An extract from

Reading Statistics

Leigh Wood and Peter Petocz

Mathematics Study Centre,
University of Technology, Sydney.

Sydney, Australia
## Contents

**Preface**  1

**How to use *Reading Statistics***  5

**Overview of articles**  9

**Reading statistics – articles and activities (extract):**

- Gambling trends  11
- Acupuncture and stress  17
- Kiama blowhole  23
- Test cricket  37
- Conceptions of environment  43

*Articles and activities included in full publication*

- Exercise and aging
- Legal education
- Language background
- Blood pressure
- Images of hospitality
- Trumpet playing
- Panoramic radiographs
- Aerial photography
- Assessing agreement
- T’ai Chi and balance
- Predicting toxicity
- Adjustment after spinal injury
- Bird migration
- Vocational education
- Interpersonal skills
- Nursing study intention
- TV watching
- Fibrinolysis
- Sand dune plants
- Batting averages
- Employment after spinal injury
- Mandibular splint
- Anti-smoking warnings
Preface

Introduction

There has been a substantial increase in the cultural and academic diversity of commencing tertiary students in Australia over the previous decade. From 1990 to 2000 the total number of students in higher education increased by 43% (to just under 700,000), while the number of overseas students increased by a factor of four (to just under 100,000), raising the ratio to one-in-seven students (Department of Education, Science and Training, 2002). With this increase in number and cultural diversity comes an increase in academic diversity. At the same time, changes in the nature and scope of professional work are placing an increasing demand on the range of skills, linguistic and numerical, that are needed by a successful graduate. A challenge for mathematics and statistics educators is the development of a curriculum that addresses the language-related difficulties of language minority students, the numerical difficulties of students with diverse mathematical backgrounds, and enhances the learning outcomes for all students.

This book represents our approach to this challenge: it is based firmly on our research into students’ conceptions of statistics [6, 7] and the language needs of professionals in the mathematical sciences [5, 9]. The notion of helping students develop their language skills in the context of a particular subject area such as statistics is not new. However, what is new is the linking of our research into learning statistics with a theory of the relation between discourse and social factors. Moreover, this approach can be applied in any other area of science, and is not limited to statistics.

Australian government and professional bodies have recognised the importance of this area. A recent review [4] with a scope of “the 15 years from 1995–2010”, finds “unequivocal evidence that, as an economic and social instrument, advanced mathematical services relying on the mathematical sciences are critically important to Australia.” Recent government initiatives have targeted declining interest, standards and resourcing in mathematics. Similarly, universities are moving towards the integration of ‘generic skills’ within the curriculum, loosely equating statistical study with a higher-level ‘numeracy’.

Tertiary students in a wide range of areas will need statistics as a tool in their professional life. Many of them will find their first statistics course to be quite a shock, not only because it represents a completely new way of using numbers and language, but even because some of them had no idea that statistics would be a component of their studies. Future researchers and professionals in many areas will need at least a working knowledge of the meaning of statistical outputs, and some may need to develop a sophisticated understanding of statistics.

The research background

Our approach to the challenge outlined here is based on four theoretical frameworks. The first arises from a study of students’ conceptions of statistics carried out by Petocz and Reid, based on a relational approach to learning known as phenomenography [2]. They found that statistics major students have qualitatively different ways of understanding and learning in statistics, ranging from limiting to expansive views. Students who describe the most atomistic and limiting views seem only to be able to focus their attention on fragmented and unrelated components in their learning environment. Conversely, students who describe the most integrated and expansive views are able to make use of a wide range of learning approaches to further their already sophisticated understanding.
Reid and Petocz introduce the abstract notion of the ‘Professional Entity’ – a way of thinking about students’ (and teachers’) understanding of professional work. It consists of three different levels: the extrinsic technical level describes a perception that professional work is constituted as a group of technical components that can be used when the work situation demands it; the extrinsic meaning level describes a perception that professional work is about developing the meaning inherent in discipline objects (for example data, in the area of statistics); the intrinsic meaning level describes the perception that professional work is intrinsically related to a person’s own personal and professional being. The significance of the Professional Entity is that there are specific conceptions of teaching and learning associated with each of its levels: a particular way of viewing the world of professional statistics corresponds to a particular approach to teaching and learning.

Moreover, teachers can help students move beyond the more limiting conceptions towards the broader conceptions. A teachers’ approach to their teaching and the sort of learning environment that they set up in their classes can encourage students who identify with the lower, fragmented levels to engage with their learning at a higher level. However, this can also work the other way if a teacher sets students tasks that are best carried out using the more fragmented conceptions of learning.

The second framework on which our work is based is the theory of critical discourse analysis, particularly the ideas of Fairclough [1]. Critical discourse analysis, “studies the way social power abuse, dominance and inequality are enacted, reproduced and resisted by text and talk in the social and political context … critical discourse analysts take explicit position, and thus want to understand, expose and ultimately to resist social inequality” [8]. Within the overall aims of critical discourse analysis, there is a diversity of theoretical frameworks. In the British context, the writings of Fairclough focus on various dimensions of power. He (and others) argue that there has been a large-scale restructuring of employment with major implications for the linguistic demands of work. The modern workplace is requiring more interpersonal communication skills, as the emphasis shifts from isolated workers to teams. A transformation is also occurring within the professions, where clients are no longer expected to adapt to the professional discourse: rather, the professionals are adapting to the language needs of their clients. In our experience, the situation seems no different in the Australian setting.

