Use of a Rule Tool in Data Analysis Decision Making

Ismail O, Muraina   (Corresponding author)

Computer Science Department, School of Science, Adeniran Ogunsanya College of Education

Otto/Ijanikin, P.M. B 007, Festac Town, Lagos

Tel: +2348036308030   E-mail: niyi2all@yahoo.com

Mukaila A, Rahman

School of Science, Lagos State University

Ojo Lagos

Tel: +2348028652711   E-mail: rahmalade@yahoo.com

Imran A, Adeleke

Computer Science Department, School of Science, Adeniran Ogunsanya College of Education

Otto/Ijanikin, P.M. B 007, Festac Town, Lagos

Tel: +2348055603413   E-mail: ade1424@yahoo.com

Edward A, Aiyegbusi

Computer Science Department, School of Science, Michael Otedola College of Primary Education, Noforija
Epe-Lagos

Tel: +2348035683229   E-mail: aiyegbusied@yahoo.com

Abstract

Without any doubt, research work is an integrate part of any educational pursuit. However, students engaging in researches often find it difficult to choose an appropriate statistical analysis instrument for their selected data. This paper presents Research Statistical Analysis – Expert (RSA-Expert) which can be employed in selecting 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 using Logic Programming Model. The RSA-Expert discussed in this work can be of great use to researchers in making a firm decision in utilizing suitable statistical data analysis in researches.

Keywords: Research, Visirule, Analysis, Univariate, Bivariate

1. Introduction and Background of the Study

Students in higher institutions usually take a course in research methods and embark on research in their final year. Beginning researchers are scared of embarking on research either as a result of ignorance of what to do or laziness and the procedures to take couple with the cost implication. It has been observed by project supervisors that most undergraduates apparently copy old research report/analysis not bothering whether the analysis suite the research or

not and submit to their supervisors or sometimes contract someone who could conduct and report research on their behalf. To some young researchers, the major challenge comes from which scale or variable brings about a particular analysis/statistics to be used.

The author thinks of possible way of solving this problem by considering the application of decision tree called RSA-Expert (Research Statistical Analysis-Expert).

To many scholars, research could be seen as a process of seeking solution to a problem. Being a process implies that it is an engagement or activity that goes on continuously and can take place anywhere, anytime, in every field of human endeavor (Ejifugha, 1998). It is a careful, scholarly, purposeful and systematic investigation or study that is designed to add or delete from a body of knowledge (Akinde and Owolabi, 2009).

In research, there are certain areas that do pose a little difficulty on the road of researchers. One of these areas is data analysis. After gathering data from the respondents either through questionnaire or test or any other means such as interview, observation etc. the next important aspect is to find meaning to those data collected. The two major statistics are: univariate statistics and bivariate statistics.

2. UNIVERIATE STATISTICS

As Muijs (2006) puts, if researchers are interested to know how the respondents have replied to particular questions or how many times a teacher has asked a particular question or sometimes the researcher might often just what to know how many boys and girls are in the sample. This kind of descriptive information can give us useful information on the variables and the research questions, because the researcher is looking at individual variables, such questions like: How many of students in the senior class are male? This type of analysis is called univariate.

One of the most obvious things to do at the start of an analysis is to look at the frequency distribution of the variables. To do this one need to calculate measure of central tendency. Researcher should also ask him/herself to know at what level the variable is measured. There are three levels of measurement: nominal, ordinal and continuous.

- *Nominal Variable* e.g. Ethnicity, gender. This does not allow researcher to order the categories; the numbers or categories assign are just labels.
- *Ordinal Variables* e.g. Agreement to disagreement, strongly agree, agree, disagree, strongly disagree. This allow researcher to order categories from low to high or from less to more, but he cannot measure precisely what the distance is between scale points.
- *Continuous Variable* e.g. Measuring tape, ruler. This allow researcher to both order categories and to say that the distance between all categories is exactly the same.
  There are three measures of central tendency that go along with these three levels of measurement discussed above:
  1. *Mode*: It means the most common value in a dataset. It is the most suitable measure for nominal variables.
  2. *Median*: It implies the middle value in a set of data ordered from low to high. It is best measure for ordinal variables.

