Data Interpretation in the 21st Century: Issues in the Classroom

James Nicholson & Gerry Mulhern

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

Abstract: In the information rich environment of the 21st Century, the ability to interpret data, both raw and statistically summarised, will be increasingly important, as will the capacity to critically analyse arguments based on such data. In this context, this paper gives a brief overview of the different approaches evident in the new A-level Statistics specifications in the UK and examines considers the implications for conceptual development and topic sequencing at this level. We consider ways of making use of technology to allow students to build experience-based models of situations they currently treat only on a theoretical or abstract basis, and of making use of concrete scenarios with which students are familiar to develop critical analysis and interpretative skills.

Introduction:

We are currently involved in a development project funded by the Nuffield Foundation, the aim of which is to produce diagnostic and support materials for teachers and students for the teaching and learning of statistics. Nicholson & Mulhern (in press) reported that in interviews with teachers and examiners, one of the major issues which arose repeatedly was the difficulty that students have with interpretation.

It is widely acknowledged that today's, and tomorrow's, students will require much more in the way of data handling skills than ever before. We believe that an increasing proportion of that information will rely on inference from data or numerical evidence of some sort, and is often put in the public domain by those who wish to persuade the recipient of a particular view. Therefore there will be a premium on the ability to be aware of the validity of such arguments, including those put forward by vested interests.

This paper aims to:

 (i) outline the rationale for the approach taken in our materials, namely

    a) the use of technology to allow students to build experience-based models of situations they currently treat only on a theoretical or abstract basis

    b) the use of concrete scenarios with which students are familiar to develop critical analysis and interpretative skills

Our approach is illustrated with specific examples.

(ii) briefly provide an overview of the (quite radically) different approaches evident from the various examination boards in their new A-level specifications

(iii) briefly examine what hierarchical structures really exist in conceptual development / topic sequencing at this level

(iv) consider the implications for the 21st Century classroom.

Rationale for diagnostic and support materials:

Once statistics goes beyond the descriptive, and the calculation of summary measures, which characterise the early curriculum in many countries, students face a considerable challenge in identifying the role of stochastic behaviour.

In our experience a stylised, procedural approach to the instances in which students first encounter inference masks the non-deterministic nature of inference. For example, once a sample of data has been identified, and the level of confidence specified (and both are normally provided by the textbook or the examiner) there is an identifiably correct, unique, confidence interval. This obscures the underlying dependence on the data whereby confidence intervals are genuinely stochastic quantities. Without a firm grasp of this stochastic nature it is not possible to understand what the estimator does, and perhaps more importantly does not, tell you.

In a similar manner, an hypothesis test has a single correct 'outcome' once the size of the test, and the sample of data, are known - yet we constantly have to remind students that statistical tests never prove anything – the conclusion may be 'wrong' as a Type I or Type II error.

In fact, this can be generalised to any decision-making in the face of uncertainty. The thinking required is really quite subtle – a decision may be 'correct' in the sense that you would always choose the same course of action in the same circumstances, yet it may prove to have been 'wrong' in the sense that another choice would have been made with the benefit of hindsight. Insurance premiums for policies on which no claim is made is one example, and investment strategies are another.

Students also encounter situations which are essentially the same as the mathematics they have encountered up to this point at school, in that these situations are ones where there is a 'correct' answer at the end. There is, however, some interference from aspects of each situation where uncertainty plays a role, so some students fail to appreciate the reality of the deterministic nature of these situations. For example, in calculating expectations, the probability distribution needs to be used - but the uncertainty of outcomes observed from this distribution is not relevant to the calculation of expectations. This is essentially the same difficulty, but in the opposite sense, to that outlined above and the two-way nature of this difficulty poses particular challenges in the teaching and learning process.

There are also situations where there is an element of reasoning or judgement to be exercised, and which therefore do not necessarily have a single 'correct response'. Undoubtedly these situations are harder to provide instruction for than for situations where the response can be broken down procedurally. Moreover, they are much harder to provide appropriate assessment for, since the possibility exists that an 'inaccurate evaluation' (as the examiner sees it) may be due to incorrect reasoning, or could be due to correct reasoning using a different value system - and unless the examiner can see both the conclusion which a student has drawn, and some evidence as to the process by which that conclusion was arrived at, it is not possible to evaluate reasonably the worth of the response.

