LETTER FROM THE PRESIDENT

National, European, international… and face-to-face

By Shirley Coleman

As we approach the fifth ENBIS International conference on Business and Industrial Statistics in Practice, it is interesting to review the importance of face-to-face meetings. I recently returned from the International Symposium on Business and Industrial Statistics, ISBIS4, organised by the International Statistics Institute (ISI) in Cairns, Australia, where I was invited, as president of ENBIS, to give a presentation.

The meeting, like ENBIS conferences, was very motivational because it brought together statisticians interested in business and industry from all over the world; it also it gave me a fresh perspective on the subject, its applications and depth. There were 150 attendees from 33 countries; a similar number to those attending ENBIS conferences.

The meeting launched the International Society for Business and Industrial Statistics (ISBIS) as a free-standing entity, independent of the ISI. In 2003, the third ENBIS conference was held jointly with the business and industry section of ISI in Barcelona. The timing for Europeans was rather difficult; in some countries, staff are allowed to take holidays from work only in August. Now that ISBIS is an independent entity and can meet at any time, we can consider holding joint meetings again.

So now we have ENBIS and ISBIS, and most people are also active in their own country's societies; personally, I chair the quality improvement section of the Royal Statistical Society in the UK.

What is unique to ENBIS?

Members of ENBIS are mostly Europeans. We are joined together in a shared economy and are geographically close. Many of our businesses have branches in other European countries; many ENBIS members travel throughout Europe. One of our main reasons for being interested in each other is that we can work together on joint projects and apply for EU funding. It was the EU's fifth framework pro-ENBIS project, set up by past president Dave Stewardson, including 45 joint workshops and 29 industrial visits that really helped to get ENBIS members working together. ENBIS must focus on European issues and be a real resource to European business and industrial statisticians. For this, we need to share ideas about promising innovative projects and funding for collaborative work. The ENBIS website – www.enbis.org – is the ideal place to do this. Our website is vital and it is important that, following its excellent first five years under Jeroen de Mast, it should continue to evolve and flourish.

ENBIS
European Network for Business and Industrial Statistics
www.enbis.org

We have had an excellent response to the call for papers for the fifth ENBIS conference, to be held in Newcastle from 14 to 16 September. There will be a full programme with presentations on: robust designs; Bayesian models; reliability; measurement; capability; Six Sigma; risk; finance; surveys; knowledge trading; optimisation; and process control, with speakers from many companies and institutions.

This year there are also nine workshops before, during and after the conference. In time order, these are:

- Data mining;
- Simulation of clinical trials;
- Operational risk management;
- Statistics in innovation and the design process;
- Research methods in practice – opening up a toolbox;
- Reliability studies;
- Survey of advanced methods; and
- A demonstration of designed experiments and statistical consulting skills.

Interest groups have been especially active in putting together very exciting workshops, including multiple speakers. The diverse and enthusiastic contributions to the conference show that face-to-face meetings between members are well-liked – and that there is great energy in ENBIS, which promises a vibrant future.

ARE YOU A MEMBER OF ENBIS?

Vision

The vision of the European Network for Business and Industrial Statistics is:

- to promote the widespread use of sound, science-driven, applied statistical methods in European business and industry;
- to attract statistical practitioners from business and industry into membership;
- to emphasise multidisciplinary problem-solving involving statistics;
- to facilitate the rapid transfer of statistical methods and related technologies to and from business and industry;
- to link academic teaching and research in statistics with industrial and business practice;
- to facilitate and sponsor continuing professional development;
- to keep its membership up to date in the field of statistics;
- to seek collaborative agreements with related organisations.

ENBIS has:

- A general assembly
- A council
- An executive committee
- A permanent office in Amsterdam, the Netherlands

Members

- **No membership fee**
- Fast growing number of members
- To apply for membership use registration form at the website: www.enbis.org

Corporate Members

- Costs: €500 per year;
- Exclusive membership;
- To apply for corporate membership contact the webmaster at: enbiswebmaster@ibisuva.nl

Interest Groups

- Design of Experiments (DoE);
- Reliability & Safety;
- Data mining/warehousing;
- General statistical modelling;
- Process modelling and control;
- Quality Improvement;
- Statistical Consultancy;
- Measurement Uncertainty.

Local Networks

- bENBIS (Belgium);
- dENBIS (Denmark); and
- nENBIS (the Netherlands).

