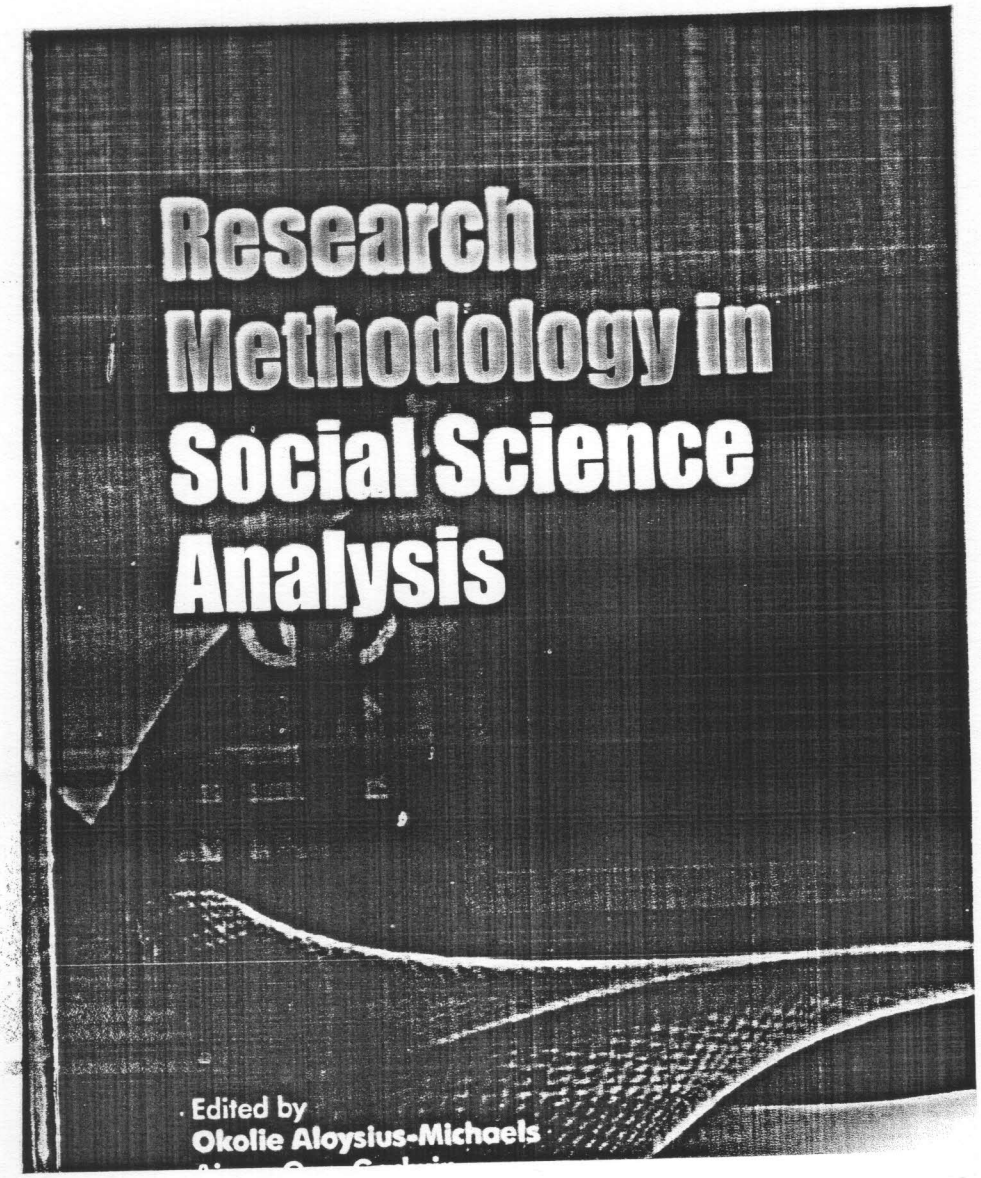

Research
Methodology in
Social Science
Analysis
Edited by
Okolie Aloysius-Michaels

