STATISTICS FOR THE SOCIAL SCIENCES

Edited by
Adeniyi Gbadegesin
Razaq Olopoenia
Afeikhena Jerome

Ibadan University Press
2005

STATISTICS FOR THE SOCIAL SCIENCES

Edited by

Adeniyi Gbadegesin
Razaq Olopoenia
Afeikhena Jerome

IBADAN UNIVERSITY PRESS
2005

Published by
Ibadan University Press
Publishing House
University of Ibadan
Ibadan, Nigeria

First Published 2005

Reprinted 2007

ISBN: 978 – 121 – 405 – 8

Printed by: Okley Printers (Nig.) Ltd
08037152927, 08052110067

CONTENTS

1

INTRODUCTION

Adeniyi Gbadegesin and Olugboyega A. Oyeranti

Social Sciences and Empirical Analyses

Social science-based disciplines such as economics, geography, political science, psychology and sociology, to mention but a few, are now generally considered as empirical sciences. These disciplines are empirical because they share a common methodology that includes the following as its most important elements:

- Qualitative and quantitative observations of phenomena, either directly or by carefully designed experiments.
- Numerical and statistical processing of the observed data.
- Constructing theoretical models that describe the observed phenomena and explaining the relationships between them.
- Using these theoretical models in order to derive predictions.
- Correcting and improving models so that they predict better.

Empirical sciences thus rely on processes of observation, modeling, and verification. For any academic enquiry to qualify fully as an empirical science, each of the foregoing points is important. Observations without theory can give purely descriptive pictures of reality that lack explanatory power. But theory without observation equally risks losing contact with the reality that it is trying to explain.

It is not a coincidence that social science disciplines are empirically inclined. These disciplines are concerned with the analysis of human behaviour. To analyse human behaviour, they all follow a common trend, which is developing and using systematic, causal explanations of human behaviour. Indeed, systematic, causal relationships of behaviour form the cornerstone of any social science. Explanations about human behaviour are offered and they are conventionally referred to as models since they provide mere statements about the most important determinants of the behaviour being investigated. To a social scientist, models, by design, represent simplifications and abstractions of reality and they help to describe the most important systematic aspects of behaviour seen in a wide variety of circumstances, rather than completely accounting for specific events.

An important tool in the hands of social scientists for analyzing human behaviour is statistical analysis. Generally, statistical analysis is used in the study of occurrences where the true value or relation cannot be measured directly or is hidden by other things. Put differently, the numerical statement on an occurrence or on a relationship cannot be obtained directly from the original or 'raw' figures available. Instead, the data must be analyzed to determine the values desired.

The special need for analytical methods in the social sciences has been clearly stated by an eminent Englishman–Julian Huxley (1940) as follows:

> Causation in social science is never single as in physics or biology, but always multiple and complex. It is of course true that one–to–one causation is an artificial affair, only to be unearthed by isolating phenomena from their total background. Nonetheless, this method is the most powerful weapon in the armoury of natural science: it disentangles the chaotic field of influence and reduces it to a series of single causes, each of which can then be given due weight when the isolates are put back into their natural interrelatedness, or when they are deliberately combined (as in modern electrical science and its

> applications) into new complexes unknown in nature. This method of analysis is impossible in social science. Multiple causation here is irreducible.
>
> The problem is a two–fold one. In the first place, the human mind is always looking for single causes for phenomena. The very idea of multiple causation is not only difficult, but definitely antipathetic. And secondly, even when the social scientist has overcome this resistance, extreme practical difficulties remain. Somehow, he must disentangle the single causes from the multiple fields of which they form an inseparable part. And for this, a new technique is necessary.

In terms of research, social scientists employ statistics to do a number of things. First, to design surveys and experiments, bearing in mind the cost of obtaining a specified quantity of information. Second, to make inferences from a given sampling situation. Beyond these two major uses, social scientists do rely on statistics to measure the goodness of an inference and also to ensure that all designs and inference–making procedures are selected to achieve an inference of specified quality, but at minimum cost. At this point, there is the obvious need to address a question: What is statistics?

Meaning and Nature of Statistics

The aspect of decision-making that has to do with numerical information is known as statistics. The word *statistics* is often used to mean any of the following: numerical information, a summary of numerical information, and a discipline. 'Statistics', in general, is used to describe a collection of numerical data. For example, population data, health statistics, and school enrolment data, among others. It is also used to refer to a summary of a collection of numerical data, such as the total, the minimum, the maximum, the range, and the average of such data of interest.

