Evolution of Statistical Software and Quantitative Methods in Educational Research

Brandon LeBeau[1] & Ariel Aloe[1]

[1] University of Iowa

Author Note

Correspondence concerning this article should be addressed to Brandon LeBeau, Psychological and Quantitative Foundations, University of Iowa, Iowa City, IA 52245.

E-mail: brandon-lebeau@uiowa.edu

Abstract

Statistical software is the cornerstone of quantitative research studies and the availability and use of the software can greatly shape which methods are used by researchers. Software that is more accessible is likely to have more users and the methods implemented within the software limits the methods accessible to researchers. Open source software, (e.g. R), has reduced these barriers by making cutting edge statistical methods available to researchers through add-on packages. In addition, the idea of reproducible analyses has grown significantly within the statistics and medicine disciplines. This paper aims to explore the evolution of statistical software within educational research using a research synthesis to establish the state of affairs.

Keywords: Research Synthesis, Statistical Software, Quantitative Methods

Evolution of Statistical Software and Quantitative Methods in Educational Research

Objectives

The purpose of this paper is to explore the evolution (or lack thereof) of statistical software usage over time. As this usage is likely tied closely to the methods they are employing, the interaction between software usage quantitative research methods will also be explored. Research synthesis methods will be used to explore these trends over time in published educational research journals.

Theoretical Framework

Research Questions

1. To what extent has the statistical software usage shifted over time in published analyses?
    - If there is evidence of a shift, is there evidence this shift differs based on quantitative method or journal?
2. To what extent are published analyses citing statistical software?
    - Has this changed over time and across journals?
3. To what extent are open-source software tools used?
    - Is there evidence of reproducible analyses being employed?

Methods

Research synthesis methods will be used to explore the evolution of statistical software and quantitative methods in educational research. More specifically, the statistical software used for the analysis will be coded in additional to the specific quantitative methods (i.e. linear regression, hierarchical linear model, etc.). Additional meta data will also be coded including, journal, article title, author information, article keywords, and year

published. This information will be used to explore descriptive trends in the data over time, by journals, and methods.

The research synthesis will gather data from a handful of education journals that primarily publish empirical data analysis research. The search will not include journals that the primary focus is methodological, the use of software in these journals would likely be a different population. Therefore the following journals were selected to be searched from 1995 onward:

- American Educational Research Journal
- Educational Researcher
- Educational Evaluation and Policy Analysis
- Higher Education
- Journal of Educational Psychology
- Journal of Experimental Education
- Journal of Teacher Education
- Journal for Research in Mathematics Education
- Sociology of Education

Data and Software

All journal articles published between 1995 through 2016 will be organized into EndNote. Within EndNote, the find pdf feature will be used to gather the published documents from each journal. This pdf database will then be searched using the *pdfsearch* R package (LeBeau, 2016, R Core Team (2017)). This package allows for keyword searching directly within pdf documents. This will be the primary data collection method. The software keywords searched for can be seen in Table 1.

Table 1

Search keywords used in search of published journal documents

Search	Keywords
Software	"SPSS Statistics", "SPSS Modeler", "SPSS", "R", "R-project", "R project", "SAS", "JMP", "STATA", "MATLAB", "Statistica", "Statsoft", "Java", "Hadoop", "Python" , "Minitab", "Systat", "Tableau", "Scala", "Julia", "Azure Machine Learning", "Mplus", "LISREL", "AMOS", "BILOG", "BILOG-MG", "Multilog", "PARSCALE", "IRT PRO", "HLM[0-9]", "HLM [0-9]"
Quantitative Methods	Still formulating this list

A handful of articles will be randomly selected to be coded manually by reading the document to evaluate the accuracy of coding from the *pdfsearch* package.

Initial Results

Scholarly Significance

References

LeBeau, B. (2016). *Pdfsearch: Search tools for pdf files*. Retrieved from https://github.com/lebebr01/pdfsearch

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/