We have found it useful to consider Fairclough’s three dimensions of critical discourse analysis:

(a) description of the text (with ‘text’ interpreted widely, and including spoken words)

(b) interaction with the text, involving processes of producing and interpreting the text

(c) explanation of the interaction with the text, by referring to its social and discipline context

This defines three levels at which students can work with text. We have used these levels in designing our learning materials.

The third research framework that informs the design of our learning materials is the increasing emphasis on graduate profiles, professional competencies and generic skills. Universities and professional societies are agreed on the importance of fundamental professional skills, or “graduate competencies”, that help to prepare students for an increasingly uncertain future. The development of academic and professional discourse skills (“communication skills”) accords well with the capabilities needed by all graduates, but sometimes lacking in graduates of quantitative disciplines. The siting of the language of the professions within informal language
when dealing with clients, and within formal language when dealing with colleagues, places extra importance on the diverse communication skills needed by graduates.

The fourth and final framework that we have used connects the ideas of equity, equal opportunity and non-discrimination which have become an explicit part of teaching and learning at university. For many lecturers, these are principles to which they have always adhered. However, the changes have been in the legislative and social acceptance of these principles. It is widely accepted that universities are no longer only for an elite, and importance is given to teaching and learning initiatives that are inclusive and that assist students to reach their full academic potential whatever their background. For example, an extract from the University of Technology, Sydney’s Equity Plan 2000–2003 states: “UTS is committed to the right of all students to study and access services in a university environment which is equitable, free from discrimination and harassment, and in which everybody is respected and treated fairly. A central objective of the UTS Mission Statement is to improve educational provision for students from a diversity of backgrounds.” We believe that good curriculum design that develops academic language skills and statistical skills meets the requirements of such an Equity Plan.

**Curriculum design**

Our focus in this book is on enhancing language, numeracy and communication skills in culturally and academically diverse student cohorts, using the results of our research on student learning and communication needs. We have designed a flexible curriculum that uses real sources (published journal articles, conference papers, academic subject notes) from the disciplines of the participating students. In *Reading Statistics*, we have examples from the areas of tourism and sport, environmental science, medicine and the health sciences, music, physical sciences, engineering, education and orthodontics. In each example, the source material is the basis of a series of questions focusing on the language of the article and the statistical aspects of the study. Although our questions do not neglect the lower-level skills, they are designed to encourage students to use the highest level of the taxonomy of discourse, and to view the subject material at the most expansive and holistic level statistically. These teaching and learning experiences can be adapted to any situations where students need to interpret and understand statistical material – the typical situation that many students in “servicing statistics” courses will find themselves. They are encouraged to develop critical reading skills and understand the statistical content of research papers. They are also supported in making connections between the topics of the articles and their own personal and professional lives.

Previously, we have been involved in writing a wide range of learning materials that aim to extend students’ ideas about the nature of statistics and mathematics, and to develop their communication skills in these areas. Such materials include textbooks, video packages and laboratory materials. In this introduction to *Reading Statistics*, we are making explicit the theoretical background on which our approach is based, showing how we use our and others’ research to prepare materials that will enhance the skills of our diverse groups of students. We are also allowing for the possibility of designing a whole statistics course that is built around the use of a series of readings in applications of statistics.

It is interesting that when we were preparing questions at the highest levels of both the hierarchy of conceptions of statistics and the taxonomy of discourse, we often came up with very similar ideas for questions that focused on language and those that focused on statistics. Although there is no evidence that the theories underlying the two approaches are related, and the fact that they each have three levels is simply coincidental, it seems that they can result in the same types of high-level questions. Looking at the descriptions of the highest levels of discourse and statistical
conceptions, this is not surprising since they both focus on placing the student in the context of their discipline and their personal connections with society. In this context, it is interesting to note that early studies using the phenomenographic method [3] focused on different ways of understanding written text, and led to the identification of “deep” and “surface” approaches to learning, described by Marton and Booth [2].

References


An expanded version of this paper can be found at http://science.uniserve.edu.au/pubs/procs/
How to use *Reading Statistics*

The 28 articles in this book have all been written by researchers who make use of statistics. Every article is accompanied by a set of activities which cover aspects of communication and statistics. These become progressively more difficult as you move through the articles. The first section (articles 1-6) has articles that use elementary descriptive statistical techniques; the second section (articles 7-19) has articles using introductory techniques of statistical inference; and the last section (articles 20-28) uses advanced statistical techniques. The language difficulty varies depending on the style of the authors and the type of publication.

*Reading Statistics* is designed to help you develop skills in statistics and communication and to be able to use these skills in your discipline area. Each of the sources or readings is preceded by a set of questions that direct your learning. The questions are organised under two main headings, ‘Preliminary reading’ and ‘In-depth reading’. The introductory questions are the same for all the readings, and these are followed by questions and activities focusing on specific aspects of the source document. In general, the introductory questions for each reading cover the descriptive and technical aspects of the text and the statistical analysis. The specific questions address higher order skills: they ask you to generate text or analyses, and place them within a wider context, in your own discipline.

The readings are taken from a wide range of disciplines. While it is interesting and instructive to study an article from your own area, you will still learn by reading and answering the questions for other articles. In these days of interdisciplinary knowledge, being able to read in other disciplines is an important skill.

