EXPERIMENTAL RESEARCH IN A STATISTICAL CONCEPTS COURSE

Alice M. Richardson
University of Canberra
Australia

This paper describes the experience of synthesis of concepts learnt in a semester-long statistical concepts course into a major piece of work. In the University of Canberra course entitled "The World of Chance", the major assessment item is a group project. Groups of two or three students identify a research question to study via a small experiment or observational study, carry out the data collection, calculate descriptive statistics and draw simple conclusions on the basis of those statistics. In 2001, in an attempt to enhance the integration of topics across the course in their assessment, all students were directed to carry out a project involving an experiment. This paper describes the material covered in class on experimental design, and the workshop activities used to support this material. This paper also describes a selection of the topics investigated by students, and discusses the degree of integration achieved by the students in their projects.

INTRODUCTION

The integration of statistics and research designs into a single course is an ambitious goal that can be attacked in a range of ways, at a range of levels. This paper will focus on a particular case of integrating statistics and research designs not into a mathematical statistics course at an advanced level, but rather into a statistical concepts course at an introductory level. These courses are frequently inspired by the recommendations of the ASA/MAA Joint Curriculum Committee (see Moore, 1997), which maintain that the way to teach statistics in the 21st century is not with an emphasis on statistical techniques, formulae and memorizing but with an emphasis on statistical thinking, data and active learning.

In this paper we give a brief overview of published experiences and advice concerning the integration of experimental research into statistical concepts courses. We then describe a particular statistical concepts course at the University of Canberra, called "The World of Chance". In 2001 a special effort was made to integrate experimental research into the lectures, tutorials and assessment of the course. We describe the process, the results of this effort, and finally we offer recommendations for raising the degree of integration achieved.

OTHERS' EXPERIENCES

It is virtually impossible to identify the originators of the idea of integrating research projects into statistics courses, although Hunter (1977) seems to be one of the earliest authors to suggest using projects to enliven the teaching of experimental design. Lock and Moore (1992) advise that (a) students work in groups to more closely mirror real-world approaches to research; (b) that they submit and have approved a proposal in order to screen out inappropriate or infeasible topics. Roberts (1992) classifies student projects into three main types: sample surveys, predictive time series and process improvement. His classification reflects his experience as a professor in a business school: we would add a fourth type to Roberts' list, namely experiments, which are the focus of this paper. As will be seen in later sections, when students are given a free choice of topics some will naturally choose a predictive time series type topic, which still fits under the broad category of experimental research.

Mackisack (1994) reports her experiences with including experiments as part of a statistics course taken by mathematics majors, some of whom go on to become practicing statisticians. She notes that the main points students learn in the process of carrying out the experiment include (a) the distinction between an experiment and an observational study; (b) the difference between factors and blocks; (c) the importance of minimizing uncontrolled variability; (d) not all errors are Gaussian; (e) ways of dealing with factors outside the experimental design; and (f) data do not always become available as planned.

Issues associated with the use of research projects in statistics courses can be embedded in the broader debate about assessment of statistics course in general. For example, Garfield (1994) gives two guiding principles for assessment. Firstly, assessment should reflect the statistical content that is most important for students to learn and secondly, assessment should enhance the learning of statistics and support good instructional practice.

THE WORLD OF CHANCE

"The World of Chance" has been taught since 1998 at the University of Canberra. "The World of Chance" was inspired by the courses described in Snell and Finn (1992) and online on the Chance Database (http://www.dartmouth.edu/~chance/). "The World of Chance" is a one-semester course taken by third year students, usually as part of a "general education" requirement in their degrees. There are no pre-requisites. The stated goals of the course are to enable students to assess critically the statistical content of articles and other presentations in the media, and to understand and appreciate the judicious use of statistics in their own field of study. Most of the students who enrol are information technology majors, along with a small number of students from other disciplines ranging from nursing to communication.

"The World of Chance" is not like a standard introductory statistics course, hence the use of the phrase "statistical concepts" in the title of this paper. There is a flexibility of content, as there is no syllabus set by other departments or outside bodies. A high level of mathematical skill is not required of, nor developed in, the students: virtually no formulae are introduced. Not as many statistical techniques are taught either. In particular, hypothesis testing is omitted, although confidence intervals are introduced very early in the course through margins of error in surveys. After all, the aim of this course is to teach statistical concepts, not statistical techniques. A number of students have done a "standard" introductory statistics course and, later on, "The World of Chance", and have found the fresh approach to statistics quite enough to challenge them. Moore (2001) is used as the textbook. Lecture material is taken from that book and supplemented with activities from Scheaffer et al. (1996). A list of broad topics covered is given in Appendix 1. Most of the tutorial activities are taken from Spurrier et al. (1995), and a list of these is given in Appendix 2.

