British Journal of Education, Society & Behavioural Science
14(2): 1-10, 2016, Article no.BJESBS.23003
ISSN: 2278-0998

SCIENCEDOMAIN *international*
www.sciencedomain.org

Statistical Approaches and Decision Making Towards Bivariate and Multivariate Analyses with Visirule

Ismail O. Muraina[1*], Mukaila A. Rahman[2] and Imran A. Adeleke[1]

[1]School of Science, Adeniran Ogunsanya College of Education, Otto/Ijanikin, Nigeria.
[2]Faculty of Science, Lagos State University, Nigeria.

Authors' contributions

This work was carried out in collaboration between all authors. Author IOM designed the study, sketched summary graphs for both bivariate and multivariate statistics and applied visirule in making decisions. Author MAR also involved in applying visirule, generated visirule screenshots, wrote abstract, conclusion and beneficiaries. Author IAA collated relevant literature, wrote the first draft of the manuscript and tested the visirule applications to ensure its correctness. All authors involved in searching for relevant literature, editing the manuscript as well as reading and approving the final manuscript.

Article Information

DOI: 10.9734/BJESBS/2016/23003
Editor(s):
(1) Stan Weeber, Professor of Sociology, McNeese State University in Lake Charles, Louisiana, USA.
Reviewers:
(1) Baidyanath Pal, Indian Statistical Institute, India.
(2) Derya Sevim Korkut, Duzce University, Duzce, Turkey.
(3) Tsung-Yu Hsieh, MingDao University, Taiwan.
Complete Peer review History: http://sciencedomain.org/review-history/12982

Original Research Article

Received 9th November 2015
Accepted 4th January 2016
Published 15th January 2016

ABSTRACT

Without any doubt, research is sometimes designed to solve a particular problem. In social sciences and education, data collected during research studies are often subjected to statistical analysis for deducing desirable results. Reported research that involved data collection without proper analysis to back these results often makes the work non-robust enough for wider acceptance in the research community. One of the greatest challenges often faced by amateur researchers and students is how to select appropriate statistical tools which are suitable and relevant to the data at hand. Having a firm knowledge of suitable statistical tool goes a long way in motivating and enhancing researcher's attitudes towards the work. This paper presents procedures necessary for researchers to select appropriate statistical instruments for a desired purpose. The

**Corresponding author: E-mail: niyi2all@yahoo.com;*

Visirule software, being free software, was used as a decision enabling tool, presented graphically based on a logic programming model. In research, knowledge of bivariate and multivariate analyses would assist researchers – both undergraduate and postgraduate – to make viable decision before a particular statistical data analysis is being used.

Keywords: Statistics; visirule; bivariate; multivariate; decision support tool.

1. INTRODUCTION AND BACKGROUND

Research studying and writing is a sine-qua-non to all students in higher institutions to embark upon at last stage/level of their schooling. There are courses and topics that equip students to understand all basic requirements in research and statistical analysis yet many students still find it confusing while other see it as a threat to their graduation. Reports have shown that students' writing projects move at a fast rate from chapter one to chapter three before many of them boycott or abandon the project totally or continue sporadically. To some students, taking clues after old projects is an alternative. These and many more reasons interest researchers in this study. Reading meaning to already collected data has remained a serious drawback to students' writing projects. In this study, the researchers take a further step to look into proper ways of selecting most relevant statistical tool to analyze a particular data at hand: focus is set towards bivariate and multivariate analysis/statistics using visirule as the decision tree software.

