Statistical Software for Curriculum and Careers

William C. Adams and Donna Lind Infeld
George Washington University

Carli M. Wulff
U.S. Department of Health and Human Services

Abstract
What statistical software programs should be taught in public affairs graduate schools to best prepare students for prospective jobs? A nationwide survey of MPA, MPP, and related Masters programs found that introductory statistics courses most often use SPSS. In later statistics classes, MPP programs use Stata more often than SPSS. Budget and finance courses almost always employ Excel. Those priorities were fairly congruent with the market. Relevant employers on 35 job websites requested familiarity with Excel far more often than any other statistical software. SAS and SPSS were nearly tied for second place, and Stata was third. SPSS is used for data analysis in a plurality of relevant academic publications, but the use of Stata has increased sharply over the past 15 years. As public affairs faculties evaluate their curricula, findings from this research on program practices and employer priorities can help inform their deliberations.

Over the past several decades, students of public administration, public policy, and public affairs have often been taught statistics using various computer packages, such as SAS, SPSS, or Stata. Most MPA, MPP, and related Masters programs require at least one course in statistics and in budget and finance (Infeld & Adams, 2011; Koven, Goetzke, & Brennan, 2008; Morçöl & Ivanova, 2010; NASPAA, 2009). But in the data-rich, computer-based world of the 21st century, what is being done in those quantitative courses regarding statistical software? Despite all the student and faculty time and resources devoted to these important, challenging courses, there has been surprisingly little published research that can help inform

Keywords: Excel, SPSS, Stata, statistical software
faculty decisions on statistical software packages. To explore this topic, we conducted a multi-method study addressing the following questions:

1. Is there any prior research that indicates a particular statistical software program is superior?
2. Which statistical software programs are most widely integrated into MPA, MPP, and related Masters programs?
3. What statistical software skills, if any, do relevant employers specify in job announcements?

**Prior Research**

Despite the growing emphasis on outcome evaluations in public affairs education (e.g., Aristigueta & Gomes, 2006; Fitzpatrick & Miller-Stevens, 2009; Newcomer & Allen, 2010; Powell, 2009), few published articles assess the comparative advantages of competing software, and only one outcome experiment was identified. One reason may be that basic features are fairly similar. Leading software such as SAS, SPSS, and Stata need to match their competitors and all run the most common and many arcane statistical procedures, even if a few exotic options are sometimes missing.

Most users take for granted that statistical software calculates precise results, but, due to programming shortcuts, that is not necessarily the case. Statistical software in the late 1990s produced a surprising amount of imprecision (Altman & McDonald, 2001). Several versions later, one extensive analysis (Keeling & Pavur, 2007) still detected mistakes (mainly with nonlinear regression and autocorrelation calculations) in the nine software packages tested (including Excel 2003, SAS 9.1, SPSS 12.0, and Stata 8.1) but found notable advances over earlier releases.


> Excel 2007, like its predecessors, fails a standard set of intermediate-level accuracy tests…[I]t is not safe to assume that Microsoft Excel’s statistical procedures give the correct answer. Persons who wish to conduct statistical analyses should use some other package.

Among the errors McCullough & Heiser identified were a flawed linear regression algorithm and output, erroneous nonlinear regression results, a nonrandom random number generator, and inaccurate t-tests (incorrect results, especially when missing data are involved; wrong p-values; and even mistaken labels). Shortcomings in Excel 2007 were also identified by Yalta (2008), Hargreaves & McWilliams (2010), and Almiron et al. (2010). No published critiques of Excel 2010 were found.
When it comes to price, Excel has a structural advantage, often pre-installed on computers or seemingly complimentary as part of the Microsoft Office suite. In recent years, price competition increased thanks to “SAS OnDemand for Academics” (online, free, academic access to many SAS applications), the lower (although fluctuating) cost of the student version of SPSS, and the “Small Stata” student product. However, site licenses for many academic lab computers still run into thousands of dollars and entail complex fee formulas. While those license calculations are beyond the scope of this study, they cannot be dismissed as irrelevant.