The main assessment item in "The World of Chance" is a project, consisting of a report and a poster, worth 40% of the final mark, to be carried out individually or in groups of two or three. In previous years the topic and type of study were left completely open, which resulted in projects of widely varying standards and usefulness to the students. A particular problem was found to be students' ability to identify whether a survey or an experiment was an appropriate method of answering their research question. For example, one assignment included the question "Design a study to test the truth of the proverb 'Red sky at night, shepherds' delight; red sky at morning, shepherds' warning'." Many students answered by describing surveys in which they asked people whether they believed the saying to be true or not, rather than designing an observational study of some sort. Furthermore, having groups of students rushing about collecting data in the form of questionnaire responses was also seen as potentially risky with regard to the image and ethical behaviour of the University. Thus in 2001 the emphasis was placed on designing, carrying out and reporting the results of an experiment.

THE 2001 EXPERIENCE

Students were allowed to form their own groups of one, two or three people, and then submitted a project proposal by week 7 of semester. They also received a page of advice on carrying out projects in the form of quotes from students in previous semesters who had reflected on the process of carrying out this assignment themselves. The advice was grouped under the following headings: topic selection, planning of the whole project, timing of the project, background reading, data collection, preparation of the report and preparation of the poster. In week 12 the groups submitted a progress report, and in week 16 (one week after lectures had finished) the project and associated poster were due. Ninety-three students submitted a total of forty-five projects. Fourteen of the projects were essentially two-factor factorial designs, looking at topics ranging from plant growth to battery lifetimes. Eleven projects were essentially one-way designs, investigating anything from television commercials to vocabulary richness in novels. Seven projects involved collecting time series data, whether of gym usage, petrol prices or exchange rates. Six groups collected observational data, at locations ranging from supermarket checkout lanes to airport terminals. Five teams carried out simulations to estimate parameters of processes such as poker machine payouts. Finally, two teams collected data in order to study correlations: one between basketball players' heights and salaries, and one between crime rates and income in Australian states.

Seven main areas where the research projects might fall short of an ideal piece of research were identified. These will now be discussed in turn. First, essential details of experimental protocol were sometimes omitted, which meant that it would not be possible for another team to take the project and reproduce it exactly. For example, the team who studied how gender and age relate to shampoo preferences categorized shampoos into two types: those with chemical names on the label and those without. However they did not list which shampoos fell into each category. Another team who studied vocabulary size amongst ten authors of novels did not say which ten authors were actually studied.

Second, students sometimes attempted research which generated data that was too complicated for them to analyse fully, because they did not have the technical expertise, and they did not recognize this when they designed their data collection plan. For example, one team collected data on the colour distribution of M&Ms, but because no inferential statistics are taught in this course, they were unable to proceed beyond side-by-side pie charts of percentages of colours.

Third, a related problem arose when students collected data that they were able to analyse using complex methods that had been taught, but they did not recognize this. For example, a team who studied the effect of fatigue on a person's ability to play a computer car race game collected observations of lap times that were clearly not independent because the lap times were a time series. Another team who studied the effect of three different laundry powders on two different stains completely ignored the possibility of interaction between the factors they studied. A group who studied the lengths of television commercial breaks used television channel as a factor, but failed to include blocks such as program type or time of day.

Fourth, graphics chosen to display data collected were inappropriate or inelegant. The statistical computing package used throughout the semester was SPSS. However many students clearly used Excel to store and graph their data, and were content to use Excel's default style of graphics. A team who studied commercial breaks on television presented graphs of frequencies and percentages, where one or the other would do. A team who studied mobile phone plans presented a bar graph with cone-shaped bars, and two different variables (time and number of calls) displayed on the one vertical axis. A team who studied the Stroop Effect presented the times taken by 20 subjects under two different conditions as a scatter plot of time taken vs subject, rather than presenting side-by-side boxplots or, even better, a single boxplot of differences.

Fifth, outside forces sometimes acted upon the research topic and students coped with this with varying degrees of skill. One team were planning to study the differences in behaviour of two major airlines at Canberra airport; however, a few days before they were due to spend a day at the airport collecting data, one of the airlines was placed into receivership and many flights were cancelled. This team coped quite well with the disruption, adapting their data collection to focus on whatever data they could collect about the two airlines. Two other teams were collecting data on petrol prices over time, and the time they studied included the dramatic events in New York and Washington DC of September 11, 2001. Neither group thought to draw attention to this event and its impact upon their data in their research reports.

