As part of a Maths, Stats and OR Network mini-project at the University of Southampton, a statistics e-Learning website on the design and analysis of experiments is under development. The website has a particular focus on the application of the methods in chemistry research, although it is suitable for use in a variety of applied statistics courses. When complete, the website will incorporate formative student assessment and will be suitable for both pedagogy and independent learning. Key to the effective development and evaluation of the module is its adoption into courses at the University of Southampton in the Schools of Mathematics and Chemistry.

Background

Increasingly, statistics pedagogy is making use of real data and specialist software to bring the learning experience closer to the practice of statistics [1, 2]. Compared to the traditional approach of a worksheet of problems to be solved by applying a software package, the use of web-based modules allows a more structured development of realistic scenarios, which interest students and demonstrate the relevance and importance of the statistical methods. The inter-twining of concepts and practice, with student interaction, can produce an effective and coherent learning framework with students gaining understanding through “hands-on” examples. To date, most web-based statistics resources focus on introductory and basic statistics, as reviewed by Larreamendy-Joerns et al [3].

As part of the Combechem e-Science project (GR/R67729), funded by EPSRC, development was initiated on a statistics e-Learning website covering topics in the design and analysis of experiments. The initial aim of the website was to provide exemplars of best practice to academic chemistry researchers in modern statistical methods appropriate to their research. As such, the website is motivated by real applications and involved the use of authentic chemistry data sets. The website provides interactive design and analysis tools using the R statistical software [4] running on a webserver. Therefore, no specialist software is needed on the student’s machine, allowing the website to be used interactively from any computer. Further background and technical details of the statistics topics and web implementation are given by Woods et al [5], who also describe the available interactive examples. The latest version of the website is available at http://www.doe.soton.ac.uk/elearning.

As the website was originally targeted at academic chemists, its main focus is the application of the methods rather than an exposition of the underlying statistical theory. However, the generic nature of the topics means that the website is potentially well suited to provide self-study elements in a variety of undergraduate
and postgraduate applied statistics courses, allowing the students to interactively experience concepts and methods encountered in lectures.

Aims of the Project

The specific aims of the Maths, Stats and OR Network mini-project are:

1. To improve the presentation and interactivity of the website: during the development of the website, we have tried to follow the guidelines for online material set out by Oliver [6]. In particular, the interactive examples allow for both stimulation and planned participation of learners. Through collaboration with a computer scientist, the web-design incorporates best practice in design and implementation providing good structure and navigation and optimised graphics. The statistical analysis and simulation tools allow not only for student interaction but also for learning via repeating, with students able to reflect on results whilst trying different approaches [7].

2. To introduce formative student feedback and assessment: at the outset of the MSOR project, the feedback offered by the website was limited in scope. The student can see the results of their decisions through the incorporated R functionality, but are rarely given feedback on the appropriateness of any choices that they have made. Further, there is no attempt to assess the student’s understanding of the concepts discussed by the modules. To improve this aspect, currently two forms of formative assessment are being incorporated:

   (a) Assessment of concepts and understanding via a database of multiple choice questions, with topic-specific questions for each section of the modules.

   (b) Assessment of the application of statistical methods to real and simulated data. Specially developed scenarios, using both real data sets and simulated data, allow the appropriateness of a student’s decisions to be assessed relative to previous expert analysis or statistically valid model evaluation criteria (such as $R^2$, PRESS or outlier identification measures).

This work is ongoing. Currently, a database of multiple choice questions is being created (which can be supplemented in the future) and a number of assessment and feedback “scenarios” are under test.

3. To evaluate the modules within statistics courses: the modules are being incorporated into undergraduate and postgraduate courses at various points in their development. Feedback from students and instructors in being collated and used to improve the website.

The Website

The website has the following six sections, most of which include interactive examples that use R code embedded into the webpages.

