A Curriculum for a Data Rich World

Gai Mooney, Australian Bureau of Statistics

The ABS and Statistical Literacy

The mission of the Australian Bureau of Statistics (ABS) is 'to assist and encourage informed decision making, research and discussion within governments and the community, by leading a high quality, objective and responsive national statistical service'. This places certain obligations on the ABS including the need to produce statistics that are accurate, relevant and accessible. With the advent of publication to the web, the statistics of the ABS are now much easier to physically access. However for the data to be truly accessible, it must be able to be understood and appropriately used. The ABS has taken seriously its responsibility to assist people understand statistics and hence has established a Statistical Literacy Unit as well as the Education Services unit.

Many national statistical organisations are similarly recognising the need to improve the statistical understanding of their citizens and their role in this. At the 2nd OECD world forum on measuring and fostering progress in societies (Nov, 2008) Enrico Giavannini, the chief statistician of the OECD, noted that there needed to be 'more investment in statistical literacy' by national statistical offices.

The OECD Instanbul Declaration (2007) declared that for societies to progress, there needed to be improved evidence based decision making. As well governments are more consistently calling for policy to be 'evidence based'. To make evidence based decisions, not only does the evidence – the data – need to be available, but it needs to be understood. The ability to understand data is a skill that can no longer be restricted to a few: workers in many fields as well as citizens need to be able to grasp statistical concepts. As Steen said in 1999 'we live in a data drenched society'; how much more pertinent is this, ten years later.

It has never been easier to collect, store and disseminate vast amounts of data – from the 120 Gigabytes of data returned by the Hubble telescope every week to the easily accessible Census of Population and Housing data that can be manipulated and then downloaded from the ABS website. Despite this plethora of data, organisations such as the ABS find it difficult to recruit statisticians and statistics faculties at universities draw less and less students – despite the fact that the 'sexy job in the next 10 years will be statistician' according to Hal Varian, Google's chief economist!

Statistical Illiteracy

There is also strong evidence that many educated adults misinterpret, misunderstand and misuse statistics. Gigerenzer et al (2007) for example cite a study where doctors incorrectly estimated the chance of a patient having cancer after receiving a positive result from a screening test. In this study only 21% of the doctors gave the correct answer (from a choice of four answers which differed by an order of magnitude).

Closer to home, it has been reported that a senior executive commented to a government official, that as the male unemployment rate was 5% and the female rate was 4%, the total unemployment rate was 9%. Another story tells of a politicaion who declared that with their election 'all schools would be above average'.

Students too, have difficulty understanding statistical concepts. A study by Watson and Kelly (2003) that looked at students' understanding of statistical language found more than 60% of students tested in Grades 3,5,7 and 9 could not explain the terms 'sample', 'random' or 'variation'. In fact 67.1% of Grade 3 and 29.5% of Year 9 students showed no understanding of the terms at all.

History of 'Statistical Literacy'

The term 'statistical literacy' was first used in the academic literature in a paper by W.F. Ogburn. The first use of the term in a paper title was in 1951 by Helen Walker (president of the American Statistical Association): 'Statistical Literacy in the Social Sciences'. In 1948 the International Statistical Institute (ISI) established an education committee which became the International Association of Statistics Educators (IASE) in 1991. In 1994 the International Statistical Literacy Project (ISLP) began in recognition of the need for improved statistical literacy across the world.

As data has become ever more accessible, as evidence based decision making has become the norm (in government, business and the community) and as technology has allowed easier data manipulation, educators have realised that they must prepare their students adequately to fully participate in and contribute to, society. There was evidence that current courses were not adequately achieving this aim. A study by Mathews and Clark (2007) for example, highlighted the need for improved understanding of statistical concepts by even successful students from introductory statistics courses. These researchers interviewed a sample of students who had received a grade of A in an Introductory Statistics course in the preceding few months. The results showed that all students had only a rudimentary understanding of the statistical concepts and a high reliance on memorised algorithms.

Over recent years more and more research has looked at the acquisition of statistical understanding at younger ages, in the primary and secondary years of schooling. Much research is looking at what statistical ideas should be introduced at what stage of schooling, how best to teach statistical concepts and what the developmental stages of statistical understanding may be.

Definition

Statistical literacy does not at this stage have an agreed definition. However there is general consensus that it is about understanding not computation, that it requires knowing how and why data are produced, that it requires a familiarity with basic ideas and terms, that an understanding of basic concepts of probability is needed and that it requires knowledge of how statistical conclusions and inferences are made (see for example Holmes, 2003; Gal and Garfield, 1997). ABS Education Services considers four statistical competencies:

1. data awareness,
2. the ability to understand statistical concepts;
3. the ability to analyse, interpret and evaluate statistical information and
4. the ability to communicate statistical information and understandings,

and maps these competencies to its resources.

