Minitab: a retrospective?

James Currall
IT Services
University of Glasgow
j.currall@compserv.gla.ac.uk

Introduction

Minitab Statistical Software was developed in about 1972 by three folk who taught statistics at Penn State University. They were responding to the need that they saw in their teaching for a simple interactive tool to take the drudgery out of practical work. Much the same agenda that Robert Gentleman and Ross Ihaka had when they started development on R at The University of Auckland in the early nineties. I think that I first used Minitab in 1982, when I was looking around for something, anything, to enable environmental science students to stop worrying about doing calculations, which buttons to press on their calculators, or which columns of punched cards to put an X in to get the desired outputs. Minitab came as a breath of fresh air. At last we could structure sessions around what the students were really interested in as opposed to the technicalities of calculation, button pressing or card punching. Minitab has therefore, a good pedigree in relation to teaching statistics.

Progress

The first Minitab review that I wrote would appear to have been in 1993 and it was of Minitab 8. The conclusion of that review was:

“If you want your students to work with moderate sized data sets and want them to produce descriptive statistics, graphical representations of their data and use simple hypothesis tests or a toolbox of mathematical tools for data manipulation then you can’t do much better than the well tried and tested MINITAB. You can also have good regression facilities and simple analysis of variance. It is an excellent teaching package and very easy to use with lots of help. Don’t however expect a great deal beyond this, there are other packages that do more complex or specialised things better.”

All positive stuff, but with something of a warning in the final sentence, to which I will return later. A year later when the first MS Windows version appeared, my review concluded:

“Overall Minitab has been improved a great deal in this release. This is particularly true of the MS Windows version. Having used this version for nearly 9 months, I now find PC release 8 Minitab very frustrating in many little ways. Nowadays labs full of powerful PCs running MS Windows are popping up everywhere and the appearance of this version was most timely. Unlike some other packages Minitab has not forgotten that education is where it started and small touches like the inclusion of more data sets are appreciated.”
By 1997 we had progressed to Minitab 11 and:

"Each release of a statistics package makes things a little easier, but is enough done to meet our rising expectations? There are a number of things which are difficult to do in Minitab. Chi squared for contingency tables now supplies a p-value (about time too), but there are still no p-values provided with correlations.

Minitab has come a long way over the years in ease of use, but seems to be on the horns of a dilemma as to how many 'named' techniques to put on the menus. Put lots on and make it easy to do everything or put fewer on and make them less cluttered and leave the rest to the statistical sophistication of the user. The former approach has been taken with the Design of Experiments and Statistical Process Control and the latter with Regression and Analysis of Variance. Statistical sophisticates might say that if you have multiple regression, you can do almost anything, given the flexible toolkit approach which Minitab offers. This is of course true, but this is not the way that the menu and button bar culture which most of our students belong to see the world. The help system explains how you can work out partial correlations using Minitab, but this is going to leave many student users standing, as their statistical sophistication is not up to it and it can't be found on a menu. There is an argument for making things difficult to do so that they are less easily abused, but I am sure that this is NOT why there are no p-values with correlations."

---

"Courses for the reluctant frequently employ a fair amount of practical work to ensure that the students can apply what they have been taught. As teachers, when considering which package to use for a course of reluctant students, we have to think very carefully, because there will be enough in the course to put the students off, without the package that they have to master also being an issue."

Minitab has moved on by fits and starts over the years. The early versions had a surprisingly rich functionality and I still have working copies of Minitab 8.1 and Minitab 7. Progress was rather slow through versions 6-8, but with Windows things stepped up a gear and versions 9-12 followed each other fairly quickly and now things have slowed down again, with relatively little change in the analytical possibilities. Like many readers, most of my statistics teaching has been with groups of students who are doing statistics reluctantly. This is generally for one of two reasons, neither of which is because they really want to engage with statistics but they are doing it anyway. They are doing it because:

- it is a compulsory subject and they have been told that they will have an exam to pass at the end of it; and,
- they see the need to be able to give some statistical credibility to their research project in order to have any chance of having it accepted for their MSc/PhD or getting it into print in a reputable journal.

The key to continued suitability is therefore to be found in the question that forms the first sentence of the 1997 review for Minitab 11 (at the top of this page).