ENBIS www.enbis.org

A cost-cutting newspaper delivery strategy

Magne Aldrin and Sonia Mazzi explain how statistical modelling can help businesses to reduce costs, with a newspaper as an example

Dagbladet is a popular daily Norwegian newspaper, sold nationwide in tabloid format. In a country with a population of about four million, around a quarter of a million copies of Dagbladet are sold every day, at 11,000 outlets throughout Norway. These include newsstands, supermarkets, cafeterias, service stations, and kiosks, with widely varying opening times. Delivery of too many copies to an outlet increases costs and wastes resources. But an outlet loses sales revenue if it sells all of its copies too early in the day. So the newspaper needs good sales forecasts to plan its production and deliveries.

Accurate sales forecasts are crucial in every business. Inaccurate forecasts lead to a waste of resources, lost revenue and high costs. Directly related to the problem of sales forecasts is the problem of keeping adequate stocks of consumer goods.

Dagbladet is a general interest newspaper with some coverage of politics, social and cultural issues, and sport. People seem to read it more during the weekend, holidays, and festivities. Sales are different in an outlet in the heart of Oslo from an outlet north of the polar circle, or at a skiing resort. Forecasts have to account for these differences and for other components of variation.

The first task when trying to produce accurate forecasts is to identify as many sources of variation as possible and to incorporate them effectively in a model to minimise the amount of unexplained variation. Depending on events locally, nationally or internationally, readership may vary too. Hence, the day of the week, the time of the year or season, the occurrence of holidays and festivities, temporal trends reflecting varying interests – and different intensity of these effects according to the geographical location of the different sales outlets – are all sources of variation that must be taken into account and modelled efficiently to produce accurate sales forecasts.

Figures 1 and 2 show the observed sales of *Dagbladet* for two service stations in the region of Østfold, during 2004. During the summer time, service station B shows an increased number of buyers of *Dagbladet*, unlike service station A. Also, the day of the week with highest sales levels for service station A is Saturday, whereas for service station B it is Sunday.

The Norwegian Computing Centre (NCC) has developed a system for optimal delivery to sales outlets of *Dagbladet*. The objective was to determine the number of copies to be delivered to each outlet to maximise sales revenue on a given day. NCC used S-plus to create a statistical model named SOL, System for Optimal Leveranse, which means 'system for optimal delivery'. Historical data was needed to fit the model. The amount of historical data varies, from outlet to outlet, from a few years to a few days. Also, at the time a prediction is made, data from the previous week or two is missing, for all outlets, because of reporting delays.

> **'Accurate sales forecasts are crucial in every business. Inaccurate forecasts lead to a waste or resources, lost revenue and high costs'**

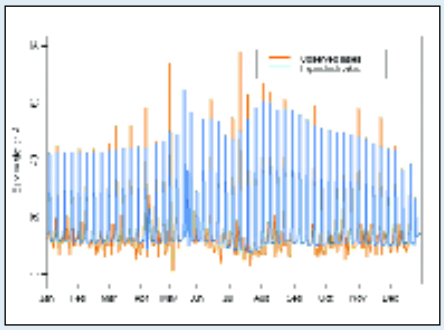

Figure 1. Observed and expected sales for service station A.

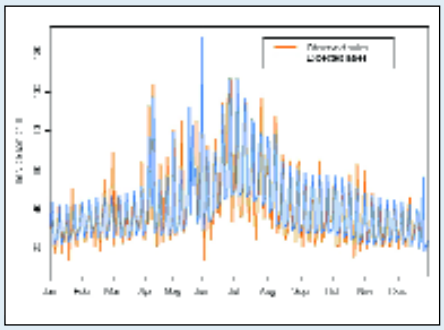

Figure 2. Observed and expected sales for service station B.

Furthermore, it is not possible to assess potential sales at some outlets during certain days because opening times vary. Some service stations, for example, open 24 hours a day, all year round, but other outlets including most cafeterias, are open from Monday to Friday. Most supermarkets are closed on Sundays. With all this complexity and varying amounts of information, is it possible to devise an effective delivery system? The answer is yes, and the key to success is good statistical modelling.

For a certain day and a certain sales outlet, which are fixed from now on, let: y_p be the potential sales, the number of copies that could have been sold if enough copies had been delivered; d be the actual number of copies delivered; I be the marginal income per sold copy; and C be the marginal cost per delivered copy.

The marginal cost, C, and marginal income, I, are determined by *Dagbladet* and are known. At the time of delivery of the newspaper to the outlets, y_p is unknown. Suppose we knew the probability distribution of y_p.

Then, $P(y_p \geq d)$ is the probability of selling the dth copy when at least $d-1$ copies have been delivered.

The expected profit is, thus, $P(y_p \geq d)\, I - C$ and the dth copy should be delivered if $P(y_p \geq d) > C/I$.

The total expected profit is maximised if d copies are delivered so that $P(y_p \geq d) > C/I$ and $P(y_p \geq d+1) < C/I$.

