IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 3, March 2015.
www.ijiset.com
ISSN 2348 – 7968

Relative Representation of Graph

Arup JyotiChutia

3rd sem, B.Tech(ME),Tezpuruniversity,Assam(India)

ABSTRACT

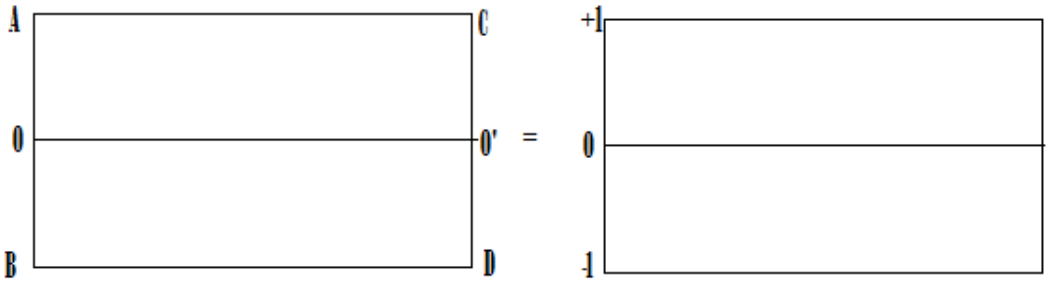
The more is the information, the better is the representation of statistical data (provided it's easily comprehensible). If we can represent much more information say central tendency, deviations and statistical constants etc. in an efficient way; it would be easier to deal with various data. The main objective of this paper is to introduce a novel approach of graphical representation.

INTRODUCTION

Thethree broad ways of presenting data are as follows-

- Textual representation
- Tabular representation
- Graphical representation

Out of these most common is the graphical representation. Each of the above ways of representation has respective advantages and drawbacks as well. Basically what we represent in a graph is the raw data collected, in a certain order which often require to observe thoroughly for a better conclusion .
Here my approach is not to represent raw data directly but to represent data that are based upon collected raw data.
DIFFERENT STEPS TO BE FOLLOWED ARE
1.Consider a rectangular box ABCD as a system such that O and O' are the mid points of AB and CD respectively. Also OA=1unit,OB=-1unit.

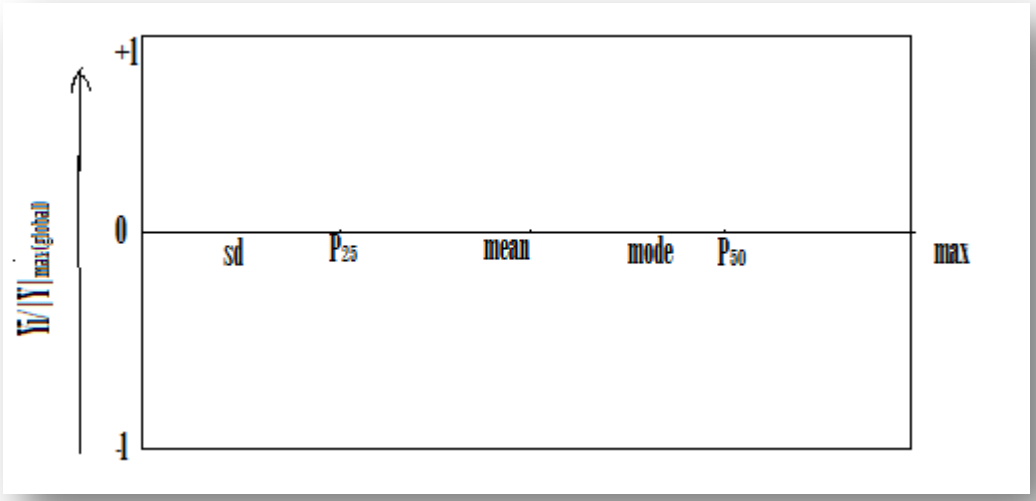


2. OO' is the base line and BOA is the value line for corresponding base coordinate plotted in OO'. Base coordinates can be percentiles(p) measures, central tendency ,deviations ,statistical constants and other quantity we prefer to choose and represent.
3. O(P1,min) and O'(P100,max) are fixed .
Besides plotting percentiles we also plot other base coordinates say (Bi) in between O and O'. In plotting the these base coordinate with their corresponding value say (Yi) we follow the flowing assumptions :

