P: ISSN NO.: 2321-290X RNI : UPBIL/2013/55327 VOL-5* ISSUE-6* February- 2018
E: ISSN NO.: 2349-980X **Shrinkhla Ek Shodhparak Vaicharik Patrika**

Behavioral Analysis of Statistical Data- Causes and Consequences

Kaustubh Narayan Misra
Associate Professor,
Deptt.of Geography,
Buddha P G College,
Kushinagar,
DDU Gorakhpur University,
Gorakhpur, U.P.

Abstract

Whereas statistics and data analysis procedures generally yield their output in numeric or tabular form, graphical techniques allow such results to be displayed in some sort of pictorial form. They include plots such as scatter plots, histograms, probability plots, spaghetti plots, residual plots, box plots, block plots and bi plots and so on. Exploratory data analysis relies heavily on such techniques. They can also provide insight into a data set to help with testing assumptions, model selection and regression model validation, estimator selection, relationship identification, factor effect determination and outlier detection. In addition, the choice of appropriate statistical graphics can provide a convincing means of communicating the underlying message that is present in the data to others. If one is not using statistical graphics, then one is forfeiting insight into one or more aspects of the underlying structure of the data. In this regard in the field of Geography, this is an important factor to exercise the analysis and explanation of the facts and findings. This study deals the importance, causes and consequences of the Graphical Representation of Statistical Data.

Keywords: Statistical, Graphical, Representation, Graphs, Diagrams, Logarithms, Practical, Bar, Smooth, Pie, Star.

Introduction

To represent the statistical data in graphical manner, first it is necessary to choice the base for that particular graphical representation. It should be-chart, sheet, graph, logarithmic graph, sheets like-butter, transparency etc. On these bases, we make graphs, diagrams and maps according to our academic needs and available figures. In practical aspect of Geography, different types of graphs diagrams and maps uses for different purposes of the study; but it must be also noted that there is a important aspect of theoretical base of representation of statistical data. Earth is habitat and earth's resources are the heritage of man. Utilization of the earth's resources by the man is the main feature of present materialistic age. An appraisal of the resources means the preparation and maintenance of records of the resource utilization. This record is invariably in statistics or figures. Generally these figures are so vague and sometimes so confusing that no definite conclusion can be deduced from those figures. So the graphical representation of the statistical data (GRSD) is a better method to represent these figures and result for more and more convenience and correct.

Statistical graphics have been central to the development of science and date to the earliest attempts to analyze data. Many familiar forms, including bi-variant plots, statistical maps, bar charts, and coordinate paper were used in the 18th century. Statistical graphics developed through attention to four problems. 1. Spatial organization in the 17th and 18th century. 2. Discrete comparison in the 18th and early 19th century. 3. Continuous distribution in the 19th century. 4. Multi-variant distribution and correlation of the data in late 19th and 20th century. Since the 1970s statistical graphics have been re-emerging as an important analytic tool with the revitalization of computer graphics and related technologies. In this process and study, it is important to note every time that there are the proper relations and correlations between the traditional cartography and recent Remote Sensing, Ariel Photogrammetric and Computer based cartographic analysis like Geographical Information System.

Aims of the Study

The main aim of this particular study in that how, graphical representation of statistical data able to explain the facts in better way and how, a lay man or common people understands things properly without any hesitation. With the help of empirical method of the study, we try to

understand and examine the reality of graphical representation. The main objective of this study deals herewith accordingly. Graphical statistical methods have four objectives. A- The exploration of the content of a data set. B- The use to find structure in data. C- Checking assumptions in statistical models. D- Communicate the results of an analysis.

Data and Methodology

Generally, there are four stages of the study in geography –

1. Discovery of the phenomenon and facts by surveying, observing and studying.
2. Classification of the phenomenon and facts,
3. Evolution, origin growth of the phenomenon and facts and
4. Correlation and interrelation between and among the phenomenon and facts – it can be divided into three 1 – Cause-effect relationship 2- functional relationship and 3- Regional/ areal relationship.

For all the three Graphical Representation of Statistical data is needed. The study is based on the data available from both the primary and secondary sources. For this study the production of different crops, crop's acreage related data have been collected by the both the sources and converted the data into different type of graphs for the result in analytical form, not in pictorial form..

