
Microsoft Excel™ As a Tool for Teaching Basic Statistics

C. Bruce Warner
Anita M. Meehan
Kutztown University of Pennsylvania

Many instructors use specialized statistics packages, such as SPSS or SAS/STAT in teaching quantitative and research methods courses. As an alternative to the specialized packages, Microsoft Excel™ offers charting capabilities and the Analysis ToolPak, which provides a set of data analysis tools beyond those normally found in spreadsheets. With Analysis ToolPak, Excel resembles what one would expect from a specialized low-level statistics program. We have successfully integrated Excel into our introductory statistics courses and developed a tutorial manual to guide students through various statistical procedures in Excel. We present student ratings of an intermediate form of the manual and its exercises and discuss the advantages and disadvantages of using Excel rather than a specialized statistical package for teaching basic statistics.

Similar to many instructors in the social and behavioral sciences, for years we relied primarily on specialized statistical software such as SPSS in our statistics and research methods courses. As computer technology has become ubiquitous and spreadsheet programs more popular, we wondered if a spreadsheet application was not a better medium for a first course in statistics. Keeping in mind that most students would be expected to use spreadsheets in their work places (Davis, 1997), we began to use Microsoft Excel™, a popular spreadsheet software package, in our introductory statistics courses.

Why Use a Spreadsheet Application Such as Excel?

One reason we embraced a spreadsheet application is that students need data analysis and computer skills that they can generalize and transfer to many contexts. Students are likely to work in environments in which general software skills, such as use of spreadsheet applications, are expected (Olsten Corporation, 1996). A recent survey found that prospective employers ranked spreadsheet skills second, behind word-processing skills, as highly desirable and expected skills for new hires (Davis, 1997). Moreover, 86% of the employers expected new hires to go beyond spreadsheet basics and be able to perform detailed analyses. Davis also found that employers highly valued the ability to create presentations and graphs.

Another reason we decided to teach with Excel is that we believe its simplicity facilitates student learning (Henle, 1995). Students can concentrate more on statistics and less on the mechanics of the software. Data entry is often easier in spreadsheet applications than in specialized statistical programs such as SPSS. Moreover, many students already have some familiarity with spreadsheets, thereby reducing the time

instructors must devote to software training. In our experience, the sophisticated output generated by advanced statistical packages can be confusing for students. The output generated by Excel often matches examples presented in introductory statistics textbook exercises or in classroom instruction.

Another advantage of Excel is that it simplifies the electronic exchange of assignments. When a student turns in an assignment on disk or via e-mail, the instructor can open the file, examine the contents, and attach comments directly to the file using the annotation tools. For instance, when a student turns in an assignment in which he or she has performed the wrong test, the instructor can insert a text balloon that points directly to the error and offers corrective advice. When we grade assignments, we save a copy of each text balloon into an open workbook, so that we can later paste the balloon into the workbook of another student who has made a similar mistake.

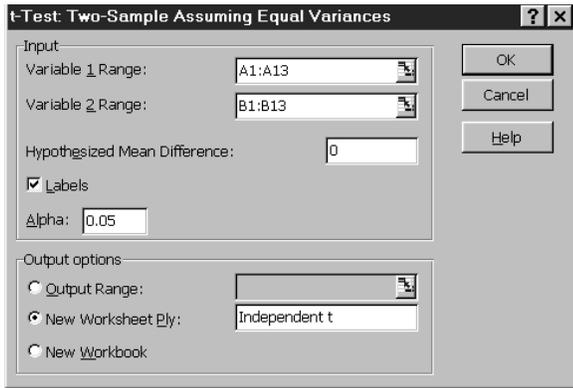
A final advantage of using Excel is economic. Students and universities already own or have access to computers with spreadsheet software, whereas specialized programs are an added expense. The extra cost also means that students have limited access to the specialized statistical software, as it is only available in selected campus locations.

Why Choose Excel?

We were familiar with Excel because it is the spreadsheet standard for our campus. Still, it is a natural choice for a statistics course, as it performs most basic analyses, creates easily customizable graphs using its Chart Wizard feature, and is the market leader in its product category.

Similar to all spreadsheet applications, Excel has the capability to solve both simple and complex equations. Excel's Paste Function (known as Function Wizard in versions earlier than Excel 95) simplifies the process by taking you through the steps necessary to paste statistical calculations into your spreadsheet. For example, you can use Paste Function to find common statistics such as mean, median, mode, standard deviation, variance, and correlation.