ISSN 2348 – 7968

a. central measures are assumed to be concentrated in the middle sections of OO' .
b. Deviations and constants like β,μ etc. are assumed to be nearer to the minimum value hence plotted near O(Pi,min). Hence the following table is to be followed:

BASE COORDINATES	REGION OF OO'
Deviations and statistical constants	(0-25)% of length
Central tendency measures like mean ,median, mode etc.	(25-75)% of length
Other base variables (as per choice whose value are observed to be greater than mean, median etc.)	(75-100)% of length



[NOTE: While plotting the data we will plot those base coordinates which are of our interest. It is not compulsory to represent all the base coordinates. But if we can represent as much possible the better informative it is. While plotting (B_I,Y_I) alphabetical order is to be followed for or as comfortable to the person]
4. Choose the maximum$|Y_i|$ (global max) of all data i.e.$|Y|_{(max)global}$. Before plotting all 'Y_i' each of its is to be divided by$|Y|_{max(global)}$. i.e. ($Y_i/|Y|_{max}$). Also Y_{max} is to be denoted in the information box such that we can directly find all the original entries as per required.(note in figure above)

5. Suppose we have to represent mean and mode as base coordinate (B_i) , now since alphabetically mean would come before mode. Next we may have say three modes in the data we would then sort them in ascending order and denote by using subscripts say $Mode_1$, $Mode_2$ and $Mode_3$ such that $Mode_1$<$Mode_2$<$Mode_3$ and represent. We may use different symbols for different B_i but while sorting in alphabetical order we would prefer letters in the original term.
6. All measures i.eY_i can be connected by a curve line except the quantities like deviations . (Which actually do not represent the entries of the data but how they are deviated from the central measures .

ISSN 2348 – 7968

Similarly constants like skewness coefficient etc.) Butthen these values are to be connected by a separate line. Otherwise If not connected then no quantity is to be connected .

These are also represented in the (0-25)%length of OO'. Often some constants may have very smaller value then also they are to be plotted after dividing by $Y_{(max)global}$. Because here we are to compare the data .

[NOTE:
A. Here all Y_i's are divided by $|Y_i|_{(max)global}$ since they are to be represented in the range [-1,1] of the length of AOB of rectangle ABCD of the above specified system.
B. It is up to the plotter to plot or not some quantities in the system . Specially in the region of (75-100)% of OO' plotter may define some quantities and then plot but these should not be central measures and deviations etc. which are already covered in the (0-75)% of OO'. While defining ,it should be noted that the new Y_i's that would be found should not be greater than $|Y_i|_{max(global)}$ otherwise that would not fit the system as well as not a good representing one.
C. Here the graph can be represented by both connecting the Y_i's or not connecting these points (rule6), but if connected accordingly then we would be able to visualise the Y_i's like functional points and the area between the connected line formed by joining Y_i's and the base line OO' would help to interpret . The greater the area better will be the distribution of raw data for the curve joining Y_i for central tendency values, min, max etc. And for the curve joining $Y_{i'}$s for deviations ; greater the area more would be the irregularity in the distribution of data.

Illustration:

1.Consider the following table of marks obtained by randomly choosen10 different students in mathematics in different branches in a university in test II out of max mark 25 and plot in the graph using the above procedure:

Branch	Marks of 10 students
Mechanical Engineering(ME)	23,10.5,11,9,9,18,16.5,2.5,13.5,7
Civil Engineering (CIE)	10,6.5,2.5,12.5,10,8,13.5,13,5.5,7
Electrical Engineering(EE)	6.5,2,17.5,10.5,4.5,3,10.5,11,5,3.5
Computer Science Engineering(CSE)	9,18,13,14,9,4.5,7,4,5.5,11

ISSN 2348 – 7968

Here we have the following table of calculated data: $|Y|_{max(global)}=23$

BRANCH	DATA OF MARKS OBTAINED								
	min	sd	P_{25}	mean	median	P_{50}	mode	P_{75}	max
ME	2.5	5.92	9	12	10.75	10.5	9	16.5	23
$Yi/\|Y\|_{max}$	.109	.257	.391	.521	.467	.456	.391	.71	1
CIE	2.5	3.59	6.5	8.85	9	8	10	12.5	13.5
$Yi/\|Y\|_{max}$	.109	.156	.282	.384	.391	.348	.434	.543	.586
EE	2	4.86	3.5	7.4	5.75	5	10.5	10.5	17.5
$Yi/\|Y\|_{max}$	.087	.212	.152	.322	.25	.217	.456	.456	.76
CSE	4	4.53	5.5	9.5	9	9	9	13	18
$Yi/\|Y\|_{max}$	.174	.197	.239	.413	.391	.391	.391	.565	.783