Don’t limit yourself to the articles in this book. One of the best ways to improve your own research is to read widely and to think about why the authors you read chose to do things the way they did. As a note of caution, it is good also to develop skills of criticism. Not everything you read is correct, even if it is published in a good peer-reviewed journal. Research is a human activity and is sometimes not free of errors in analysis or reporting.

**General questions**

The questions for each article start in the same way. We have a series of pre-reading questions to help you develop organised reading skills, and then we move to detailed reading. Pre-reading activities will save time and help you get the most out of your reading. Learning to skim read is particularly important because of the large volume of material that you will have to read during your working life. Skim-reading techniques help you deal with large amounts of material in a short time.

Try to follow the same procedure for each article. Note the author and the type of publication, search for all signposts provided by the author, skim quickly through the whole article. Pay particular attention to the title, abstract, headings, diagrams and the first and last paragraphs. Get as good an idea as you can of what it is about in the time limit. Developing these pre-reading habits enables you to get a good idea of what an article is about before you start reading it. This saves you time either way: if it doesn’t look interesting you can abandon it without a thorough reading, and if it does look interesting you will have some idea of what it is about and what its aims are. This will make reading easier, more fun and faster.

After doing the major reading you are asked to answer questions about the aim, audience and content of the article. This is to make sure that you have understood the main points.
The statistical questions also start the same way for each article. We want you to develop a standard set of questions that you ask yourself every time you read an article. In time you will ask yourself these same questions whenever you write your own articles. The following activities will help you identify the reasons why the author has used the particular research design and statistical techniques.

1. Identify the research question: Write down the research question. How is it presented? Why has the author presented it this way?

2. Identify the research methodology used: Write down the research methodology used. Is it an observational study or an experimental study, or neither? What sampling techniques are used (if any)?

3. How does the author deal with data: What data are given in the article? How are the data presented or described?

4. What statistical techniques are used and why: List the statistical techniques used or referred to. Give a reason for using each of them.

**Specific questions: What to look for**

**Aim and audience**

Academic writers, like all other writers, write in different ways depending on who they are writing for and what their aim is in writing for a particular audience. Several articles, such as Rose, Warne and Lim (1997), Murphy et al. (1997) and Athanasou (1999), are reprinted from academic journals intended to be read by researchers. Even if they are too difficult for you to understand fully you should examine the way they are constructed and the way the statistics and natural language interact. You should also notice differences between articles written for general and academic readers, especially regarding the use of abstracts or summaries. Note also whether the intended audience is Australian or international. The statistical techniques used will also give you clues about the aim of the whole text.

Articles on topics which will interest a broader section of the community, say holders of university degrees and diplomas, are printed in science magazines such as *New Scientist*. These do not follow the same format as academic journal articles, although they often report or comment on research which has been recently presented in that form.

This collection includes articles written for presentation at conferences and others that are written as teaching materials. Textbook writers and lecturers have a different approach. Whilst they too have to bear in mind what is already familiar to their audience, their aim is to extend the comfort zone of their readers by filling out background knowledge. Consider Irish (1999) and Garlin (1997) as examples of writing suitable for students.

Each article has its clues as to the type of publication from the style that it is written in. You can learn to pick up these clues and develop your ability to write in the different genres.

**Content**

Each article has one or two questions asking you to define or describe the meaning of words or abbreviations. Researchers use technical terms which are not used outside their field. It is important for writers to define technical terms and for readers to recognise and understand definitions. Semi-technical vocabulary – common words used in a specialised sense – is actually more problematic because if a writer does not
define such terms and a reader is not aware of their special meaning, then the reader will carry on reading with quite the wrong sort of understanding.

**Statistical analysis**

This section requires you to look in depth at the statistical work in the article and beyond. We often invite you to work with a partner or in a group. For many questions there is no correct answer. We require you to make judgements and to think about issues to do with the statistical aspects, including the ethical considerations of the research. Many of the published articles have errors in the statistical analysis, or problems with the presentation of data. We want you to become critical consumers of research. This should help you with your own research.

Sometimes, we ask you to find further information or to reanalyse data to draw your own conclusions.

**Activity**

This section concentrates on the communication aspects of the article. We often ask you to rewrite the article in a different form: summary, newspaper article, talk and so on. We ask you to practise writing for different audiences. You will also be encouraged to make links with your own discipline and research area.

**Making use of the library and the internet**

If you are particularly interested in a topic you have read about or need to do research for a presentation, you will need to use a library or the internet. Several activities suggest going to the library or web to search for articles or further information. If you are not familiar with your library you will need to seek the advice of the librarian. The UTS library also has packages on use of the library databases and search engines for internet use.

**Making use of the statistical packages**

Carrying out a statistical analysis will require a statistical package. Standard spreadsheet packages, such as Excel, and statistical packages such as Minitab, will perform statistical analyses. For more specialised analyses, professional statistical packages such as SPSS, S-plus or SAS may be required. UTS has site licences for Excel, Minitab and SPSS. Training programs are available at the Mathematics Study Centre.

**Finally**

We hope that you enjoy using this book and the various articles and exercises that it contains. We will be very pleased if it helps you develop your ability to read such articles, write or speak about them, and think critically about the statistical methods used. We will be even happier if you are able to use your expertise in your own professional work.