Sixth, the survey versus experiment problem persisted, despite the strong indicators given to students that their project should be based on an experiment of some sort. One group who studied television commercial breaks included a small survey of opinions about the length of commercial breaks amongst other data on length collected in a one-way design.

Seventh, some projects simply displayed a disappointing lack of imagination. One group simply downloaded some data from the Internet regarding temperature change and sea level change (United States data, not even Australian data) and presented a few bar graphs of the data.

The students' own conclusions regarding their experience in "The World of Chance" were sought in three different ways. First, students were asked to include a section in their project reports entitled "Reflections". They were asked to comment on the following questions regarding that they learnt from the process of doing their project. What went wrong? What would they do differently next time? What advice would they give future students in this subject? Some teams recognised that they had to struggle with one or more of the problems identified above: in particular, the team investigating Australian airlines "were afraid that [the collapse of Ansett] would have a significant outcome on our project and that we would have to change our idea half way through."

Some teams also recognised the importance of planning both data collection and analysis methods in advance. For example, a team using simulation to estimate probabilities associated with the children's game "Scissors, paper, stone" noted that "Future students in this subject would be advised to predict roughly in what way they would like to represent the data collected, and before collection design a template to gather the required information, as this helped us greatly during the course of this project."

Second, a survey of attitudes towards statistics was administered to students in the first and last lectures each semester. This survey is a combination of surveys described in Gal et al. (1997) and is given in Appendix 3. Detailed results of the evaluation of the course in 1999 are given in Richardson (1999). In 2001, while 58 responses to the survey were obtained in the first lecture, only 23 were obtained in the last class with only 8 pairs amongst these. It is difficult to draw strong conclusions from a small sample such as this one. Generally the change in attitude was not marked, and was split between changes which went in the desired direction and changes in the opposite direction. One explanation for this is that many of the issues raised in the survey, e.g. question 15 about the value of experimental data over anecdotal data, are not discussed explicitly during the semester and it is likely that students have not picked up any strong messages about these issues unless they have carried out independent reading of the set text.

Third, an evaluation of the subject was carried out by the University of Canberra's Centre for the Enhancement of Learning, Teaching and Scholarship. There were no specific comments regarding the project in these evaluations, beyond the message that students appreciated not having to sit an exam as part of the assessment! Overall the median rating of the course was 5 on a 7-point scale (1 = very poor, 5 = good, 7 = excellent).

CONCLUSIONS

We began the teaching of "The World of Chance" in 2001 with a desire to integrate experimental research into the curriculum, and a belief, based on the literature available, that it would be worthwhile for students and staff alike. All students were asked to carry out a piece of experimental research as their major piece of assessment, and the lectures and tutorials were structured to expose students to a variety of areas of research and experimental designs. This optimistic position was fully supported by the results achieved by the majority of the students. The experience of assessing experimental research nonetheless suggested the following changes could be made in the future to further improve the learning experience of the students.

First, a stronger emphasis on the earliest stage of research, namely answering the question "What is it that you want to know?". This should overcome the survey/experiment confusion that exists in many students' minds, and help them go from rough ideas to quantitative research questions. Second, a stronger emphasis on planning of research. This could be achieved through direct discussion of broad research plans such as the chapter headings of Robinson (2000), the online plan of Holcomb (2000) or the cyclic approach of Bishop and Talbot (2001).

Third, a greater degree of control over the project proposals and progress reports. This should help to screen out potential problems with topics or methods earlier. Other authors have commented on the importance of this, and our experience supports this conclusion. Fourth, a greater emphasis on good graphical practice. Personal experience has shown that lecturers of courses like "The World of Chance" would much rather display posters that include good graphical practice than those that don't! The importance of communication through graphical means has risen rapidly in the last decades with the wide availability of spreadsheets and other packages that produce graphs. Tufte (1983), amongst others, has attempted to raise the standard of graphical presentation, but more work remains to be done.

ACKNOWLEDGEMENTS

The author thanks Judith Ascione, Mary Hewett, Neil Porter and Janelle Wallace of the School of Mathematics and Statistics at the University of Canberra for their willingness to act as tutors in "The World of Chance". She also thanks Glenys Bishop for her editorial input and helpful discussion of the poster component of the projects while they were displayed at the University of Canberra.

REFERENCES

Bishop, G., & Talbot, M. (2001). Statistical thinking for novice researchers in the biological sciences. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 215-226). Amsterdam: International Association for Statistical Education and International Statistical Institute.

