The Role of Data Analysis Software in Graduate Programs in Education and Post-Graduate Research

Michael Harwell
University of Minnesota

The importance of data analysis software in graduate programs in education and post-graduate educational research is self-evident. However the role of this software in facilitating supererogated statistical practice versus "cookbookery" is unclear. The need to rigorously document the role of data analysis software in students' graduate coursework and research, and in post-graduate educational research studies, is emphasized along with recommendations for obtaining this information.

The appearance of commercial data analysis software in the 1970s, such as SPSS (IBM Corp., 2015), SAS (SAS Institute, 2004), and Minitab (Minitab Inc., 2010) had a major impact on data analysis by enabling users with a range of statistical expertise to plan and perform analyses themselves rather than relying on statisticians or others for assistance. The software significantly enhanced the quality and impact of data analyses in numerous ways, including minimizing computational errors, allowing multiple analyses to be performed quickly, increasing the accessibility of complex statistical models and large datasets, and facilitating supererogated practice (i.e., employing recommended statistical practices such as using data to carefully examine the plausibility of underlying assumptions of statistical models). For example, examining how closely data follow a normal distribution using plots, tail weights, skewness and kurtosis statistics, and statistical tests of this assumption represents supererogated practice that increases the likelihood of drawing valid and replicable inferences based on statistical results.

The appearance of data analysis software was quickly followed by concerns it would encourage poor statistical practice. For example, a review of data analysis software manuals by Berk and Francis (1978) outlined several concerns which generally centered on researchers using software to perform and interpret analyses with little understanding of the statistical theory underlying a procedure or its assumptions: “People can now run analyses of variance, multiple regressions, factor analyses, etc., on their data with very little knowledge of statistical theories and assumptions” (McCarthy, 1978, p. 87); “We can now only hope that we are not fostering a widespread misuse of statistics by making it easy to run an inappropriate analysis without even having to look the other way when important underlying assumptions are spelled out” (Schucany, 1978, p. 93). These concerns are consistent with the term "cookbookery” which Tukey (1962) used to describe statistical procedures applied with little or no understanding of the theory and assumptions underlying them, inviting misuse.

Similar concerns were voiced in the initial and final reports of the American Psychological Association (APA) Task Force on Statistical Inference (Wilkinson & APA Task Force on Statistical Inference, 1996, 1999).

> Elegant and sophisticated computer programs have increased our ability to analyze data with substantially greater sophistication than was possible only a short time ago. The ease of access to state-of-the-art statistical analysis packages, however, has not universally advanced our science" (p. 4),

And,

> More important than choosing a specific statistical package is verifying your results, understanding what they mean, and knowing how they are computed. .... Do not report statistics found on a printout without understanding how they are computed or what they mean" (p. 598).

The concerns of McCarthy (1978), Schucany (1978), Wilkinson and the APA Task Force on Statistical Inference (1996, 1999), and others speak to the potential of data analysis software to facilitate poor statistical practice (i.e., cookbookery). Examples of poor practice include lack of attention to underlying assumptions of a statistical model such as normality and homoscedasticity (i.e., variances computed for a dependent variable are constant across values of an independent variable or predictor), inappropriate treatment of missing data (e.g., imputing missing data without attending to how data are missing such as missing in a completely random fashion versus missing in ways related to variables in the analysis), and unpropitious analyses (e.g., estimating and testing parameters without attending to measurement scales such as whether a variable possesses an interval or ordinal scale). A common thread in these examples of poor practice is using software to perform an analysis without regard to whether it is appropriate, thus potentially compromising inferences. Cookbookery increases the likelihood of biased parameter estimates, invalid inferences in the form of Type I or Type II errors for statistical tests, and unreplicable findings (Stodden, 2015), as opposed to supererogated practice, which would increase the likelihood of valid and replicable statistical results.

Research Documenting the Role of Data Analysis Software in Graduate Programs in Education

Most graduate students in education use data analysis software in their coursework and research in some capacity. A natural source of research-based evidence of the nature, frequency, and impact of this software is statistics education but this literature is centered on secondary and post-secondary statistics instruction and learning, leaving only non-research-based recommendations for incorporating data analysis software into graduate level statistics instruction (e.g., Brogan & Kutner, 1986; Ekmekci, Hancock, & Swayze, 2012; Smith & Martinez-Moyano, 2012; Thisted, 1979). Thus the role of this software in facilitating either supererogated practice or cookbookery in graduate programs in education has not been rigorously documented.