If you have any comments about the book, we would like to hear from you at Leigh.Wood@uts.edu.au or Peter.Petocz@uts.edu.au
## Overview of Articles

(a pre-reading routine is practiced for all articles)  
(an in-depth routine is practiced for all articles)

<table>
<thead>
<tr>
<th>Article</th>
<th>author</th>
<th>date</th>
<th>statistics activity</th>
<th>language activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introductory articles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Gambling trends</td>
<td>Veal</td>
<td>1997</td>
<td>Objective presentation of data</td>
<td>Informal language</td>
</tr>
<tr>
<td>2 Exercise and aging</td>
<td>Daley &amp; Spinks</td>
<td>2000</td>
<td>Graphical presentation of data</td>
<td>Writing for different audiences – newspaper</td>
</tr>
<tr>
<td>3 Legal education</td>
<td>Stuhmcke, Booth &amp; Melano</td>
<td>1997</td>
<td>Response bias</td>
<td>Designing a survey</td>
</tr>
<tr>
<td>4 Language background</td>
<td>Logan &amp; Hazel</td>
<td>1999</td>
<td>Presentation of data – graphs and tables</td>
<td>Referencing</td>
</tr>
<tr>
<td>5 Blood pressure</td>
<td>Brown</td>
<td>1990</td>
<td>Measurement error</td>
<td>Finding original source – summarising</td>
</tr>
<tr>
<td>6 Images of hospitality</td>
<td>Barron</td>
<td>1997</td>
<td>Ethics</td>
<td>Review of article</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Graphical presentation</td>
<td>Survey design</td>
</tr>
<tr>
<td><strong>Articles using inference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Trumpet playing</td>
<td>Hunsaker &amp; Ramsey</td>
<td>1995</td>
<td>t-tests</td>
<td>Popular versus research literature</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Writing hypotheses</td>
<td></td>
</tr>
<tr>
<td>8 Panoramic radiographs</td>
<td>Stramota et al.</td>
<td>2000</td>
<td>Interpreting graphs</td>
<td>Internationalisation of research</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Multiple comparisons</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Significance</td>
<td></td>
</tr>
<tr>
<td>9 Aerial photography</td>
<td>Bayliss &amp; Yeomans</td>
<td>1990</td>
<td>Response variables</td>
<td>Critical review</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ANOVA</td>
<td></td>
</tr>
<tr>
<td>10 Assessing agreement</td>
<td>Bland &amp; Altman</td>
<td>1986</td>
<td>Correlation</td>
<td>Oral presentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Measuring agreement</td>
<td></td>
</tr>
<tr>
<td>11 T’ai Chi and balance</td>
<td>Hain et al.</td>
<td>1999</td>
<td>Measurement</td>
<td>Style</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Significance</td>
<td>Drawing conclusions from data</td>
</tr>
<tr>
<td>12 Predicting toxicity</td>
<td>Rose, Warne &amp; Lim</td>
<td>1997</td>
<td>Linear regression</td>
<td>Summary</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Outliers</td>
<td>Writing for different audiences – high school students</td>
</tr>
<tr>
<td>13 Adjustment after spinal injury</td>
<td>Craig, Hancock &amp; Dickson</td>
<td>1999</td>
<td>Experimental design</td>
<td>Verb usage</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chi-squared tests</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ethics</td>
<td></td>
</tr>
<tr>
<td>Article</td>
<td>author</td>
<td>date</td>
<td>statistics activity</td>
<td>language activity</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>------</td>
<td>---------------------</td>
<td>------------------</td>
</tr>
</tbody>
</table>
| 14 Acupuncture and stress | Garvey et al. | 1997 | Statistical significance  
Selection of sample  
Academic searches | Passive and active voice |
| 15 Bird migration | Witschko, Witschko & Munro | 1997 | Circular statistics  
Experimental design | Comparing value of different research areas |
| 16 Kiama blowhole | Irish | 1999 | Poission processes  
Stochastic models | Following procedures – modelling |
| 17 Vocational education | Athanasou | 1999 | Multiple regression  
Interpreting results | Word choice  
Survey design and analysis |
| 18 Test cricket | Croucher | 1985? | Reanalysing data  
Sports statistics | Recording of statistics |
| 19 Interpersonal skills | Garavan | 1997 | Interpreting tables  
T-tests  
Publications containing errors | Ethics |

**Advanced statistical techniques**

<table>
<thead>
<tr>
<th>Article</th>
<th>author</th>
<th>date</th>
<th>statistics activity</th>
<th>language activity</th>
</tr>
</thead>
</table>
| 20 Nursing study intention | Tang et al | 1998 | Sampling design  
Calculating chi-squared and confidence intervals | Designing research questions  
Critical analysis |
| 21 TV watching | Garlin | 1997 | Factor analysis | Debate |
| 22 Fibrinolysis | Alman-Farinelli et al. | 1999 | Experimental design  
Power  
Publication of non-significant results | Referencing measurement instruments |
| 23 Sand dune plants | Buckney & Morrison | 1992 | Data collection  
Multidimensional scaling | Writing for different audiences – environmental magazine |
| 24 Batting averages | Danaher | 1998 | Maximum likelihood ratio  
Nature of statistics | Academic style |
| 25 Employment after spinal injury | Murphy et al | 1997 | Multiple regression analysis  
Discriminant function analysis  
Experimental design | Ethics |
| 26 Mandibular splint | Mehta et al | 2001 | Crossover design  
Ethics  
Baseline comparisons | Writing for different audiences – personal |
| 27 Anti-smoking warnings | Motuk | 2000 | Experimental design  
Conjoint analysis | Writing for different audiences – science magazine |
| 28 Conceptions of environment | Loughland et al | 2001 | Logistic regression  
Interpretation of findings | Designing a research plan |
Gambling trends

Preliminary reading

Do the following pre-reading activities before you settle down to read carefully. This will save time and help you get the most out of your reading.