Has any prior research tested the comparative effectiveness of different software when teaching statistics? After all, as Moore observed (1997, p. 131), “Software designed for doing statistics is not necessarily well structured for learning statistics.” The only known randomized, controlled test was conducted among undergraduates in an introductory statistics course at Indiana University (Proctor, 2002). The dozen who used Excel scored higher than the dozen who used SPSS in terms of computational knowledge and slightly higher in conceptual knowledge. Other than this one small trial, no systematic studies of the user-friendliness, usability, or classroom impact could be found in the published literature. More comparative studies would have been expected by now, especially since the dominant pedagogical model for statistics had, by the 1990s, decisively moved from a passive lecture approach to more active student engagement in creative problem solving and practical applications, plus more interpretation and analysis beyond rote calculations (Moore, 1997; National Research Council, 1990, 1991).

All in all, prior research does not offer much guidance regarding the best software to employ in our quantitative classes. It raises questions about Excel’s precision but does not objectively identify other notable distinctions in terms of features, usability, price, and pedagogy to replace personal anecdotes. Certainly, arguments can reach the level of religious differences given some people’s attachment to and investment in their preferred statistics software package. Yet academic programs must still make decisions about software packages, so we sought to discover what programs across the country now require for their Masters students.

Program Practices

A nationwide online poll of 260 eligible NASPAA representatives was conducted in May–July 2011. Completed surveys constituted a total of 131 Masters programs affiliated with NASPAA (both accredited and not accredited). Responses encompassed over half \( n = 98; 52\% \) of the Master of Public Administration (MPA) programs, over half \( n = 16; 53\% \) of the Master of Public Policy (MPP) programs, most \( n = 5; 71\% \) of the Master of Public Affairs programs, and 12 of the many dozens of varied other Masters degrees related to the public sector. Responses from MPP programs also represent over one fourth of those with institutional membership in the Association for Public Policy Analysis and Management (APPAM).
Table 1.
*Statistical Software Use in Graduate Courses*

<table>
<thead>
<tr>
<th></th>
<th>Introductory Statistics Course</th>
<th>Later Statistics Courses</th>
<th>Introductory Budget &amp; Finance Course</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MPA Programs</td>
<td>MPP Programs</td>
<td>Other Masters</td>
</tr>
<tr>
<td>No course</td>
<td>2</td>
<td>2%</td>
<td>1</td>
</tr>
<tr>
<td>No software</td>
<td>3</td>
<td>3%</td>
<td>1</td>
</tr>
<tr>
<td>SPSS only</td>
<td>41</td>
<td>42%</td>
<td>7</td>
</tr>
<tr>
<td>SPSS + Excel</td>
<td>27</td>
<td>28%</td>
<td>1</td>
</tr>
<tr>
<td>Excel only</td>
<td>16</td>
<td>16%</td>
<td>2</td>
</tr>
<tr>
<td>Stata only</td>
<td>3</td>
<td>3%</td>
<td>4</td>
</tr>
<tr>
<td>Stata + Excel</td>
<td>3</td>
<td>3%</td>
<td>1</td>
</tr>
<tr>
<td>SAS only</td>
<td>1</td>
<td>1%</td>
<td>1</td>
</tr>
<tr>
<td>All other(^1)</td>
<td>2</td>
<td>2%</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>98</td>
<td>100%</td>
<td>16</td>
</tr>
</tbody>
</table>

\(^1\) Other than SPSS, Excel, Stata, and SAS.
Of the 131 programs, only two do not offer an introductory statistics course (Table 1). All but four programs teach statistics employing a software program. Before dismissing these exceptions, it should be noted that even a few respondents who do use software in their MPA program were skeptical.

We’re having a lively debate about whether any statistics package should be used. Do MPAs really need to be able to run regressions?…[Some alumni say] these software programs got them their first job. Others say it was a waste of three credit hours.

Most of our students don’t use stats on the job. Some use Excel but don’t need training from us. They get their real training on the job.

Despite a few such reservations, the overwhelming majority of these degree programs use statistical software. In the introductory statistics course, 97% of the MPA programs employ software, as do 100% of the MPP programs and 94% of the other Masters programs. Among those using software, most offer a companion computer lab (74% of MPA programs, 88% of the MPP programs, and 81% of other Masters programs). A majority of MPA programs (59%) and a very large majority of MPP and other Masters programs (94%) offer follow-up statistics courses; such courses almost always use statistical software and often include a computer lab.