The problem, then, is to estimate the probability distribution of potential sales $P(y_p \geq d)$ i = 0, 1, 2, . . ., using the historical data available. The optimal number of copies to be delivered can be found as a percentile of this distribution.

In reality, the total number of copies printed is determined by the management, depending mainly on the values of d suggested by the model, and adjusted according to the prevailing market situation and the front-page content. The number of copies delivered to each outlet is scaled uniformly, so that the total number of copies delivered equals the total number of copies printed. Potential sales are not observed, though. We observe only the number of copies sold and have information on whether an outlet sold out or not. Also, 82 co-variates were included, such as day of the week, season and location, to explain predictable sources of sales variation. This intricate model is estimated using the SOL software, based on S-plus. Figures 1 and 2 (see previous page) show the observed and estimated sales for two outlets during 2004.

After the implementation of SOL, Dagbladet has certified that, while maintaining sales at historical levels, about two per cent fewer copies are printed daily. There is no increase in the sold-out rate and the number of copies available to customers has been maintained. Variable costs have been greatly reduced because about 25 per cent of the cost of a newspaper is attributable to the cost of paper, ink, and printing. *Dagbladet* adopted SOL as a management strategy tool in 1999 and has been using it since.

Norsk Regnesentral (Norwegian Computing Center, NR), in Oslo, Norway, is a private, independent, non-profit foundation established in 1952. NR does contract research and development projects in information and communication technology and applied statistical modelling. Clients are industrial, commercial and public service organisations, national and international. Scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. Visit the website at http://www.nr.no .

TONY GREENFIELD IS WILLIAM G HUNTER AWARD WINNER

Tony Greenfield, a founding member and past president of the European Network for Business and Industrial Statistics (ENBIS), has been presented with the William G Hunter Award of the Statistics Division of the American Society for Quality (ASQ). He is the first Briton to win the award since its inception in 1987; the presentation was made at the ASQ's recent conference in Roanoke, Virginia.

The award is presented annually to promote, encourage, and acknowledge outstanding contributions to the creative development and application of statistical techniques to problem-solving in the quality field. Named in memory of the statistics division's founding chairman, the award recognises that person whose actions most closely mirror Bill Hunter's attributes as a consultant, educator of practitioners, communicator, and integrator of statistical thinking into other disciplines.

'I am amazed and greatly honoured by this award,' Professor Greenfield said. 'Amazed that people in a large and leading professional society in America should know about me and recognise my work. Greatly honoured simply because that society has tens of thousands of members who have contributed greatly to successes of so many small, large, and global companies, and so many of those members are influential in business and industry. But I feel especially honoured because the award was established in memory of their greatest thinker and leader, Bill Hunter. I never met Bill Hunter but, during the several years before his too-early death, he telephoned me several times to discuss my ideas and my work.'

Those who have won this award since its inception in 1987 include Soren Bisgaard, Ronald D Snee, Brian L Joiner and Roger W Hoerl.

Like Bill Hunter, Tony Greenfield has been active in promoting the use of statistical methods across industry and business. His work for ENBIS has helped bring together those people who are improving the effectiveness of many businesses in Europe and around the world. He is editor and co-author of *Research Methods: Guidance for post-graduates* (first published by Edward Arnold in June 1996, second edition in June 2002), a text used in some British universities in courses for postgraduates who intend to proceed to research degrees.

Professor Greenfield was formerly head of process computing and statistics at the British Iron and Steel Research Association, Sheffield, and professor of medical computing and statistics at Queen's University, Belfast. He is a visiting professor to the Industrial Statistics Research Unit (ISRU), the University of Newcastle-upon-Tyne. His publications include *The Pocket Statistician* (co-author), also published by Edward Arnold in 1996. This book is used by people in business and industry, who are not highly qualified statisticians, as a practical guide to the use of statistics for improving quality in their daily work. He is currently completing, with Andrew Metcalfe of Adelaide University, *Design and analyse your experiment with MINITAB*, which will be published by Arnold early next year.

> **'His work for ENBIS has helped bring together those people who are improving the effectiveness of many businesses in Europe and around the world'**

Statistics unravelled

In the first of a series of articles, John Logsden writes about the importance of statistical concepts and methods in all sciences.

I found statistics when I was a physicist. The subject so fascinated me that I changed my profession. I realised an understanding of statistical methods was essential in all scientific work and I needed that understanding to do my job.

The transition was slow and practical: I often had to learn on my feet to explain statistical modelling to those other scientists I was trying to help. No one really knows something unless they can teach it. This experience helped me to understand the difficulties that physicists, chemists, metallurgists, and other scientists have in understanding and using statistical methods.