1 Identify the background of the article:
   (a) Who is the author?
   (b) What is the title?
   (c) What is the name of the publication it appears in?
   (d) Write out a reference for this article in one of the approved styles.

2 Skim through the article:
   (a) Set a timer for 5 minutes and skim the whole article. Get as good an idea as you can of what it is about in the time limit.
   (b) Summarise the article in 3-5 dot points.

Now read the whole article carefully and answer the following questions. They will help you identify the reasons why the author has used the particular research design and statistical techniques.

3 Identify the research question:
   (a) Write down the research question.
   (b) How is it presented?
   (c) Why has the author presented it this way?

4 Identify the research methodology used:
   (a) Write down the research methodology used.
   (b) Is it an observational study or an experimental study, or neither?
   (c) What sampling techniques are used (if any)?

5 How does the author deal with data?
   (a) What data are given in the article?
   (b) How are the data presented or described?

6 What statistical techniques are used and why?
   (a) List the statistical techniques used or referred to.
   (b) Give a reason for using each of them.
In-depth reading

The preliminary questions considered the techniques used in the original research and then in writing the article. The following set of questions and activities are to help you with thinking about the text and statistical design in more depth.

1 Aim and audience

(a) What is the author’s major aim in writing the article?

(b) What clues do the statistical techniques give you about the aim?

(c) What audience is the author writing for? Give reasons.

2 Content

(a) Figure 1 shows the various components of gambling expenditure in Australia. In 1996, which component was the largest, and which component was growing the fastest?

(b) Why is NSW the most “mature” gambling market in Australia?

3 Statistical analysis

(a) In his paper, Veal presents some of the information in graphical form. Select one of the figures from the paper. Write a sentence commenting on how successful it is in visually presenting the information. Suggest three improvements to the graph.

(b) Why do you think published data are always out of date? Search the web for current information on gambling in Australia and use this to update one of the figures 1–5 in the paper. What will you do about the fact that the data are presented “in 1996 dollars”? Does the extra information follow the trends identified by Veal?

(c) Do you think statistical presentations of data should be objective? Is Veal’s paper objective in its presentation of gambling trends, in other words, does the author approve of gambling? Give reasons.

4 Activity

This is written in an informal style. List 5 reasons, quoting from the article, why you consider this to be an informal paper.
GAMBLING TRENDS IN THE 1990s

Tony Veal
School of Leisure and Tourism
University of Technology, Sydney

The latest statistics on gambling expenditure in Australia have just been published. What do they tell us about this quintessentially Australian pastime? The dimensions of the industry are impressive, revealing it to be one of Australia’s biggest service industries. Figure 1 shows that, over the last 25 years, real expenditure (net of winnings) has grown from $2.75 billion to almost $10 billion (at 1996 prices). During the 1990s the rate of growth has accelerated to the status of a boom. The boom conditions, however, have not been enjoyed uniformly by all sectors of the gambling industry. Racing has been static, while lotto and lotteries have grown steadily – but the bulk of the growth has been generated by poker (or gaming) machines and casinos.

Figure 1: Gambling Expenditure in Australia: 1973–1996 (in 1996 dollars)

Throughout the 1970s and 1990s, virtually all poker machine expenditure was accounted for by registered clubs in New South Wales, but since the early 1990s these machines began to be licensed in other states, including Queensland, Victoria and the ACT, so that they now account for $4.6 billion, or almost half of all gambling expenditure.

Casinos started from a small base, with the Wrest Point Casino in Tasmania and Diamond Beach in Darwin in the 1970s, but have made a late run with the opening of bigger and bigger casinos in more central locations, culminating in the Melbourne and Sydney temporary casinos in 1995/96 (see Figure 2). Casinos now account for expenditure of $1.7 billion a year – and the figures do not yet include the permanent casino in Melbourne, which opened a couple of months ago, or the permanent casino in Sydney, due to be opened in December.
Aggregate sums, however, tell only part of the story. Over the last 25 years, incomes and the level of population have grown, so we might expect gambling expenditure to have increased. Figure 3 and 4, however, show that gambling expenditure per head of population has more than doubled in this period, to more than $700 per annum, and expenditure as a proportion of disposable household income has also more than doubled, to around three percent. These figures should, however, be treated with a certain amount of caution, because they do not take account of the fact that some of the recorded gambling expenditure is from tourists, both domestic and international.

![Figure 3: Gambling Expenditure per Head](source: TGC, 1997)
Because of its 40-year association with poker machines, New South Wales is the most mature gambling market in Australia. Figure 5 clearly demonstrates the dominance of New South Wales in the Australian gambling scene, accounting for $3.75 billion, or almost 40 percent of all gambling expenditure. However, other states, particularly Victoria and Queensland, are rapidly catching up.