Research simply means a search for knowledge. It can also be seen as a scientific and systematic search for pertinent information on a specific topic. [1,2] define research as a systematized effort to gain new knowledge while some people consider research as a movement, a movement from the known to the unknown. In fact, it is a voyage of discovery. It is a careful, scholarly, purposeful and systematic investigation or study that is designed to add or delete from a body of knowledge [3]. To [4,5], research is the orderly investigation of a subject matter for the purpose of adding to knowledge. Research can mean "re-search" implying that the subject matter is already known but, for one reason or another, need to be studied again. Alternatively, the expression can be used without a hyphen and in this case it typically means investigation of a new problem or phenomenon. Research in an academic activity and as such the term should be used in a technical sense. According to [6] research comprises defining and redefining focused problems, coming up with hypotheses; collecting, collating and putting judgment on data; making deductions and reaching final conclusion; carefully and finally testing the conclusions to know whether those data fit the formulated hypothesis. In short, the search for knowledge through objective and systematic method of finding solution to a problem is research.

Statistics is a part of the everyday language of education and the other social sciences [7,8]. These research based disciplines routinely use statistics to express knowledge and to discuss theory and research. Students as well as educationists should be literate in the vocabulary of research, data analysis and scientific thinking. Fluency in statistics will go a long way to help understanding the research reports one encounters in everyday life. This will also help one to be able to conduct quantitative research, contribute to the growing body of education and reach full potential as an educational scientist. Although, essential learning statistics can be challenging. Students in writing projects should be equipped with knowledge necessary to conduct a successful statistical analysis in research.

In research, the kind of variable available determines the type of approach to the analysis. A variable according to [9] is any characteristics, in form of numbers or quantities that can be counted or measured. A variable can be regarded as a data item. Examples such as sex, age, marital status, date of birth, house type, height and class can be called variables because they may change as data units change or change in value as time passes. Variables have types: Numeric variables – Which can be a continuous variable (Observations can have values between series of real numbers) or a discrete variable (Observations can have values from series of distinct whole values), categorical variables. Numeric variables explain a measurably quantified number, in form of "how". Hence, numeric variables are quantified while categorical variables can be nominal (not be able to organize in a logical sequence) or ordinal (that can be ranked or ordered). Therefore, categorical variables are to be represented by a non-numeric value.

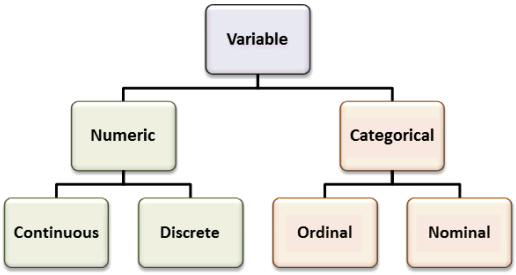

Fig. 1. Summary of variable types under numeric and categorical

Similarly, variables can be classified into dependent and independent variables. When conducting research, according to [10] researchers do manipulate variables. For instance, a researcher may change independent variable(s) to dependent variable(s). If this is done, it would have a great effect on the result or output expected from such research. Another way of looking at types of variables is qualitative variables and quantitative variables. Quantitative variables are expressed in numerical ordering such as height, weight and scores in a course while qualitative variables are those expressed by qualifying or explaining the attributes of a variable such as gender, religion and hair colour. After researchers look into the type of variables on ground it is pertinent to consider the type of statistics available for the type of variables. In research, the two major common classifications of statistics are parametric statistics and non-parametric statistics. In [11] it was clearly stated that the term parametric statistical test assumes the parameters of the population distribution(s) where the data are drawn. Examples of parametric tests include:

- ✓ All forms of T-tests
- ✓ Pearson product Moment Correlation Coefficient Test
- ✓ All forms of ANOVA Test
- ✓ All forms of ANCOVA Test
- ✓ MANOVA Test

A non-parametric statistical test makes no assumption like parametric statistical test. This is a null category which assumes one thing or another about the properties of such population(s). Examples of non-parametric tests include:

- ✓ The various forms of chi-square tests,
- ✓ The Fisher Exact Probability test,
- ✓ The Mann-Whitney Test,
- ✓ The Wilcoxon Signed-Rank Test,
- ✓ The Kruskal-Wallis Test,
- ✓ And the Friedman Test.