ISSN 2348 – 7968

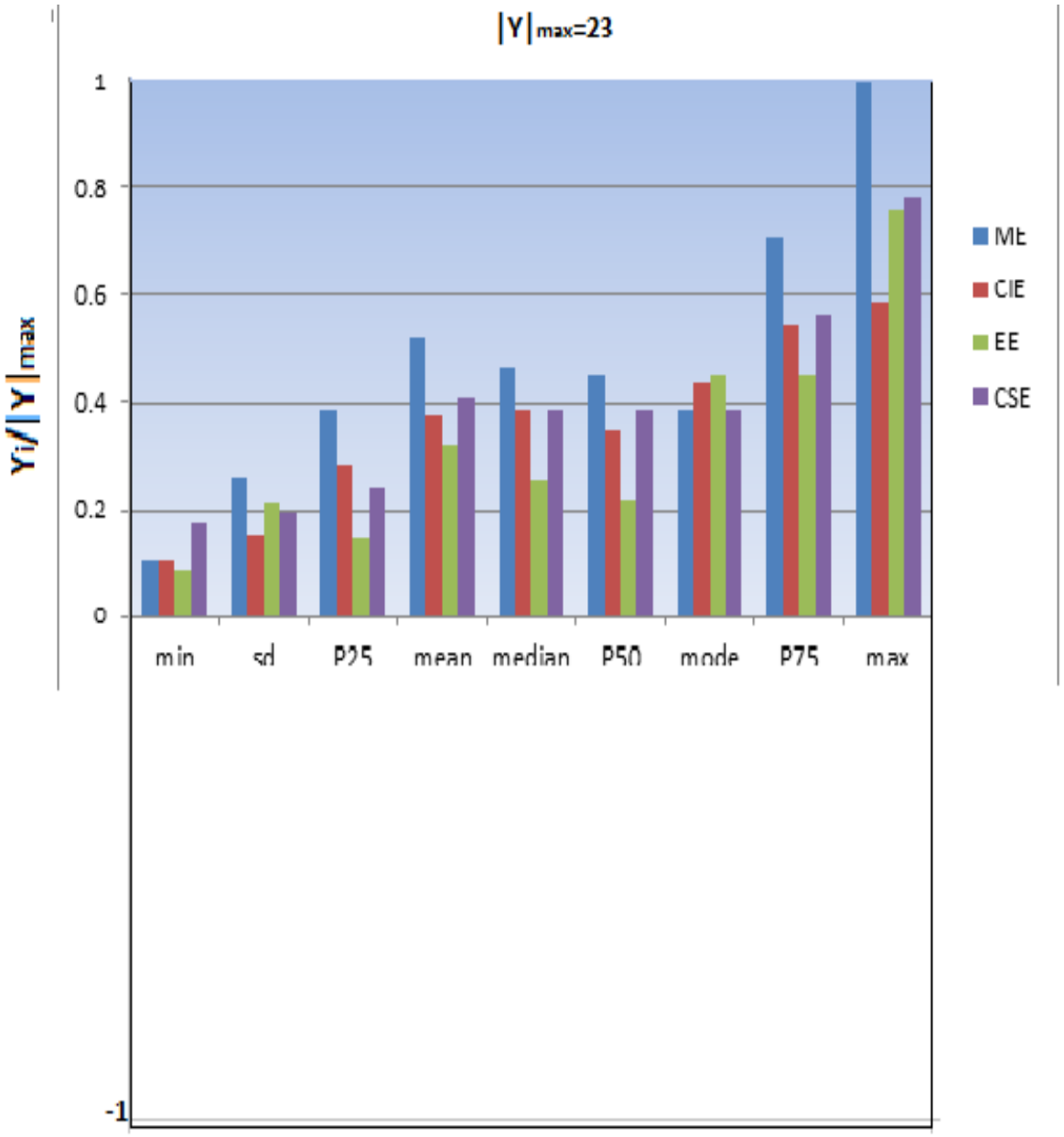


Fig: Graph to illustrate the above procedure but not connecting Y_i 's by a curve line .