3. *The Mean*: It is sum of all value divided by the number of observations. This is best measure for continuous variables (except where there are outliers, when it may be better to use the median).

As well as measure of central tendency we often want to look at measures of spread of the values around the central, examples are as follows:

1. *Range*: It is simply the difference between the highest and lowest values. As it is sensitive to outliers, researcher often uses the inter-quartile range instead; this is the difference between the third and first quantities. Both are good measures when one is using ordinal variables.
2. *Standard Deviation*: This is a measure of the spread of all the values around the mean. It is best suited for continuous variables.

3. BIVARIATE STATISTICS

In educational research, researchers often want to look at the relationship between two variables. For instance, do boys do better than girls in reading? Is there a relationship between attendance at school and pupils self-concept? Is there a relationship between girls taking Technology Class and their mathematics scores? These and many other questions will necessitate the use of bivariate analysis.

It is seen that the methods researchers use depend strongly on the measurement level of the variables: are they nominal, ordinal or continuous?

When a researcher has two nominal, a nominal and an ordinal (with a limited number of categories) or two ordinal variables (with a limited number of categories) Cross-Tabulation tables can be used together with chi-square test and the *phi* measure of effect size. Cross-Tabulation allows researchers to compare the actual response in the sample to what everyone would expect to find if there were no relationships between the data. The chi-square test allows people to see whether or not relationship is statistically significant and the effect size measure gives a clear look at how strong the relationship is. Sometimes, Researcher may want to compare the means of a continuous variable between two groups, the **t-test** and **Cohen's d** can be a good instrument to do this. The t-test gives statistical significance while Cohen's d is a measure of effect size. In both cases one needs to take account of the assumptions that need to be met.

A times, researcher might be interested in looking at the relationship between two continuous or ordinal variables. Whichever one, correlation co-efficient is the appropriate instrument. This looks at whether or not high scores (in the case of continuous variables) or high rankings (in the case of ordinal variables) on variable X goes together with scores or rankings on variable Y. Where we have two continuous variables we use the **Person's r** correlation co-efficient otherwise (two ordinal variables) use **Spearman's rho (or kendall's tau-b)**.

	Measure of Central Tendency	**Measure of Spread/ Variation**
Nominal	Mode	-
Ordinal	Median	Range, Inter-quartile Range
Continuous	Mean or Median (there are outliers)	Variance and Standard Deviation

Fig.(i) Univariate Summary Table

		Independent		
		Nominal	**Ordinal**	**Continuous**
Dependent	**Nominal**	Cross-Tabulation + Chi-Square + phi	Cross-Tabulation + Chi-Square + phi	Two nominal Groups – T. Test
	Ordinal	Cross-Tabulation + Chi-Square + phi	Cross-Tabulation + Chi-Square + phi or Spearman' rho	Spearman's rho
	Continuous	T. Test (2 Groups) + Cohen's D	Spearman's rho	Pearson's r

Fig.(ii) Bivariate Summary Table

Summary of Differences between Univariate and Bivariate Statistics / Data

The table below shows the differences between univariate and bivariate statistics / data, according to Donna (2012)

