Supporting statistics teaching and learning at A-level: Using computer-based materials.

Nicholson, James
Queen's University of Belfast, School of Psychology
University Road
Belfast BT7 1NN, Northern Ireland
j.r.nicholson@qub.ac.uk

Mulhern, Gerry
Queen's University of Belfast, School of Psychology
University Road
Belfast BT7 1NN, Northern Ireland
g.mulhern@qub.ac.uk

1. Introduction and background to the development project at Queen's University, Belfast

Over the past few years statistics has become increasingly prominent in national curricula across the world. In the United Kingdom the Government has attached a specific priority to the use of Information and Communications Technology (ICT) in schools as a vehicle for delivery of the curriculum.

While there may be a general consensus regarding the potential value of ICT in the teaching and learning of subjects like statistics, as yet there is little empirical evidence identifying whether, and how, the use of ICT improves the quality of teaching and learning in classrooms. In our view, it is not a cardinal truth that all ICT is inherently effective simply because it uses clever technology. Courseware must be much more than a perfunctory electronic manifestation of information through which the pupil has to wade. Such 'page-turning' is likely to achieve no more than the average textbook. In fact, more often than not, a textbook may well do the job better. In developing and integrating ICT into any school curriculum, careful consideration should be given as to whether the technology is providing genuine 'added value'.

The National Council for Educational Technology (NCET) (1996) identified six ways in which ICT can provide opportunities for students learning mathematics:

- Learning from feedback
  - the computer is non-judgemental and impartial, and so creates an environment in which students can make conjectures, test them and modify their own ideas.
- Observing patterns
  - the speed and accuracy of computers enables students to produce many examples, supporting the making and justifying of generalizations.
- Seeing connections
  - formulae, tables of values and graphs can be easily linked: changing one representation and seeing changes in the others helps students to understand the connection between them.
- Working with dynamic images
  - encouraging pupils to visualise the geometry as they generate their own mental images.
- Exploring data
  - work with real data - supports interpretation and analysis.
- Teaching the computer
  - encouraging the student to 'instruct the computer' using formulae and other code.

With the exception of the final point, which is not a focus of our project, these principles have provided a useful framework for the development of our materials. The rest of this paper provides details of those materials and ventilates a number of pertinent issues.

2. Our approach

Throughout the development of our materials we have been guided by several principles regarding style and presentation:

Navigation. As far as is practicable, users should be able to move around material flexibly and with no apparent restriction. At all times, users should 'know where they are' within modules and in the overall courseware scheme. In the case of our own material, we have attempted to achieve this by ensuring that the 'Home' button is always available and visible, and that it returns the user to intuitively sensible points depending on their current location - allowing access to any location in a given module with 3 mouse clicks. We have also used location buttons that are meaningful and transparent.

Language. The volume of text should be kept to a minimum, and informal and non-threatening language should be used at all times, albeit without compromising the statistical sophistication or accuracy of content.

Structure. Courseware should be more than a ragbag of information linked up in an *ad hoc* manner. Rather, the organisation of content should reflect a well-specified pedagogy and the structure of material should reflect the growth of students' conceptual understanding, as we would wish it. In our materials we have undertaken careful consideration of pedagogic factors and an interative process of formative evaluation of material.

Contiguity. There should be a close association between dynamic simulation and textual commentary in 'real time' and the integration of the two should facilitate the development of the appropriate conceptual knowledge. In short, the whole should be greater than the sum of the parts.

Data. Examples and simulations should use data that is meaningful and relevant to the user. The issue of 'real data' should be fully explored when constructing examples.

Self-monitoring. Students should have ample opportunity for personal reflection and self-assessment.

Throughout the development process, students and teachers have used the materials, and their views have been used as part of an iterative process of development, improving the navigation and the conceptual structure and determining the scope of the conceptual challenges to be addressed.

3. Outline of scope of materials

Our materials provide comprehensive coverage of the following:

- how the points within a sample behave and the amount of variability within a sample
- the similarities and differences in sampling behaviour based on different distributions and comparing different summary statistics, including sampling distributions
- different sampling methods
- how the relaxation of key assumptions in a distribution model may affect the appropriateness of the model
- the stochastic nature of the line of regression and the concomitant implications for its use in prediction and interpretation – by having a large number of 'recorded observations' from which the user can take repeated samples the roles of the sample size and of the underlying correlation between the variables is explored
- the discriminatory power of different sample sizes in identifying the existence of correlation between variables

We have adopted a modular structure so that the various aspects of sampling and estimation are dealt with together, as are the various aspects of regression and correlation, while sampling methods and modelling distributions form two other modules. All of these are fundamentally based on simulation and directed activities, and an overarching theme in all topics is variability.

Consolidation of ideas and self-reflection are encouraged through appropriate probe questions. Answers or comments for almost all questions are available on-screen at the click of the mouse, and the language used in these constantly reinforces the role of variability, and its nuances. Since simulations use randomly generated 'real data', the language used recognises that, where many students work through such simulations, occasionally unusual outcomes will be observed.

