

# Statistical Decision Making for Bivariate and Multivariate Analysis with Visirule

## 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 instrument for a desired purpose. The Visirule software was used as a decision supporting tool, in which the rules are basically and precisely presented graphically based on logic programming model. The bivariate and multivariate analysis discussed in this work can be of great use to researchers and to both undergraduate and post-graduate students in making a productive and realizable decision in utilizing suitable statistical data analysis in researches.

**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 project 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 way 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 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 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 problems, formulating hypothesis or suggested solution; collecting, organizing and evaluating data; making deductions and reaching conclusions; and at last carefully testing the conclusions to determine whether they fit the formulating 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 every day 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.

## Condition for the use Parametric and Non-Parametric Statistics

### Parametric Statistics

The parametric statistics is used when

- The dependent variable is measured at the interval or ratio level, that is, using a continuous scale rather than discrete categories.
  - The samples should be selected randomly from the population
  - The observations that make up data must be independent of one another. That is, each observation or measurement must not be influenced by any other observation or measurement.
- Examples of parametric statistics are: T-TEST, ANOVA, ANCOVA, MANOVA, etc

### 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-Whitney V test, Spearman Rank order Correlation, etc

The table1 below shows differences between parametric and non-parametric statistics according to [9].

**Table1.** Lists of Parametric and Non-parametric statistics

| NON-PARAMETRIC                         | PARAMETRIC                           |
|----------------------------------------|--------------------------------------|
| <b>Chi-square</b>                      | No equivalent test                   |
| <b>Mann-Whitney V test</b>             | Independent-sample t-test            |
| <b>Wilcoxon Signed Rank Test</b>       | Paired Sample t-test                 |
| <b>Kruskal-Wallis Test</b>             | One way between group ANOVA          |
| <b>Friedman Test</b>                   | One-way repeated measures ANOVA      |
| <b>Spearman Rank order Correlation</b> | Pearson's product moment correlation |

### Correlation Techniques- Non-Experimental Design

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

- (1) Explore the association between pairs of variables – **Correlation**
- (2) Predict scores on one variable from scores on another variable – **Bivariate regression**
- (3) Predict scores on a dependant variable from one score of a number of independent variables –

### Multiple regression

- (4) Identify the structure underlying a group of related variables – **Factor analysis**

All these are used to test models and theories, predict outcomes and access reliability and validity of scales

### Correlation Techniques- Experimental Design

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

- 99       • Non-parametric – spear-mans rank other correlation
- 100       Partial Correlation
- 101       • To explore the relationship between two variables while statistically controlling for a third
- 102       variable. This is useful when one is suspecting that the relationship between two variables of
- 103       interest may be influenced or confounded by the impact of the third variable
- 104       • Three variables: all continuous, two variables to explore the relationship and one variable to
- 105       control
- 106       Multiple Regression
- 107       • To predict a single dependant continuous variable from a group of independent variables. It can
- 108       be used to test the predictive power of set of variables and to access the relative contribution of
- 109       each individual variable.
- 110       • One continuous dependant variable and two or more dependant variables or one dichotomous
- 111       independent variables e.g (sex: male & female)
- 112       Logistic Regression
- 113       • To allow one to test models to predict categorical outcome with two or more categories. Your
- 114       predictor (independent) variables can either be categorical or continuous or a mix of both in one
- 115       model
- 116       • One categorical (dichotomous) dependent variable (problem with sleep:) and two or more
- 117       continuous or categorical predictor (independent) variables (note: we must code dichotomous
- 118       variables as 0 and 1) (e.g sex, trouble getting to sleep, trouble staying asleep, coded as 0 and 1)
- 119       Factor Analysis
- 120       • This is useful in reducing a large number of related variables (items that make up a scale) to a
- 121       smaller, more manageable, number of dimensions or components
- 122       **Comparing Group – Experimental Design**
- 123       Independent-Sample Test
- 124       • It is used when you want to compare the mean score of two different groups of people or
- 125       conditions
- 126       • One categorical independent variable (males / females or experimental /control) and one
- 127       continuous dependent variable (achievement score)
- 128       Paired Sample T-Test (Repeated Measure)
- 129       • It is used when you have only one group of people (students, companies etc) and one collects
- 130       data from them on two different occasions or under two different conditions (pretest / posttest)
- 131       • One categorical independent variable (time 1 and time 2) and one continuous dependent
- 132       variable(performance scores measured on two different occasions or under different conditions)
- 133       One-Way ANOVA(Between-group)
- 134       • When we are interested in comparing the mean scores of more than two groups. That is, when
- 135       you have different subjects or cases in each of your groups(independent group)
- 136       • One categorical independent variable with three or more distinct categories (this can also be a
- 137       continuous variables that has been recoded to give three equal groups e.g age group: 20-24 age,
- 138       25-29 age and 30-44 age.) and one continuous dependant variables(performance test score)
- 139       One-Way Repeated Measures ANOVA
- 140       • It is designed to measure ANOVA design that each subject is exposed to two or more different
- 141       conditions, or measure on the same continuous scale more than one occasions.
- 142       • One independent variable (categorical: e.g time1/ time2/ time3) and one dependent variable
- 143       (continuous: performance test score)
- 144       Two Ways Between Groups ANOVA
- 145       • It is used when there are two independent variables and between-groups indicates that different
- 146       people are in each of the groups. The techniques allows us to look at the individual and joint effect
- 147       of two independent variables on one dependant variable
- 148       • Two categorical independent variables (sex: Males/Females, Age: young, middle, old) and one
- 149       continuous dependant variable(achievement scores)

