Importance of R program and SPSS program, which is required in Biostatistics for Analysing in public health.

MD.Shahadat Hossain
@Sena Kalyan Sangstha, Bangladesh

I.   IMPORTANCE OF THIS TOPIC

Researches on public health issue are essential in examining the root causes for creating possibilities for solutions. Research on public health issues and consequences helps integrate high values into society and provide relevant information about the diseases to the general public (Frizzelle et al., 2009). Public health factors are important to aid the good health of society by detecting their issues and resolving them as early as possible (Nass et al., 2009). Different approaches and ideas help researchers provide complementary insights by performing experiments and analysing the facts (Kar et al., 2020). This research topic is significant in analysing the importance of the widely used statistical programs such as the R program and SPSS in biostatics for analysing public health-related data. It focuses on the effectiveness of the two selected statistical tools in terms of their accuracy of results, methodology, access and cost-effectiveness. Past studies have revealed that public health data comes under sensitive data because it might affect a person's health or a large group of people (Price et al., 2019).

Biostatistics are the development and the use of applications related to statistical methods for a wide range of biology-related topics (Reiczigel et al., 2019). It encompasses the design of the experiments, collection and analysis of the collected data from the performed experiments, and interpretation of those results. The rationale behind the selection of this research study is to provide a better understanding to the readers regarding the use of the selected program in biostatistics for the study of public health-related components. The health-related studies are higher chances of errors that might affect the ground realities with severe issues. Sometimes, researchers do not have enough time to examine the probability and statistics of the disease prevalence (Van Norman, 2019). For such issues, the statistical tools reduce the time of manual examination and promote ease for the examiners to evaluate the root causes. Health research can lead to bringing revolutionary changes in traditional treatment interventions (Nass et al., 2009). These can work in the development of new therapies and improvement in primary and public health care. SPSS program provides data analysis for bivariate and descriptive statistics, numeral outcomes and predicting the identifying groups (TechTarget, 2022). Whereas, Association for Psychological Science (2017) mentioned that R is a programming language for graphics and statistical computing that can be used to analyse, clean and data graphs.

Biostatics in public health is considered the best way to identify resources to treat the population. To control the epidemic, the role of biostatics is not only limited to finding the best way of treating the people who are infected by the disease, but it helps to identify the way which can be used to prevent the people from life treating infections. In both infectious disease and behavioural research, interventions provided to the individuals can significantly impact the others in the community. It requires the research method to investigate the relationship based on individual patients and the population levels. These types of research can be performed on a micro level. Nan Larid and Harvey V.Fineberg, professors of public health, believe that biostatics is the technique that can be used to mount the missing data by designing that data (Ryan, L., 2015). Biostatics is responsible for observing the data in different ways, which can be considered where the facts are missing. It is one of the widely used and advanced methods being applied in various settings in the biomedical field of research.

Whereas according to professors Francesca Dominici and Brent Coull, assistant professors at Crown Ziglar, believe that biostatics is the new quantitative developed method used for the analyses while considering the quality policy and regulations (Kianian, B., 2021). This is an innovative form of data collection which helps to

understand the issues and their solutions in public health closely. Biostatics has been playing a significant role in AIDS, where the study by Professor Miceal Hughes developed and designed the statistical method (Beigel,2019). This statistical method was used to prove the important development of highly effective strategies for the treatment of this infectious disease along with its prevention. The biostatistical data has made a greater contribution that was considered to be important for helping to create change in HIV infection. Several other researchers have considered biostatistical data in public health has greatly benefited the public and researchers.

According to Chan B.K. (2018), after 1996, the open-source programming language programming R gained the population in the environmental statistical analyses and graphical output. R has been free, providing several advantages to biostatics, including strong graphic capabilities, the ability to write and customise functions and extensibility. R has several important roles in biostatistics. It plays a significant role in nutrition, health-related environment, and behavioural health. The basic understanding of R. R Program usage for public health is an important concept or term. Covid-19 is considered to play a greater role in raising the profile of public health workers at all levels, including nurses, doctors, and other front-line workers in hospitals. The situation shows a significant requirement for continuous improvement in public health. To create progress, the public health system review the scientific data, which is responsible for making a greater contribution to research, logic, management of the data, reporting and public communication. R program has been making a difference in the collection of the data and providing statistical information related to the situation, and highlighting the areas which need to be significantly improved (Picoli, 2018).

