

Original Article

Impact of hands-on training in STATA for data management and data analysis: How much knowledge gained among health-care professionals?

Hira Ballabh Pant¹, Tripura Batchu¹, Raashi Raj², AY Nirupama¹, Varun Agiwal¹

¹Indian Institute of Public Health, ²Center for Health Outcomes Research and Economics, Indian Institute of Public Health, Hyderabad, Telangana, India.



*Corresponding author:

Varun Agiwal,
Indian Institute of Public
Health, Hyderabad, Telangana,
India.

varun.agiwal@iiph.org

Received : 14 November 2022

Accepted : 13 December 2022

Published : 17 January 2023

DOI

10.25259/IHOPEJO_30_2022

Quick Response Code:



ABSTRACT

Objectives: Data management and statistical analysis are integral parts of any research. But there is difficulty in cleaning, organizing, and maintaining data in a systematic and meaningful way for accurate analysis without any tool. Hence, there is a need to learn a tool such as SPSS, STATA, R, SAS and so many for understanding the data. This study focused on the impact of short-course training on STATA on healthcare eye professionals and found an improvement in their competency to use computer-based statistical software packages for data management and analysis and identify the major challenges for participants and suggestions to improve the training.

Materials and Methods: We organized a 10-day short course on data management and data analysis using STATA. A pre and post-test evaluation was done every day to find the impact. The questionnaire covered several areas of research methodology from data management to statistics. Categorical data were described in frequency and percentage whereas continuous data was described in median and interquartile range (IQR). Wilcoxon signed-rank test was applied to find out the significant improvement in the knowledge.

Results: The 10 days cumulative median score for pre- and post-test was 43 (IOR = 39-47.5) and 72.5 (IOR = 69-77) respectively and was statistically significant ($P = 0.012$). Male participants, younger age groups, experience, and prior knowledge were identified as major determinants to improve the knowledge and performance in the usage of STATA. We observed that advanced training in the usage of STATA is highly beneficial to people interested/involved in research to gain autonomy in data management and analysis.

Conclusion: The evaluation of the training confirms that hands-on training in the use of STATA software does have a significant impact on the knowledge, self-efficacy, and attitude of the learner. Since Stata is very user-friendly, we recommend that researchers and medical students should be offered a comprehensive training course in STATA.

Keywords: Data management, STATA, Data analysis

INTRODUCTION

Data management and statistical analysis are integral parts of any research. Data management is a process of collecting, cleaning, organizing, and maintaining data in a systematic and meaningful way for accurate analysis.^[1] It is very critical to have a quality data to provide best analysis which means data quality problems such as manual data entry errors, missing data, and equivocal data have to be resolved before analysis. There are many tools to identify errors and missing observations. One of the tools to identify this is STATA. According to StataCorp (2016), Stata

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, transform, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

©2023 Published by Scientific Scholar on behalf of IHOPE Journal of Ophthalmology

is “a complete, integrated statistical software package that provides everything for data management, data analysis, and graphics.”^[2] Once, the data management is over, data analysis is performed.

For data analysis, researcher must have some basic knowledge of statistics. Statistics is a branch of science that helps to organize, describe, classify, and analyze the data and interpret the outcomes.^[3] Statistics is essential in every field ranging from commerce to healthcare, psychology to social sciences. Biostatistics is the use of statistical tools in health care and biological sciences. In medical education, biostatistics is an integral part of the curriculum. An innovative and computer-application-based learning is always aspirational and orients the learner in applying the tools toward relevant research and developing evidence-based interventions. It helps health-care professionals in accurate analysis and interpretation of results. Lack of biostatistical knowledge can lead to unethical research, and misinterpretation of results leading to a negative impact on evidence-based medical practice. There are many statistical software for statistical computing such as R (www.r-project.org), SAS (www.sas.com), SPSS (www.spss.com), and Stata (www.stata.com). Stata is both command line and point-and-click menus software. Stata is the most practical and effective software for data management and analysis in health-care research. Therefore, there is a need to introduce a training course on Stata for data management and data analysis for the health-care professionals and medical students.