## Mixed Between-Within Subject ANOVA

- It is used when you have two independent variables: one is a between-subject variable (gender: males/females), the other is a within-subjects variable (time). In this case you would expose a group of both males and females to the treatment at time1 (pre-test) and time2 (post-test)
- One categorical independent between-subjects variable with two or more levels (group1/group2)
- One categorical independent within-subjects variable with two or more levels (time1/ time2/ time3) and one continuous dependent variable (scores of students in mathematics test measure at each time period)

## Multivariate ANOVA

- MANOVA is an extension of ANOVA for use when you have more than one dependant variable. The dependent variables must be related in some way or there should be some conceptual reason for considering them together
- One categorical independent variable (e.g sex) and two or more continuous dependant variables. MANOVA can also be extended to two-way and higher-order designs involving two or more categorical independent variables

## One Way ANCOVA

- It is used when you have 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, are treated as a covariate to 'control' for pre-existing differences between the groups
- One categorical independent variable with two or more levels (group1/group2); One continuous dependant variable (score on mathematics test at time 2) and One or more continuous covariates (scores on mathematics test at time 1)

## Two Way ANCOVA

- ANCOVA will control for scores on your covariate and then perform a normal two-way ANOVA. This will tell you if there is a significant main effect on your first independent variable (e.g group). A main effect for your second independent variable (sex). A significant interaction between the two.
- Two categorical independent variables with two or more levels (sex Males/Females), group: Mathematics / Computer). One continuous dependent variables, one or more continuous covariates.

