

Statistics as breadth: the Melbourne experiment. II: assessment

Sue Finch, Ian Gordon, Robert Maillardet
University of Melbourne, AUSTRALIA

“Critical thinking with data” is a new “University Breadth Subject” developed for first year students under The University of Melbourne’s “Melbourne Model”. It aims to teach important elements of statistical science, with minimal mathematics, and was taught in first semester 2008. We present our approaches to assessment of the subject. This has required the use of approaches that are quite distinct from mainstream statistical subjects, since students are not really being taught to do statistical work. They are required to make astute judgments of material with quantitative information, including such texts as a short article about some research in the newspaper. We have used a variety of forms of assessment, including weekly quizzes, (very) short assignments, and a larger project. The style of assessment is more consistent with that used in humanities subjects, and therefore has some important challenges for staff involved in marking, for example.

WHAT IS CRITICAL THINKING WITH DATA?

Critical thinking with data (CTWD) is a new subject available to all first-year University of Melbourne students in 2008. Students learn about important aspects of statistical science. As a “University Breadth Subject” it draws on perspectives from a very broad range of disciplines and covers stages of data-based investigations from asking a research question to interpreting and communicating findings. Our goal is to help students develop statistical dispositions and an appreciation of how core concepts of probability and statistics inform real-world arguments based on empirical data. Gal and Garfield (1997) characterise statistical dispositions as involving an appreciation of chance and randomness, of the use of data-based methods in making decisions in uncertain situations, and of the value of data-based investigations over anecdote and subjective opinion; statistical dispositions prompt us to question and interrogate data-based arguments.

With statistical dispositions, students can learn to critically assess data-based arguments in media reports and academic materials, and construct sound data-based arguments. They should have an awareness of the potential flaws in reasoning about data and the likelihood of events, and in communicating information about data-based investigations.

A ‘breadth’ subject at The University of Melbourne is broad in many senses; it draws on perspectives and knowledge from many disciplines, and provides students with knowledge and skills that have broad application across disciplines. Breadth subjects also teach students more generic skills including those relating to presentation, written communication, and teamwork.

The content of *CTWD* covers the purpose and logic of data-based investigations, the processes involved in conducting such investigations including collecting, summarising and drawing conclusions from data. Students learn about the variety of approaches that arise across disciplines, and the different processes involved. The course includes some fundamental ideas of probability, and material about how people understand probability. There is virtually no teaching of procedural and computational skills in carrying out statistical analysis in *CTWD*. However a well-tuned statistical disposition requires the ability to question and interrogate numerical information and to make sense of summary reports of data; this requires some simple computational skills. Our companion paper – Statistics as breadth: the Melbourne experiment. II: content and delivery – provides more detail about the rationale for the subject.

Student learning is often driven by assessment. There are important messages to students in the content we choose to assess, and in the way we ask questions and set projects. The assessment and exam in one semester can set the agenda for students in subsequent semesters. The framework briefly outlined above required assessment tasks that allowed students to demonstrate their developing statistical dispositions as well as some generic skills. Assessment materials needed to be strongly grounded in real-world problems; richness without ambiguity or too much statistical complexity would be ideal.

OUR STRATEGY

A team of staff in consultation with academics from a range of disciplines developed the course materials for *CTWD*; the course was developed from the ground up. A framework of lecture topics with some details of content was outlined. We then described the specific learning outcomes that we wanted to achieve for each topic. By considering the learning outcomes at this stage, we ensured that lecture material was well-tailored to the types of knowledge, skills and dispositions we wished to assess.

For example, one of the topics *CTWD* covered was graphics. The lecture content provided illustrations of good and bad graphical practices, and discussed important features of good graphs and research on interpretation of graphs. Lectures gave examples of media reports including graphical displays and modelled how to critique and improve graphical displays. Five principles for good graphics were presented, and a number of standard forms were recommended. The principles were developed from the work of Tufte (1983) and Cleveland (1994).

The learning outcomes required students to develop limited skills in producing graphs for small data sets. However, they did not use software. Rather, students should understand:

Good graphical presentation of data rarely happens automatically.

