

UNDERGRADUATE PROJECT SUPERVISION: ANALYSIS OF DATA

Prof. Fidelis A. Onwiodouokit
Department of Science Education
University of Uyo.

Being a paper presented at the Seminar on Project Supervision, organized by Maritime Academy of Nigeria, Oron on Thursday, June 12, 2008.

PREAMBLE: **Project method** is one of the fundamental methods in a teaching/learning process. It is required, generally, that at the end of programmes at our tertiary level of Education, students are expected to carry out project works duly guided and supervised. My experience with students and staff over the years have shown that most students and even supervisors, in some cases, do not seem to be very clear on what is expected of them.

This seminar, arranged by the authority of this institution is therefore very worthwhile and commendable.

To ensure a proper conduct and supervision of project at the first degree level is to ensure a proper research foundation in the future. A student who has a firm grasp of project is most likely to be a good researcher in his/her field of study in the future.

SUBSET OF PROJECT: A project, though a concrete whole, has many subsets or components. Prominent among these are: the topic, statement of the problem, research questions and hypotheses; significance of the study, literature review, research methods, analysis of data and the discussion of findings.

In this seminar, I am asked to address the issue of data analysis, and this, therefore, is the focus of this paper.

DATA ANALYSIS IN STUDENTS PROJECTS|

Introduction:

A good and meaningful project that is capable of contributing positively to knowledge is not without data collection. The essence of this is

either to answer a research question or to test a stated hypothesis or both. Whenever data are collected in the course of the study, there is need to use the relevant statistical tools to analyze them.

METHOD OF DATA ANALYSIS

The methods required for analyzing any data are predicated on the following:

1. The Project topic;
2. Research question and hypothesis;
3. Research designs adopted; and
4. the nature of data generated for the study.

RELEVANT STATISTICS FOR PROJECT DATA ANALYSIS

Statistics is generally seen as a scientific way of collecting, organizing, summarizing, presenting, and analyzing data in order to draw valid conclusion and make reasonable and meaningful decision based on such analysis. There are basically two types of statistics, namely, descriptive and inferential (inductive) statistics.

Descriptive statistics seeks to analyze and describe the characteristics features of a given group or sample without an intention to draw conclusion or make inference about the larger group or population. On the other hand, inferential statistics seeks to go beyond the sample under study, to generalize on the population. It should however be noted that for the outcome of the inferential statistical analysis to be meaningful and useful, the sample selected and studied must be a good representation of the population.

DESCRIPTIVE STATISTICS

In Project Supervision, descriptive statistics can be very useful in summarizing the information or data collected. This can even be in Pictorial form. Descriptive statistics at this level has to do mostly with the measures of central tendency which include:

- (i) Arithmetic mean
- (ii) Geometric mean
- (iii) Harmonic mean
- (iv) Median.
- (v) The Time Series Analysis

Arithmetic mean,

This is simply referred to as mean with symbol, \bar{X}

$$\begin{aligned}\bar{X} &= \frac{X_1 + X_2 + X_3 \dots + X_N}{N} \\ &= \frac{\sum_{j=1}^N X_j}{N} \\ &= \sum \frac{X}{N}\end{aligned}$$

Example 1: Find the mean of the following numbers
10, 25, 35, 15, 5, 20.

$$\begin{aligned}\bar{X} &= \frac{\sum X}{N} \\ &= \frac{(10 + 25 + 35 + 15 + 5 + 20)}{6} \\ &= \frac{110}{6} = 18.33\end{aligned}$$

WEIGHTED MEAN, X_w

In weighted mean, the relative weights of scores are considered. For example, let us consider a year one student offering the following general courses and earning the scores shown below:

COURSE CODE	CREDITS HOUR (WEIGHT)	SCORES, X
PHY 111	3	60
CHM 111	4	45
MATT 111	4	40
PHY 112	3	50
CHM 111	3	55
MATT 112	3	30

4

In calculating the weight mean, \bar{X}_w , the relative weight of scores are brought into the calculation, hence the table:

COURSE CODE	WEIHT, W	SCORES, X	WX
PHY 111	3	60	180
RHM 111	4	45	180
MTH 111	4	40	160
PHY 112	3	50	150
CHM 112	3	55	165
MTH 1/2	3	30	90

$$\bar{X}_w = \frac{WX}{W} = \frac{925}{20}$$

$$= 46.25.$$

Where W = total weight. This is similar with frequencies.