An indication of the style is given by the following extract from the Sampling and Estimation module, at the point at which the user is ready to explore the behaviour of the sample mean:

> *"You have considered how the points within a sample behave, the amount of variability which there is within a sample, and how one sample differs from another. It is quite complicated to compare one sample with another simply by looking at the values, or even a graph showing the values as points.*
> *One thing we can do is to take the sample mean as a way of summarising all the values in each sample. An inevitable drawback is that you lose information about the individual values, but the plus side is that you can readily make comparisons between different samples."*

4. Conclusions and future directions

As our project draws to an end, we have had an opportunity to consider some fundamental questions and issues:

Was it worthwhile? If we have learnt anything, it is that development of ICT materials cannot be done on a whim. Given the resource implications, it is essential that specific needs are identified, full consideration is given to pedagogy and the approach should ensure genuine added value, rather than merely adding 'bells and whistles' to what could be achieved without ICT.

What are the practical issues regarding integration into the curricula? Our view is that materials should be as technically generic as possible, both in terms of platform (Mac / PC) and use of commercial software. Hunt and Tyrrell (2000) have outlined the pros and cons of using web-based resources for learning statistics, with recent advances in the technical integration of web pages and applications software such as Excel allowing a much more flexible environment than was previously available. Because Excel can operate smoothly in a web-based environment, and is universally available in our schools through the Government provisions for the use of ICT, it provided the accessibility we sought for our materials.

Software such as Minitab, or more recently Fathom and Autograph provide an excellent environment for statistical exploration and analysis, but not all schools will have access to them, nor do they have teachers who are familiar with their use. Even to provide activities for students to program themselves has difficulties, as features may vary considerably between versions of the same software, and of course currently the notation for functions is usually very different than that employed for conventional mathematics. Of course, technology is not standing still. In the future, publishers will be able to provide web-based support so that different platforms can be catered for, and the development of different user-interfaces should allow mathematical notation to be less of a stumbling block than it is currently.

Real data or not? This is an important question, in both practical and philosophical terms. Hodgson and Burke (2000) have found instances where some students acquired incorrect beliefs which appear to be attributable to the simulation activities they undertook. They suggest that the salient features of those activities may not have been apparent to those students, and that the focus on non-salient features of the activities is one source of statistical misconceptions. Our view is that there is a need for more clarity on the issue. In particular, we need to be specific about what we mean by 'real data'. In considering this question we have identified three broad categories:

- real-time data: these are 'open-ended' data generated by computer from scratch each time a simulation is run or an example is presented. The drawback is that, since such data are prone to the vicissitudes of random generation, on occasions they may be unconvincing in demonstrating a statistical concept, or may even mislead. The manner in which this aspect is addressed is crucial to the effectiveness of the use of such datasets.
- authentic data: these would be the most common perception of 'real data', comprising 'closed' data sets that are genuine and publicly available, 'warts and all'. Such data may be carefully selected to illustrate particular statistical idea and to be relevant to the learner. A drawback is that, as well as being difficult to find, the data may have too many 'warts' for convincing illustration of the statistical ideas in question.
- realistic data: these are also closed data, that are plausible, concrete, and meaningful to the student, but essentially artificial. The advantage is that appropriate randomness and variability can be 'engineered' by the instructor, so that while the data may appear authentic, they are sufficiently 'clean' to allow ideas to be communicated clearly.

What next? Statistical reasoning presents particular challenges for individuals who are not formally trained, irrespective of their level of education. The nature of probability is such that statistical reality is often counter-intuitive and this can lead to widely held misconceptions and conceptual difficulties that can hinder the teaching and learning of statistics. While in recent years there has been a steady increase in the use of ICT in the teaching of statistics, there is as yet no consensus about the most effective means of using these technologies (Batanero, et. al., 2000). A widely acknowledged problem is that, while students may acquire working knowledge of statistical procedures, such knowledge does not inevitably translate into statistical insight involving conceptual understanding and decision making. Major questions remain regarding the types of computer-based activities that help students construct a deep understanding of statistical concepts, as opposed to procedural competency. Nicholson and Mulhern (2000) have argued there is a need to develop this type of ICT materials further in this area, and to extend the scope of the development of higher-level statistical reasoning skills.

5. Acknowledgements

This work was supported by a grant from the Nuffield Foundation. The simulations are programmed by, and web-based materials developed in collaboration with, Neville Hunt of Coventry University, UK.

REFERENCES

Batanero, C., Garfield, J.B., Ottavani, M.G. and Truran, J. (2000). Research in statistical education: Some priority questions. Statistical Education Research Newsletter, 1, 2-5.

Hodgson, T. and Burke, M. (2000). On simulation and the teaching of statistics. Teaching Statistics, 22 (3), 91-96

Hunt, N and Tyrrell, S. (2000). Learning statistics on the web – DISCUSS. Teaching Statistics, 22 (3), 85-90.

National Council for Educational Technology (1996). Mathematics and IT - A pupil's Entitlement. NCET. Coventry

Nicholson, J.R. and Mulhern, G. (2000). Conceptual challenges facing A-level statistics students: Teacher and examiner perspectives. Papers on Teaching and Learning Statistics – ICME9

RESUME

On se sert de plus en plus de l'ordinateur dans l'enseignement et dans l'apprentissage de la statistique. Cet exposé examine le développement de certaines activités assistées par ordinateur, afin de faciliter la compréhension des notions fondamentales de la statistique au niveau supérieur de l'enseignement du second degré. L'exposé présente l'analyse raisonnée de la méthode employée, indique l'étendue de l'utilisation des activitès, et considère les démarches à prendre à l'avenir.