## 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

## 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

## 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

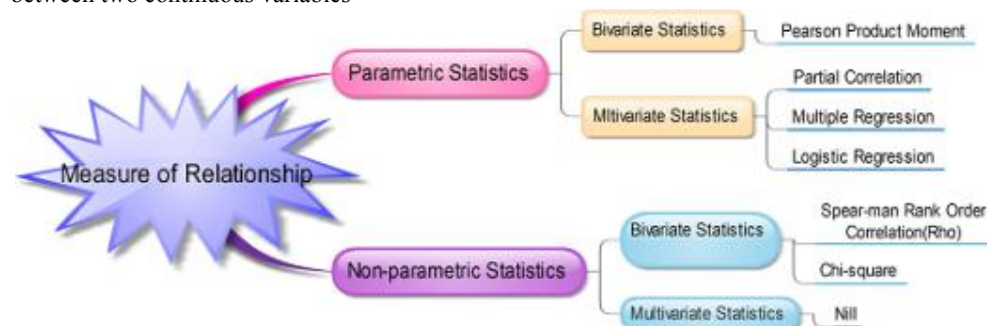
## kruskal-wallis test

- 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

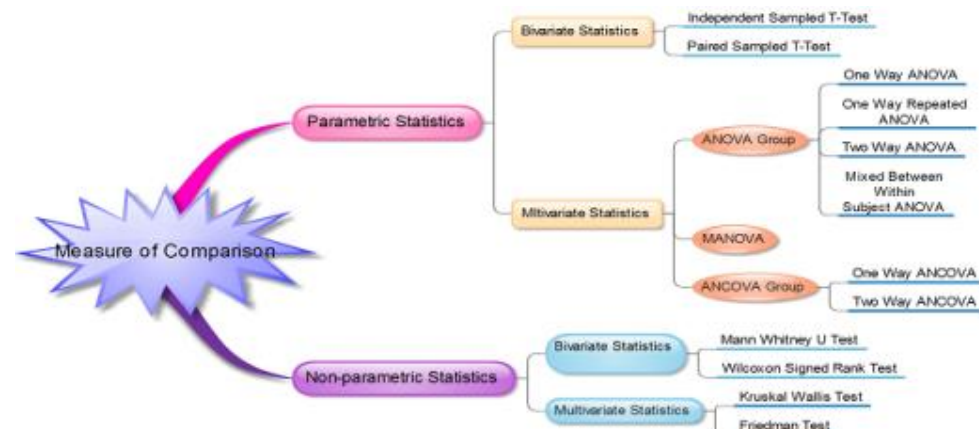
## Frie Sman Test

- 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 three different time or periods or under three different conditions

Spearman's Rank Order Correlation (rho) is used to calculate the strength of the relationship between two continuous variables



Summary Graph for Bivariate Statistics: **Figure 1**



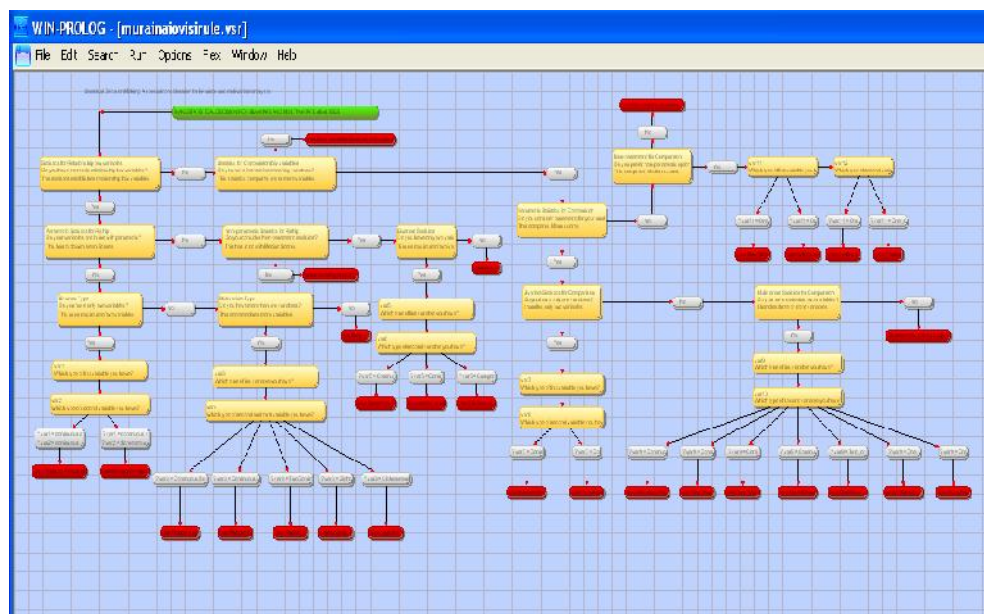
Summary Graph for Multivariate Statistics: **Figure 2**

## 2. Methods

This paper presents procedures necessary for researchers to select appropriate statistical instrument for a desired purpose. The Visirule software was used as a decision supporting tool, 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 ([10,11]). 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.

