Research in Education

Edited by:

A. O. Jaiyeoba
A. O. Ayeni
A. I. Atanda

Research

in

Education

Edited by:

Adebola O. Jaiyeoba
A.O. Ayeni
A. I. Atanda

Edited by:

ISBN: 978-978-55702-8-1

Published by:

The Department of Educational Management
Faculty of Education, University of Ibadan, Ibadan

Printed by:

HIS LINEAGE PUBLISHING HOUSE
09 Alli Street, Mokola Ibadan, Nigeria.
Tel: 0803 3596 818
E-mail: awemakin@gmail.com

Preface

Undergraduate and Postgraduate students in tertiary institutions more often than not, face some challenges in the course of their research work and these often result in delay in the completion of such work. This concern necessitated the production of this book

The book "Research in Education" is made up of thirteen chapters and designed to provide modern, straightforward and easy to read guides for readers in writing acceptable projects, dissertations and theses. The book has been written by experts, based on their rich and long standing experiences in their different areas. Also, the writers had written on different topics to guide readers through in the course of writing their projects.

Topics that were covered in the book included: meaning and purpose of research in education; criteria for the identification and diagnosis of a viable research problem; research questions and hypotheses; types, uses of research variables and definition of terms. Other topics covered in this book are: literature review; theoretical and conceptual framework in research; research design and procedure; instrumentation; issues of validity and reliability. Additionally, other issues such as: types and methods of data analysis; techniques of writing of abstract; methods of referencing in research; vetting and editing in writing and finally, preparation and presentation of research reports were also addressed in the book.

The Department acknowledges the administration that initiated this book project under the headship of Dr. F. S. Akinwumi (Now Professor F. S. Akinwumi). Also, efforts of the contributors to this book are well recognized, while the editors are also appreciated. I believe that the book will go a long way to assist the students and researchers to understand and appreciate the fundamental nature of research in education. Other experienced researchers too will find the book useful in meeting their immediate professional needs.

Ayeni, Abiodun Olumide
Department of Educational Management
Faculty of Education
University of Ibadan, Ibadan

Contents

Chapter 1

Meaning and Purpose of Research in Education
Akinwumi, F. S.; Odinko, Monica N. & Akintayo, A. I. 1

Chapter 2

Criteria for the Identification and Diagnosis of a Viable Research Problem
Adedeji, S. O. 13

Chapter 3

Research Questions and Hypotheses
Adegbesan, O. A. & Emunemu, B. O 25

Chapter 4

Research Variables: Types, Uses and Definition of Terms
Abiodun – Oyebanji, Olayemi J. 43

Chapter 5

Literature Review
Fakolade, O. A & Atanda, A. I. 55

Chapter 6

Theoretical and Conceptual Framework in Research
Otunla, A. O. 71

Chapter 7

Research Design and Procedure
Salami, I. A. & Isah, E. A. 93

Chapter 8

Instrumentation and Issues of Validity and Reliability
Adeleke, J. O. 121

Chapter 9

Types and Methods of Data Analysis
Adeleke, J. O. 145

Chapter 10

Techniques of Writing Abstract
Akinsola, M. K; Ayeni, A.O. & Kester, K.O 175

Chapter 11

Methods of Referencing in Research
Babalola, J. B. &. Onuka, A. O.U 185

Chapter 12

Vetting and Editing in Writing
Kolawole, C. O. O. & Oduwole, Yetunde A. 211

Chapter 13

Preparation and Presentation of Research Reports 221
Jaiyeoba, Adebola O.

Chapter 9

TYPES AND METHODS OF DATA ANALYSIS

Adeleke, J. O.
Institute of Education,
University of Ibadan, Nigeria.
+2348033510688
joadeleke@yahoo.com

Introduction

Statistics is concerned generally, with scientific methods for collecting, organizing, summarizing, presenting, and analyzing data as well as with drawing valid conclusion and making reasonable decisions on the basis of such analysis. Hence, every university student will employ statistics to analyse collected data for completion of project/dissertation/thesis which is a compulsory requirement for any programme. Three major types of data analysis are discussed under this chapter, (i) graphical representation of data (ii) Descriptive statistics and (iii) Inferential statistics. Various methods of data analysis under each type are equally addressed.

Graphical Representation of Data

The graphical representation of data makes the reading more interesting, less time-consuming and easily understandable. The disadvantage of graphical presentation is that it lacks details and is less accurate. There are so many ways data can be represented, only four common graphs will be presented in this chapter. They are 1. Bar Graphs 2.Pie Charts 3.Frequency Polygon 4. Histogram.

Bar Graphs

This is the simplest type of graphical presentation of data. The following types of bar graphs are possible: (a) Simple bar graph (b) Double bar graph (c) Divided bar graph.

Example 1

Table 1

Item	Food	Rent	Education	Savings	Misc.	Total
Amount (Rs.)	3000	800	1200	1500	700	7200