Along with the R program, SPPSS also plays a veritable role in the statistical procedure. SPSS is a window-based program which can be used for data management and analyses to create tables or graphs (Kumar, R., 2019).SPSS can handle a large amount of data, and it can perform the analyses covered within the text. SPSS has been commonly used by the public health and pharmaceutical industries to facilitate data collection and improvement in the system. SPSS can be treated as a medical statistic that can provide an understanding based on clinical research. SPSS provide guidance related to the statistical research and interpret the data on how the participants were required and the study was designed. What types of variables were used in the study. It also helps to understand the effects and size of the data collected. It is responsible for guiding the researcher through the process of selecting the relevant statistics and showing the results.

SPSS provides a clear and concise explanation that is narrowed and covers all data to provide the exact results in the medical statistics. Data collection in the medical field can be complex and sensitive. These results can affect the lives of the people or the health of the people (Tao,2018). Therefore these tools are considered to play a significant role in collecting data and information suitable for the results. These tools are effective because it reduces the chances of human errors, and it helps to gather a large amount of data which can be a problem while collecting without using these tools. R and SPSS are useful for narrowing the data and simplifying it to make it easier the understand the reader (Patridge, 2018).

Research aims and objectives
Following are the aims and objectives of the research study:
Aims
This study has aimed to examine the importance of the R program and SPSS program, which are required biostatistics for analysing public health. It also evaluates the program's efficacy by analysing public health factors.
Objectives
● To know about the accuracy of the R program and SPSS program results for analysing the public health factors.
● To examine the R and SPSS program's role in biostatics to analyse the public health elements.
● To learn about the requirements of biostatics while using a statistical application to analyse public health data.
Research questions
The enlisted research questions will be answered throughout the research project:
● What is the importance of the R and SPSS programs for the biostatistical analysis of public health data?
● What is the efficacy of using the R and SPSS programs in analysing the collected data?

Research outcomes
An understanding of biostatics is important in the field of public health. The following study provided the details based upon the use OF SPSS and R programs in biostatics in the field of public health. The study provides the details regarding the SPSS programs that it is

responsible for providing the data analyses for the descriptive and bivariate statistics. SPSS is also useful for providing the numerical outcomes of the data prediction based upon the group identification. SPSS is the biostatistics software that provides the transformation through graphics and direct data featuring (Abbasnasab, 2021). Through this software, the search identified that the use became able to open the data similarity to the spreadsheets and its major views. Along with the second variable, using the metadata by the SPSS helps to describe the variables and data entries that are present in the data file. Data collection in public health is considered a sensitive matter; therefore, the study collected data provides details regarding the importance of using SPSS in the public health system.

The data collected through the research suggested that SPSS is one of the most popular software tools used by the health care community to analyse the data (Dolezel & McLeod,2019). The study used secondary research, including the previously collected data within the 5 years, to gain up-to-date information. Through secondary information, the study provided details that SPSS has been used in public health since 1995 (Bogale,2018). The study found that SPSS, which has been operating for stats, is one of the popular software for conducting the data. This present study provided information on using SPSS and R programs in the public health sector. The controlled study from the secondary data shows that this software, both SPSS and R, is considered the most commonly used statistical software. In recent years, the characteristics have also increased with the update in this software. The updated version of this software has also played a greater role in making the statistical data more effective and easier for the researchers in the health department. Through these software updates, users can get the accessibility to advanced methods, which can reduce the chances of error and increase the efficiency of the data. The collected data reveals that the SPSS is disparately used at a greater level in several countries for conducting public health data in biostatistics.

The following study mainly focused on why this software was used to collect statistical information. The findings and outcomes of the study showed that it is important to use this software as they help the researcher collect the error fee data. There are several errors if humans collect the data, which also increases the time and cost of collecting these data. The data collected by humans can also be in a wide range, and it isn't easy to interpret that information. Therefore this software has been created, and this software is free

of cost, which decreases the cost barrier for data collection in public health. This software is based on Artificial intelligence and includes updated features that can help reduce the chances of errors and make the data reliable and efficient for the public health sectors. Through these SPSS and R software, a large amount of data can be collected within a limited period of time through numerical (Mertler,2018). The software not only gathers the information within a limited period but also performs the data analyses within a few hours, which can take several days or months to be analysed by human resources.