We conducted a study to find out if providing a short-course training on STATA to healthcare eye professionals will achieve the objective of improving their competency to use computer-based statistical software packages for data analysis in addition to an increase in theoretical knowledge of data management and biostatistics. We also assessed effectiveness of the training using some key performance indicators such as facilitator’s knowledge, participant’s self-efficacy, and attitude toward training. We requested the learners’ feedback on challenges identified and suggestions to improve the training program.

Objectives of the training

The objectives are as follows:

- To determine the impact of hands-on-training on STATA among health-care eye professionals and program manager
- To evaluate the performance of training based on facilitator’s knowledge, participant’s self-efficacy, and attitude toward training
- To identify the major challenges for participants and suggestions to improve the training.

MATERIALS AND METHODS

We organized a 10-day short course on data management and data analysis using STATA at the Indian Institute of Public Health, Hyderabad (IIPHH) under Operational Research Capacity Building project funded by Seva Foundation, USA. Eight participants from four eye care institutes (Nepal-3, India-1) received the training in STATA. It was a heterogenous group, comprising ophthalmologists, optometrists, outreach, and program managers.

The facilitators presented a brief overview of the topic at the beginning of each session. Care was taken to make the sessions engaging and memorable with frequent interactions through questions, exercises, and discussions using real-world examples. Using Google forms a pre and post-test evaluation was done every day. The questionnaire covered several areas of research methodology from data management to statistics. STATA software test assessment was done each day before beginning the session to assess pre-existing knowledge of the participants on the topics to be presented. The post-test questionnaire helped to evaluate the improvement in knowledge and skill gained through the course on that particular day.

On the last day of the training, the participants were asked to fill out the feedback survey form. The responses evaluated the facilitator’s knowledge, participants’ self-efficacy and attitude post-training, challenges encountered, and scope for improvement. No sampling method was followed. All participants included in the study were anonymized. Data were recorded in MS Excel and statistical analysis was carried out using STATA.

Categorical data were described in frequency and percentage whereas continuous data were described in median and interquartile range. Wilcoxon signed-rank test was applied to find out the significant improvement in the knowledge regarding the STATA software. Institutional ethical committee clearance was taken by the IIPHH. $P < 0.05$ was considered statistically significant.

RESULTS

Eight participants took part in the short course on STATA training. All participants held a graduate or post-graduate degree and worked in eye care institutions with an average work experience of 12.5 years ($SD \pm 6.0$). The mean age of the participant is 37.9 years ($SD \pm 4.1$ years). The female to male ratio is 1:3. Almost two-thirds of the participants had prior knowledge in managing data and applying basic statistical tools using other software such as SPSS, Epi-Info, and MS Excel.

Before providing training in STATA, pre-test evaluation was carried out every day. At the end of each day, a post-test evaluation was conducted. The 10 days cumulative median score

for pre- and post-test was 43 (IOR = 39–47.5) and 72.5 (IOR = 69–77), respectively, and was statistically significant ($P = 0.012$). Male participants, younger age groups, experience, and prior knowledge were identified as major determinants to improve the knowledge and performance in usage of STATA [Table 1].

Self-efficacy and attitude evaluation

About 75% of the participants agreed with self-efficacy regarding their expectations from the training. About 87.5% of the participants agreed that assignments, quizzes, knowledge checks in each session, and reading materials and tools are more helpful in participants learning. About 75% of the participants indicated that they had more confidence to apply knowledge and skills gained from the course to their jobs. About 25% of the participants were neutral and in disagreement regarding the reasonable time allowed to cover the course information [Table 2].

Facilitator evaluation

All participants completely agreed about the technical support provided by facilitators. About 87.5% of the participants agreed that the presentation given by the facilitators was clear, and valuable with various real life applications whereas 25% of the participants were neutral regarding the consistency among lectures, discussions, and exercises. The major reasons are basic knowledge of statistics and prior knowledge regarding systems/technology [Table 3].