Excel also contains a less commonly known feature: Analysis ToolPak. Once you have performed the simple steps that install this add-in as a working part of your Excel application, you have an array of tools for performing descriptive statistics, *t* tests, ANOVA (one-way, one-way repeated measures, and two-factor), correlation, and regression. It also adds statistical and graphing capabilities that are unavailable in the basic Excel application. To illustrate the relative simplicity of



	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		<i>meditation group</i>	<i>control group</i>
4	Mean	4.25	5.583333333
5	Variance	0.75	1.537878788
6	Observations	12	12
7	Pooled Variance	1.143939394	
8	Hypothesized Mean Difference	0	
9	df	22	
10	t Stat	-3.05360497	
11	P(T<=t) one-tail	0.002911237	
12	t Critical one-tail	1.717144187	
13	P(T<=t) two-tail	0.005822473	
14	t Critical two-tail	2.073875294	

Figure 1. Independent-samples *t*-test dialog box (top) and sample output produced by the procedure (bottom).

the Analysis ToolPak, the top panel of Figure 1 contains an example of the dialog box that appears when you select an independent-samples *t*-test analysis. In the dialog box, “input variable range” refers to the spreadsheet location for the data from each group. In our example, we had data in two spreadsheet columns representing the hypothetical pain-pill consumption of a meditation group (A1–A13) versus a control group (B1–B13). From the dialog box, you can also choose to include column labels with the data being input. Labels help to clarify the output from the procedure. The Excel output appears in the bottom panel of Figure 1.

Additional resources exist to support instruction with Excel, including various statistical add-ins and simulations for Excel written in Visual Basic. Internet discussion groups have evolved around the topic of using spreadsheets in mathematics and science education, including teaching statistics. Good starting points for Excel resources on the Web are the Association of Statistics Specialists Using Microsoft Excel (ASSUME; <http://www.jiscmail.ac.uk/lists/assume.html>) and Spreadsheets in Education (<http://sunsite.univie.ac.at/Spreadsite/>).

Student Evaluation of Excel

Beginning around 1996, we incorporated a few Excel assignments into our introductory statistics course and developed accompanying handouts. For the Fall 1998 semester, we expanded our use of Excel and developed an instruction manual for students. The manual explained how to use the Paste Function, Analysis ToolPak, and Chart Wizard to perform

many of the basic procedures covered in an introductory statistics course. The manual included screen shots of Excel (e.g., those in Figure 1), sample problems with associated statistical output, and homework assignments. An example of a homework assignment follows:

Make up a data set that is appropriate for an independent-samples *t* test and label your two groups. Use 15 scores per group. Do the analysis and save the output as Sheet 4. Beneath the output generated by Excel, type in your interpretation of the results. This must include: the value of the observed *t* and its actual *p* value, the critical *t*, the means for each group, and a decision about the null hypothesis. Use alpha = .05, two-tailed. See the model provided for interpreting results under Step 4 of the section on computing an independent-samples *t* test using Excel.

We required students to generate simulated data to ensure that each submission was original and to provide students with practice in generating simple data sets appropriate for analysis. By examining the data and interpretations students had generated, we could often discover and correct misconceptions that students had about various statistical procedures. For instance, some students persisted in generating scores from two imaginary independent samples when performing a related-samples *t* test.

Students completed five required Excel assignments during the semester. In addition, almost all students chose to turn in at least one bonus assignment, with the majority completing three. To improve our teaching of statistics as well as our instruction manual, we decided to conduct a formal evaluation of student reactions to Excel.

Method

Participants

We surveyed 49 of the 52 students enrolled in our Fall 1998 Introductory Statistics classes who had used the manual and completed the Excel assignments during the semester. Participation in the survey was voluntary and anonymous.

Measures

We asked the students to express their opinions about the homework assignments and the course manual using a four-part questionnaire. In Parts A through C, students rated the assignments and manual. Students responded to semantic differential descriptions of the assignments (Part A) and the course manual (Part C) on a 5-point scale, ranging from 1 (*complete agreement with the description on the left*) to 5 (*complete agreement with the description on the right*). In Part B, participants responded to course goal-oriented statements about the Excel assignments. Ratings on Part B were also based on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The complete set of items and the mean responses appear in Table 1. Part D prompted students to provide free-form comments and suggestions.