[12] observed that one should use nonparametric tests when data don't meet the assumptions of the parametric test, especially the assumption about normally distributed data. This assumption may not be enough to use any test; it is also necessary to note that parametric analysis is used to test group means while non-parametric analysis is to test group medians. In research, a statistical hypothesis can be setting assumption of a particular population parameter. In this, the assumption may either be true or false depending on the inference made at the end. Hypothesis testing means the formal stepwise procedures required by researcher to accept or reject statistical hypotheses [13]. Hypothesis is used as a statistical test to find out whether there is ample evidence in a sample data to ascertain that a condition is true for the entire population. There are two common types of hypothesis tests which one is opposing the other: Null hypothesis and alternative hypothesis. Usually, the null hypothesis has statements of "no effect", "no difference", "no significance" while alternative hypothesis is the statement when the conclusion is true. P-value stands as a determination of a null hypothesis being rejected or not. If P-value is less or equal to the level of significance, then researcher can reject such hypothesis.

2. CONDITION FOR THE USE PARAMETRIC AND NON-PARAMETRIC STATISTICS

2.1 Parametric Statistics

The parametric statistics is used when

- The dependent variable is having interval or ratio scale, this means having a continuous scale instead of discrete scales.
- The samples should be selected randomly from the population
- The observations that make up data should be independent type. No measurement or observation should influence other measurement or observation.

Examples of parametric statistics are: T-TEST, ANOVA, ANCOVA, MANOVA, etc

2.2 Non-Parametric Statistics

The non-parametric statistics is used:

- When measured on nominal (categorical) and ordinal (ranked) scale
- When they have very small samples
- When data do not meet assumptions of parametric techniques

Examples of non-parametric statistics are: Chi-square, Mann-Whiteney V test, Spearman Rank-order Correlation, etc.

The Table 1 below shows differences between parametric and non-parametric statistics according to [14].

Table 1. Lists of parametric and non-parametric statistics

Non-parametric	Parametric
Chi-square	No equivalent test
Mann-Whitney V test	Independent-sample t-test
Wilcoxon signed rank test	Paired Sample t-test
Kruskal-WalliS test	One way between group ANOVA
Friedman test	One-way repeated measures ANOVA
Spearman rank order correlation	Pearson`s product moment correlation

2.3 Correlation Techniques- Non-Experimental Design

Correlation techniques are used by researcher engaged in non-experimental designs WHERE predictor should correlate highly with dependent variable not other independent variables. Different kinds of correlation are as follow:

(1) Finding the association between pairs of variables – Correlation
(2) Predicting scores on one variable from scores on another variable – Bivariate regression
(3) Predicting scores on a dependent variable from one score of a number of independent variables – Multiple regression
(4) Identifying the structure underlying a group of related variables – Factor analysis

All listed above are usually used to put models and theories into test as well as predicting outcomes.

2.4 Correlation Techniques- Experimental Design

2.4.1 Correlation

- To describe the strength and direction of the relationship between two variables
- Two continuous variables or one continuous and dichotomous variables (e.g two values, sex: male and female)
- Pearson Product-Moment co-efficient
- Non-parametric – spear-mans rank order correlation

2.4.2 Partial correlation

- To explore the relationship between two variables by controlling statistically the third variable. This can be done when a researcher is suspecting that existed relationship within two variables at hand may be influenced or confounded by the impact of the third variable
- Three variables: all continuous, two variables to explore the relationship and one variable to control

2.4.3 Multiple regression

- To predict a single dependent continuous variable from a group of independent variables. It can be used to test the predictive power of set of variables and to access the relative contribution of each individual variable.
- One continuous dependent variable and two or more independent variables or one dichotomous independent variables e.g (sex: male & female)

2.4.4 Logistic regression

- This allows one to predict categorical result with two or more categories. Researcher predictor (Independent) can either be categorical or continuous or combination of the two in one model
- One categorical (dichotomous) dependent variable (problem with sleep) and two or more continuous or categorical predictor (independent) variables (we must code dichotomous variables as 0 and 1) (e.g sex, trouble getting to sleep, trouble staying asleep, coded as 0 and 1)

2.4.5 Factor analysis

- This is used when larger number of related variables is to be reduced into a smaller, more manageable, number of dimensions or components.

