2 + 2f_1 f_3 + \dots + 2f_{n-1} f_n)}$$

$$= \frac{f_1 x_1^2 (N-f_1) + f_2 x_2^2 (N-f_2) + \ldots + f_n x_n^2 (N-f_n) - (2f_1 f_2 x_1 x_2 + 2f_1 f_3 x_1 x_3 + \ldots 2f_{n-1} f_n x_{n-1} x_n)}{\frac{1}{2}(2f_1 f_2 + 2f_1 f_3 + \ldots + 2f_{n-1} f_n)}$$

$$[2]\ M^2 = \frac{N(f_1 x_1^2 + f_2 x_2^2 + \ldots + f_n x_n^2) - (f_1^2 x_1^2 + f_2^2 x_2^2 + \ldots + f_n^2 x_n^2) - 2(f_1 f_2 x_1 x_2 + 2f_1 f_3 x_1 x_3 + \ldots + 2f_{n-1} f_n x_{n-1} x_n)}{\frac{1}{2}(2f_1 f_2 + 2f_1 f_3 + \ldots + 2f_{n-1} f_n)}$$

IN MATRIX NOTATIONS

Let us define two matrices G and $\triangle$ both of order n x n such that G is matrix with diagonal elements o and with all off-diagonal elements equal to unity, and $\triangle$ is a diagonal matrix with (i, i)th element equal ti f_i ; i.e.

$$\underset{n \times n}{G} = \begin{bmatrix} 0 & 1 & 1 & . & . & . & 1 \\ 1 & 0 & 1 & . & . & . & 1 \\ 1 & 1 & 0 & . & . & . & 1 \\ 1 & 1 & 1 & . & . & . & 1 \end{bmatrix}, \quad \text{and } \underset{n \times n}{\triangle} = \begin{bmatrix} f_1 & 0 & 0 & . & . & . & 0 \\ 0 & f_2 & 0 & . & . & . & 0 \\ 0 & 0 & f_3 & . & c. & . & 0 \\ 0 & 0 & 0 & . & . & . & f_n \end{bmatrix}$$

Then from (1)

$$[2]\ M^2 = \frac{N X' \triangle X - X' \triangle^2 X - X' \triangle G \triangle X}{\frac{1}{2} F' GF}$$

$$= \frac{2X' [N \triangle - \triangle^2 - (\triangle G) \triangle] X}{F' GF}$$

$$[3] \text{ or } F'GF\ M^2 = 2X' [N \triangle - \triangle^2 - \triangle G \triangle] X$$

$$\text{or } F'GF\ M^2 = 2X'AX = \text{a quadratic form}$$

Where A is matrix obtained by the simplification of the expression in the paranthesis.

Rewritten [1] we have (for computational purposes)

$$M^2 = \frac{N\sum^{n} fx^2 - \sum^{n} f^2 x^2 - (2f_1 f_2 x_1 x_2 + 2f_1 f_3 x_1 x_3 + \dots + 2f_{n-1} f_n x_{n-1} x_n)}{\frac{1}{2}(2f_1 f_2 + 2f_1 f_3 + \dots + 2f_{n-1} f_n}$$

$$= \frac{N\sum^{n} fx^2 - \sum^{n} f^2 x^2 - \left[(\sum^{n} fx)^2 - \sum^{n} f^2 x^2)\right]}{\frac{1}{2}\left[(\sum^{n} f)^2 - \sum^{n} f^2)\right]}$$

$$= \frac{N\sum^{n} fx^2 - \sum^{n} f^2 x^2 - (\sum^{n} fx)^2 + \sum^{n} f^2 x^2}{\frac{1}{2}(N^2 - \sum^{n} f^2)}$$

[4] $$M^2 = \frac{N\sum^{n} fx^2 - (\sum^{n} fx)^2}{\frac{1}{2}(N^2 - \sum f^2)}$$

At first it seems that Mutual deviation is not based on all the observations because it does not consider the deviations among the same values but [4] shows that it uses all the observations.

From [4]

$$M^2 = \frac{N\sum^{n} fx^2 - (\sum^{n} fx)^2}{\frac{1}{2}(N^2 - (\sum f^2)}$$

$$\text{or} \quad M^2 = \frac{N^2 \sigma_x^2}{\frac{1}{2}(N^2 - \sum f^2)}$$

[5] $$\text{or} \quad M^2 = \frac{2N^2}{N^2 - \sum f^2}\,\sigma_x^2 \quad , \left[\sigma_x^2 = \frac{1}{N}\sum^{n} f(x - \bar{x})^2\right]$$

[5] shows the relation of Mutual deviation with standard deviation.