The 1990s have seen a massive increase in gambling opportunities around Australia, with a number of state and territory governments legalising poker machines and casinos being licensed to operate in every capital city and a number of tourism centres. Gambling has been a significant feature of Australian culture since 1788, but why this explosion of activity in recent years? For an explanation, we need look no further than the following figures: in 1972–73, total state and territory government revenue from taxes on gambling amounted to $168 million (at 1996 prices); by 1995–96 this had increased to $3.2 billion. Enough said.
Notes

1. This paper was submitted to *Australian Leisure Management* (1997).

2. All data on expenditure and turnover in this paper are taken from *Australian Gambling Statistics 1972-73 to 1995-96*, compiled and published by the Tasmanian Gaming Commission on behalf of all state and territory governments.

3. Gambling *expenditure* is the total net losses incurred by gamblers. The incomes of gambling businesses and government taxes are derived from this expenditure. The total amount staked is referred to as ‘turnover’. For some forms of gambling, such as poker machines, turnover is known precisely; in other cases, such as table games in casinos, turnover is estimated. In Australia in 1995/96, gambling turnover was estimated to have been $72.9 billion, or over ten times the level of expenditure.
Acupuncture and stress

Preliminary reading

Do the following pre-reading activities before you settle down to read carefully. This will save time and help you get the most out of your reading.

1 Identify the background of the article:

(a) Can we tell who are the authors of this paper? Who is the presenter and who are the co-investigators?

(b) What is the title?

(c) What is the name of the publication it appears in?

(d) Write out a reference for this article in one of the approved styles.

2 Skim through the article:

(a) Set a timer for 5 minutes and skim the whole article. Get as good an idea as you can of what it is about in the time limit.

(b) Summarise the article in 3-5 dot points.

Now read the whole article carefully and answer the following questions. They will help you identify the reasons why the authors have used the particular research design and statistical techniques.

3 Identify the research question:

(a) Write down the research question.

(b) How is it presented?

(c) Why have the authors presented it this way?

4 Identify the research methodology used:

(a) Write down the research methodology used.

(b) Is it an observational study or an experimental study, or neither?

(c) What sampling techniques are used (if any)?

5 How do the authors deal with data?

(a) What data are given in the article?

(b) How are the data presented or described?

6 What statistical techniques are used and why?

(a) List the statistical techniques used or referred to.

(b) Give a reason for using each of them.
In-depth reading

The preliminary questions considered the techniques used in the original research and then in writing the article. The following set of questions and activities are to help you with thinking about the text and statistical design in more depth.

a. **Aim and audience**

(a) What is the authors’ major aim in writing the article?

(b) What clues do the statistical techniques give you about the aim?

(c) What audience are the authors writing for? Give reasons.

2 **Content**

(a) What is clinical significance?

(b) What is a sham group? Explain the difference between a sham group and a non-treatment control group.

3 **Statistical analysis**

(a) The pre to post differences for the treatment and sham groups are "very close to being significant with the p value = 0.06". However, "The difference between the treatment and sham groups was not statistically significant." Write one or two paragraphs to explain the meaning of these statements to a fellow student who has not studied any statistics at all.

(b) Who were the subjects for this trial, and how were they selected? At what time of the year was the research conducted? How might these features introduce bias into the results? How many subjects were there in each group? The idea of a non-treatment control group seems to have been suggested by the results coming from stage 2 of the trial (i.e. the second group of acupuncture and sham groups). The pilot for this was ‘hurriedly’ put together at this time. What influence could the timing of this have on the results? Could you suggest a reason for the similarity in the stress values of the control pre-group and the treatment and sham post-groups?

(c) The conference paper is a report on a study in progress. Many questions are raised by the paper. Search the journals that are referred to in the publication to answer some of the questions raised and to find an end to the story! In addition to the papers listed as being “in press”, it is likely that further papers reporting on the whole project were also published in the same journals.

4 **Activity**

The first three paragraphs make use of the passive voice, for example: “Results for the treatment and sham acupuncture groups will be presented”. Rewrite the paragraphs using active voice. Why do you think the author used the passive voice in these paragraphs? Do you think that active or passive voice is better in this context? Give reasons.
Acupuncture and Stress Clinical Trial Results

Presenter: Mary Garvey
Associate Lecturer, College of Traditional Chinese Medicine
University of Technology, Sydney

Co-Investigators: Mary Garvey
Carole Rogers
Damien Ryan
Yang Congxing
Chris Zaslavski
Zhang Shiping

The UTS College of Traditional Chinese Medicine has been conducting a clinical research project for the last two years. The overall goal of the trial has been to test the hypothesis that traditional Chinese acupuncture (TCA) is effective in the treatment of stress.

This paper is a continuation of the reporting of the research which has been done at the two national conferences held by the University of Technology, Sydney, and the Victorian University of Technology in 1995 and 1996. Since the research can be conducted only during the mid-year non-teaching periods it has been carried out in stages and reporting so far has concentrated on design and development issues as well as reporting preliminary results. A third group of subjects has been tested this year. This group is a non-treatment control group, the inclusion of which is a response to the preliminary data.

Results for the treatment and sham acupuncture groups will be presented. The rationale for the inclusion of a non-treatment control group (and trends from a small pilot group) will be discussed. There has been a good deal of interest in the clinical trial so proposed publication and those currently in press will be advised.