ISSN 2348 – 7968

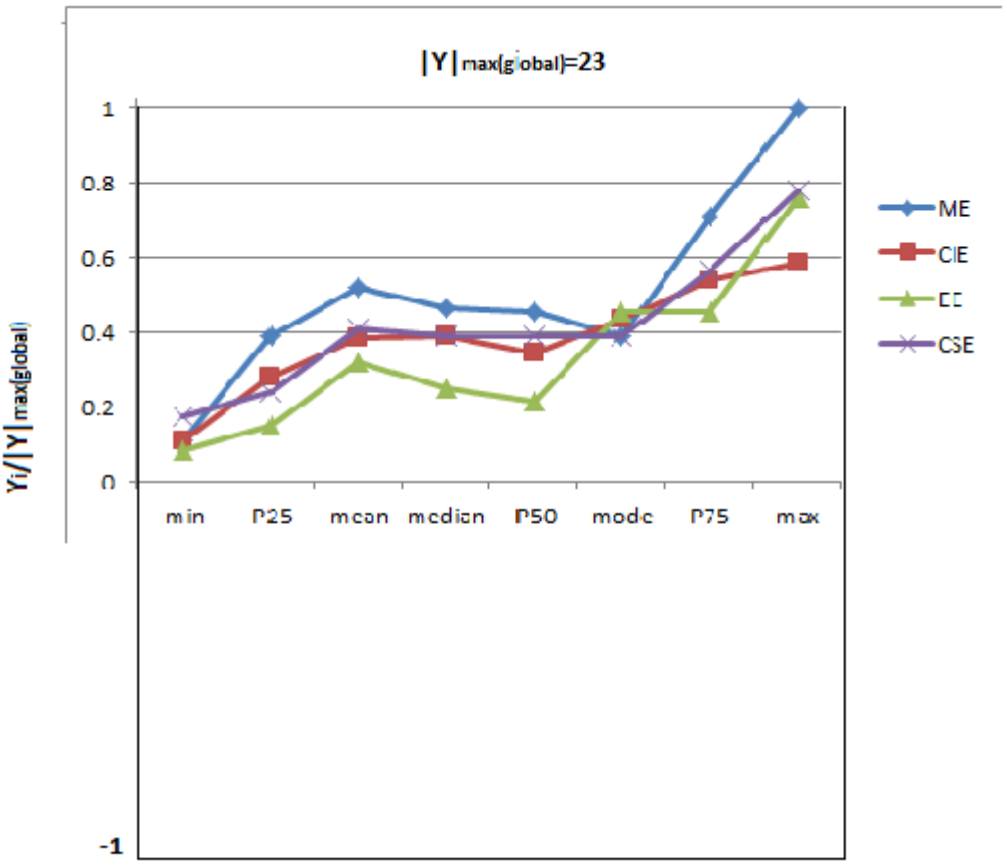
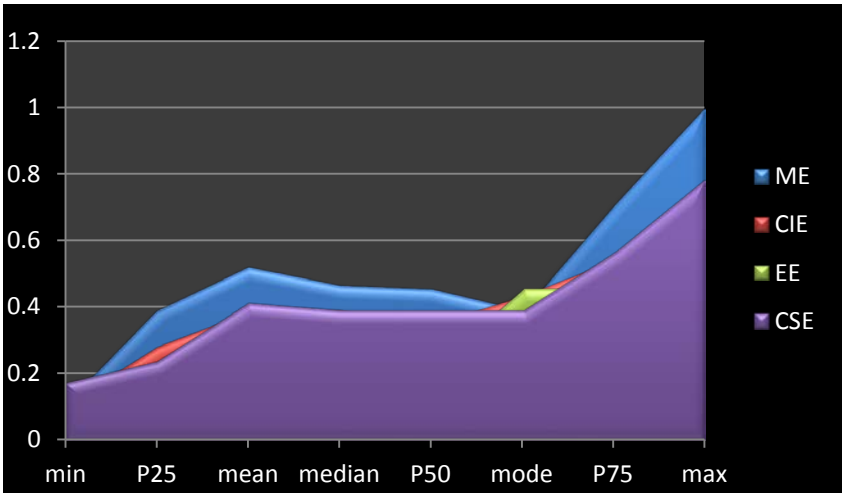


Fig: Representation of the tabulated data excluding sd . Joining all the Y_i's plotted in the graph.