# Research Methodology in Social Science Analysis

Edited by
**Okolie Aoysius-Michaels**
**Ajene Oga Godwin**

Published by the Nigerian Political Science Association

## Chapter Eight

# QUANTITATIVE TECHNIQUES IN POLITICAL SCIENCE RESEARCH

**Aremu Fatai Ayinde** PhD

### Introduction

For decades political scientists have debated the merits of statistical applications and in-depth qualitative techniques. Much as proponents of quantitative approach contend that statistical analysis are easily replicable and are therefore the surest route to "systematic" and "scientific" study of politics, supporters of qualitative approach fervently reject that position (King, Keohane and Verba, 2010:4). As a result of the intense contention between both seemingly irreconcilable positions, an imaginary wall was erected between the so-called 'quantitativists' and 'qualitativists' with each holding steadfastly to its forte. Until recently when mixed methods and multi-methods traditions began to gain palpable traction, both approaches hardly communicate across the aisle.

Meanwhile, quantitative technique is yet to gain significant footing in mainstream political science research in some African institutions. The generational 'phobia' or 'indifference' to numbers, equations and models is particularly observable in Nigerian political science community. The belief that political variables such as behaviour, attitude, motivation, interests and others, are largely unquantifiable remain pervasive. Therefore, this chapter begins with the questions: is quantitative technique important, necessary and desirable? Can (or should) political scientists quantify political variables? In the context of global and national reality, answers are 'yes' and 'yes.' Not only has global political science gone largely quantitative (and there is much catching up to do in this regard), other social science disciplines have incorporated quantitative methodological tools to the extent that disciplines like economics is hardly distinguishable from Psychology or Geography. From the angle of interdisciplinary and multidisciplinary intellectual conversation, political science community in this clime needs to halt its decoupling from international peers and immediate family of disciplines by imbibing and upgrading its application of advanced quantitative techniques.



In any case, 'politics,' as it is generally known, is a game of numbers. Elections, votes, voter turnout, bills passed, two-third majority, first-past-the-post, party membership, violence casualties, and so on are examples of political concepts that are inherently numerical. Indeed, suffice to say that without numbers, without counting and quantification, politics is empty, boring and meaningless. Making sense of politics and offering widely graspable interpretations requires sufficient grounding in numerical calisthenics and application of quantitative tools. Therefore, if the discipline of political science is to regain its prime of place in domestic policy and global arena, it has to factor in the critical role of quantitative techniques in its research designs. It goes without saying that numbers, figures, graphs and statistical results are highly compelling and makes our arguments and the storyline that undergird them more convincing and easily comprehensible irrespective of the readers' intellectual and philosophical background.

### Types of Quantitative Design

There are three main types of quantitative designs. These are Exploratory, Descriptive and Causal Designs.

***Exploratory Design***: In some instances, very limited information is known about a phenomenon that could be used to frame a precise problem and formulate useful hypothesis. Exploratory research is often conducted to gain preliminary insights into the social reality that is about to be investigated. Also known as 'formulative' research, it could take the form of a pilot survey which reveals the broad pattern on the subject of inquiry prior to designing and executing the research. Because it is hardly useful as the basis for decision making, exploratory research is often criticized for its lack of generalizability and utility (Hartwig & Dearing, 1979).

***Descriptive***

***Design***: Descriptive design generally answers the "what" question. As the name implies, descriptive statistical analysis lays out observable pattern that are discernible from social phenomena. Graphs, tables, charts and simple statistical tools (such as mean, median, mode, etc) could be employed to aid the description of observed information. This could take the form of time series (trend), cross-sectional (longitudinal) or panel data. Time series data applies to a situation whereby a unit of analysis is observed over a period of time. Spatial unit is constant while the temporal component varies.

**Table 1: Example of Time Series Data (hypothetical)**

| State | 1999 | 2003 | 2007 | 2011 | 2015 |
|---|---|---|---|---|---|
| Kwara | 560,000 | 620,000 | 700,000 | 800,000 | 600,000 |

Table 1 contains hypothetical data on voter turnout in Kwara State between 1999 and 2015. Kwara State is the unit of analysis while the data covers 1999-2015. The data varies over time, hence, time series data.

