

Making Statistics Significant in a Short Course for Graduates with Widely-Varying Non-Statistical Backgrounds

H. L. MACGILLIVRAY

h.macgillivray@fsc.qut.edu.au

School of Mathematical Sciences, Queensland University of Technology, Australia

Abstract. With reference to a short compulsory module for MBA students, this paper discusses how, despite a range of pressures and an extraordinary range of student backgrounds, fears and motivations, it is possible to quickly facilitate understanding and confidence in handling data and basic statistics for graduates in business and associated workplaces. Links and contrasts with a variety of undergraduate teaching scenarios help to enrich teaching and learning strategies, and the strategies and materials developed through interaction with the range of MBA students are transferable to other postgraduate coursework situations. In particular, the need for understanding and confidence in handling data and its presentation, in considering categorical and simple continuous data, and in understanding correlation and association, are common across many different workplaces and applications.

Keywords: business, graduates, MBA, statistics, workplaces

1. Introduction: The Background

The specific teaching and learning scenario that is the main impetus for this discussion was a 6-week module in introductory data and statistics given to 14 groups of Masters of Business Administration (MBA) students during 1999-2000 at Queensland University of Technology. It was a compulsory module to be taken early in their MBA course, and was needed for both marketing and finance modules, as well as being of general value in itself. The size of each group was supposed to be capped at 35-40, with multiple groups per 6-week session as necessary, but a few of the group sizes were up to 50, with one of 60. All the students were graduates, with an extraordinary range of backgrounds, from languages, music, law, architecture and psychology, to medicine, science and engineering, as well as all the different business backgrounds. All but a few of the students were part-time, and their workplaces and work experiences ranged from research labs, hospitals, schools, aviation, and private practice, to selling for or managing small companies, to multinationals, government and industry.

When the module was initially introduced to the MBA course, materials from a first year (undergraduate) business faculty data analysis unit had been extracted and used, with a business statistics textbook set as a compulsory text. However, this initial version of the module experienced high failure rates and student complaints, including, apparently paradoxically, that the material was either too hard, or trivial and irrelevant. There is no paradox because graduates in the workplace, no matter what their quantitative backgrounds or lack thereof, have different needs to undergraduate students, again no matter in what course. The converse is also true and unfortunately increasingly not recognized in areas such as engineering - that undergraduate needs are different to graduates'. In particular undergraduates are not yet what they are studying to be; they are learning about their chosen course at the same time as trying to understand selected concepts and techniques from universal areas such as statistics.

After responding to a call for help, and listening to the situation and requirements expressed by staff in marketing and finance, the author developed materials and strategies intended to take students from virtually nothing through to ANOVA and regression, using Excel with occasional reference to other packages as needed. The marketing representative had good understanding of the needs of MBA students and their workplaces, in comparison to the mathematical, statistical and computing wish list from finance. At first teaching the use of Excel was included at the request of finance, but this was quickly changed in response to student feedback, to interpreting output with instructions on using Excel provided but not assessable. The wide variety of backgrounds and workplaces also included an extraordinary range of IT and/or Excel familiarity, but the students' wishes were all the same - they wanted every minute of class to be spent on the statistical concepts and usage. As one manager commented, "I can ask other people to produce that output; I want to know how not to be fooled by it". Those with little IT or Excel background were grateful for instructions and information without pressure added to the statistical learning.

The design, delivery, learning experiences and assessment of such a subject involve a range of complex and sometimes conflicting considerations. This paper discusses just some aspects of 'introductory' statistics for post-graduates highlighted by the above MBA module, including some comparisons with undergraduates.

2. Materials, Texts and Resources

In the previous, initial version of the MBA module, the materials that had been taken from half the first year business subject were detailed, prescriptive, and fiddly. Despite reference to Excel output there was calculation emphasis, a large textbook, almost no data, and the content did not progress beyond single sample normal mean inference. One of the most interesting aspects in the teaching of statistics is that lack of quantitative confidence in students tends to increase their requests for detail and exact answers, which are the opposite of what they need. Responding to this with more and more detail and prescriptions tends to produce cumbersome courses that work against facilitating statistical understanding and thinking.