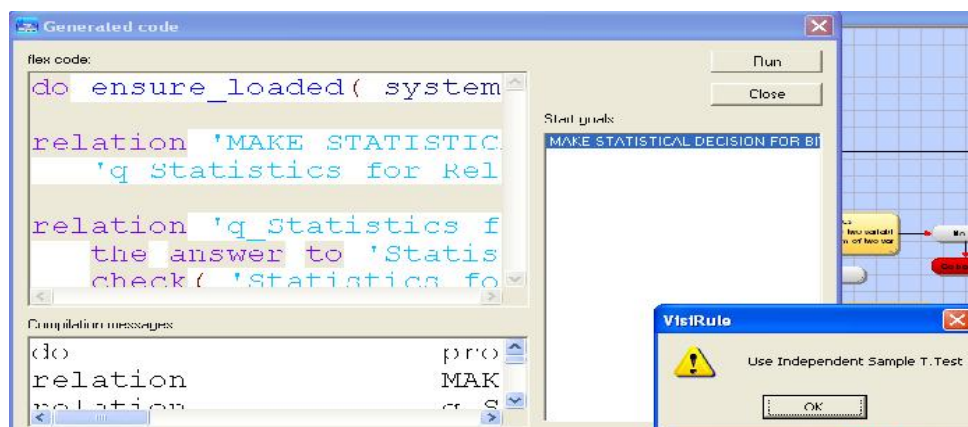


Visirule Application Screenshots: **Figure 3**

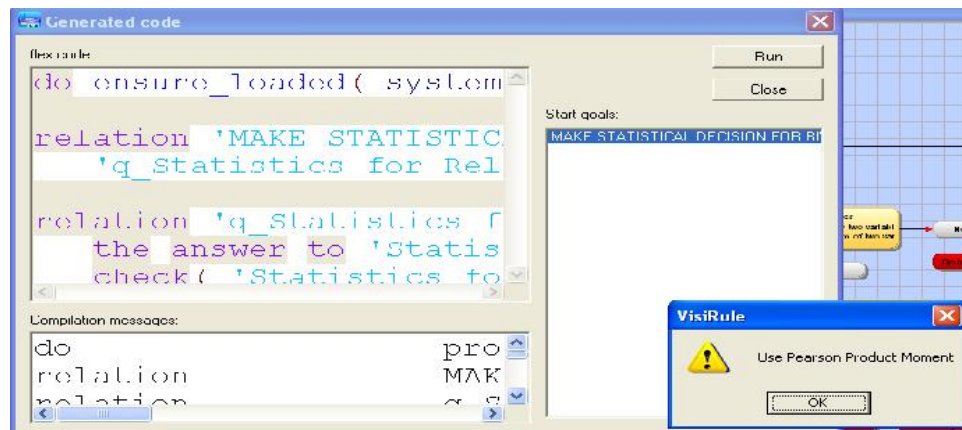
### 3. Results and Discussion

The bivariate and multivariate analysis discussed in this work can be of great use to researchers and to both undergraduate and post-graduate students in making a productive and realizable decision in utilizing suitable statistical data analysis in researches.

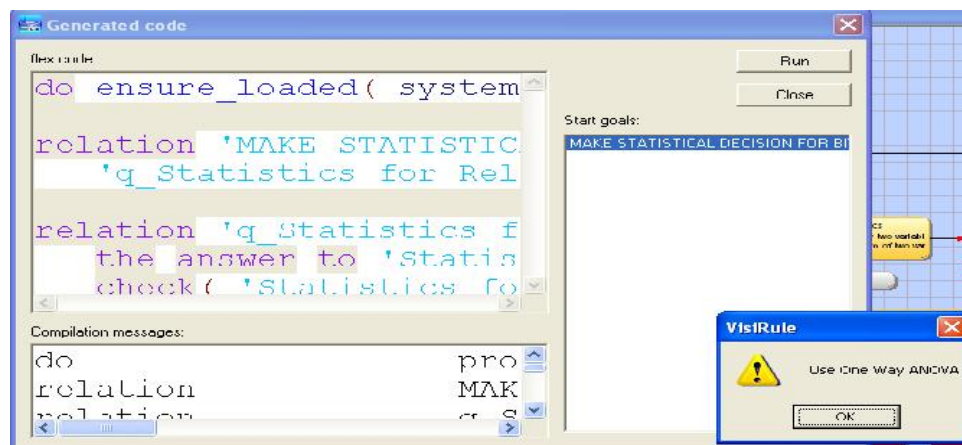
This study will be of an advantage to both lecturers and research students who at the end of their course would write project on their fields of study. Also, students do not need any special knowledge of programming before selecting appropriate statistical tool for analysis. It can also serve as a supportive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



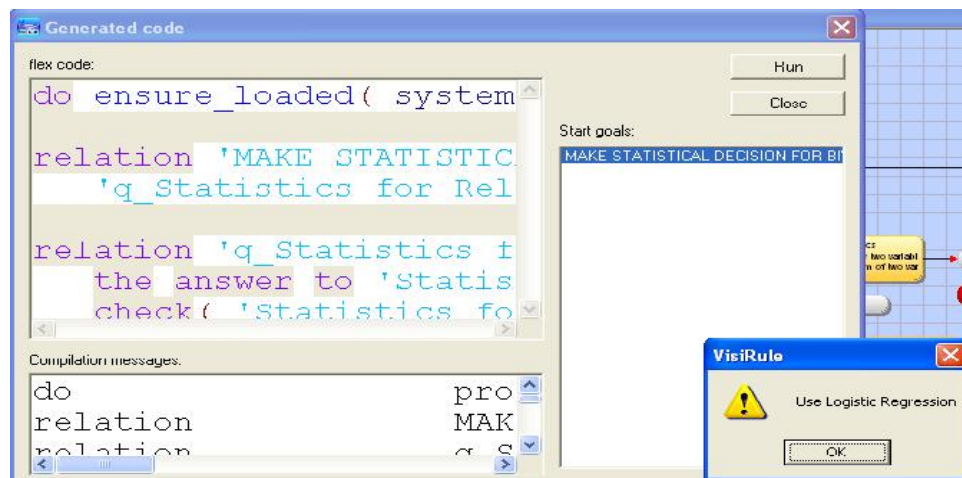
Visirule Output Screenshots: **Figure 4**



