# TECHNOLOGY, STATISTICAL THINKING AND ENGINEERING STUDENTS

H. L. MacGillivray
Queensland University of Technology
Australia

*The importance of the statistical sciences in modern engineering is apparent but not necessarily its many roles. Although the second of the US Engineering Criteria 2000, "an ability to design and conduct experiments as well as to analyse and interpret data" has been quoted as stating that statistics has a dotpoint "on its own", (Phillips, 1998) the wording tends to underestimate the diversity and extent of statistics in engineering. The roles of technology in statistics education are also diverse, and must be considered within the student and course context. This paper considers engineering statistics education and where and how statistical technology can facilitate students' conceptual structure, statistical thinking and confidence. There is no magic wand, but technology facilitates and integrates with good teaching and learning strategies based on teachers' statistical understanding and their understanding of the needs of their students.*

## INTRODUCTION

From regression to experimental design, from SPC to MCMC and large datasets, from reliability to queueing, from risk analysis to time series and image analysis, every engineering context/area comes into contact with some aspects of statistical thinking and techniques. Data and variability are inherent in engineering problems, real and theoretical. Because of the diversity of engineering statistical needs, students need an introduction to statistical thinking, concepts and techniques that they can use immediately in real contexts, and a coherent and logical development that optimises understanding at that stage as well as providing a basis for ongoing learning.

Recent comments (see, for example, Davies, 2001; Evans, 2001) have emphasised that engineering is about applying knowledge, creative thinking, design and teamwork. But the applications, creations and designs are of the types that bring together knowledge, skills or components from a range of areas and synthesise them to produce something that works, whether it be a tangible, system or theoretical product. Contexts can range from medical techniques to heavy equipment, from constructions to maintenance or transport schedules, from ecosystems to computer systems, from electronics to power grids, from individual speech to space signals, from chemical processes and biotechnology to technology management of all types. These contexts can also range from the most hands-on of practicalities to the most theoretical.

Thus engineering strives to "sort out messes" and "pin things down" to create products with purpose whether tangible, system, or conceptual. Because engineering models are oriented to end-use and seek to use new knowledge or technologies, their development can be heuristic. It is often in the asking "why", and the linking across models and ideas, that mathematical and statistical scientists are of great assistance in cooperative work with engineering. It is therefore not surprising that engineering students are themselves diverse - in their motivations, inclinations and abilities. This natural diversity has been both increased and ignored in some countries by a variety of extra external pressures (MacGillivray & Moody, 2001).

Engineering students are not yet engineers – an aspect that unfortunately seems to be overlooked more in engineering education than many other disciplines. They are learning about engineering at the same time as learning in the many areas and skills on which engineering calls. It is a serious educational mistake in other disciplines to introduce statistical concepts within contexts in the other discipline unless students are comfortable in those contexts.

This paper considers how introductory statistical data handling/analysis, oriented to standard basic techniques, can and should fit into engineering education, and how the judicious and integrated use of statistical technology can contribute to immediate and longterm student learning. It focuses on technology's roles in development of student knowledge, skills and understanding, and does not consider use of technology for delivery of materials. Technology contributes to and supports many integrated strategies that use teachers' statistical understanding, their understanding of the needs of, and pressures on, their students, and their understanding of how technology should work for the educators, not the educators for the technology.

## LEARNING AND ENGINEERING EDUCATION

Learning involves increasing knowledge - including self-knowledge; increasing capabilities - both specific and generic; and increasing ability for ongoing learning. The relationships and balances amongst these depend on the subject area and the students. To use technical, scientific, mathematical and other knowledge to synthesise, design and create with purpose, it is necessary to have sufficient skills and understanding in both breadth and depth. This is why engineering courses are so valuable to society but difficult to design and update. The balance of breadth and depth - of broad brush and detail – is difficult to achieve by the educators and the students in their learning. The difficulties are not helped by extra pressures from administrations about prerequisites and student numbers, nor from employers about more skills without support. For example, the managements of technology, of systems, and of knowledge are inherent to engineering, but these are in danger of being swamped in the recent emphasis on the inclusion also of management of people and finances in engineering courses.

