Chapter 6

Statistical Analysis and Software

Statistics as a wide subject is used in almost all disciplines especially in research studies. Statistics provides scientific tools for representative data collection, appropriate data analysis and summarization of data and inferential procedure for drawing valid conclusion in the face of uncertainty. Each and every researcher should have some knowledge in Statistics and must use statistical tools in his or her research, one should know about the importance of statistical tools and how to use them in their research or survey. The quality assurance of the work must be dealt with: the statistical operations necessary to control and verify the analytical procedures as well as the resulting data making mistakes in analytical work is unavoidable. The emergence of statistical software in the twenty-first century has helped different researchers in the physical and social science to improve in the quality of research. Most renowned researchers in adopting this software in their data analysis have been able to identify the immense contribution to research findings (Adetola, 2013). Any quantitative research cannot be done effectively without Statistical Software. Moreover, it enables research data for easy presentation. It helps professionals to interact

with data thereby paving way for creativity and innovation. Some are user friendly interface with drop-down tips for beginners (ATS, Ucla Edu, 2014). Advances in technology have improved all our lives (Akindutire, 2013) and has allowed experts greater freedom to come out with results within a twinkle of an eye than ever before where it takes time to finish analysis. This same technology has offer tremendous opportunity to research and keep research as a more interesting field of study. This study is significant, because it is an attempt to measure the direct impact of different statistical software on research analysis.

6.1 Common statistical software and their application

Stephanie D. (2009), Standford PhD Statistical Consulting and Karen (2013), Wikipedia library and other scholars have identified different popular statistical software programs, which are SPSS, Eview, SAS, MATLAB, MINITAB, STATA, Mathematica and lots more have been utilized by people across all disciplines for many years and are quite user friendly.

a) *Statistical Package for the Social Sciences (SPSS)* - (Statistical Package for the Social Sciences now Statistical products and Solution services) is most widely used in social science disciplines and courses. SPSS is the oldest software programs developed and made available in 1960s and has been redeveloped over the years, the latest version is SPSS 24.0. Many sociologists, psychologists and social workers use this program to enter their research data and formulate results. Although social science uses SPSS more widely than other fields, many find it easy to navigate with SPSS because it is a package that many beginners enjoy due to its very

easy to use nature. SPSS has a "point and click" interface that allows you to use pull down menus to select commands that you wish to perform. Odusina (2011) disclosed that working with SPSS demand some background knowledge of statistics. There are slight variations in the difference version of SPSS e.g. version 10, 11, 12, 13, 14, 15, 16, 18, etc. SPSS assists the user in describing data, testing hypotheses and looking for a correlation or relationship between one or more variables. SPSS is very suitable for most regression analysis and different kinds of ANOVA (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis but not suitable for time series analysis and multilevel regression analysis)-Wikipedia (2014). Many students, both undergraduate and graduate, are taught SPSS during research analysis classes in demography, psychology, sociology and other social sciences.

b) *Statistical Analysis System (SAS)* - is a software suite developed by SAS Institute for advanced analytics, multivariate analyses, business intelligence, data management, and predictive analytics. In order to use Statistical Analysis System, Data should be in a spreadsheet table format or SAS format. SAS programs have a DATA step, which retrieves and manipulates data, usually creating a SAS data set, and a PROC step, which analyses the data. Each step consists of a series of statements. The DATA step has executable statements that result in the software taking an action, and declarative statements that provide instructions to read a data set or alter the data's appearance. The DATA step has two phases, compilation and execution. In the compilation