**Packages for courses**

Courses for the reluctant frequently employ a fair amount of practical work to ensure that the students can apply what they have been taught. As teachers, when considering which package to use for a course of reluctant students, we have to think very carefully, because there will be enough in the course to put the students off, without the package that they have to master also being an issue. Ease of use is clearly important, but the problem is that what one person might find easy, another might not. To a degree this is a function of how often the package is used. I know that I struggle with some relatively straight-forward software that I only use once or twice a year. In terms of course content and syllabus design there are perhaps two starting points. One can start from a point of view of:

- confining one’s teaching to what a particular package can handle in a straight-forward manner; and,
- working out the syllabus based on the techniques that are required to deal with the likely design and data needs of the students concerned.

Whichever end of the scale one starts at, there are likely to have to be some compromises, in order to arrive at course and software that together can approximate what is required.

For a long time I have used Minitab for teaching biology postgraduates. It is very easy to get started, which quickly builds confidence. It has a wide range of straight-forward graphs and even a spattering of dynamic graphics. Input and manipulation of data are now quite reasonable, especially since the introduction of multiple worksheets in a single project and the facilities to split, subset, etc. Minitab does things in bite-sized chunks, so it is not a question of wading through acres of output to find a simple piece of information as is the case with some other packages.

From where I sit, most packages are problematic. They are too difficult for a casual user to really get to grips with or too limited in scope, lacking important elements, etc. There has never been an ideal solution and there never will be
(except the one that I will never get around to writing). One can of course do almost anything in R, but casual users simply do not get enough exposure to its syntax to develop any degree of fluency.

**Trends in requirements**

One of the more difficult aspects of giving advice to people about the use of software in any field, is to help someone who is faced with being comfortable with what they know, but whose needs cannot now be fully satisfied with what they currently use. They have a choice:

- learn something new; or,
- constrain what they do to the limitations of what they already know.

In my experience, students now have variables which are glorious mixtures of categorical and continuous variables (and so they should) and have a much better understanding of what that means. As a result I no longer talk about t-tests, anova and regression, but treat such problems as General Linear Models, as per Grafen and Hails [1]. The students don't get any more confused than they used to and have started to move away from a ‘which test should I use’ type of thinking. Minitab has been very able to handle this change and has a very workable interface to such an approach. It was fairly easy to update notes (including example output) to reflect this change and required only minor change to examples, etc.

There have been some other trends in what is needed by students that have been rather less easy to accommodate. Two areas that give me trouble with regard to Minitab and ultimately may force me to move away from it are:

**Non-linear relationships**

Projects have increasingly been employing designs in which the responses are not linearly related to the explanatory variables and both students and supervisors have been turning to packages (such as GraphPad Prism) that can throw a wide (300+) range of functions at a set of data and find the ‘best fit’. It is relatively easy to explain the difficulties of such an approach and how other approaches are likely to produce more reliable and robust fits, but unless there is a simple linearising transformation, Minitab simply doesn't have any tools in its box to implement an alternative. It simply isn't good enough to suggest that if there is no suitable linear transformation that there is nothing that can be done and in order to do something, Minitab must be abandoned in favour of something else and the learning curve that is implied by the change.

**Non-normal errors**

A similar issue surrounds problems where the errors clearly are not normal and no simply transformation can help with the problem.

Maybe I should just confine what I teach to a more conservative agenda which would be relatively easily handled by Minitab, but I would rather that the students went away with a reasonable idea about how to deal with the actual experiments that they are undertaking in their discipline than give them a bunch of techniques that don’t quite fit with what they need to do.

“In my experience, students now have variables which are glorious mixtures of categorical and continuous variables and have a much better understanding of what that means.”

**Conclusion**

For me calculation is no longer what it is about (if it ever was), it is teaching students to be able to understand what their data are trying to tell them rather than a more mechanistic mode of operation. There may be an element of course redesign implied by this, given we now have tools that really do have the potential to take ‘the drudgery out of practical work. At various points, in reviews and elsewhere, I have drawn attention to facilities that Minitab does not have that it needs to in order to continue to be a tool of choice for ‘service’ courses. Over the years many improvements have been responses to such entreaties. Neither of the areas that I highlighted above are new, I have mentioned both of them on several occasions. I have found over the years that discussion of statistical packages can arouse very strong passions – rather like religion and politics. Everyone reading this will therefore have their own views on how to deal with the issues raised. We may well differ as to how we address such matters, but I am certainly getting close to a time when I am going to have to make a tough choice between teaching what I feel I ought to teach and sticking with what I have used for the best part of 25 years. A tough call – how long now until retirement?

**Reference**