Subjects within the courses need to be coherent within themselves, but link efficiently with minimum redundancy. Students can find it difficult to provide the linkages themselves, but resent repetition at the expense of time for new work and learning experiences. Sometimes, in efforts by course designers to remove subject separations, the tendency can be to the other extreme - too much broad brush in a problem immersion scenario that can appear to be effective in the short term for small homogeneous groups but does not meet long term needs for the full diversity of engineering and its students.

Some of the most valuable gifts tertiary teachers can give engineering students are structure and coherence to help them assimilate and link skills and understanding across the plethora of engineering knowledge, contexts, models and techniques. These facilitate student ownership of their learning and foster the conditions for developing the creative thinking, design and technical synthesis that characterise engineering achievements. In particular, mathematical and statistical scientists can help both students and staff in this respect, as mathematics and statistics are ideal vehicles for structure, coherence and linkages for engineering students.

## STATISTICS IN ENGINEERING EDUCATION

The challenge of statistics teaching is that the learning involves: combining knowledge, skills and understanding for ongoing learning; judgement in usage; and integration into other contexts. In comments from Garfield (1995),

- “Students learn by constructing new knowledge, using their prior knowledge”
- “Students learn by active involvement in learning activities”
- “Students learn to value what they know will be assessed”,

we see reference to knowledge, skills and ongoing learning, but the last dotpoint needs more analysis, particularly for engineering students. Students will spend time on what they know (or think) will be assessed, but to learn to value it requires confidence in the teacher’s knowledge of the subject and of how it does and will contribute to the students’ current and future learning and work, whether specifically or generically. Such knowledge must be demonstrated within the teaching rather than asserted in motherhood statements.

Statistics can be something of an enigma for engineering students. It is a way of thinking that is both context-free and context-based. It appears to have its own jargon and concepts that seem somewhat familiar but are not. It resists their attempts to reduce it to a simplistic set of black-and-white rules. At the introductory level, it appears simple mathematically but its thinking and concepts are deceptively sophisticated.

Humans in general appreciate knowledge and like to receive well-presented information – if we didn’t, TV programs such as David Attenborough’s or quiz shows would not survive, let alone thrive. Engineering students are highly appreciative of information, provided it is efficiently and effectively given in useable forms. They are critics and consumers of information. They appreciate universality but only if it helps make the information coherent and more useable.

There are a number of stages in an engineering student’s absorption of statistics and statistical thinking that are not necessarily sequential or in the same order for different individuals. These include seeing how statistical thinking and techniques can help in a situation with which the student can identify to a reasonable extent; an increasing sense of familiarity with the structure underlying the techniques and with the concept of a statistical model; realising that allowing variables to be random can in fact simplify some problems; and seeing how statistical thinking and techniques can help in a situation within their own engineering area of interest.

In areas of engineering where design and analysis of experiments, or process control are essential, the last stage above can occur much earlier than in others. But in areas such as communications, computer engineering, avionics, robotics, signal processing, hydrology, traffic, medical, mining, structures and surveying, where the statistical techniques used move quickly beyond those of the introductory level, students need their statistical literacy for its innate value across all workplaces *and* to form their basis for more specialised statistical learning. Unfortunately in engineering courses, it is common that either the statistical literacy is sacrificed or the bridge from statistical literacy to such techniques is not given sufficient attention. Note that referring to contexts that the students will meet is different to embedding new statistical concepts within engineering contexts not yet familiar to the students.

One of the interesting side-effects of the strategy of own-choice group projects of MacGillivray, 1998, MacGillivray and Hayes, 1997, is the range of choices. Although students in all areas of engineering choose contexts ranging from the everyday through to those from their workplaces or their engineering areas, it tends to be more common for students in areas such as electrical, computer, avionics, communications, to either choose non-engineering contexts or to have difficulty with an engineering topic chosen from their area. One of the main motivations and outstanding successes of the own-choice strategy has been to help engineering students own their statistical learning and development at the introductory level. Students' choices of topics illustrate the types of examples in which they want to see how statistical thinking and techniques can help.