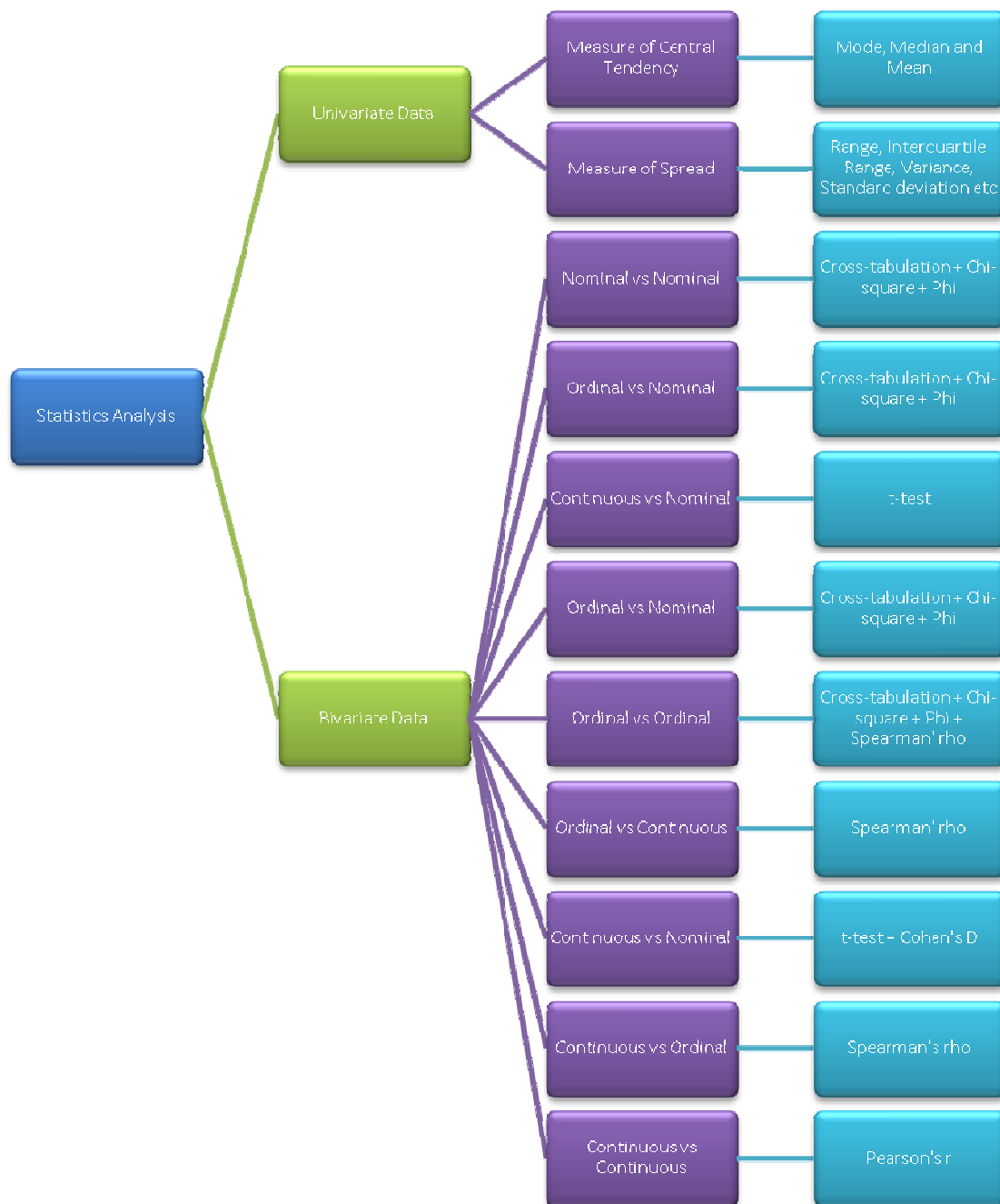| Univariate Data	Bivariate Data
• involving a **single variable**	• involving **two variables**
• does not deal with causes or relationships	• deals with causes or relationships
• the major purpose of univariate analysis is to describe	• the major purpose of bivariate analysis is to explain
• central tendency - mean, mode, median • dispersion - range, variance, max, min, quartiles, standard deviation. • frequency distributions • bar graph, histogram, pie chart, line graph, box-and-whisker plot	• analysis of two variables simultaneously • correlations • comparisons, relationships, causes, explanations • tables where one variable is contingent on the values of the other variable. • independent and dependent variables
Sample question: How many of the students in the freshman class are female?	**Sample question:** Is there a relationship between the number of females in Computer Programming and their scores in Mathematics?

Fig. (1) Diagrammatical representation of both univariate and bivariate data

4. APPLICATION OF VISIRULE IN DATA ANALYSIS

Visirule allows experts / researchers to concentrate on explaining and establishing the structure of the logic correctly using their chosen tools - those embedded materials that can assist researcher to accomplish his mission (Bilgi, Kulhkarni and Spenser, 2010). It is also seen to be a powerful tool that helps to avoid or scrape from some likely errors or bugs which can come into play when trying to code logic in a text based rule language. Identifying the knowledge used in decision making or problem solving is a very crucial component of the expert system design (Giarratano, 2007).

The Visirule is majorly used as decision support tool to select a data analysis suitable for a particular research grounded on the kind of variables researchers have collected.