Certainly in the last decade statisticians have increasingly written about the need for emphasizing statistical understanding, thinking and a data-driven approach in introductory statistics ([3], [2], [7]). The effects of statisticians' work in teaching and learning can be seen in a range of strategies and materials such as STEPS (Statistical Education through Problem Solving, <http://www.stats.gla.ac.uk/steps/>), independent student projects, ([9], [5]), and online/multimedia materials such as Discuss, ([4]), and ActivStats, ([11]). However the penetration of such approaches is slow, particularly in other disciplines and particularly where the development of cumbersome or prescriptive detail was in response to apparent rather than real student needs.

Although there are written support materials for strategies such as the above, (<http://www.stats.gla.ac.uk/steps/>, [10], [6], [11], [4]), the influences on texts is still marginal, again particularly in user areas such as business and social sciences. Soweby [8], includes discussion on the sameness and inappropriateness of many introductory texts. The situation is not helped by graphics calculators and spreadsheet software which tend to produce a plethora of single-valued 'statistics' but whose graphics are not oriented to user-friendly and statistically-oriented exploratory data analysis. Introductory statistics for business texts tend to be close to the winners in size, minutiae and rules. Such texts can sometimes be useful as references as it is important for students to know where to go if they want more information. It is also very valuable for students, particularly postgraduates, to learn to be critical in reading about statistics. Thus, provided the core notes/resources from the instructor/lecturer are coherent and well-structured, the use of a number of references rather than the prescription of single compulsory text, can greatly assist the learning process.

Non-quantitative postgraduates can seem to be, paradoxically, both more insistent on, and impatient of, detail and prescriptive approaches than undergraduates. This comes from their simultaneous desires for as much understanding and as much instantly-useable knowledge as possible under time pressure. Their needs are best served by well-structured, full but succinct core material that tells a coherent logically-linked statistical story driven by data and real problems without complex contexts, and that emphasize key concepts, with reference books for further detail if desired. In efforts to help non-algebraic students, a trap can be to become over-anxious with formulae. The critical need for algebraic thinking for non-quantitative students is in modelling, and this is particularly evident at the postgraduate level. The emphasis should be on the power of algebra to capture and represent concepts and key points. Trivial, repetitive and non-essential formulae and definitions should be minimized, and the emphasis placed on how mathematical language and symbols help sort out the ‘messes’ and focus on ‘essentials’.

3. Delivery and Resources

Feedback from all levels and types of students in statistics subjects consistently demonstrates the aspects they value most include organization, interaction with staff who are completely at home with the material and ‘connect’ with students, well-structured and coherent courses with clear elucidation of concepts, and good supporting learning materials. Debates about the balance between web-based and hard copy materials and other aspects of the web, can tend to obscure considerations of the multiple roles of human interaction so vital to learning. Human interaction includes whole class, class subsets (for example, tutorials), student group and individual student/teacher. All are important. In the rush to decry ‘the lecture’ in learning, it has often been forgotten that what is being criticized is overdependence on the role of ‘lectures’ in the whole learning package. Students regard ‘lectures’ as essential components of an overall learning structure, and state that their needs of ‘lectures’ are to give structure, guided tours, emphasis on important signposts, vividness, class identity, and class interaction. This is particularly so in an area like statistics where concepts and structure are essential; student reaction to a lecture-less strategy in statistics was “It’s the no-teach, no-learn method”.

There are some major differences between undergraduates and postgraduates here. Considerable thought and effort can be required in facilitating engagement by undergraduates, particularly at the first and second year levels. Postgraduates tend to engage much more readily, but this seems

not the only reason they value the ‘lecture’ elucidation and interaction very highly. While undergraduates benefit considerably from closely-organized learning experience sessions and activities, postgraduates prefer the freedom to do their learning experiences in their own time, even though they may choose sometimes to work together or to attend extra help sessions. Despite the extraordinary diversity of backgrounds and workplaces across 14 different MBA groups, it was almost unanimous that they wanted every minute of their class time given to the teacher teaching, provided the teacher had expertise in the material and communication of it, and encouraged interaction. They are also more intolerant than undergraduates of ‘time-hogging’ by individual students. Balancing across the variety of students in a postgraduate class requires judgement and statistical breadth, and, most importantly, finding out the workplaces and backgrounds within each group to enable comment when appropriate within discussion to individuals’ ‘reference spaces’. Such comments are of interest to the whole class, simultaneously extending experiences, strengthening class interaction and the vividness of new concepts and techniques.