Review of the Literature

Andrienko, N and Andrienko, G in 2005 gives its details in his work 'Exploratory Analysis of Spatial and Temporal Data- a Systematic Approach. A Buja, D Temple Lang, H Hofmann, H Wickham and M Lawrence has explains the importance and theoretical behavior of Graphical Representation of Statistical Data in his contribution 'Interactive and Dynamic Graphics for Data Analysis'. In the edited book of Hoaglin, D C, Mosteller, F & Tukey, John Wilder in 1985, they emphasized the exactness of the data analysis in their work 'Exploring Data Tables, Trends and Shapes'. Hoaglin, D C, Mosteller, F & Tukey, John Wilder's in their edited book in 1983 'Understanding Robust and Exploratory Data Analysis'. Isenberg, Alfred in 2009 in 'Parallel Coordinates: Visual Multidimensional Geometry and its Applications' and Leinhardt G, Leinhardt S in their work 'Exploratory Data Analysis: New Tools for the Analysis of Empirical Data, Review of Research in Education' explains the graphical methods in details. Martinez W L and Martinez A R & Solka J in 2010 in the work 'Exploratory Data Analysis with MATLAB; Theus M, Urbanek S in 2008 in 'Interactive Graphics for Data Analysis: Principles and Examples'; Tucker L, Mac Callum R in 1993 in the work 'Exploratory Factor Analysis' and Tukey John Wilder in 1977 presented the 'Exploratory Data Analysis. P F Hoaglin D C 1981 in the book 'Applications, Basics and Computing of Exploratory Data Analysis'; Young F W Valero-Mora P and Friendly M in 2006 by 'Visual Statistics: Seeing your data with Dynamic Interactive Graphics'; Jambu M in 1991 by 'Exploratory and Multivariate Data Analysis'; S H C Dutoit, A G W Steyn, R H Stump in 1986 by 'Graphical Exploratory Data Analysis' and Paul J. Lewi in 2006 'Speaking of Graphics' have presented the pictoral scale based graphical representation.

Edward R Tufte in 2001 and 1983 with the work 'The Visual Display of Quantitative Information'; Denis, Dan in 2005 in 'The early origins and development of the scatter plot'; Stephen B in 1994 with work 'Basic Statistics' explained about, when we search for a relationship between two variables, a standard graph of the available data pairs (X, Y), called a scatter diagram, frequently helps this explanatory note. William in 1993 in the work 'Visualizing data'; Nancy R Tague in 2004 by 'Seven Basic Quality Tools' and Green Walton A, Schoerke Barrette, Crowley Jason in 2013 by the work 'The Generalized Pairs Plot' given the behavioral explanation of representation of statistical data. Singh Raghunandan & Kanaujia Lekhraj Singh in 1972 in his work 'Map work and Practical Geography'; Robinson A H & Bryson R A in 1957 in the work 'A Method for Describing Quantitatively the Correspondence of Geographical Distribution' and Srivant S and Phadke V S in 2001 by their work 'Cartographic Representation of the results of Chi-square, Analysis for Spatial Explanation' are pioneer works in the field of cartography. Yadav H L in 2000 in his Hindi book 'Pryogatamak Bhoogol Kay Aadhar'; Singh R L & Singh Rana P B in 1991 in 'Prayogatamak Bhoogol Kay Mooltatwa are the main contemporary work in this field.

Study Area

Mahrajganj district acquired the status of an independent district on 2nd Oct. 1989. Earlier, it was the sub-division of Gorakhpur, known as Maharajganj Tahsil. At present Mahrajganj is a chief town and headquarter of the district.

The region has the location advantage in the heartland of Bihar. It extends from 25^0 29' N to 26^0 02' N latitudes and from 85^0 05'E to $85^0$38' longitudes and comprises and area of 1995. I square km. with total population of 2112716 (2001). It touches Nepal in the North; district of Kushinagar in the East, Siddharthnagar in the West and Gorakhpur in the south. This District separated from Nepal by International Boundary.

Discussion

Following explanation, verification of different graphs on the basis of particular data of a district has been made. On the basis of these explanations, we will be able for the conclusion of the study matter.

Smoothed Graph

In this study the base uses, is logarithmic graph and the data plotted on this base is, block wise area under different cereals in hectare. The diagram shows a clear-cut picture of the said data and common man can also evaluate facts in detail with the help of this diagram. But carefulness require for two purposes – 1. More than three types of data should not be plotted on single X, Y axis. 2. If plotting is necessary of more than three categories of data, one should take bigger size of logarithmic sheet; otherwise we cannot find complete and clear-cut result.