GEOMETRIC MEAN, G

The geometric mean of a set of N numbers, $X_1, X_2, X_3, X_4 \dots X_N$ is the Nth root of the product of the numbers.

$$G = \sqrt[N]{X_1 X_2 X_3 X_4 \dots X_N}$$

For example, the geometric mean of the numbers 2, 4, 8 is $\sqrt[3]{2 \times 4 \times 8}$
 $= \sqrt[3]{64} = 4.$

HARMONIC MEAN, H

The harmonic mean, H of a set of N numbers $X_1, X_2, X_3, \dots, X_N$ is the reciprocal of the arithmetic mean of the reciprocals of the numbers:

$$H = \frac{1}{\frac{1}{N} \sum_{j=1}^N \frac{1}{X_j}}$$

$$= \frac{N}{\sum \frac{1}{X}}$$

5

For example, the harmonic mean, H of 2, 4, 8, is given as

$$H = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}}$$

$$= \frac{3}{\frac{7}{8}} = 3.43$$

MEDIAN

Median is a middle score in a distribution when arranged in an order of magnitude.

Example:

Find the median scores of the following distributions.

- (i) 12, 8, 10, 5, 15, 6, 3
- (ii) 5, 28, 30, 2, 16, 14

Median of frequency data

Example

X	f	f Above	f below	f A - B
0	2	98	0	98
1	8	90	2	88
2	9	81	10	71
3	9	73	19	54
4	14	58	28	30
5	11	47	42	5
6	14	33	53	-20
7	11	22	67	-45
8	10	12	28	-66
9	8	4	88	-84
10	4	0	96	-96

$$\text{Median} = \frac{\sum f = 100}{2} L + \left\{ \frac{N}{2} - \sum f_B \right\} C$$

f_{median}

b

Where L = Lower limit of the median class

N = Number of items in the Data

fl = Sum of frequencies of all cases
lower than the median class

C = Size of the mean class

f_{median} = Frequency of the median class.

$$L = 4.50$$

$$N = 100$$

$$fl = 42$$

$$f_{\text{median}} = 11$$

$$C = 1$$

$$\begin{aligned} \text{Median} &= 4.50 + \frac{(50 - 42)}{11} \\ &= 5.23 \end{aligned}$$

It should be noted that data involved in descriptive statistics can be described pictorially or graphically using pie charts, bar charts, and histograms.

INFERENCIAL STATISTICS

Under inferential statistics, we shall consider

- (a) Correlation analysis and production
- (b) t = Test analysis,

(a) Correlation Analysis

The essence of this analysis is to investigate the extent of relationship existing between two or more sets of variables. For instance, it is possible to find out whether or not there exists a relationship between students' performances in Further Mathematics and their performances in an Engineering course.

Further
Maths

7
Engineering
Course

X	X ₂	Y	Y ²	XY
30	900	25	625	750
32	1024	60	3600	1920
10	100	20	400	200
30	900	65	4225	1950
58	3364	40	1600	2320
60	3600	50	2000	3000
20	400	35	1225	700
70	4900	80	6400	5600
55	3025	45	2025	2475
48	2304	53	2809	2544
35	1225	40	1600	1400
60	3600	55	3025	3300
70	5625	80	6400	6000
38	1444	48	2304	1824
40	1600	50	2500	2000
Σ661	34011	746	41238	35983

$$\bar{X} = 44.07, \quad \bar{Y} = 49.73$$

$$Y = a + bx$$

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\{[N \sum x^2 - (\sum x)^2] [N \sum y^2 - (\sum y)^2]\}^{1/2}}$$

$$= \frac{15 \times 35985 - (661 \times 746)}{[(15 \times 34011) - (436,921)][618570 - 5565165]}$$

$$= \frac{46,639}{\sqrt{4,549,083.17}}$$

$$r = \frac{46639}{67417.23}$$

$$= 0.69$$

Because the relationship may not be a linear one, it is always advisable to convert the r - value to t - value j as.

