EXPERIENCES WITH REAL AND ACCESSIBLE RECENT DATA IN CONTEXT TO MOTIVATE STUDENT LEARNING AT HIGHER LEVELS IN STATISTICS

John Harraway
University of Otago, Dunedin, New Zealand
jharraway@maths.otago.ac.nz

Motivational data and statistical analyses arise in research in the workplace and at university. Videos of researchers developing context are recorded. As well as designed experiments researching food authentication, the videos show data from surveys reviewing social issues. These include a postal survey of 2200 citizens investigating use of taxes for building a controversial sports stadium, a telephone survey of attitudes of 1200 women to alcohol consumption during pregnancy, and contrasting opinions of 5000 tourists from Japan, Australia and Germany to the attractions of New Zealand. The studies are practical, relevant and locally generated which capture student interest. Survey design is discussed. Statistical techniques used are regression modelling and multivariate procedures. A data set is given to students who have eight weeks to analyze it and write a report of 25 pages for the organization which commissioned the study.

INTRODUCTION

Successful approaches to training both specialist statistics students and users of statistics for evidence based decision making in the workplace will be illustrated. Context and Inquiry Based Learning (IBL) are important components.

Context motivates learning in statistics. Cobb (2007) stated that "context is an essential part of statistical thinking, and some of the worst teaching of statistics occurs when the teacher or textbook tries to treat context as irrelevant". Finch and Gordon (2010) report student feedback from questionnaires emphasize the importance of context. Neumann et al (2013) report: "When real-life data is used to illustrate statistical concepts, practice calculations, and show the application of statistics, there is some association with specific types of student learning experiences such as relevance, understanding and learning, interest, motivation, and engagement", concluding that "- it might be fruitful for statistics teachers to incorporate real-life data into their teaching". There is therefore either a demand for access to interesting recent data with context described, or training in methods for designing experiments, conducting surveys and constructing questionnaires.

Once data are identified students can be left to analyse it and write a summary for management. Spronken-Smith et al. (2012) report that such an IBL approach enhances student learning and introduces "problem-based learning", "undergraduate research" and "research-based teaching". For the two studies reported in the following students learn the necessity of cleaning a large data set, how to summarize data visually, how to analyse data either individually or in teams, how to use software and how to report results, conclusions and recommendations for management in a workplace.

SOURCES OF DATA

There are a several sources of interesting data such as GAISE (2010) which describes a curriculum including suggested activities, CAUSE (2011) which is a compendium of resources and the Data and Story Library (DASL, 2011) which is an online library of data files and stories that illustrate the use of statistical methods. Statistics New Zealand (http://www.stats.govt.nz/) also provides data based on the National Census, such as a SURF (Synthetic Unit Record File) http://www.stats.govt.nz/tools_and_services/schools_corner.aspx. Larger files such as a CURF (Confidentialized Unit Record File) are also available for research purposes. There are numerous other files available from Statistics New Zealand for use such as the Household and Labour Force Survey. Harraway and Forbes (2013) describe a complete postgraduate course in Official Statistics which uses such files and others in the areas of data visualization, demography, health statistics confidentiality, GIS and economic statistics.

Data reported here come from a set of 19 videos which can be uploaded from http://www.maths.otago.ac.nz/videos/statistics/. Nineteen researchers have been recorded describing their research and presenting their data for analysis. Designed experiments for

measuring shell fish growth in tidal inlets, measuring the effect of high altitude on sleep apnea, assessing the survival of the rock wren, a small endangered native New Zealand bird, and investigating food origin and authentication of oysters involve small data collections. Another study, a cluster survey of iron levels in 330 new born New Zealand infants and toddlers, examines reasons involving diet and infection for low iron levels, a health problem for young children in New Zealand, possibly related to high consumption of cows' milk rather than breast feeding. A further study is based on a postal survey of 2200 people investigating opinions on use of taxes for building a controversial sports stadium.

