EXPLORING DATA ANALYSIS IN PHYSICAL SCIENCE RESEARCH

BY

DR. MCCHESTER ODOH

DEPARTMENT OF COMPUTER SCIENCE

MICHAEL OKPARA UNIVERSITY OF AGRICULTURE, UMUDIKE, UMUAHIA, ABIA STATE OF NIGERIA

ABSTRACT

Data is the representation of facts, observation and occurrences. It is the input of the statistical system, it is also the input of the data processing system, it is also the input of the integrated computer system. It could be in the form of words, figures or even computer codes. Data analysis involves the splitting of the units into their paths and also the aggregating the facts in similar variables. It is at the analysis stage of a research project that meaning is given to the data that is collected. By research is meant the act of trying to find out solutions to problems so that proper inferences can be made. In this paper, a theoretical review is to be done on data analysis and research and this objective is fully achieved.

INTRODUCTION

According to Osuala [1], scientific problems can be solved only in the basis of data and major responsibility of the investigator is to set up a research design capable of providing the data necessary for the solution of his problem. The more clearly and thoroughly a problem and its many ramifications are identified, the more adequately the study can be planned and carried to a successful completion, thus the task of the investigator is to synchronize the statement of the problem with the design to be used in its solution. Every aspect of the study down to the last details of execution must be planned before the study is undertaken. It is not wise to select a topic, no matter from adequate, if circumstances render the collection data required for its solution impossible [2]. However, to [3], the issue of data is at very center in base of research, the nature of data for study depends entirely on the objective of the research and the type of research undertaken. Equally, relevant are the research hypotheses and cost.

MEANING OF DATA ANALYSIS

Data is defined as quantitative information; analysis is the breaking of data into constituent parts in order to make them meaningful. The frame for analysis is

usually set prior or so the collection of data. The set prior to the collection of the analysis function comprises the number of stages including the ordering of the data into meaningful categories, determining whether significant differences exist between the categories, explaining why the differences exist and finally dreaming conclusions or making recommendation, it also involves searching for trends and patterns of associations and relationships among these data or groups of them [10].

MEANING OF RESEARCH

Business practitioners are often confronted with numerous problems which demand a thorough investigation on the matter. To solve these problems, the most appropriate scientific approach is to embark on highly objective research subject. The word research is derived from the French word recherché, which means to search closely, while chercher means to search, its literary meaning is to investigate thoroughly [5].

Basically, research is an organized and systematic way of finding solutions/answers to questions. To appreciate this definition, we shall provide meanings to some of the key concepts used in the definition above:

Systematic – This means there is a definite set of procedures and steps which must be followed. There are certain things which are always done in order to get the most accurate results.

Organized – This implies that there is a structure or method in going about doing research. It is a planned procedure, not a spontaneous one. It is focused and limited to a specific subject and scope.

Finding solutions/answers – This is the end result of all research endeavor. Whether it is the answer to a hypothesis or even a simple question, research is successful when we find answers to a question. The answer may even be no, but it is still an answer.

Questions- They are fundamental to research endeavor. If there is no question, there cannot be an answer. Research is focused on relevant, useful, and important questions. Without questions, research has no focus, drive or purpose [6].

Research is equally defined as a systematic investigation to establish facts. Furthermore, research is an inquiry, that is, a search for knowledge. Research is also viewed as a human activity based on intellectual application in the investigation of mater [7].

Broadly, research is the systematic process of collecting and analyzing information to increase our understanding of the phenomenon under study. Eight characteristics of research can be identified as follows:

i. Research originates with a question or problem
ii. Research requires a clear articulation of a goal

iii. Research follows a specific plan of procedure
iv. Research usually divides the principal problem into more manageable sub problems
v. Research is guided by the specific research problem question, or hypothesis
vi. Research accepts certain critical assumption
vii. Research requires the collection and interpretation of data in attempting to resolve the problem that initiated the research
viii. Research is, by its nature, cyclical, or more exactly, helical [5]. A researcher is someone who is professionally engaged in research. Furthermore, a researcher is a scholar who can, or will in time, through learning and experience, demonstrate the following:

1. Specialized knowledge or expertise, conceptual and intellectual capacities such as the ability to identify and frame key problems, to think critically and analytical, and to generate and communicate interesting and original insights.
2. Academic skills such as the ability to produce scholarly high-quality written work and research papers – clearly composed so that the argument, and the evidence that supports it, can be grasped by the intended audience (whether specialist or more general, a conference delegate, or a reader with time to reflect).
3. Research skills such as the ability to use sources effectively, to gather and organize information, to analyze text, data and theory.
4. Personal attributes such as the ambition and ability to work to high standards, to take initiative and responsibility, to be well organized in one's procedures and balanced in one's judgments, to collaborate well with others where appropriate, and to take on board and incorporate constructive criticism.
5. Social skills such as the ability to liaise with students, colleagues and academics from the institutions in an effective and appropriate way, to be able to adjust to different circumstances required by academia and to integrate into the larger community of scholars [6].