Let us consider two cases :

CASE - I

Let $f_i = 1, i = 1, 2, \ldots\ldots, n.$

Then $N = \Sigma f_i = n$ and $\Sigma f^2 = n$

From [5]

$$M^2 = \frac{2N^2}{N^2 - \Sigma f^2} \delta_x^2 = \frac{2N^2}{n^2 - n} \delta_x^2$$

[6] $\therefore M^2 = \frac{2n}{n-1} \delta_x^2$ where n = total number of observations

CASE - II

When the mutual deviations are taken among the same values too i.e. when the deviations ($x_i - x_i$) S are also considered (x_i has a frequency of f_i and thus) the number of ($x_i - x_i$) will be $f_{i_{C_2}} = \left[\frac{f_i (f_i - 1)}{2} \right]$, then adds nothing to the numerator but adds $\sum^{n} f_{C_2}$ to the denominator. Then from [4]

$$M^2 = \frac{N \sum^n fx^2 - (\sum^n fx)^2}{\frac{1}{2}[N^2 - \Sigma f^2] + f_{1_{C_2}} + f_{2_{C_2}} + \ldots\ldots + f_{n_{C_2}}}$$

$$= \frac{N \sum^n fx^2 - (\sum^n fx)^2}{\frac{1}{2}[N^2 - \frac{1}{2}\Sigma f^2 + \frac{f_1 (f_1 - 1)}{2} + \frac{f_2 (f_2 - 1)}{2} + \ldots + \frac{f_n (f_n - 1)}{2}}$$

$$= \frac{N^2 \quad \delta_x^2}{\frac{1}{2}(N^2 - \Sigma f^2 + \Sigma f^2 - \Sigma f)}$$

$$[7] \quad = \frac{2N^2 \quad \delta_x^2}{N^2 - \Sigma f} \qquad = \frac{2N^2}{N-1} \quad \delta_x^{2*}$$

This expression is similar to that of [6]. This relation would have directly obtained dividing the numerator by N_{C_2}, total number possible pairs of the values.

Thus M, the Mutual deviation is given by (from 4)

$$[8] \quad M = \sqrt{\frac{N \sum^{n} fx^2 - (\sum^{n} fx)^2}{\frac{1}{2}(N^2 - \Sigma f^2)}}$$

$$[9] \text{ or, } M = \sqrt{\frac{2X'(N\Delta - \Delta^2 - \Delta G \Delta)X}{F'GF}}$$

Thus M may be considered as a measure of dispersion, for it possesses several characteristics of an ideal measure of dispersion.[1] Some of them are:

(a) It is rigidly defined

(b) It is easily understandable and is as easier to calculate as standard deviation.

(c) It is based on all observations

(d) It is amenable to further mathematical treatment. Besides these, Mutual deviation, unlike the standard deviation seeks no reference to the deviation from a central value which in turn is supposed to possess the above mentioned characteristics

* This is nothing but the average of the squares of the terms of Gini's mean difference (1912).

1. Kapoor, J.N. and Saxena, H.C., S. Chand & Co. Ltd. Pvt., Ram Nagar, New Delhi 55, 1972 p. 29.

D_1 COMBINED DATA

To find the combined mutual deviation of two set of values.

Let $\underset{k \times 1}{x_1} = (x_1, x_2 \ldots\ldots x_k)^1$ and $\underset{(n-k)\times 1}{x_2} = (x_{k+1}, x_{k+2} \ldots\ldots x_n)^1$

be two set of values with respective frequency vectors

$\underset{k \times 1}{F_1} = (f_1, f_2, \ldots\ldots f_k)^1$ and $\underset{(n-k)\times 1}{F_2} = (f_{k+1}, f_{k+2} \ldots\ldots f_n)^1$, and

$\Sigma f_1 = N, \Sigma f_2 = N_2$ such that $N_1 + N_2 = N$. The combined vectors will be

$\underset{n \times 1}{X} = (x_1, x_2, \ldots\ldots x_k, x_{k+1} \ldots\ldots x_n)^1$ and $\underset{n \times 1}{F} = (f_1, f_2, \ldots\ldots f_k, f_{k+1},$

$\ldots\ldots f_n)^1$. Let M, M_1 and M_2 denote combined Mutual deviation, Mutual deviation for the first set and Mutual deviation for the second set, respectively.