Good graphs are simple in design.

Graphs should, above all else, show the data clearly.

Good graphs have transparent visual encoding.

Good graphs have accurate titles and well-labeled axes with measurement units defined.

Axes should generally maintain constant measurement scales.

Good graphs identify the source of data.

Pie charts and pictograms are poor choices for representing data.

Aligning data along a common horizontal axis facilitates accurate comparisons.

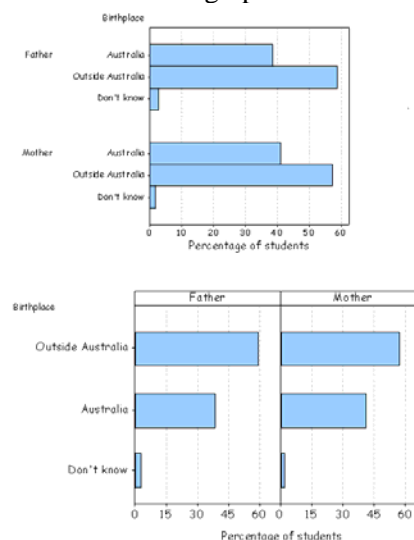
Standard forms have been developed for particular purposes.

Unusual values should be investigated carefully for information they may provide; they are not necessarily mistakes.

A limited search for some very simple graphics software capable of supporting the production of well-designed graphics was unsuccessful. Some commonly available software was avoided because it encouraged the unthinking production of some very poor designs and non-standard forms.

The assessment tasks we designed evaluated students' capacity to apply this knowledge appropriately to real-world problems. Hence we asked questions that related to recognising the instantiation of the principles we wished students to learn rather than requiring the production of 'statistical output' according to these principles. For example:

Consider the two graphs below. They show responses to the questions about the birthplace of parents from a 'Getting to know you' survey designed collaboratively with students in the course. Which important graphical design principle described in lectures is used in the top graph but not the bottom graph?



OUR CHALLENGES

There were many challenges in developing and choosing methods of assessment and appropriate materials for *CTWD*. A very tailored approach was needed; existing assessment resources, including exercises and activities, could not really contribute to the course. The assessment tasks should require the application of generic skills and data-related thinking skills; traditionally assessment in statistical courses had focussed more on the data-related skills.

We chose to use examples and materials that students themselves might have easily come across, in order to emphasize the relevance of the course to material met in the media as well as academic contexts. There are many beautiful historical examples and case studies that are ideal lecture material, some of which we used, including stories of the trials of the Salk vaccines, the decision to launch leading to the Challenger disaster, and John Snow's mapping of the 1854 cholera epidemic. Assessment materials needed to be contemporary and have a strong local flavour. Local material we used included the story of the breast cancer cluster in the ABC studios in Toowong, research on cannabis use in Australian adolescents and a Herald-Sun survey of Victoria Police.

Traditional statistics courses often rely on good examples of the methods they are trying to teach. We needed to find examples with a mix of good and bad features, and with the right degree of complexity to allow students to apply their critical enquiry skills. Simple flawed examples of data-based investigations can assess an important principle but have limited scope for assessing varying depths of students' understanding; examples that are too complex can include features that are not relevant to the assessment task but may mislead students to focus on them.

The student cohort in *CTWD* is mixed; in Semester 1 2008, 28% were Arts students, 27% were Science, 18% and 15% were from Commerce and Biomedicine respectively. The capacity of this mix of students to deal with the course content and assessment was unknown. Interestingly, the course attracted a broader spread of students than many other University breadth subjects launched at the same time. The majority of students in many of these other subjects were from just one or two faculties corresponding to a specific discipline interest relevant to their main course of study.

Traditional statistics courses often rely on a mathematical approach to teaching and assessing many fundamental statistical ideas. *CTWD* included topics such as variation and statistical modelling, but without formal mathematical treatment. Developing appropriate lecture content required innovation and finesse; setting appropriate assessment tasks was even more challenging.