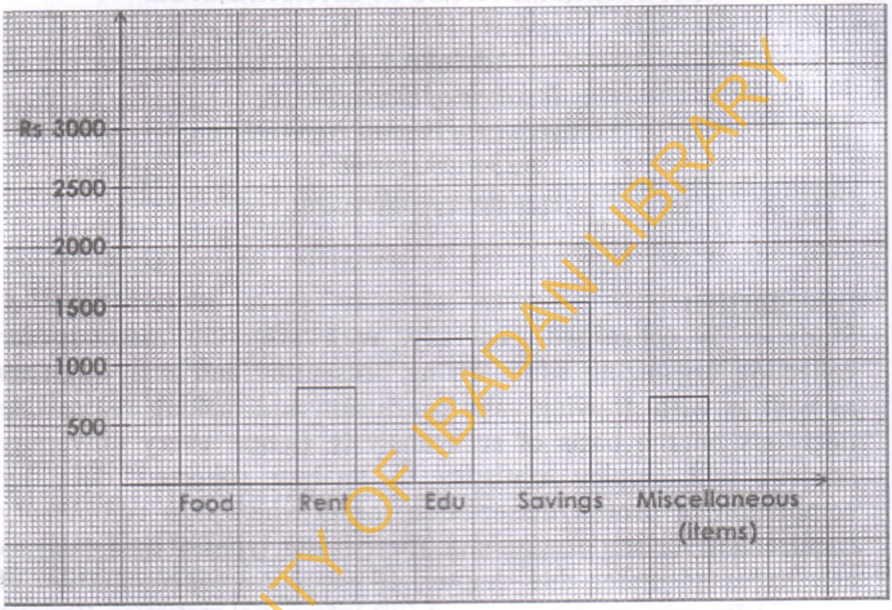

Fig. 1.Bar Chart

Pie Graph or Pie Chart

Sometimes a circle is used to represent a given data. The various parts of it are proportionally represented by sectors of the circle. Then the graph is called a Pie Graph or Pie Chart.

To find the angle of each sector

Total of data corresponds to 360º.

Let xº = the angle at the centre for item A, then

$$x^0 = \frac{\text{Value of item A}}{\text{Total value of all the items}} \times 360^0$$

Example 2

The data given in example 1 can be used to draw a pie graph.
Calculation of Angles

Food:

Angle at centre

$$= \frac{360}{7200} \times 3000 \quad = 150^{o}$$

Rent:

Angle at centre

$$= \frac{360}{7200} \times 800 \quad = 40^{o}$$

Similarly we can calculate the remaining angles, and the total of angles column should always come to 360^{o}.

Table 2:

Item	Amount (Rs.)	Angle
Food (A)	3000	150^0
Rent (B)	800	40^0
Education (C)	1200	60^0
Savings (D)	1500	75^0
Miscellaneous	700	35^0
Total	**7200**	**360^0**

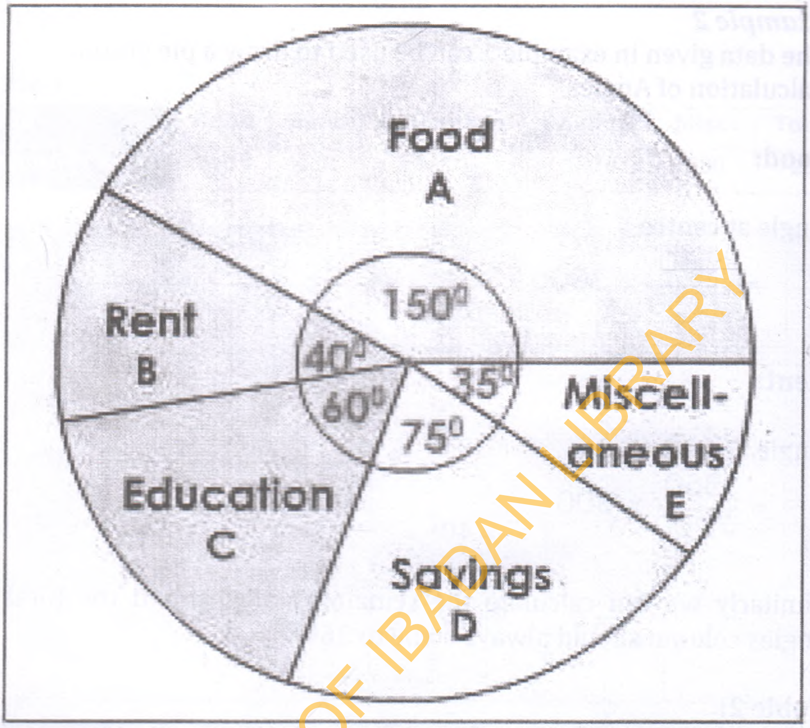

Fig 2: Pie Chart

Frequency Polygon

In a frequency distribution, the mid-value of each class is obtained. Then on the graph paper, the frequency is plotted against the corresponding mid-value. These points are joined by straight lines. These straight lines may be extended in both directions to meet the X - axis to form a polygon.

Example3

The weights of 50 students are recorded in Table 3. Draw a frequency polygon for this data.

Table 3:

Class	Mid-mark	Frequency
40 - 44	42	3
45 - 49	47	10
50 - 54	52	12
55 - 59	57	15
60 - 64	62	7
65 - 69	67	5

Suggested answer:

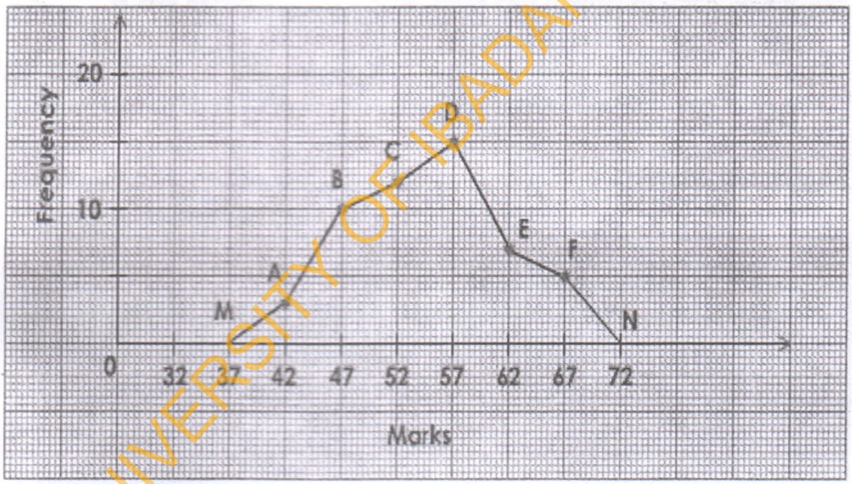

Fig. 4: Frequency Polygon

Example 4

The marks scored by 120 students in an examination are as given in Table 5, form a frequency polygon.