An introductory budget and finance course is offered by large majorities (92% of MPA programs, 75% MPPs, 88% all others) and typically employs statistical
software (89% of those MPA programs offering such a course; 92% MPPs; 82% others), although most do not have an accompanying computer lab (only 30% among MPA programs; 45% MPPs; 14% others). Subsequent budget and finance courses are less common, but, if offered, usually use statistical software without accompanying computer labs.

Thus many students in these MPA, MPP, and other Masters programs are likely to have worked with statistical software in at least two courses and perhaps more since program evaluation, policy analysis, and capstone courses sometimes entail the use of statistical software as well. What software are these programs featuring?

For budget and finance courses, the answer is easy: Excel. Having long ago vanquished its challengers like Lotus 1-2-3, Quattro Pro, and PlanPerfect, Excel dominates the academic spreadsheet world without any major rival. Only a handful of programs do not use Excel in their budgeting and finance courses; a few add exposure to SPSS, Stata, or some other software along with Excel; two outliers use SPSS alone.

For introductory statistics courses, software choices are less uniform. SPSS is most widely taught. In MPA programs, a large majority (70%) use SPSS, as do a

Figure 1.
Software Used in the Introductory Statistics Course

Note: Totals for each type of Masters program sum to more than 100% because some courses use more than one statistical software package; most often, SPSS is coupled with Excel.
majority in MPP (63%) and other Masters programs (59%; see Table 1 and Figure 1). A plurality in MPA programs (42%) and MPP programs (44%) use SPSS alone, but it is also used in conjunction with Excel, especially in MPA programs (28%). In MPP programs, Stata supplants Excel as a main rival to SPSS. Other software (R, JMP, gretl, and Crystal Ball) was rarely mentioned.

For subsequent statistics courses, Stata emerges as a stronger contender, markedly surpassing SPSS among MPP programs and showing a nontrivial presence in MPA and other Masters programs (see Table 1 and Figure 2). SPSS continues to be the software of choice among MPA programs, used in over two thirds of the later statistics classes, sometimes along with Excel or other software.

Excluding Excel, respondents were asked to rank the demand for SAS, SPSS, and Stata among employers: “How would you estimate the current usage of these three software packages at the jobs your Masters students seek?” Over three fourths (76%) thought that SPSS was the most widely sought, far surpassing Stata (12%) or SAS (9%). However, in terms of trends, the position of SPSS is not so secure. Respondents were asked if they had “noticed any trends in the popularity of these software programs over the past decade.” Stata was the clear winner in perceived

Note: Totals for each type of Masters program sum to more than 100% because some courses use more than one statistical software package; most often, SPSS is coupled with Excel.

Figure 2.
Software Used in Later Statistics Courses
momentum; 52% said Stata has gained popularity, and only 6% said it has lost popularity. That net positive 46% compares with a net negative 15% for SAS and a net negative 11% for SPSS. The widespread impressions of Stata’s strides suggest that SPSS ought not to rest on its laurels. In this vein, one respondent commented:

We relied on SPSS as our stat package for years. However costs are making us consider alternatives. Since some of us use Stata in our own work, and it is higher education friendly, I think we will be moving more in that direction.

In response to questions about these three statistical software packages, some respondents volunteered that this trio was inconsequential because nothing matters but Excel. For example:

Most of our students are using Excel or similar, job specific applications, not these stats software.

In polling our students and graduates, [we found that] only Excel is used in the work world.

**Job Listings**

Even if all other things (such as cost and ease of instruction) are not equal, perhaps extra consideration should still be given to statistical software that will be more valuable after graduation. Why focus on software that is rarely used outside the halls of ivy?

Indeed, the literature on diffusion and network effects repeated indicates that—when it comes to adopting software and hardware—people place an enormous premium on connectivity and minimizing transaction costs (Brock, 1975; Economides, 1996; Katz & Shapiro, 1985). Working collaboratively, exchanging files, and discussing issues are all enormously facilitated when everyone is using a common language, network, and system. Once interconnectivity is established with a shared framework, a new improved framework must be dramatically superior, not just modestly better, to dislodge the advantages of network connectivity and overcome switching costs (Farrell & Saloner, 1986; Katz & Shapiro, 1986).

