

SPSS: An Imperative Quantitative Data Analysis Tool for Social Science Research

Arifa Rahman¹, Md. Golam Muktadir²

¹Ph.D. Research Fellow, Dr. Wazed Research and Training Institute, Begum Rokeya University, Bangladesh

²Independent Researcher

Abstract:- The purpose of this paper is to elaborate on the importance of the Statistical Package for the Social Sciences, widely known as SPSS in the field of social sciences as an effective tool for quantitative data analysis. This paper includes the broader aspects of SPSS from its foundation as a statistical analysis tool to how it has advanced as a major choice for researchers from different fields of study. It is necessary for researchers, especially new researchers to know the ins and outs of SPSS as to why they should use it. This study expressed the arguments from personal experiences of using SPSS and what other SPSS users have described in various pieces of literature. Different works of literature related to SPSS and social sciences have been discussed throughout the paper. The most relevant ones have been talked about. The features and major advantages are portrayed in such a way that anyone can relate the tool with their research works. In the end, it has been evident that SPSS is considered one of the most important and influential statistical tools for quantitative data analysis.

Key words: SPSS, Quantitative Data Analysis Tool, Data Analysis Tool, Statistical Tool for Data Analysis, Research Tool

I. INTRODUCTION

Technology has taken an important place in social science research. Earlier researchers had to do a lot of analytical jobs for analyzing quantitative data. With the same thing in mind, *Statistical Package for the Social Sciences* widely known as SPSS was initially developed by Norman H. Nie, a Social Scientist himself, along with his two fellows, Dale H. Bent and C. Hadlai Hull in 1968 at the Stanford University (McCormick et al., 2016). The software program was later acquired by IBM in 2009 and is now called IBM SPSS (Horber, n.d.). With the ease of use features of statistical analysis for ordinary researchers in mind, the initial manual for SPSS written by the developers themselves has been described as one of the most influential books of the sociology of all time (Wellman, 1998). Other significant features of SPSS include data management and data documentation. SPSS is a powerful and user-friendly software package for all sorts of statistical analysis of data (Levesque, 2007). This program is mostly used by students and researchers from the fields of sociology, psychology, economics, business studies, medicine, engineering, and other disciplines. Apart from that, various public, private, and non-government organizations also use SPSS for their various projects. SPSS is a strong choice for marketing and survey companies for analyzing consumer behavior and forecasting (Vorhies, 2017).

Social science researchers are often needed to deal with a large volume of data. Primarily these data are collected from online or offline surveys, interviews, group discussions, or observations. The biggest advantage of SPSS is that the program is designed to handle a large set of data with multiple variables associated with it (Jasrai, 2020). It also has all the flexibilities of multiple analyses of data along with graphical representation (Garth, 2008).

In today's world data means everything and people are more inclined towards using computer software for any kind of data analysis. Data analysis software should have all the basic features that an analyst requires. There are several computer software packages available for quantitative data analysis, but SPSS stands out due to its usability and superlative features that bound a researcher to use it even after having some free alternatives available in the market (Arkkelin, 2014).

II. METHODOLOGY

No specific fieldwork is done for the study. This paper is designed with a logical explanatory model that has been adopted from various empirical data and available secondary data sources related to SPSS and data analysis. In particular, our objective guided us for such a methodological approach. The primary data source relies on the authors personal experiences of using SPSS for different social-science related research works done before. Any other methodological approach was not suitable for such suggestive work. The secondary data sources like journal articles, seminar papers, and other related publications selected to advance this study are considered based on the properties such as authentic data source, accuracy and consistency of data, and author's background. The fair intention of the study was to develop direct explanations from the available data.

III. STATISTICS AND QUANTITATIVE ANALYSIS

According to Upton and Cook (2014), statistics is the practice that collects, organizes, and analyzes numerical data and further interprets and presents it to be applied in scientific, industrial, or social problems. Statistics in another way is a mathematical procedure to interpret required information from gathered data, especially for using a group of samples from those in a representative population (La Trobe, 2020).

Quantitative analysis is a statistical method of analyzing numerical data collected online or offline, polls, questionnaires, and surveys, or by considering pre-existing

statistical data more mathematically with the help of computational techniques. The analyzed findings further generalized across different groups, or explain a particular phenomenon (Muijs, 2010; Babbie, 2020).

In social science research, within a given population, quantitative analysis is used to determine the relationship between an independent variable and a dependent variable. Values of the independent variable can be manipulated; more precisely values can be controlled or changed to observe the effect on the dependent variable (Niño-Zarazúa, 2012).

IV. SPSS: THE TOOL FOR STATISTICAL ANALYSIS OF QUANTITATIVE DATA

In today's world, quantitative data analysis is done by most social science researchers for their studies. The list of software packages to do such analysis is not small either. R, PSPP, Stata, SAS, SPlus, JASP, and BMDP are few general-purpose statistical analysis software packages available in the market (NYU, 2021). Some of these are open source and the rest are under the commercial license. Excel, widely used spreadsheet software by Microsoft also allows quantitative data analysis through an add-on module. Though these choices are not so user-friendly, but available alternatives to SPSS. Most educational and non-educational institutes use SPSS due to its user-friendly features (Ong and Puteh, 2017). It is not necessary that SPSS is the best tool out there for quantitative data analysis but it has some advantages over other available software packages. Even some other packages have features that are way better than SPSS. SPSS is available for all three major computer platforms, i.e. Windows, Mac OS, and LINUX (URI, 2019).

