HEURISTICS OF APPLYING STATISTICAL TESTS USING APPROPRIATE MEASUREMENT SCALES

***Kamran Ahmed Siddiqui, Ishtiaq Ahmad Bajwa, Muhammad Ather Elahi and [1]**Syed Muhammad Fahim**
University of Dammam, SAUDI ARABIA
**E-mail: KASiddiqui@uod.edu.sa*
[1]DHA Suffa University, Karachi, PAKISTAN
***Email: SMFahim@dsu.edu.pk*

ASTRACT: *This paper aims to present the abridged guidelines for the usage of various measurement scales in social research. Selection of measurement scales play a pivotal role in social science research and most of statistical techniques and tests require a specific measurement scales to be employed in the research. This paper starts with a summary of measurement scales and then provides useful guidelines for permissible arithmetic operations, permissible descriptive statistics, and permissible graphical presentations. It also provides summary of statistical tests for independent and dependent variables containing different measurement scales and permissible uni-variate, bivariate and multivariate statistical analysis for different measurement scales. At the end it provides limitations of this study.*

I. INTRODUCTION

In this paper we have made an attempt to give one-stop solution for the permissible statistical test depending on the type of data and purpose of analysis. We discuss important characteristics of the four measurement scales that must be well-understood before choosing any statistical test and making inference. The paper makes a valuable contribution in terms of a checklist for social scientist about the selection of measurement scale and the permissible arithmetic operations, descriptive statistics, graphical presentations, statistical tests in the context of dependent / independent variables and multivariate analysis. We expect researchers from disparate fields like psychology, medicine, marketing, anthropology etc. are major beneficiaries of this contribution.

II. LITERATURE REVIEW

The measurement remains a critical factor in any scientific experiment. It encompasses the assignment of verifiable identities (mainly numerals) to various outcomes according to rules. As the numbers are labelled under different set of rules therefore measurement scales varied in terms of their coverage and usage. This is why it is important to understand and correctly interpret the difference in general characteristics, mathematical properties, nature of data and statistical operations of the measurement scales [1].

This study is based on the seminal contribution of Stevens [2], who classified the measurement scales into four categories: Nominal, Ordinal, Interval and Ratio scales. The classification is based on the properties of numbers and how these numbers are measured and interpret in the context of social experiments. This is important because the statistical inference only make sense if it is applied properly on relevant data with known statistical properties. Therefore, there is a need to have one-stop solution regarding the do's and don'ts of statistical analysis in social research with respect to the scale of measurement.

Despite weaknesses in the use of permissible statistical analysis on Stevens' classification of measurement scales, it provides basis for scientific reasoning and hence cannot be ignored [2]. The debate of the relationship between scales of measurement and appropriate statistics is still continued even after seven decades of the seminal publication of Stevens. The simple typology was further explained by illustrating the permissible statistical procedures for all measurement scales [3]. Statistical procedures were developed for univariate and multivariate analyses employing various scales of measurement [4]. Notwithstanding the usefulness of Stevens' classification, it is cautioned that data analysis in all social experiments should be guided by the relevant hypothesis, not by the rigid adherence to Steven's classification [5]. Even for the selection of most appropriate statistical tests for the hypothesis under consideration, the meaningfulness of statistical inference is more important than the mere representation or uniqueness of the method [6].

Although the measurement theory has gained roots in interpreting the statistical parameters and make inferences, but a blind faith on Stevens' typology for all statistical analysis cannot be recommended. The automated decisions to pick a statistical model based on the measurement scale of available data may lead to wrong assertions [7]. Therefore, a cautious approach to put all available information in perspective with respect to the research question(s) before choosing the right statistical test. The application of Stevens' measurement scales in statistical computer packages was also assured [8].

III. METHODOLOGY

Experienced researchers developed heuristics for quickly applying the statistical methods based on their experiences. These heuristics are termed as rules-of-thumb for particular research method, technique or issue [9,10,11,12].