A larger study assessing opinions of 3664 tourists from Japan, Australia and Germany to New Zealand is described in the next section. This is followed by a discussion of a second case study under development which includes a questionnaire administered to a large cohort of students whose attitudes on sustainability are being monitored over four years. The two studies feature Inquiry Based Learning and are used in a third year course on multivariate statistics where they are designed to introduce students to research at the undergraduate level.

ATTITUDES TO TOURISM IN NEW ZEALAND: AN IBL APPLICATION

A survey of actual tourists and potential tourists to New Zealand with 1043 people from Australia, 1526 from Japan and 1095 from Germany interviewed previously in their home countries or at airports as they left New Zealand has been carried out. The purpose of the investigation was to produce a report for tourist and travel agencies in New Zealand which summarises opinions of potential and actual visitors to the attractions available. New policies are to be recommended for attracting visitors to New Zealand through appropriate marketing in the three countries and to identify potential development of new activities.

Context and Data

Context is described by the lead researcher in one of the 19 videos which can be viewed directly at http://www.maths.otago.ac.nz/videos/statistics/Tourists/index.html. Professor Gnoth, a colleague from the Marketing Department at the University of Otago, describes data presented in nine files. The three files for each country are opinions of people who were interviewed at home but who did not visit New Zealand, people who were only interviewed after they had been to New Zealand, and people interviewed at home who also visited New Zealand. There are 127 variables recorded for each of Australia and Germany and 124 variables for Japan. There are 31 variables measuring the strength of intention to pursue 31 main tourist activities when in the country, 31 variables recording the number of times actual visitors undertook each of the 31 activities and 31 binary variables recording whether the activity was considered a highlight. In addition, the profiles of all interviewed include age, gender, education level, income and occupation while other variables record type of accommodation, mode of travel and time spent in New Zealand. Because tourism is a major source of income for New Zealand and students, being local, are familiar with all the tourist attractions, the context is motivational with student responses confirming this.

The data are released for student investigation six weeks into a 15 week semester. The data are presented in Excel and SPSS since SPSS has been used in the first part of the course covering multivariate analysis of variance, principal components, exploratory factor analysis, discrimination and cluster analysis. Subsequently, logistic regressions for discrimination, confirmatory factor analysis, scaling, canonical correlations and correspondence analysis are taught and can be used for analyses.

Instructions for students

Each student in the class is told that after 8 weeks they are expected to produce a report up to 30 typed pages including results and recommendations for tourist agencies in an executive summary. Students are advised to begin their investigations early which can involve simple visualisation techniques and data summaries. Students can work in approved pairs provided the contributions to the report are shared. Computer output can be added in a set of appendices and further research into context or use of methods not taught in the lectures must be referenced. The final document is bound as an undergraduate mini thesis.

Limited advice is given with students left to their own devices. There are many missing values and small differences in the sets of 31 activity variables in the different countries making some comparisons difficult. The first steps are therefore to clean the data allowing country attitude comparisons to be made. This can involve formulating an approach to estimating missing values and deciding a policy on merging different files. It is useful to select data subsets, recode variables, reorder variable values, create new variables and investigate approaches to missing value estimation.

After cleaning the data visitor attitudes, both within and between the three countries, can be investigated. Profiles with respect to occupation, age, gender, accommodation, mode of travel and time spent in New Zealand can be reported. Activity attitudes for different groups of tourists can be compared and features discriminating between different countries summarised. There is a natural interest in the relationship between what people thought they would do in New Zealand before coming and what they actually did during their visit. Students are told to formulate at least six research questions and they are free to use whatever SPSS (or other software) options they wish in their analyses but their appendices should make it clear what they have done and some key tables or graphs may be copied from the appendices into the report. Marks are awarded for the breadth of the hypotheses investigated, for the clarity of the report and for its general structure. Extra marks are awarded for ingenuity in the investigation. Some unexpected results always emerge.