V. FEATURES OF SPSS

The first and foremost feature of SPSS is that it has been developed for people with non-technical backgrounds especially social sciences, so no need for prior knowledge for a programming language is required to start with the program. As a result, the program is not only user-friendly but also easily applicable in any kind of quantitative analysis (Arkkelin, 2014). Its ability of data management is another great feature that makes it a top-notch choice for researchers around the world. The program has exceptional report generation capabilities (Wagner, 2019). Comprehensiveness and flexibility in quantitative data analysis are a must feature to choose this program upon any other alternatives (Aljandali, 2016). Authorization, collaboration, and deployment of survey data for strategic data mining for statistical analysis are reasons to use SPSS in social science researches. By knowing basic concepts of operating SPSS, it becomes easier for the researcher to use it for the effective analysis of quantitative data. Thus it helps the researcher to face any difficulties during the process. SPSS requires defining a set of variables and further create cases by suitable input of data within these variables. There are mostly four types of variables in the SPSS datasheet. These are independent variables, dependent variables, intervening

variables, and moderator variables (Bala, 2016). In laymen's terms, an independent variable is a cause. Its value is independent of any other variables in a study. Whereas, a dependent variable is an effect, whose value depends on any changes in the independent variable. In the case of an intervening variable (sometimes called mediating variable), it refers to a hypothetical variable, that is usually used to explain causal links between other variables within research. Lastly, the moderating variable is that variable that can alter the association between independent and dependent variables. Identifying variables is very important.

VI. ADVANTAGES AND DISADVANTAGES OF SPSS

There are various advantages of SPSS as far as its usability is concerned. The biggest advantage of using this software is its Graphical User Interface (GUI) and thus easy to learn and use (Landau and Everitt, 2004). It has in-depth statistical capacity. It has the unique quality of creating variables from existing information (MacInnes, 2016). Most of the complex statistical tests are available built-in. One can open a variety of file formats in SPSS such as Excel, SAS, Stata, tab-delimited, and few other known formats. In today's world platform compatibility for software is a huge issue and here also SPSS has all the advantages over other statistical packages. SPSS runs on Windows, macOS, and LINUX platforms (URI, 2019). As a result, researchers can work on their data irrespective of the computer system they are using. SPSS is a wonderful choice for cluster analysis as well (Field, 2013, 2017). SPSS addresses all facets of the analytical process from data preparation and management to data analysis and reporting. The use of regression and expectation-maximization also improve forecasts and plans by imputing missing values with expected values. The program has features that provide automated methods to identify anomalies and statistical transformations to address outliers if any. Tables and effective visualizations of results are unmatched features of SPSS. It can also predict values of target variables based on values of predictor variables. Accurate modeling of linear and non-linear relationships enables another door for researchers (IBM, 2020). Moreover, SPSS is a complete package for Quantitative Data Analysis.

Like all other software, in addition to advantages, SPSS has few disadvantages too. The first disadvantage is the commercial license of the program which means you have to pay for the software to use it. Even students are not going to get the program for free. Discounts are available but it has some cost attached to them. In some cases especially for regular MS Excel users, the usability might not feel the same. Editing the graphs is a bit difficult. Not many options for graphs as well. Graphic quality could also be a major nay. The output is not easy to edit (UT Austin, 2012). The program can run slow depending on the machine it is installed. Certain add-on modules are not as easy as they should be.

VII. TYPES OF SPSS

The stable version of IBM SPSS 28 has three different types of licensing systems. Subscription-based which does not require an authorization code, the second one is non-subscription based which requires an authorization code and also comes with one-year technical support, and the final one is the academic license (IBM, n.d.). Some of the new features of SPSS 28 are meta-analysis, power analysis, ratio statistics, relationship maps, statistics workbook, search, table side-pane editor, and high contrast support (IBM, 2021). SPSS 28 is the most stable SPSS version available on the market at present and has the most number of tests that can be performed.

List of few major tests that can be done using SPSS

Analysis of covariance	Factorial ANOVA	Multivariate multiple regression	Paired t-test
Binomial test	Factorial logistic regression	Non-parametric correlation	Repeated measures logistic regression
Canonical correlation	Fisher's exact test	One sample median test	Simple linear regression
Chi-square goodness of fit	Friedman test	One sample t-test	Simple logistic regression
Chi-square test	Kruskal Wallis test	One-way ANOVA	Two independent samples t-test
Correlation	McNemar test	One-way MANOVA	Wilcoxon signed rank sum test
Discriminant analysis	Multiple logistic regression	One-way repeated measures ANOVA	Wilcoxon-Mann-Whitney test
Factor analysis	Multiple regression	Ordered logistic regression	