phase, declarative statements are processed and syntax errors are identified. Afterwards, the execution phase processes each executable statement sequentially. Data sets are organized into tables with rows called "observations" and columns called "variables". Additionally, each piece of data has a descriptor and a value. The PROC step consists of PROC statements that call upon named procedures. Procedures perform analysis and reporting on data sets to produce statistics, analyses and graphics. There are more than 300 procedures and each one contains a substantial body of programming and statistical work. PROC statements can also display results, sort data or perform other operations. SAS Macros are pieces of code or variables that are coded once and referenced to perform repetitive tasks. SAS data can be published in HTML, PDF, Excel and other formats using the Output Delivery System, which was first introduced in 2007. The SAS Enterprise Guide is SAS' point-and-click interface. It generates code to manipulate data or perform analysis automatically and does not require SAS programming experience to use.SAS is one of the packages that are difficult to learn. To use SAS, one must write SAS programs that manipulate the data and perform data analyses. If you make a mistake in a SAS program, it can be hard to see where the errors occurred or how to correct it. However, it can take a long time to learn and understand data management in SAS than many other packages with simpler commands line. However, SAS can work with many data files at once SAS can handle enormous data files and the number of records is generally limited to the size of user's hard disk. SAS performs most

general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis). The greatest strengths of SAS are probably in its ANOVA, mixed model analysis and multivariate analysis, while it is probably weakest in ordinal and multinomial logistic regression (because these commands are especially difficult), and robust methods (it is difficult to perform robust regression, or other kinds of robust methods)- ATS Ucla Edu(2014). While there are some supports for the analysis of survey data, they are quite limited as compared to STATA.

c) *Econometric Views (EViews) i*s a statistical package for Windows, used mainly for time-series oriented econometric analysis. It is developed by Quantitative Micro Software (QMS), now a part of IHS. Version 1.0 was released in March 1994, and replaced MicroTSP. The TSP software and programming language had been originally developed by Robert Hall in 1965. The current version of EViews is 9.5, released in March 2016. EViews can be used for general statistical analysis and econometric analyses, such as cross-section and panel data analysis and time series estimation and forecasting. EViews relies heavily on a proprietary and undocumented file format for data storage. EViews combines spreadsheet and relational database technology with the traditional tasks found in statistical software, and uses a Windows GUI. This is combined with a programming language which displays limited object orientation. However, for input and output it supports numerous formats, including databank format, MS-Excel format,

SPSS/PSPP, DAP/SAS, STATA, RATS, and TSP. EViews can access ODBC databases.

d) **MINITAB -**is statistical software used by educators, students, scientists, business associates and researchers to provide statistical software in a multitude of areas. MINITAB, developed around 1990, and remains one of the oldest statistical software programs available. MINITAB has compatibility with PC, Macintosh, Linux and all other major platforms. As one of the easiest statistical software programs to use, MINITAB remains a popular choice with those new statistical software. With drop-down menus and dialog boxes describing how and what to do next, MINITAB persists as a popular choice for teaching students about statistics and data analysis. MINITAB primarily has a user base of educators using the program to show students research methods and analysis in college and graduate-level courses. MINITAB performs most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, but has its weaknesses in general linear model (GLM) and Multilevel regression).

e) **STATA-** is a powerful statistical package with smart data-management facilities, a wide array of up-to-date statistical techniques, and an excellent system for producing publication-quality graphs. STATA latest version was produced June 24, 2013 which is a fast and easy to use data management package. STATA is available for Windows, Unix, and Mac computers. The standard version is called STATA/IC (or Intercooled STATA) and can handle up to 2,047 variables. There is a special edition called STATA/SE that can handle up to 32,766 variables

(and also allows longer string variables and larger matrices), and a version for multi-core/multiprocessor computers called STATA/MP, which has the same limits but is substantially faster. STATA performs most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis and time series analysis).

f) **R & MATLAB**– Stanford (2014**)** identified R and MATLAB as the richest statistical systems by far. They contain an impressive amount of libraries, which is growing each day. Even if a much desired specific model is not part of the standard functionality, you can implement it yourself, because R and MATLAB are really programming languages with relatively simple syntaxes. As "languages" they allow you to express any problem. In terms of modern applied statistics tools, R libraries are somewhat richer than those of MATLAB. R (http://cran.at.r-project.org) is a comprehensive statistical environment and programming language for professional data analysis and graphical display. It provides additional packages for statistical data analysis in different areas. The R software is free and runs on all common operating systems and it can be used to carry out some simple and complex analyses. On the flip side, MATLAB has much better graphics, which you will not be ashamed to put in a paper or a presentation. MATLAB and R perform most general statistical analyses (regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis). The greatest strengths of both are

probably in its ANOVA, mixed model analysis and users creative freedom in analysis.