2.5 Comparing Group – Experimental Design

2.5.1 Independent-sample test

- It is used when you want to compare the mean score of two different groups of people or conditions
- One categorical independent variable (males / females or experimental /control) and one continuous dependent variable (achievement score)

2.5.2 Paired Sample T-Test (Repeated Measure)

- It is useful when researcher has one group of respondents (teachers, students, drivers etc) that are to be tested in two occasions or under two different conditions (pretest / posttest)
- One categorical independent variable (time 1 and time 2) and one continuous dependent variable (performance scores measured on two different occasions or under different conditions)

2.5.3 One-Way ANOVA (Between-group)

- When we are interested in comparing the mean scores of more than two groups. That is, when you have different subjects or cases in each of your groups (independent group)
- One categorical independent variable of three or more distinct levels (this can also be a continuous variable that has been recoded to give three equal groups e.g age group: 20-24 age, 25-29 age and 30-44 age.) and one continuous dependent variables (performance test scores of students)

2.5.4 One-way repeated measures ANOVA

- It is the type of ANOVA that measures each subject is exposed to two or more different conditions or measure on the same continuous scale more than one occasions.
- One independent variable (categorical: e.g. time1/ time2/ time3) and one dependent variable (continuous: performance test score)

2.5.5 Two ways between groups ANOVA

- This means that there are two independent variables and between-groups which indicate that different respondents are in each of the groups. The statistical tools allow us to look at the individual and joint effect of two independent variables on one dependent variable
- Researcher must have two categorical independent variables (sex: Males/Females, Age: young, middle, old) and one continuous dependent variable (achievement scores in a test)

2.5.6 Mixed between-within subject ANOVA

- A researcher should have two independent variables: one is a between-subject variable (gender: males/females), the other should be a within-subjects variable (time). In this situation, researcher would expose a group of both males and females to the treatment at time1(pre-test) and time2(post-test)
- Researcher also need one categorical independent between-subjects variable with two or more levels (group1/group2)
- Finally, researcher need one categorical independent within-subjects variable with two or more levels (time1/ time2/ time3) and one continuous dependent variable (scores of students in computer science test measure at each time period)

2.5.7 Multivariate ANOVA

- MANOVA is an extension of ANOVA which is useful when a researcher has more than one dependent variable. The dependent variables must be related in some way or having conceptual reason for using them together
- One categorical independent variable (e.g. sex) and two or more continuous dependent variables. MANOVA can also be extended to two-way and higher-order designs involving two or more categorical independent variables.

2.5.8 One way ANCOVA

- This is a type of statistical tool that researcher use when he/she has a two-group pretest/post-test design (e.g. comparing the impact of two different interventions, taking before and after measures for each group). The scores on the pretest should be treated as a covariate to `control` for pre-existing differences between the groups
- This tool requires one categorical independent variable with two or more levels (group1/group2); also, one continuous dependent variable (score on computer science test at time 2) and finally, one or more continuous covariates (scores on mathematics test at time 1)

2.5.9 Two way ANCOVA

- ANCOVA is a statistical tool that a researcher needs to control scores on covariate and then perform a normal two-way ANOVA. This will tell the researcher whether there is a significant main effect of the first independent variable (e.g. group) on the main effect of the second independent variable (sex). There should be a great significant interaction between the two independent variables.
- Two categorical independent variables with two or more levels (sex Males/Females), group: Mathematics / Computer). One continuous dependent variables, one or more continuous covariates.

2.6 Comparing Group – Non Experimental Design

Chi-Square: The test is used when you wish to explore the relationship between two categorical variables.