IV. ANALYSES

This paper offers reliable literature for understanding measurement scales followed by the use of permissible arithmetic operations, descriptive statistics, graphical presentations and statistical tests for each type of measurement scale; then it discusses the permissible statistical tests for different combinations of measurement scales within the dependent and independent variables; and finally it presents the permissible multivariate analysis available for each measurement scale.

Heuristic # 1 Understanding the Hierarchy of Measurement Scales

The social science literature mainly span on four measurement scales, each illustrate a different set of information. They all use numbers at the core level but the interpretation of these numbers varies from scale to scale. The four scales are nominal, ordinal, interval, and ratio. Table

1 summarizes the nature of data, characteristics and examples of the four measurement scales [13].

Nominal scale is the most simplistic scale in which various outcomes are labelled with some identification marks (may be numbers or any other alphabets). These labels are mutually exclusive and without any numerical significance on each other. In other words we can say numbers/ alphabets are used for identification purpose only and have no real quantitative value. Examples of nominal scale parameter include Employees ID, National ID number, Bank Account number, Gender, colour of skin/hair, nationality etc.

Ordinal scale of measurement has an edge over the nominal scale as it allows data to be ordered. It classifies the preference of one data value on the other; however, the quantum of this preference is not known. For example, socioeconomic class of inhabitants in an economy can be classifies as high, medium, or low but the exact difference between these classes is not quantifiable. Also it is not necessary that successive classifications are equidistant from each other. Other examples of ordinal scale classification could be the top ten tourist destinations in Europe, satisfaction level of medical students in the graduating class or an individual's state of happiness. These rankings allow us to order the positions of successive sample points but the extent of their differences is not known. The ordinal scale works perfect with non-numeric qualitative data that measure the relative position, but the difference between them is not measurable.

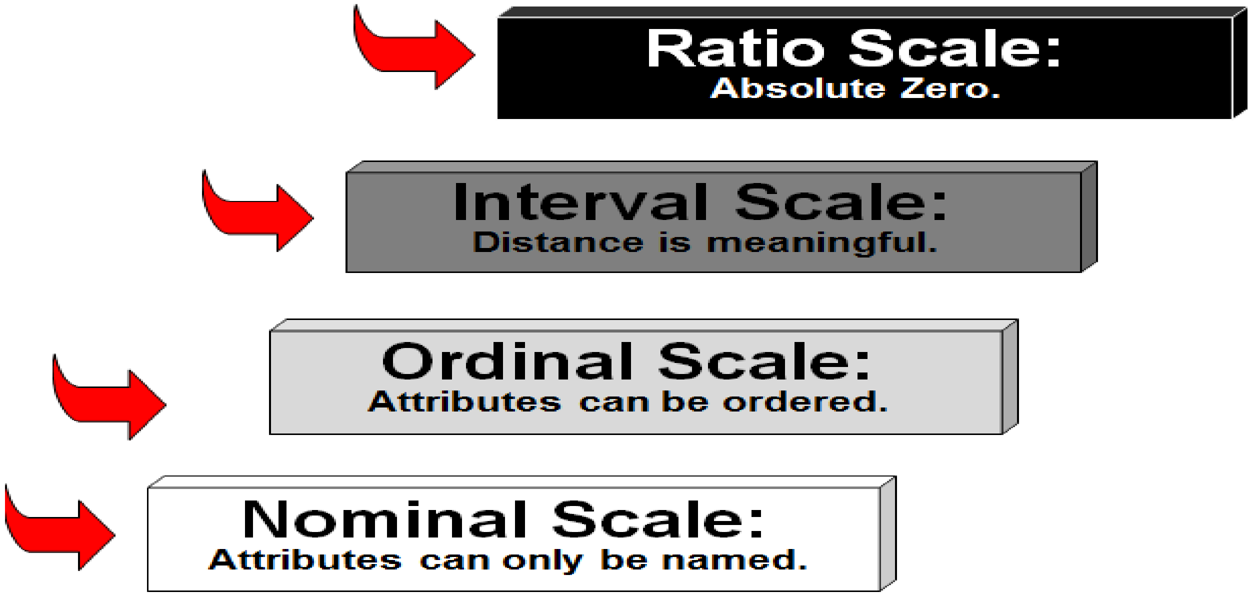
Next in line is the interval scale, which has identifiable value that can be ordered and also the difference between these values is measurable. The difference between the successive values is called an interval that is same across all values. For example, time is measured on interval scale wherein increments are known, consistent and measurable. Thus we can measure precisely how much more time is needed for minute-hand of the clock to move from point 1 to 2 and so on. Similarly temperature is measured on interval scale because the difference between each value is precise and measurable. We can apply statistical analysis on interval scale and measure mean, median, mode, standard deviation etc. Interval scale though solve the issue of measurability but suffers with a deficiency of not having a true zero value. For example, there is no such thing as any temperature; therefore, it is not possible to compute ratios in interval scale [14].