g) *MS-EXCEL-*Microsoft Excel is part of the Microsoft Office productivity suite has applications worldwide and one can use Excel to analyze data, for example, in accounts, budgets, billing and many other areas. Excel allows you to explore the menu bar and the different tasks that can be done with it. You can work on sample spreadsheets doing basic math, adding and deleting columns and rows, and preparing the worksheet for printing. You can run your data visually to show trends, patterns and comparisons between the data in a chart, table or other template, and Excel performs most general statistical analyses but weak in regression, logistic regression, survival analysis, analysis of variance, factor analysis, multivariate analysis.

h) **ARTFISH** - Approaches, Rules and Techniques for Fisheries statistical monitoring-ARTFISH has been developed as a standardized tool adaptable to most fisheries in the developing countries. Its design was driven by the need to provide users with robust, user-friendly and error-free approaches and computer software, and achieve the implementation of cost-effective fishery statistical systems with minimal external assistance.

i) **BEAM 1 and 2**-Bioeconomic Modeling of Artisanal and Industrial Sequential Shrimp Fisheries Software packages for bio economic modeling of artisanal and industrial sequential shrimp fisheries based on an age structured Thompson and Bell (in Ricker, 1975) yield per recruit biological model and a simple input output microeconomic model.

BEAM1 gives its simulated results by age groups. BEAM2 gives them by standard commercial categories as used in the shrimp fishing industry.

j) **BEAM 3**- A Bio-economic Simulation Model of Tropical Shrimp Fisheries Using Fixed or Random Recruitment - BEAM3 is a stochastic model that aims at determining the optimum size, in the long term, of the level of fishing capacity corresponding to the management goals.

k) **BEAM 4**- Analytical Bio-economic Simulation of Space structured Multispecies and Multi fleets Fisheries – BEAM 4 is a versatile tool for the rational management of exploited living aquatic resources. It can deal with a fishery system of several stocks, several fleets, several areas (fishing grounds) and several processing plants and can account for migration of the animals as well as seasonality of recruitment.

l) **CLIMPROD**-Experimental Interactive software for choosing and fitting surplus production models including environmental variable - Various equations allow for the introduction of an environmental variable into surplus production models. CLIMPROD helps the user select the most appropriate model for a particular case, according to objective criteria. It resembles a simple expert system and uses artificial intelligence language (PROLOG) to converse with the user.

m) **CLIMPROD-PLUS**-Experimental Interactive software for choosing and fitting surplus production models including environmental variable - Various equations allowing for the introduction of an environmental variable into surplus production models are available. CLIMPROD helps the user to select the model corresponding to a particular case

according to objective criteria. It looks like a simple expert-system and uses artificial intelligence to converse with the user.

n) **FAST:** Fishing Activity Simulation Tool-The Fishing Activity Simulation Tool, called FAST has been developed as an extension for the ESRI ArcView GIS Software (v.3.1) in combination with the Spatial Analyst v. 1.0 extension. It is based on two models focusing on the analysis of the spatial components of the fishing effort distribution.

o) **FISAT II**-Stock Assessment Tools - The FAO-ICLARM Stock Assessment Tools (FiSAT) is a program package consisting of robust methodologies for use with microcomputers, enabling users to formulate management options for fisheries, especially in data-sparse, tropical contexts.

p) **FishStatJ** - Software for fishery statistical time series-The system provides users with access to Fishery Statistics of various sorts. Any data having time series structure can potentially be stored and processed by FishStatJ. The system consists of a main module and the datasets. Currently one set of data is included within the set up module, while other sets are provided separately.