Cross-sectional data on the other hand applies when the unit of analysis is more than one while the time is constant. In other words, multiple units are observed at a single point in time as shown in the example contained in Table 2.

**Table 2: Example of Cross-sectional/longitudinal data (hypothetical)**

| State | 2015 |
|---|---|
| Kwara | 800,000 |
| Osun | 300,000 |
| Lagos | 1,500,000 |
| Oyo | 1,200,000 |
| Kogi | 450,000 |

Panel data describes the situation whereby there is variation in the spatial and temporal components of the information. The unit as well as the time of observation varies. Table 3 contains example of panel data.

**Table 2: Example of Panel data (hypothetical)**

| State | 1999 | 2003 | 2007 | 2011 |
|---|---|---|---|---|
| Kwara | 560,000 | 620,000 | 700,000 | 800,000 |
| Osun | 250,000 | 150,000 | 220,000 | 300,000 |
| Lagos | 800,000 | 750,000 | 900,000 | 1,500,000 |
| Oyo | 650,000 | 720,000 | 860,000 | 1,200,000 |
| Kogi | 320,000 | 360,000 | 400,000 | 450,000 |

***Causal Design*:** Causal design seeks to establish causal pathway between an outcome (dependent variable) and a cause (independent variable). The overarching concern of causality is to provide evidence that indeed outcome (Y) is a consequence of a cause (X) through the mechanism (Z).



For example, if we consider the hypothesis that money (X) may (not) cause happiness (Y) to increase or decrease depending on what use a person deploys his/her money (mechanism). An individual may use his/her money for productive venture like investment or charity which gives him/her happiness. Another person may spend his/her money on alcohol or drug which could propel violent behaviour that may earn jail terms. Causal mechanism is therefore the link between a cause and an outcome. It is however important to note that a cause may produce different outcomes (equifinality) just as multiple causes may produce an outcome (conjunctures) (Gerring, 2012; Marini & Singer).

In order to establish causality between a set of variables, counterfactuals (what if...?) are often used to test whether X actually causes Y. This implies that what if X was not there, would Y still exist? In a sense, causal design could also set the framework for experimental design where counterfactuals are generally embedded in the design, *ab initio*. For the purpose of policy inclined decision making, however, causal designs are often more compelling because it allows the decision maker to tackle the causes of a problem, unlike studies that end up with description.

**Quantitative Data Collection**

Data is an important element of executing quantitative research design. Without data, it is impossible to achieve any research objective regardless of the methodological tradition. Basically, data refers to pieces of information that allows the researcher to answer the questions raised in the research. Such information could be in the form of symbols, speeches, recordings, documents, numbers, etc. The nature of data to be generated depends on the research objectives and the questions being explored. The popular cliché among methodologists is that inappropriate data is as bad as no data! In other words, data is not to be sought for its own sake; it must be appropriate and suitable to the research question. A researcher is often faced with two sources of data: primary and secondary.

*Primary data* is the type of data that originates directly from the researcher. This implies that the researcher is the original producer of the information which could be in quantitative form (if it is numerical) or in qualitative form (if it is non-numerical). Suppose a researcher is interested in a research titled: "Charisma and Political Influence." Depending on the case study, the nature of this research would require the researcher to delve into archives, adopt

ethnographic observations, conduct in-depth interviews or survey. It could also entail a combination of various techniques. To the extent that the researcher is the 'originator' of the data, this is primary data.

*Secondary data*, on the other hand, is the type of data that the researcher captures from existing pool of repositories. In other words, the researcher is **NOT** the 'original producer' of the data. Is secondary data important? Secondary data could be very useful whenever it is practically impossible or unnecessary to generate primary data. There are occasions where existing data repositories have very useful data on the subject of inquiry. Except where existing data is not appropriate or suitable for the research question being explored by the researcher, utilizing secondary data saves time and conserves resources. Secondary data could be in quantitative or qualitative forms. Secondary data sources that are commonly used in political science research include databases like POLITY IV, Afrobarometer, Failed State Index, State Fragility Index, World Development Indicators, UNDP's Human Development Report, etc. Studies on aggregate issues such as democracy, violence, conflict, freedom, etc could benefit from secondary data which could be accessed from existing data warehouses as well as media sources (traditional and 'new') and literature.