Finally the ratio scale is the ultimate scale that has all the characteristics of the nominal, ordinal, and interval scales besides having a true zero value. On a ratio scale, zero means none that indicates a complete absence of the thing measured. The scale is represented by quantitative data that can be added, subtracted, multiplied or divided (ratios) is a meaningful way. For example, we can use statistical analysis on variable like height, weight, age, sales, income, cost etc. Ratio scales offers opportunity to apply both descriptive as well as inferential statistics at its potential. We can measure central tendency by mean, median, or mode; whereas dispersion can be measured by standard deviation or coefficient of variation from ratio scales.

Table 1 Understanding Measurement Scales.

Scale	Nature of Data	Characteristics	Examples
Nominal	Non-Parametric/ Qualitative	Numbers to identify and classify sample points.	Employee ID, National ID #, Bank Account #, Players #
Ordinal	Non-Parametric/ Qualitative	Numbers to order the positions of successive sample points but the extent of their differences is not known	Ranking of teams, ranking of students in class, Market position, Social Class, Preference ranking, Satisfaction levels
Interval	Parametric/ Quantitative	Measures also the magnitude of difference between successive sample points	Altitudes, Temperature, Time
Ratio	Parametric/ Quantitative	True zero value exist; Ratios of sample points are comparable	Age, Sales, Income, Cost

The inherent hierarchy of measurement scales is shown in Figure 1. Nominal scales are placed at the bottom of the hierarchy as attributes can only be named. Second level of hierarchy is measured by ordinal scales where attributes can be named and ordered in a particular sequence. Third level is occupied by the Interval scales where attributes can be named and ordered in a particular sequence and the distance between two items is meaningful. Highest level of hierarchy is occupied by the Ratio scales where absolute zero makes the measurement perfect. .



Heuristic # 2 Measurement Scales & Permissible Arithmetic Operations

The differences among the four scales of measurement in terms of arithmetic, nature and permissible arithmetic operations are highlighted in Table 2. Since all four measurement scales identifies objects in a distinct manner, they can be used for comparison purpose. Therefore, equality/inequality operations can be applied in all measurement scale with ease. As nominal scale is the mere labelling of distinct objects we cannot use nominal scale for ordering/ranking or any other arithmetic operations. Although ordinal scale allows us ordering/ranking, it is restricted in terms of interval quantification with respect to the characteristic being measured. This inability to determine differences between objects prohibits the use of any arithmetic operation on an ordinal scale. Interval scale is an improvement on ordinal scale with verifiable interval quantification, thus allows for addition/subtraction; however, the non-existence of true zero value limits its usage for multiplication/division of sample points. Finally, the ratio scale solves the mystery of zero-value existence and hence the ultimate scale that permits all arithmetic operations [16].