We can have area under the curves plotted above as follows:



One can easily observe the area under the curve of respective branches(ME,CIE etc.)

ISSN 2348 – 7968

It is distinct that the area under ME curve is maximum and hence represents better set of data which can be verified by observing the table also.. [note "NOTE : C " above]

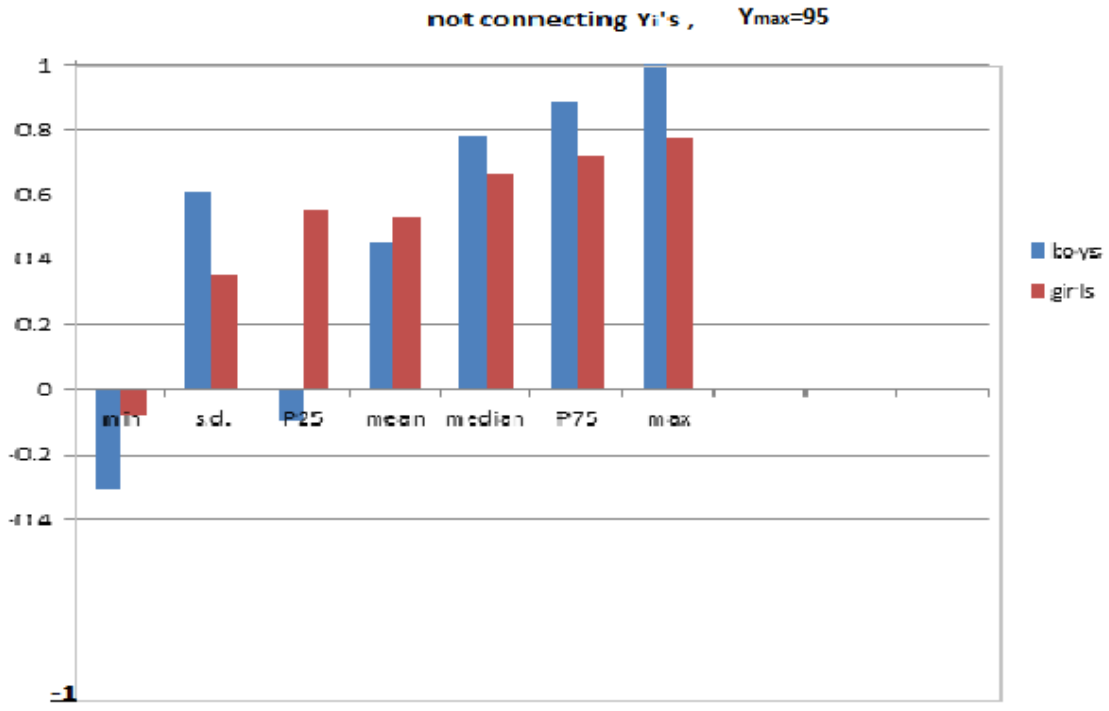
2.Let us consider the following table of marks obtained by 5 male and 5 female Candidates in a competitive exam out of 100 marks be as follows:

CANDIDATES	MARKS SECURED
Boys	-30,-10,75,85,95
Girls	-8,54,64,69,74

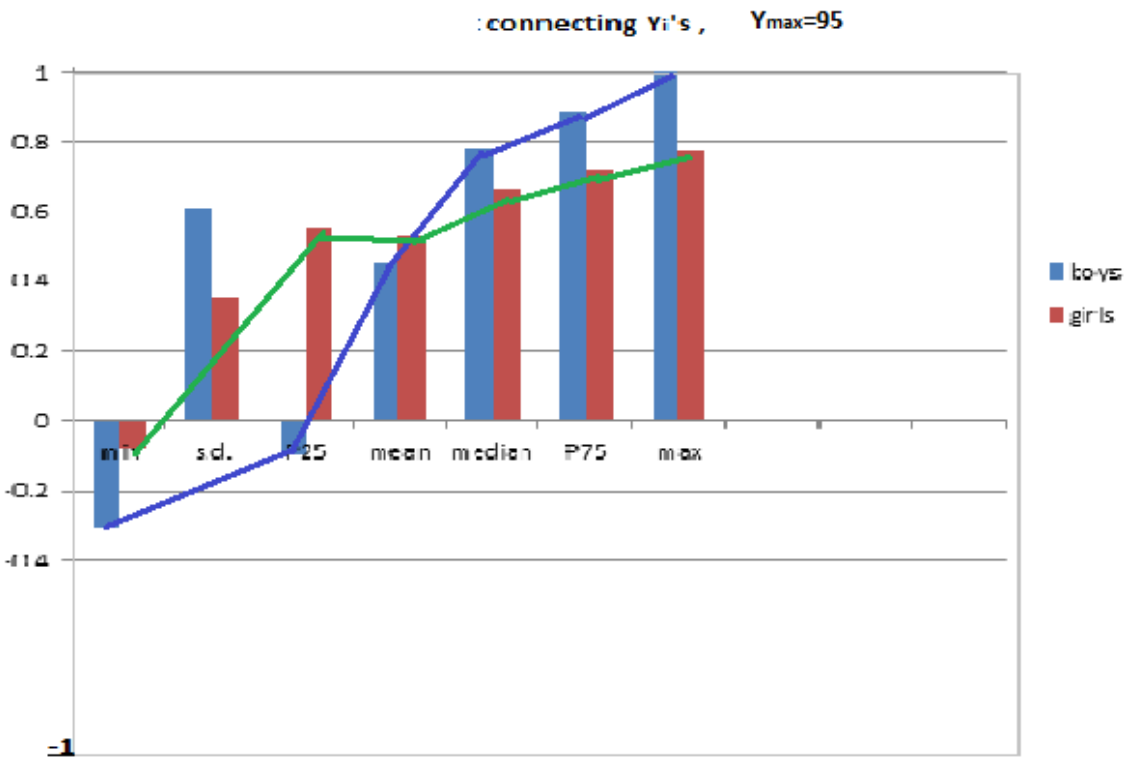
Here we have the following table of calculated data:

Candidates	Data of marks obtained						
	min	s.d.	P_{25}	mean	median	P_{75}	max
Boys	-30	58.3	-10	43	75	85	95= $\|Y\|_{max}$
$(Y_i/\|Y\|_{max})$	-0.31	.61	-0.10	.45	.78	.89	1
Girls	-8	33.5	54	50.6	64	69	74
$(Y_i/\|Y\|_{max})$	-.08	.35	.56	.53	.67	.72	.77

Applying the above rules in bar graph not connecting the Y_i to each other we Have the following graph



ISSN 2348 – 7968



In the first graph Y_i's are not connected but in the second graph these are connected except for deviations which could have been connected to the origin or to another deviations point by an another line if there had been other such point. In the second graph it can clearly be seen that the positive area under the curve for girls Is greater than the curve for boys performance of girls can be said to be better than the Boys. It also proved by standard deviation bar also that for boys values are more deviated.

DRAWBACK OR NOT:

It always represent the data with the motion from minimum to maximum. Infact this feature is very helpful for many situations . Now let us consider the following table:
For two farmer of two different countries

Country	Money earned	
	January (2014)	February(2014)
India	Rs 3000	Rs 6000
USA	$3000 (=Rs3000*61.28= Rs183840 , as on 26-10-2014)	$2500 (=Rs2500*61.28= Rs 153200)

ISSN 2348 – 7968

It can be easily observed that numerically both the farmers earned the same in January but the American farmer earned less than the Indian farmer in February. But obviously the American farmer earned more money than the Indian farmer. Actually here the performance of the indian farmer is increasing but it won't be expressed well in the procedure so far discussed . But the above procedure is to have a overall representation of data which would obviously be expressed in the graph that the American farmer hasearned much more than the other in terms of rupees.

If we want to focus our attention in performance not bothering about the actual money earned
Then we can represent the graphs following steps or rule (1,2,4) mentioned and since we are now related with individual data set so Y global maximum is to be replaced by Y local maximum. One thingto be noticed is that while representing the base coordinate we would follow order of base coordinate
we choose (eg. here the order of month)

Now $|Y|_{max(global)}$ would change to $|Y|_{max(local)}$ then we would have representation as following:

Case1: for the Indian farmer:

$|Y|_{max(local)}$=Rs 6000.