Although the statistical educator must keep in mind the differences amongst the engineering areas, it is of great benefit in introductory data analysis for as many different strands of engineering to be in the same class as possible. This helps to provide awareness of the underlying statistical structure and the wide applicability of statistical thinking. Comments about what one strand of engineering can expect to see more of in their area, are of benefit to all the students. A possible exception to this all-in-together strategy is if a subset, for example electrical/computer engineering, have already done some introductory probability/distributions.

## STATISTICAL TECHNOLOGY IN PROBLEM TACKLING, DATA COLLECTION, RECORDING, EXPLORING AND PRESENTING

The obvious contributions of technology here are the reduction of calculation and the visualisation capabilities. Both are clearly important, but technology's contributions can be even more subtle and fundamental. In dealing with real problems from their very beginning, it is essential to identify the observational units (in experiments, the experimental unit), the variables, and their types. In all engineering problems/scenarios, the identification of variables is crucial, but it is surprising how reluctant engineering students generally are to do so explicitly in a data situation. Perhaps the identification of deterministic variables in engineering is a subliminal process. Whatever the reasons, emphasising the importance of identifying variables, observational units and types of variables is extraordinarily powerful in helping them tackle real problems. Where does technology help? Encouraging students to think in terms of their core spreadsheet, with each row an observational unit and each column a variable, has proved of great value. Not only does it focus their attention on the variables and observational units, but it also helps them:

- to design their experiments and studies;
- to recognise the difference between their core data records and derived presentations;
- to recognise when they have more than one experiment/study in the sense that they are dealing with different experimental/observational units;
- to understand the importance of recording data in the order of observation – "in case"; and
- in the case of software such as Minitab, to use numeric and text columns to emphasise that important next step in variable identification – categorical versus continuous or count.

To go even closer to the start of a real problem is to ask questions such as, what are we interested in, what can we measure, can we measure what we want, and are there any other variables we should observe/record at no extra cost in case they are of use. Again, in investigations that start "from scratch", thinking of the spreadsheet can contribute significantly in the initial planning stages.

The above approach has proved highly beneficial in developing statistical literacy for all students. It is remarkable how often the points must be emphasised, but it leads to greater student confidence in choosing and exploring graphical and tabular presentations, and in moving on to inferential procedures for their data. The experience gained in helping students with own-choice projects "from scratch" has also led to development of assessment items and styles that more closely mimic the requirements of tackling real projects in real contexts.

Real contexts and real problems, particularly engineering ones, have more than just a few variables. Although it is important to introduce any concept, procedure or tool, through simple examples that focus on the new idea/technique, the inclusion of extracts from larger examples provides an excellent step in encouraging students to explore many-variabled problems in a systematic way. By making data manipulation and exploration so much easier, technology facilitates this confidence-building approach and improves the student's systematic approach, choice of tools and judgement. These are key elements in problem-solving.

Considering real contexts from the point of problem conception through the setting up of the fundamental data record and into data exploration, graphical and tabular presentation, also helps students understand similarities and differences between the capabilities of spreadsheet software and statistical software.

## STATISTICAL TECHNOLOGY IN INTRODUCING INFERENTIAL PROCEDURES AND CONCEPTS

The traditional approach to introducing statistical inferential procedures for a general student group tends to move from data shapes to distributional shapes, and then, whether implicitly or explicitly, to using sampling distributions of estimators. The first step, from data shapes to distributional shapes, is a sound strategy whether restricted to the normal or not, and is an opportunity for facilitation and visualisation using statistical software. For engineering students, depending on students and course, this can also provide an excellent opportunity for considering other distributional models for data and perhaps introducing them to the concepts of simulation, the sample distribution function, and estimating probabilities "non-parametrically" and "parametrically" using a distributional model.

