

A Step-Wise Evaluation of Three Commonly used Statistical Packages in Institutional Research

Introduction

Over the course of summer 2014, a conversation regarding statistical software packages emerged on the TAIR listserv. The conversation included questions, about cost, feasibility, the use of open source software, and general experience and usability of various software packages. One member initiated a poll, and soon thereafter, a summary by institution and software had been compiled and disseminated through the listserv. This summary of software packages by institution sparked our interest to examine the usability of each of the software packages listed and to share with other users in institutional research (IR). As such, the present usability study emerged.

Usability tests are routinely conducted on software, applications, widgets, and various other computer interfaces in order to obtain a *real-feel* sense and assessment of its functionality. The tests are generally categorized by whom the tester is, that is, expert-based or user-based. The user-based usability test considers domain or content experts to be the evaluators. This approach has been found to be a highly reliable means to evaluate the functionality of information systems (Sjoberg et. al., 2002).

There are several approaches to conducting a usability test. Among the most common user-centered tests are heuristic evaluations, cognitive walkthrough, and think aloud methods. In all cases, the results of such tests yield quantitative and qualitative data that allow for robust conclusions regarding usability.

Rogers (1995) suggests that one important advantage of user-based usability test is that the findings can easily be transferred to other end users. Sjoberg and colleagues (2002) add to Rogers' (1995) assertion that results from user-based usability tests are as reliable as those conducted by expert evaluators.

Given the interest by IR, TAIR members in statistical software packages, and the credibility afforded by user-centered usability tests, the purpose of this study was to conduct a step-wise evaluation of three commonly used statistical software program by IR personnel. To this end, we examined statistical software in regards to tasks that are germane to institutional research (IR). Specifically these tasks were: data import/export, data manipulation, output, visualization, versatility and help documentation.

Method

The present study employed a step-wise or heuristic usability test of three commonly used statistical software packages used in Institutional Research. Criteria selection for the

statistical packages was based on the TAIR listserv online poll conducted by Meyers (2014) where he summarized the statistical software used by IR offices in Texas (see table 1). These included two commercial packages and one open-source statistical analysis package.

Table 1. Statistical Package use by Institution

	CP1	CP2	OS
Total	30	19	3
4 Year	13	13	0
2 Year	17	5	3
Public	27	15	3
Private	3	3	0

Specifically, our usability test employed a modified form of the step-wise method outlined by Kushniruk, Patel, and Cimino (1997) which includes 1) development of the testing plan, 2) selection of testers, 3) selection of specific tasks to evaluate, 4) conducting the test, and 5) data analysis. For each statistical package, the evaluator completed specified tasks once or twice. For all tasks, testers used the same data set. Issues emerging during the completion of a task were rated with: M= Minimal (minor delays or obstacles when performing the task); S= Serious (major delays or obstacles when performing the task); C= Critical (not able to complete the task (or part of it) without assistance).

Results

Results of our usability test indicated that only in a few instances a rating of *minimal* was assessed to a task. A summary of the usability rubric is provided in table 2.

Table 2. Task by Issues

Evaluation Category	CP1	CP2	OS
Data Import/Export			
Import data		M	M
Export data	M	M	M
Data Manipulation			
View data		M	
Merge a file		M	M
Filter data		M	
Data Analysis			
Conduct a t-test			
Conduct an ANOVA			
Conduct a regression analysis			
Visualization			
Produce a graph (QQ plot for normality)			
Produce a graph (Histogram for descriptive)			
General			
Help documentation is easy to understand	M	M	M (2)
Help documentation is readily available			
Rater Expertise			
Novice			✓(2)

Table 2. Task by Issues

Evaluation Category	CP1	CP2	OS
Competent	✓(3)	✓	✓(3)
Expert			

CP refers to Commercial Product

OS refers to open source

Number in parenthesis indicates iterations

As the table suggest, CP1 had the lowest number of issues reported (2), CP2 had the highest (6). While the evidence suggests that CP2 had the highest number of issues reported, it should be noted that two testers considered their expertise with this package to be novice. Most of the issues arose in the categories of *data manipulation*, while *help documentation* appeared to be an issue with all packages.

From a qualitative perspective, tester notes revealed some of the nuances experienced while testing the statistical packages. Two testers admitted that *googling* for assistance or seeking help through a video sharing website (i.e., Youtube.com) was more helpful than using the help menus. Tester number one and two admitted to instinctively using other sources as a means for help rather than the help menu. Tester number two stated: *I wasn't sure how to rate this item because I use Google or YouTube for help documentation or try to figure it out on my own.* The open source statistical package help menu was found to be confusing as the help provided is mostly for users who have expertise in coding, and is not necessarily user friendly for those who are more comfortable using the graphic user interface. Tester two commented: *if you don't have programming skills it is difficult to understand how to apply it to your data/variables. I rated it M because the documentation is readily available in the software yet not easy to understand.* Data manipulation appears to be a matter of practice as tester number one suggested that while some of the functions were problematic, with repeated use, these issues would be surpassed.

Conclusions and Implications

The results of our usability test lead us to conclude that all three statistical packages are adequate to meet the demands and requirements of an office of institutional research. As such, we suggest that the decision to select a statistical software package for IR may very well lie within the scope of other factors (i.e., experience, fear of the unknown, cost), and not necessarily be based on whether the software can accomplish tasks germane to IR.

It should not be discarded that tester experience with a statistical software or user knowledge and/or expertise in statistics may have biased our results. As such we proffer this as a limitation. We also are mindful of the role we played as researchers designing the usability test and also in conducting the usability test of the software. This latter admission should also be considered a limitation of this study.

We feel that there is value in the results presented herein. The value of these results are potentiated by the transferability of the domain-users experiences and findings to other

users in similar positions (i.e., IR). We further assert that the findings from this study can aid in the decision making process of IR offices in regards to adopting a statistical software package suitable for their work.

We also recommend that more usability tests be conducted with information systems pertinent to institutional research. Offices of IR rely heavily on the use of information systems and it is imperative that well founded decisions be made in order to match their expectations with their office needs.

Lastly, we encourage offices of institutional research to seek out training opportunities that will enhance the knowledge and skills sets of their employees.

References

- Kushniruk, A. W., Patel, V. L., & Cimino, J. J. (1997). Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. *In Proceedings of the AMIA annual fall symposium* (p. 218). American Medical Informatics Association.
- Rogers, E. M. (2004). A prospective and retrospective look at the diffusion model. *Journal of Health Communication, 9*(S1), 13-19.
- Sjoberg, D. I., Anda, B., Arisholm, E., Dyba, T., Jorgensen, M., Karahasanovic, A., ... & Vokác, M. (2002). Conducting realistic experiments in software engineering. In *Empirical Software Engineering, 2002. Proceedings. 2002 International Symposium n* (pp. 17-26). IEEE.

Recommended citation:

Schmidt, M., Allayne-Bayne, G., & Xiao (2015). A Step-Wise Evaluation of Three Commonly used Statistical Packages in Institutional Research. Proceedings from 2015 TAIR Conference. *Texas Association for Institutional Research*. Addison, TX.