5. VISIRULE IMPLEMENTATION STRUCTURE

In this structure, more than eighteen (18) questions were raised relating to particular structural analysis relevant to data collected, which is depicted by yellow single choice box. Followed by white boxes indicating options to be taken by the proceeding questions. Finally, there are red boxes implying the outcome or conclusion (i.e. Appropriate statistical analysis tools for a particular purpose) which usually terminate series of inferences. The Visirule used by authors did not cover all the statistical analysis, it is only limited to both univariate and bivariate analysis.

6. CONCLUSION AND RECOMMENDATION
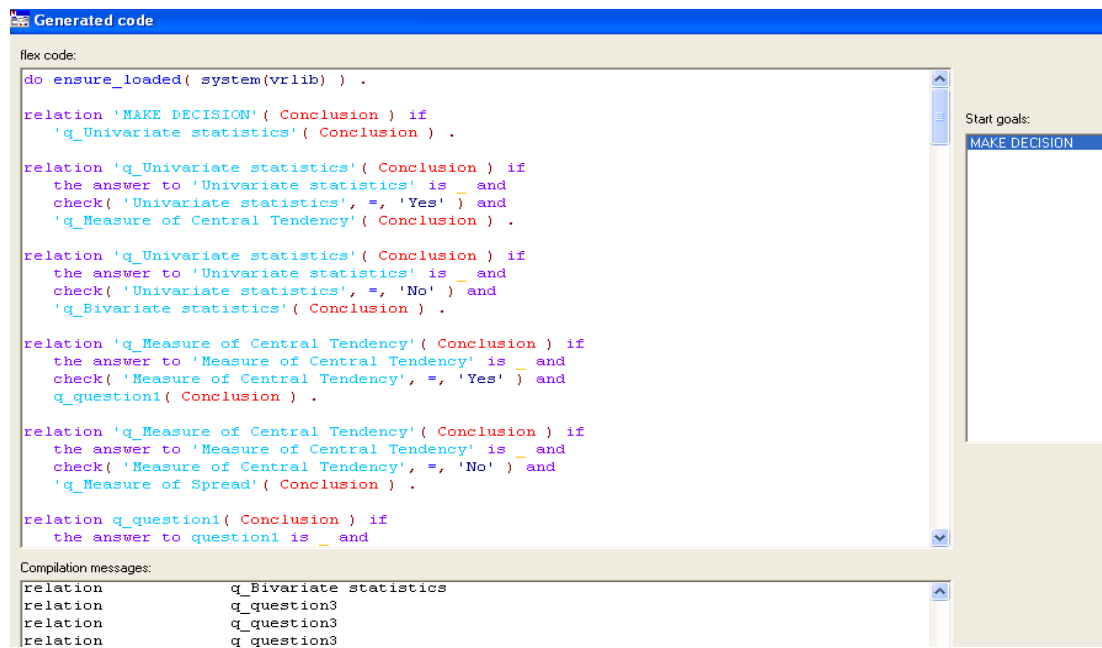
This research work emphasizes on how problem of selecting an appropriate statistical tool for data analysis would be overcome. Alongside, the authors give details of their major levels of measurement that are very crucial in determining the kind of suitable data analysis to use. More lights are shed on condition for using those available statistical data analysis such as measure of central tendency measure of spread, compare means of groups and measure of relationship. To crown it, a Visirule Decision Support Approach was used to assist in realizing this, by just selecting boxes rather than writing codes.

Hence the authors felt that RSA-Expert research work can benefit both lecturers and research students who at the end of their course would write project on their fields of study.

It will also be helpful for the users with very limited knowledge of programming and research, since the use of graphics simplifies the understanding of choice of statistical tools for data analysis. The system can be further improved by considering the limitation of this research.

REFERENCES

Akinade, E.A and Owolabi, T. (2009). Research Methods. A pragmatic approach for social sciences, Behavioral Science Education. Lagos, Connel Publications.

Alwan, R.H and Benrachi, B. (2009). Motivating learners using graphical Visirule based system International Journal of Information Studies Volume (1): issue (4).

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).

Donna R (2012). Examine the differences between univariate and bivariate data. Retrieved on the 5th May, 2012 from http://regentsprep.org

Ejifugba, A.U (1998). Fundamentals of Research in Health Education Benin, Barloz publishers Inc.