- Section 1: Introduction to designed experiments;
- Section 2: Classical designs for multi-factor experiments;
- Section 3: Optimal designs for multi-factor experiments;
- Section 4: Running an experiment in practice;
- Section 5: Fitting a model to experimental data;
- Section 6: Evaluating and choosing a model;

Only a fairly limited background in statistics is assumed, with perhaps some experience of the Normal distribution and simple linear regression (although these topics are briefly revised). For fuller details of the statistical content, see [5].

The use of embedded R code allows interactive content to be provided, following the principle that students learn best from dynamic activities, rather than didactic exposition [8]. Use is made of simulations to allow learners to generate experimental data, fit regression models and visualise their results, all in real time. This allows the user to “try out” the methods discussed on real chemistry examples and provides demonstration of the potential application of the statistical methods.

To embed output from the R system into the webpages requires a fast and secure interface between the web-browser and R. This is achieved using the JavaServer Page (JSP) system, which allows Java code to be embedded within the HTML of a website. JSP allows the development of custom tags, which resemble HTML or XML tags. When the webpage is retrieved by a browser, the custom tag is replaced with the output of the Java code associated with the tag. This code is invoked on the server and so no Java software is required on the client machine.

A library of custom JSP tags has been developed to allow the embedding of R code within the website. Through tags, input boxes can be added to pages to allow values to be submitted to scripts, and the output from the scripts presented to the user, see Fig 1. All R computations are done on the server, giving the advantage that the R software is not required on the client machine. For security reasons, users can only run the embedded R code and not upload their own.

Evaluation of the Website

A fundamental part of the project is exposing the website to students during the development process. In the current academic year, the website has been trialled in three different courses at Southampton:

1. A chemometrics course to advanced undergraduate chemists in the School of Chemistry: this was part of the initial target audience and the website was used in conjunction with several lectures and computer labs to provide exemplars and additional self-study material. One aim of the website was to demonstrate the usefulness of the statistical methods to the students’ field of study.
2. A Masters-level course on the design of experiments in the School of Mathematics: as students on this course have a wide variety of previous experience in the subject, the website provided a useful source of background and revision material. By using the website to introduce the topic of computer-generated and optimal design, lecture time was freed up for class discussion and review. The website also provided an efficient means of distributing data for a coursework problem on response surface designs.

3. A third-year BSc course on the design and analysis of experiments in the School of Mathematics: the website was introduced to the students late in the course as a reinforcement and revision tool. Some students appreciated the more intuitive approach taken to some difficult concepts, such as aliasing.

Fig 1  Output from an interactive R example to demonstrate the fitting of a first-order regression model to data from an experiment using two replicates of a $2^2$ factorial design, with data simulated from a second-order polynomial model.
1. Content, 2. Interactivity, 3. Presentation, 4. Delivery and support, and 5. User satisfaction. Given the limited nature of the initial deployment of the website, the questionnaire focused on the first three of these categories. Initial instructor feedback also focused on these topics. The result from student and instructor feedback from each of the three courses is currently being collated and will be used to inform further work. Differences in responses from the three student groups will also be considered, in order to identify changes to the website which would make it more relevant to different cohorts and courses.

**Future Work**

Current progress on the website will be presented at the Compstat 2006 meeting in August. Development is continuing, guided by the experiences and feedback of users of the website. Incorporation of additional assessment exercises and formative feedback will enhance the interactive experience and allow the website to evolve into more than a purely didactic resource. By the completion of the MSOR project next year, these additional features should have been trialled with the primary target audience of chemistry students.

Development is also underway of a companion website on modern regression methods, again with application to chemistry research. This website uses a different web-to-R interface and will allow a comparison of approaches. Initial use of the module with chemometrics students, employing a mix of examples and interactive exercises similar to that on the design of experiments website, has proved promising.

**Acknowledgments**

The development team at Southampton is Dave Woods, Dan Grove and Sue Lewis (Southampton Statistical Sciences Research Institute), Ilaria Liccardi (School of Electronics and Computer Science) and Jeremy Frey (School of Chemistry).

**References**