Appropriate partitioning

$$\underset{n \times 1}{X} = \begin{bmatrix} \underset{K \times 1}{X_1} \\ \underset{(n-k)\times 1}{X_2} \end{bmatrix} \qquad F = \begin{bmatrix} \underset{K \times 1}{F_1} \\ \underset{(n-k)\times 1}{F_2} \end{bmatrix}$$

$$\underset{n \times n}{G} = \begin{bmatrix} 0 & 1 & 1 & \ldots & 1 & \ldots & 1 & 1 \\ 1 & 0 & 1 & \ldots & 1 & \ldots & 1 & 1 \\ 1 & 1 & 0 & \ldots & 1 & \ldots & 1 & 1 \\ 1 & 1 & 1 & \ldots & 0 & \ldots & 1 & 1 \\ - & - & - & -- & - & -- & - & - \\ 1 & 1 & 1 & \ldots & 1 & \ldots & 0 & 1 \\ 1 & 1 & 1 & \ldots & 1 & \ldots & 1 & 0 \end{bmatrix} = \begin{bmatrix} \underset{k \times k}{G_{11}} & \underset{k \times (n-k)}{G_{12}} \\ \underset{(n-k) \times k}{G_{21}} & \underset{(n-k) \times 9n-k)}{G_{22}} \end{bmatrix}$$

$$= \begin{bmatrix} f_1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & f_2 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & f_k & \cdots & 0 & 0 \\ - & - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & f_{k+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & f_n \end{bmatrix} = \begin{bmatrix} \underset{k \times k}{\Delta_{11}} & \underset{k \times (n-k)}{O} \\ \underset{(n-k) \times k}{O} & \underset{(n-k) \times (n-k)}{\Delta_{22}} \end{bmatrix}$$

Now, (Where Ω is a null vector)

$$M^2 = \frac{X'(N\Delta - \Delta^2 - \Delta G \Delta)X}{\frac{1}{2} F' GF}$$

$$\text{or } \tfrac{1}{2} M^2 F'GF = X'(NI - \Delta - \Delta G)\Delta X$$

(Where $\underset{n \times n}{I}$ is an identity matrix.)

$$\text{or } \tfrac{1}{2} M^2 F'GF = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}^1 \left[(N_1 + N_2) \begin{pmatrix} I_1 & 0 \\ \Omega & I_2 \end{pmatrix} - \begin{pmatrix} \Delta_{11} & \Omega \\ \Omega & \Delta_{22} \end{pmatrix} - \begin{pmatrix} \Delta_{11} & \Omega \\ \Omega & \Delta_{22} \end{pmatrix} \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \right] \begin{pmatrix} \Delta_{11} & 0 \\ 0 & \Delta_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\text{or } \tfrac{1}{2} M^2 \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}^1 \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = 1/2 M_1^2 F_1^1 G_{11} F_1 + 1/2 M_2^2 F_2^1 G_{22} F_2 + N_1 X_2^1 \Delta_{22} X_2 + N_2 X_1^1 \Delta_{11} X_1 - 2X_1^1 \Delta_{11} G_{12} \Delta_{22} X_2$$

$$\text{or } M^2 (F_1^1 G_{11} F_1 + F_2^1 G_{22} F_2 + 2F_1^1 G_{12} F_2) = M_1^2 F_1^1 G_{11} F_1 + M_2^2 F_2^1 G_{22} F_2 + 2N_1 X_2^1 \Delta_{22} X_2 + 2N_2 X_1^1 \Delta_{11} X_1 - 4X_1^1 \Delta_{11} G_{12} \Delta_{22} X_2$$

$$\text{or } M^2 = \frac{M_1^2 F_1^1 G_{11} F_1 + M_2^2 F_2^1 G_{22} F_2 + 2N_1 X_2^1 \Delta_{22} X_2 + 2N_2 X_1^1 \Delta_{11} X_1 - 4X_1^1 \Delta_{11} G_{12} \Delta_{22} X_2}{F_1^1 G_{11} F_1 F_2^1 G_{22} F_2 + 2F_1^1 G_{12} F_2}$$

Part II

A_2 For an univariate distribution standard deviation or variance gives a good measure of dispersion. If M be the mutual deviation and 6_x standard deviation, then they are related by the equation.

$$(10) \quad M^2 = \frac{2N}{N-1} \delta_x^2$$

Where N = number of observation.