q) **FishStat Plus** - Universal software for fishery statistical time series-The system provides users with access to Fishery Statistics of various sorts. Any data having time series structure can potentially be stored and processed by FishStat Plus. The system consists of the main module and the datasets. Each dataset can be installed and uninstalled separately. This application, developed in 1998 (FishStat+), is NOT

compatible with 64 bit operating system and Windows 7 and has now been replaced with FishStatJ.

r)　**MTBASE 1.1**-The Model of Trophodynamics of Black - Azov Ecosystem - An ecology software package for the study of the impact of exotic species introduction on fish stocks: Mnemiopsis-Anchovy relationships case (MTBASE 1.1).

s)　**NANSIS**-Software for Fishery Survey Data Logging and Analysis - NANSIS is a Survey Information System for logging, editing and analysis of scientific trawl survey data (trawl/catch data and length/frequency data).

t)　**SPATIAL**- Space time Dynamics in Marine Fisheries-SPATIAL is a simulation package developed to model the space-time distribution of fishing intensity using alternative approaches.

u)　**THOMPSON AND BELL YIELD ANALYSIS**-This application refers to the use of a group of models loosely based on the work of Thompson and Bell (1934). The models are essentially mathematical depictions of fisheries, or parts of fisheries.

6.2 FEATURES OF STATISTICAL SOFTWARE

Statistical software has some common characteristics that make it reliable and suitable for data analysis: Data editor is in rows and columns which make it very easy to enter numeric data, There is availability of menu bar comprises drop-down menu, quick analysis as well as brief user manual, statistical level of measurement is put into consideration in data entry, They follow the initial steps in research project -

(a) Getting your data ready to enter into the software.

(b) Defining and labeling variable

(c) Entering data appropriately with each row containing each case and each column as variable.

(d) Data checking and cleaning is possible.

All data should be numeric, although it may not be all variables it is not desirable to use letter or word (String variable) as data. This can be achieved by recoding the letter or word (string data) into desirable numeric and labeled appropriately. Data exploration can be done to check for errors and other accuracy. The statistical level of significance for rejecting null hypothesis (Ho) is when your p-value significance is less than 0.05.

6.3 COMMON STATISTICAL TOOLS AND THEIR ANALYSIS BY SPSS

Statistical analysis generally begins with the calculation of a number of summary statistics such as the mean, median, standard deviation, etc., and by creating informative graphical displays of the data such as histograms, box plots, and stem-and-leaf plots. The aim at this stage is to describe the general distributional properties of the data, to identify any unusual observations (*outliers*) or any unusual patterns of observations that may cause problems for later analyses to be carried out on the data. Following the initial exploration of the data, statistical tests may be applied to answer specific questions or to test particular hypotheses about the data. According to Shamoo and Resnik (2003) various analytic procedures ―provide a way of drawing inductive inferences from data and distinguishing the signal (the phenomenon of interest) from the noise (statistical fluctuations) present in the data.

The Statistical analysis depends on the objective of the study. The objective of a survey is to obtain information about the situation of the population study. The first Statistical task is therefore is to do a descriptive analysis of variables. In this analysis it is necessary to present results obtained for each type of variable. For qualitative and dichotomous variables, results must be presented as frequencies and percentages. For quantitative variables, the presentation is as means and deviations. After this analysis, you can access the association between variables and predictive analysis based on multiple regression models. An essential component of ensuring data integrity is the accurate and appropriate analysis of research findings. Improper statistical analyses distort scientific findings, mislead casual readers (Shepard, 2002), and may negatively influence the public perception of research. Integrity issues are just as relevant to analysis of non-statistical data as well. In deciding which test is appropriate to use, it is important to consider the type of variables that you have (i.e., whether your variables are categorical, ordinal or interval and whether they are normally distributed. Some common statistical tools/techniques are furnished below:

Mean

The arithmetic mean, more commonly known as - the average is the sum of a list of numbers divided by the number of items on the list. The mean is useful in determining the overall trend of a data set or providing a rapid snapshot of your data. Another advantage of the mean is that it's very easy and quick to calculate.