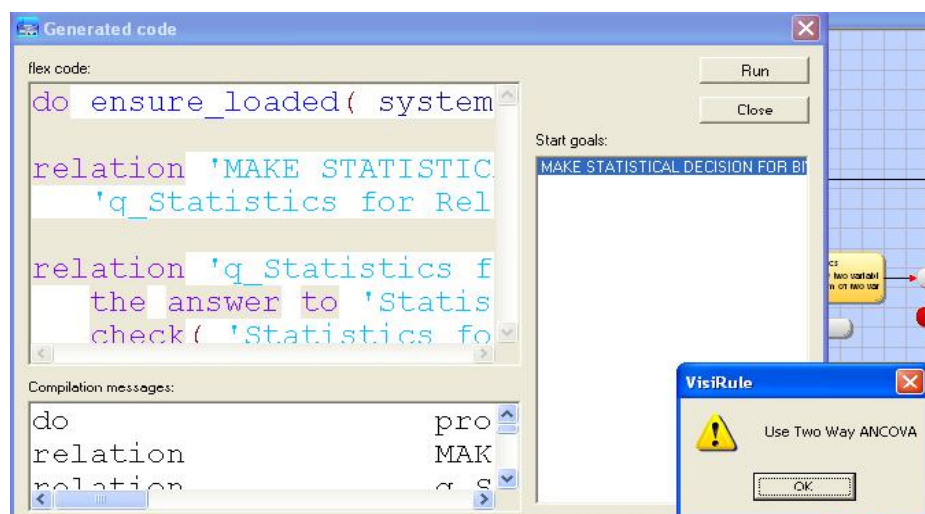
VisiRule Output Screenshots (cont.): **Figure 5**



VisiRule Output Screenshots (cont.): **Figure 6**



VisiRule Output Screenshots (cont.): **Figure 7**



Visirule Output Screenshots (cont.): **Figure 8**

#### 4. Conclusion and Beneficiaries

This study is a step further of earlier research that was primarily on univariate and bivariate statistics [12], 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 part of institutions would find it more helpful to teach students statistics.

#### References

- [1] Redman, L. V. and Mory A. V. H. (1923). Research Methodology: An Introduction. Retrieved on the 12-10-2014 from [www.google/research/methodology](http://www.google/research/methodology)
- [2] Muijs, D. (2006). Doing qualitative research in Education with spss. New Delhi, SAGE Publications Ltd.
- [3] Akinade, E.A and Owolabi, T. (2009). Research Methods. A pragmatic approach for social sciences, Behavioral Science Education. Lagos, Connel Publications.
- [4] Ross R.H (2014). Motivating learners using graphical Visirule based system  
i. International Journal of Information Studies Volume (1): issue (4).
- [5] Ejifugba, A.U (1998). Fundamentals of Research in Health Education Benin, Barloz publishers Inc.
- [6] Woody, C. (2014). Meaning of Research. Retrieved on the 12-10-2014 from [www.google/research](http://www.google/research)

- [7] Healey, J. F. (2014). The essential of Statistics: A tool for Social Research. Retrieved on the 12-10-2014 from [www.askme.com](http://www.askme.com)
- [8] Giarratano, J.C (2007). Principles and programming Fourth Indian Edition Thomson.
- [9] Donna R (2012). Examine the differences between univariate and bivariate data. Retrieved on the 5th May, 2012 from <http://regentsprep.org>
- [10] Alwan, R.H and Benrachi, B. (2009). Motivating learners using graphical Visirule based system International Journal of Information Studies Volume (1): issue (4).
- [11] Bilgi, N.B, kulkarni, R.V and Spenser .C (2010). 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), Volume (1): issue (2).
- [12] Muraina I.O; Rahman M.A; Adeleke I.A and Aiyegbusi E.A. (2013). Use of a Rule Tool in Data Analysis Decision Making. *International Institute of Science, Technology and Education (IISTE) (Information and Knowledge Management)*. Vol.3, No.3, ISSN 2224-5758 (Paper) ISSN 2224-896X (Online), Pp 131-142.

287  
288  
289

290