However, some of the researchers show in the study that this software has different benefits for the data collection that they have become the major part of biostatics for collecting and analysing the information to get better and more reliable results for the public health sector. Several researchers have been conducted to identify the effectiveness of the historical and use of R and SPSS programs in public health. The researchers also show that this software has been used to create changes in people's lives and has helped improve the public health sector. These improvements have been taking place through identifying the areas of weakness and identification of the areas of weakness. This software is useful for identifying the opportunities that can be applied to gather the data, which can inform the researcher what steps can be taken to improve public health. Other outcomes of the study provide the importance of the SPS and R software in public health by considering the example of AIDS (Tesfaye,2019). The researchers discussed that SPSS and R software had been used to collect the information and opportunities that can be utilised to improve the health sector and create chances to prevent people from getting this infectious disease. The primary reason behind collecting this research was to identify the statistical software used in the biostatistics, which can provide the statistical results to create a difference in the health sector and disease identification and prevention.

Like other industries, clinical laboratories also rely on statistical data. The outcomes of the study show how clinical laboratories are getting benefits for the data statistics and how this software is providing the benefits and helping them meet their aims and objectives to improve the system. The study found that the clinical laboratories generate the process and store the transactional data with higher quality and efficiency. These data are collected by using the R program and SPSS, which is important for patient care and the quality assurance of the activities and increases operational

decision-making (Athiraman,2022). The study found that all these data heavily rely on the spreadsheet or the specialised data application such as R program and SPSS. However, other programs can be used to collect and analyse the study. However, these programs can be functionally limited and unsuitable for the complex statistical analyses and visualisation of the data on larger scales and dimensions. The other software or programs can also be considered as having a lack of transparency of the data. In contrast to those R and SPSS are considered as the open surface and independent positions that are freely availing with the programming language. These platforms are also having the ability to deal with massive data and are used by different countries around the world for data collection.

These identified characteristics are considered to be important for making this software useful and suitable for clinical liberties, the Application R in the medical field is specifically used to create the difference in clinical liberties, and its usage has been increasing because of the increase in this versatility and availability for the relevant and focused training.

These tools are considered to be conceivably applications which enable to adopt the variety of user interests. Further, the study outcomes show that the R program has different features that make this application more suitable for the public health such as use of R is highly customise and reproduces the data as well. Along with SPSS, the study also identified that are are different benefits of using R in the public health. As R programming is known for allowing the public health communities and others to analyse and communicate the data result. It is based upon a wide range of capabilities that proves the greater level of facility in working with the Microsoft Excel and some other commercial data analyses programs. R is based upon the text commands in order to process the data it is used for citing with the full-fledge programming language for advanced users in public health (Amalina,2019). These outcomes of the study are also based on showing that SPSS and R, are both playing significant role IN public health because of their advanced features and system that allows the users to conduct the reliable and effective data.

II. CONCLUSION

Biostatics has been playing a significant role in collecting the data in public health. The study was based upon identifying the importance of R and SPSS in the field of public health through applying biostatics. The study shows the importance of R and SPSS which has been playing a greater role in the field of public health. Both R and SPSS are software application that are used for data collection and data analyses. These software are available for free of cost, and they have the advance feature which helps the researcher to collect the information which is reliable and efficient. The public health is now heavily relying on these software because of the increase efficiency and decrease the chances of human errors. the study outcomes shows that biostatics has been using these software since 19s and till now with the advance features and updates these software are considered to play the greater role in the field of public health. Therefore the study can be concluded that R and SPSS are considered as the important tools for data collection. These are not only the tools which help to collect the relevant data, but they are also responsible for certain the opportunities to improve the health care centers. These improvements are important for both the public and the staff of health care centers.

REFERENCE

[1]. Association for Psychological Science, 2017. Why You Should Become a UseR: A Brief Introduction to R. available at: https://www.psychologicalscience.org/observer/why-you-should-become-a-user-a-brief-introduction-to-r#:~:text=R%20is%20a%20programming%20language,of%20statistics%20and%20research%20methods.