Compound bar Graph

When the bars show one item of data, it is called a simple bar Graph; but when they are

employed for showing two or several items, they are known as Compound Bar Diagrams. They are intended for graphic representation of not only the variation in the total of a given classification of values, but also the changes in the component parts of the total. On the plane graph, block wise percentage area under different cereals – Rice, Wheat, Maize, Barely, Bajara and Jwar; plotted with the help of compound bar graph. If we are careful on three points, we can touch the target authentically.

Multiple Line and Bar Graph

In this graph both bars and smoothed line graph used jointly. Block wise per hectare production of different cereals in quintals have been represented on vertical axis and areas on x-axis. For the presentation of all the six crop in one graph was critical; but it was necessary, because of comparative study among all the cereals. When, however, time is fixed and the unit of area varies, lines are replaced by bars. The data involved are shown by bars of equal width and proportionate lengths. In some cases if different unit of data is necessary to represent on one graph, bar and line may be represent in one graph. The arrangement can be easily interchanged and bar & line will become horizontal-vertical inspite of vertical-horizontal.

Multiple Lines Graph

In this graph, total production of cereals of Maharajganj District present in tons by multiple line graphs. Blocks are plotted on x-axis, while different cereals are presented on vertical axis. Line graph is most suitable for representing a sequence of values corresponding to successive points. In this graph the trend of production or utilization is represented by a smooth crop of region, it will be plotted on simple line graph. But when such production of different crops, if required to represent, it will be plotted on poly or multiple graph. The polygraph is most suitable for showing comparative trends or casual relationship of Agricultural Production.

Band Graph

A band graph is virtually a line graph, sub divided to show the components thereof and consequently consists of several bands formed by sub-divisions and shaded in different tints as shown. When the variations in components are not too divergent, the graph gives a good impression, regarding their trends, but when the fluctuations are two wide, its legibility is marred. In the latter case, it is better to take recourse to separate line graphs. In this particular case block wise are under different cash crops in hectare of Maharajganj District for different blocks are presented. There are five cash crops, which can be plotted in this diagram. Blocks are on x axis area in hectare on vertical axis.

Vertical Compound Bar Graph

Block's percentage area of Maharajganj District under different cash crops can be presented by the help of Vertical Compound Bar Graph. Blocks shows on horizontal x-axis, while percentage crops on vertical axis. When the bars shows only one item, it is called a simple vertical/ horizontal Bar Graph; but when they are plotted for showing two or more items, they are known as compound vertical/ horizontal bar graph. They are intended for graphic representation of not only the variation in the total of a given classification of values; but also the changes in the component parts of the total. This is the exact example of horizontal compound bar graph. The differences between the vertical compound bar graph and horizontal compound bar graph can be easily seen. In this figure block wise per hectare production of different fruit crops in quintals are presented on y axis and blocks' of Maharajganj District are presented on x-axis.

Logarithmic Multiple Line Graph

In this graph block wise per hectare production of cash crops of different blocks of Maharajganj District are presented. Where blocks are on x-axis and production values of different cash crops are on y-vertical axis. In general graphs, scale increases according to geometric ratio, while in logarithmic graph it increases according to mathematical ratio. Generally simple graph divided into the 10^{th} part of inch; but logarithmic graph divided/ increases by the static ratio of 10. For the example – If first circle start from 0.1 and end 1.0, then second start 1.0 and end to 10 and so on. In logarithmic graph, the scale of both axes is logarithmic; but in semi-logarithmic, only vertical or y-axis is logarithmic. In this semi-logarithmic graph, block wise total production of different cash crops in tons of different blocks of Maharajganj District are presented. In this figure total production of different cash crops in tons are presented by y-vertical axis, while blocks are on x-horizontal axis. Also in another figure, block wise production of different fruit crops in tons are presented; where production presented on vertical axis and blocks are on horizontal axis.

Star Graph

Basic difference between these two is that in first values of different fruit crops area in hectare are presented on different diverted (according to determined scale) x-axis; while in second block wise percentage are under different fruit crops are presented. It is a known fact about the star graph that, in this graph, there is no x-axis; from 0 point according the number of x-axis, base data (y-axis) extends radial lines according to the scale. If we joint all the values of different base data, we can find a star. Three stars may be seen for different three crops. Star graph can be used by both the graphs methods-General Graph and logarithmic graph according to the need and available data.