Gal (1996) identified two distinct aspects of handling data - *generative* skills, where students act upon data (i.e. *doing* statistics), and *interpretative* skills, where students form opinions about the meaning of the data. It is crucially important to develop what Gal (1998) terms a 'culture of explaining' in the classroom, so that the process by which judgements are arrived at can be evaluated, and students helped to develop their reasoning skills, rather than it appear that just the outcome is important. The situations where reasoning and judgement are required are under-represented in current textbooks, and even where they do appear, they tend to be treated superficially - textbooks provide numerical solutions, but very brief, or even no solutions to questions which require judgement - often giving the appearance of the existence of a single correct opinion, and revealing none of the process by which a decision might be arrived at.

We aim to address these key issues in the materials we are producing. In the next section we outline two of the topic areas being covered in our project and how our materials address the difficulties in an innovative fashion.

Sampling Methods:

This is one topic where the approach in textbooks has often been confined to stating the mechanics of each procedure, i.e. defining what the terminology means and stating abstract generalisations of the advantages and disadvantages of the various procedures. Students are expected to learn this 'factual' material and be able to regurgitate it in examination settings.

The reality is that this information is only of use if it can be applied to choosing what sampling

method is appropriate in a specified context, yet there is little or no guidance available as to the process by which such a decision could be made, and even less is available in relation to the quantitative aspects of that process e.g. how the cost of different sample methods is related to the sample size which can be used for a survey at a fixed price.

In our experience, students seem to have little difficulty in appreciating the rationale for the standard properties of a good estimator i.e. unbiased and small variance, and even the notion of consistency is intuitively accessible. Where a population distribution is known it may be possible to calculate mathematically what the mean and variance of different possible estimators are and so identify an order of how good the estimators are. Much more often, however, a population exists without a known distribution, and the information is to be generated by the sampling procedure.

Again, we have found that all too often students are rather blasé about the size of samples involved when working with simulations - a larger sample gives better information, so just choose a large enough sample to give whatever quality of information you would like. The cost of the sampling process, and even the time required to undertake it, do not give the student pause for thought because the simulation does not address these issues.

Our materials will produce different types of populations generated on specified characteristics of concrete scenarios. For example, the audience at a sporting or cultural event is normally made up of groups of different sizes, and within any particular group the individual members are more likely to share the same opinion than in the population as a whole. Random generators are used to construct the size of groupings in the audience, and then the opinions within that group. Different sampling methods can then be explored by taking repeated samples, and comparing the variability of the outcomes. The cost of conducting each type of sampling is displayed and the student uses this in conjunction with the observed outcomes to decide what sampling method (s)he feels is most appropriate in that context. A variation on this automatically determines the sample size for each sampling method which could be employed in a fixed cost survey - so the sample size available for use in a systematic sample would be much larger than for a stratified sample.

It is, however, important that such decisions are robust. For example, in another of the scenarios, the output from a production line is simulated. Random defects appear in every population generated, but in some populations there is also a regularly occurring defect, simulating what happens if one gripper in ten on a rotating wheel is faulty. In this scenario, the systematic sampling method will look equally as good as simple random sampling on occasions where there is no regularly occurring defect but this conclusion will not survive exposure to the case where the regularly occurring defect is present.

Since populations with pre-determined specific characteristics may, by chance, also exhibit other characteristics for which a particular sampling method gives good results, students need to be realise that conclusions about methods should be tested on a number of populations of that type.

Interpretation in regression using real data contexts:
In our view this is one area where the computational aspect should receive less assessment credit in the near future, since even basic calculators will now do two-variable statistics, and it is therefore merely an exercise in data entry. The interpretation becomes correspondingly more important, and as with the examples outlined earlier, there is a dependence on the data which means that the regression line is a genuinely stochastic quantity, whose nature is masked by the deterministic, and normally automated, calculation once the sample of data is known.

There are a number of issues which we feel the interactive Web environment, with simulations incorporated, can help to build stronger conceptual understanding. The interpretation of

the coefficients in linear regression models is an area where textbook discussion has often been extremely limited, and Web-based materials offer the opportunity for greater detail to be requested through hot links, and a larger variety of real life data contexts to be explored. For example, the meaning of residuals and their role in a linear regression model is more sophisticated, and more important, than its status in most textbooks would suggest.

Our materials will allow random samples to be taken from a population of bivariate observations. The amount of variability observed, between repeated samples from the population, in the regression lines obtained depends on the number of data points being sampled and the correlation of the underlying population.

Exploring this in a systematic manner, in a variety of real data contexts will, we believe, build a more accurate conceptual model of both what the regression line is and what the uncertainty is associated with predictions or estimates made using the regression line. For example, residuals are the total effect of all factors other than the explanatory variable used, but it is the context (rather than some abstract or theoretical notion) which determines the relative contributions of the residuals and the explanatory variable.