DATA COLLECTION PROCESS

To [1], data collection is the processes through which the researcher searches for information to help him solve the problem on hand, every research effort therefore and it centers on search for information either primary or secondary sources. To types of data can be collected: primary data and secondary

data. Primary data is the data in which the researcher is the original collector. Secondary data is the data in which the researcher is not the original collector.

[6], has observed that there are two different research methods for collecting data: the research method for primary data and the research method for secondary data. The research methods for collecting primary data include the survey, experiment, a case study and model modification. The research method for collecting secondary data include: the use of secondary data, content analysis, library search and also model modification.

CRITERIA FOR EFFECTIVE DATA COLLECTION PROCESSES

According to [3] certain criteria could be adopted in assessing the merits of the data collection method.

These include:

1. The cost of collecting data.
2. The period of time spent in collecting data.
3. The accuracy of the observations.
4. The practicability or otherwise of the method.
5. The appropriateness of the sampling techniques.
6. The representativeness of the sample.
7. The ability to assess inherent errors and biases.
8. The ability to minimize (and if possible eliminate) these errors.
9. The suitability and relevance of the data to the desire of the researcher.

Note, on secondary data, these must be used with data collected by other people i.e., they are other peoples' data and researcher using them should to a large content be interested in whom they (the data) were collected for. Often, secondary data may not provide the same kind of information the study requires and may have to be re-organized or re-arranged by the researcher in an attempt to meet his needs.

SOURCE OF DATA

According to [2] data sources are fall-outs from the nature of desired data. In the management and behavioral sciences, the two main methods are the primary and secondary.

PRIMARY DATA COLLECTION

[2], by primary data is meant data that are collected from original sources for a particular purpose, such data does not exist in computed form and so the researcher must set out to collect the data himself or, herself. However, primary source include personal interview and the questionnaire as instruments. The questionnaire is the structured form that contains questions. The questions could be in the form in which the respondents are given alternative answers to pick. They could also be in the form of Likert scale questions. The questions are of the form of

strongly agree, agree, undecided, disagree and strongly disagree. The answers to the questions can be analyzed using percentages, relative frequencies, Z-test, Z-test of population proportions, coefficient of variation, etc.

THE INTERVIEW METHOD

According to [3], personal interview is a process by the use of recording gadgets using a cassette or any other scientific equipment and played back if required. Questions could be prepared in advance but because of the persons to person contact required, question easily flow and the position of the respondents on the issue in question is easily understood. The disadvantage of this method is that the interviewing situation may change from one occasion to another especially if more than one field data collector is used to do the fieldwork. This disadvantage can be minimized by the researcher doing most of the fieldwork himself/ herself.

THE QUESTIONNAIRE METHOD

According to [3], the second method of collecting primary data is through the use of the questionnaire. The questionnaire is a body of questions targeted at the respondents with a view to satisfying the purpose of the study and providing information for analysis.

1. They are normally designed in order to gather information from people generally by beyond the physical reach of the researcher.
2. The question must be designed in very simple and clear terms.
3. Questions must be directed at providing data to answer the research hypotheses.
4. They can be administered by mail through the post or the computer (i.e., by e-mail) or by hand.
5. Scaling must be taken into consideration when designing the question.

DESIGNING OR CONSTRUCTION OF THE QUESTIONNAIRE

According to [3], a typical questionnaire should contain three or more sections. In short a minimum of three types of information are provided. Thus, a break-down of thee major things is important. Namely:

1. Research questions (or problem statement).
2. Objectives of the study.
3. Research hypotheses.

However, section one of the questionnaire normally seeks information on the respondent, some information bordering on his broad data such information and classification are necessary for analysis. Section two deals with research questions as well as the objectives.