$$\begin{aligned}
 t &= \frac{r - 0}{\frac{\sqrt{1-r^2}}{N-2}} \\
 &= \frac{0.69}{\frac{(1-0.48)}{13}} \\
 &= \frac{0.69}{\frac{(-0.48)}{\sqrt{0.04}}} \\
 &= \frac{0.69}{0.2} = 3.45 \\
 t &= 3.45
 \end{aligned}$$

In order to predict whether the score of a particular student in one subject, given the other, the line of best fit (that is regression line), the slope and the intercept involved must be determined in $y = a + bx$

Note a is a Point at which the line crosses the vertical axis, y , while b is the slope of the line.

$$\begin{aligned}
 b &= \frac{\sum xy - (\sum x)(\sum y)}{N} \\
 &\quad \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N} \\
 &= \frac{SS_{xy}}{SS_x}
 \end{aligned}$$

$$= \begin{array}{r} 35983 \\ \hline 34011 \end{array} - \frac{493106}{15}$$

$$= \frac{3109.27}{4882.92}$$

$$b = 0.64$$

Using the means values of x and y

$$\bar{y} = a + b\bar{x}$$

$$\begin{aligned} 49.73 &= 0.64 \times 44.07 \\ a &= 49.73 - 28.21 \\ &= 21.52 \end{aligned}$$

Hence the predicted score in Engineering course, y^1 for a student who scored 68% in Further Mathematics is given as

$$\begin{aligned} y^1 &= a + 21.52 \times 68 \\ &= 21.52 + 43.52 \\ &= 65.04\% \end{aligned}$$

THE TIME SERIES ANALYSIS

A time series is a set of observation taken at specific times, usually at equal intervals. Examples of time series are the total production of goods over a number of years, hourly temperatures announced by the weather bureau of a city, a total monthly sales receipts in a departmental store, etc.

MOVING AVERAGE

Given a set of numbers y_1, y_2, y_3 in a moving average or order N to be given by the sequence or arithmetic mean

$$\frac{y_1 + y_2 + \dots + y_n}{N}, \frac{y_2 + y_3 + \dots + y_n + y_{n+1}}{N}$$

$$\frac{y_3 + y_4 + \dots + y_{N+2}}{N}$$

for example for a given set numbers 2,6,1,5,3,7,2, a moving average of order 3 is given as

$$\frac{2+6+1}{3}, \frac{6+1+5}{3}, \frac{1+5+3}{3}, \frac{5+3+7}{3}, \frac{3+7+2}{3}$$

That is 3,4,3,5,4

Original: 2,6,1,5,3,7,2

Moving average of order 3: 3, 4,3,5,4

The moving average tends to reduce the amount of variation passed in a set of data. In time series this process is used to remove unwanted fluctuation and is called smoothing of time series.

t- Test Analysis

There are two versions of the t test analysis, viz, for uncorrelated (independent) and correlated dependent data.

t – test for uncorrelated data.

There are two forms of this:

- (a) Pooled variance model
 - (b) Separate variance model
- (a) Pooled variance model

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}^{1/2}$$

- (b) Separate variance model

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}^{1/2}$$

t – test for correlated data

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\left\{ \left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right] - 2r \left[\frac{s_1}{\sqrt{n_1}} \right] \left[\frac{s_2}{\sqrt{n_2}} \right] \right\}^{1/2}}$$

To calculate this, both standard deviation, s and variance s² must be calculated besides the respective means of the two groups.

$$S = \sqrt{\frac{\sum(x-\bar{x})^2}{N-1}}$$

$$= \left\{ \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N-1} \right\}^{1/2}$$

Conclusion

As mentioned earlier, the use of appropriate statistics in analyzing data generated in research and project works is very crucial. It is not the sophistication but appropriateness of the statistical tools employed. It should also be noted that Statistical packages are available for analysis of various data using computer but both the student and the Supervisor should know what the computer does.

Reference:

Johnson, M. K and Liebert, R. M. (1992). Statistics; tool of the Behavioral Science. New York: Holt, Rinehart and Winston

Spiegel, M. R. (1972). Schaum's Outline Series: Theory and Problems of Statistics. New York: McGraw-Hill publishing Co. Ltd.

Onwioduokit, F. A. (200). Educational Research Methodology and Statistics. Uyo: Dorand publishers.