- Two categorical variables, with two or more categories in each
- For 2 x 2 table use Fisher`s Exact Probability Test

2.6.1 Mann whitney V test

- This test is used to test for differences between two independent groups on a continuous measure. Mann Whitney V Test actually compares medians
- One categorical variable with two groups e.g. sex. One continuous variable.

2.6.2 Wilcoxon signed rank test

- Also called Wilcoxon Matched pairs signed rank test is designed for use with repeated measures; that is when your subjects are measured on two occasions or under two different conditions. It converts scores to ranks and compares them at time1 and time2
- One group of subjects measured on the same continuous scale or criterion on two different occasions

2.6.3 Kruskal-wallis test

- This test (also known as kruskal wallis test) allows you to compare the scores on some continuous variables for three or more groups. It converts the scores to ranks and the mean rank for each group is computed
- One categorical independent variable with three or more categories (age 18-29, 30-44, 45+). One continuous dependent variable

2.6.4 Frie Sman Test

- This test is used when you take the same sample of subjects or cases and you measure them at three or more points in time or under three different conditions
- One sample of subjects measured on the same scale or measured at their different time or periods or under three different condition

Spear Man`s Rank Order Correlation (rho) is used to calculate the strength of the relationship between two continuous variables

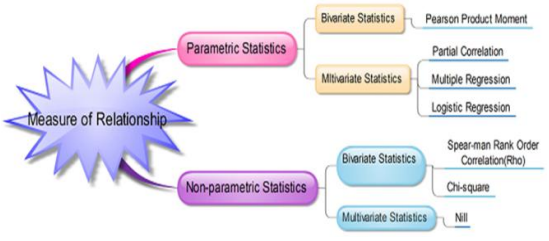

Fig. 2. Summary graph for bivariate statistics

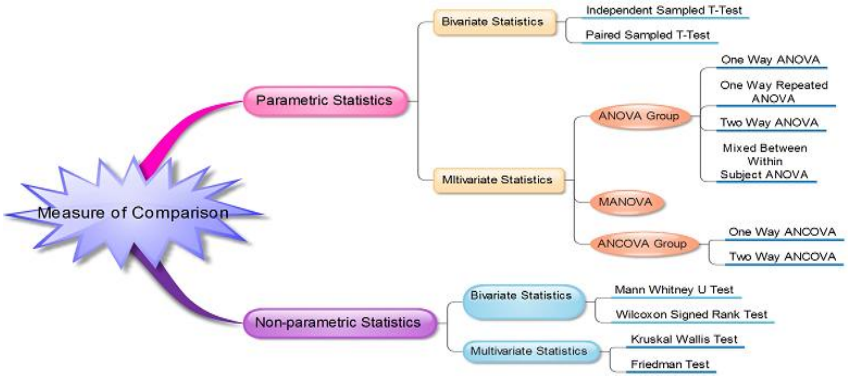

Fig. 3. Summary graph for multivariate statistics

3. METHODS

This paper presents procedures necessary for researchers to select suitable and relevant statistical instrument for a desired purpose. A free software (Visirule) was used as a support tool for making viable and logical decision in which the rules are basically and precisely presented graphically based on logic programming model. Graphical illustrations were also given which clearly depicted the relationships between the statistics discussed. Subsequently, Visirule software was used to implement the easy selection of those discussed tools.

In this structure, more than twenty-three (23) questions were raised relating to particular structural analysis relevant to data collected, which are depicted by yellow single choice boxes. Followed by forty-five (45) white boxes indicating options to be taken by the proceeding questions. Finally, there are twenty-nine (29) red boxes implying the result or outcome (conclusion) which usually terminate or end series of inferences [15,16]. Visirule was used by authors only covered bivariate and multivariate analysis.

The application of visirule will greatly assist the researchers, students and other interested people to just follow the questions based on the kind of data or variable they have. The software (Visirule) will generate the appropriate statistical tool that is best fit the kind of data or variable the researcher has. This software is chosen because it helps in making good decision without necessarily memorize any commands.