New A-level specifications from September 2000:
[In the following, the UK refers only to England, Wales and Northern Ireland as Scotland has a different examinations system.]
Boland and Nicholson (1996) considered the changes in curriculum in the USA, Ireland and the UK after the last revision of A-level courses in 1995. Probability and Statistics had been given a much more prominent place in the compulsory school curriculum (up to 16 years old), and the Mathematics of Uncertainty had been included in the Common Core for A-level.
In the 1995 revision, courses became more modular, but Boards offered A-levels based on 2, 4 or even 6 modules. Nonparametric tests were included in a number of courses for the first time, and comparative analysis of the different courses showed a considerable variety in the order of certain topics e.g. conditional probability and correlation and regression.
The A-level specifications being taught in UK schools and colleges from September 2000 appear to show some important differences. All Boards now have to offer a six module course, and only one of these does not use equal weightings. Schools and candidates can generally choose to include 0, 1, 2 or 3 Statistics modules as part of a Mathematics A-level, together with 3 Pure Maths and either Mechanics or Discrete Maths modules to complete the Applied requirement; only one Board includes a compulsory Methods course maintaining what had been compulsory for all candidates by its inclusion in the Common Core.
The traditional order of a first course in probability and statistics has started with a grounding in probability and the collection and summarising of data, followed by discrete and continuous random variables, the parameterised families of distributions (particularly Binomial, Poisson and Normal), and expectation algebra. Then inference is introduced with confidence intervals and hypothesis testing for parameter values. The variations previously have been where conditional probability, correlation and regression, and permutations and combinations were introduced. These variations might be viewed as reflecting different beliefs as to the relative importance of these topics. One consequence of the traditional ordering was that students taking A-level Mathematics with Statistics as well as either Geography or Biology would have to use Students' t-test or correlation and regression, introduced usually as 'black box techniques' with no theoretical underpinning, in their other A-levels and might not encounter them at all in their Mathematics and Statistics course.
Boards now offer from 3 up to 7 statistics modules which can be taken together. Almost all of these have dependency requirements, so that modules S1 and S2 are required to take S3 etc. Two

exceptions are the specification from Mathematics in Education and Industry (MEI) and the Assessment and Qualifications Alliance (AQA) specification B. MEI has S4, S5 and S6 which each require only S1 - S3. MEI also has a module entitled Commercial and Industrial Statistics which only requires S1 and S2 and so can be taken at any stage from then on, including as part of a single A-level. For later Statistics modules all boards except AQA 'B' also require increasing amounts of Pure Mathematics.

AQA (2000) in their specification B aim to offer a full A-level to those who want to do Statistics without Mathematics, as well as those taking the traditional Mathematics and Statistics combination. As a result the dependencies are more complex and less rigid and the later modules do not move into the sort of mathematical statistics, such as pgf's and mgf's and maximum likelihood estimation, which other boards include. Instead, the rubric for their later modules states: "The emphasis is on using and applying statistics. Appropriate interpretation of contexts and the outcomes of statistical procedures will be required." (pp 69 & 71). This implies that these are higher order skills. However, Figure 1 shows a substantial difference in the proportion of marks allocated for interpretation in all of the AQA 'B' modules (in Group 2) when compared with standard modules (Group 1), illustrating that there is considerable scope for rewarding interpretation skills more highly, across all module levels, than is the current practice with most specifications.

The following remarks are based on an informal analysis of the new A-level specifications (a term which replaces 'syllabus' in the new structure), and highlight the aspects which we feel are most important. They should not be viewed as an exhaustive treatment.

There appears to be more in terms of applications of statistics early on in the new courses. Correlation and regression is met in the first module in three specifications, and in the second module in another two specifications. Hypothesis testing and / or estimation are introduced early - in the first statistics module in two specifications and in the second module in all of the others. In four of the seven specifications, inference is introduced in the same module (one-sixth of the A-level course) as the distribution(s) to be considered inferentially. In all cases inference is first met with parameterised distributions, and five of the seven specifications now include non-parametric methods, usually in the third or fourth module.

Hierarchical structures in conceptual development / topic sequencing.

Mathematics is widely recognised as involving multiple, interlinking perspectives. It follows that children's mathematical development is enhanced, at least in part, by revisiting conceptual areas regularly, in a variety of contexts. We believe that statistics has much in common with mathematics, but has important further dimensions to the thinking involved once inference is encountered.

As we have outlined previously, we believe the stochastic nature of inferential statistics requires a firm grasp of some subtle ideas. The pressure to introduce applications of statistics into the courses earlier than previously is understandable as the curriculum broadens, and statistics becomes more widely used in the professions, and in other areas of employment. However, there is a lack of empirical evidence available currently about the progression in statistics courses at this level. We suggest that there is an urgent need for such work to be carried out so that distinctions between useful and necessary precursors to new concepts can be identified, and future course development can be underpinned by such principles.