Surely having familiarity with a specific software package (sufficient to put on one’s resume) is an asset when applying for a job with an employer who uses that same software, especially when the employer advertises that such a specific skill is preferred. Which statistics software do relevant employers most often mention? Job announcements were collected from the following online sites (April 20–30, 2011):
Although hiring had not rebounded to pre-recession levels, searches for full-time jobs using the keywords SAS, SPSS, Stata, and Excel (always checking to confirm that Excel was not being used as a verb) yielded a total of 409 unique job listings. A close inspection of listings found that few other statistical packages were requested, so the summary here is confined to these four. Eleven job postings were eliminated as specialized outside an area appropriate for typical MPA and MPP students (e.g., actuary, statistician, information technology programmer). Of the remaining 398, only a dozen specified “proficiency” or “mastery” of at least one of these packages. Most descriptions used vague but less demanding phrases such as “familiarity with,” “skill with,” “knowledge of,” or often simply “experience with.”

Figure 3.
Relevant Job Announcement Requesting Software Familiarity (SAS, SPSS, and Stata only)
Because the precise distribution of cited software is likely to shift somewhat from week to week, the value of this snapshot is the big picture that it provides, and that picture is quite clear:

1. Relevant employer requests for experience with Excel \((n = 290)\) surpassed that of the other three statistical software combined \((n = 108\) positions for SAS, SPSS, or Stata). 
2. In the race for the second-place ranking behind Excel, no single software package dominates (see Figure 3).
3. SPSS and SAS outrank Stata (especially with half of the “Stata only” listings due to its popularity at just one organization, the Urban Institute).
4. Familiarity with SAS fits two thirds of the listings citing one of these packages (68%), SPSS is sufficient for almost two thirds (64%), and Stata satisfies about four out of 10 (38%).
5. Familiarity and experience with any two of these three packages would qualify an applicant for at least eight out of 10 of these openings where familiarity with a non-spreadsheet package is sought.

Searches of other databases were conducted at a later date (May 10, 2011) to test and validate the initial findings. Results confirmed the same basic pattern: Again, Excel excelled. Among the others, SAS and SPSS were consistently ahead of Stata.10

In theory, job announcements might mention specific software as examples but not as an exclusive list, and thus a candidate familiar with a rival package would still be on equal footing. If so, employers could easily avoid needlessly narrowing their recruitment by simply inserting “such as,” “e.g.,” or “etc.” before illustrative software. In fact, only 22% of those who mentioned SAS, SPSS, or Stata suggested any openness to alternatives, and a large majority (78%) did not indicate any flexibility. Thus it seems fair to conclude that an advertisement recruiting someone with familiarity with SPSS, for example, is indeed intended to hire someone with familiarity with SPSS. And very few of the job advertisements requesting Excel familiarity could be construed as open to alternative software (less than 1%).

Along with gauging the preferences of relevant employers, we examined statistical software used in academic research. After all, some Masters graduates go on to earn doctorates and pursue academic careers. Moreover, software used for faculty research signals preferences that may spill over, sooner or later, into the classroom. Using Google Scholar’s extensive collection of books and journal articles, full-text searches were run for relevant fields from 1996 through 2010. In the 2006–2010 period, over 153,000 scholarly publications cited one of these four software products: 30% SPSS (up slightly from 25% in the 1996–2000
period), 27% Stata (up from only 7%), 25% SAS (down from 37%), and 18% Excel (down from 31%).11 The striking surge in citations of Stata in both social science journals and in administration and policy journals helps explain the view of many academics, noted earlier, that the use of Stata has grown markedly over the past decade.

**Conclusion**

Overall, for purely pedagogical purposes in graduate courses, no compelling basis was found for promoting any particular one of the leading statistical software packages. Faced with the need to make choices, what are graduate programs doing, and what do employers want?