Table 2 Measurement Scales & Permissible Arithmetic Operations

Measurement Scale	Inequality	Ordering/ Ranking	Interval Quantification	Addition/ Subtraction	True Zero Value	Multiplication/ Division
Nominal	Yes	No	No	No	No	No
Ordinal	Yes	Yes	No	No	No	No
Interval	Yes	Yes	Yes	Yes	No	No
Ratio	Yes	Yes	Yes	Yes	Yes	Yes

Heuristic # 3 Measurement Scales & Permissible Descriptive Statistics

Typically, social science experiments are based on sufficiently large data to implement statistical analysis and conduct hypothesis testing. In the process we need to summarize trends, estimate dependence and make inference from available data for the unknown population. For the first step, we use descriptive statistics to understand the shape and form of the distributions (e.g., mode, median, mean, standard deviation). Then we conduct regression/factor analysis to estimate dependence through functional relationship and lastly, we use statistical tests for getting inference from the available information.

Table 3 Measurements Scale & Permissible Descriptive Statistics

Measurement Scale	Count	Mode	Median	Mean	Standard Deviation	Sum
Nominal	Yes	Yes	No	No	No	No
Ordinal	Yes	Yes	Yes	No	No	No
Interval	Yes	Yes	Yes	Yes	Yes	Yes*
Ratio	Yes	Yes	Yes	Yes	Yes	Yes

* With caution

Table 3 shows the types of descriptive statistics that can be used for each type of measurement scale. The frequency distribution and mode can be used with all measurement scale. Since nominal scale has no quantitative meaning therefore no other statistics is applicable in this case. Ordinal scale has defined ranking so median is useful but others statistics are not available. Other two scales can handle all statistics, but we need to be cautious in summing sample points in interval.

Heuristic # 4 Measurement Scales & Permissible Graphical Presentation

Graphical presentation is very effective tool in any univariate analysis. In table 4 we mention the compatibility of various graphical tools with respect to measurement scales. Bar charts and Pie charts are commonly used graphs for comparison purpose. Though these graphs can work in all four measurement scales but it is more recommended for categorical scales (i.e. nominal and ordinal scale). Pie chart can also work with interval scale with limited applications; however, it is not suitable for ratio scale due to existence of the true zero value[17]. Stem and Leaf plot is used to identify typical values and their spread (i.e. modes), the distribution of data with gaps and outliers. Nominal scale cannot be presented using Stem and leave graph; however; it is applicable to all other measurement scales. Both the line chart and histogram can be used in interval and ratio scales and are useful in depicting statistical properties of the data [18].

Table 4 Measurement Scales & Permissible Graphical Presentation

Measurement Scale	Bar Chart	Pie Chart	Stem and Leaf	Line Chart	Histogram
Nominal	Yes	Yes	No	No	No
Ordinal	Yes	Yes	Yes	No	No
Interval	Yes	Yes	Yes	Yes	Yes
Ratio	Yes*	No*	Yes	Yes	Yes**

* (Only if Discrete) ** (Only if Metric)

Heuristic # 5 Measurement Scales & Permissible Statistical Tests

Validity of statistical inference in social experiments is crucially dependent on the choice of correct testing technique. Table 5 shows the statistical tests that are available for each measurement scale. Due to the nature of data, we have limited choice available for nominal scale; however, testing methods increases as we move towards the ratio scale. As nominal scale is dealing with non-parametric (category) variables we need fewer assumptions about sampling process. For dichotomous data in nominal scale (i.e. for two classifications e.g. all school age children go to school: yes or no) we use binomial test if there is one sample, whereas Z test for two samples. But if there are more than two classifications in a one sample we have to use Chi-square test and compare the proportion of observed occurrences with the theoretically expected value under null hypothesis. Chi-square is also valid for nominal data in case of two or more than two but mutually independent samples. For paired (related) samples we use McNemar for two samples and Cochran Q test when there are more than two samples [19].