While there may be no pressing need to reopen the puerile debate on the superiority of one over the other, it is nonetheless useful to stress that generating primary data confers unquestionable originality on the research, especially if in cases of degree-seeking research. This does not necessarily imply that originality in research is not possible or attainable with secondary data. There are several original studies that utilize secondary data simply because they are not amenable to the production of primary data. An example of such studies is one entitled: "The Geopolitics of Civil Wars: External Military Assistance, Competitive Interventions and Duration of Intrastate Conflicts." The researcher relied on secondary data on civil war duration (in days) drawn from the UCDP Conflict Termination Dataset (CTD).[1] It would have been practically impossible for the researcher to embark on generating primary data on civil wars which occurred in different countries and different times.

[1] Noel Anderson's presentation to the 2013 Institute for Qualitative and Multimethods Research (IQMR) at Syracuse univerty, New York, USA 17th -28th June.



Meanwhile, one of the popular means of generating primary data is survey. *Survey* is the systematic process of observing, counting and recording of items of interest in a research. It is a very useful tool for gathering data on attitude, opinion, perception, preference, motivation, knowledge, demographics, etc. Survey data could be generated with the use of questionnaire. A questionnaire is a document containing questions and other types of items that are designed to solicit information. Questionnaires are used primarily in survey but are also used in experiments, field research and other modes of observation. It could contain open-ended or closed-ended questions depending on whether the researcher is interested in gathering qualitative or quantitative data. Open-ended questions allow the respondents to provide in-depth comments freely while close-ended questions confine the respondents to a range of options provided by the researcher.

For the purpose of data quality and reliability, the process of constructing questionnaire is expected to follow some guidelines. It is important to choose appropriate question forms (questions or statements; open or closed-ended) and make each items very clear. It is generally recommended that double-barreled questions are avoided. Of course, the questions should be relevant to the subject of interest to the researcher and respondents must be competent and willing to answer. Shorter questions are generally better than longer questions and researcher's interest is better served when negative items are avoided. In order to ensure high quality data, it is often recommended that survey instruments are pretested before application on the field. This will allow the shortcomings of the questionnaire, including logical and consistent ordering of question items, to be discovered and addressed before proceeding to actual fieldwork.

Administration of questionnaire on the field presupposes that the researcher figures out well in advance the population of the study, sample size and sampling technique. *Population* refers to the total number of observable units within the area of study. Examples of population include people, voters, legislators, women groups, homes, cities, tribes, companies, etc. A study of voting behaviour among persons with special needs in Oyo State needs the total population of persons with special needs in Oyo State who are within voting age. It is from this population that a *sample,* a smaller unit of the population upon which the questionnaire will be administered can

be drawn. There are two broad category of sampling techniques. These are probability and non-probability sampling.