Our large nationwide survey of MPA, MPP, and related Masters programs found that Excel has a near monopoly in budget and finance courses. For statistics courses in MPA programs, SPSS dominates, but sometimes shares the spotlight with Excel. Among MPP programs, SPSS is the most widely used in the initial statistics course, but Stata is used more often in later courses. Other Masters programs were in between; SPSS was again on top, but both Stata and Excel also were popular. SAS, R, and other programs are rarely used in these Masters courses. These priorities, especially regarding Excel and SPSS, are not far out of line with what relevant employers want, although employer requests for SAS familiarity were not matched by much classroom exposure to SAS. Overall, employers sought familiarity with Excel far more than with any other quantitative software. Among non-spreadsheet options, SAS and SPSS skills were most widely requested.

Ideally, courses using statistical software will both strengthen students’ statistical literacy and provide software skills that will be of practical value in the future. Individual programs have to make software choices, and, given the time, expense, and energy devoted to teaching quantitative courses and the potential impact on job opportunities of graduates, the decisions are not insignificant. For curriculum decision-making purposes, this research offers some provocative findings but does not purport to offer a precise blueprint. However, some additional interpretations will be proposed for consideration.

To be fair, note that most of the employers did not advertise that they required any statistical software familiarity. For example, in the authors’ school’s “Career Central” website for MPA and MPP graduates, two thirds of the listings (485 out of 733) did not specify any software background. Maybe employers assume that good candidates these days are at least acquainted with spreadsheets, but perhaps some do not care, especially for certain jobs. Yet, for us to reject statistical software altogether requires a gamble that our graduates can find good jobs without being eligible for the roughly one third of the positions that do call for particular software skills, that statistical software familiarity would be a trivial asset on their resumes, and that they can acquire adequate statistical literacy without the active learning of engagement with software.
Specific software experience does appear to matter. Few of the job advertisements mentioning particular software hinted that substitutions would be acceptable. In today’s “buyer’s market,” employers can easily screen for particular skills and avoid the costs of remedial on-the-job training.

Excel has extraordinary currency. Employer requests for Excel experience surpassed that for all non-spreadsheet statistical software combined. And, despite its imprecision in some advanced calculations, as often documented in the journal *Computational Statistics & Data Analysis*, Excel’s accuracy has improved since the 1990s.

SPSS remains a highly attractive non-spreadsheet choice—both frequently requested by employers and widely taught in our graduate programs. Assuming something is to be said for the “wisdom of crowds” and the advantages of connectivity, SPSS is a prime choice. Unlike SAS, SPSS has an available menu format that may help with its popularity. SAS skills, on the other hand, could be an interesting strategic advantage since SAS is also popular with employers but is now rarely taught to MPAs and MPPs, thereby reducing the competition for some jobs.

Stata has made dramatic gains, almost quadrupling its share of academic citations in one decade, and now has a toehold among at least some prospective employers. It is also noteworthy that a majority of the MPP programs surveyed, along with many other programs, turn to Stata in intermediate and advanced statistics classes. Yet, while Stata may have an increasingly prominent future, it seems premature to jettison older packages that are still markedly more popular among employers.

Nationwide, the dominant practice in our graduate programs is “mixed methods”: MPAs tend to gain familiarity with Excel and SPSS, and MPPs tend to gain familiarity with Excel, SPSS, and Stata. Research reported here indicates that this is a highly defensible approach and should yield a broad range of job opportunities. Few employers request mastery or proficiency, so experience with Excel plus exposure to at least one statistical software package like SPSS may be the optimal configuration for MPAs. The additional statistics course that MPPs often take offers an opportunity to add Stata or SAS to their repertoire. Aiming for more research-oriented careers, that larger analytical tool chest may be of special value to MPP graduates.

Excel is a registered trademark of Microsoft Corporation, Redmond, Washington.

SAS is a registered trademark of the SAS Institute Inc., Cary, North Carolina.

SPSS is a registered trademark of IBM Corporation, Armonk, New York.

Stata is a registered trademark of StataCorp LP, College Station, Texas.
FOOTNOTES

1 During its acquisition by IBM, SPSS was known for a while during 2009 and 2010 as PASW (Predictive Analytics SoftWare) but has reverted to its SPSS name, which once stood for “Statistical Package for the Social Sciences.”