INTRODUCTION

The UTS College of Traditional Chinese Medicine has been conducting a clinical research project over the last two years. The overall goal of the trial has been to test the hypothesis that traditional Chinese acupuncture (TCA) is effective in the treatment of stress. This paper is a continuation of the reporting of the research which has been done at the two national conferences held by the University of Technology, Sydney, and the Victorian University of Technology in 1995 and 1996. Since the research can be conducted only during the mid-year non-teaching periods, it has been carried out in stages and reporting so far has concentrated on design and development issues (1995) as well as reporting preliminary results (1996).
The study began with a standard single-blind repeated-measures experimental design using two subject groups: a treatment group and a sham acupuncture group to control for the placebo effect. This year, a third group of subjects is currently undergoing testing. This group is a non-treatment control group, the inclusion of which is a response to the preliminary data reported at the conference last year.

RESULTS

Base-line measurements were taken for treatment and sham groups before the commencement of the study. The post-trial measurements were taken the day after their final session of treatment. This year, results include those from the 1996 subject groups which have now been added to the 1995 results.

Some of the indicators show that acupuncture, both real and sham, has produced a positive effect, particularly with respect to the measurements for psychological indicators and blood pressure. The most notable change again was in the General Health Questionnaire readings, that is, in the participants’ own perceptions of their individual stress levels.

The graph shows the GHQ variations – given our time constraints today the GHQ results will serve to illustrate a number of general points about the study’s results. Figure 1 gives a pictorial representation of the GHQ’s comparative variations: the means of the GHQ readings are plotted for the pre and post-trial phases. The first two bars (on the left) are for the treatment and sham groups respectively. A reading above “8” shows that our subjects’ perceptions of their own stress levels were quite high and that the two groups were comparable in this regard. Figure 1 shows that both groups improved over time. The improvement over time came very close to being significant with the p value = .06. The difference between the treatment and sham groups was not statistically significant.

**Figure 1: GHQ – perceptions of health**

<table>
<thead>
<tr>
<th>GHQ</th>
<th>Pre-</th>
<th>Post-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>9.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Sham</td>
<td>10.2</td>
<td>3.5</td>
</tr>
<tr>
<td>Control</td>
<td>3.2</td>
<td>5.7</td>
</tr>
</tbody>
</table>
The third bar in Figure 1 is a pilot group for a second control group – a passive, or non-treatment, control group. It was decided to include this third group to provide data to measure the variance between the treatment and non-treatment (non-active) control – data from a non-treatment group should minimise the influence of sham needling and the therapeutic encounter.

We decided to pilot the non-treatment control idea whilst running Stage 2 of the trial and hurriedly put together a pilot of 12 subjects. These were not selected via our normal route which was to advertise for people who felt they were stressed. To pilot the non-treatment control group in 1996 we just enlisted whoever was available for testing at the time and ran the measurements approximately with the major groups’ testing regime. As you can see the first set of GHQ readings for this group do not match their treatment and sham group colleagues – we had not selected according to the trial’s criteria, and this meant that the perceived stress levels for this group are not comparable with the other two subject groups. However, they clearly do not improve over time which is the kind of comparison needed to demonstrate the improvements that do occur with treatment. This is why we are running a third stage now so as to include data for a third (the non-treatment control) group. We have advertised and selected according to the trial criteria and currently have another 15 subjects for this group.

After adding the Stage 2 subject data to Stage 1 data for the treatment and sham groups, the only other measures which give clear indications of a treatment effect are the diastolic and systolic blood pressure readings. Although the changes over time do demonstrate an effect, the differences between the treatment and sham groups once again were not statistically significant. Nevertheless, there is a notable clinical significance. This has been represented in the following table by calculating the percentage reduction of measures (clinically, an improvement in measures) over time for the blood pressure readings for the treatment and sham groups.

The table shows that the improvement for the treatment group is greater than for the sham group. From results from the pilot non-treatment group, we anticipate that acupuncture, both treatment and sham, will be shown to have a greater clinical effect than no treatment.

<table>
<thead>
<tr>
<th>BLOOD PRESSURE</th>
<th>Treatment gp</th>
<th>Sham gp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diastolic</td>
<td>3% improvement</td>
<td>2.7% improvement</td>
</tr>
<tr>
<td>Systolic</td>
<td>4% improvement</td>
<td>2.5% improvement</td>
</tr>
</tbody>
</table>

Lewith and Aldridge (1993, p.96) say that there has been an over-reliance on statistical probability measures, and that these have often been employed inappropriately by clinical researchers leading to an impoverished interpretation of results. Discussing these results in terms of clinical significance rather than statistical significance does make more sense given the small sample sizes in the study. Stux and Pomeranz (1991, p.22) suggest that to determine statistical significance, subject numbers would have to be in the vicinity of 122, that is, 122 participants in each of the subject groups.
SUMMARY

Our analysis of the data found that the most significant change in stress indicators was in terms of just three of the measures undertaken. The participants’ self-perception of their stress level as shown in the pre and post-test GHQ readings (the subjective measure) was one of these, and the other two were the diastolic and systolic blood pressure readings (two of the objective physiological measures).

Participants were included in the study on the basis of their own perception as being stressed. They were not selected on the basis of abnormal physiological factors, so it stands to reason that changes in the GHQ measures were going to be the most significant stress-indicator change. The group’s physiological indicators in fact lay well within normal ranges, which underlines the clinical significance of the changes recorded over time given that just four treatment sessions were offered.