By definition M^2 is the arithmetic mean of squares of the differences of x_i 'S, the difference being taken for each possible pair. The set of observations x_i , x_2 , ... x_N can be plotted in a graph.

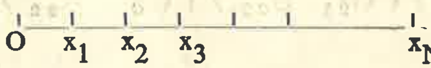


Fig. 1.

Now it can be seen that $(x_i - x_j)^2$ is the square of the distance from the point x_j to x_i $(i \neq j)$. Thus M^2 is the arithmetic mean of these squares. But variance is the arithmetic mean of the squares of the deviations taken from the mean. Though δ_x^2 and M^2 apparantly are based on different thinking, more or less they express same thing and are related as in (10).

Let us consider the case of bivarate distribution.The set of N observations (x_1 , y_1), (x_2 , y_2) ...... (x_N , y_N) is shown in a two-dimension plane.

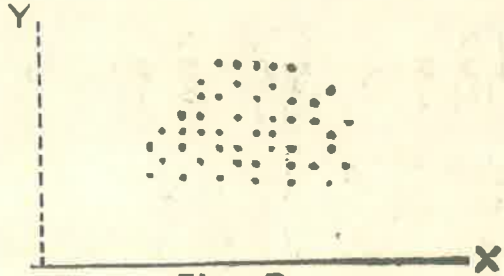


Fig. 2.

The dots in Fig. 2 are scattered more densely than the dots in Fig. 3 delow.

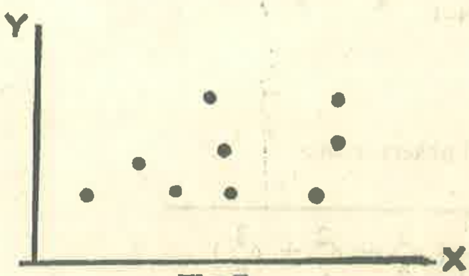


Fig.3.

The dots in Fig. 2 and Fig. 3 are measured with reference of two frames and thus standard deviation can not measure the scatterness. If the dots were measured with reference of a single (Fig. 4) frame then both the standard deviation and mutual deviation could have measured the scatterness.

Fig. 4 (a)

Fig. 4 (b)

As has been mentioned above M^2 (square of mutual deviation) is the average of the squares of distances between each possible pair of points. Similarly for two dimensional points M^2 canbe defined as the average of distances between each possible pair of points.

Now the distance between (x_i, y_i) and (x_j, y_j) $(i > j$ and $i = 1, 2, \ldots\ldots N)$

$$= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Squaring these expression and then taking average

$$M^2 = \frac{1}{{}^N C_2} \left[\sum_{i>j}^{N} \sum_{j} (x_i - x_j)^2 + \sum_{i>j}^{N} \sum_{j} (y_i - y_j)^2 \right]$$

$$= \frac{2N}{N-1} \delta_x^2 + \frac{2N}{N-1} \delta_y^2$$

$$= \frac{2N}{N-1} (\delta_x^2 + \delta_y^2)$$

(11) $\therefore \; M = \sqrt{\frac{2N}{N-1} (\delta_x^2 + \delta_y^2)}$

Similarly,

For 3 dimensional observations

$$M = \sqrt{\frac{2N}{N-1} (\delta_x^2 + \delta_y^2 + \delta_z^2)}$$

If n dimension be represented by 1, 2, and n than,

$M = \sqrt{\frac{2N}{N-1} \sum_{i=1}^{n} \delta_i^2}$, where δ_i^2 = standard deviations for the component of i^{th} dimension.

Thus standard deviation does not measure the scatterness or dispersion of data which are not one dimensional.