Table 5:

Marks	Frequency
0 - 10	2
10 - 20	8
20 - 30	10
30 - 40	15
40 - 50	24
50 - 60	36
60 - 70	14
70 - 80	6
80 - 90	5

Suggested answer:

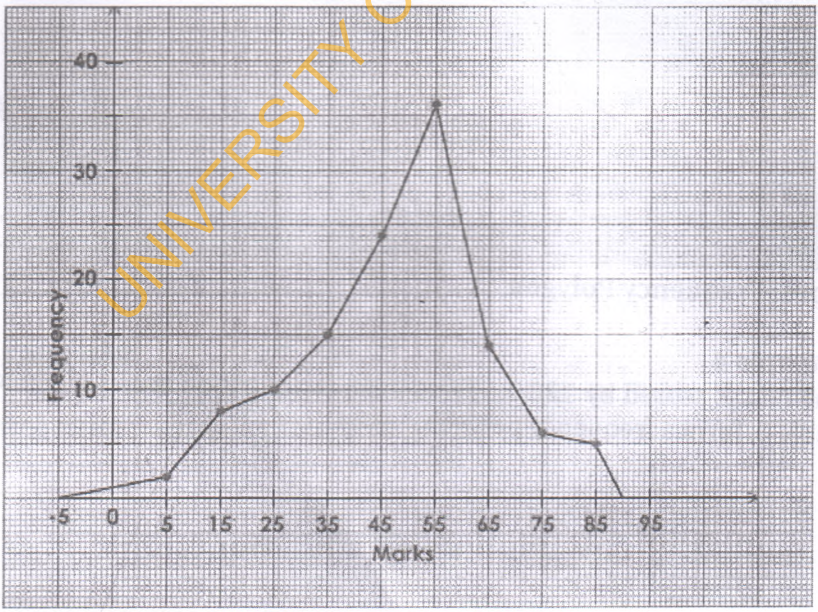

Fig 5: Frequency polygon

Histogram

A histogram is a diagram which represents the class interval and frequency in the form of a rectangle.

To draw a histogram, follow the steps stated below

(1) Mark class intervals on X-axis and frequencies on Y-axis.
(2) The scales for both the axes need not be the same.
(3) Class intervals must be exclusive. If the intervals are in inclusive form, convert them to the exclusive form.
(4) Draw rectangles with class intervals as bases and the corresponding frequencies as heights.

The class limits are marked on the horizontal axis and the frequency is marked on the vertical axis. Thus a rectangle is constructed on each class interval.

If the intervals are equal, then the height of each rectangle is proportional to the corresponding class frequency.

If the intervals are unequal, then the area of each rectangle is proportional to the corresponding class frequency.

Example 5:

Draw a histogram for the data in Table 6:

Table 6:

Class Interval	Frequency
0 - 5	4
5 - 10	10
10 - 15	18
15 - 20	8
20 - 25	6

Suggested answer:

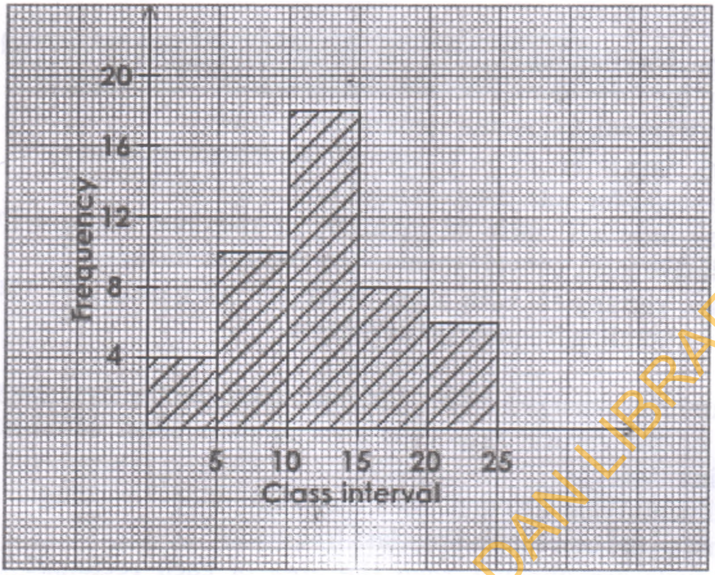

Fig. 6: Histogram

Note:

In Fig 6, the intervals are exclusive. Now, let us consider an example with inclusive intervals.

Example 6

The daily wages of 50 workers, in rupees, are given below:
In table 7, the class intervals are inclusive. So we convert them to the exclusive form as shown in table 8.

Table 7

Wages (in Rs.)	Frequency
51 - 60	4
61 - 70	12
71 - 80	8
81 - 90	16
91 - 100	4
101 - 110	6

Table 8

Wages (in Rs.)	Frequency
50.5 - 60.5	4
60.5 - 70.5	12
70.5 - 80.5	8
80.5 - 90.5	16
90.5 - 100.5	4
100.5 - 110.5	6

Suggested answer:

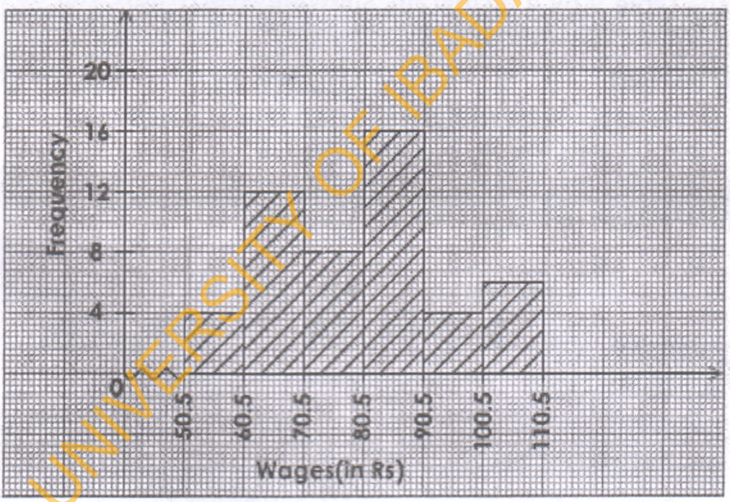

Fig. 7: Equal width Histogram

Note:

(i) The class intervals are made continuous and then the histogram is constructed.