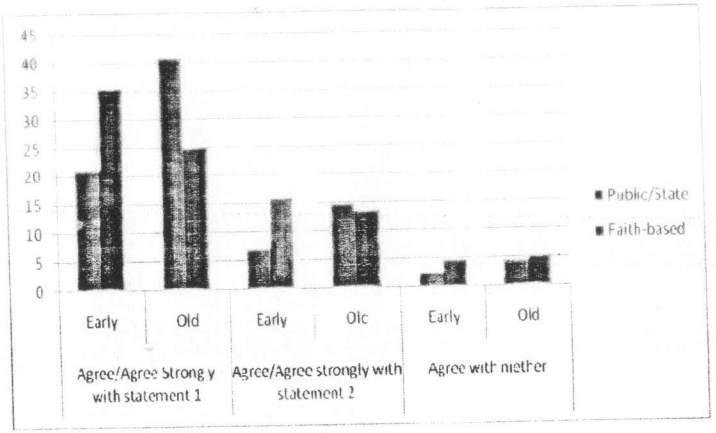
*Probability sampling* procedure gives every unit of the population equal chance of being selected while the element of randomization is absent in non-probability sampling. Probability sampling technique include simple random sampling, systematic sampling, stratified random sampling, cluster sampling and probability-proportionate-to-population sampling (PPPS) which is the technique used in Afrobarometer surveys. *Non-probability sampling* techniques include accidental sampling, purposive sampling, snowball sampling and panel sampling. While probability sampling techniques are more suitable for Large-N studies, non-probability are usually applied in Small-N studies where the researcher has prior knowledge of the category of respondent with useful information that are relevant to the study. A study of gang behaviour or drug abuse among prostitutes may find purposive and snowball sampling more suitable than simple random sampling. Similarly, a researcher studying public perception of government subsidy policy may find probability sampling technique more appropriate than purposive sampling. Hence, the type of sampling is a function of the nature of research question and the research objective.

**Quantitative Data Analysis**

Once quantitative data is generated either from primary or secondary sources, the next step is to decide the nature of data presentation and analysis to be used. The first cluster of analytical tools contains a range of descriptive techniques. Descriptive analysis could take the form of data summary and presentation through tables, graphs and charts using simple percentages.

**Fig. 1:** Example of Descriptive Analysis using Primary Survey Data in Graphical form

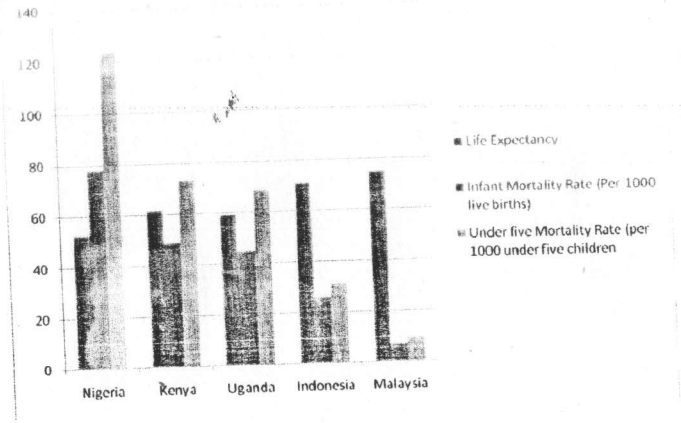
Question: Statement 1- I will follow state orders and directives at all times.
Statement 2- I will ignore and reject state orders if I have the choice




Aremu F. A. (2015) Faith-Based Universities in Nigeria and the Consequences for Citizenship, Africa Today Vol. 62 No.1:3-30 (August)

Fig. 1 contains original quantitative data that was generated from a survey of 1,886 university students in 12 Nigerian universities. The chart describes the pattern of responses to the question on willingness to comply with state orders among students in public and faith-based universities depending on their length of stay on the cluster of university campuses. Although the chart reveals significant variation in the responses of 'early' and 'old' students across public and faith-based universities, it does not mean that length of stay in those campuses 'causes' the variation in political attitude. This is the inherent limitation of descriptive analysis. The simple fact that two variables are related does not necessarily imply that one is causing the other.

Fig. 2: Example of Descriptive Analysis using Secondary Data in Graphical form




UNDP Human Development Report 2014 culled from Hamalai, L. Aremu, F. A. & Umar A. (2015) Industrial Strikes in the Health Sector: Incidence, Causes and Solutions, Policy Paper Series No. 6, National Institute for Legislative Studies, National Assembly, Abuja.

Fig. 2 is a descriptive (graphical) illustration of secondary data on cross-country comparison of basic health indicators using data from UNDP's Human Development Report. Graphical depiction of data allows for simplicity in viewing datasets to illustrate a point or support a claim. It has the advantage of passing the message across in a compelling fashion thereby arresting the attention of readers much more than wordy narratives could do.