Case2: for the American farmer

$|Y|_{max(local)}$=$3000

Country	Money earned(in respective) month	
	january	february
India	Rs 3000	Rs6000
$Y_i/|y|_{max(local)}$	0.5	1
Usa	$3000	$2500
$Y_i/|y|_{max(local)}$	1	.833

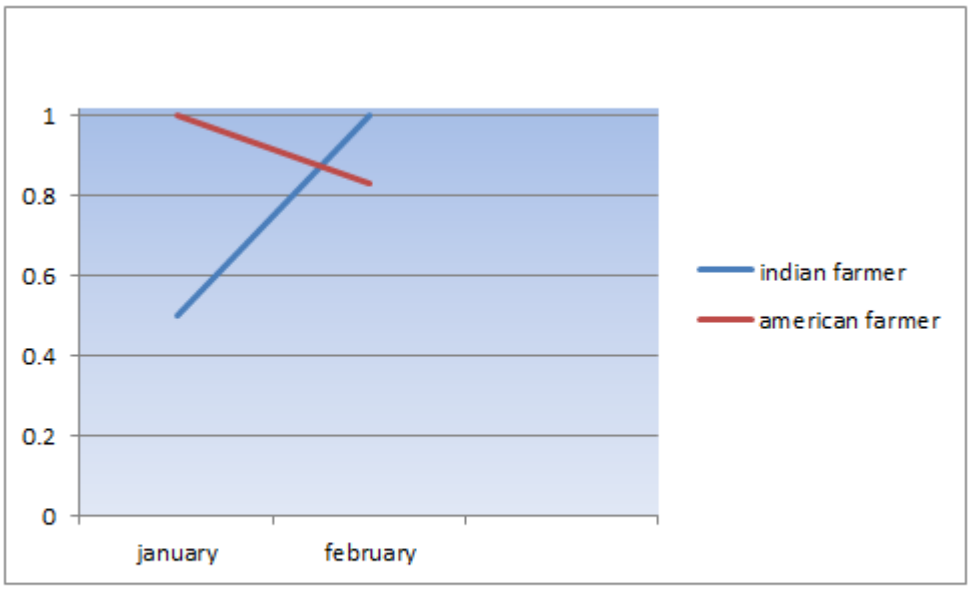


Fig: relative representation of performance of two farmer in terms of improvement without bothering the amount of actual money earned.

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 3, March 2015.
www.ijiset.com
ISSN 2348 – 7968

REPRESENTATION OF GRAPH TO DETERNMINE BETTER DISTRIBUTION OF DATA

(INDEPENDENT OF TIME) :

If we have to compare the earning of farmer of two different period of time , it would not be fair to directly plot the graph because value of money differ along with time. For e.g Rs1000 in 1947 is economically not same with Rs 1000 in 2014.Let us consider the following table of money earned by 5
different farmer of a place in two different period of time:(assumed to be in a remote area)

Period of time	Money earned in rupee
1947-1948	185,200,240,250,300
2013-2014	7500,8000,12000,15000,25000

For representing this situation we would represent the graphs using steps or rules (1-6) except Y global maximum is to be replaced by Y local maximum. Now $|Y|_{max(global)}$ would change to $|Y|_{max(local)}$ and each entries in respective time period is to be divided by $|Y|_{max(local)}$ before plotting in the graph . Since both groups of farmer are independent in earning as well as money value also varies from time to time so it is better not to divide the quantities by $|Y|_{max(global)}$.

Here we have the following table of calculated data for the given table:

Period of time	Earning in rupees				
	min	sd	mean	median	Max
1947-1948	185	45.27	235	240	300
$Y_i/\|Y\|_{max(local)}$	.61	.15	.78	.8	1
2013-2014	7500	7123.9	13500	12000	25000
$Y_i/\|Y\|_{max(local)}$	.3	.28	.54	.48	1