Postgraduates want the learning experience materials, activities and assessment support, to have nothing that is not absolutely necessary, with succinct instructions, and easily-accessible help. For all students, formal group work should not be compulsory except on tasks for which a group is needed, but for postgraduates any group activity requires great care and alternatives should be considered. All students value assistance via email, but the value for postgraduates is immense. The web is ideal for backups, and items to interact with or to return to. For all students, the value of statistical educational software is a function of the combination of statistical and teaching expertise of the strategy designer, but for postgraduates it is of no value unless it is also highly purposeful and time-efficient.

4. Content, Structure and Data Types

The single most important aspect of content and its structure in teaching statistics, no matter what audience or level, is that it be purposeful and coherent. There are a number of useful questions to consider in designing modules and courses. For example: ‘For what situations are we preparing the students?’, ‘What do we want them to be able to do?’ and ‘What basis do we want them to have for future development - in understanding and techniques?’. The importance of purposeful and coherent structures is why electronic educational resources that are strongly modular for user selection are useful, and why printed resources need to be coherent and steadily developmental in structure, with clearly identified purpose. The increas-

ing choice provided by modular electronic educational resources further underlines the need for greater variety in statistics textbooks.

If the purpose of the learning unit, module or subject (hereafter referred to as unit) is to facilitate handling, exploring, interpreting and basic analyzes of data, then the first considerations are what types of data and what types of questions are relevant for the data. Coherence can be achieved in a number of ways. If the postgraduate group is in one industry or workplace, a single relevant case study can provide the coherence (see for example, [1]). But if the group comes from mixed industries or workplaces, coherence can be achieved by using a number and variety of smaller general interest examples and case studies and building on the themes of types of data and associated procedures from handling, exploring and analyzing. For both undergraduates and postgraduates, it should be remembered that it is a serious educational mistake to introduce statistical concepts within contexts unless students are comfortable in those contexts. We can use what they are going to see in other contexts for problem motivation, but context motivation is different to immersion in which transferability and linking is harder to achieve. All students prefer a range of small examples to illustrate new concepts and techniques, and it is of great value in illustrating the tackling of real situations to use extracts from data sets with at least five variables. Postgraduates tend to have had more contact with data and situations needing statistics than undergraduates, so in a mixed class, flexibility and finding out about the participants' workplaces and backgrounds, as described in 3 above, can be very useful.

If the starting point is handling data, a theme of identifying variables, observational subjects or units, and types of variables, can take the structure from the spreadsheet of the raw data through to choice of exploration, graphical and analysis tools. For example, types of data summaries, graphs and plots can be classified as appropriate for categorical, count or continuous data. Most postgraduate business students have surprisingly little knowledge or understanding of many types of plots, particularly scatter-plots. The concepts of variables, distributions and parameters are essential, but notions of probability need be called on only as needed.

An aspect of introductory statistics that has always caused difficulties has been the introduction of a number of concepts simultaneously; this may have been necessary before ready access to computing, but current technology plus re-thinking can help distinguish 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 using statistical software for visualization. It is the jump 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 introducing too many concepts too quickly.

Technology facilitates student grasp of the collecting, recording and exploring of the whole of real data sets 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. A very effective strategy for postgraduate business students (and indeed for almost all students) is to move from EDA (exploratory data analysis) to 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 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 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 involve decisions about categorizing and combining categories, so this is also an early opportunity for students to consider practical data decisions and the balance between model detail and numbers of observations.

The concepts of sampling distributions, standard errors and interval estimates are very difficult for non-quantitative postgraduates, even with statistical software to illustrate. But having tackled this, moving to one-way ANOVA is a simple step from 'two-sample t ' by emphasizing that both involve one continuous and one categorical variable. Such emphasis also maintains integrity of approach and coherence throughout the unit, with overall emphasis on variables and data in real, often large-scale, scenarios. Randomized blocks are a natural extension to paired samples from the practical point of view, and, thanks to technology-aided plots and output, interaction is trivial to explain. For postgraduates, interaction is a far simpler concept to understand than the effects 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.