2 For example, SPSS lacks quantile regression, autoregressive conditional heteroscedasticity analysis, and generalized autoregressive conditional heteroscedasticity analysis. The only recent published comparison listing the features of many statistical programs (excluding Excel) regarding regression analysis, time series analysis, ANOVA, and selected other statistics was found in Wikipedia (2011) and confirmed basic feature comparability.

3 Excel seems to be making inroads beyond budget, finance, and business. Articles are appearing about ways to use Excel, usually with its Analysis ToolPak add-in, to teach statistics in fields as varied as psychology (Warner & Meehan, 2001), nursing (DiMaria-Ghalili & Ostrow, 2009), and engineering (Prvan, Reid, & Petocz, 2002). Likewise, a few statistics textbooks now focus on Excel (e.g., Dretzke, 2011; Carlberg, 2011), although some strongly argue against using Excel (or any other spreadsheet) as a vehicle for teaching statistics (Nash, 2008; Su, 2008).

4 Two personal evaluations argued that Stata was superior to SAS and SPSS, but both authors had written books about Stata (Acock, 2005; Mitchell, 2005).

5 This figure excludes a total of 19 representatives who were from NASPAA affiliates outside the United States, had bounced or invalid e-mail addresses, or had previously opted out of all online surveys conducted via SurveyMonkey.com. Eligible representatives who were not certain about software issues were asked to invite an “appropriate faculty colleague” to participate in the short survey.

6 Four programs conduct statistical software training in computer workshops that, although not formally linked to a specific course, were added to the computer lab total.

7 With direct network effects, the value of an interactive good increases as more people purchase the same good (see Katz & Shapiro, 1985). Common examples include telephones and fax machines. If few people have telephones or fax machines, they are not nearly as valuable as when many people have them.

8 Of the 10 most populous states, four (Georgia, North Carolina, Pennsylvania, and Texas) lacked keyword searchable listings for state government jobs or did not run full-text searches for any keywords. Overall, state government hiring in 2010 was not strong.

9 Six of the 10 most populous cities had online job listings that could either be searched for keywords or had so few jobs listed that they could be searched individually.

10 Data from these later, additional sites were not added to Figure 3 because of a few duplicate listings and a slightly different time period: NACElink.com (under “government”)—Excel 105, SAS 9, SPSS 7, Stata 3; PolicyJobs.net (with Anglosphere international listings)—Excel 133, SAS 49, SPSS 32, Stata 26; and the authors’ school’s “Career Central” website, which lists jobs from nonprofits, think tanks, consulting firms, and government—Excel 205, SPSS 20, SAS 16, and Stata 7.

11 See Adams (2010) for more on the advantages and disadvantages of Google Scholar. Google Scholar categorizes most public administration and public policy journals into its subject area called “Business, Administration, Finance, and Economics.” Political science journals are found, as expected, under “Social Sciences, Arts, and Humanities.” A sample of 150 articles using the word Excel or excel found that 56% were referring to the software, so that proportion was used to adjust the estimate of articles mentioning Excel.
Acknowledgments

The authors thank Stephanie Celinni, Dylan Conger, and Laura Minnichelli for their comments and suggestions. An earlier version of this paper was presented at the Association for Public Policy & Management Teaching Workshop on November 2, 2011, in Washington, DC.

References


William C. Adams, professor of Public Policy and Public Administration at the Trachtenberg School at The George Washington University, teaches research methods and applied statistics courses to both MPA and MPP students. He has been a frequent contributor to the *Journal of Public Affairs Education*. His most recent book is *Election Night News and Voter Turnout*.

Donna Lind Infeld is a professor at the Trachtenberg School of Public Policy and Public Administration at The George Washington University, where she is also director of the PhD program. She teaches graduate courses in policy analysis and research methods. Her recent publications have focused on public administration education in the United States and China.

Carli M. Wulff is a Presidential Management Fellow at the U.S. Department of Health and Human Services. She earned her BA at Hamline University and her MPP at George Washington University while contributing to this research. She served as a Peace Corps volunteer in Kyrgyzstan.