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal and J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37-51). Amsterdam: IOS Press.

Garfield, J. (1994). Beyond testing and grading. *Journal of Statistics Education,* 2*(1).* Available at http://www.amstat.org/publications/jse/v2n1/garfield.html

Holcomb, J. (2000). Steps in a statistical study. Available at www.cis.ysu.edu/~jholcomb/statsdir/Project_Guidelines_Hope.htm

Hunter, W.G. (1977). Some ideas about teaching design of experiments, with 2^5 examples of experiments conducted by students. *American Statistician, 31*, 12-17.

Lock, R.H., & Moore, T.L. (1992). Low-tech ideas for teaching statistics. In F. Gordon and S. Gordon (Eds.), *Statistics for the 21st Century* (pp. 99-108). Washington DC: Mathematical Association of America.

Mackisack, M. (1994). What is the use of experiments conducted by statistics students? *Journal of Statistics Education, 2(1).* Available at http://www.amstat.org/publications/jse/v2n1/mackisack.html

Moore, D.S. (1997). New pedagogy and new content: the case of statistics. *International Statistical Review, 65*, 123–165.

Moore, D.S. (2001). *Statistics: concepts and controversies* (5th edn). New York: W.H. Freeman.

Richardson, A.M. (1999). *The world of chance*. University of Canberra Division of Management and Technology Research Report ISE 99/114.

Robinson, G.K. (2000). *Practical strategies for experimenting*. Chichester: John Wiley & Sons.

Roberts, H.V. (1992). Student-conducted projects in introductory statistics courses. In F. Gordon and S. Gordon (Eds.), *Statistics for the 21st Century* (pp. 109-121). Washington, DC: Mathematical Association of America.

Scheaffer, R.L., Gnanadesikan, M., Watkins, A., & Witmer, J. (1996). *Activity-based statistics*. New York: Springer-Verlag.

Spurrier, J.D., Edwards, D., & Thombs, L.A. (1995). *Elementary statistics laboratory manual, MS-DOS version*. Belmont, CA: Wadsworth.

Snell, J.L., & Finn, J. (1992). A course called "Chance". *Chance, 5,* 12-16.

Tufte, E.R. (1983). *The visual display of quantitative information.* Cheshire, CO: Graphics Press.

APPENDIX 1: SCHEDULE OF LECTURE TOPICS FOR "THE WORLD OF CHANCE".

Surveys: Design, summary statistics, margins of error, ethical issues.

Experiments: Taste-tests with Coke and Pepsi, principles, one-way and two-way designs, factorial designs.

Graphics: Time series, categorical data, correlation and regression, good and bad graphics.

Probability: Streaky behaviour, the law of averages, simulations, applications to estimation.

APPENDIX 2: SCHEDULE OF TUTORIAL ACTIVITIES FOR "THE WORLD OF CHANCE".

Descriptive statistics: pulse rates

Random sampling: cars in a carpark

One-way designs: absorbency of paper towels

Blocking: walking styles

Factorial designs: paper planes
Time series: measuring traffic flow at traffic lights
Correlation: price and quality of chocolate biscuits
Regression: measurement of body parts
Calibration: distances to nearby landmarks
Estimation: population size via distance sampling
Estimation: population size via residual trap-catch
Probability: simulations to estimate probabilities

APPENDIX 3

This appendix contains the Survey of Attitudes Towards Statistics administered at the start of semester. Students were directed to circle the number (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree) that best represented their point of view about each of the 15 statements below. The version of this survey administered at the end of the semester was identical except that the first statement became "I enjoyed taking a statistics course".

I think I will enjoy taking a statistics course.
Statistical skills will make me more employable.
Because it is easy to lie with statistics, I don't trust them at all.
Understanding probability and statistics is becoming increasingly important in our society, and may become as essential as being able to add and subtract.
Statistics is not particularly useful to the typical professional.
You need to be good at mathematics to understand basic statistical concepts.
To be an intelligent consumer, it is necessary to know something about statistics.
Statements about probability (such as what are the odds of winning a lottery) seem very clear to me.
I can understand almost all of the statistical terms that I encounter in newspapers or on television.
I could easily explain how an opinion poll works.
Given the chance, I would like to learn more about probability and statistics.
I often use statistical information in forming my opinions or making decisions.
I feel insecure when I have to do statistics problems.
Statistics should be a required part of my professional training.
When buying a new car, asking a few friends about problems they have had with their car is better than consulting an owner satisfaction survey in *Choice*.