Standard Deviation

The standard deviation, often represented with the Greek letter sigma, is the measure of a spread of data around the mean. A high standard deviation signifies that data is spread more widely from the mean, where a low standard deviation signals that more data align with the mean. In a portfolio of data analysis methods, the standard deviation is useful for quickly determining dispersion of data points.

Steps in SPSS: Enter data in SPSS Data Editor. There are two views in SPSS Data Editor. In variable view, one can define the name of variables and variable types such as string or numeric and data view gives the spreadsheet in which data pertaining to variables may be entered in respective columns. In the present case, we enter data in numeric format. Choose Analyze from the Menu Bar. Now select **Analyze → Descriptive Statistics → Frequencies….** Click on the **Statistics…** tab to get the Frequencies: Statistics dialog box. Select the desired options (Mean, Std. Deviation, Minimum, Maximum, Skewness and Kurtosis) required for analysis. Click on the **Continue** tab to return to the Frequencies dialog box. Click on **OK** to get the output.

Correlation and Regression

Correlation and Regression models the relationships between dependent and explanatory variables, which are usually charted on a scatter plot. The regression line also designates whether those relationships are strong or weak. Regression is commonly taught in high school or college statistics courses with applications for science or business in determining trends over time.

Steps in SPSS: Obtain correlation coefficient between each pair of the variables using the following steps. Once the data entry is complete, Choose Analyze from the Menu Bar. Now select **Analyze → Correlate → Bivariate.** For partial correlation use the following steps. Choose Analyze from the Menu Bar. Now select **Analyze → Correlate → Partial.**

For fitting a multiple linear regression equation by taking yield as dependent variable and biometrical characters (for example) as explanatory variables. Use the following steps: **Analyze → Regression → Linear.** In the Linear Regression **dialog** box select the Dependent Variable and the independent(s) variable box. For the Durbin-Watson statistics click **Statistics** on the Linear Regression dialog box. In the Linear Regression: Statistics dialog box check the Durbin-Watson check box under Residuals and other options as required. Click **Continue** to return to the Linear Regression dialog box. Click **Save** in the Linear Regression dialog box. In the Linear Regression: Save dialog box check the desired options under Predicted Values, Residuals, Distances and Influence Statistics. Click **Continue** to return to the Linear Regression dialog box. In the Linear Regression dialog box Click **Options.** Click **Continue** to return to the Linear Regression dialog box. Click **OK**.

Hypothesis Testing

Also commonly called *t* testing, hypothesis testing assesses if a certain premise is actually true for your data set or population. In data analysis and statistics, you consider the result of a hypothesis test *statistically significan*t if the results couldn't have happened by random chance. Hypothesis tests are used in everything from science and research to business and economic.

One sample t-test

A one sample t-test allows us to test whether a sample mean (of a normally distributed interval variable) significantly differs from a hypothesized value i.e. the mean of the variable for this particular sample of students which is statistically (significantly) different from the test value. We would conclude that this group of students has a significantly higher mean on the writing test than the given.

Steps in SPSS: In the SPSS menu, select Analyze>Compare Means>One Sample t-test Select the variable(s) from the list you want to look at and click the button to move it into the "Test Variable(s)" area. Then enter the test value. In this example, we're testing the hypothesis that the median house value is 200,000. Select the Options button and check that the confidence interval is where you want it (the default is 95%, which is what we normally use. Select "Continue" and then OK on the main window. You should get the following output.

Independent Samples T-test: This test is similar to the one-sample test, except rather than testing a hypothesized mean, we're testing to see if there is a difference between two groups. For the grouping variable, you can choose a demographic trait (such as gender, age, ethnicity, etc) or any other variable that classifies your groups. (In an experimental design, it is a good way to test the differences between the control group and the manipulation group.) In this example, we'll use gender.