Table 1. Summary of Student Responses Concerning the Fall 1998 Excel™ Homework Assignments and the Fall 1998 Excel Manual

Evaluation Item	<i>M</i>	<i>SD</i>
Assignments—Part A: Semantic differential items		
Extremely bad—extremely good	3.59	0.61
Boring—interesting	3.40	0.89
Difficult—easy	3.51	0.98
Unpleasant—pleasant	3.39	1.02
Not frustrating—frustrating	3.06	1.22
Imposing—friendly	3.65	0.81
Confusing—clear	3.43	1.00
Too much work—not too much work	3.84	0.85
Not useful—useful	3.86	1.04
Assignments—Part B: Likert-scale items		
The Excel assignments helped me understand statistical concepts better	3.37	1.04
The Excel assignments improved my computer skills	3.94	0.99
The Excel assignments allowed me to be creative	3.65	0.90
The Excel assignments required critical thinking on my part	3.69	0.96
The Excel assignments gave me skills I can use	3.90	1.02
Excel manual—Part C: Semantic differential items		
Extremely bad—extremely good	3.96	0.64
Not helpful—helpful	4.20	0.89
Hard to follow—easy to follow	3.69	1.07
Not detailed enough—too detailed	3.04	0.71
Too short—too long	2.98	0.32
Imposing—friendly	3.84	0.85
Confusing—clear	3.49	1.00
Not useful—useful	4.25	0.80

Note. $N = 49$. For the 5-point ratings on parts A and C, scores ranged from 1 (*complete agreement with the description on the left*) to 5 (*complete agreement with the description on the right*). The poles for several items on the survey were reversed to reduce participant response bias. In Part B, scores ranged from 1 (*strongly disagree*) to 5 (*strongly agree*).

Results and Discussion

Table 1 presents the results of the student ratings of the Excel homework assignments. Students indicated that the assignments were useful, not too much work, and required some degree of creativity and critical thinking. They expressed agreement that the assignments promoted the development of useful skills. Although students agreed that the assignments improved their knowledge of statistical concepts, they rated the assignments more highly in terms of improving their computer skills. Student frustration with the assignments was moderate, which we believe is a fairly reasonable level given the typical frustrations associated with any statistics assignment, plus the added headaches associated with using any software application. In many instances, frustrations were not specifically related to Excel but were tied to general computer issues such as getting access to campus labs or forgetting to save a homework file to disk. Reactions to the instruction manual were also generally positive, as shown in Table 1. Students indicated the appropriateness of both the length and comprehensiveness of the manual. They found the manual helpful for completing assignments and generally found it useful and user friendly.

Considerations for Using Excel to Teach Statistics

Deciding whether to use Excel rather than a specialized statistical package not only requires a careful comparison of

the specific merits of each software package, but also demands a methodical examination of student needs, course and department goals, and instructor expectations. Excel makes sense when time for teaching software is limited or access to specialized software is limited, when course goals include the development of general computer skills, when most students will enter the job market directly and not pursue graduate training, or when students come from a variety of departments that emphasize different statistical packages in training their majors. An instructor considering the adoption of Excel should also consider whether students would have an opportunity to learn a specialized statistical package in another course. In our opinion, Excel training should not constitute a student's sole exposure to computerized data analysis. Although we may use Excel in introductory statistics, we use both Excel and SPSS in our methods courses.

Excel simply cannot supplant specialized packages for two reasons. First, the standard set of statistical procedures for Excel is quite limited compared to packages such as SPSS. For instance, Excel can perform only three ANOVA procedures: (a) one-way ANOVA for independent samples, (b) one-way repeated-measures ANOVA, and (c) two-way ANOVA for independent samples. Excel's ANOVA procedures will not perform planned comparisons or post-hoc tests, and the two-way ANOVA requires equal N s in all cells. Second, the standard statistical procedures included in Excel are not as robust as procedures in specialized statistical packages (e.g., McCullough & Wilson, 1999; Sawitzki, 1994). Based on Excel's performance on a variety of data sets, McCullough and Wilson, and Sawitzki, recommended that Excel not be

used for scientific data analysis. They cited weakness in Excel's implementation of several algorithms, which will lead to serious errors on large data sets, especially those containing large values. McCullough and Wilson also observed that Microsoft does not correct flawed algorithms, even though the flaws have been exposed in earlier versions of Excel.