On the other hand, the anecdotal experiences of educators and researchers in statistics instruction, service on committees for student theses and dissertations, consulting with students, etc., provide multiple opportunities to assess the role of software in promoting supererogated practice or cookbookery. The nature of anecdotal evidence limits inferences but it seems likely many educators and researchers would conclude there is evidence of both supererogated practice and cookbookery and these are facilitated by data analysis software although its nature, frequency, and impact is unclear. What is needed is rigorous documentation of the role of data analysis software in graduate programs in education. Evidence that supererogated statistical practice is common would suggest few changes are needed whereas evidence this software plays an important role in promoting cookbookery would suggest non-negligible changes are needed, such as a renewed focus on using software to promote supererogated practice in graduate statistics coursework.

Research Documenting the Role of Data Analysis Software in Post-Graduate Educational Research

The introduction of software in the 1970s produced dramatic increases in the volume and complexity of data analyses as documented in summaries of statistical procedures appearing in educational journals from 1971 through 2010 (Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Keselman et al., 1998; Kiefer, Reese, & Thompson, 2001; Koppe & Dammeyer, 2014; Warne, Lazo, Ramos, & Ritter, 2012). Direct evidence of the role of data analysis software in promoting supererogated practice versus cookbookery in post-graduate educational research is not available but indirect evidence can be found in reviews of statistical practice in published articles (e.g., Dedrick et al., 2009; Fath, 2014; Keselman et al., 1998; Kiefer et al., 2001; Namasivayama, Yana, Wong, & Lieshout, 2015; Osborne, Kocher, & Tillman, 2012).

A focus of many methodological reviews is evidence of model checking, informed in part by the guidance provided by Wilkinson and the APA Task Force on Statistical Inference (1999): “You should take efforts to assure that the underlying assumptions required for the analysis are reasonable given the data. Examine residuals carefully” (p. 598). Data analysis software plays a critical role in model checking via data cleaning, plots and simple descriptive statistics, regression diagnostics, and specialized procedures. Data cleaning focuses on detecting, correcting, and/or removing inaccurate records or values such as outliers (Van den Broeck, Cunningham, Eeckels, & Herbst, 2005) and data cleaning software is available through programs like OpenRefine (formerly Google Refine), Trifecta Wrangler (formally Data Wrangler), and the R package Tidyr (R Core Team, 2015). Plots and descriptive statistics such as QQ-plots and distribution tail weights along with regression diagnostics (Neter, Kutner, Nachtsheim, & Wasserman, 1996) are available in several programs, whereas more specialized model-checking procedures are typically limited to a few programs, for example, Mahalanobis distance statistics for assessing normality of random effects in multilevel models (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011).

Available evidence suggests data analysis software has not produced widespread model checking, raising questions about its role in promoting supererogated practice. For example, Dedrick et al. (2009) reported 4% of their sample of studies considered outliers, Fath (2014) reported 12% of sampled studies described model-checking results, Osborne et al. (2012) reported 10%-32% of sampled studies checked various distributional assumptions, and Namasivayama, Yana, Wong, and Lieshout (2015) reported 23% of their sampled studies checked the assumption of normality. Relatedly, Veldkampf, Nuijten, Dominguez-Alvarez, van Assesn and Wicherts (2014) argued that the tradition in psychological research of relying on one person to conduct all data analyses and write-up the findings has played an important role in the quality of the statistical work, and reported 60.3% of data analyses in their sample of studies were conducted by one person. A similar tradition likely exists in educational research.

Methodological reviews generally equate missing methodological information in an article, such as evidence of normality, with poor statistical practice. It is possible editorial policy plays a role in the omission of methodological information (e.g., model-checking information is purposefully omitted due to space limitations) but there is no evidence of how often this occurs. On the other hand, it is likely the absence of methodological information is sometimes attributable not to editorial practice but to researchers with minimal understanding of the assumptions underlying a statistical procedure or researchers who understand the assumptions but choose to bypass this step, both of which are consistent with poor statistical practice. For example, credible inferences based on using software to fit a multilevel students-within-classrooms model to continuous cross-sectional data requires evidence of normality of classroom residuals (random effects) to help ensure estimates of variance component(s) and associated significance tests are not biased. A failure to report model-checking information whether due to cookbookery or an editorial decision represents poor practice.