What the literature tells us

A review of the literature in 2007 by Garfield and Ben-Zvi (2007) found that the research demonstrated:

- students learn by constructing knowledge;
- students require active involvement in their learning;
- students need practice;
- it is easy to underestimate the difficulty students have with statistical concepts;'
- it is easy to overestimate students' understanding;
- technological tools can assist students explore data and develop understanding;
- students require consistent and helpful feedback.

Most of these are not surprising and do not differ from the findings in most areas of teaching and learning. However the underestimation of student difficulty and the overestimation of student understanding have implications for the teaching of statistical ideas and need to be consciously kept in mind.

What students need to improve statistical literacy

To be statistically literate students must gain an understanding of key concepts such as variability and uncertainty and they must learn to accept ambiguity. There is a difference between the deterministic knowledge of mathematics and the stochastic knowledge of statistics. Students must also learn how to question data and its assumptions and they need to know and understand the differences and different uses of empirical and observational data.

If we are going to improve the statistical literacy of students there needs to be a focus on all areas of teaching and learning: on curriculum, pedagogy, technology, assessment, training and research. In 1952, the Royal Statistical Society argued that statistics needed to be 'taught in practical contexts' (cited in Holmes, 2003, p441) and this has been echoed many times since. In 1991 Radke-Sharpe for example, noted that 'statistics is an interdisciplinary subject'.

Curriculum

Statistics are already in all areas of the curriculum. Chance and Data in Maths of course, but also such areas as population density in geography, environmental statistics in SOSE/HSIE and science, economic indicators in economics, voting patterns in civics, persuasive argument in English, experimental design in science, to name a few. However it has been pointed out by many (see for example, Holmes, 2000) that these statistical concepts are introduced in various subjects with little coordination and with the assumption that students are already familiar with the basic statistical ideas. This is often not the case. /There needs to be coordination regarding how and when statistical concepts are introduced in the various subjects. It is not sufficient to recognise the need for statistical understanding across the curriculum.

Further, someone needs to take responsibility for statistical literacy or its delivery will not occur. Best (2005) for example cites the case of 'critical thinking' in the USA. There was he argues, general agreement that this skill was relevant in all areas, essential to have and should be taught in all areas. However with no coordination the 'critical thinking movement' all but

disappeared. Since it was everybody's responsibility, it became nobody's. Statistical literacy is too important for the future of today's students to allow this to happen.

To further complicate the matter however, the developmental stages of statistical understanding are poorly understood (notwithstanding the leading work of Jane Watson at the University of Tasmania in this area). The need for further research in this area is critical.

Pedagogy

Genuine statistical literacy (rather than the learning of algorithms) requires real not just realistic data (see for example Schield, 2007), it requires that students grapple with real problems and it requires explicit teaching. Teachers need to ensure that students can construct their understandings of statistical ideas by engaging meaningfully with data. Students' misconceptions need to be uncovered and addressed and time needs to be available for students to build understanding rather than just familiarity with formulae.

As well as the problem mentioned above of an uncoordinated introduction to statistical ideas across the curriculum, there is much evidence (see for example Jordan and Haines, 2006; Garfield and Ben-Zvi, 2007) that students do not readily transfer knowledge from one area to another. This too has implications for pedagogy, since it means teachers of all subject areas need to introduce the relevant statistical concepts in ways that improve understanding. The statistical concepts therefore must be overtly covered in all contexts in which they arise. This scaffolding to learning needs to be removed slowly and carefully.

Technology

With the availability of computer technology in the classroom, teachers have not only real, relevant data at their disposal, but also a tool that can enhance understanding. Technology can be used as more than a quick way to arrive at a numerical answer. Work by Ridgeway and his colleagues (see for example Nicholson, Ridgeway and McCusker, 2006; Ridgeway, McCusker and Nicholson 2008) has shown that students can successfully manipulate multivariate data to improve both their statistical and content understanding. Technology not only allows this manipulation, but it allows students to engage meaningfully with real data in which they are interested.

If students are to adequately gain experience (and practice) asking questions of data, they also need to be exposed to 'dirty data'. Here too technology allows this to happen. The CensusAtSchool datasets from ABS Education Services for example, contain the responses of over 112,000 students from across Australia. As this is raw data, students need to ask particular questions in order to examine it (for example about outliers versus errors).

Curriculum, pedagogy and technology are not in practice separated. The available technology and a curriculum that is not overcrowded, allow new and potentially more effective pedagogies to be used.

Assessment, Research, Training

These three areas also need further attention, and I will only be briefly mentioning them here.