TYPES OF ASSESSMENT

Assessment needed to be continuous for both course developers and the students. Early in the course, students expressed uncertainty about what they were supposed to be learning from lectures, or indeed what the subject was "all about". Course developers needed feedback about students' capacity to deal with the scope of material presented. *CTWD* used four types of assessment.

Weekly quick quizzes of 8 to 10 questions could be attempted repeatedly (until the closing date). These quizzes were worth 5% of the total marks; they were important for revision of lecture content and concepts. There were 5 short assignments worth 4% each and totalling 1000 words. A major assignment was worth 25%, and had two components – group work and an individual write up. The final exam was worth 50% and was modelled on the other assessment components. The quizzes, short assignments and major project are described in detail below. We also provide an example from the exam.

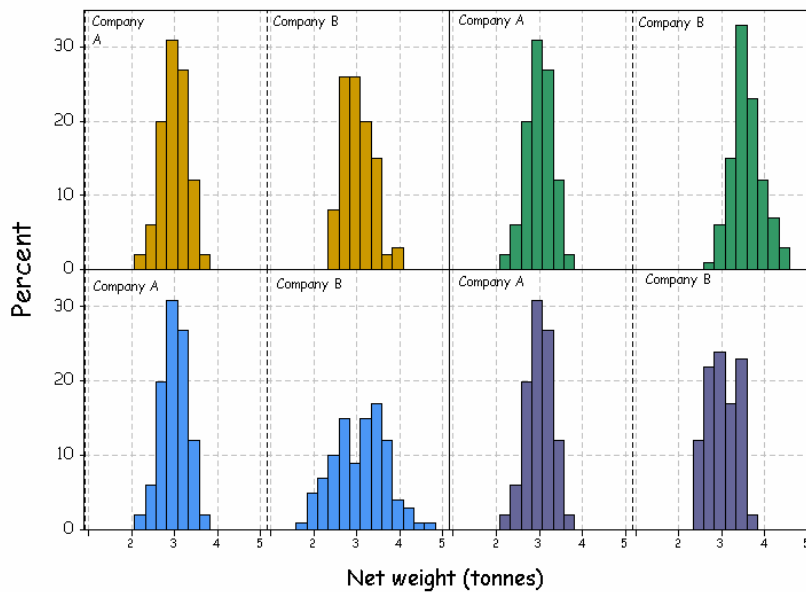
QUICK QUIZZES

Students completed the weekly quick quizzes via the University's Learning Management System. This allowed a variety of different types of questions to be asked including multiple choice, multiple answers, exact numerical, matching, ordering, fill-in-the-blank and 'hot spot'; all of these types of questions were used in *CTWD*. A 'hot spot' question gives a visual

representation, and students need to click on the area of the representation that gives the answer to the question.

The use of visual representations was one approach used to present and assess concepts and principles traditionally given a formal mathematical treatment. For example, this hot spot question examined the idea of modelling location differences in relation to background variation:

There are four pairs of histograms below. Each compares the net weight of garbage dumped by two companies. Click on the pair that shows the strongest evidence of mean differences between the two companies, relative to the background variation within companies.



SHORT ASSIGNMENTS

The length of the five short assignments was deliberately very constrained – students could only write 200 words on each. The due dates for the assignments were spaced throughout the semester. Broadly, the assignments asked for a critical evaluation of a data-based argument, report or representation; in each case, students were guided in terms of which aspects to focus on. For example:

This assignment is about a small item in The Age’s “Diary” column, which is published on the back of the front section of The Age newspaper. The item appeared on 29 November 2007 as follows:

Yeah, right

... And how about No Idea Magazine’s latest poll discovery: ‘one in five’ women in Australia felt threatened by violence in their homes. Check the bottom line and you find the poll quizzed just 1500 of the nation’s 20.4 million people, thereby smearing at least 2 million blokes in a poll covering just 0.000073% of the population.

Students were asked to write two emails: one to The Age diarist pointing out the misconceptions and errors in the short quote, and a second to the magazine involved (New Idea, derisively referred to as No Idea) to ask for important information to assist in an evaluation of the survey.