Steps in SPSS: In the SPSS menu, select Analyze>Compare Means>Independent Samples T-test Select your Test variable from the list. This is the variable for which you want to compare means. Now select the

grouping variable, which is the trait you're using to divide the groups. For this example, we will select gender. Select it from the list and click the arrow next to "Grouping Variable." Then, click "Define Groups …" and enter the values for the two groups. In this example, 1=Female and 2=Male. Also notice the "Cut Point" option. What if we wanted to divide the sample into two groups based on home value? The cut point would be the value where you split the sample. For example, if you entered 100,000, it would create two groups – one for home value less than 100,000 and another for more than 200,000. For this example, let's stick to gender, though. Select continue, and then click "OK."

Paired Samples t-test: With the paired samples t-test, we're not testing for differences between groups. Instead, we're testing for means of different variables within the sample. For example, we want to compare the mean for user-created videos and the mean for company generated videos. Go to Analyze>Compare Means>Paired Samples T-test Select the two variables you want to compare, and click the arrow to move them into the "Paired Variables" pane. Under options, make sure that you're using a 95% confidence interval. Clicks continue and then OK and you'll get the following output.

One-Way ANOVA: What if we want to test for differences between more than two groups. ANOVA (which stands for "Analysis of Variance") is the way to go. Analysis>Compare Means>One-Way ANOVA Select the variable(s) you want to test and move into the "Dependent List" pane. Now, move the variable that you are using to separate them into groups into the "Factor" pane. For example, in our data set, there are three age groups

("1",")2" and "3") for the 15-18, 19-24, and 25+ groups, respectively, in the Age Group variable. Click "OK" and you get the following output. To determine which one of the means is significantly different, we must use a post-hoc test. Go back to Analyze>Compare Means>One Way ANOVA but this time, click Post-Hoc, Then, select "Tukey" and "Duncan".

One sample median test: A one sample median test allows us to test whether a sample median differs significantly from a hypothesized value. The one-sample median test is used to test a hypothesized median value against an observed median value in a representative sample.

For instance, let's say researchers believe that the college-age students will rate the look of a new social networking site at as a "4" on a 5-point Likert scale. Researchers collect a representative sample from a Psychology 101 class and find a median value of "1." This difference of ratings between what researchers expected (Median = 4) and the observed value (Median = 1) may be statistically significant according to a one-sample median test. The data is entered in a between-subjects fashion and is specified as a **scale** variable in the **Variable View**. Click **Analyze**. Drag the cursor over the **Nonparametric Tests** drop-down menu. Click **One Sample**. Click on the **Settings** tab. Under the **Customize tests marker**, click on the **Compare median to hypothesized (Wilcoxon Signed-rank test)** box to select it. In the **Hypothesized median:** box, type in the *a priori* **hypothesized median value**. Click on the **Run** button.

Binomial test

A one sample binomial test allows us to test whether the proportion of successes on a two-level categorical dependent variable significantly differs from a hypothesized value. Go to Analyze > Nonparametric Tests > Legacy Dialogue > Binomial Test

Chi-square goodness of fit

A chi-square goodness of fit test allows us to test whether the observed proportions for a categorical variable differ from hypothesized proportions.

First step: Select Data → Weight Cases. A box pops up with a list of the input variables. Click the "Weight cases by" button, and then move the count variable into the box under "Frequency Variable:" Finally, click "OK."

Second step: Select Analyze → Nonparametric Tests → Chi-Square. Move the variable indicating categories into the "Test Variable List:" box. If you want to test the hypothesis that all categories are equally likely, click "OK." Otherwise, categories should be labeled in ascending numerical order, say 1, 2, 3, ..., and you must specify the expected count for each category by first clicking the "Values" button under "Expected Values." Then, in the box to the right of "Values," enter the expected count for category 1 and click the "Add" button. Now enter the expected count for category 2 and click "Add." Continue in this way until all expected counts have been entered. Then click "OK."

Notes: (1) By default SPSS assumes that the chi-square degrees of freedom are k-1, where k is the number of categories. If your null hypothesis does not completely specify all probabilities, then you will have to determine the

correct P-value yourself. (2) SPSS will not perform a chi-square test if one or more observed counts are 0. In this case, use other software or combine categories.

