Spreadsheets Statistics for Teaching and Production

Brian Murphy, G A Bartlett, Jane Klobas and Grant Keady - Perth, Australia

1. Introduction

We have found spreadsheets very useful for teaching statistics, operations research, and other quantitative methods in commerce: students learn quickly when typing and debugging formulae and macros, and get a real feeling for the relation between computation and theory. However, after a while, it is necessary to turn to a more traditional statistical package to find both standard and advanced procedures needed for real-world data examination. WESTAT Associates has developed the *MASS* system for some years, a stand-alone statistical package of some size and power, and we have used it extensively for teaching, at second course but not at introductory level. In the past six months we (1.5 persons) have ported it to sit on the new spreadsheet *WingZ* (by Informix Inc) on the Mac to form a new program *StatZ*, which combines the advantages of a spreadsheet and a specialised program for the teaching and practice of statistics. (See Section 3(ii) for other computers.) Porting a program from one language or machine or system to another is usually a painful exercise. However, the power and flexibility of the "scripting" and of the facilities for linking external code provided by *WingZ* and its *HyperScript* macro language, which we have used for incorporating our MASS code into their highly commercial, well-tested base product, seems destined to have a profound effect on related areas of program development. This will yield new research and teaching software, as already pioneered by Apple's *HyperCard*, the technical precursor of *WingZ*'s *HyperScript*.

2. Some teaching considerations

A basic teaching usage of spreadsheets in statistics is teaching simple formulae like those for the mean, standard deviation, confidence bounds, t, chi-values, etc., and

providing spectacular yet simple graphical methods of summary. Statistical packages may do many of these types of computations even better. However, in as much as canned routines can be too-readily called up without thought, such statistical packages can be less effective than a spreadsheet in helping the student to think about the options, and to derive insights when actually calculating with some numbers. A spreadsheet retains this teaching role, and also guarantees correct calculation of the numbers put into it: an unexpected answer can be carefully studied, and the problem's origin revealed - whether the concept, the formula, or the data is in error. This contrasts with the situation in many statistical packages, where the data is, in effect, unseen and unseeable, and the data's part in unexpected answers is often unimaginable without some sophistication in the user's understanding of both programming and mathematical technicalities.

Eventually, real data - and our commerce students possibly get to that faster than students taking a more theoretical mathematical statistics course - leads on to real statistical computing needs, and this is where our *StatZ* add-in to *WingZ* is important: see Section 2.4.

2.1 *Teaching the spreadsheet*

At the University of Western Australia, the spreadsheet approach has been used, successfully, for teaching statistics to commerce and other non-mathematical students. There is the economy that the students already know the spreadsheet for other purposes. Furthermore, the spreadsheet comes (as does the *StatZ* part) with good on-line helps, an item which is crucial in the teaching applications of software where provision of manuals, at the point of use, to very large numbers of users is impractical. Associated with economies of scale arising from the large number of users, there are good training videos, good books for one's university library, and so on. One of the *WingZ* training videos is very good indeed.

There are various levels at which a spreadsheet can be used. The easiest level at which to start is where one drives it, step-by-step. For repeated tasks it soon becomes clear that programming, *WingZ* scripting, is useful. *WingZ* has a learn facility, whereby *WingZ* records a script (and, if desired, displays it as it develops on a window on part of the screen), while the students perform the task they want scripted on their worksheet (in a neighbouring window). The learn facility of *WingZ* scripting, indeed, can be used both by the computer to learn what it is expected to repeat, and by the students to learn programming in a painless way in a platform they see as useful.

Incidentally, the teacher matters too! *WingZ* scripting, and the facility to put in control buttons, is useful for live demonstrations in front of classes. *WingZ* is used for "presentation graphics", and this is useful in the lecture theatre and tutorial room as well as outside in business.

2.2 *Elementary statistics, from WingZ without scripts or StatZ*

Returning to matters statistical, some of these commerce students seem to refuse to learn in the traditional "mathematical" way. However, they can be seduced to learn the required formulae and concepts by using a computer package they see as useful. Some important details we have successfully taught this way are:

(i) *t-tests.* The pooled estimate of error, and the F-test of the assumption of variance equality, are easily computed using a spreadsheet, and the nature of the relationship between them easily seen. The comparison between paired and unpaired t-tests is also easily illustrated.

(ii) *Analysis of variance.* It is very easy to construct one and two-way ANOVA using spreadsheet formulae. Of course, the equations are not general, usually being restricted to tables of a particular size, but the ideas generated are. It is also easy to construct the phases of the ANOVA analysis by sweep techniques, so that the relation of the model to the analysis becomes clear.

(iii) *t-test and ANOVA equivalence.* Using a simple two column data set, the students can show that $F = t^2$ when comparing one and two-way ANOVA respectively with unpaired and paired t-tests.