Implications for the 21st Century Classroom:

Analysis of the mark distribution in specimen papers published by the Boards show that the standard modules (group 1 in Figure 1 below) have between 75% and 97% of the credit allocated to computational work. Only 3% to 25% of the credit was allocated to interpretation, describing, stating test conclusions in the problem context or stating assumptions. Group 2 is comprised of the

papers offered by AQA 'B' and the Commercial and Industrial Statistics module offered by the MEI specification.

Coursework is one mechanism by which broader statistical skills can be assessed. In 4 of the 7 specifications, there is no coursework possible in Statistics. The specification from the
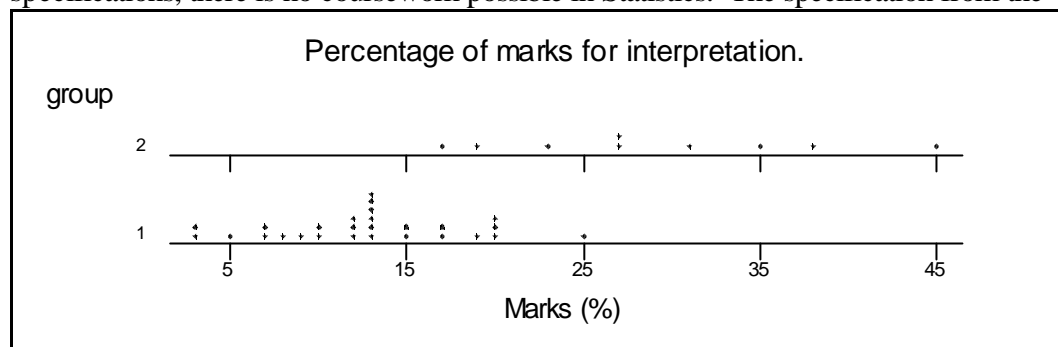

Figure 1.

Oxford, Cambridge and RSA Examinations (OCR) offers an optional project module; Edexcel has projects for 25% of modules S3 and S6 and MEI has projects for 20% of modules S1, S2 and the Commercial and Industrial Statistics modules.

If the computational side of statistics is genuinely to reduce its profile, as the improvements in technology make it less of an imperative, then we believe this must be mirrored in the assessment profile over the next few years. While teachers and examiners acknowledge that students have difficulty with interpretation and view this as an important part of statistics (Nicholson & Mulhern, in press), the emphasis in assessment on computational skills is likely to make these the primary focus of classroom activity. The pressure on both teachers' and students' performance - as measured purely by grades achieved, is intense. Moreover, the time which will be available for teaching is likely to be reduced by the increase in frequency of assessment, with all A-level courses now having six modules.

There is a need for teachers to be comfortable in leading discussions in which variability plays a key role, and in which real data provides the evidence on which arguments are to be based. This is very different to the traditional mathematics classroom, and statistics is still largely taught in schools by mathematicians. To do this successfully, they need to have experience of the scale of variability which will be encountered in different contexts, otherwise they may commit themselves to positions which subsequently prove to be untenable, and the knowledge of this may explain in part why teachers find the tackling of such situations to be a daunting prospect.

We aim to produce guidance for teachers as to how to use these materials in class groups and how to mediate discussions generated by them. We would like teachers to be able to use them as diagnostic tools and so the guidance will also identify common difficulties, and the sort of responses students with particular difficulties normally generate in specific contexts.

REFERENCES

AQA (2000). *General Certificate of Education: Mathematics and Statistics, Specification B.* Guildford, UK. Assessment and Qualifications Alliance.

Boland, P.J. and Nicholson, J.R. (1996). The statistics and probability curriculum at the secondary level in the USA, Ireland and the UK. *The Statistician, 45,* 437-446

Gal, I. (1996). Assessing students' interpretation of data. In B. Phillips (Ed.) *Papers on Statistical Education -*

ICME8. Hawthorn, Aus: Swinburne University of Technology.

Gal, I. (1998). Assessing Student Knowledge as it Relates to Students' Interpretation of Data.. In S.P. Lajoie (Ed.) *Reflections on Statistics: Learning, Teaching, and Assessment in Grades K-12*. Mahwah, NJ: Lawrence Erlbaum Associates.

Nicholson, J.R. and Mulhern, G. (in press). Conceptual Challenges Facing A-Level Statistics Students: Teacher and Examiner Perspectives. *Papers on Teaching and Learning Statistics - ICME 9.*