(ii) A kink or a zig - zag curve is shown near the origin. It indicates that the scale along the horizontal axis does not start at the origin.

(iii) The horizontal scale and vertical scale need not be the same.

Example 7

Distribution of shops according to the number of wage - earners employed at a shopping complex is given in table 9:

Table 9:

Number of wage-earners	Number of Shops	Frequency density
Under 5	18	3.6
5 - 10	27	5.4
10 - 20	24	2.4
20 - 30	20	2.0
30 - 50	16	0.8

Illustrate the table 9 by a histogram, showing clearly how you deal with the unequal class intervals.

Suggested answer:

Note:

When the intervals are unequal, we construct each rectangle with the class intervals as base and frequency density as height.

$$\text{Frequency density} = \frac{\text{Frequency of the class interval}}{\text{Class size of the interval}}$$

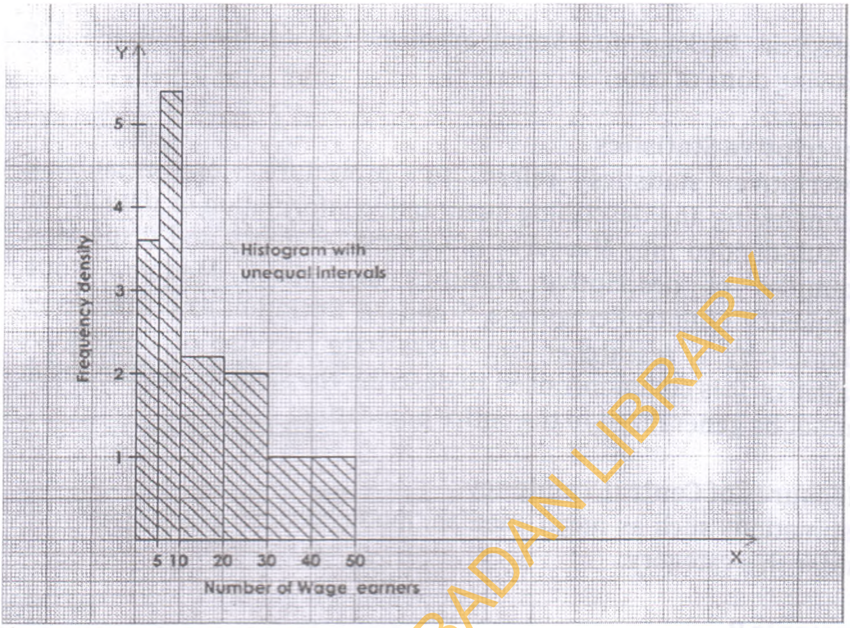

Fig 8: Unequal Width Histogram

Table 10: Comparison of Histogram and Bar Graph

Histogram	Bar graph
1. It consists of rectangles touching each other.	1. It consists of rectangles, normally separated from each other with equal space.
2. The frequency is represented by the area of each rectangle.	2. The frequency is represented by height. The width has no significance.
3. It is two dimensional (width and height are considered)	3. It is one dimensional (only height is considered)
	4. It is used a visual aid to represent data.

Note:

Manual approach to the drawing of the graphs is needful to make students acquire basic knowledge in data analysis. There are however, computer programme/software such as Microsoft excel, Statistical Package for Social Sciences (SPSS) that can get the drawing done at ease.

Descriptive Statistics

Descriptive statistics refers to procedure for organizing, summarizing and describing quantitative information or data. It can be sub-divided into Measures of Central Tendency and dispersion. Examples of descriptive statistics are Mean, Median, Mode, Frequency Counts, Percentage, Standard Deviation e.t.c.The design adopted by a researcher can necessitate the use of descriptive statistics. It is commonly used to describe the sample, but the result cannot be generalized to the entire population

Measures of Central Tendency

Estimations such as mean, median, mode, decile and percentile fall under this category.

Example 8

Table 11 presents the distribution of annual income in millions. The information in the table was used to estimate different measures.

Table 11: Frequency Table

Income (₦ million)	No of worker (f)	Class mark (x)	Class Boundaries	fx	$X-\bar{X}$	$(X-\bar{X})^2$	$f(X-\bar{X})^2$
9.3-9.7	2	9.5	9.25-9.75	19	-1.592	2.534464	5.068928
9.8-10.2	5	10	9.75-10.25	50	-1.092	1.192464	5.96232
10.3-10.7	12	10.5	10.25-10.75	126	-0.592	0.350464	4.205568
10.8-11.2	17	11	10.75-11.25	187	-0.092	0.008464	0.143888
11.3-11.7	14	11.5	11.25-11.75	161	0.408	0.166464	2.330496
11.8-12.2	6	12	11.75-12.25	72	0.908	0.824464	4.946784
12.3-12.7	3	12.5	12.25-12.75	37.5	1.408	1.982464	5.947392
12.8-13.2	1	13	12.75-13.25	13	1.908	3.640464	3.640464
Total	**60**			**665.5**			**32.24584**

$\bar{X}$= the mean of the distribution.

Mean

Mean usually denoted by $\bar{X}$

Mean $= \bar{X} = \frac{\sum fx}{\sum f}$ or $\frac{\sum fx}{N}$

Where: X is the class mark

f is the frequency

N is total number of cases.

$\therefore \bar{X} = \frac{665.5}{60} = 11.0917$

≈ 11.09

Median

Median simply refers to the mark at the middle for grouped data.

$$\text{Median} = L + \left(\frac{\frac{N}{2} - Cf_b}{f_w}\right)C$$

Where: L is the lower limit of the median group.

