METHODOLOGY OF

BASIC AND APPLIED RESEARCH

Proceedings of a Workshop

Edited by
A. I. OLAYINKA. L. POPOOLA & A. OJEBODE

METHODOLOGY OF BASIC

AND

APPLIED RESEARCH

Proceedings of a Workshop

Edited by

A.I. Olayinka, L. Popoola & A.Ojebode

Published by: The Postgraduate School
University of Ibadan,
Ibadan.

E-mail: postgrad@mail.ui.edu.ng
Website http://www.postgraduateschool.ui.edu.ng

First Published 2008

ISBN 978-075-071-1

CONTENTS

Chapter 11

Justification for the Use of Statistics in Research

Gbenga Adewale

Introduction

Statistics as used in this chapter refers to the techniques and theories employed in transforming raw data into useful and meaningful information for making logical and tenable decisions and conclusions. In this chapter, we justify the use of statistics in research. Specifically, we shall examine the relevance of statistics in research, the different statistical approaches and specific statistical techniques.

Relevance of Statistics in Research

In quantitative research, statistics helps the researcher to:

- acquire the techniques of data collection;
- understand statistical procedures or methods for analyzing data; and
- acquire competence in planning and carrying out research.

Descriptive Statistics

This is an aspect of statistics used to describe the characteristics of a group. It involves the process of determining the mean or average ages of a group or data. There are other descriptive measures like the mode, median, which could be used based on the need. In addition, measures of variability like range, variance and standard

deviation could also be used in describing the characteristics of a sample.

Inferential Statistics

Inferential statistics rest on the strength of deductive reasoning. It has to do with making inferences about a population from studying a sample. For instance, we can use a sample of one thousand (1000) students, we can use it to infer the population of the students. This deals with making deductions or inferences based on the values from a sample.

Data Preparation

It is proper to find out if the respondents provided useful information that could be used to trace their responses long after they have finished responding to the research instruments. The first thing to do is to enumerate all the copies of the questionnaire collected from the field serially. In addition, demographical data should also be checked to make sure there are no empty spaces. For example, if a respondent forgets or decides not to include his/her gender, but responded that she had been pregnant before, this would give us a clue that the respondent is female.

Furthermore, if a respondent decided not to indicate his socio-economic status (in terms of high or low) but indicated that he has 4 cars, has built 7 houses, and employs more than 1000 employees, this may also help the researcher to identify the respondent as a high SES respondent. Second, it is important to code the information either directly into the computer or using coding sheets (these are available in the University Computing Centre).

Scale of Measurement.

The assignment of numbers or values to the variables under study (e.g. weight is equal to 45kg) is not done arbitrarily; a certain procedure is followed. We cannot use a thermometer to measure

weight or a ruler to measure temperature; each of these physical characteristics has its own instrument and scale for measurement.

In the physical sciences, the appropriate instrument is used for the appropriate characteristic of interest. This is also applicable to the social sciences and humanities. There are various ways in which numbers or values are assigned to variables. In some cases, simple classification is done while in others we arrange in order of size. There are 4 different scales of measurement. These are Nominal Scale, Ordinal Scale, Interval scale and Ratio scale.

Nominal Scale

Data in nominal scale do not have magnitude, equal interval or absolute zero. Nominal scale is used for identification or for categorization purposes. For example, in a football match, players wear jerseys with different numbers (usually 1 to 11). That a player wears jersey number 1 does not indicate that he is the best or the worst player. Hence, we often categorize and make frequency counts on the group e.g. categorizing male and female as M and F or 1 and 2. Another example is brands of cars. They can also be categorized like 1=Jeep, 2=Benz, 3=Nissan, 4=Toyota. The value attached to each of these brands of cars does not connote serious meaning because it cannot be concluded that a jeep is the best or worst car because we are using nominal data.

Ordinal Scale

Here the order or arrangement is important. The order is more than a mere listing. It is a progression from the highest to the lowest or from lowest to the highest.

Consider these average scores

1.	76.5	1
2.	76.3	2
3	73.0	3
4.	71.0	4
5.	64.0	5

Ordering here, It is done by ranking. It has magnitude but no equal interval and absolute zero. In ordinal scale of measurement, we have ranks to indicate positions which do not give any information about distance/interval between points on the scale.