Giarratano, J.C (2007). Principles and programming Fourth Indian Edition Thomson.

Muijs, D. (2006). Doing qualitative research in Education with spss. New Delhi, SAGE Publications Ltd.

**Generated code**

flex code:

```
do ensure_loaded( system(vrlib) ) .

relation 'MAKE DECISION'( Conclusion ) if
    'q_Univariate statistics'( Conclusion ) .

relation 'q_Univariate statistics'( Conclusion ) if
    the answer to 'Univariate statistics' is _ and
    check( 'Univariate statistics', =, 'Yes' ) and
    'q_Measure of Central Tendency'( Conclusion ) .

relation 'q_Univariate statistics'( Conclusion ) if
    the answer to 'Univariate statistics' is _ and
    check( 'Univariate statistics', =, 'No' ) and
    'q_Bivariate statistics'( Conclusion ) .

relation 'q_Measure of Central Tendency'( Conclusion ) if
    the answer to 'Measure of Central Tendency' is _ and
    check( 'Measure of Central Tendency', =, 'Yes' ) and
    q_question1( Conclusion ) .

relation 'q_Measure of Central Tendency'( Conclusion ) if
    the answer to 'Measure of Central Tendency' is _ and
    check( 'Measure of Central Tendency', =, 'No' ) and
    'q_Measure of Spread'( Conclusion ) .

relation q_question1( Conclusion ) if
    the answer to question1 is _ and
```
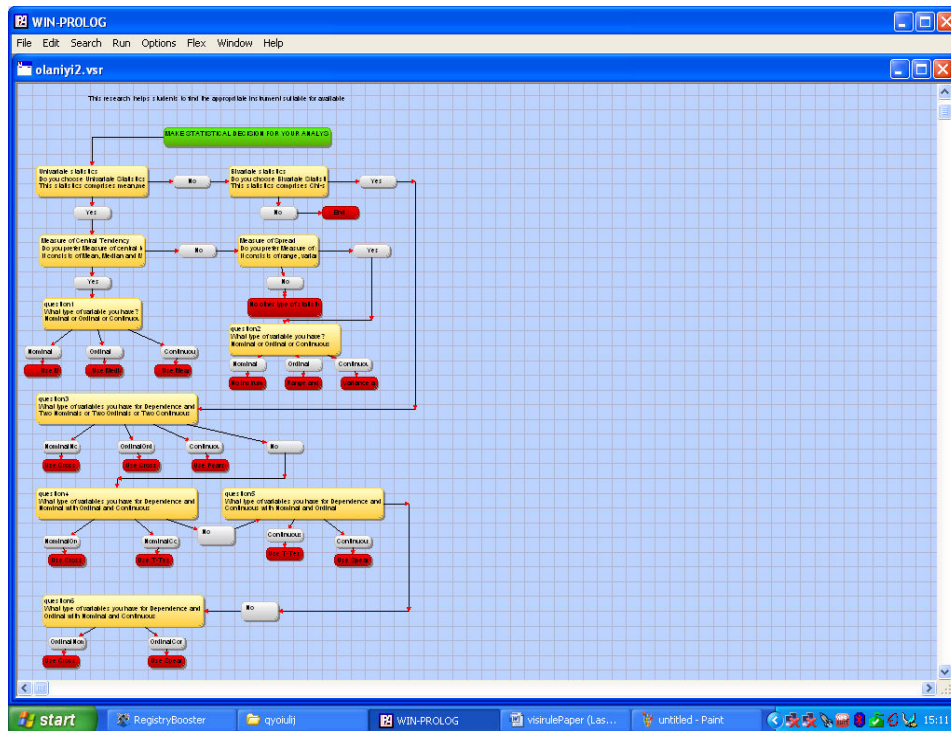
Start goals:

MAKE DECISION

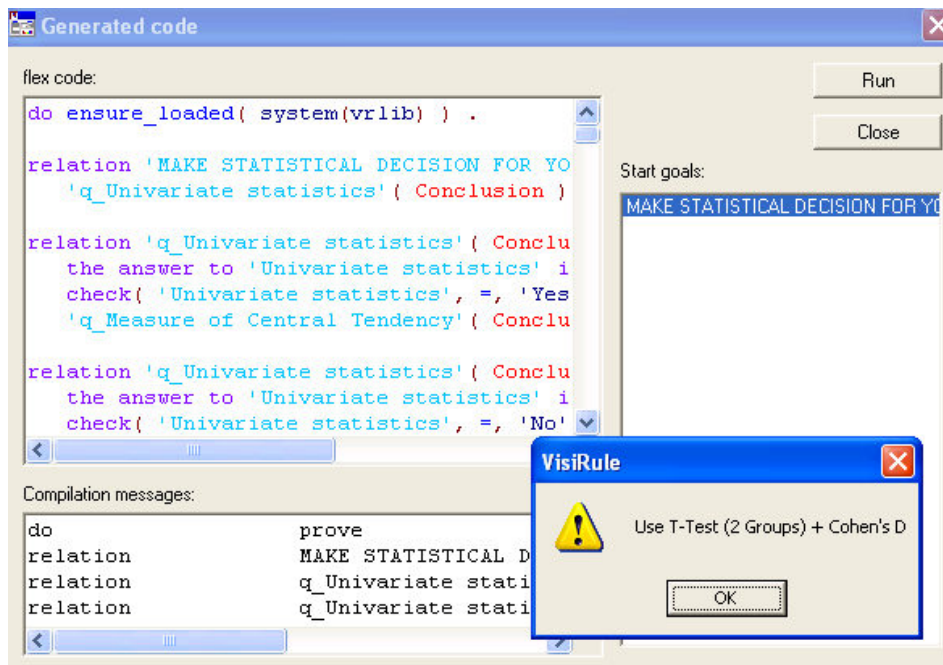Compilation messages:

```
relation            q_Bivariate statistics
relation            q_question3