N is the total number of cases

Cf_b is the cumulative frequency before the median group.

f_w is the frequency within the median group

C is the class size.

Hence

$$\text{Median} = 10.75 + \left(\frac{30 - 19}{17}\right) 0.5$$
$$= 10.75 + \left(\frac{11}{17}\right)0.5$$
$$= 10.75 + 0.3235$$
$$= 11.07$$

Mode

Mode refers to the most occurred cases in a distribution.

$$\text{Mode} = L + \left(\frac{D_x}{D_x + D_y}\right)C$$

Where L = lower limit of the modal group.

D_x = difference in frequency between the modal group and the group before it

D_y = difference in frequency between the modal group and the group after it.

C = class size

Hence,

$$\text{Mode} = 10.75 + \left(\frac{5}{5+3}\right)0.5$$

$$= 10.75 + \left(\frac{5}{8}\right)0.5$$

$$= 11.0625$$

$$\approx 11.06$$

Quartiles

Quartiles refer to values which divide a distribution into four equal parts. Q_1, Q_2, and Q_3 are referred to as first, second and third quartiles respectively.

Hence,

$$\text{First Quartile} = Q_1 = L + \left[\frac{N/4 - Cf_bC}{f_w}\right]$$

$$= 10.25 + \left(\frac{15-7}{12}\right)0.5$$

$$= 10.25 + 0.3333 = 10.5833$$

$$\approx 10.58$$

Third Quartile = $Q_3 = L + \left[\frac{3N/4 - Cf_bC}{f_w}\right]$

$= 11.25 + \left(\frac{45-36}{14}\right)0.5$
$= 11.25 + \left(\frac{9}{14}\right)0.5$
$= 11.25 + 0.3214$
$= 11.5714$
≈ 11.57

Inter Quartile Range = $Q_3 - Q_1$
$= 11.57 - 10.58$
$= 0.99$

Semi Inter Quartile Range = $\frac{Q_3 - Q_1}{2}$

$= \frac{0.99}{2}$
$= 0.495$
≈ 0.5

Decile
Deciles refer to values which divide a distribution into ten equal parts .

Hence;

Seventh Decile = $D_7 = L + \left[\frac{7N/10 - cf_bC}{f_w}\right]$

$= 11.25 + \left(\frac{42-36}{14}\right)0.5$
$= 11.25 + \left(\frac{6}{14}\right) 0.5$
$= 11.25 + 0.2143$
$= 11.4643$
≈ 11.46

Percentile
Percentiles refer to values which divide a distribution into four equal parts.

Hence;

Sixty Seventh Percentile = P_{67}

$$\therefore P_{67} = L + \left[\frac{67N/100 - cf_b}{f_w} C\right]$$

$= 11.25 + \left(\frac{40.2-36}{14}\right)0.5$
$= 11.25 + \left(\frac{4.2}{14}\right)0.5$
$= 11.25 + 0.15$
$= 11.4$

Measure of Dispersion
This aspect is concerned with how scores are dispersed from the mean. Examples are range, variance and standard deviation.

Range equals the difference between the highest and the lowest scores in a distribution.
Hence;
Range from above distribution = 13.2 – 9.3 = 3.9

Variance and Standard Deviation.
Variance and standard deviation give the picture of how far apart are the scores. The square root of the estimated variance gives the standard deviation of the distribution.

Hence;

$$\text{Variance} = \frac{\sum f(\bar{X} - X)}{N}$$

$= 32.24584$
$= 0.5374$
≈ 0.54

Standard Deviation = SD = $\sqrt{variance}$

$$\therefore SD = \sqrt{0.5374}$$
$$= 0.7331$$
$$\approx 0.73$$

Inferential Statistics

Inferential Statistics concerns the methods by which induction is made to a larger group (population) on the basis of observations made on a smaller subgroup (Sample).

Table 12: Inferential Statistics

Parametrics	Non - Parametrics
(I) t – test (dependent group)	(I) Wilcoxon Test for two corrected samples
(II) t – test (independent group)	(II) Mann whitney U – test
(III) ANOVA (One way)	(III) Kruskal – Wallis test for K indep. Sample.
(IV) ANOVA (Two way)	(IV) Fres man's test
(V) ANCOVA	(V) Spearman Rank order correlation Coefficient
(VI) Pearson product moment correlation Coefficient.	(VI) Chi - Square
(VII) Multiple regression	

The commonly used among these statistical tools shall be discussed in this section

Correlation

Correlation is a bivariate measure of association (strength) of the relationship between two variables. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship). It is better reported in terms of its square (r^2), interpreted as percent of variance explained. For instance, if r^2 is .34, then the independent variable is said to explain 34% of the variance in the dependent variable.

According to Garson, (2008), there are several common pitfalls in using correlation. Correlation is symmetrical, not providing evidence of which way causation flows. If other variables also cause the dependent variable, then any covariance they share with the given independent variable in a correlation may be falsely attributed to that independent. Also, to the extent that there is a nonlinear relationship between the two variables being correlated, correlation will understate the relationship. Correlation will also be attenuated to the extent there is measurement error, including use of sub-interval data or artificial truncation of the range of the data. Correlation can also be a misleading average if the relationship varies depending on the value of the independent variable ("lack of homoscedasticity"). And, of course, a theoretical or post-hoc running of many correlations runs the risk that 5% of the coefficients may be found significant by chance alone.

Beside Pearsonian correlation (r), the most common type, there are other special types of correlation to handle the special characteristics of such types of variables as dichotomies, and there are other measures of association for nominal and ordinal variables. Regression procedures produce multiple correlation, R, which is the correlation of multiple independent variables with a single dependent. Also, there is partial correlation, which is the correlation of one variable with another, controlling both the given variable and the dependent for a third or additional variables. And there is part correlation, which is the correlation of one variable with another, controlling only the given variable for a third or additional variables.