Some results

Factor analysis and cluster analysis on the 31 activities identified factors within each country and across the three countries with students reporting names and descriptions. Some students identified typical factors as follows:

- Intellectual Tourism: includes visiting *museums & galleries*, *botanic gardens*, *historic places* and taking *sightseeing tours*, *short walks* and *boat tours*. These activities were popular for older tourists.
- City & Nightlife: includes *shopping*, *meeting locals*, trialling *local cuisine* and *evening entertainment.* All age groups participated as expected.
- Wilderness & Adventure Sports: includes *rafting*, *jet boating*, *scenic flights, parachuting* and *visiting glaciers.* Young tourists, mainly male, found these activities popular.
- Winter Sports: includes *skiing* and *winter sports* and was found to be particularly popular among young tourists.
- Maori Culture: includes *Maori performances* and *meeting Maori* which was popular for the three countries and all age groups.
- *Bungy-jumping* was not popular for visitors from all three countries

Discriminant function analysis distinguished between countries. Australians were more interested in theatre and winter sports than the Germans and Japanese. Germans were more attracted to swimming and viewing marine life than the Australian and Japanese while the Japanese were more interested in hiking, golf and casinos than the Germans and Australians.

German tourists are on average younger than Australian and Japanese tourists. The majority of German tourists are young adults between 20 and 40 and with very few retirees. By contrast Australian and Japanese visitors are older between 40 and 60 with many retired. The majority of German and Japanese tourists are male, whereas the majority of Australians are female. German tourists had a higher level of total satisfaction with their visit to New Zealand than the Australians.

Student reaction

The results reported are a sample. Other approaches generated many more conclusions. One student carried out an investigation of the missing value options in SPSS which was unexpected but an excellent learning experience. This has encouraged me to include a module on missing values in the course in future. Several of the students said "These data are interesting. Can we have more time to analyse the data as I wish I had started earlier?" Although I found this response encouraging, it was not possible to allow more time as the projects had to be read and marks allocated.

A NEW IBL CASE STUDY

For continued student motivation and training using Inquiry Based Learning it is important to update and develop new contexts with data which are current, relevant and interesting. Harraway et al (2012) in 2009 surveyed the attitudes to sustainability of 541 students in a first year undergraduate statistics class using the Revised New Ecological Paradigm abbreviated here as the NEP (Dunlap et al., 2000). The NEP includes 15 statements that relate to limits to growth, the position of humans in the environment, the fragility of nature, and the imminence of ecocrisis. Respondents were asked to record their agreement with 15 items on 5-point Likert-like scales. The cohort surveyed included students from Surveying, Human Nutrition and Ecology. The respondents were anonymous but each participant was asked to write a confidential code on their survey to allow follow up during subsequent years at university in their chosen disciplines. Over the next three years therefore the same instrument could be used to monitor changes in ecological worldviews and contributory sustainability tendencies of the cohort while studying at university. Changes in the attitudes within the three disciplines can be monitored and gender differences investigated. The combined data which will be released to the multivariate statistics class is complicated but lends itself among many possibilities to a longitudinal analysis using mixed models.

It will be suggested to the class that they consider factor analysing the initial data to identify common factors. Whether there should be three, four or five factors or even more should be reviewed using a confirmatory factor analysis package such as SPSS/AMOS which is taught in lectures. Combinations of the 15 items should also be considered. The class is aware of the four factors originally identified by Dunlap and they can research published work about these factors. Participant student responses can be combined into a summated overall NEP score where higher overall scores indicate stronger pro-environmental attitudes. Similarly the responses for the items forming each identified factor can be summed.

The data for the cohort of first year participants in 2010 has also been collected and this can be examined similarly to verify whether there is factor agreement with the results from 2009. Students will be able to report on this if the opt for this as one of their analyses. Another study (Jowett et al., 2013) uses all of the data from the calendar year 2010 and includes those attending Year 1 (104 participants) and Year 2 (311 participants). The three responses are whether participants have weak sustainability attitudes, or are unsure about it or whether they hold strong attitudes to sustainability and these can be modelled in terms of year of study and discipline using an appropriate multinomial regression.