relation            q_question3
relation            q_question3
```
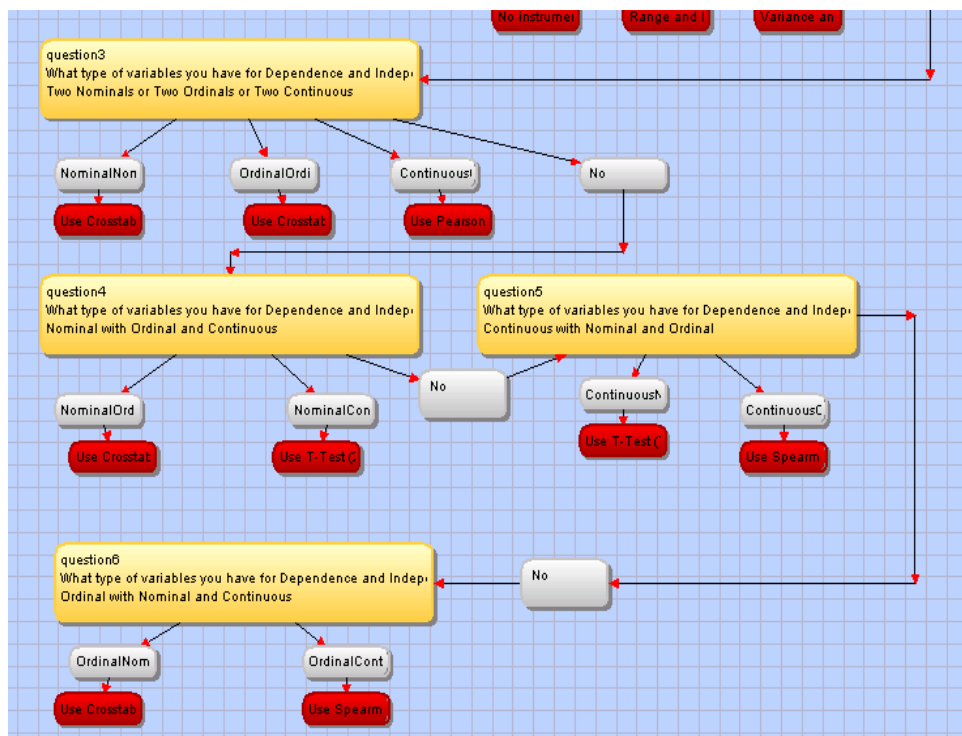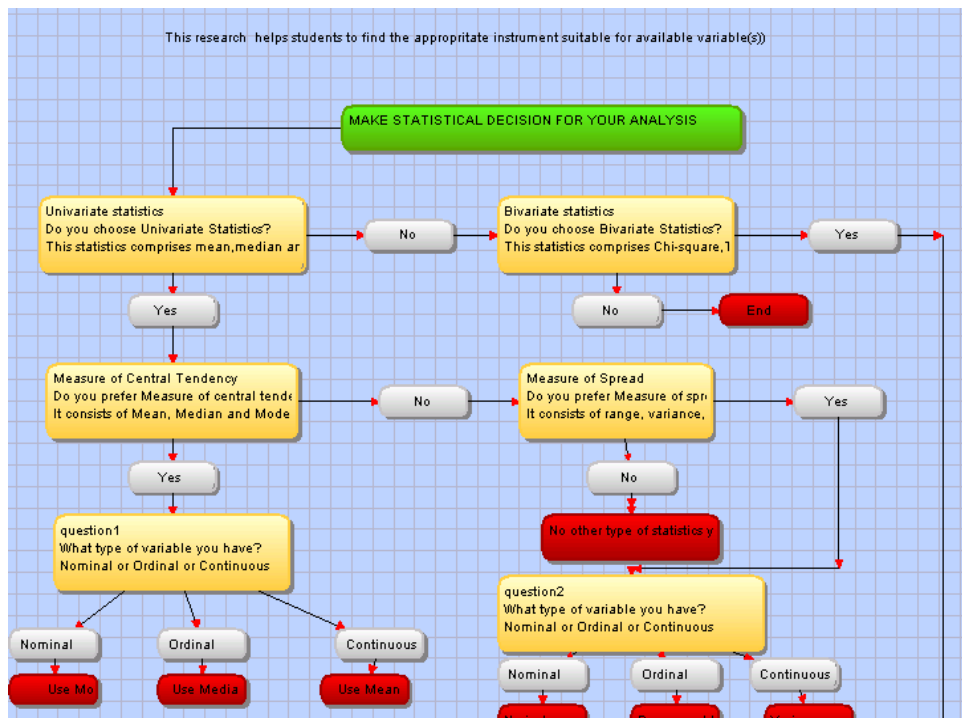