As a discipline, statistics is the scientific method of decision making under uncertainty when numerical data (information in

numerical form) and calculated risks (because of uncertainty) are involved. In sum, statistics is defined as the science of collecting, organizing, presenting, analysing and interpreting numerical data. Thus, statistical methods are those procedures used in the collection, presentation, analysis, and interpretation of data.

Attempts at defining statistics from a social science perspective often lead to the considerations of two types of statistics: descriptive statistics and inferential statistics.

Descriptive Statistics: Simply defined, descriptive statistics constitutes an aspect of statistics that is concerned with the description of large masses of data where the social scientist has information on all measurements in a population as well as all statistical techniques employed to carry out such description. This implies that descriptive statistics only provides information about the collected data and in no way draws inferences or conclusions concerning a larger set of data.

Inferential Statistics: Inferential statistics is a theory of information concerned with the acquisition of data, sampling, and the use of data in making inferences about a population from which a sample of data is selected. In the field of statistical inference, social scientists are interested in arriving at conclusions concerning a population. Whenever it is practically impossible to make the entire set of observations that make up a population, what is done is to depend on a subset of observations from the population to help in the making of inferences concerning the same population. This brings a social scientist in contact with the theory of sampling in statistics.

Objectives and Structure of the Book

The objectives of this book are:

(a) to help students acquire an understanding of the prominent statistical methods underlying research methods commonly used by social scientists;
(b) to help students develop, modify and, where necessary, improve their views about social research methodology by reading (both in this book and others) and examining case

studies that will allow them to express their personal views;

(c) to help students distinguish between disciplinary, subject–matter, and problem–solving research in terms of:
  (i) their main characteristics
  (ii) difficulties commonly encountered in organizing, conducting and evaluating different types of research;

(d) to expose students to more recent ways and means of solving statistically related problems, both in terms of information gathering and processing.

The book is not entirely structured in the order of these objectives, though all the objectives constitute the focus of the book. As argued previously, social science-based disciplines are genuinely interested in causal relationships. Hence, there is always the need to go beyond the collection of observations and the construction of the necessary variables. The inclusion of some topics, such as correlation analysis, regression analysis, and the use of information technology in statistical analysis, adds a unique value to the book. For example, the basic conceptual approach underlying correlation coefficients in the social sciences is the development of a standardized measure of association between variables.

What makes empirical work in the social sciences different, and sometimes often more difficult than in the physical sciences, is the lack of an experimental methodology by which to control the human and societal variables being investigated. Beyond the problem of control, observations (data) sometimes fail to explain the true behaviour of the population This is always so when it is found that a portion of the observed behaviour is unexplained by the hypothesized model, thereby rendering a simple deterministic model of the type 'X is sufficient to predict Y' [Y = F (X)] irrelevant. When this happens, instead of modeling the observed behaviour of Y solely as a function of X, social scientists translate a deterministic model to a stochastic model by adding a random or stochastic variable popularly written as μ. The specification of a stochastic model yields Y = F* (X, μ). It is important to know that

the stochastic element is always unobserved, as only Y and X are observed. There is one important implication of adding the unobserved component in the social sciences: social scientists must concentrate on analytical tools that are appropriate for probabilistic models. In other words, even when the value of X and the functional form F*( ) are known, to predict the value of Y with certainty remains a problem because of the presence of μ. The need to take care of stochastic models in social science research provides the justification for topics on probability and regression analysis. This constitutes another area of value-addition of the book.

Definitions of some Concepts

Definition 1:1

Statistic: A statistic is any numerical information describing a characteristic of a sample, e.g., the age of a student.

Definition 1:2

Parameter: A parameter is any numerical value or information describing a characteristic of a population.

Definition 1:3

Population: A population is the set representing all measurements of interest to the research.

Definition 1:4

Sample: A sample is a subset of measurements selected from the population of interest.

Definition 1:5

Estimate: An estimate is a number computed from sample data used to approximate a population parameter.

Definition 1:6

Estimator: An estimator is a rule that tells how to compute an estimate based on information contained in a sample. An estimator is usually given as a mathematical formula.

Definition 1:7
Random Sample: A sample of n measurements selected in such a way that every different sample of n elements in the population has an equal probability of being selected.

Definition 1:8
Random Variable: A random variable is associated with an experiment. Its values are numerical events that cannot be predicted with certainty.

REFERENCE

Julian Huxley (1940), "The Science of Society", *Virginia Quarterly Review*, Vol. 16, No. 3, Pp. 348–365, Summer.