Indeed, for any student, the introduction of the normal, and possibly the binomial, models should be used to compare data and theoretical pictures, and probabilities (percentiles) estimated directly from data with probabilities (percentiles) obtained from models with estimated parameters. Whatever the discipline or quantitative background of the student, the short and long term benefits of this are significant – and technology makes it easy. It is the traditional hurrying to the second step above, usually focussed on the sample mean and normality, that has the potential for long-lasting student confusion, simply because of the dangers involved in introducing too many concepts too quickly, no matter to what school of statistical thought the educator is inclined.

Another less recognised casualty in a too quick jump to inference based on the sample mean, are categorical variables, which may suddenly (from a student's point of view) surface again in ANOVA. They might also re-appear in an apparent sidetrack into contingency tables and perhaps testing sets of proportions, although unfortunately far too often data in these areas are presented to students in derived tabular form rather than going back to the original datasets. Technology facilitates student grasp of the collecting, recording and exploring of the whole of real datasets with a number of variables. Having established a holistic approach with the help of spreadsheets and/or worksheets, the coherence of this approach should be maintained as we move to inference, whether the educator's preference is to consider categorical or continuous variables first.

My preference, which I started with engineering students but now use successfully with all types, ages, and levels of students, is to start with categorical variables, comparing estimated and theoretical probabilities, and then observed and expected frequencies. It is ideal for introducing the simple concept of "how likely were we to observe … if …" There are no sampling distribution distractions, and the subtleties (and controversies) of hypothesis testing can be left to later. Technology allows considerations of a single variable with two categories (that is, a "binomial" situation) without the distraction of approximate normality. In considering a model (that is, theoretical proportions) for a single variable with more than two categories, the standard test statistic is algorithmic and easy to justify intuitively, including why we can't have the "expecteds" too small. Moving to ask if two categorical variables are related touches on two extra notions (independence and the degrees of freedom) but the test statistic is still student-friendly.

Many real situations that potential statistical users see involve decisions about categorising and combining categories. Because statistical software such as Minitab deals without fuss with either the original data or derived tables, and provides cautionary messages, technology helps this scenario to be an early opportunity for students to consider practical data decisions and the balance between model detail and numbers of observations.

With engineering students I continue on to goodness of fit for discrete variables, and then to continuous variables using Kolmogorov-Smirnov and any appropriate tests provided by the statistical software. Engineering students are the most likely non-mainstream students to see a range of distributions. Again technology permits great freedom to explore a range of models, and always with instant visual comparisons available for illustrating the formal procedure.

Moving to the sample mean, standard errors and interval estimates, statistical educators have been using standard statistical software for more than a decade to bring to life the central limit theorem and the interpretation of interval estimation. No extra educational software is needed for this. What *is* needed are good demonstrations of the variation of the sample variance, and a way of bringing to life for students the difference between interval estimates of parameters, and intervals which (with a level of confidence/belief) cover a required proportion of individual observations.

## BEYOND TWO VARIABLES

Because one-way ANOVA is only two variables – one continuous and one categorical – moving from "two-sample t" to one-way ANOVA maintains integrity of approach and coherence with the exploration of an original dataset and spreadsheet. Randomised blocks are a natural extension to paired samples from the practical point of view. There seems to be increasing consensus amongst statistical educators that it is beneficial for students to learn how to calculate ANOVA for these two situations "by hand" (calculator) and then move to statistical software. Two of the simplest concepts to explain now thanks to technology-aided plots and output, are interaction, and the effect of ignoring information provided by a second factor.

It is of course in regression and beyond that computing technology has opened the statistical doors and windows for so many students at the introductory level – provided their introduction has given them confidence with variables, the concept of a model, and p-values. Illustration of least squares is an optional extra, and the simplicity of Bowman and Robinson (1990) is still fresh. For engineering students, I supply appendices in their course notes, for the mathematics of regression in matrix form, in case they ever need it for reference. Considerable emphasis is needed to help the students become accustomed to consider plotting and residual considerations as integral parts of regression analysis. Many have somehow acquired (graphics calculators? schooling? engineering staff?) a great fondness for $R^2$ as a determinant of "good" and "bad". Perhaps it is desire for a non-thinking button that gives a black or white answer. Other possible areas for engineering are reliability and lifetime data, correlated variables, SPC, and special topics. The ideal situation would be to split the cohort into groups at this stage, not by engineering strand, allowing the students to choose a topic from, say, three offerings.

