



Statistical Software Applications and Statistical Methods Used In Community Medicine and Public Health Research Studies

Krishna P Joshi¹, Deepak C Jamadar²

¹Professor&HOD, Dept. of Community Medicine, SVS Medical College, Mahabubnagar

²Assistant Professor cum Statistician, Dept. of Community Medicine, SVS Medical College, Mahabubnagar

ABSTRACT

Background: The utilization of software in statistical methods and tools has increased very much in last two decades. The easily availability of statistical application software on internet, made easy to analyse, interpretation and conclude the study variables.

Aim and Objectives: To know the statistical software applications and various statistical methods used in the community medicine and public health research studies.

Materials & Methods: A cross sectional retrospective study, We had included online community medicine and public health journals of four years (2017 to 2020).By cluster sampling method were used and each cluster includes articles of four different year, 400 randomly selected research articles were downloaded & reviewed. The data was analysed by using SPSS version 23.

Results: 400 articles were reviewed. 317 (79.25%) were used Ms-Excel software followed by 261 (65.25%) descriptive statistics and 145(36.25%) SPSS and 143 (35.75%) applied cross sectional study design. 91 (22.75%) applied chi-square test, 77(19.25%) applied ANOVA test. 59(14.75%) t-tests. Only 41(10.25%) articles were reported with confidence intervals.

Conclusion: The current study reveals, Microsoft excels, and SPSS and R-programming were more frequently using software in the community medicine and public health, whereas descriptive statistics, chi-square tests, t-tests and non-parametric tests were commonly using statistical methods.

Key word: Statistics, Community Medicine & public health, software, SPSS, ANOVA etc.

INTRODUCTION

Statistics became the integral part of the community medicine and public health research studies. Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.^[1] There is no question that community medicine and public health research studies come to rely heavily on the computer and statistical applications software. This dependence has become so great that it is no longer possible to understand community medicine and public health research studies

without substantial knowledge of statistics and without at least some rudimentary understanding of statistical software.

From the past three to four decades in the scientific and research studies the statistics and various statistical methods have been accepted as a powerful tools and it is documented in the medical journals.^[2-4] The importance of statistical applications and interpretations has been identified as one of the leading contributors to current research studies.^[5-7]

How to cite this article: Joshi K, Jamadar D. Statistical Software Applications and Statistical Methods Used In Community Medicine and Public Health Research Studies. *Natl J Community Med* 2021;12(3):53-56. DOI: 10.5455/njcm.20210329094615

Financial Support: None declared **Conflict of Interest:** None declared

Copy Right: The Journal retains the copyrights of this article. However, reproduction is permissible with due acknowledgement of the source.

Date of Submission: 27-03-2021; **Date of Acceptance:** 30-03-2021; **Date of Publication:** 31-03-2021

Correspondence: Dr. Deepak Jamadar (Email: deepak3march@gmail.com)

Many researchers including community medicine and public health experts are using the various software for data analysis which bridges the gap between data generation and data analysis to make meaningful results and conclusions. The statistical software tools empower the researchers without depth knowledge of statistics. There are number of statistical applications software are available like MS-Excel, SPSS, Graph pad R-programming etc. And some of the most widely-used statistical analysis software is provided on several internet sites.^[8] And they are upgraded time to time and statistical software can be used as a conclusion tools in research studies without having deep knowledge of statistics.

The number of types of statistical software packages continues to grow each year and the demand of statistical software has been increased some of the software like excel is used widely in statistical analysis. The program has functions that perform simple and complex mathematical and statistical functions one at a time.^[9]

Similarly SPSS statistical software is most comprehensive of the statistical tools is a cross-disciplinary tool used in medical science, biology, statistics, social sciences, etc. The methodology to be used for analysis is at times understood by the software but in most cases to be defined specifically and not that direct for a novice in statistics.

Graph pad is one of the statistical software it is used both in academic and biologist the software helps the researcher perform basic statistical methods needed in laboratory researchers and clinicians like t-tests, nonparametric comparisons, one-way ANOVA and two-way ANOVA, analysis of contingency tables, and survival analysis.^[10]

Now R-programming language is free statistical software easily available through online it is specifically used for the statistical analysis and data plotting.^[11-12] Hence the demand of statistical application software has been increased due to easy understanding and easy availability.

Taking all above points into consideration we try to analyse how frequently the researchers of community medicine and public health are using the statistical software applications and various statistical methods in the community medicine and public health research studies.