Pearson's r: This is the usual measure of correlation, sometimes called *product-moment correlation*. The most common correlation for use with two continuous variables. Pearson's r is a measure of association which varies from -1 to +1, with 0 indicating no relationship (random pairing of values) and 1 indicating perfect relationship, taking the form, "The more the x, the more the y, and vice versa." A value of -1 is a perfect negative relationship, taking the form "The more the x, the less the y, and vice versa."

Table 13: Sample Data

Physics Score	Mathematics Score
73	54
64	47
42	34
68	67
77	89
56	66

Note: In older works, predating the prevalence of computers, special computation formulas were used for computation of correlation by hand. For certain types of variables, notably dichotomies, there were computational formulas which differed one from another (ex., phi coefficient for two dichotomies, point-biserial correlation for an interval with a dichotomy). Today, however, SPSS will calculate the exact correlation regardless of whether the variables are continuous or dichotomous.

Ordinal correlation

Correlation for ordinal and dichotmous data: Variations of correlation have been devised for binary and ordinal data. Some studies suggest use of these variant forms of correlation rarely affects substantive research conclusions. Rank correlation is nonparametric and does not assume normal distribution. It is less sensitive to outliers. Examples of correlation for ordinal and dichotomous data are Spearman's rho and Kendall's tau.

Spearman's rho: The most common correlation for use with two ordinal variables or an ordinal and an interval variable. Rho for ranked data equals Pearson's r for ranked data where d is the difference in ranks.

Table 14a: Sample Data

Academic Qualification	Knowledge Adequacy
1	2
2	3
2	3
3	2
3	1
1	1
1	2
3	1

or

Table 14b

Academic Qualification	Monthly Income (₦)
1	60,000
2	45,000
2	52,000
3	70.000
3	43,000
1	34,000
1	40,000
3	66,000

Academic Qualification:
Below First Degree= 1
First Degree = 2
Above First Degree = 3

Knowledge Adequacy:
Not Adequate= 1
Adequate = 2
Very Adequate = 3

Kendall's tau: Another common correlation for use with two ordinal variables or an ordinal and an interval variable. Prior to computers, rho was preferred to tau due to computational ease. Now that computers have rendered calculation trivial, tau is generally preferred. Partial Kendall's tau is also available as an ordinal analog to partial Pearsonian correlation

Correlation for dichotomies

Point-biserial correlation is used when correlating a continuous variable with a true dichotomy. It is a special case of Pearsonian correlation and Pearson's r equals point-biserial correlation when one variable is continuous and the other is a dichotomy. Thus, in one sense it is true that a dichotomous or dummy variable can be used "like a continuous variable" in ordinary Pearsonian correlation. (Special formulas for point-biserial correlation in textbooks are for hand computation; point-biserial correlation is the same as Pearsonian correlation when applied to a dichotomy and a continuous variable).

Table 15: Sample Data

Sex	Mathematics Score
1	54
1	47
2	34
1	67
2	89
2	66

Sex:
Male= 1
Female = 2

However, when the continuous variable is ordered perfectly from low to high, then even when the dichotomy is also ordered as perfectly as possible to match low to high, r will be less than 1.0 and therefore resulting r's must be interpreted accordingly. Specifically, point-biserial correlation will have a maximum of 1.0 only for the datasets with only two cases, and will have a maximum correlation around .85 even for large datasets, when the independent is normally distributed. The value of r may approach 1.0 when the continuous variable is bimodal and the dichotomy is a 50/50 split.

Rank biserial correlation: Used when an ordinal variable is correlated with a dichotomous variable.

Table 16: Sample Data

Sex	Academic Qualification
1	1
1	3
2	3
1	2
2	1
2	2
1	2
1	1

Sex:
Male= 1
Female = 2

Academic Qualification:
Below First Degree= 1
First Degree = 2
Above First Degree = 3

Phi: Used when both variables are dichotomies. Special formulas in textbooks are for hand computation; phi is the same as Pearsonian correlation for two dichotomies in SPSS correlation output, which uses exact algorithms.

Table 17: Sample Data

Sex	Meeting Attendance
1	2
1	1
2	1
1	2
2	1
2	2
1	2

Sex:
Male= 1
Female = 2

Meeting Attendance:
Attended= 1
Not Attended = 2

Coefficient of determination, r^2: The coefficient of determination is the square of the Pearsonian correlation coefficient. It represents the percent of the variance in the dependent variable explained by the independent. Of course, since correlation is bidirectional, r^2 is also the percent of the independent accounted for by the dependent. That is, the researcher must posit the direction of causation, if any, based on considerations external to correlation, which, in itself, cannot demonstrate causality. It is not sufficient for a researcher to make conclusion based on the value of **r** but take a step further to estimate the coefficient of determination ($\mathbf{r^2}$).

Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is used to establish the effect and interaction effects of categorical independent variables (called "factors" or "Grouping variable") on an interval dependent variable.For example the effect of Treatment (Independent variable) on student achievement in Mathematics (Dependent Variable).A *"main effect" is the direct effect of an independent variable* on the dependent variable. An "interaction effect" is the joint effect of two or more independent variables on the dependent variable.

The key statistic in ANOVA is the F-test of difference of group means, testing if the means of the groups formed by values of the independent variable (or combinations of values for multiple independent variables) are different enough not to have occurred by chance. If the group means do not differ significantly then it is inferred that the independent variable(s) did not have an effect on the dependent variable. Based on the level of significance set (α = .01 or .05) if significant difference of group means is observed then independent variable(s) has/have effect on the dependent variable.

Note that analysis of variance tests the null hypotheses that group means do not differ. It is not a test of differences in variances, but rather assumes relative homogeneity of variances. Thus some key ANOVA assumptions are that the groups formed by the independent variable(s) are relatively equal in size and have similar variances on the dependent variable ("homogeneity of variances"). Like regression, ANOVA is a parametric procedure which assumes multivariate normality (the dependent has a normal distribution for each value category of the independent(s)).