Because Excel has a limited variety of procedures and because it may produce inaccurate results on large data sets, it is simply not a good choice for advanced courses or general scientific data analysis. Realistically, though, most introductory courses will not cover many procedures beyond those included in Excel's Analysis ToolPak, and sample data sets will be no larger than those that can be computed conveniently by hand. Excel's performance on small data sets, in our experience, is generally very acceptable (cf. McCullough & Wilson, 1999). What Excel does offer for introductory courses is simplicity of both input and output, which Henle (1995) described as transparency, meaning that students can easily see the relation between the numbers and the results of operations performed on those numbers. Also, add-ins are available for Excel that may overcome limitations of both flexibility and accuracy. A list of available add-ins can be found at the ASSUME Web site (<http://www.jiscmail.ac.uk/lists/assume.html>).

Regardless of whether the instructor chooses Excel or a specialized package, students expect training on the software and expect guidance in completing computer assignments. In a recent survey of over 1,000 Cornell University undergraduates, students rated credit-bearing classes and peer support as the most effective means of acquiring computer skills (Davis, 1999). A small student focus group also expressed how frustrating and stressful it is when professors assume students possess certain software skills yet fail to provide students with support or training (Davis, 1999). At the same time, the focus group also indicated that students do not want professors to devote significant class time to teaching software at the expense of course content. Based on our experience, and consistent with the student views expressed in Davis (1999), we believe instructors need some sort of student guide (e.g., Berk & Carey, 2000; Dretzke & Heilman, 1998; Meehan & Warner, 2000) to help students complete assignments efficiently, effectively, and with minimal frustration.

Conclusions

We believe that Excel is a useful tool for teaching elementary statistics. Business schools have been at the forefront of using spreadsheet applications in educating their students, and many resources for business statistics exist (e.g., Levine, Berenson, & Stephan, 1999; Middleton, 2000; Neufeld, 1998; Pelosi & Sandifer, 1999), but spreadsheet-oriented statistics

guides for broader audiences have also been appearing (e.g., Berk & Carey, 2000; Dretzke & Heilman, 1998; Meehan & Warner, 2000). Individual instructors will need to weigh the costs and benefits of using a spreadsheet program versus a specialized statistical package, but we believe a spreadsheet, such as Excel, will prove more attractive in many situations.

References

- Berk, K. N., & Carey, P. (2000). *Data analysis with Microsoft Excel: Updated for Office 2000*. Pacific Grove, CA: Brooks/Cole.
- Davis, P. (1997). What computer skills do employers expect from recent graduates? *T.H.E. Journal*, 24, 74–78.
- Davis, P. (1999). How undergraduates learn computer skills: Results of a survey and a focus group. *T.H.E. Journal*, 26, 68–71.
- Dretzke, B. J., & Heilman, K. A. (1998). *Statistics with Microsoft Excel*. Upper Saddle River, NJ: Prentice Hall.
- Henle, M. G. (1995). Forget not the lowly spreadsheet. *The College Mathematics Journal*, 26, 320–328.
- Levine, D. M., Berenson, M. L., & Stephan, D. (1999). *Statistics for managers using Microsoft Excel*. Upper Saddle River, NJ: Prentice-Hall.
- McCullough, B. D., & Wilson, B. (1999). On the accuracy of statistical procedures in Microsoft Excel 97. *Computational Statistics & Data Analysis*, 31, 27–37.
- Meehan, A. M., & Warner, C. B. (2000). *Elementary data analysis using Microsoft Excel*. New York: McGraw-Hill.
- Middleton, M. R. (2000). *Data analysis using Microsoft Excel* (2nd ed.). Pacific Grove, CA: Brooks/Cole.
- Neufeld, J. L. (1998). *Learning business statistics with Microsoft Excel 97*. Upper Saddle River, NJ: Prentice Hall.
- Olsten Corporation. (1996). *Managing and developing the new workforce*. Melville, NY: Author.
- Pelosi, M. K., & Sandifer, T. M. (1999). *Doing statistics for business with Excel: Data, inference and decision making*. New York: Wiley.
- Sawitzki, G. (1994). Report on the reliability of data analysis systems. *Computational Statistics & Data Analysis*, 18, 289–301.

Notes

1. Support for the initial development of an Excel instruction manual was provided by a grant from the Kutztown University Professional Development Committee.
2. A version of this article was presented at the meeting of the American Psychological Society, Denver, CO, June, 1999.
3. We thank Megan Mumma, who student tested our tutorials. We also thank David Pittenger, Randolph A. Smith, and three anonymous reviewers for their helpful comments on earlier drafts of this article.
4. Send correspondence to C. Bruce Warner, Department of Psychology, Kutztown University, Kutztown, PA 19530; e-mail: warner@kutztown.edu.