[2]. Athiraman, N.K., Patience, A., Onwuneme, C., Zalewski, S. and Skeath, T., 2022. Impact of COVID- 19 on maternity and neonatal services–Three year- on- year review data from the North East of England. Acta Paediatrica, 111(5), pp.1039-1041.

[3]. Beigel, J.H., Aga, E., Elie-Turenne, M.C., Cho, J., Tebas, P., Clark, C.L., Metcalf, J.P., Ozment, C., Raviprakash, K., Beeler, J. and Holley Jr, H.P., 2019. Anti-influenza immune plasma for the treatment of patients with severe influenza A: a randomised, double-blind, phase 3 trial. The Lancet Respiratory Medicine, 7(11), pp.941-950.

[4]. Bogale, T.Y., Bala, E.T., Tadesse, M. and Asamoah, B.O., 2018. Prevalence and associated factors for stunting among 6– 12 years old school age children from rural community of Humbo district, Southern Ethiopia. BMC public health, 18(1), pp.1-8.

[5]. Chan, B.K., 2018. Data analysis using R programming. In Biostatistics for Human Genetic Epidemiology (pp. 47-122). Springer, Cham.

[6]. Frizzelle, B.G., Evenson, K.R., Rodriguez, D.A. and Laraia, B.A., 2009. The importance of accurate road data for spatial applications in public health: customising a road network. International journal of health geographics, 8(1), pp.1-11.

[7]. Kar, A.K. and Dwivedi, Y.K., 2020. Theory building with big data-driven research–Moving away from the "What" towards the "Why". International Journal of Information Management, 54, p.102205.

[8]. Kianian, B., 2021. Statistical Methods for Spatial Data in Public Health (Doctoral dissertation, Emory University).

[9]. Kumar, R., 2019. Introduction to Statistical Tool: SPSS. Hands-on training on "Statistical Tools and Database Management In Agriculture", p.15.

[10]. Mertler, C.A., Vannatta, R.A. and Lavenia, K.N., 2021. Advanced and multivariate statistical methods: Practical application and interpretation. Routledge.

[11]. Nass, S.J., Levit, L.A. and Gostin, L.O., 2009. The value, importance, and oversight of health research. In Beyond the HIPAA privacy rule: enhancing privacy, improving health through research. National Academies Press (US).

[12]. Patridge, E.F. and Bardyn, T.P., 2018. Research electronic data capture (REDCap). Journal of the Medical Library Association: JMLA, 106(1), p.142.

[13]. Picoli, M.C.A., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., Coutinho, A., Esquerdo, J., Antunes, J., Begotti, R.A. and Arvor, D., 2018. Big earth observation time series analysis for monitoring Brazilian agriculture. ISPRS journal of photogrammetry and remote sensing, 145, pp.328-339.

[14]. Price, W.N. and Cohen, I.G., 2019. Privacy in the age of medical big data. Nature medicine, 25(1), pp.37-43.

[15]. Reiczigel, J., Marozzi, M., Fábián, I. and Rózsa, L., 2019. Biostatistics for parasitologists–a primer to quantitative parasitology. Trends in parasitology, 35(4), pp.277-281.

[16]. Ryan, L., 2015. A conversation with Nan Laird. Statistical Science, 30(4), pp.582-596.

[17]. TechTarget, 2022. SPSS (Statistical Package for the Social Sciences). Available at: https://www.techtarget.com/whatis/definition/SPSS-Statistical-Package-for-the-Social-Sciences#:~:text=SPSS%20provides%20data%20analysis%20for,spreadsheet%20in%20its%20main%20view.

[18]. Tesfaye, D.J., Hibistu, D.T., Abebo, T.A., Asfaw, F.T., Lukas, K., Laelago, T., Turuse, E.A., Kebede, H.G., Altaye, A.A. and Bekele, F.B., 2019. Option B plus antiretroviral therapy adherence and associated factors among HIV positive pregnant women in Southern Ethiopia. BMC pregnancy and childbirth, 19(1), pp.1-8.

[19]. Van Norman, G.A., 2019. Limitations of animal studies for predicting toxicity in clinical trials: is it time to rethink our current approach? JACC: Basic to Translational Science, 4(7), pp.845-854.