Aside tables, graphs and charts, descriptive analysis could take the form of statistical estimation and inference involving measures of central tendency (mean, median and mode), measures of dispersion (standard deviation, ANOVA and t-tests) as well as measures of relationship such as Correlational and regression analysis. In some cases, quantitative designs extend beyond simple description to hypothesis-testing and exploration of causal relations.



In such circumstances, the researcher inclines towards statistical modeling. Let us begin with the familiar example of demand function which stipulates that demand for a commodity will vary as price changes. This could be written using simple linear equation as:

$$QTY = f(P)$$

i.e. quantity demanded is a function of price.
The same equation could be converted to statistical model:

$$Q = \alpha + \beta P + \varepsilon$$

Where $Q$ is quantity demanded; $\alpha$ is constant (quantity demanded when price is zero); $\beta$ is effect of price on additional quantity demanded; and $\varepsilon$ is the "noise" meaning other factors that may affect $Q$ other than price.

Essentially therefore, statistical model building is a collection of symbols which are used to represent reality. If we consider an hypothesized relationship between education and earning such that earning increases with higher education. The statistical (linear) model could be written as:

$$I = \alpha + \beta E + \varepsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\{1\}$$

Where $\alpha$ = a constant amount (what one earns with zero education);
$\beta$ = the effect of an additional year of schooling on income, hypothesized to be positive; and
$\varepsilon$ = the "noise" term reflecting other factors that influence earnings.
The variable $I$ (income) *is termed the "dependent" or "endogenous" variable;*
$E$ (education) *is termed the "independent," "explanatory," or "exogenous"* variable; $\alpha$ is the "constant term" and $\beta$ the "coefficient" of the variable $E$.

Suppose the researcher wishes to extend the model by introducing more causal factors, for instance, age. Since higher earnings could be associated with age, the model could be specified as:

$$I = \alpha + \beta_1 E + \beta_2 A + \varepsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\{2\}$$

The model specification extends as more explanatory variables are added. It is expected that the error margin ($\varepsilon$) will reduce as more relevant explanatory variables are included in the model provided the factors actually influence earning. It is important to stress that there are various tests that are required to determine the 'fitness' of the model as well as ensure that the problems of multicolinearity, reverse causality and heteroskedasticity are minimized (Wooldridge, 2009; Baum, 2006). Various analytical softwares such as STATA,

**References**

Aremu, F. A. (2015) Faith-Based Universities in Nigeria and the Consequences for Citizenship, *Africa Today* Vol. 62 No.1:3-30 (August).

Bartels, L. M. (2010) Some Unfulfilled Promises of Quantitative Imperialism, in Brady, H. E and Collier D. (eds.) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, (2nd Ed.) Maryland: Rowman & Littlefield Publishers, Inc. pp. 83-88.

Baum, C. (2006) *An Introduction to Modern Econometrics using STATA*, Brighton: MIT Press.

Gerring, J. (2012) *Social Science Methodology: A Unified Framework*, (2nd Ed.) Cambridge: Cambridge University Press.

Hamalai, L., Aremu F. A. & Umar, A. (2015) Industrial Strikes in the Health Sector: Incidence, Causes and Solutions, Policy Paper Series No. 6, National Institute for Legislative Studies, National Assembly, Abuja.

Hartwig, F. & Dearing, B. E. (1979) *Exploratory Data Analysis*, Series: Quantitative Applications in the Social Sciences, Sage University Paper 16, Newbury Park: Sage Publications Inc.

King, G., Keohane, R. O. & Verba, S. (1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton, New Jersey: Princeton University Press.

Mahoney, J. (2010) After KKV: The New Methodology of Qualitative Research, *World Politics* 62, no. 1 ( January 2010), 120–47.

Marini, M. M. & Singer, B. (1988) Causality in the Social Sciences, *Sociological Methodology*, Vol. 18, 347-409.

Woodridge, J. M. (2009) *Introductory Econometrics: A Modern Approach*, South-Western Cengage Learning.