ANALYSIS OF AGGREGATE DATA

*K. S. Shukla**

Introduction

HUMAN BEINGS have always been anxious and curious to know and understand the environment around them. The kind and degree of interest regarding environment has generally been conditioned or guided by the nature of philosophy for life which a nation, a region or an individual adheres to. The level of anxiety and curiosity has invariably been influenced by the immediate or long term professional or personal involvement of the people. Depending upon the interest, resources, ability and potentiality, the shades and dimensions of interest may be narrow or wide. Accordingly, one or more dimensions of an issue have been explored/understood/assimilated/adopted by the human race in space and time.

The quest for understanding the environment has been perpetual because human beings have always been concerned (a) to manage their environment and plan strategies to make optimum utilisation of societal resources and to contain the recurrence of deleterious situations on the health—ecological, physical, social *etc.* of a nation or a region, and (b) to understand various dimensions of social functioning affecting their living patterns. The ramifications stated earlier had been of interest to all but more attention to these aspects has been, more often than not, given by those who had a greater stake—altruistic or selfish—in an orderly and smooth functioning of a society or a region. In this regard the perspective reflection synthesis of the views of the thinkers and practitioners of earlier generations are made available to the posterity in the epics or in other relevant works.

Depending upon the region, responsibilities, orientation, training, resources and constraints, the available aggregate data** have been used regarding a segment of a society or on a social issue or on an area. These data may relate to one or more events or situations that attempt to analyse the ethnic/racial/religious/caste characteristics or may be a detailed case study of persons.

---

*Reader, Institute of Criminology and Forensic Science, New Delhi.

**Available detailed account in the form of a biography, autobiography or communications, or overall figures at local/regional/state/central level that give an indication regarding the nature of administration or mode of operation of a particular sector/segment/issue/phenomenon/time would constitute aggregate data. These could be generated by a specialised agency/department—official or non-official—at mandatory or voluntary level by recording a major or a minor incident/case/event, as it comes to the notice of that agency/department for scrutiny on a continuous basis.

Available aggregate data may provide a focus on, and important insight into, a particular social issue, because it is one of the precise methods that could form a base for further discussion. Consequently, the similarities and differences in the mode of operation, effectiveness, relevance *etc.* of a system under study could easily be known, as also the relations between various variables of a system or data highlighted.

These data could be analysed on an extensive or intensive level depending upon the priorities, resources and objectives of a social investigation. Their analyses have generally been undertaken to assess the nature, evaluate the extent, understand the trend and scrutinise the pattern of a particular phenomenon/process with a view to plan strategies, formulate long or short term policies and take effective and appropriate steps for control/prevention.

It has generally been argued that depending upon the utility and efficacy of the steps taken by a system, this reality invariably reflects or is an indicator of state's capability to be sensitive to the needs and demands of its citizenry and maintain order or guarantee smooth operation within its territorial jurisdiction. Therefore, all governments, in space and time, have used aggregate data to project their effectiveness.

Sources of data

Different types of records and writings, *i.e.*, autobiography, biography, correspondence, descriptive accounts, *shastras.*, *neetis etc.* could form a part of aggregate data, but available empirical data on national or regional level have been one of the important and more often used sources of reference both by intellectuals and administrators and by those who are interested in the welfare of a society or a region.

Aggregate data could be collected, collated and compiled by a researcher with the help of various tools of social research like an opinionnaire, inventory, questionnaire/schedule and interview. They may also be collected and compiled, on a continuing basis, by official or non-official agencies with the help of infrastructure available with a department at field level. These data could be forwarded to a centralised agency which may collate, codify and publish them for a wider use. The secondary sources of data could be a private organisation like a voluntary agency, a private concern, or a specialised department of the local/state/central administration. The utility of data should generally depend on the objective and plan of a study, coverage, period for which such data are available, nature of data *etc.* The objectives of compilation of data or design of a study would ultimately influence the phenomenon of compilation and collection of data. The earlier stated variables would reflect one or more dimensions of a pehnomenon which is being studied thereby affecting the nature and level of utility of data.

There have been various types of official and non-official sources of data. They may relate to population, national income, gross national product, food, agriculture, health, irrigation, industry—heavy, small, automobile *etc.*, commerce, textile, housing, transport, defence, labour, post and telegraph, education, crime *etc.* Partial data available with various administrative departments—private or public—could also form a part of aggregate data. Each one of the sources stated earlier provides vital data regarding one or more dimensions or shades of social life of a nation. All comparisons, therefore, are made on the basis of the data available with different agencies. The periodical use of aggregate data has been extensive as every system has attempted to show its effectiveness on the basis of a comparative study of the data in a specified place at a particular point of time by focusing on the similarities or differences/fluctuations in the trend and pattern of a particular phenomenon.

Utility

The aggregate data are generally of interest to various sections of society and have been used by them to make their argument valid with partisan or non-partisan motivations. Such data have proved their usefulness to all sections of society particularly to (*a*) administrators—religious, political, social, bureaucratic, educational *etc*; (*b*) researchers; and (*c*) citizens.

The administrators have used aggregate data to assess existing social, political, religious or other ethos of a nation with a view to locate gaps in social responses in space and time and to determine priorities for action. Invariably these exercises help evaluate existing measures that assist in planning future strategies and provide indicators regarding effective remedial measures.

Researchers have generally made use of these data to understand the prevailing ethos; assess the nature of a social organisation and social realities; indicate the priorities in research as per the compulsions and needs of a social organisation; and provide projections regarding the nature of social issues including their intensity, extent, patterns and dimensions in terms of the shape and duration as also regarding emerging social realities. Researchers have attempted, with the help of aggregate data, to point out or critically evaluate the policies, precepts and practices of a system in terms of their efficacy and relevance. They have also attempted to gauge the impact of research efforts in the light of existing social realities and demands. Research undertaken with the help of aggregate data, coupled with realistic interpretation, could be useful both to administrators and researchers in future.

The interest of the citizenry in aggregate data has generally been conditioned by the position of the evaluator in a social order or the ideological commitment of the commenter. This interest has, therefore,

generally been superficial. As a consequence it may not be advisable to attach undue importance to such views for they may not have been considered at all. The nature and interpretation of aggregate data, in general, by citizens either infuses confidence in the people or demoralises them or develops a sense of helplessness. Such exercises, devoid of adequate home work, generally lead to rumour mongering which may have extensive and intensive social repercussions.

Analysis and interpretation

Data in any form are generally neutral. Their direction and trend is generally highlighted/focused/reflected with the help of analysis and interpretation. Dependability, relevance and efficacy of data would invariably be conditioned by the scienticism, understanding, sensitivity, knowledge and insight of the researcher regarding the ethos of the social reality of the segment which is being studied at a particular point of time.

The initial step, before the analysis and interpretation work has been launched, would be to define the concepts in clear-cut terms. Then the objectives of research could be carved out by evolving realistic hypothesis which invariably should be done after scrutinising and making a preliminary study of all available raw data.

After an initial scrutiny of the raw data, the researcher should endeavour to prepare his plan for analysis to indicate the variations or similarities in data. Initiation of analysis, of course, would presuppose that the researcher has pondered over the subject and done adequate home and library work. While analysing data, total involvement and concentration of the researcher is essential. He should devote sufficient time on each table so that he may work out all possible combinations for the interpretation of each table.

It is, therefore, essential that the researcher should have a total grasp of the data which he proposes to undertake for interpretation. When data have been collected from all the sources, in consonance with the objectives of study, they need to be placed in different categories. The categorisation should be utilitarian. After placing them in different categories, they need to be arranged in an order. Ordering of data should be done from an ascending order to descending one or *vice versa*. The tabulation and ordering should have internal consistency or logic. The data so analysed could be arranged in the form of tables and/or graphs/charts *etc.* with a view to sharply project the qualities and characteristics of the subject under study.

Since all aggregate data are intertwined, the researcher should learn to interpret data. As per the design or planning, he should manipulate data in such a way that it projects and highlights those aspects/dimensions that require further manipulation/analysis. The inferences, as drawn by him after arranging data in different categories

and groups, should be pondered over. Should he so desire, he may consult earlier available studies on the subject, or interview a section of informed people on the topic. From all the manipulations and placement of data in a logical consistent order, he should evolve a summary of data. After evolving a mental summary of the report, he may be able to scrutinise further the available data with a view to determine as to how much of the available data would be relevant which could be used in a particular analysis/interpretation.

During interpretation of data, the researcher may start from the generalisations to concrete specifics and from specifics to concrete generalisations. After working out all possible interpretations he may develop paradigms/concepts from the concretes. These concepts may be aligned or linkages provided with concrete situations. Existing theories/principles on a particular issue could be evolved after an extensive analysis of data backed by other sources of information.

Precautions

It is desirable to undertake an analysis of aggregate data regarding a segment/issue/person with which the researcher has familiarity which may be in respect of its ethos or mode of operation or any other aspect.

The researchers should study and scrutinise all available literature, and record on an issue which he proposes to analyse before launching the analysis and interpretation work.

Data should be analysed in different ways with a view to work out various combinations. He should endeavour to work out all possible arrangements by benefiting from the experiences of others and should be prepared to modify his plan as per the suggestions/experiences of knowledgeable people. While interpreting data, he should continue to be a keen observer with a view to verify and test interpretations in the context of existing social realities.

During analysis of data he should keep in view the interest of the society and organisation (whose data is being analysed) and, as far as possible, should not compromise with his interests. He should be equally conscious of the consequences of his interpretation. Therefore, the researcher should be very cautious regarding interpretation of sensitive issues which may have wider social implications. The circulation of such a study should also be done on a restricted basis.

The researcher is competent to analyse data, but as far as possible, he should desist from giving suggestions or predictions. After the study has been completed, it could be discussed at a forum or by a panel. This forum or panel should evolve/work out suggestions or predictions.

Interpretation of data, particularly of that which is to be used by administrators, should be written in a language which is familiar to them. Use of terminologies or complex words should, as far as possible, be avoided.

Every aggregate data is capable of highlighting plausible generalisations only; therefore, the processes involved in the operation of a segment may not be clearly brought out by the available secondary data. Consequently discussion/interpretation regarding processes should be avoided.

Limitations

One has to keep in view the following limitations while relying upon the aggregate data:

First, aggregate data should be taken only as an indicator reflecting a portion of existing social reality. Undue reliance or stress on it alone should be avoided.

Second, it would be desirable to keep in view that any information which forms a part of the record or statistics, as collected by different agencies, would generally be a crystallised record/statistics which had to be recorded by a particular agency, in spite of all pulls and pressures operating on that agency, because the pressure/compulsion for recording was greater.

Third, recording of data takes place after all the forces against it have been neutralised. Therefore, there is always a scope that a part of facts may not have been recorded or may have even been mutilated. Consequently the recorded figures invariably form part of the total reality.

Fourth, the recorder's prejudices/biases/ideological or social commitments may influence the mode of registration and consequently the overall figures.

Fifth, recorded data may have fallacies as the job is not done by trained professionals and biases involving age, sex, religion, caste, education, occupation *etc.* are likely to influence the aggregate data.

Sixth, organisation itself may have a latent stake or interest in the recording of data. Therefore, inflation or burking in the recording of data should not be ruled out.

Seventh, classified data may not be accessible to every researcher due to the compulsions operating on an organisation.

Eighth, data for only a particular number of years is preserved. Earlier data is invariably destroyed.

Ninth, every social issue cannot be analysed or understood with the help of aggregate data; therefore, a combination of tools of research may provide more realistic and relevant data.

References

Claire Selltiz, Marie Jahoda, Morton Deutsch and Stuart W. Cook, *Research Methods in Social Relations* 385-440 (revised volume 1964) (Holt, Rinehart and Winston, New York).

Claire Selltiz, Lawrence S. Writesman and Stuart W. Cook, *Research Methods in Social Relations* 455-507 (3rd ed. 1976) (Holt, Rinehart and Winston, New York).

Fred N. Kerlinger, *Foundations of Behavioral Research* 134-56 (1973) (Holt, Rinehart and Winston, New York).

Henry L. Manheim and Bradley A. Surion, *Sociological Research : Philosophy & Methods* 298-318 (1977) (Dorsey Press, Illinois).

James A. Black and Dean J. Champion, *Methods and Issues in Social Research* 403-25 (1976) (Wiley, New York).

J.P. Guilford and Benjamin Fruchter, *Fundamental Statistics in Psychology & Education* 1-14 (1973) (McGraw-Hill Book Company, Inc., London).

K.S. Shukla, "Studying Social Protest Situation: Some Methodological Implications", *Indian Journal of Criminology and Criminalistics* 40-46 (1981).

Margaret Staecy, *Methods in Social Research* 108-33 (1969) (Pergamon Press, Oxford).

S. Venugopal Rao, *Dynamics o, Crime* 24-44 (1981) (Indian Institute of Public Administration, New Delhi).

Thorsten Sellin and Marvin E. Wolfgang, *The Measurement of Delinquency* 1-26 (1970) (Wiley Eastern Pvt. Ltd., New Delhi).

Willian J. Goode and Paul K. Hatt, *Methods in Social Research* 313-59 (1952) (McGraw-Hill Book Company, Inc., New York).