This was the second short assignment; it assessed students’ understanding that it is the sample size rather than the sampled proportion of the whole population that is important to the statistical reliability of a study. Lectures had discussed the 1936 Literary Digest poll extensively. Students needed to check the percentage reported from the figures provided in the article in order to identify a computational error. Most students took the calculation at face value; they did not

expect that it required checking. This is an example of the kind of checking and simple computational skill needed.

Another short assignment asked students to evaluate a graphical representation found on a website describing the capacity in Melbourne's water storages over time. Students were provided with a number of alternative representations of the data, and asked to choose the one best suited to a particular purpose and to write a description based on this.

The short assignments assessed critical thinking skills in an applied context, and also required good reading comprehension and writing skills. The assignments gave very clear guidance about the nature of the task, the content to consider, and the points to be addressed in the word limit. Some students still failed to evaluate the materials within these guidelines.

MAJOR ASSIGNMENT

The major assignment gave students the opportunity to make a detailed review of a single research study. The lectures illustrated how media reports often only tell part of the story, and how many important and interesting questions are raised when the background material for the study is examined. Students worked on one of five case studies; each case study had a newspaper report and a published article.

There were two parts to the major project. Students worked in groups on the first part; groups were formed within tutorials to prepare for a (tutorial) poster display and presentation session. This session was modelled on a conference poster session; each group gave a brief presentation of their poster for 3 minutes, and then assessors and fellow students circulated the room to learn more about the research study. Both the verbal and poster presentation were assessed. One aim of this task was to help students understand the research study they were reviewing quickly and efficiently. We chose studies that were relatively straightforward in their design, but the complexity of the statistical analysis was quite varied. We did not expect students to understand all of the details of the analysis in the papers provided; in the tutorial time provided for establishing the group work, students were given guidance about this.

The second and individual part of the major project asked students provide a 1200 word critical assessment of the reporting of the study in the news item and in the published article, and of other aspects of the design, implementation and analysis of the study. This critique required description and explanation of both the strengths and weaknesses of the study. Students could comment on the analysis of some of the studies, for example, by critiquing the graphical presentation of the results.

The major project assessed students' capacity to work well in groups as well as their oral presentation and written communication skills.

EXAMINATION

The examination included quick quiz type questions, short answer questions and longer questions similar to those used on the short assignment. Here is an example of a longer question:

In April 2008, The Herald Sun published a series of articles over several days about a poll of Victorian police officers. The headline on the first day was "POLL: POLICE FACE CRISIS". There had been heightened tensions between the Police Commissioner, Christine Nixon, and the secretary of the Police Association, Senior Sergeant Paul Mullett.

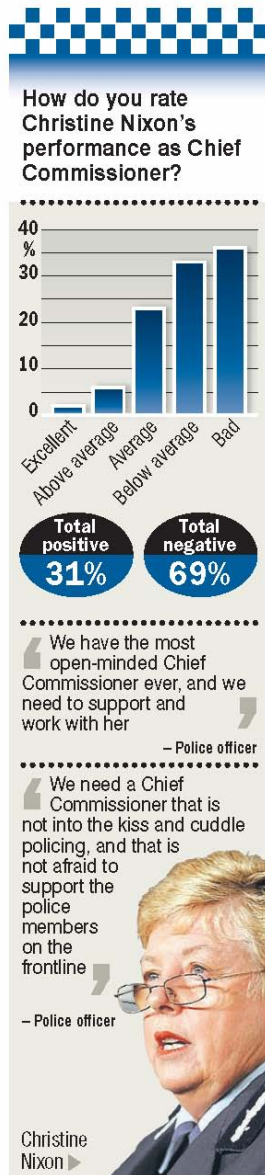
The Herald Sun published the following information about the survey:

"The Herald Sun wrote to more than 11,000 Victoria Police officers and invited them to complete the Herald Sun survey – 3459 responded. That means 30 per cent of the force's sworn officers took part in the unprecedented poll.

The survey sample of 3459 is much larger than those commonly used by pollsters. Galaxy Research vetted the questions to ensure they conformed to accepted standards of conducting research in Australia.

The Herald Sun provided each police officer with a unique password, which expired after it was used once. That ensured they could not submit multiple entries.