A key analysis involves modelling the responses of the participants over the period 2009 when arriving at university to 2012 when they are in the final year of undergraduate study. The unique identifier for each respondent makes possible this analysis which will identify trends in the different disciplines, taking account of the correlations present with the use of such longitudinal data.

DISCUSSION

Context is known to motivate learning in statistics. It is important to continue to update case studies by collaborating with colleagues from a wide range of disciplines which apply statistics. In the cases described here staff in the Department of Marketing and staff in the University Higher Education Development Centre are called upon showing that there are always possibilities of colleagues contributing interesting data.

Inquiry Based Learning is less known for its success with case studies like those described here. The notion of introducing research in statistics classes at the undergraduate level for both statistics specialists and users of statistics from other disciplines appears to be well received by the students.

Preparation for the workplace is an important bonus if external agencies provide the data for analysis and propose hypotheses for investigation. In the case of the tourism data some very clear marketing objectives were identified and it was New Zealand Tourism which funded the Department of Marketing. Fortunately they allowed the data to be used.

A problem highlighted is the importance of designing a questionnaire or survey document carefully as this will simplify "cleaning" the data, a process which results in students appreciating the problems of a poorly designed questionnaire.

The size of tourism study described here was initially viewed with alarm by the class but towards the end of the semester it was common to receive comments like "this is interesting. I wish I had started earlier. Can I have more time?"

Smaller assignments covering specified techniques and contexts like some of those described in the earlier videos are set for the students as each technique is taught in the course. There are tutorials to help with these regular assignments but little help is given with the large project.

Final reports are word processed and include an executive summary with recommendations, an introduction, chapters developing analyses and references. One student wrote an exceptional report which was shown to the lead researcher in the Marketing Department who said he would like to work with the student to write a paper for publication.

REFERENCES

Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) (2011). Alexandria, VA: American Statistical Association. www.causeweb.org

Cobb, G. (2007). One possible frame for thinking about experimental learning. *International Statistical Review*, *75*(3), 336-347.

The Data and Story Library (DASL) (n.d.). Ithaca, NY: Cornell University. *http://lib.stat.cmu.edu/DASL/* accessed 8 Dec 2011.

Dunlap, R., van Liere, K., Mertig, A., & Jones, R. (2000). New trends in measuring environmental attitudes: Measuring endorsement of the New Ecological Paradigm: A revised NEP Scale. *Journal of Social Issues*, *56*(3), 425-442.

GAISE (2010). *Guidelines for assessment and instruction in statistics education, college report.* Alexandria, VA: American Statistical Association. *www.amstat.org/education/gaise/*

Finch, S. & Gordon, I. (2010). Lessons we have learned from post-graduate students. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society.* ICOTS8 Proceedings, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

Harraway, J., Broughton-Ansin, F., Deaker, L., Jowett, T., & Shephard, K. (2012). Exploring the use of the revised new ecological paradigm scale (NEP) to monitor the development of students' ecological worldviews. *The Journal of Environmental Education, 43*(3), 177-191.

Harraway, J. A., & Forbes, S. D. (2013). Partnership between national statistics offices and academics to increase official statistical literacy. *Statistical Journal of the IAOS*, *29*, 31-40.

Jowett, T., Harraway, J., Lovelock, B., Skeaff, S., Slooten, L., Strack M., & Shephard, K. (2014). Multinomial-Regression Modeling of the Environmental Attitudes of Higher Education Students Based on the Revised New Ecological Paradigm Scale. *Journal of Environmental Education*, *45*(1), 1-15.

Neumann, D. L., Hood, M., & Neumann, M. M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal*, *12*(2), 64-75.

Spronken-Smith, R., Walker, R., Batchelor, J., O'Steen, B., & Angelo, T. (2012). Evaluating student perceptions of learning processes and intended learning outcomes under inquiry approaches. *Assessment & Evaluation in Higher Education*, *37*(1), 57-72.