Anecdotal evidence is again the main source for evaluating of the role of data analysis software in facilitating either supererogated practice or cookbookery in post-graduate educational research. This is available through the experiences of educators and researchers as reviewers for journals, books, and conferences papers, and membership on panels reviewing grant applications, etc. Although the nature, frequency, and impact are unclear, it is probably fair to again conclude that data analysis software facilitates both supererogated practice and cookbookery in post-graduate educational research. Evidence of supererogated statistical practice suggests data analysis software is playing an important role in increasing the likelihood of valid and replicable findings, whereas evidence of poor practice should prompt substantial changes, such as those suggested by the APA Task Force (1999), which hoped their report would “induce editors, reviewers, and authors to recognize practices that institutionalize the thoughtless application of statistical methods" (p. 603). In addition to Wilkinson and the APA Task Force on Statistical Inference (1999), several resources are available to discourage thoughtless applications of statistical methods (facilitated by data analysis software) including the APA Publications and Communications Board Working Group on Journal Article Reporting Standards (American Psychologist, 2008), Osborne (2013), and What Works Clearinghouse (2014).

Recommendations

The need for direct evidence of the nature, frequency, and impact of data analysis software in promoting either supererogated practice or cookbookery in graduate programs in education and post-graduate educational research points to several recommendations:

> **The role of software in promoting supererogated statistical practice versus cookbookery in graduate programs in education should be rigorously documented using multiple methods**. These include (a) case studies or other qualitative methods, (b) online, paper-and-pencil, or telephone surveys of students and faculty, (c) evaluation studies of the impact of data analysis software, and (d) meta-analyses. The latter might involve coding student work in conference papers, technical reports, and master's theses and dissertations on characteristics such as data analysis software used, type of statistical analysis, topical focus, nature and extent of a student's statistics coursework, program size, institutional characteristics, and evidence of supererogated practice with the latter serving as an effect size (e.g., creating a checklist of items reflecting supererogated practice with the proportion of checked items serving as an effect size).
>
> **Documenting the role of software should involve sampling a variety of graduate programs in education.** Properly characterizing the role of data analysis software should involve obtaining information from multiple programs varying in (a) topical focus (e.g., learning and cognition, special education, measurement and statistics), (b) statistical course offerings (e.g., number of statistics courses offered, introductory and specialized statistics courses), (c) program size reflected in number of enrolled students, and (d)

institutional characteristics (e.g., enrollment of fulltime students, selectivity based on average GRE scores, whether an institution is publicly or privately supported).

The role of data analysis software in promoting supererogated practice versus cookbookery in post-graduate educational research should be rigorously documented. This work could take various forms (e.g., qualitative methods, meta-analyses) and should represent a deeper probing of the nature, frequency, and impact of data analysis software in promoting supererogated practice versus cookbookery than that provided by existing methodological reviews.

Evidence data analysis software promotes poor statistical practice should prompt changes in its role in graduate programs in education and/or post-graduate educational research. Changes in graduate programs in education might include a greater departmental focus on using data analysis software in coursework and student research in ways that promote supererogated practice, a separate course devoted to using software that is centered on facilitating good statistical practice, and an important role for organizations like the American Educational Research Association (AERA) and its divisions (e.g., Measurement and Research Methodology, Learning and Instruction) in promoting good statistical practice. The latter might involve a special issue of an AERA-sponsored journal (e.g., Educational Researcher), invited symposia at the annual AERA meeting, and funding for research conferences devoted to improving statistical practice among graduate students in education. Perhaps the most important change in post-graduate educational research would be to require authors who submit papers to journals to explicitly address how their work incorporates good statistical practice. There may also be an important role for the argument of Veldkampf et al. (2014) of the value of co-piloting in both graduate programs in education and post-graduate educational research. Having at least two persons involved in all facets of data analysis is an intriguing framework that could potentially enhance the role of data analysis software in promoting supererogated statistical practice.

Author Notes

Michael Harwell is a Professor in the Department of Educational Psychology at the University of Minnesota.

Correspondence regarding this article should be addressed to Michael Harwell at Harwe001@umn.edu.

References

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63*(9), 839-851. doi: 10.1037/0003-066X.63.9.83

Berk, K. N. & Francis, I. S. (1978). A review of the manuals for BMDP and SPSS: Rejoinder. *Journal of the American Statistical Association, 73*(361), 94-95. doi: 10.2307/2286517

Brogan, D. & Kutner, M. H. (1986). Graduate statistics service courses. *The American Statistician*, 40, 252-254.