(iv) *Interaction in ANOVA.* The interaction SS (sum of squares) can be related to the row differences in data. The linked line graph of the observations can be displayed. One can even hear "oohs" and "aahs" of dawning comprehension from the class as the SS calculation, and the shape of the line graph, change as the data is modified!

(v) *Simple regression.* This is an obvious place to use a spreadsheet and linked graph. Unfortunately, in contrast to most good statistical packages - including *StatZ* - no spreadsheet alone gives easy outlier identification. (*WingZ* could be scripted to do so, but it may be better to move to a statistics package proper, *StatZ*, already coded and tested, rather than re-scripting.)

(vi) *Multiple regression.* Spreadsheets can build the formulae for the SSP matrix with *Copy Down/Right* and *Copy* and *Paste*, from only a few original formulae. However, from there on we usually move into the package world. The *WingZ* spreadsheet, however, has the further facility of providing simple matrix computation, whence the whole of multiple regression. (We don't show the students this until they have had one try at solving the normal equations by high school algebra.)

In all the above cases, the main advantage is that the student can easily compute - with all steps spelt out - and compare multiple approaches to the real problem in question, and usually understand what has been done. There is a further, often unrecognised, advantage - the students tend to work on the computers in groups when doing statistical assignments, and although this can lead to mindless copying, by some, of others' work, more often one has the gratification of seeing students working with camaraderie and obvious pleasure.

2.3 *Elementary statistics, WingZ scripts but without StatZ*

At a more sophisticated level, one may wish to teach students how to provide their "own" specialised mini-statistics package, by creating macros. This, incidentally, is nearly unrealistic in many spreadsheet macro systems, but not in *WingZ* where the scripting is relatively pleasant. By setting the students a succession of problems, one in a first week and say six the next week, which are the same except for the numbers, the students are motivated to script the task. The power of the *WingZ* scripting facility, which incidentally is a factor in certain special aspects of *StatZ*, also makes it possible

to expect more serious "programming" effort from statistics students, without demanding formal study of computing. An obvious sort of task, which we have found within the capacity of second-course students, is to "script" a one or two-way ANOVA procedure. This raises the computing content of statistics courses (and places extra burdens on teachers), but students seem to enjoy this challenge more than following the same steps by hand. It also teaches a useful level of programming skills on the side.

After the students have completed one such task, they appreciate the provision of a well-commented library of scripts for particular tasks, e.g. statistical graphics.

2.4 *More statistics, WingZ with StatZ*

StatZ is a compiled statistics package. It is decidedly much more than a library of scripts. The code is compiled in Pascal and C. For large data sets such compiled code runs faster than do scripts written for the same task.

When the inevitable point is reached where a formal statistics package must be employed, the introduction of the *StatZ* menu on the *WingZ* spreadsheet causes no worries to student or teacher. The ease of use of this spreadsheet thus extends to complex statistical problems, and with the scripting facility and linked libraries, students can be pressed to learn advanced topics by more direct means - including developments of the methods described earlier - than via a conventional package, though that (as in *StatZ*) still remains for comparison and advanced research. The spreadsheet's database structure eases problems of data storage, editing, and input/output, and makes access to substantial data such as arises in a real survey or scientific enquiry, a practical possibility for all students.

The *StatZ* menu fits seamlessly into the *WingZ* menu bar, and the items and sub-menus of each perform as a Mac user expects. Seamless as the appearance is, its design as a single item at the right top of the main menu bar, is a reminder that *StatZ* items call a physically different program to *WingZ*. Samples of the *StatZ* menus, and much more, are given in all the references at the end of this article.

3. Questions of porting

(i) *Porting a program to the Mac WingZ platform.* Macs are our students' first choice of machine, and we have 40 SE20s in our laboratory. (There are 30 substantially unloved IBM-type PCs there also!) Porting *MASS* to *StatZ* was only a smallish problem; *WingZ* is an effective development platform, possessing complete facilities required of a program hosting external code.

(a) The platform has its own language, compiler, and links to other (relocatable) compiled code by a standard third party linker. (The latter half of Chapter 11 of the *WingZ HyperScript* manual is the key reference here.)

(b) The platform provides a visual interface to accompany the structural interface enabling the new code to be accessed via the platform.

(c) Modular construction, segmentation, and overlaying exist to overcome memory constraints.

It is probably not surprising that, for mass-market microcomputers, the first really successful such platform has been developed on the Macintosh. Apple Computer have kept a close rein on Mac programming, providing many development tools which insist that standards are followed. In particular, most developers use the Apple *MPW* (Macintosh Programmers' Workbench) system, and most Mac compilers produce relocatable code with which the *MPW* linker can combine. Modular construction and the use of overlays or segments are integral with *MPW*, so expansion causes few problems.

