

LECTURE 12 DATA COLLECTION

DR GAURAV SANKALP

DEFINITION

“Data are facts, figures and other relevant materials past and present serving as bases for study and analysis”.

MEANING OF DATA

The search for answers to research questions calls collection of Data. “Data are facts, figures and other relevant materials, past and present, serving as bases for study and analysis”.

TYPES OF DATA

The Data needed for social science Research may be broadly classified into:

- Data pertaining to human beings
- Data relating to organizations
- Data pertaining to territorial area.

Personal Data (relating to Human beings) are of two types.

- Demographic and socio-economic characteristics of individuals. Like name, sex, race, social class, relation, education, occupation, income etc.

Behavioural Variables: Attitudes, opinion knowledge, practice, intensions etc.

Organisation Data: - Consist of data relating to an organizations, origin ownership, function, performance etc.

- Territorial Data: - are related to geo-physical characteristic, population, infrastructure etc. of divisions like villages, cities, taluks, distinct, state etc.

IMPORTANCE OF DATA

The data serve as the bases or raw materials for analysis without Data no specific inferences can be drawn on our study. Inferences based on imagination or guesswork cannot provide

correct answers to research questions. The relevance, adequacy and reliability of data determine the quality of the findings of a study. The reliability of data determines the quality of research.

Data form the basis for testing the hypotheses formulated in a study. Data also provide the facts and figures required for constructing measurement scales and tables, which are analysed with statistical techniques. Inferences on the results of statistical, analysis and tests of significance provide the answers to research questions. Thus the scientific process of measurement, analysis, testing and inferences depends on the availability of relevant data and their accuracy. Hence the importance of data for any research studies.

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities, business, etc. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same.

The importance of ensuring accurate and appropriate data collection

Regardless of the field of study or preference for defining data (quantitative, qualitative), accurate data collection is essential to maintaining the integrity of research. Both the selection of appropriate data collection instruments (existing, modified, or newly developed) and clearly delineated instructions for their correct use reduce the likelihood of errors occurring.

Consequences from improperly collected data include

inability to answer research questions accurately

inability to repeat and validate the study

distorted findings resulting in wasted resources

misleading other researchers to pursue fruitless avenues of investigation

compromising decisions for public policy

causing harm to human participants and animal subjects

While the degree of impact from faulty data collection may vary by discipline and the nature of investigation, there is the potential to cause disproportionate harm when these research results are used to support public policy recommendations.

Issues related to maintaining integrity of data collection:

The primary rationale for preserving data integrity is to support the detection of errors in the data collection process, whether they are made intentionally (deliberate falsifications) or not (systematic or random errors).

Most, Craddick, Crawford, Redican, Rhodes, Rukenbrod, and Laws (2003) describe 'quality assurance' and 'quality control' as two approaches that can preserve data integrity and ensure the scientific validity of study results. Each approach is implemented at different points in the research timeline (Whitney, Lind, Wahl, 1998):

Quality assurance - activities that take place before data collection begins

Quality control - activities that take place during and after data collection

Quality Assurance

Since quality assurance precedes data collection, its main focus is 'prevention' (i.e., forestalling problems with data collection). Prevention is the most cost-effective activity to ensure the integrity of data collection. This proactive measure is best demonstrated by the standardization of protocol developed in a comprehensive and detailed procedures manual for data collection. Poorly written manuals increase the risk of failing to identify problems and errors early in the research endeavor. These failures may be demonstrated in a number of ways:

Uncertainty about the timing, methods, and identify of person(s) responsible for reviewing data

Partial listing of items to be collected

Vague description of data collection instruments to be used in lieu of rigorous step-by-step instructions on administering tests

Failure to identify specific content and strategies for training or retraining staff members responsible for data collection

Obscure instructions for using, making adjustments to, and calibrating data collection equipment (if appropriate)

No identified mechanism to document changes in procedures that may evolve over the course of the investigation .

An important component of quality assurance is developing a rigorous and detailed recruitment and training plan. Implicit in training is the need to effectively communicate the value of accurate data collection to trainees (Knatterud, Rockhold, George, Barton, Davis, Fairweather, Honohan, Mowery, O'Neill, 1998). The training aspect is particularly important to address the potential problem of staff who may unintentionally deviate from the original protocol. This phenomenon, known as 'drift', should be corrected with additional training, a provision that should be specified in the procedures manual.

Given the range of qualitative research strategies (non-participant/ participant observation, interview, archival, field study, ethnography, content analysis, oral history, biography, unobtrusive research) it is difficult to make generalized statements about how one should establish a research protocol in order to facilitate quality assurance. Certainly, researchers conducting non-participant/participant observation may have only the broadest research questions to guide the initial research efforts. Since the researcher is the main measurement device in a study, many times there are little or no other data collecting instruments. Indeed, instruments may need to be developed on the spot to accommodate unanticipated findings.

Quality Control

While quality control activities (detection/monitoring and action) occur during and after data collection, the details should be carefully documented in the procedures manual. A clearly defined communication structure is a necessary pre-condition for establishing monitoring systems. There should not be any uncertainty about the flow of information between principal investigators and staff members following the detection of errors in data collection. A poorly developed communication structure encourages lax monitoring and limits opportunities for detecting errors.

Detection or monitoring can take the form of direct staff observation during site visits, conference calls, or regular and frequent reviews of data reports to identify inconsistencies, extreme values or invalid codes. While site visits may not be appropriate for all disciplines, failure to regularly audit records, whether quantitative or qualitative, will make it difficult for investigators to verify that data collection is proceeding according to procedures established in the manual. In addition, if the structure of communication is not clearly delineated in the procedures manual, transmission of any change in procedures to staff members can be compromised

Quality control also identifies the required responses, or 'actions' necessary to correct faulty data collection practices and also minimize future occurrences. These actions are less likely to occur if data collection procedures are vaguely written and the necessary steps to minimize recurrence are not implemented through feedback and education (Knatterud, et al, 1998)

Examples of data collection problems that require prompt action include:

errors in individual data items

systematic errors

violation of protocol

problems with individual staff or site performance

fraud or scientific misconduct

In the social/behavioral sciences where primary data collection involves human subjects, researchers are taught to incorporate one or more secondary measures that can be used to verify the quality of information being collected from the human subject. For example, a researcher conducting a survey might be interested in gaining a better insight into the occurrence of risky behaviors among young adult as well as the social conditions that increase the likelihood and frequency of these risky behaviors.

To verify data quality, respondents might be queried about the same information but asked at different points of the survey and in a number of different ways. Measures of 'Social Desirability' might also be used to get a measure of the honesty of responses. There are two points that need to be raised here, 1) cross-checks within the data collection process and 2)

data quality being as much an observation-level issue as it is a complete data set issue. Thus, data quality should be addressed for each individual measurement, for each individual observation, and for the entire data set.

Each field of study has its preferred set of data collection instruments. The hallmark of laboratory sciences is the meticulous documentation of the lab notebook while social sciences such as sociology and cultural anthropology may prefer the use of detailed field notes. Regardless of the discipline, comprehensive documentation of the collection process before, during and after the activity is essential to preserving data integrity.