Statistical software packages have been recognized to have contributed immensely to research analysis by helping to minimizing human and experimental errors in data analysis. Conclusively, it is the opinion of the researcher, after analyzing the field data to a very large extent, that statistical software has contributed immensely to social research especially in the area of demographic and data analysis. This was achieved by using a scientific approach to solve fundamental problems in research which is data analysis.

1. It has been discovered that some analysis such as post Hoc, complex analysis in time series, regression and variance analysis cannot be calculated manually effectively without statistical software packages.

2. Statistical software packages have been discovered to help academic staffs in higher institution to improve their research expertise by attending training on usage of statistical packages.

3. Statistical packages make research work robust and faster.

4. It was discovered that maximum efficiency of researcher in statistical software is determine by the years of experience in usage and the area of specialization.

5. Most reason for using statistical software was it easy usage, suitability for many statistical analysis and the likes.

6. While reason for non usage range from lack of attention to learn, difficult usage, cost of licensing, etc. It was discover that statistical

software are not expensive neither are they too difficult to use but people do not give attention to its learning.

Reference:

[1] Adetola O.G (2013) Learning Statistical Package Workbook. Nigeria (Unpublised).

[2] Akindutire, A.F. (2013) Usage of Mobile Phone: Deception, Deviance and Fraud (A case study of Ambrose-Alli) Chris McGann, Role of Statistical Software, eHow Contributor http://www.ehow.com/facts_6923793_role-statistical-software.html#ixzz2xoZFcTZX visited on 3rd April, 2014.

[3] Daniel, S. (2009). *The Most Popular Statistical Software.* accessed on 3rd April, 2014: eHow Contribution Eview Users Manual (2009) History and usage of Eview visited on 28th April, 2014

[4] Gottschalk, L. A. (1995). Content analysis of verbal behavior: New findings and clinical applications. Hillside, NJ: Lawrence Erlbaum Associates, Inc International Organization of Scientific Research, IOSR.

[5] Karen (2013) Choosing a Statistical Software Package or Two Source:http://www.theanalysisfactor.com/choosing-statistical-software/ VISITED ON 7TH April, 2014

[6] McCullough, B.D. (1999). "Econometric software reliability: EViews, LIMDEP, SHAZAM and TSP". Journal of Applied Econometrics 14 (2):191–202. HYPERLINK

[7] McGann, C. (2009). Role of Statistical Software. eHow Contribution accessed by 3rd April, 2014

[8] Resnik, D. (2000). Statistics, ethics, and research: an agenda for educations and reform. Accountability in Research. 8: 163-88

[9] Savenye, Robinson, 2004 Clinical significance of research: A growing concern. Canadian Journal of Nursing Research, 24, 1-4.

[10] Schroder, K.E., Carey, M.P., Venable, P.A. (2003). Methodological challenges in research on sexual risk behavior: I. Item content, scaling, and data analytic options. Ann Behav Med, 26(2): 76-103.

[11] Shamoo, A.E., Resnik, B.R. (2003). Responsible Conduct of Research. Oxford University Press.

[12] Silverman, S., Manson, M. (2003). Research on teaching in physical education doctoral dissertations: a detailed investigation of focus, method, and analysis. Journal of Teaching in Physical Education, 22(3): 280-297.

[13] Smeeton, N., Goda, D. (2003). Conducting and presenting social work research: some basic statistical considerations. Br J Soc Work, 33: 567-573.

[14] STATA PRESS site (2014) Stata history and usage.visited 28th April, 2014.

[15] W.N. Venables and B.D. Ripley. Modern Applied Statistics with S. Springer, first edition, 1994.

[16] W.N. Venables and B.D. Ripley. S Programming. Statistics and Computing.Springer Verlag, New York, N.Y., 2000.

[17] Zaslavsky Alan (2014) HYPERLINK "http://www.hcp.med.harvard.edu/statistics/survey-soft/ Summary of Survey Analysis Software wikipedia library