Table 18: Sample data

Sex (IV)	Treatment (IV)	Mathematics Score (DV)
1	2	54
1	1	47
2	1	34
1	2	67
2	1	89
2	2	66
1	2	72

Sex:
Male= 1
Female = 2

treatment:
Computer assisted= 1
Lecture method = 2

IV= independent variable; DV= Dependent variable

Analysis of covariance (ANCOVA)

Analysis of covariance (ANCOVA) is used to test the main and interaction effects of categorical variables on a continuous dependent variable, controlling for the effects of selected other continuous variables which covary with the dependent. The control variable is called the "covariate." There may be more than one covariate. One may also perform planned comparison or post hoc comparisons to see which values of a factor contribute most to the explanation of the dependent.

Table 19: Sample data

Sex (IV)	Treat-ment (IV)	Pretest score in Maths (Covariate)	Posttest score in Maths (IV)
1	2		54
1	1		47
2	1		34
1	2		67
2	1		89
2	2		66
1	2		72

IV= independent variable; DV= Dependent variable

Sex:
Male= 1
Female = 2

Treatment:
Computer assisted= 1
Lecture method = 2

Data requirements

For both ANOVA and ANCOVA, the dependent(s) is/are numeric. The independents may be categorical factors (including both numeric and string types) or quantitative covariates. Data are assumed to come from a random sample for purposes of significance testing. The variance(s) of the dependent variable(s) is/are assumed to be the same for each cell formed by categories of the factor(s) (this is the homogeneity of variances assumption).

One-way ANOVA tests differences in a single interval dependent variable among two, three, or more groups formed by the categories of a single categorical independent variable. Also known as univariate ANOVA, simple ANOVA, single classification ANOVA, or one-factor ANOVA, this design deals with one independent variable and one dependent variable. It tests whether the groups formed by the categories of the independent variable seem similar (specifically that they have the same pattern of dispersion as measured by comparing estimates of group variances). If the groups seem different, then it is concluded that the independent variable has an effect on the dependent measure.

Two-way ANOVA analyzes one interval dependent in terms of the categories (groups) formed by two independents, one of which may be conceived as a control variable. Two-way ANOVA tests whether the groups formed by the categories of the independent variables have similar centroids. Two-way ANOVA is less sensitive than one-way ANOVA to moderate violations of the assumption of homogeneity of variances across the groups.

Multivariate or n-way ANOVA. To generalize, n-way ANOVA deals with n independents. It should be noted that as the number of independents increases, the number of potential interactions proliferates. Two independents have a single first-order interaction (AB). Three independents have three first order interactions (AB,AC,BC) and one second-order interaction (ABC), or four in all.

Four independents have six first-order (AB,AC,AD,BC,BC,CD), three second-order (ABC, ACD, BCD), and one third-order (ABCD) interaction, or 10 in all. As the number of interactions increase, it becomes increasingly difficult to interpret the model. An example of SPSS output on 3-way ANCOVA is presented in table 20. Significant interaction effect of Treatment and Aptitude was observed.

Table 20: Tests of Between- Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig	Partial Eta Squared
Corrected Model	1092.134	12	91.011	10.050	.000	.537
Intercept	3042.486	1	3042.486	335.961	.000	.764
Pretest	5.214	1	5.214	.576	.450	.006
Treatment	557.305	1	557.305	61.539	.000	.372
Aptitude	42.506	2	21.253	2.347	.101	.043
Gender	41.346	1	41.346	4.566	.035	.042
Treatment* Aptitude	62.115	2	31.057	3.429	.035	.042
Treatment* Gender	1.912	1	1.912	.211	.647	.002
Aptitude* Gender	19.784	2	9.892	1.092	.339	.021
Treatment* Aptitude* Gender	4.534	2	2.267	.250	.779	.005
Error	941.831	104	9.056			
Total	9554.000	117				
Corrected Total	2033.966					

Independent t-Test

Independent t-test is similar in function to one-way ANOVA. It tests differences in a single interval dependent variable between only two groups. The key statistic is the t-test of difference of group means, testing if the means of the two groups formed by values of the independent variable are different enough not to have occurred by chance. If the group means do not differ significantly then it is inferred that the independent variable(s) did not have an effect on the dependent variable. Based on the level of significance set (α = .01 or .05) if significant difference of group means is observed then independent variable(s) has/have effect on the dependent variable. An example of a research question that can require the use of t-test is "will there be significant effect of sex on monthly income of University of Ibadan academic staff?"

Table 21: Sample data

Sex	Monthly Income (₦)
1	60,000
2	45,000
2	52,000
1	70.000
1	43,000
2	34,000
1	40,000
2	66,000

Sex:
Male= 1
Female = 2

Multiple regression,

Multiple regression is a statistical tool a researcher employs when he intends to account for (predict) the variance in an interval dependent, based on linear combinations of interval, dichotomous, or dummy independent variables. Multiple regression can establish that a set of independent variables explains a proportion of the variance in a dependent variable at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the independent variables (by comparing beta weights). One can test the significance of difference of two R^2's to determine if adding an independent variable to the model helps significantly. Using hierarchical regression, one can see how most variance in the dependent can be explained by one or a set of new independent variables, over and above that explained by an earlier set. Of course, the estimates (b coefficients and constant) can be used to construct a prediction equation and generate predicted scores on a variable for further analysis.

The multiple regression equation takes the form $y = b_1x_1 + b_2x_2 + ... + b_nx_n + c$. The b's are the regression coefficients, representing the amount the dependent variable y changes when the corresponding independent changes 1 unit. The c is the constant, where the regression line intercepts the y axis, representing the amount the dependent y will be when all the independent variables are 0. The standardized version of the b coefficients are the beta weights, and the ratio of the beta coefficients is the ratio of the relative predictive power of the independent variables. Associated with multiple regression is R^2, multiple correlation, which is the percent of variance in the dependent variable explained collectively by all of the independent variables.