The assessment of students' statistical knowledge is an area that needs further attention. If we don't want students to simply learn how to plug numbers into formulae without understanding, then assessment tools need to be readily available that assess deeper understanding. Here too, current technologies can be a useful tool to address this issue but there is still much work to do to develop assessments and to understand just what they are assessing.

Many teachers may also require further professional development in both the teaching of statistics and the technology that can facilitate the effective teaching of statistics. While a potential need for professional development is not a sufficient reason to postpone statistical content, it is an issue that needs tacklin.

Conclusion

If it is accepted (and it seems that it is), that statistical literacy is an essential skill for the 21st century, then there are a number of requirements for students to gain this skill.

- statistical concepts need to be explicitly taught across the curriculum;
- there needs to be a coordinated approach to teaching statistical ideas;
- data needs to be real, to both engage students and to improve their understanding;
- teaching statistical ideas needs to start early;
- technology needs to be exploited to assist teaching and learning of statistical concepts; and
- there needs to be a recognition that statistical literacy is about understanding not calculation

The introduction of a national curriculum offers an opportunity for this essential skill to be appropriately addressed to the advantage of today's students and the benefit of society.

References

Best, Joel (2005) 'Lies, Calculations and Constructions: Beyond How to Lie with Statistics', **Statistical Science**, Vol **20**, No 3, 210-214

Gal, Iddo and Joan Garfield (1997) 'Curricular Goals and Assessment Challenges in Statistics Education' in Gal, I and JB Garfield (eds) *The Assessment Challenge in Statistics Education* IOS Press; Accessed from http://www.stat.auckland.ac.ne/~iase/publications.assesbkref

Garfield, Joan and Dani Ben-Zvi (2007) 'How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics' *International Statistical Review* 75, 3, 372-396

Giavannini, Enrico (2008) presentation to the 2nd World Forum on Measuring and Fostering Progress in Societies, available from http://www.minedu.fi/export/sites/default/OPM/Tapahtumakalenteri/2008/11/setu/Enrico_Giovannini.pdf

Gigerenzer, G, W. Gaissmaier, E. Kurz-Milcke, LM Schwartz and S. Woloshin (2007) 'Helping doctors and patients to make sense of health statistics' *Psychological Science in the Public Interest*, Vol 8 53-96.

Holmes, Peter (2000), 'Statistics Across the English National Curriculum' RSS Centre for Statistics Education, http://www.rsscse.org.uk/resources./natcur.html

Holmes, Peter (2003), '50 Years of Statistics Teaching in English Schools: Some Milestones' *The Statistician*, Vol 52, No. 4 439-474 http://www.jstor.org/stable/4128125 Accessed 25/08/2008

Jordan, Joy and Beth Haines (2006) 'The Role of Statistics Educators in the Quantitative Literacy Movement', Journal of Statistics Education, Vol14, No 2 Accessed from http://www.amstat.org/publications/jse/v14n2/jordan.html on 21 March, 2007

Mathews, David and Julie M. Clark (2007) 'Successful Students' Conceptions of Mean, Standard Deviation, and The Central Limit Theorem' Paper presented as part of the Research in Undergraduate Mathematics Education Community; accessed from http://www1hollins.edu/faculty/clarkjm/stats1.pdf

Nicholson, James, Jim Ridgway and Sean McCusker, (2006), 'Reasoning with Data – Time for a Rethink? In *Teaching Statistics*, Vol 28, No. 1, 2-9

OECD (2007) Istanbul Declaration. http://www.oecd.org/dataoecd/0/24/39070305.pdf accessed 19/11/2008

Radke-Sharpe, Norean (1991), 'Writing as a Component of Statistics Education' *The American Statistician*, Vol 45, No. 4 292-293 http://www.jstor.org/stable/2684457 Accessed 25/08/2008

Ridgway, Jim, Sean McCusker and James Nicholson (2008), 'Alcohol and a Mash-up: Assessing Student Understanding' EARLI SIG Berlin http://www.dur.ac.uk/resources/smart.centre/Publications/AlcoholandaMash-upAssessingStudentUnderstanding.pdf

Schield, Milo (2007) 'Teaching the Social Construction of Statistics' Accessed from http://www.statlit.org/pdf/2007SchieldMSS.pdf

Steen, Lynn Arthur, (1999) 'Numeracy: The New Literacy for a Data-Drenched Society *Educational Leadership*, Vol 57, No. 2, 8-13

Varian, Hal (2009) cited in McKinsey Quarterly, January 2009 http://www.mckinseyquarterly.com/Hal_Varian_on_how_the_Web_challenges_managers_2286

Watson, Jane M and Ben A Kelly (2003), 'The Vocabulary of Statistical Literacy' Proceedings of the AARE conference, Accessed from http://aare.edu.au/03pap/wat03297.pdf