ISSN 2348 – 7968

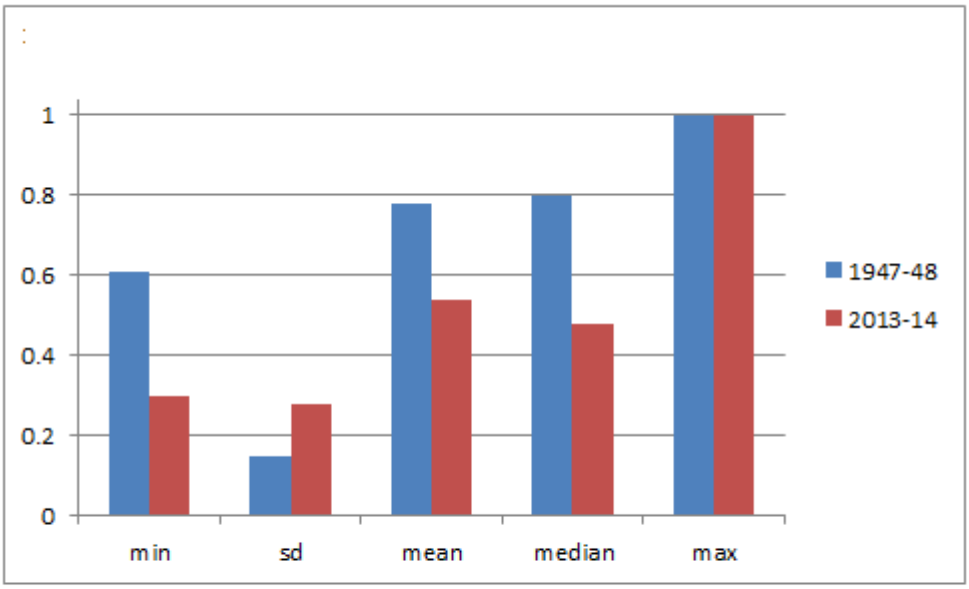


Fig: relative representation of graph independent of time. The graph shows that for 1947-48 distribution is better which even can be verified practically by comparing the economic condition of market in the respective period of time.

We can also visualise the graph under area graph connecting all Y_i's except s.d. as follows:

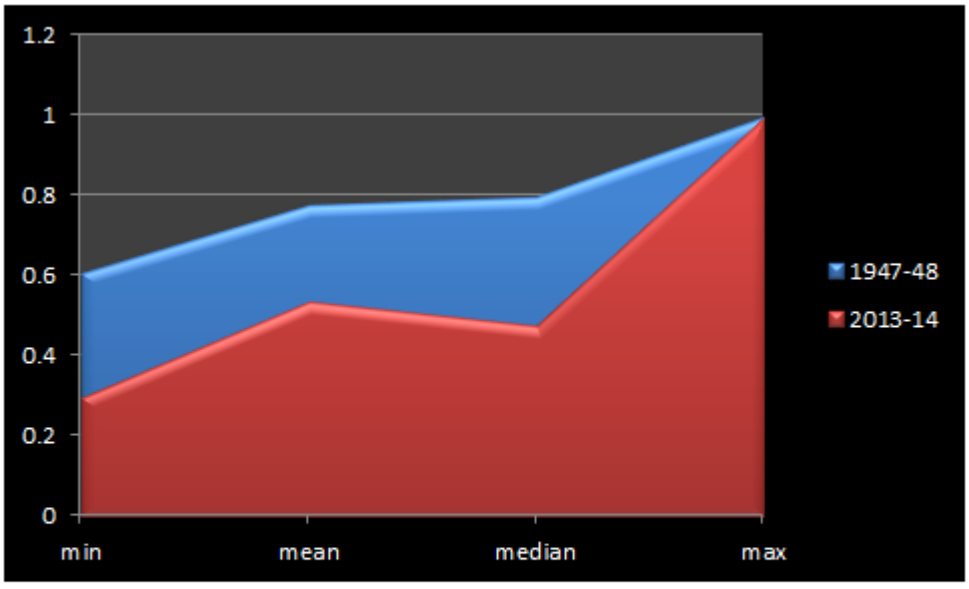


Fig: Graph showing the area under the curve joining respective Y_i excluding sd . it is distinct that area under the curve for 1947-48 is more hence better distribution of data than for 2013-14

ISSN 2348 – 7968

CONCLUSION:

In most of the graph (for e.g. bar graph or histogram), entries are shown in interval with frequencies. This makes it difficult to determine maximum value, minimum value, mean etc.This problem can be circumvented by representing a graph(following the already mentioned first six steps) which results inthe estimation of values formean, median etc. by studying the graph directly without performing any calculation. Moreover, values for maximum value, minimum value, and deviations measures etc. presented in the graph makes it easier to have a better conclusionabout any set of data. One important feature ofthis representation is that the graph represents the corresponding values of basecoordinatesstarting from minimum to maximum (it may be referred to as a directionalrepresentationof data) creating an order in the data. Also the user of this procedure candefine quantitiesand include as per convenience following the specified rules or steps.