MATERIALS & METHODOLOGY

Present cross sectional, retrospective study was conducted during January 2021 to March 2021. Institutional Ethics Committee clearance was obtained prior to start the study. We had included around 10 various journals of last four years (national & international) of the community medicine and public health which were convenient & easily available in online (year 2017 to 2020). No other inclusion or exclusion criteria were used while

selecting the journal. By using cluster sampling procedures, the articles were divided into four clusters with respective different years like 2017, 2018, 2019 & 2020.

Each cluster includes randomly selected 25 published articles from each year. So in each cluster 100 published research articles were included from the various journals, the process was continued until we reached the study sample size (N=400). Further we reviewed all the downloaded articles and verified statistical software and various statistical methods were used in the articles.

The data was collected by pre designed & pre tested form which includes closed-coding system for quantifying statistical methods reported by the authors of each published article. The statistical software application SPSS version 23 was used for analysis.

The names of selected journals are not been disclosed here to avoid any contradictory statement or claim regarding inclusion or exclusion of particular journal.

RESULTS

The study comprised of a total 400 published research articles, Table:1 reveals that 79.25% researchers used Ms-Excel software for statistical reporting followed by 36.25% were using the SPSS followed by 6.25% researchers used some other software whereas around 7.25% researchers did not used any statistical software in their research articles. Table: 2 reveals that around 65.25% researchers reported simple summaries such as percentage, mean, median, and standard deviation followed by 22.75% research articles were used chi-square test & by 35.75% researchers used a cross sectional study design followed by 19.25% used analysis of variance techniques. Around 27.5% were used t-test. 10.25% researchers calculated the confidence intervals and 6.25% were used the other applications software which are not statistical. 5.5% researchers concluded their results without using any statistical software applications and statistical methods.

In this study we found that, the number of researchers used multiple number of statistical software applications and multiple number of statistical methods in their research articles, hence we considered all the above points, the collected data was analysed statistically.

DISCUSSION

The present study is an attempt to explore the information on statistical software applications and various statistical methods used in the community medicine and public health research articles. Here some of the research articles were used multiple number of statistical applications software and mul-

multiple numbers of statistical methods in their studies; taking all these data into consideration we were calculated the data and presented the result statistically. 261 (65.25%) reported simple summaries such as percentage, mean, median, and standard deviation are commonly used for reporting, similar findings from the Rigby *et al.* Our analysis shows that, the most commonly used study design was the cross-sectional study. This finding is consistent with other studies and Robert Muenchen was also used various similar statistical methods for data analysis in his research. [13-14]

In our study it shows that 145(36.25%) were using the SPSS, similar findings from a survey conducted by Rigby *et al.* and Scotch *et al.* found that SPSS was the most popular software used for community health assessment data analysis. [15-16]

Hayat M.J mentioned in this study that how to test the significance process and how to determine the statistical significance.[17] Mary Johnson *et al.* clearly explained uses and importance of statistical software applications in biomedical research article similarly we also tried to elaborate uses and importance of statistical applications like MS-Excel, SPSS, Graph pad and R-programming in the community medicine and public health research studies. Several authorities mentioned that, uniform guidelines for reporting of statistical methods and results. These guidelines generally advise authors to identify the statistical software and version used, when applicable.[18-22]

CONCLUSION

In the present study the researcher has intention to provide brief reports towards the most common types of statistical software applications and various methods used in community medicine and public health studies. This study reveals that Microsoft excels, SPSS and R-programming were more frequently using software in the community medicine and public health research studies. Descriptive statistics (Mean, Median, SD etc.), chi-square tests, non-parametric tests were commonly used statistical methods in the community medicine and public health studies.

Our study did not attempt to qualitatively assess the merits of particular software applications relative to one another or to evaluate their suitability for different analytical uses. Because we randomly selected few journals for the study, so our results might not be representative of software usage patterns throughout the entire community medicine and public health research studies. Further attempt should be made to know more about statistical software applications and various statistical methods used in the community medicine and public health research studies.