Table 5 Measurement Scales & Permissible Statistical Tests

Measurement Scale	One Sample	Two Samples		More than 2 Samples	
		Independent	Paired	Independent	Paired
Nominal	**Binomial Chi-Square**	**Z test Chi-Square**	**McNemar**	**Chi-Square**	**Cochran Q**
Ordinal	Kolmogorov-Smirnov	Mann-Whitney test	Wilcoxon matched-pairs sign-test	Kruskal-Wallis test	Friedman test Kendall-coefficient of concordance
Interval	z test [if N>30]; t test [if N<30]	z test [if N>30]; t test [if N<30]	t test for related groups	ANOVA	ANOVA
Ratio	z test [if N>30]; t test [if N<30]	z test [if N>30]; t test [if N<30]	t test for related groups	ANOVA	ANOVA

Statistical significance in ordinal data can be analysed using non-parametric Kolmogorov-Smirnov test for the univariate (one sample) case. Kolmogorov-Smirnov test is considered more powerful than the chi-square test for univariate analysis in ordinal data. In the hypothesis testing of differences in the rankings (ordinal scale) within two independent samples we can use Mann-Whitney U test [20]. Whereas Wilcoxon matched-pairs signed-rank testis more appropriate for paired samples (or repeated measurements in a single sample) to assess the difference in mean-ranks of respective populations [21]. In comparing the difference in the sum of ranks for more than two independent samples the Kruskal-Wallis H test is a better choice for ordinal data. In case of more than two correlated samples Friedman test is more appropriate to determine whether the total ranks in two or more groups are significantly different. The Friedman test is also applicable in non-parametric analysis of repeated sample points where ANOVA cannot be applied [22]. A normalized version of Friedman test is Kendall's coefficient of concordance whose value lies between 0 (i.e. no agreement) and 1 (i.e. complete agreement).

The more powerful parametric tests (i.e. z-test, t-test and ANOVA) are available for interval and ratio scales. The efficacy of these tests is, however dependent upon the normality of the distribution of variables. z-test can be used in one or two samples when $N \geq 30$ else we use t-test in case of interval and ratio scales. If the two samples are correlated then we use t-test in both interval and ratio data. The Paired Samples t-tests is used for comparing means of sample points from the same population at two different times (e.g. before and after intervention) or two different but related conditions or units (e.g. male and female). We test the null hypothesis that the sample means are same in both cases [23]. Analysis of variance (ANOVA) is the ultimate choice for more than two samples in both types of measurement scales.

Heuristic # 6 Measurement Scales as Dependent/ Independent Variables

Table 6 provides a general guideline for permissible statistical tests for different combinations of measurement scales as dependent and independent variables. In case of a nominal type dependent variable whereas the independent variables variable is either nominal or ordinal the most common statistics to be used is Chi-square. Similarly the regression techniques, either linear or multiple, are used with interval or ratio scale type measurement on both sides of the equation. On the other hand, analysis of variance is more appropriate when dependent variable belong to interval or ratio scale and independent variables is measured on nominal or ordinal scales [24,25,26].

Heuristic # 7 Measurement Scales & Permissible Multivariate Analysis

The table 7 discusses possible multivariate analyses; we may use for different types of data categories. The table provides a general guide for what type of multivariate analyses should be considered based on the nature of data. It does not mean that other analyses could not be done but simply that the usual analyses are the ones that are listed. One can observe that all given multivariate analysis are applicable to interval or ratio scale because in these two categories the nature of data is quantitative [31,32,33,34].