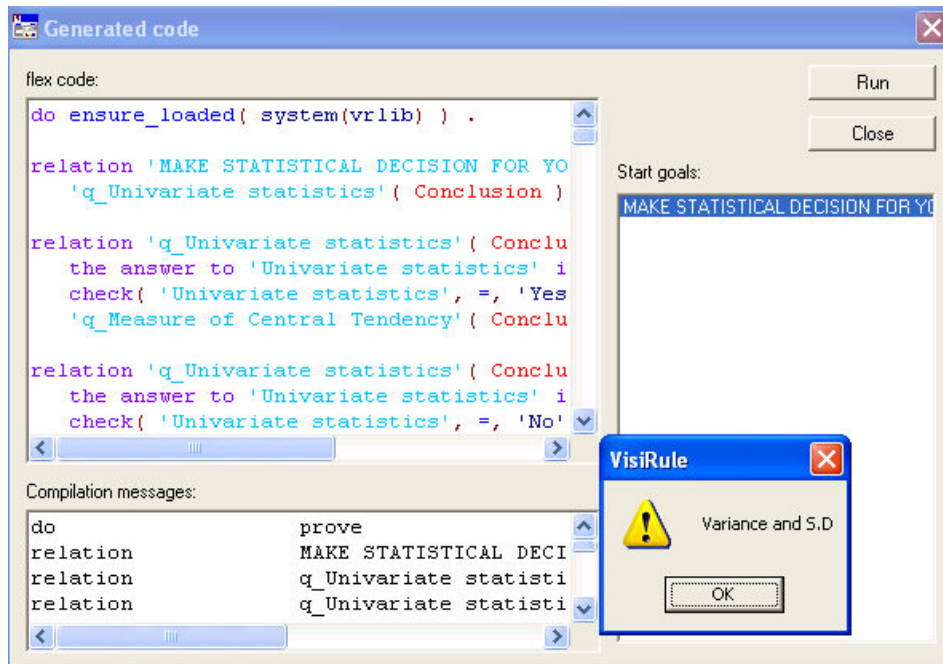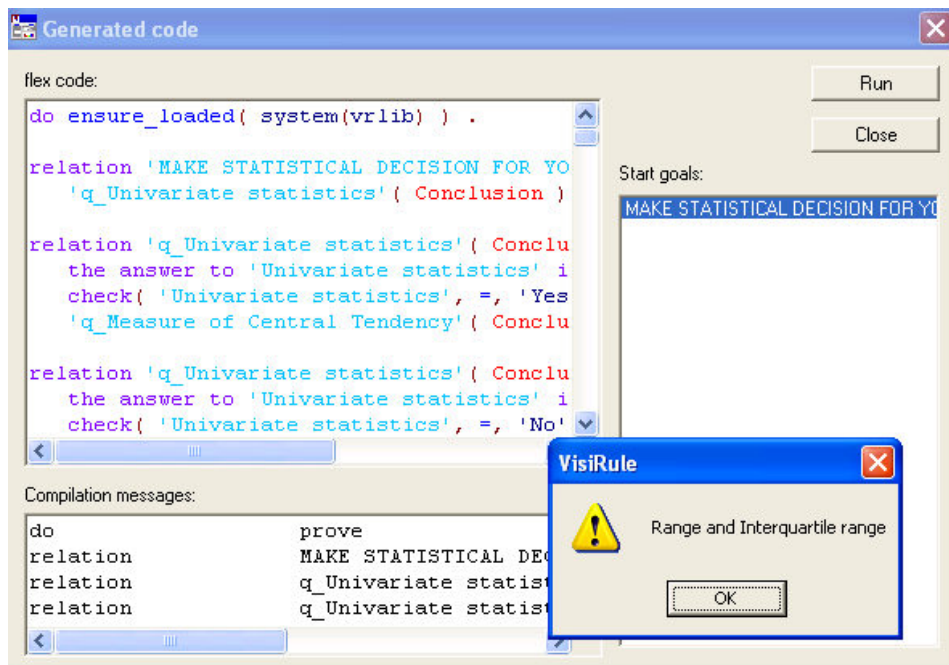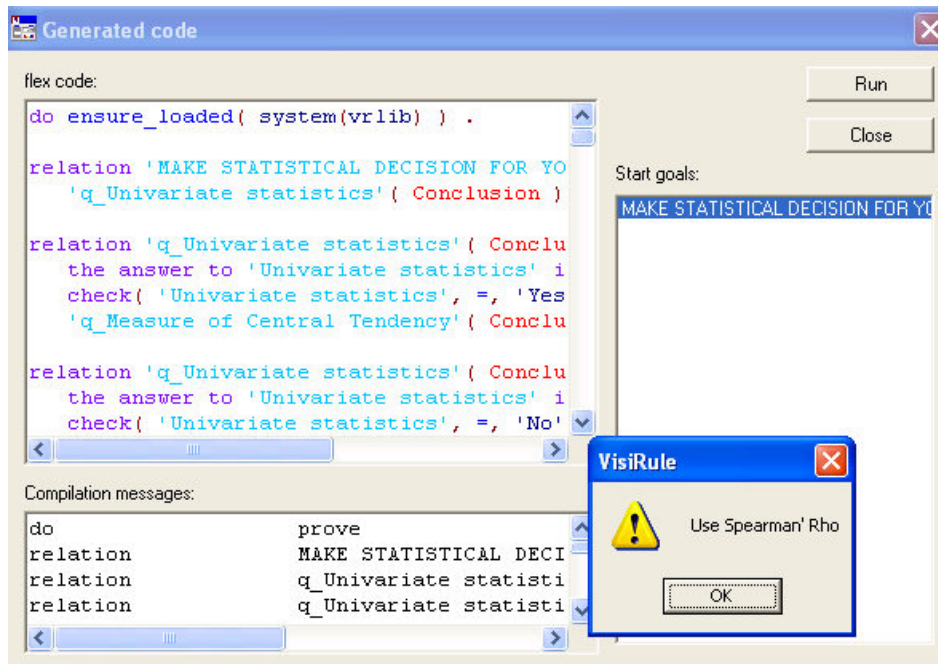
CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar