1a

Ogwumike, F.O; D. D. Ajayi and U. Isiugo-Abanihe (2005): Data Collection. In: Gbadegesin, A.S; R. Olopoenia and A. Jerome (eds), Statistics for the Social Sciences, Chapter 3, pp. 19-24. Ibadan University Press. –Nigeria (33.3%)

STATISTICS FOR THE SOCIAL SCIENCES

Edited by
Adeniyi Gbadegesin
Razaq Olopoenia
Afeikhena Jerome

Ibadan University Press
2005

CONTENTS

3

DATA COLLECTION

Fidelis O. Ogwumike, Dickson 'Dare Ajayi and Uche Isiugo-Abanihe

The Standard English Dictionary defines data as facts and figures from which conclusions may be drawn. In general, however, statistical data refer to numerical descriptions of quantitative aspects of situations. Data are useful in providing an informed understanding of situations, with an overriding view to better decision making. While data, as a matter of definition, are quantitative, they may in fact represent either quantitative or qualitative facts. Quantitative facts, which are already expressed as numbers, lead directly to data in the form of these numbers. For example, the quantity of maize in tons purchased over time and the price of maize in naira. Note the units of measurement, tons and naira. Qualitative facts for which no measure exists can also be expressed in the form of data. For example, you are either a male or a female, married or unmarried.

There are three main types of data: cross-sectional, time series and panel data. Cross sectional data refer to data collected from selected units of a population at one point in time. Time series data show observed values of a variable over a period of time; and panel data refer to data collected from the same group of selected units over a period of time.

Data collection is an activity or group of activities aimed at getting information, facts and figures to satisfy given decision objectives. Data collection is perhaps the most critical of all the stages of data analysis. This is so because once the data collected

are of questionable quality, we cannot expect more from their output. There are three (3) basic methods of collecting primary data, namely, (A) Observation, (B) Experimentation, (C) Surveys.

In the observational method, the data collector records facts and figures regarding the subject of concern by observing the relevant actors. The experimental method consists of introducing stimuli into a controlled environment and systematically varying the stimuli. Extraneous factors are eliminated or controlled so that observed changes can be related to the variations in the stimuli. Compared with either direct observation or experimentation, the survey method yields a broader range of information and is effective for a greater number of research problems.

As an activity, the outcome of data collection mainly depends on the nature of the decision to be taken, the units providing the information, the method or approach adopted, and the degree of sophistication.

Statisticians aim at collecting adequate data required for taking action under uncertainty. The process of data collection must, therefore, satisfy certain criteria. It requires a proper design or a set of integrated strategies. The design of a data collection tool is influenced by the nature of the data to be collected. In particular, it depends on whether the information sought for concerns (A) animate or inanimate units or (B) human beings and their activities and environments.

Information relating to (A) is usually collected via controlled experiments because the units can easily be put under control. The data collection technique for doing this is called experimental design and analysis. Regarding (B) above, data are usually collected by means of surveys or through administrative records. This can be done through sampling of the units involved or by complete coverage of all the units if the whole population can be determined, reached and surveyed. However, the use of complete coverage lacks sophistication and may not produce better or more accurate results than sampling.

There is a subtle difference between a sample survey and a sample census. Surveys are mainly adopted where detailed information is being sought, while a census is used when the information required is relatively not detailed. Hence, a data

collection technique based on sampling may be referred to as a sample survey or sample census.

It is imperative to have a design for an impending data collection. This is a logical starting point since it forces data collectors to plan the process of data collection. A good data collection design must cover and articulate the following:

(a) the objective(s) of the exercise;
(b) the population/sample involved and its frame;
(c) the set of data required and its form of presentation;
(d) the methods of measurement;
(e) the sample design and selection, including the required precision;
(f) the field work; and
(g) the summary and data analysis.

Surveys are essentially conducted at two levels, namely,

(a) the primary data level, and
(b) the secondary data level.

At the primary data level, we collect first hand numerical/ qualitative information from respondents, for example, household surveys, census surveys, opinion poll surveys etc. The quality of information collected here depends highly on the efficiency of the survey design, the investigator and the extent of cooperation of the respondents.

The secondary data level involves the collection of numerical information which has already been generated but may or may not necessarily be in the form required by the investigator. The quality of the secondary data does not depend on the efficiency of the survey design or the investigator. Most common surveys of this type are establishment surveys.

One of the most useful forms of survey is the two-stage cluster sampling. This is obtained by first selecting a simple random sample of clusters and then selecting a simple random sample of elements from each sampled cluster. A random sample is one drawn in such a way that every possible element of the population has an equal chance (probability) of being selected.

Methods of Data Collection

In collecting data, there are various means a researcher can adopt. Whichever method is adopted is informed by the type of data or information required and the accessibility of each method to the researcher. Cost can also be a determinant of the method selected. The methods broadly include

(a) personal interviews;
(b) mails and e-mails;
(c) telephone interviews/conversations;
(d) experiments/observation; and
(e) electronic media.

These will be discussed in turn.

Personal Interviews

For personal interviews, the researcher may have to recruit field workers, enumerators and supervisors who are expected to go out and meet the respondents in face-to-face interviews. These enumerators visit respondents in their homes or offices and conduct the interview based on predetermined questions listed on a questionnaire. They are also to complete the questionnaire according to the answers provided by the respondents. In certain situations an interpreter may be required if the respondents are not literate.

Mail Method

In the mail method, questionnaires are sent to selected respondents by mail or e-mail. Although it is a cheap method, you cannot ensure that the respondents give correct and up to date answers. Besides, respondents may delay returning the answers.

Telephone Interviews

Telephone interviews are essentially applicable where telephone networks are not only available but reliable. The method is very fast but can be a little expensive in the generation of data.

Observation Media

Here, information is recorded directly into a computer. For example, a measurement of the distribution of television audiences can be carried out with devices attached to a sample of television

sets. These devices automatically record in a computer the channels being watched.

Planning Data Collection

The pivotal role of data collection in any research endeavour cannot be overemphasized. The result of any research is usually as good as the data which facilitated it. Hence, careful and in-depth planning should precede data collection. In this connection, planning will demand attention to the following:

(1) clearly defined objectives;
(2) a defined scope and coverage;
(3) the identification of possible respondents;
(4) the selection of suitable frames and an appropriate design;
(5) the design and printing of questionnaires, if they are needed;
(6) the assembling of supporting staff;
(7) the training of investigators, enumerators, collectors etc.;
(8) the estimation of costs; and
(9) a consideration of administrative and incidental arrangements.

A significant aspect of planning is the identification of objectives. Once objectives are clearly identified, they can be a pointer to the most appropriate method of data collection to adopt. Moreover, the design of the questionnaire can be informed by the objectives being pursued. Respondents should be identified and their number enumerated. This serves two main purposes. The first is that it suggests the number of copies of the questionnaire the researcher will need, thereby reducing immense waste. Also, it suggests the quality of the respondents. Hence, if the need for an interpreter is identified then one can be provided. However, the researcher may use respondents rather than spend more money on interpreters in all aspects of the work and he must have a working knowledge of the whole project. This can be achieved through adequate staff training. This kind of training has the potency of improving the psyche of the staff as they see the project as theirs too. Essentially, every aspect of planning has the ultimate goal of achieving accurate and hitch-free data collection.

Problems of Data Collection

In Nigeria, the collection of data is hampered by a host of problems which inevitably affect the quality and quantity of data collected. Although peculiar problems are associated with different categories of data, the general problems of data collection include:

(1) the uncooperative/unreliable attitudes of respondents, e.g., companies;
(2) the lack of coordination between data collecting agencies;
(3) insufficient coverage;
(4) a lack of adequate funding;
(5) the poor communication with rural respondents arising from a poor transport system;
(6) the inexperience of some data collectors;
(7) the poor record-keeping in ministries and government agencies;
(8) an inadequate legal framework for data rendition; and
(9) the high rate of illiteracy of respondents.

Sources of Data in Nigeria

The Federal Office of Statistics (FOS) has the mandate for data collection and publication on all sectors in the country. It has an extensive data gathering network in the country with its presence in every state capital and in many towns and villages in the federation. The FOS collects data at both the primary and secondary levels through sample surveys and censuses. Its surveys and censuses cover virtually every sector of the economy, particularly agriculture, industry, education, health, un-employment, prices, trade and transport, communication, etc.

Other agencies, apart from the FOS, which collect and dissemiate data include the Central Bank of Nigeria, Federal/State ministries, National Population Commission (NPC), National Universities Commission (NUC), National Manpower Board, Nigerian Ports Authority (NPA), National Directorate of Employment (NDE), Joint Admissions and Matriculations Board (JAMB) and Nigerian National Petroleum Corporation (NNPC). These agencies collect and produce statistics in their areas of interest and operation.