About 62% of participants said that 80–100% of the training is linked to a real-world application followed by 25% who said 60–80% and 13% who said 40–60% [Figure 1]. After

the training course, the major challenges for participants are too many competing priorities, followed by no time to apply the skills and no opportunity to use the skills [Figure 2]. Overall, 87.5% are likely and more likely to share this training with your colleagues. The participants also provided some suggestions to improve the course. These suggestions are: increase the duration time for this training, have basics knowledge on statistics, need a refresh course after some period and after that an advance level course on the same.

DISCUSSION

Our 10 days hands on training workshop in data management and data analysis using STATA showed that there was a statistically significant improvement in knowledge of STATA software based on pre-test and post-test. A focused workshop on introduction to statistical methods and SPSS hands-on training to enhance analytical skills among research professionals held at Vijayapura by Tellur *et al.* showed improvement where the median score of pre-test was 12.5 and post-test was 14.0.^[4] In another study, Prabhu *et al.* showed an increase in mean scores from 3.42 (pre-test) to 10.53 (post-test) and Sari found an average pre-test and post-test score was 58.37 and 84.85, respectively,^[5,6] whereas in our study there was a substantial increase from 43.0 (pre-test) to 72.5(post-test).

We also assessed the facilitator's knowledge, participants' self-efficacy, and attitude at the end of the training. Majority of the (87.5%) of the participants agreed that adequate learning happened through assignments, quizzes, and knowledge checks. They also agreed that the facilitators had adequate knowledge of

Table 1: Evaluate the performance based on the participant's characteristics profile.

Characteristics	Pre-test score	Post-test score	P-value
	Median (IQR)	Median (IQR)	
Overall ($n=8$)	43 (39,47.5)	72.5 (69,77)	0.012
Gender			
Male ($n=6$)	41 (38,44)	72.5 (69,77)	0.027
Female ($n=2$)	47.5 (46,49)	73 (69,77)	0.18
Age			
30–39 ($n=5$)	44 (42,49)	77 (74,77)	0.042
40–49 ($n=3$)	38 (37,46)	69 (67,71)	0.109
Experience			
≤ 10 ($n=3$)	42 (40,51)	77 (69,79)	0.103
> 10 ($n=5$)	44 (38,46)	71 (69,74)	0.043
Prior knowledge regarding data management, basic statistics, and other software			
Yes ($n=5$)	44 (38,46)	71 (69,74)	0.043
No ($n=3$)	42 (40,51)	77 (69,79)	0.103
Professional			
Ophthalmologist/Optometrist ($n=4$)	41 (37.5,47.5)	70 (68,72.5)	0.066
Program Manager/Officer/coordinator ($n=4$)	44 (41,47.5)	77 (73,78)	0.068

IQR: Interquartile range

Table 2: Feedback survey for self-efficacy of participants and attitude toward this training.

	Agree (%)	Neutral (%)	Disagree (%)
Self-efficacy of participants			
I knew what was expected of me in this program	75	25	0
Assignments, quizzes, or knowledge checks helped to learn	87.5	12.5	0
Regularly used course materials and tools in between sessions	87.5	12.5	0
Generally received helpful feedback after practicing a task	100	0	0
There were frequent opportunities to interact and learn from others	100	0	0
Attitude towards training			
Confidently apply knowledge and skills from this course to a job	75	25	0
The amount of time allowed to cover course information was reasonable	50	25	25

Table 3: Feedback survey to evaluate the facilitator’s knowledge.

Facilitators evaluation	Agree (%)	Neutral (%)
Technical support was easily available when or if it is needed	100	0
There was a good balance between lecture, discussion and exercises	75	25
Various methods were used to present concepts or ideas	87.5	12.5
The course material was of high quality	75	25
Course faculty presented information clearly	87.5	12.5