Table 6 Measurement Scales as Dependent/ Independent Variables

Dependent Variable	Independent variable	Permissible Statistics Test
Nominal	Nominal	Chi-square
Nominal	Ordinal	Chi-square, Mann-Whitney, Poisson regression
Nominal	Interval/Ratio	Binominal Distribution, Logistic Regression, Discriminant Analysis, Poisson regression
Ordinal	Nominal	Chi-square, Fisher's exact test
Ordinal	Ordinal	Chi-square, Spearman rho
Ordinal	Interval/Ratio	Poisson regression, Logistic Regression, Discriminant Analysis
Interval /Ratio	Nominal	ANOVA, Multivariate analysis of Variance
Interval /Ratio	Ordinal	Pearson r, Multiple Regression, Independent – Sample t Test
Interval /Ratio	Interval/Ratio	Linear Regression, Multiple Regression, Correlation, Independent – Sample t Test

Judgment will be called for when the investigator has, for example, five independent variables, three of which are interval, while one is ordinal and one is nominal, with one dependent variable that is interval. Most investigators would use multiple regressions in this case. Many other tests are given based on different pairs of dependent and independent variables [27 28 29,30]. However, due to qualitative nature of data, in case of nominal and ordinal measurement, most of these multivariate analyses are not applicable. But one must consider the following exceptions.

In case of limited dependent variable we use logistic regression model (i.e. logit/probit) if the data has binary choice and multinomial logistic regression for more than two possible values for dependent variable on nominal scale. While dealing with ordinal data, Poisson regression model can be used provided that the dependent variable follows Poisson distribution. The data used in cluster analysis can be interval, ordinal or categorical. Categorical data, which is a group of information collected, may include the data categories made on the basis of preferences, groups and yes or no responses etc. [35]. Similarly, discriminant analysis may be used for nominal data and ordinal data [36].

Table 7 Measurement Scales & Permissible Multivariate Analysis

Measurement Scale	Regression Analysis	Factor Analysis	Cluster Analysis	Discriminant Analysis	Correspondence Analysis	MDS
Nominal Scale	No*	No	Yes	Yes	No	No
Ordinal Scale	No*	No	Yes	Yes	No	No
Interval Scale	Yes	Yes	Yes	Yes	Yes	Yes
Ratio Scale	Yes	Yes	Yes	Yes	Yes	Yes

*Regression analysis allows the use of nominal and ordinal scale in independent variable through a dummy variable

V. LIMITATIONS

Joint exposure of authors is limited to social sciences only, therefore researchers from other discipline must also consult other similar literature relevant to their research area and their research needs. Secondly this research paper only provides the guidelines for measurement scales usage for statistical techniques rather than mathematical computation.

VI. IMPLICATIONS

Social scientists are always perplexed about the application statistical techniques. Most often they have designed the survey instrument without keeping any specific statistical technique in mind, and after data collection they found out that the measurement scales were not appropriate for their desired statistical technique. This research paper provides guidelines for naïve researchers and social scientists to decide upon the measurement scales even at the time of questionnaire design. This research paper also highlights the limitations of various measurement scales for application of statistical techniques.

REFERENCES

[01] Stevens, S.S., (1946) On the theory of scales of measurement. Bobbs-Merrill, College Division.

[02] McCrum-Gardner, E., (2008). Which is the correct statistical test to use? British Journal of Oral and Maxillofacial Surgery, 46(1): p. 38-41.

[03] Stevens, S.S., (1951) Mathematics, measurement, and psychophysics, In Stevens (ed.): Handbook of experimental psychology. NY: Wiley.

[04] Andrews, F.M., Klem, L., Davidson, T.N., O'Malley, P.M., Rodgers, W.L., (1981). A guide for selecting statistical techniques for analyzing social science data 2nd ed, University of Michigan ..

[05] Velleman, P.F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. The American Statistician, 47(1): p. 65-72.

[06] Lord, F.M. (1953). On the statistical treatment of football numbers. American Psychologist, 8, pp: 750-751

[07] Velleman, P.F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. The American Statistician, 47(1): p. 65-72.

[08] Velleman, P.F. and D.C. Hoaglin, (1981) Applications, basics, and computing of exploratory data analysis. Vol. 142. Duxbury Press Boston.