For non-quantitative postgraduates, it is essential to check first that they understand what a straight line relationship is, including the effects of non-zero slope and sign of slope.

5. Assessment

Statistics lends itself to a range of types of assessment. Care is needed in choosing from an increasing range of forms of assessment to avoid over-assessment or over-expectations in marking. Managing student workload helps also to manage staff workload. Assessment during semester without quality feedback causes problems for students and staff. Assessment during semester is best as combination formative/summative, either through contributing a small proportion or by using 'best of' approaches. Students' workloads should always be considered and feedback can be a combination of universal and individual. Provision of sample solutions is a must, particularly for postgraduates, but judgement is needed in the balance between the timing and amount of provision of information in solutions and the students' learning by trying. Provision of solutions does not imply limiting to questions with exact answers, as indications of types of solutions or what solutions should include are satisfactory.

Commenting on features of data and computer output involves judgement and communication, and the marking of students' comments is demanding work, particularly for less experienced markers. Non-quantitative students tend to have a general inability to interpret data matched by an inability to communicate effectively and to learn from feedback in quantitative situations. It became clear that (a) non-quantitative students tend to have difficulty in taking their cue from examples and (b) it is very difficult to give feedback on what is absent or lost in waffle. A successful innovation has included statements about data based on plots and computer output, with students asked to choose appropriate statements from given lists. Feedback has been that these exercises/questions are a great help in learning, often identifying 'holes' that students did not realize they had. Assignments and exams used the strategy of single scenarios with data on five or more variables, and a range of questions to be answered using computer output provided, with questions requiring choice of relevant section of output as well as choice of appropriate procedure and/or comments.

6. Conclusions

The success of the six week unit for MBA students that provided the motivation for the above comments, was evidenced by formal and informal student reaction, participation, feedback, results, and student confidence and pride in their learning and the transfer of this learning to their workplaces. Considerable difficulties arose from attitudes amongst business staff towards texts and criteria-based assessment. However the materials and strategies developed have influenced teaching in other courses, both undergraduate and postgraduate, and are currently being adapted for both online and face-to-face teaching and learning in other postgraduate courses. The combination of observing students across a range of levels and courses, and identifying similarities and contrasts in their learning needs and approaches, is invaluable in informing transferable and individual learning and teaching developments in statistics.

References

1. N. Davies. A new way to teach statistics to engineers. *MSOR Connections*, 1(2):7-9, 2001.
2. J. Garfield. How Students Learn Statistics. *International Statistical Review*, 63(1):25-34, 1995.
3. R.V. Hogg. Statistical Education: Improvements are Badly Needed. *The American Statistician*, 45:342-343, 1991.
4. N. Hunt and S. Tyrell. DISCUSS, <http://www.coventry.ac.uk/discuss/>, 1997.
5. H.L. MacGillivray. Developing and synthesizing statistical skills for real situations through student projects. *Proceedings of the 5th International Conference on Teaching Statistics*. International Statistical Institute, pp. 1149-1155, 1998.
6. H.L. MacGillivray and C. Hayes. Practical Development of Statistical Understanding: a project based approach. Queensland University of Technology, Brisbane, 1997.
7. E.R. Sowey. Statistical Vistas: Perspectives on Purpose and Structure. *Journal of Statistics Education* [Online], 6(2): www.amstat.org/publications/jse/v6n2/sowey.html, 1998a.
8. E.R. Sowey. Statistics teaching and the textbook - an uneasy alliance. *Proceedings of the 5th International Conference on Teaching Statistics*. International Statistical Institute, pp. 151-165, 1998b.
9. S.B. Vardeman. Independent student projects in undergraduate engineering statistics and quality control courses. *Communications in Statistics—Theory and Methods*. 25:2633-2646, 1996.
10. S.B. Vardeman and J.M. Jobe. *Basic Engineering Data Collection and Analysis*. Duxbury, Pacific Grove, CA, 2001.
11. P. Velleman. *ActivStats*, <http://www.awlonline.com/ActivStats/>, Addison-Wesley, 2002.