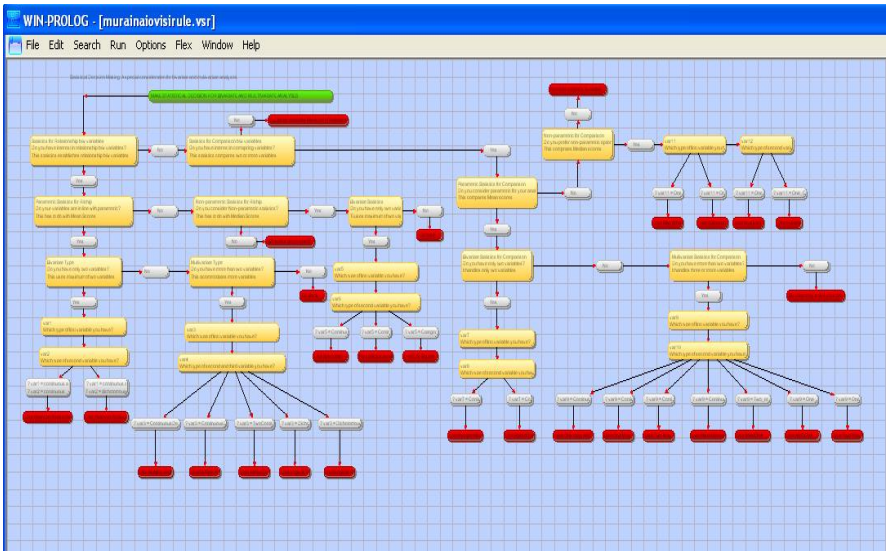

Fig. 4. Visirule application screenshots

4. RESULTS AND DISCUSSION

The bivariate and multivariate analysis presented in this work can be helpful to everyone especially young graduates and researchers in making logical, productive and fulfilling decision in analyzing and interpreting suitable statistical data in researches.

This study will be of an advantage to both lecturers and research students who at the end of their course would write project on their fields of study. Also, students do not need any special knowledge of programming before selecting appropriate statistical tool for analysis. It can also serve as a supportive material for teacher teaching research related courses.