The completed surveys were analysed by an independent market research company, which provided the results to the Herald Sun for publication.”



Unless otherwise specified, use the extracted information from the Herald Sun article to answer the following questions.

- What was the sample frame for this survey?
- The day the survey was released, Chief Commissioner Christine Nixon was critical of it: “I don't intend to go anywhere,” Ms Nixon said this morning after a landmark survey raised a raft of serious issues about the state of the force. Speaking on radio 3AW this morning, Ms Nixon attacked the results because just 30 per cent of sworn officers – 3459 – had responded to the survey. “That is despite the sample being much larger than that commonly used by pollsters.” (Herald Sun)

What was the basis of Christine Nixon's criticism?

- What argument is the Herald Sun using in its comment: “That is despite the sample being much larger than that commonly used by pollsters.”?

Provide an analysis of the argument.

- What strengths and weaknesses of the survey are apparent from the Herald Sun's description of the survey? Comment on any features you have not discussed in (b) and (c) above.

Parts (e) to (h) refer to the extract from the Herald Sun shown at the left.

- What type of information is in the two quotes supplied at the bottom? What weight should be attached to these comments as evidence?
- Examine the bar chart at the top of the extract. What type of data is represented?
- Comment on the descriptions of the categories of response (“Excellent”, “Above average” ... “Bad”) and the effect they could have had on the responses.
- How were the two simpler categories – “Total positive” and “Total negative” constructed? Comment on this construction.

This relatively simple example allowed questions to be asked about many aspects of the processes involved in a data-based investigation as well as to assess examples of statistical reasoning. In the examination, questions about this type of material were quite structured give the time constraints within which students had to work. In short assignments there was less explicit guidance.

MARKING

Tutorial staff from three different departments supported CTWD. Just as the students were required to make astute judgements about the material provided, tutors were required to make astute judgements about student arguments. The style of assessment is more consistent with that used in humanities subjects. Constraints on short assignments were set to encourage students to write clearly and succinctly, but also to limit marking loads. Second markers attended poster display and presentation sessions to assist with on-the-spot marking of presentations and student responses to questioning.

STUDENT FEEDBACK

Formal feedback about assessment and other aspects of the course was sought in an online survey during the fourth week of the course; 93 students (63% response rate) responded. At this time, over 50% of students indicated that the short assignments were beyond the right level of challenge. About half of the students found the word limit too restrictive. One student gave this perspective: "Short assignments are perfect I think, and the 200 word limit really forces you to develop concise yet relevant points." Informal feedback towards the end of the course suggested that students did appreciate the challenge and constraints of short assignments, and they felt the assignments had helped them develop their writing skills. We are yet to receive results from students' evaluations in the final week of the course.

LESSONS

The development of any new subject is time consuming. The development of *CTWD* had unusual demands of needing specialised materials and innovative approaches to teaching and assessment. The contribution of staff working in applied fields meant that a great deal of potential contemporary local material could be identified for lecture material and interested assessment tasks. Often our investigations of the details and background behind the stories we found revealed richer material that could also be used in teaching and assessment. In two cases the analysis work done to prepare a case study for presentation in the subject led to important discoveries about the published research, which were communicated back to grateful original authors.

The challenge to teach a mixed and different cohort of students forced us to consider the many different ways in which important statistical concepts, principles and ideas can be represented and explained – through examples, counter-examples, visual representations, simulations, and (very occasionally) with formulae. The need to consider the broad relevance of statistical dispositions to many disciplines without relying on mathematics brought new perspectives on teaching and assessment – perspectives that will inform and enhance our teaching in a broad range of courses. Our own assessment of the approaches we used to assessment in *CTWD* has only just begun.

REFERENCES

- Cleveland, W.S. (1994) *The elements of graphing data*. Monterey Park, CA: Wadsworth.
- Gal, I., & Garfield, J. (1997) Curricular goals and assessment challenges in statistics education. In Gal, I., & Garfield, J. (Eds.) *The assessment challenge in statistics education*. Amsterdam: International Statistical Institute.
- Tufte, E. (1983) *The visual display of quantitative information*. Graphics Press.