## REAL PROBLEMS AND ENGINEERING STUDENT PROJECTS

Many comments above either refer to, or have arisen from, teaching and learning experiences oriented to using statistics in real problems from conception to report. Objectives include logical development of the material and coherence with progress through real investigations. The own-choice group project is a semester-long activity that runs in parallel to an orderly carefully-structured course with many small examples and exercises, but with constant reference to how each concept and technique *might* contribute to different types of investigations.

The original intention of the own-choice project was to give students the experience of identifying and setting up an investigation, collecting, exploring and commenting on data. It was not possible to stop the engineering students from trying out analysis and inferential techniques as they acquired them, and so the project grew to become a semester-long activity, with emphasis on choice and judicious use of appropriate techniques. This process, and the feedback into teaching of experience from helping with the projects, have been significantly assisted by technological developments even in just the past six years. The assistance has come not just from software availability and developments, but also in smarter teaching spaces and rapidly increasing electronic communications with students – individually, with groups and the whole class. However the strategy would not have been as successful as it has without grants that enabled student support material to be developed, and indepth student input and feedback on what they thought worked in course materials, course structure, project support materials and assessment.

## PROBABILITY AND DISTRIBUTIONAL MODELLING

Although this paper focuses on data analysis and modelling, the role of stochastic modelling in engineering education, and technological support of student learning in this area, are also important topics. In many areas of engineering, students are not given sufficient background in probability and distributions for the engineering needs, sometimes forcing students and staff to a hybrid of deterministic/stochastic interpretations. Under time pressure, good simulations of stochastic scenarios, and technological learning experiences that emphasise the assumptions and circumstances of different distributions help to form a sound basis for future learning.

## CONCLUSION

Technology in statistics education has received considerable attention, but more focus is needed on its different roles. It is as much a servant in teaching and learning in statistics as it is in *using* statistics, and its roles in any course must be considered within the student and course context. Some questions are how and to what extent technology can foster the individualism, judgement and the “careful thinking” (Hogg, 1991) of good statistics.

An aspect of introductory statistics that has always caused difficulties is the introduction of a number of concepts together. In pre-technology times this was unavoidable in some parts of statistics. As well as providing calculation relief and tremendous visualisation support, technology enables concepts to be more clearly identified, better developed in a coherent structure, and hence better understood. Used wisely and integrated with the educator’s understanding of statistical science and of students in introductory courses, technology is a servant for the holistic, coherent approach. Without such understanding, it is like the broom in the Fantasia version of “The Sorcerer’s Apprentice” – multiplying everywhere, fantastically busy, and achieving nothing.

## REFERENCES

Bowman, A.W., & Robinson, G.R. (1990). C.I.T. IOP Publishing Ltd.

Davies, N. (2001). A new way to teach statistics to engineers. *MSOR Connections, 1*(2), 7-9.

Evans, D. (2001). What will engineering education of the future look like. *Proceedings of the Australasian Engineering Education Conference*, IEAust.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, *63*(1), 25-34.

Hogg, R.V. (1991). Statistical education: Improvements are badly needed. *The American Statistician*, *45,* 342-343.

MacGillivray, H.L. (1998) Statistically empowering engineering students. *Proc. 3rd Engineering Maths and Applications Conf*, 335-338, IEAust.

MacGillivray, H. L., & Hayes, C. (1997). Practical development of statistical understanding: A project based approach, Queensland University of Technology.

MacGillivray H.L., & Moody, J. (2001). Enabling student ownership of mathematics and statistics in their engineering learning. *Proc. Australasian Engineering Education Conference*, 111-117, IEAust.

Phillips, W.M. (1998). Engineering opportunity in the information age and the influence on engineering education. *Proc. 3rd Engineering Maths and Applications Conf*, 47-48, IEAust.