

Letters to Editor

Data Analysis - Make it Easy

Stewart Andrews

VSN International

I was intrigued to read the report on 1st year practicals and their role in developing future bioscientists, and would like to contribute to the debate. As a developer of data analysis software for the bioscience sector, obviously our considerations are, by necessity, limited strictly to the data analysis aspect of any bioscience course, hence the highly specific focus of the following letter.

One of the underlying problems in bioscience is that quite a high number of first year undergraduates don't appreciate that they have to take data analysis or investigative studies as part of their chosen course. Obviously this isn't true for every student but for those students it is true for, part of the reason for this lack of understanding is the students themselves. From the students' perspective, they've enrolled on a course either because they just happen to be good at it, or in some cases, wish to pursue a personal interest in, for example animal behaviour, plants or the environment, but what they haven't done is read the small print. On the other hand, some students could rightly argue that quite often 'data analysis' doesn't feature as part of their course descriptions even when it is a compulsory module. But whatever the reason, for many it will have come as an unwelcome surprise to find that, far from leaving maths and stats behind them, data analysis is a compulsory part of their chosen course. This issue is further compounded by the fact that the data analysis practical that many first year students encounter can often have little relevance to their chosen course.

One solution is obviously to offer students practicals that are relevant to the issues we face today. This is a popular route taken by many universities. On the other hand, given the finite time and resources available to lecturers, I appreciate that not every university can afford the time and/or cost involved in sourcing up-to-date data sets and then developing relevant and interesting practicals for all first year students. For those universities, I would like to suggest another way that this might be achievable.

It's already common practice for many universities to develop partnerships with commercial agri-businesses or research institutes such as BBSRC and NERC. This is because it is a win-win for everyone involved: it makes good commercial sense for universities to support their PhD students by giving them access to current research provided by these partnerships. This kind of proactive and collaborative support helps universities to attract more students, the PhD students in turn get the opportunity and experience of working on real and relevant data sets, and in doing so improve their chances of getting a job with the institute or business that supplied the data; and of course the institute or business providing the data, benefits from being able to outsource its research cost-effectively to the country's best post-grad students.

Now, since this is already accepted practice, would it not make sense for the universities to make maximum use of this data by using it (albeit pared down) for their first year student practicals? By offering first year students real, relevant and therefore 'exciting' data, surely this would enable the students to see the purpose of what they are doing and so take their lab and practical work more seriously? At the same time, would it not offer some students an insight into a possible career path opportunity and in doing so, go some way to encouraging students of bioscience subjects to pursue a bioscience career and improve the percentage of students

remaining in bioscience after their degrees? I hasten to add here that I am not suggesting that universities do not already offer their students exciting data sets. I know that many do an admirable job. All I am merely trying to do is offer suggestions of approaches to those universities who do not have the resources.

The best tools

Another way to help motivate students is to ensure they are taught data analysis using the best tools available. By 'best' I mean, tools that offer unrivalled levels of support and guide users to the most correct and appropriate analyses for any data. When confronted for the first time with data analytic software, naturally students can feel overwhelmed by what may appear to be a whole 'new language' not to mention a completely new way of thinking that they need to get to grips with. In response, students can often resort to simply memorising which buttons to click and which menus to use to 'get the right answer', without necessarily understanding what they are doing and why. By 'progressing' their use of data analysis in such a manner, however, students miss the point that different analyses are required for different experiments. Worse still, they run the risk of making incorrect choices of analysis, which in turn are likely to produce invalid results. While choosing the correct statistical method for analysing a set of data can be daunting, it doesn't have to be - if students are encouraged to think critically from the very beginning. This can only be achieved with the *right* data analysis software.

GenStat for Teaching (www.vsni.co.uk/genstat/teaching) is a data analysis tool that has recently been made available free to educators and students worldwide. (GenStat is available to students up to a Masters). It developed from our belief that financial limitations should not be a barrier to gaining access to expert tools and follows in the footsteps of GenStat Discovery, a free version of our software that is available to all education and non-profit organisations in the world's poorest 100 countries. Indeed many students find their plight not wholly dissimilar to those in developing countries with limited budgets.

But of course, nobody should use a product just because it's free. The reason GenStat should be considered for use by the education market is because it has been professionally developed by a respected organisation with its roots in research and because it actively encourages critical thinking in users. Take for example its well-planned menus that offer a consistent 'analyse - check assumptions - save details' approach. This ensures users think logically and interactively about data analysis rather than race to get the right answer.

To help students think logically and fearlessly about data analysis, probably one of the first things that they need to appreciate is that they don't necessarily need to understand the syntax of the underlying statistics (that's what the software's for). With GenStat, you know that the underlying statistics is well-formed, accurate and reliable so there is no longer any need for long-hand manual calculations. However, what students do need to understand from the very beginning is the importance of knowing how to design their experiments. To achieve this, students will need a good grounding in the principles of statistics. The key principles students need to be grilled in are : recognising and understanding valid data; choosing appropriate analysis techniques; and critically analysing output. Another point I'd like to make in favour of GenStat's usefulness for education is the range of supporting materials that accompany the software. In addition to a whole host of start up guides and introductory manuals, GenStat also has hundreds of worked examples and tutorials to help students with their analysis.

But this isn't just about the students' experience. Lecturers also deserve to have tools that make teaching easier and relevant. While there is quite a number of software tools available to the lecturer, some that are free, others that require a hefty subscription fee, very few are 'fit for purpose' or are designed in a way that actually encourages students to take a greater interest

in what they are studying and helps them understand what they are doing and why. GenStat is one of these. Its history and current links with the bioscience industry across the world has meant that much of the terminology used within the software reflects the language used in all areas of bioscience: the language and documentation in GenStat is written from a bioscience perspective and even the results are displayed in precisely the ways required by bioscientists. Teaching students with the very software that is used in the real world will of course make the experience far more compelling.

At this juncture, I'd like to introduce a view from academic users overseas that may be of use. As a British company with an international user base, we come into contact with a great many universities from all over the world and naturally benefit from their feedback. It was with this in mind that I approached a number of our academic users in Australia for their insights on teaching first year practicals. Mario D'Antuono, lecturer in biostatistics at the University of Western Australia had this to say:

"Introducing statistics too early can run the risk that the students will never truly get into their core subject. In our experience, teaching statistics in the first year is not worth the effort largely because at this stage of their education students should be focus on understanding their core subject. It is better to bring statistics in as a module when students actually need it and can understand its relevance to their subject."

This begs the question to UK universities: while you may have the right components and the right modules on your respective courses, have you considered whether they are taught in the right order?

Finally, I would like to say should anyone reading this letter like to talk to us about their use of data analysis in their courses, we would be delighted to help. Being jointly owned by BBSRC via Rothamsted Research and NAG Limited, a not-for-profit software company, means we are uniquely positioned to provide expert commercial advice allied with a deep understanding of research and collaboration. And of course, being a UK-based company means we can also respond quickly to your needs.

Stewart Andrews
CEO, VSNi
www.vsnico.uk

Letters to the Editor are those of the contributor and do not necessarily represent the opinions of the journal editorship