B_2 If x_i and y_i (i = 1, 2, ... N) is related by the expression

$$y_i = a + bx_i$$

Then,

$$\delta_y^2 = 1/N \sum (y_i - \bar{y})^2 = 1/N \sum (a + bx_i - a - b\bar{x})^2$$

$$= 1/N.\ b^2 \sum (x_i - \bar{x})^2$$

$$\therefore \delta_y^2 = b^2 \delta_x^2$$

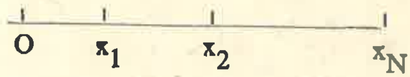
Now, $$M^2 = 2N/N\text{-}1\ (\delta_x^2 + \delta_y^2)$$

$$= 2N/N\text{-}1\ (\delta_x^2 + b^2 \delta_x^2)$$

(12) $$M^2 = 2N/N\text{-}1\ \delta_x^2\ (01 + b^2)$$

(Where b is the slope)

$C_{2.1}$ Mutual deviation gives a measure of dispersion or scatterness of two (or more) dimensional data. One is apt to think of a point around which the whole system of data tends to cluster or where the body of data is located? For one dimensional data the point of location of the data is obtained by summing the distance of each point from origin and then dividing number of terms.



$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_N}{N}$$

Similarly for two dimensional data the average distance of points (x_i, y_i) from $(0, 0)$ is given by: $(2 = 1, 2, \ldots\ldots N)$

$$\frac{\sum^{N} (x_i^2 + y_i^2)^{1/2}}{N}$$

Let us suppose that the point about which the data tends to cluster is (X,Y), such that this point lies at an average distance from the origin. Then

$$\frac{\Sigma (x_i^2 + y_i^2)^{1/2}}{N} = (x^2 + y^2)^{1/2}$$

(13) or $$\left(\frac{\Sigma (x_i^2 + y_i^2)^{1/2}}{N} \right)^2 = X^2 + Y^2$$

Again each point (x_i , y_i) will lie near to the X axis if x_i is relatively greater than y_i . (For one dimensional data this should not be considered because the force exerted on the point will be single and thus it will be located at distance $\bar{x}$ from origin one the axis. The ratio y_i / x_i may be considered as measure of nearness to the axis. If y_i / x_i is greater than one then the point will be nearer to Y axis, if the ratio is one then the point will be located at equal distances from the axes and if the ratio is less than one then the point will be located nearer to X-axis. Thus it may be thought that the average nearness of the data with respect to the axes is.

$$1/N (y_1 / x_1 + y_2 / x_2 + \ldots + y_N / x_N) = 1/N \sum_{i=1}^{N} y_i / x_i$$

But as it has already been assumed that the point needed is (X, Y) which is located at (with respect to axis) Y/X.

$$\therefore 1/N \Sigma y_i / x_i = Y/X \text{ or } X/N \Sigma y / x_i = Y$$

(14) or $Y = X/N \Sigma y_i / x_i$

Squaring

$$y^2 = X^2 / N^2 (\Sigma y_i / x_i)^2$$

Substituting this in (13)

$$\left(\frac{(x_i^2 + y_i^2)^{1/2}}{N} \right)^2 = X^2 + X^2 / N^2 (\Sigma y_i / x_i)^2$$

$$\text{or} \quad \left(\frac{\sum (x_i^2 + y_i^2)^{1/2}}{N} \right)^2 = \frac{N^2 X^2 + X^2 (\sum y_i / x_i)^2}{N^2}$$

$$\text{or} \quad \left[\sum (x_i^2 + y_i^2)^{1/2} \right]^2 = X^2 \left[N^2 + (\sum y_i / x_i)^2 \right]$$

$$\text{or} \quad X^2 = \frac{\left[\sum (x_i^2 + y_i^2)^{1/2} \right]^2}{N^2 + (\sum y_i / x_i)^2}$$

$$(15) \quad \text{or} \quad X = \frac{\sum (x_i^2 + y_i^2)^{1/2}}{\sqrt{N^2 + (\sum y_i / x_i)^2}}$$

Substitution of this value in [14] gives the value of Y which is

$$(16) \quad Y = \frac{\sum (x_i^2 + y_i^2)^{1/2}}{N \sqrt{N^2 + (\sum y_i / x_i)^2}} \sum y_i / x_i$$

Hence (the average value or) the system of the data is located at (X, Y) where X and Y are given by the equations [15] and [16] respectively.

CONCLUSION

An effort has been made in this article to introduce a Statistical technique which possesses several merits advantagous to measure dispersion of Statistical data more conveniently as compared to the existing technique. The relationship between the commonly used standard deviation technique and this technique has also been discussed. Besides, some of the advantages of this technique over standard deviation have also been pointed out.