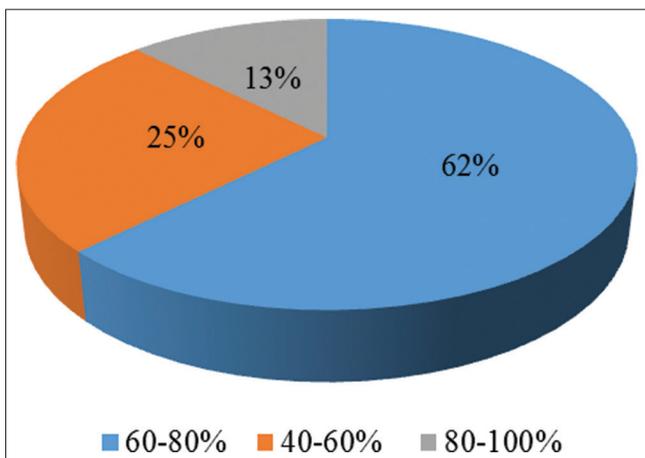


Figure 1: Proportion of training devoted to real-world application.

the course and presented the information clearly with various concepts and ideas. Similar results were found in a study by Prabhu *et al.* where the majority of the participants rated the sessions as “good.”^[5] We observed that an advanced training in usage of STATA is highly beneficial to people interested/involved in research to gain autonomy in data management and analysis.

In the present study, participants reported significant barriers to impending research and hence utilization of the STATA tool such as too many competing priorities in your work, no time to apply the skill and no opportunity to use the skill. Similar barriers to research had been reported in a study done by Giri *et al.* where lack of time was the major obstacle followed by inadequate financial support.

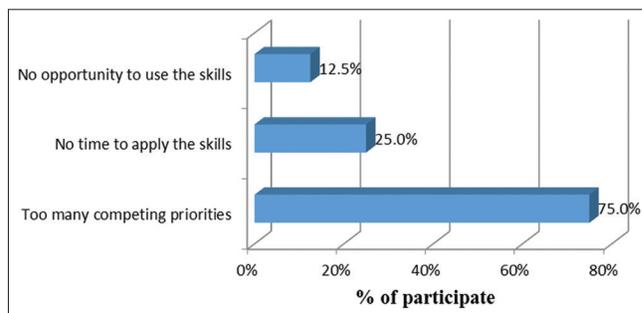


Figure 2: Different challenges after the course training.

Limitations

The duration of the training could have been longer to make the sessions less intense and more relaxed and the participant pool was small. We did not find any other study similar to ours, this is the first study to assess the impact of STATA usage with a scope to do similar evaluation on larger pool of participants from different professions.

CONCLUSION

The evaluation of the training confirms that hands on training in the use of STATA software do have a significant impact on knowledge, self-efficacy, and attitude of the learner. Since Stata is very user friendly, we recommend that researchers and medical students should be offered a comprehensive training course in STATA.

Declaration of patient consent

Patient's consent not required as there are no patients in this study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Briney K. Data Management for Researchers: Organize, Maintain and Share your Data for Research Success. Exeter: Pelagic Publishing Ltd.; 2015.
2. StataCorp LP. STATA Software (version 14.1). College Station, TX: StataCorp LP; 2016. p. 77845.
3. Brase CH, Brase CP. Understandable Statistics. Boston:

Cengage Learning; 2014.

4. Tellur L, Sorganvi V, Yadavannavar MC. Experience of workshop on introduction to statistical methods and SPSS hands-on training to enhance analytical skills among research professionals. *Int J Community Med Public Health* 2017;4:4337-40.
5. Prabhu GR, Rao BS, Sridhar MS, Reddy KA, Chandrasekharan PA, Reddy DS, *et al.* Evaluation of research methodology workshop for postgraduates in a medical college, Tirupati. *J Evol Med Dent Sci* 2015;4:15868-73.
6. Sari SU. The Effectiveness of Statistics Software Training to Improve the Understanding of Quantitative Research Methodology Concepts for Students of Tadris Mathematics, Faculty of Tarbiyah and Teacher Training, UIN Maulana Malik Ibrahim Malang. In: International Conference on Engineering, Technology and Social Science (ICONETOS 2020). Netherlands: Atlantis Press; 2021. p. 53-8.

How to cite this article: Pant HB, Batchu T, Raj R, Nirupama AY, Agiwal V. Impact of hands-on training in STATA for data management and data analysis: How much knowledge gained among health-care professionals? *IHOPE J Ophthalmol* 2023;2:4-8.