It also includes information bordering for achieving the administration of questionnaire. Section three dwells on the research hypotheses. In short section two or three must contain questions on critical variables.

Though these sections may not be spelt out as shown above, they nevertheless form the background for questionnaire. In the event of asking multiple questions appropriate (viable) alternatives should the provided to adequately guide the respondent.

SECONDARY DATA COLLECTION

[8], has observed that secondary data in contrast to primary data are those that are derived from what other researchers have left behind in the form of recorded material either in Journals, Magazine, Government Document or Books include secondary data are usually cheap and easy to find, primary data are more costly time consuming to collect. Generally, secondary data can be used in problem recognition, problem clarification of feasible alternatives and actual solution of the researcher problem.

DATA ANALYSIS TECHNIQUES

[3], has observed that data analysis is very critical in research. Apart from exposing the skills of the researcher by showing how rigorous the exercise it shows the extent to include an array of data. It has to be well done in order to arrive at meaningful conclusions and relevant recommendations. The whole process of analysis in business and economic research as well as research in the social sciences and physical sciences, involves the preparation of the data to be presented in the form that the data can later be analyzed and an assessment explanation of the data or various groups of data including their trends and patterns of relationship/association. Arising from these will be findings, and subsequently recommendations aimed generally at solving the problems at hand or surmounting the stated challenges and satisfying the objectives of the study.

DATA PRESENTATION

[9], has observed that the processing of generated data (from either primary or secondary sources) in to a fine relevant state for presentation (tabulation) or other stages analysis is called data presentation. Data can also be presented using frequency distributions such as bar-charts and histograms. Pie charts are also used in presenting data.

DATA PRESENTATION AND ANALYSIS

[7], has observed that data is presented to make it amenable for further analysis. It is meant to generate information for analysis and obtaining results in a large volume of statistical information which is mostly in its raw stage. In order to use data for the objective of a research, they have to be reduced to manageable dimensions. Thus, data analysis follows data tabulation. There are two types of data analysis according to Asika: descriptive and casual analysis.

DESCRIPITIVE ANALYSIS

According to [7], the study handles the distribution of the variables of study (in relation to subject) such as the profiles of respondent organizations, groups or any other subject. Descriptive analysis may be either qualitative or quantitative.

Qualitative Descriptive

Qualitative descriptive analysis is used to verbally summarize the information generated in the research. For example the researcher may choose to summarize the data on participation by organizations. He can simply do that without using any table, and state that out of organization that participated in the research, 13.4% were in manufacturing, 13.4% in commerce and distribution, 30.4% in services organization, 26.8% in government ministries and parastatals,

13.4% as unclassified organization and 2.7% as using some quantitative information. The description can also be done without quantitative information by merely stating that the majority of the participating organizations were services and government organization, others were manufacturing and commercial organizations [7].

Quantitative Description

According to [9] quantitative descriptive analysis is used to summarize a mass of information generated in the study, so that appropriate analytical methods could be used to further discover relationships among the variables. Quantitative descriptive analysis includes such tools as frequency distribution, measures of central tendency and measures of dispersion.

ANALYTICAL TOOLS

MEASURE OF CENTRAL TENDENCY

These are also known by such titles as measures of location or central values. A distribution is the set of data that may have been generated from one or more related sources. It is often necessary to describe the distribution or describe their profile by condensing the data into a single figure which left us to know what

is typical about the distribution. To do this one needs to calculate different measures of central tendency in the distribution [7].

The Mean

[1], has observed that the mean represents an average or the arithmetic mean i.e., central point in a distribution. This for ungrouped data is shown below as

$$\overline{X} = \frac{\sum x}{N}$$

Where $\overline{X}$ is the mean

X: a value

$\sum$: sum,

N: number of observation or values.

An example of an arithmetic mean is where we want to find average monthly wage paid to five (5) staff in a small business outfit given by the following data

S/N	***Name of staff***	***Monthly wages Bill (N)***
1.	Peter Oman	25000
2.	Victor Musa	26000
3.	Ama Joseph	25500
4.	Iyanu Bode	27000
5.	Eyeno Okon	27500

Here monthly wages represent our *X* while *N* shows the number of observations.

Therefore; $\sum x$ = 131,000

N 5

$$\frac{\sum x}{N} = \frac{N131{,}000}{5} = 26{,}200$$

Thus *X*= N26 200.00 i.e. the average monthly wage.