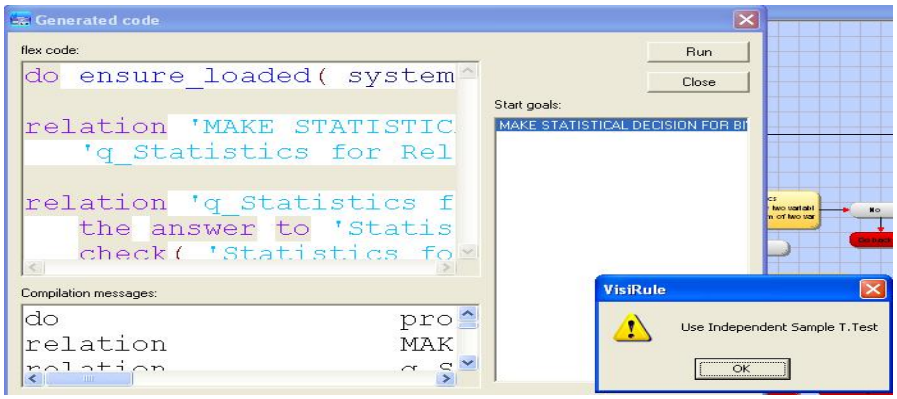

Fig. 5. Visirule output screenshots

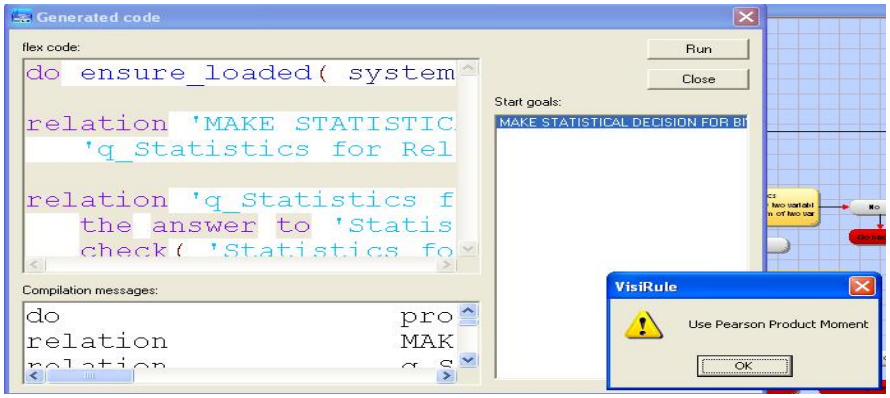

Fig. 6. Visirule output screenshots (cont.)

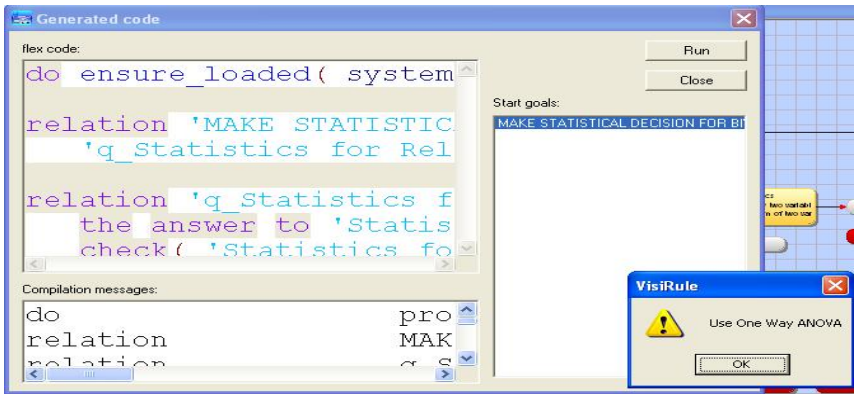

Fig. 7. Visirule output screenshots (cont.)

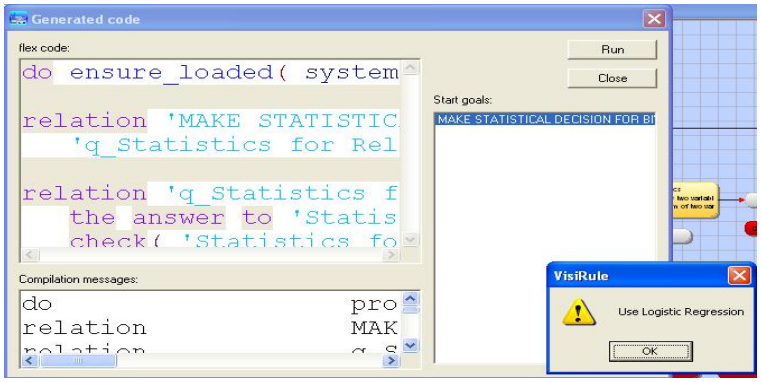

Fig. 8. Visirule output screenshots (cont.)

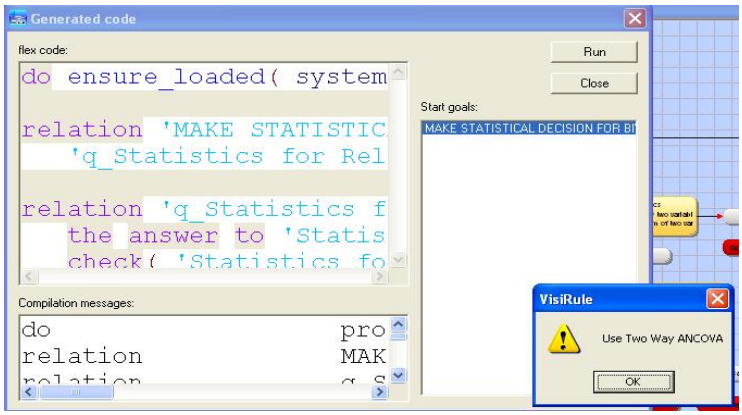

Fig. 9. Visirule output screenshots (cont.)

