PaceXL: Statistics Add-in for Excel

Robin Boyle, Faculty of Business and Law, Deakin University, Australia

Abstract PaceXL is a statistics add-in for Excel, meaning that it operates inside Excel using its own menu and toolbars. The full capabilities of Excel are retained, but PaceXL provides a significant range of extra statistical and graphical options.

PaceXL is designed for introductory-intermediate courses at the university level, but it also has widespread application for pre-university courses. It has a number of important adoptions in Australia and is now widely marketed internationally through two major publishers. It harnesses the power of Excel so that numerous complicated calculations and clear illustrative graphs can be generated with just a few clicks of the mouse. Students spend little time on computing, and instead can focus on, and enjoy, statistics.

The two Melbourne based authors have designed PaceXL to be easy to learn and easy to use, but with very powerful data exploration and model building capabilities. PaceXL is a professional looking package with applications in the workplace. Dialogs, calculations and charts are written in 'textbook' language. The approach to data analysis provides encouragement to beginning students, while preparing others for using more advanced statistics packages should they move into detailed research.

The paper examines some of the issues in statistics courses, as well as providing a summary of the key features of PaceXL and the desirable features of software for modern courses.

PaceXL: Statistics Add-in for Excel

PaceXL: What is it?

PaceXL is a statistics add-in for Excel, meaning that it operates inside Excel using its own menu and toolbars. The full capabilities of Excel are retained, but PaceXL provides a significant range of extra statistical and graphical options. Version 1 was published in 1999. Version 2 is due for publication at the end of 2002. (See publishing details later.)

PaceXL is a "number cruncher" for analysing large data sets, a teaching tool for illustrating difficult concepts, and a "text book" online (through its Help system). As a data analysis tool, it is also designed to be used in the workplace.

PaceXL has been developed principally for introductory-intermediate courses in statistics, especially business, economics, and management. The main market is for 1st and 2nd year university courses and MBA studies. But it is also suitable for other statistics courses (such as maths/science) and some TAFE and High School courses.

The authors of the software recognise that, in many university courses, only a small proportion of students proceed from introductory statistics to more advanced statistics. (On anecdotal evidence, the author estimates less than 5% in business and related courses). Thus a principal aim of PaceXL is to ensure that this large majority of students, who do not proceed to advanced studies, take away with them a useful understanding of the role of statistics in their further studies, in their research, in their workplace and in their personal lives. It also provides them with a software tool that they can confidently use for their statistical needs.

Statistics courses and the need for software

Statistics is becoming increasingly important because of the greater availability and diversity of data in modern society, and the consequential and naturally increasing role of data in decision making. Thus it is important that students and practitioners are well-grounded in the basics of generating descriptive tables, graphs and summary measures, in understanding statistical tests, in exploring for relationships, and in developing models.

Statistics has a reputation for being unpopular, no matter which field (business, psychology, biology, medicine, engineering, etc.). The author postulates that it is perhaps the most disliked subject ever. The explanation for its unpopularity may lie in statistics courses having a history of difficult to understand textbooks, poor teaching, being too mathematical, and lacking relevance.

Change has been required, otherwise there was the risk of statistics offerings being scaled back in many courses or being dropped altogether. Statistics courses had to be modernised and improved. Today's courses would seem to indicate that this is occurring: through better textbooks, instructors and software.

If statistics is to become more popular with students, to become more widely used in the workplace and to play its rightful role in decision making, then software is likely to be a key reason, if not the main one.

Role of software and data in modern statistics courses

A course in statistics without software is now almost unthinkable. Software means the focus can be removed from mathematics and formulae and directed towards the analysis of data and its role in decision making. Software means that statistical concepts can be easily explained by way of interesting data sets, while statistics courses can become more alive, interesting and relevant for students. Thus a key advantage of software is being able to teach statistics by concentrating on data.

However, the choice of package is important. Statistics software can be categorised in two broad ways:
- "Illustrative software" is specifically designed by its authors for illustrating concepts, such as the line of best fit, or the Central Limit Theorem.
- "Number crunching software" is designed to explore data, and to quickly generate tables, graphs and summary measures, carry out tests and generate models.

Illustrative software, while very useful for teaching and learning, has limited use for analysing data. Thus, once a technique or concept is understood by the student, that aspect of the software is not needed again. Number crunching software, however, has repeated use, since different data sets and tasks occur in a student's course or in the workplace. Such software is used time and time again. Illustrative software clearly has a place in statistics courses, but textbooks and courses seem to prefer number crunching packages. A key advantage of number crunching software packages is that they also often have illustrative features in them and/or they can be manipulated adequately to illustrate difficult statistical concepts.

The well-known commercial packages (such as SPSS, SAS, Minitab) are 'workhorses' for number crunching. The principal advantages of these highly regarded packages are speed of use for large data sets, fast switching between variables and between options, a wide range of graphical features, and the confidence the user has in their accuracy and thoroughness. Many statistics instructors would like to use such packages in introductory courses. However, such packages are often avoided as they are perceived to be designed more for research and industry use, they may be too powerful and too intimidating for new students, or the price may be excessive. Further, instructors would prefer they were more widely used/available in the workplace.

Because of its widespread use in the workplace (and in non-statistics subjects), business and other tertiary courses in statistics are switching to the use of Microsoft ® Excel as the main software. Further, Excel is widely owned by students for use on their home computers, and it is also widely available on institute networks for on-campus use.