Table 1: Frequency distribution of statistical software applications used in various research articles

Statistical software	Frequency N (%)*
Microsoft-EXCEL	317(79.25)
SPSS	145(36.25)
STATA	1(0.25)
M-TAB	2(0.5)
SigmaPlot	0(0.00)
Graphpad-Prism	11(0.00)
G-Power	2(0.5)
SAS	2(0.5)
R	3(0.75)
**OTHERS	25(6.25)
Not used any software	29(7.25)

*Multiple software used by various researchers; **Others-G-power, Statistica, MATLAB, sigmastat

Table 2: Distribution of various statistical methods and techniques used in the research article

Statistical Methods and study types	Frequency (n=400)*
Statistical Tests	
Descriptive Statistics & frequency tables (summaries such as percentage, mean, median, and standard deviation)	261(65.25)
Chi-square test	91(22.75)
Pearson correlation	33(8.25)
Regression	43(10.75)
Analysis of Variance (ANOVA)	77(19.25)
Non-Linear Regression	25(6.25)
Kruskal walli test	6(1.5)
Cronbach's alpha	3(0.75)
Kappa	2(0.5)
Sensitivity and specificity	16(4.00)
Mantel-Haenszel	0(0.00)
Linear Regression	13(3.25)
t-test	110(27.5)
Odds ratio/relative risk	28(7.00)
Confidence intervals	41(10.25)
**Others	102(25.5)
Study type	
Cross-sectional study	143(35.75)
Prospective study	49(12.25)
Retrospective study	13(3.25)
Cohort study	2(0.5)
Case-control study	3(0.75)
Longitudinal study	00(0.00)
Others	22(5.5)

*Multiple Statistical methods were used by many researchers

**logistic regression, multivariate analysis, sign test, ROC curve etc.

REFERENCES

1. <https://en.wikipedia.org/wiki/Statistics> accessed on 01.02.2021.
2. Altman DG. Statistics in medical journals. *Stat Med.* 1982; 1:59-71.
3. Altman DG. Statistics in medical journals: developments in the1980s. *Stat Med.* 1991; 10:1897-913.
4. Altman DG. Statistics in medical journals: some recent trends. *Stat Med.* 2000; 19:3275-89.
5. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature.* 2016; 533:452-4.

6. Kass R, Caffo B, Davidian M, Meng X, Yu B, Reid N. Ten Simple Rules for Effective Statistical Practice. *PLoS Comput Biol*. 2016; 12:e1004961.
7. Brown E, Behrmann M. Controversy in statistical analysis of functional magnetic resonance imaging data. *Proc Natl Acad Sci U S A*. 2017; 114:E3368-E3369.
8. Wegman EJ, Solka JL. Statistical software for today and tomorrow. 2005.
9. Clairfeuille T, Cloake A, Infield D, Llongueras J, Arthur C, Li Z, et al. Structural basis of α -scorpion toxin action on Nav channels. *Science*. 2019; 363.
10. Steichen J, Lin Y, Havenar Daughton C, Pecetta S, Ozorowski G, Willis J, et al. A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses. *Science*. 2019.
11. Becattini S, Littmann E, Seok R, Amoretti L, Fontana E, Wright R, et al. Enhancing mucosal immunity by transient microbiota depletion. *Nat Commun*. 2020;11:4475.
12. Wu P, O Malley J, De Gruttola V, Liberman M. Age-Related Hearing Loss Is Dominated by Damage to Inner Ear Sensory Cells, Not the Cellular Battery That Powers Them. *J Neurosci*. 2020; 40:6357-6366.
13. Wang Q, Zhang BH. Research design and statistical methods in Chinese medical journals. *JAMA*. 1998; 280:283-5.
14. Muenchen RA: The popularity of data analysis software. 2010.
15. Rigby AS, Armstrong GK, Campbell MJ, Summerton N. A survey of statistics in three UK general practice journals. *BMC Med Res Methodol*. 2004 4:28.
16. Scotch M, Parmanto B, Gadd CS, Sharma RK. Exploring the role of GIS during community health assessment. *International Journal of Health Geographics*. 2006; 5(36):1-10.
17. Hayat M.J. 2010. "Understanding Statistical Significance," *Nursing Research*, May/June: 59(3), 219-223.
18. Mary Johnson 2014. Statistical Analysis Software Programs in Biomedical Research. Labome the world of laboratories.
19. Bailar JC, Mosteller F: Guidelines for statistical reporting in articles for medical journals. *Annals of Internal Medicine*. 1988, 108 (2): 266-273.
20. Curran-Everett D, Benos DJ: Guidelines for reporting statistics in journals published by the American Physiological Society. *Physiological Genomics*. 2004, 18: 249-251. 10.1152/physiolgenomics.00155.2004.
21. Lang TA, Secic M: How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors and Reviewers. 2006.
22. Sarter M, Fritschy J-M: Reporting statistical methods and statistical results in EUR. *European Journal of Neuroscience*. 2008.