The Median

[7], has observed that the median is the statistic that describes the midpoint or centre of a group of data arranged in an order of magnitude. It is the value that lies midway between the lowest and the highest values.

Example 10, 15, 17, 18, 20, 23, 26, 30, 33, 35, 36

Median = $\frac{11+ 1\text{th}}{2}$ = the 6th item

=23

The Mode

[7], has observed that the mode represents the number in a set of scores with the highest frequency while the mean and median measure the position within a distribution, the mode measures the structure (nature) of the distribution.

MEASURE OF DESPERSION

[1], has observed that the range measures the spread of all values from the most extreme values in the distribution. It is defined as the difference between the highest and lowest values in the array or distribution. The range has no specific formula but is calculated by subtracting the smallest value from the highest value in the distribution.

Variance and Standard Deviation

According to [3], variance is a square of standard deviation it is seen as the average of the squared deviation of the values from the shapes depending on whether a sample or population is involved.

The sample variance for grouped data is given by the formula: $S^2 \frac{= \sum f(x-\bar{x})^2}{n-1}$

Where
f = frequency
x = variate
$\bar{x}$ = sample mean
n = total number of observations

SUMMARY AND CONCLUSION

Scientific problems can be solved only on the basis of data and the major responsibility of the investigator is to properly collect the data, analyze it and summarize and conclude and make recommendations and also state the areas for future research [4]. For primary data collection, the questionnaire and oral interview schedule are two very powerful instruments. The questionnaire has the disadvantage that the structured nature may compel some respondents to give answers that they do not fully endorse. The oral interview schedule has the disadvantage that the interviewing situation may change from one occasion to another. The demerits of the questionnaire can be minimized by also using an oral interview schedule that contains open-ended questions. However, the oral interview schedule has unstructured questions whose answers are very difficult to analyze. This limitation is minimized by also using frequencies which is got as a ratio of the number of respondents that gave a particular answer divided by the total number of respondents that returned the instrument.

The survey research design has the limitation that some respondents are unwilling to give answers to probes. This limitation is minimized by giving the respondents some motivation. Experimentation has the limitation that a laboratory must have experimental facilities and another set of facilities for comparing. In the management and social sciences, it is difficult to experiment because human beings are difficult to study, so the human factor makes it difficult to use

experimentation. Once a human being knows that he or she is being studied, the excitement affects the result.

The case study has the limitation that only some few variables are studied but in-depth, so it is very difficult to generalize the findings from a particularly case study to other organizations. This limitation is minimized by doing a cross sectional study in which a lot of organizations are studied. The model has the limitation that it is only an abstraction of reality and not reality itself. So not all the factors in the real system are in the model but only the significant ones, otherwise, the model will be very difficult to implement and its purpose to study, analyze and predict the real system will be defeated. This limitation is minimized by interviewing the senior and junior staff of the organizations studied and analyzing the data, collating the results, summarizing them, concluding, making recommendations and stating the areas for future research.

The use of secondary data, content analysis and even, model modification using the secondary data all have the limitation that the researcher is not the original collector. So it is possible to inherit mistakes, miscalculations and biases. This is why in the management sciences and even in the social sciences, it is advisable to use primary data and not secondary data. Also it is difficult to establish the reliability and validity when secondary data instruments are used. But if primary data is used, reliability can be established by test-retest, split-half and equivalent half methods. Validity can be established by face or even content method or other methods.

REFERENCES

[1] Osuala E.C. (2005) *Introduction to Research Methodology.* Nsukka; Africana FEP Publishers Limited.

[2] Kish, J.J. (1995). *Business Research Method* Glence; Free Press.

[3] Onwumere J.V.J. (2005) *Business Research* Lagos: Longman Nigeria Plc.

[4] Davis, (1994). *Contribution of Research to Business* Glenco: Free Press.

[5] Enikanselu, S.A. (2006). *Management and Business Research Seminar*. Yaba, Lagos: Enykon Consults.

6] Unyimadu, S.O. (2006). *Research Methods in the Social Sciences, Management Sciences, Education and Engineering*. Benin City: Harmony Publishers.

[7] Asika, A. (2006). *Research Methodology in Behavioral* Science.Lagos: Longman Nigeria Plc.

[8] Bartlet, C.O. (2001). *Research Concept in Human Behaviour* New York; Appleton Century.

[9] Sidney S.C.(1999) *Understanding Educational Research.* New York: McGraw Hill.