PaceXL is principally a "number cruncher". But it also has significant capabilities as an "illustrator" of concepts. It is priced to be affordable for students and institutes.

Microsoft ® Excel and Excel add-ins

Excel has several features that can be used for statistics:
- Worksheet features
- In-built statistics functions
- Data Analysis add-in tools
- Chart wizard
- Pivot tables

But there is dissatisfaction in using "just Excel" (that is, in relying on just those features listed above). The reasons include:
- A dearth of statistical graphs and statistics options in Excel.
- Mistakes by the user, or repeating calculations, with the Data Analysis tools require repetition of the same steps over and over again.
- Inadequate or cryptic output.
- Users need to be very Excel competent, otherwise it is very time consuming to use for detailed statistical analysis.

- A Help system that itself requires some sophistication to use and learn from.

The end result may be that students spend too much time on Excel and not enough time on statistics. The subject is not necessarily made more enjoyable, more understandable or more relevant. It may be assumed that Excel was not designed for statistical data analysis, nor, in particular, for teaching statistics.

Thus instructors have sought more from Excel, with add-ins gaining favour.

Add-ins (*.xla files) are programs that can be written by independent authors using macros and Visual Basic in Excel. Add-ins must be installed on a computer, like other software, and are loaded into Excel when required. An add-in generally appears as a Menu item and may have dialog boxes and toolbars. Add-ins can improve Excel's options. They can add extra functions and graphs. They can fill out the rest of a statistics course. In summary, an add-in retains all the features of Excel, but "adds" to Excel's capabilities.

There are several statistics add-ins for Excel available through major publishers, and they may be included on CDs in the back of textbooks. These add-ins are improvements on Excel alone in that they add extra statistical functions and options. But otherwise, they generally still reflect most of the statistical deficiencies of Excel. Thus:
- The method of handling data sets and variables is cumbersome and time consuming.
- Mistakes or repeating an option require back-tracking through a whole series of steps.
- Output can be minimal, cryptic, inconvenient and not clearly linked to the data used.
- Graphs lack options and statistical variations.
- Dozens of worksheets may be added to a workbook file from just a relatively few calculations, leading to confusing output.
- Mistakes or bugs might occur, causing frustration and anxiety for the student.

PaceXL as a statistics add-in to Excel

PaceXL has been designed around the key points made earlier in this paper. In particular, it:
- Is designed for modern courses in statistics, focusing on data.
- Overcomes many of the data analysis deficiencies of Excel, and of other add-ins.
- Provides number crunching power (by operating like a mini version of the workhorse packages).
- It is thoroughly tested in order to be as error-free and trouble-free as possible.
- Is easy to learn and to use.
- Saves students a great deal of computing time, enabling them to focus on statistics.

Key features of the operations of PaceXL are as follows:
- PaceXL is an integrated system operating inside Excel, on standard PCs.
- PaceXL appears as a Menu item in Excel. A PaceXL toolbar is added.
- Dialog boxes are used for statistics routines.
- Coverage of techniques matches topics in most introductory textbooks.
- Help and tutorials are available from the main menu.
- The full capabilities of Excel are also retained.
- PaceXL uses the Data Area concept for all routines. Thus, once a Data Area is set, those variables are available for all statistics and graphs routines, until another Data Area is chosen.

A key strength of PaceXL is quick and convenient data exploration and analysis:
- PaceXL's Data Area and Results Sheets concepts are modelled on the Minitab/SPSS approach to data analysis of working with a given data set and storing results.

- By using a single Data Area with references by variable name, it is simply a matter of clicking variables on and off as required for an operation.
- Dialogs and routines are recalled by a mouse click, saving repetition of data selection and instructions.
- The amount of time saved via the Data Area and dialogs is very significant compared to other add-ins. For example, a five minute task in PaceXL might take one hour, say, with other software.
- PaceXL has a single Results Sheet with all calculations, tables and graphs generated underneath each other as they are performed. Unwanted output can be deleted. Notes can be added to the Results Sheet to annotate useful output. The Results Sheet can be renamed (for example, as "Question 1"); a new one will then be opened automatically.
- Unstack and select options enables exploration by sub-categories.
- There is an extensive range of statistical Graphs and Charts. Features include combination graphs such as boxplot-dotplot, histogram-normal and histogram-boxplot; live probability plots, such as the normal distribution with shaded areas; and a "scatter matrix" option whereby each variable can be plotted against each other at a speed limited largely by the time taken to click the mouse.

PaceXL has also been designed for teaching and learning:
- Output is written in familiar textbook language, with coloured output and graphs.
- An "on-top" Tutorial-Tours system means students can easily teach themselves how to use the add-in.
- The extensive Help system is like a textbook and provides a useful backup to hardcopy texts.
- The formatting and style of graphical and calculated output result in high calibre assignment work that students take pride in.
- PaceXL can be used to illustrate concepts in lectures or in computing laboratories.

Conclusions

Statistical software will play an increasing and very important role in statistics courses and in the workplace. Aimed principally for introductory-intermediate courses, PaceXL provides the student with a professionally oriented but easy-to-use package running inside Excel. A key advantage of PaceXL is that it enables teaching to focus on data rather than on formulae and mathematics. Further, the ability to explore and analyse large, interesting, real-life data sets, engenders the student with a sense of being a "statistical detective", as they look for meaningful patterns, relationships and explanations. PaceXL is designed to make the teaching and learning of statistics more enjoyable and meaningful, and the application of statistics more widespread.