The UK power industry, where I worked, was full of highly numerate scientists and mathematicians, but very few had a clue about statistics, and there were hardly any statisticians. I saw results that were clearly wrong. How, for example, could only one out of 144 residuals fall outside a 95 per cent confidence interval? The researcher didn't even reply to the question. In another case, a 'correlation' – a word often wrongly used for 'calibration' – in a study of a material's coefficient of expansion had used a temperature range so small that errors swamped any real trend. Yet the fitted line was used to predict a component's diameter at temperatures well outside the range.

Carelessness, or lack of understanding, may cause such blunders, but problems are often too complex to be solved in the available time. A common view is that statistics is just a matter of pressing buttons; a view once expressed by a chartered accountant, no less! But the questions are: which buttons, which order? Garbage in, garbage out is just as true for statistics as for computing. When a statistical program tells you that something is significant, you need to understand what it means, and how you can be sure that the model is correct, before you accept the result.

In an earlier article I defined statistics as the science of uncertainty: a coherent framework that enables us to characterise measurements and events that we cannot otherwise understand.

Statistics is a methodology that starts with collecting the data, and leads, through analysis, to answers to real problems. The analysis may use complex models of underlying structures to draw inferences. How we collect the data, what models we apply, and how we interpret the conclusions in the field of application, are all responsibilities of the statistician – always in close partnership with the experts in the field.

We have loads of data; it pours out from all sources. As storage capacity gets cheaper, and data logging cards and databases get more sophisticated, there is a temptation to store data without thinking, 'just in case we need it', or even 'so we can say we have it'.

This can be useful. Databases, languages, and other tools have been developed to enable the investigator to extract information from data with minimum effort; data mining has become the province of computer scientists. But how often do they consider the statistical properties of the data? Because we have much more data, we should be able to draw many more inferences. We should be able to consider the distribution form, investigate complex higher-level interactions, and consider the structure of the data. But I have heard of cases where discrete and sometimes ordinal data are analysed as continuous rather than using the correct statistical model; of subsets being used because this was more convenient. As data becomes more and more complex, there is a great danger that the richness leads people to imagine that they do not need a model. They do. Just collecting data for its own sake does not make anyone a statistician, nor do trivial tabulation exercises. The more complex the data, so also is the model and its interpretation; therefore, we need to merge the skills of the statistician, the computer scientist and the applications scientist.

Statistical theory and methods have grown rapidly over the past century; techniques have been developed for many applied areas. Yet we retain the original vocabulary. We still have t-tests, F-tests, chi-squared tests, and techniques such as analysis of (co-) variance, regression, random effects, multilevel, mixture models, principal components, discriminant analysis, and multivariate analysis. The skill of a statistician is to understand these techniques, and more, and when and how to apply them and interpret the answers. A recent problem reviewed for a client showed that the considerable combined intellectual might of the company, plus that of an outside consultancy of mathematical physicists, had failed to recognise a simple generalised linear mixed model. What buttons? What order? Not all applications are so simple.

About half the questions I was asked in industry, often informally and sometimes with a little embarrassment, arose because the questioners had not understood the difference between a standard deviation and a standard error. I sometimes replied that one is a measure of the variation in data and the other is to do with an uncertainty in a derived parameter, but I developed a very simple answer that I pass on.

Suppose, you ask your questioner, you had 10 times the amount of data. Would you expect the standard whatever-it-is to be about the same size or smaller? If they realise that much more data would enable them to get a more precise answer, that is: the number would be smaller, then it is a standard error. On the other hand, if the number is about the same, it is a standard deviation. Enabling the questioners to think practically meant that they could go away and imagine the correct answer without getting tied in knots.

So here is an exercise. For your next set of data, start with a mentally clean sheet of paper, close your eyes and try to imagine how the data were derived. Ignore the actual data you have and imagine what links would be expected between the observations. Imagine what would happen if you recorded the data in a different order or used different measuring instruments. Would you expect things to change? For example, the future cannot affect the past. What would happen if you doubled the number of measurements or halved them? Or did so for only one part of the data? Would you expect a stable answer? And in your subsequent analysis have you added any hidden connections? Invented any additional data? Made any unjustifiable assumptions?

We welcome letters from statisticians, computer scientists and applications scientists. Please email letters@enbis.org . We hope a regular letters column will help us to help you to understand and use statistics and may suggest further articles.

John Logsden runs a statistical consultancy in Manchester, UK. He is an associate editor of ENBIS magazine and on the council as assistant webmaster. Evidence of his interest in furthering industrial statistics may be seen at www.industrial-statistics.com.