Conclusion

After the completing the serious examination of all the said graphs and figure, one can easily conclude following conclusions.

1. Selection of graphs, diagrams among the different graphs and diagrams must be very careful according to the need available data. Correlation between the data and selected graph, where the data is plotting is necessary for better result, which is the main aim full result of graphical representation of statistical data.
2. Appearance of the complete graph or diagram is also a necessary thing in the case of graphical representation of statistical data. It must be as

much as clear from the tabulated/ calculated data. With which layman can understand the need of graphical representation.

3. Proper use x and y axis is another important point and it should be taking consideration carefully. If one should use horizontal and vertical axis, it creates problem in some cases. If you selected vertical bar graph for a data and some time you select horizontal bar graph for the same or other data, it creates problem. In any circumstances x axis represent base data, while y represent dependent data.
4. However, it should be noted that all the methods for graphical representation of statistical data are not equally appealing and this is because of the limitations associated with individual method used for this case chose, when it is highly required. Graph, diagram or map should in that case chosen, when it is highly required; other than this, more simple method of graphical representation of statistical data should apply. Here it is also necessary to mention that GRST must not be used unless or until, it is not highly required.
5. Carefulness require for two purposes – 1. More than three types of data should not be plotted on single X, Y axis. 2. If plotting is necessary of more than three categories of data, one should take bigger size of logarithmic sheet; otherwise we cannot find complete and clear-cut result. Here it has been proved that Geographical facts and Graphical Representation of statistical data are closely associated with one another and it (GRST) provides chances for common public of the appearance of Geography.

References

1. *Singh Raghunandan & Kanaujia Lekhraj Singh (1972): Map work and Practical Geography, Central Book Depot, Allahabad. pp. 17-38.*
2. *Robinson A H & Bryson R A (1957): A Method for Describing Quantitatively the Correspondence of Geographical Distribution, Annals AAG, 47. pp. 30-34.*
3. *Srivant S and Phadke V S (2001): Cartographic Representation of the results of Chi-square, Analysis for Spatial Explanation Transactions 11G-23 (1 & 2). pp. 07-22.*
4. *Yadav H L (2000): Pryogatamak Bhoogol Kay Aadhar, Radha Publications, Dariaganj, New Delhi. pp. 23-78.*
5. *Singh R L & Singh Rana P B (1991): Prayogatamak Bhoogol Kay Mooltatwa, Kalyani Publications, New Delhi. pp. 100-122.*
6. *Andrienko N & Andrienko G (2005): Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach. Springer. ISBN 3-540-25994-5, pp. 34-78.*
7. *Hoaglin D C; Mosteller F & Tukey John Wilder (Eds) (1985): Exploring Data Tables, Trends and Shapes. ISBN 0-471-09776-4. pp. 105-134.*
8. *Leinhardt G, Leinhardt S (1980): Exploratory Data Analysis: New Tools for the Analysis of Empirical Data, Review of Research in Education, Vol. 8, 1980 (1980), pp. 85–157.*
9. *Martinez W L, Martinez A R and Solka J (2010): Exploratory Data Analysis with MATLAB, second edition. Chapman & Hall/CRC. ISBN 9781439812204. pp. 13-47.*
10. *S H C Dutoit, A G W Steyn and R H Stump (1986): Graphical Exploratory Data Analysis. Springer, ISBN 978-1-4612-9371-2. pp. 25-65.*
11. *W S Cleveland (1994): The Elements of Graphing Data. Summit, NJ, USA: Hobart Press. ISBN 0-9634884-1-4. pp. 12-19.*
12. *Edward R Tufte (1983 & 2001): The Visual Display of Quantitative Information (2nd ed.). Cheshire, CT, USA: Graphics Press. ISBN 0-9613921-4-2. pp. 16-23.*
13. *Brus Jan, Vondrakova Alena, Vozenilek Vit (2015): Modern Trends in Cartography; Springer, ISBN 978-3-319-07926-4, pp 10-14.*
14. *Gartner georg, Jobst Markus, Huang Haoshency (2016): Progress in Cartography, Springer, ISBN 978-3-319-19602-2, pp 13-17.*
15. *Riva Sanseverino Eleonova, Riva Sanseverino Raffaella (2017): Smart Cities Atlas, Banglore, ISBN 978-3-319-47361-1, pp 21-26.*