Conference Board of the Mathematical Sciences. (2010). Survey of undergraduate programs in the mathematical sciences. Retrieved from http://www.ams.org/profession/data/cbms-survey/cbms2010-work.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79*(1), 69-102. doi: 10.3102/0034654308325581

Ekmekci, O., Hancock, A. B., & Swayze, S. (2012). Teaching statistical research methods to graduate students: Lessons learned from three different degree programs. *International Journal of Teaching and Learning in Higher Education, 24*(2), 272-279. doi: http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Elmore, P. & Woehlke, P. (1988). Statistical methods employed in *American Education Research Journal, Educational Researcher,* and *Review of Educational Research* from 1978 to 1987. *Educational Researcher, 17*(9), 19-20. doi: 10.3102/0013189X017009019

Fath, K. Q. (2014). *Reporting methods and analyses in higher education research: Hierarchical linear and OLS regression models*. Unpublished dissertation, Loyola University, Chicago, IL.

Goodwin, L. & Goodwin, W. (1985). Statistical techniques in *AERJ* articles, 1979-1983: The preparation of graduate students to read the educational research literature. *Educational Researcher, 14*(2), 5-11. doi: 10.3102/0013189X014002005

IBM Corp. (2015). *IBM SPSS Statistics for Windows* (Version 22.0) [computer software]. Armonk, NY: IBM Corp.

International SAS Institute. (2004). *SAS/STAT 9.1* [Computer software]. Cary, NC: SAS Institute.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, D., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their

ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*(3), 350-386. doi: 10.3102/00346543068003350.

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in *AERJ* and *JCP* articles from 1988 to 1997: A methodological Review. *The Journal of Experimental Education, 69*(3), 280-309. doi: 10.1080/00220970109599489

Køppe, S., & Dammeyer, J. (2014). The change and development of statistical methods used in research articles in *Child Development* 1930–2010. *Integrative Psychological and Behavioral Sciences, 48*(3), 332–340. doi 10.1007/s12124-014-9261-7

McCarthy, J. L. (1978). A review of the manuals for BMDP and SPSS: Comment. *Journal of the American Statistical Association, 73*(361), 86-89. doi: 10.2307/2286522

Minitab Inc. (2010). *Minitab 17 Statistical Software* [Computer software]. State College, PA: Minitab, Inc.

Namasivayama, A. K., Yana, T., Wonga, W. Y. S., & van Lieshouta, P. (2015). Quality of statistical reporting in developmental disability journals. *International Journal of Rehabilitation Research, 38*(4), 364–369. doi: 10.1097/MRR.0000000000000138

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Chicago, IL: Irwin.

Osborne, J. W., Kocher, B., & Tillman, D. (April, 2012). *Sweating the small stuff: Do authors in APA journals clean data or test assumptions (and should anyone care if they do).* Paper presented at the Annual Meeting of the Eastern Education Research Association, Hilton Head, SC.

Osborne, J. W. (2013). *Sweating the small stuff: Does data cleaning and testing of assumptions really matter in the 21st century*? Lausanne, Switzerland: Frontiers Media SA.

Raudenbush S. W., Bryk A. S., Cheong Y. F., Congdon R. T., & du Toit M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling* [computer software]. Chicago, IL: Scientific Software.

Schreiber, J. B. & Griffin, B. W. (2004). Review of multilevel modeling and multilevel studies in *The Journal of Educational Research* (1992-2002). *The Journal of Educational Research, 98*(1), 24-34. doi: 10.3200/JOER.98.1.24-34

Schucany, W. R. (1978). A review of the manuals for BMDP and SPSS: Comment. *Journal of the American Statistical Association, 73*(361), 92-93. doi: 10.2307/2286522M

Smith, A. E. & Martinez-Moyano, I. J. (2012). Techniques in teaching statistics: Linking research production and research use. *Journal of Public Affairs Education,* *18*(1), 107-136.

Stodden, V. (2015). Reproducing statistical results. *Annual Review of Statistics and Its Applications, 2,* 1-19. doi: 10.1146/annurev-statistics-010814-020127

Thisted, R. A. (1979). Teaching statistical computing using computer packages. *The American Statistician, 33*(1), 27-30. doi: 10.2307/2683065

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics, 33*(1), 1-67. doi: 10.1214/aoms/1177704711

Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Medicin*e, *2*(10), 966-970. DOI: 10.1371/journal.pmed.0020267

Veldkamp, C. L. S., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A. L. M., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PLOS one, 12*(9), 1-19. doi:10.1371/journal.pone.0114876

Warne, R. T., Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly*, *56*(3), 134–149. doi: 10.1177/0016986212444122

What Works Clearinghouse. (2014). *Procedures and standards handbook* (Vol. 3). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf.

Wilkinson, L. & APA Task Force on Statistical Inference. (1996). Initial report. Retrieved from https://www.apa.org/science/leadership/bsa/statistical/tfsi-initial-report.pdf.

Wilkinson, L. & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604. doi: 10.1037/0003-066X.54.8.594