Apart from linear multiple regression, other forms are:

Logistic regression: It is used for dichotomous and multinomial dependents, implemented here with logistic procedures and above in generalized linear models.

Weighted least squares (WLS) regression this may be used when the assumption of homoscedasticity has been violated.

Cox regression may be used to analyze time-to-event as well as proximity, and preference data.

function analysis is used when the dependent variable is a dichotomy but other assumptions of multiple regression can be met, making it more powerful than the alternative, which is logistic regression for binary or multinominal dependents.

Partial least squares regression may be used even with small datasets to predict a set of response variables from a set of independent variables.

Logit regression uses log-linear techniques to predict one or more categorical dependent variables.

Poisson regression for count data in event history analysis and more, implemented here with general loglinear procedures and above in generalized linear models.

Categorical regression is a variant which can handle nominal independent variables, but now largely replaced by generalized linear models..

Chi-Square Significance Tests
Chi square test is a non parametric statistics that can be used mainly to test association of non-interval variables and also to establish if an observed distribution conforms to any other distribution. Examples are:

Pearson's chi-square is by far the most common type of chi-square significance test. If simply "chi-square" is mentioned, it is probably Pearson's chi-square. This statistic is used to test the hypothesis of no association of columns and rows in tabular data. It can be used even with nominal data. Note that chi square is more likely to establish significance to the extent that (1) the relationship is strong, (2) the sample size is large, and/or (3) the number of values of the two associated variables is large. A chi-square probability of .05 or less is commonly interpreted by social scientists as justification for rejecting the null hypothesis that the row variable is unrelated (that is, only randomly related) to the column variable.

Table 22: Sample Data for running Pearson's chi-square Analysis

Academic Qualification	Knowledge Adequacy
1	2
2	3
2	3
3	2
3	1
1	1
1	2
3	1

Academic Qualification:
Below First Degree= 1
First Degree = 2
Above First Degree = 3

Knowledge Adequacy:
Not Adequate= 1
Adequate = 2
Very Adequate = 3

For this example, the SPSS output of the Analysis, Descriptive Statistics, Crosstabs menu choice look like this:

Table 23: **Level of Aptitude * Gender Crosstabulation**

Count

		Gender		Total
		Male	Female	
Level of Aptitude	Low Aptitude	33	8	41
	Average Aptitude	47	52	99
	High Aptitude	22	18	40
Total		102	78	180

Table 24: **Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.926[a]	2	.002
Likelihood Ratio	13.809	2	.001
Linear-by-Linear Association	5.443	1	.020
N of Valid Cases	180		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 17.33.

Chi-square goodness-of-fit test The goodness-of-fit test is simply a different use of Pearsonian chi-square. It is used to test if an observed distribution conforms to any other distribution, such as one based on theory (e.g., if the observed distribution is not significantly different from a normal distribution) or one based on some other known distribution (e.g., if the observed distribution is not significantly different from a known national distribution based on Census data). The Kolmogorov-Smirnov goodness-of-fit test is preferred for interval data, for which it is more powerful than chi-square goodness-of-fit.

Likelihood ratio chi-square test, also called the likelihood test or G test, is an alternative procedure to test the hypothesis of no association of columns and rows in nominal-level tabular data. It is supported by SPSS output and is based on maximum likelihood estimation. Though computed differently, likelihood ratio chi-square is interpreted the same way. For large samples, likelihood ratio chi-square will be close in results to Pearson chi-square. Even for smaller samples, it rarely leads to different substantive results. SPSS will print likelihood ratio chi-square in the "Chi-Square Tests" table of output from the Analyze, Descriptives, Crosstabs menu selection.

References

Box, G.E. P. , Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design and data analysis.* NY: John Wiley. General introduction.

Brown, Steven R. & Lawrence E. Melamed (1990). *Experimental design and analysis.* Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series no. 74. Discusses alternative ANOVA designs and related coefficients.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* NY: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Second ed., Hillsdale, NJ: Erlbaum.

Dunteman, George H. & Moon-Ho R. Ho (2005).*An introduction to generalized linear models.* Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences, Vol. 145.

Garson, G.D. (2008) Online Text and Notes in Statistics for Economists. www.economicsnetwork.ac.uk/teaching

Girden, Ellen R. (1992). *ANOVA Repeated Measures.* Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series no. 84.

Iverson, G.R. and Helmut, N. (1987).*Analysis of variance.* Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series no. 1.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance.* Quantitative Applications in the Social Sciences Series No. 118. Thousand Oaks, CA: Sage Publications.

Levin, I. P. (1999).*Relating statistics and experimental design.* Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series.

McCall, R.B. (1990) Fundamental Statistics for Psychology, Harcourt Brace Jovanovich. Inc.

ABOUT THE BOOK

There has been a great demand for the development of indigenous books that would assist and meet the needs of students, teachers and researchers in the area of project writing and supervision both at the Undergraduate and Postgraduate levels in tertiary institutions. This book has been written to meet the need, taking into consideration that there are some challenges that individuals face while conducting researches, writing dissertations and thesis(es) at various education levels. This book titled "Research in Education" has thirteen chapters and provides contemporary, simple and trouble-free guides for readers to enable them write acceptable projects, dissertations and theses. Scholars that are enriched with various research skills from different departments contributed to the book.

The book covers topics such as: meaning and purpose of research in education; criteria for the identification and diagnosis of a viable research problem; research questions and hypotheses; types, uses of research variables and definition of terms. Other topics covered in this book are: literature review; theoretical and conceptual framework in research; research design and procedure; instrumentation; issues of validity and reliability. In addition, other issues such as: types and methods of data analysis; techniques of writing abstract; methods of referencing in research; vetting and editing in writing and finally, preparation and presentation of research reports were also addressed in the book.

Research in Education is therefore bequeathed to all researchers, students, various practitioners not only in education, but also in other disciplines requiring application of research.

ISBN: 978-978-55702-8-1

Printed by:

HIS LINEAGE PUBLISHING HOUSE
Mobile: +234 08033596818