Diagram 01: List of major tests that can be done using SPSS (Mehta and Patel, 1989)

VIII. CONCLUSION

In recent years, it has been noticeable that researchers in the field of social sciences have demonstrated a strong involvement of various software programs in their regular methodical research activities, whether it is for writing their research paper or thesis, searching relevant content online, or doing any sort of analytical works. Realizing the same, IBM SPSS or Statistical Package for the Social Sciences is one of the most used statistical analysis software by social scientists all over the world. After its inception fifty-odd years ago, it has gone through several changes according to the need of social science researchers. The need for accuracy and perfection in data representation in quantitative data analysis forced its creators to make it not only feature-rich but user-friendly as well. Within this phase, it has also evolved as a major data analysis tool for other genres of studies too. The dimensions of its use are so vast that people from all different fields ranging from medical, engineering, business, education, the government is using SPSS for their statistical analysis needs. The scopes of using a program like SPSS are huge. SPSS has the quality of digging deeper into analyst's data and making it a much more efficient instrument than other available spreadsheets, databases, or standard multi-dimensional tools. SPSS Statistics excels at drawing conclusions and predictions of user data. It is indeed one of the fastest in handling tasks like data manipulation and statistical procedures as compared to many non-statistical programs. Along with the statistical analysis, its features also cover data management and documentation. Enabling researchers to achieve any complex analyses and further plotting the same to various charts and graphs available within the program. Its user-friendly interface, online tutorials, and available documentation from the parent company and online user

communities have made it one of the most easy-to-learn statistical analysis software available in the market. In spite of being having few disadvantages, its unique features make it outstood in comparison to other statistical tools for quantitative data analysis. SPSS is capable of conducting all major tests required for quantitative data analysis in the field of social sciences. With all that being said, in today's time, realizing the need, it is not only the choice but in some cases considered essential for social researchers to use SPSS as their quantitative data analysis and representation tool.

REFERENCES

- [1] Aljandali, A. (2016). Quantitative analysis and IBM® SPSS® statistics: A guide for business and finance.
- [2] Arkkelin, D. (2014). Using SPSS to Understand Research and Data Analysis. Psychology Curricular Materials.
- [3] Babbie, E. R. (2020). The practice of social research. Cengage AU.
- [4] Bala, J. (2016). Contribution of SPSS in Social Sciences Research. International Journal of Advanced Research in Computer Science, 7(6), 250-254.
- [5] Field, A. P. (2013). Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll (4th ed.). SAGE.
- [6] Field, A. P. (2017). Cluster Analysis.
- [7] Garth, A. (2008). Analysing data using SPSS. Sheffield Hallam University.
- [8] Horber, E. (n.d.). History of SPSS. Université de Genève.
- [9] IBM. (2020). IBM SPSS Statistics: What's New.
- [10] IBM. (2021). What's New in SPSS Statistics 28.
- [11] IBM. (n.d.). IBM SPSS Statistics: Features and modules.
- [12] Jasrai, L. (2020). Data analysis using SPSS. Sage Publications Pvt.
- [13] La Trobe. (2020). Statistics. LibGuides at La Trobe University.
- [14] Landau, S., & Everitt, B. S. (2004). A handbook of statistical analyses using SPSS. Chapman and Hall/CRC.
- [15] Levesque, R. (2007). SPSS programming and data management: A guide for SPSS and SAS users.
- [16] MacInnes, J. (2016). An introduction to secondary data analysis with IBM SPSS statistics. SAGE.
- [17] McCormick, K., Salcedo, J., & Poh, A. (2016). The origin of SPSS statistics. dummies.com.
- [18] Mehta, C. R., & Patel, N. R. (1989). IBM SPSS Exact Tests (2011 ed.). IBM Corp.
- [19] Muijs, D. (2010). Doing quantitative research in education with SPSS. SAGE.
- [20] Niño-Zarazúa, M. A. (2012). Quantitative analysis in social sciences: A brief introduction for non-economists. SSRN Electronic Journal.
- [21] NYU. (2021). Quantitative analysis guide: Which statistical software to use?. Research Guides at New York University.
- [22] Ong, M. H. A., & Puteh, F. (2017). Quantitative Data Analysis: Choosing Between SPSS, PLS and AMOS in Social Science Research.
- [23] Upton, G., & Cook, I. (2014). A dictionary of statistics (3rd ed.). Oxford University Press.
- [24] URI. (2019). Minimum system requirements for SAS and SPSS software. University of Rhode Island.
- [25] UT Austin. (2012). SPSS: Displaying data for windows. The Department of Statistics and Data Sciences, The University of Texas at Austin.
- [26] Vorhies, B. (2017). SPSS Statistics to Predict Customer Behavior [Webinar]. Data Science Central.
- [27] Wagner, W. E. (2019). Using IBM® SPSS® statistics for research methods and social science statistics (7th ed.). SAGE.
- [28] Wellman, B. (1998). Doing It Ourselves: The SPSS Manual as Sociology's Most Influential Recent Book. In Required Reading: Sociology's Most Influential Books, edited by Dan Clawson (pp. 71-78). Amherst: University of Massachusetts Press.