Interval Scale

The differences between successive points on the scale are equal. Hence, they have equal intervals. For example, the difference between 10cm and 15cm in a ruler is 5cm. This is possible because equal intervals are used in the measurement. There is no absolute zero point. In educational measurement, there is a difference of 10 between a pupil who scored 65 and another who scored 75. The assumption is that the result obtained from the instrument used in measuring these pupils' ability is reliable (error is minimized). This is usually difficult and that is why it is emphasised that instruments in educational measurement should be valid and reliable.

Ratio Scale

A ratio scale, it has all the quantities such as absolute zero and equal intervals. For example, in the thermometer the absolute zero is 273 Kelvin. This makes comparison between points in the scale possible and valid.

Data Organization: Frequency Distribution

A frequency distribution is a list showing the number of times each score value (or interval of score value) occurs in a distribution. It provides a summary of test scores in a form which accurately depicts the group as a whole.

How to Construct Frequency Distribution

- Arrange the data in descending order (see table 1).
- Use a letter (e.g. x) to symbolize a score value.
- Make a tally against each occurrence of the score value (see table 1).
- Record the number of tallies made against each score value (i.e. frequency).Frequency is represented by the letter *f*.

Measures of Central Tendency

A measure of central tendency is an index whose value depicts the typical performance of the whole class.The three common indices of central tendency are the *mean, median* and *mode.*

When to use Mean, Median and Mode

The mean score is used as a typical score representing an entire distribution when the distribution is not skewed. This condition is illustrated using the distribution (3,4,5,6,6,7,8,8,9) which has a mean of 6.22, a median and a mode of 6. Since the distribution is not skewed, it is best to use the mean as a measure of central tendency of the distribution.

When the distribution is skewed, however, the median is used. The modal score is rarely used as a typical score representing an entire distribution. It should be used to supplement any information from the mean or median, especially when the distribution has more than one mode.

Concept of Dispersion

Dispersion or variability refers to the extent to which scores in a distribution deviate from each other and their central tendency. There are different measures of variability namely: range, semi-interquartile range, variance and standard deviation.

The Range

Range only takes into consideration the two extreme scores and therefore it is not a true reflection of the variability of all the scores in the distribution. Again, all other scores may be changed provided the change does not affect both the upper and the lower scores. Hence, we do not have a good reflection of the variability of all the scores in the distribution.

Standard Deviation

Consider the following illustration with 2 distributions A and B.

(15,18,21,24,27) = A

(5,13,21,29,37) = B

These two distributions have the same mean (a value of 21). However, the difference between any two scores in distribution A is 3 whereas in distribution B it is 8. Distribution A disperses less than distribution B. Again, though the mean value of both is 21, the distributions have different deviations from their central tendency measures.

These deviations are computed as follows:

Table 1

Distribution A	Deviation from mean	Distribution B	Deviation from mean
15	6	5	16
18	3	13	8
21	0	21	0
24	3	29	8
27	6	37	16
	Total deviation=18		Total deviation = 48

The total deviation of distribution B is more than that of distribution A. Distribution B disperses more than distribution A.

Correlation

Correlation means *co* and *relation*. *Co* means joint while *relation* means association. Therefore, correlation means joint association.

There are three possible considerations for correlation – positive, negative and zero.

Positive: examples are height and weight (as height increases, weight also increases), time and temperature, I.Q. and achievement, age and fertility, area and radius of circle.

Negative: examples are the age of a car and resale value, force of attraction and distance etc.

Zero: examples of things that are not related are shoe sizes and income, weight and I.Q, height and I.Q etc.

When two things are related, it does not mean that one causes the other. There are also two things to note about a correlation index. They are the strength or magnitude of the correlation index and direction (sign either +ve or –ve) of the correlation index.

Uses of Correlation Index

A correlation index helps to predict the performance of students. Thus correlation is used as an instrument of prediction For instance, if JAMB entrance examination scores used for selecting students for Engineering happen to correlate with their performance in the university examination in engineering, we would say the examination is valid for selecting students for engineering. This sort of validity is called predictive validity.

It helps to ascertain the reliability of a test is i.e. how stable and consistent the students performance is with time over a test. We do this by Test-Retest method and this determines stability coefficient. It is also used to test for equivalence. Through this, the coefficient of equivalence can be determined. Coefficient of Equivalence tells us how comparable the two (2) tests are. It is also used to determine the internal consistency of a test. This can be done by split-half reliability. This in turn tells us how each item in a part is related to another item in another part of the test.