5. CONCLUSION AND BENEFICIARIES

This study is a step further of earlier research that was primarily on univariate and bivariate statistics [17], here this particular study emphasizes on how problem of selecting an appropriate statistical tool for data analysis would be resolved. At the same time, authors gave details of conditions required of using one statistical tool. Explanations were given to parametric and non-parametric, the statistics that are under parametric and non-parametric and finally the statistics were grouped into bivariate and multivariate.

Graphical illustrations were also given which clearly depicted the relationships between the statistics discussed. Subsequently, Visirule software was used to implement the easy selection of those discussed tools.

Hence, below are the beneficiaries:

1. This study will be of an advantage to both lecturers and research students who at the end of their course would write project on their fields of study.
2. Also, students with little or no knowledge of programming and research will find the study more useful.
3. Teacher at various parts of institutions would find it more helpful to teach students statistics.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Redman LV, Mory AVH. Research methodology: An introduction; 1923. Available:www.google/research/methodology (Retrieved on the 12-10-2014)
2. Muijs D. Doing qualitative research in Education with spss. New Delhi, SAGE Publications Ltd; 2006.
3. Akinade EA, Owolabi T. Research methods. A pragmatic approach for social

sciences, behavioral science education. Lagos, Connel Publications; 2009.
4. Ross RH. Motivating learners using graphical Visirule based system. International Journal of Information Studies. 2014;1(4).
5. Ejifugba AU. Fundamentals of research in health education Benin, Barloz Publishers Inc; 1998.
6. Woody C. Meaning of research; 2014. Available:www.google/research (Retrieved on the 12-10-2014)
7. Healey JF. The essential of statistics: A tool for social research; 2014. Available:www.askme.com (Retrieved on the 12-10-2014)
8. Giarratano JC. Principles and programming Fourth Indian Edition Thomson; 2007.
9. Australian Bureau of Statistics. What is a variable; 2015. Available:http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+what+are+variables (Retrieved on the 4 December, 2015)
10. Heidi Ziemer. Online statistics education: An interactive multimedia course of study. Developed by Rice University (Lead Developer), University of Houston Clear Lake, and Tufts University; 2015.
11. Jim Frost. Choosing between a Non-Parametric Test and a parametric test; 2015. Available:http://blog.minitab.com/blog/adventures-in-statistics/choosing-between-a-nonparametric-test-and-a-parametric-test (Retrieved on the 5 December, 2015)
12. Vassarstats_Online_Textbook. Parametric and non-parametric; 2015. Available:http://vassarstats.net/textbook/parametric.html (Retrieved on the 5, December 2015)
13. Stattrek. What is hypothesis testing? 2015. Available:http://stattrek.com/hypothesis-test/hypothesis-testing.aspx (Retrieved on the 6 December, 2015)
14. Donna R. Examine the differences between univariate and bivariate data; 2012. Available:http://regentsprep.org (Retrieved on the 5th May, 2012)
15. Alwan RH, Benrachi B. Motivating learners using graphical Visirule based system. International Journal of Information Studies. 2009;1(4).
16. Bilgi NB, kulkarni RV, Spenser C. An expert system using of decision logic charting approach for Indian legal domain with specific reference to transfer of property Aet. International Journal of Artificial Intelligence and Expert System (IJAE). 2010;1(2).
17. Muraina IO, Rahman MA, Adeleke IA, Aiyegbusi EA. Use of a rule tool in data analysis decision making. International Institute of Science, Technology and Education (IISTE) (Information and Knowledge Management). ISSN 2224-5758 (Paper) ISSN 2224-896X (Online). 2013;3(3):131-142.

© 2016 Muraina et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/12982