[09] Siddiqui, K., (2013). Heuristics for Sample Size Determination in Multivariate Statistical Techniques, World Applied Sciences Journal, 27(2): pp 285-287

[10] Siddiqui, K.A., (2015). Heuristics of using Factor Analysis in Social Research, Science International (Lahore), 27(5), 4625-4628

[11] Siddiqui, K.A., (2015). Heuristics of using Structural Equation Modeling in Social Research, Science International (Lahore), 27(6), 6381-6384

[12] Siddiqui, K., Fahim, S.M., & Mirani, M.A., (2015). Model Generation using Structural Equation Modeling, Journal of Scientific Research and Development, 2(10): pp 112-116

[13] Gaito, J., (1980) Measurement scales and statistics: Resurgence of an old misconception. Psychological Bulletin, 87(3), pp: 564-567.

[14] Salkind, N.J.,(2012) Statistics for people who (think they) hate Statistics: Excel 2010 Edition, Sage

[15] Trochim, W.M., (2016) Levels of Measurement page from Research Methods Knowledge Base by : http://www.socialresearchmethods.net/kb/measlevl.htm.

[16] Koul, L., (2009) Methodology Of Educational Research: Vikas Publishing House Pvt Ltd.

[17] Babbie, E., Wagner W.E., and Zaino, J. (2016) Adventures in social research: Data analysis using IBM® SPSS® Statistics. 2015: Sage Publications.

[18] Neill, J., Summary of The Four Levels of Measurement: Appropriate Descriptive Statistics and Graphs. http://www.wilderdom.com/research/Summary_Levels_Measurement.html, 2009.

[19] Bower, J.A., (2009) Statistical Methods for Food Science: Introductory procedures for the food practitioner. 2009: John Wiley & Sons.

[20] Whitney.,U.M., (2016) Testing differences with ordinal data. http://www.harding.edu/sbreezeel/460%20files/statbook/chapter12.pdf, Chapter 12.

[21] SPSS Tutorials, (2016) Mann-Whitney U, Sign Test, and Wilcoxon Tests, . SPSS Help.

[22] Privitera, G.J., (2015) Essential Statistics for the Behavioral Sciences. Sage Publications.

[23] SPSS Tutorials, (2016) Paired samples t test retreived from http://www.spss-tutorials.com/spss-paired-samples-t-test/.

[24]Sirkin, R.M., (2006) Statistics for the social sciences. Sage Publications .

[25]Tukey, John W.: (1977) Exploratory Data Analysis. Addison-Wesley Publishing Company

[26] Hayes, B.E., (1998) Measuring customer satisfaction: Survey design, use, and statistical analysis methods. ASQ Quality Press.

[27] Afifi, A., May, S., & Clark, V. A. (2011). Practical multivariate analysis. CRC Press.

[28] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., (2010). Multivariate Data Analysis, 7th ed. NJ: Prentice Hall

[29] Meyers, L.S., Gamst, G., Guarino, A.J., (2006). Applied Multivariate Research: Design and Interpretation, Sage Publications

[30] Tabachnick, B.G. & Fidell, L.S. (1996). Using Multivariate Statistics. 3rd ed. NY: Harper-Collins

[31] Babbie, E., Wagner W.E., and Zaino, J. (2016) Adventures in social research: Data analysis using IBM® SPSS® Statistics. 2015: Sage Publications.

[32] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., (2010). Multivariate Data Analysis, 7th ed. NJ: Prentice Hall

[33] Meyers, L.S., Gamst, G., Guarino, A.J., (2006). Applied Multivariate Research: Design and Interpretation, Sage Publications

[34] Tabachnick, B.G. & Fidell, L.S. (1996). Using Multivariate Statistics. 3rd ed. NY: Harper-Collins

[35] Rosie Cornish, (2016) Cluster Analysis. Mathematical learning support centre, at http://www.statstutor.ac.uk/resources/uploaded/clusteranalysis.pdf, 2007.

[36] Mertens, D.M., (2014) Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods. Sage Publications.