In this context, it is also not surprising that menus too can be easily united, and that the same menu may call different segment modules. The strength of *WingZ* is that the modules called may be its own, or external, i.e. written in the user's favourite language from *MPW*. *WingZ* practices what it preaches in that it is constructed in the same modular way: thus facilities one becomes familiar with at the top level of usage of *WingZ* are seen as examples of what a lower level of *WingZ* makes available from users' scripts, and so on. It is so simple and so fundamental that it is surprising that this level of interface has not, as yet, been widely seen on micros (Unix mainframes had it 20 years ago). The monolithic construction of some of the competing spreadsheets will, we regret, delay significantly the time it takes for them to offer equivalent facilities. Further details are given by Murphy and Bartlett (1990a,b) describing the porting of *MASS* to *StatZ*.

(ii) *Porting to MSDOS and elsewhere.* *Unix WingZ* is available, and *Unix StatZ* is imminent. We look forward to being able to provide *StatZ* under *WingZ* on a range of machines.

As we also have 30 MSDOS machines in the Commerce Faculty laboratory, we have long had to consider MSDOS. For some years the software house MicroSoft has produced a suite of compilers and a unifying linker capable of making a first attempt at the collaboration described above within their MSDOS. Under their multi-tasking program Windows, it is possible to run our statistics package *MASS* and a database/spreadsheet in adjacent windows, and easily transfer data and output between them, interacting with many other programs' windows too. This is not as ideal a solution as the full integration on the Mac which led to *StatZ*, but those students whose computing background causes them to stay with DOS, which they have learnt elsewhere, find it suffices (and it is the best we have at present). Now that *WingZ* for MSDOS and OS/2 has been announced, with the interface mechanism to be retained as for the Mac, the corresponding ports of *MASS* to *StatZ* should be able to be completed.

A question remains as to when and which other spreadsheet developers will provide a similar modular linking facility. This is of great importance to potential third-party developers of software and courseware of all kinds: *WingZ* is the newest player in the crowded spreadsheet game and there is a strong installed base of competitor products headed by *MicroSoft*'s *Excel* (for the Mac) and *Lotus*'s *1-2-3* (for MSDOS). Whether or not *WingZ* becomes a market leader, and on what range of machines, we can already see it influencing its competitors. Brand loyalty, existing skills, and software already purchased, are often powerful factors, and many users will retain *Excel*, *Lotus*, etc. The rapid acceptance of *StatZ* by our user community, and the interest shown by other developers in our methods, highlights the facts that many more third-party developers will follow our path in producing their own specialist packages hosted by *WingZ*-type platforms. This will, for the Mac, be accelerated by the recently announced *XTND*

system, a "stand-alone" version of *HyperScript* available soon from Apple publisher *Claris*. In the face of such commercial pressures, we expect to see other programs, database and spreadsheet, moving quickly to provide these powers.

4. Statistical and other packages and the platform concept

A variety of platforms - though we hope not too many - will emerge, and some will survive. An example of a platform familiar to many professional statisticians is *S*, a platform, for Unix machines, with which one can link one's C or FORTRAN routines. *S* is a platform oriented to professional statisticians and aimed at exploratory data analysis uses.

We are convinced of the viability of spreadsheet platforms - covering a range of machines - like *WingZ*. The easy acquisition of a well-understood interface must have a profound effect on statistical (and other) package developments, both for teaching and production. It is unclear at this time, however, whether it will lead to many more small specialised programs, or substantially fewer with many alternative additions. Many authors will contribute these additions, not in tortured macro forms as often occurs at present for current packages (e.g. for *Excel*), but in the form of substantial original working code, so that the latter path of consolidation looks the more likely.

These movements may not affect the mainframe world greatly, where the large, old established packages may choose to do little more than providing tools for data conversion from spreadsheets to the input file formats the statistics packages use, but it must considerably influence emerging microcomputer-based systems. Such systems have a growing share of the market. The introduction of the *WingZ* linking concept is a factor ensuring that statistical computing will remain, for a considerable time ahead, very interesting indeed.

References

Klobas, J and Murphy, B P (1989) *The STATZ Manual.* Westat Associates Pty Ltd, Nedlands, Western Australia.

Murphy, B P and Bartlett, G A (1990a) Getting it flying on WingZ. *Australian Macworld*, January 1990.

Murphy, B P and Bartlett, G A (1990b) Statistical platforms - from dreams to reality. *Statistical Software Newsletter* **16(1)**, 2-6.

WingZ, MSDOS, Windows, Excel, Lotus 1-2-3, MASS, HyperCard, XTND and *MPW* are proprietary products of the software houses indicated.