PUBLICATIONS

There has been a good deal of interest in the clinical trial so proposed publications and those currently in press are as follows:

1995 Conference proceedings (UTS) – Design and development issues [available from the College of TCM office at 645 Harris Street, Ultimo]

1996 Conference proceedings (VUT) – Preliminary results (from the 1995 trial), and development of Stage 2 (the 1996 trial) [in press - VUT]


“Methodological issues related to the enhancement of Credibility of Sham Acupuncture when used as a control treatment in clinical trials”, The Journal of Alternative and Complementary Medicine [in press]

Proposed:
- Data analysis and discussion of results (when Stage 3 data become available).

This will be the next and the final stage of the clinical trial.

REFERENCES


Kiama blowhole

**Preliminary reading**

Do the following pre-reading activities before you settle down to read carefully. This will save time and help you get the most out of your reading.

1. **Identify the background of the article:**
   (a) Who is the author?
   (b) What is the title?
   (c) What is the name of the publication it appears in?
   (d) Write out a reference for this article in one of the approved styles.

2. **Skim through the article:**
   (a) Set a timer for 5 minutes and skim the whole article. Get as good an idea as you can of what it is about in the time limit.
   (b) Summarise the article in 3-5 dot points.

Now read the whole article carefully and answer the following questions. They will help you identify the reasons why the author has used the particular research design and statistical techniques.

3. **Identify the research question:**
   (a) Write down the research question.
   (b) How is it presented?
   (c) Why has the author presented it this way?

4. **Identify the research methodology used:**
   (a) Write down the research methodology used.
   (b) Is it an observational study or an experimental study, or neither?
   (c) What sampling techniques are used (if any)?

5. **How does the author deal with data?**
   (a) What data are given in the article?
   (b) How are the data presented or described?

6. **What statistical techniques are used and why?**
   (a) List the statistical techniques used or referred to.
   (b) Give a reason for using each of them.
In-depth reading

The preliminary questions considered the techniques used in the original research and then in writing the article. The following set of questions and activities are to help you with thinking about the text and statistical design in more depth.

a. Aim and audience

(a) What is the author's major aim in writing the article?

(b) What clues do the statistical techniques give you about the aim?

(c) What audience is the author writing for? Give reasons.

2 Content

(a) What is the principle of parsimony of parameters?

(b) What are the sources of error in the measurement?

(c) You will have noticed that this article is a teaching module. List 5 language features of this article that alert you to this.

3 Statistical analysis

(a) Work with a partner on these questions. Irish considers five questions in order to check whether a Poisson model is appropriate for the blowhole eruptions. Questions 1 and 2 are preliminary questions. What is the difference between a deterministic and a stochastic model? Why does the author believe that a stochastic model is appropriate here? What features of the measuring process need to be considered? What does the author conclude about the quality of the measurements? The next two questions are technical. What does it mean to say that the intervals between eruptions were stationary, and why is this important? What does it mean to say that the intervals between successive eruptions were uncorrelated, and why is this important? How does the author establish that the intervals were stationary and uncorrelated?

(b) Find out information about Poisson processes and their properties. Use this to check Irish's conclusion that the blowhole eruptions form a Poisson process in several different ways. You could check graphically using a probability plot with an appropriate statistics package. You can enter the interval data by hand or copy them from http://www.statsci.org/data/oz/kiama.html.

4 Activity

Work in a group of three or four people for this question. Investigate a situation in your professional area where events seem to be happening “at random”. Take measurements of times at which the events occur. Use a method similar to Irish's to check whether the times of the events can be modelled by a Poisson process.
CHOOSING A STATISTICAL MODEL:
IS THE KIAMA BLOWHOLE A POISSON PROCESS?

By James Irish

The ocean swell\textsuperscript{1} produces spectacular eruptions of water through a hole in the cliff at Kiama, about 120km south of Sydney, known as the Blowhole.

The times at which 65 successive eruptions occurred from 1340 hours on 12 July 1998 was observed using a digital watch. These are shown in the data table in the Appendix, together with the intervals between eruptions.

These intervals are highly variable, and there doesn’t appear to be an obvious pattern. A stochastic model may therefore be appropriate to describe the complex physical processes which result in eruptions.

The simplest such model is to assume that the eruptions occur randomly. In other words, the delay until the next eruption is independent of the intervals between recent eruptions. The principle of \textit{parsimony}\textsuperscript{2} of parameters states that we should select the simplest possible model consistent with the data and the physical behaviour of the system. (Why?)

\textit{If} the intervals between successive eruptions are independent, \textit{ie} random, then a theorem of mathematical statistics tells us that these time intervals have an exponential distribution; the occurrence of eruptions is a Poisson process, because the number of eruptions in a period of time, such as one minute or fifteen minutes, follows the Poisson distribution.

If a Poisson model is inappropriate, then we’ll need to consider what further complexity is warranted to arrive at an adequate model. The data may demonstrate that a simple model isn’t adequate but be insufficient to determine what form of model is appropriate; further investigation, including the possibility that more data should be obtained, may be indicated. Or we may be forced to conclude that we don’t have the time or the resources to properly answer the question; any model we select may be highly uncertain (as well as the parameters we estimate). Indeed, we may be truly ignorant of the appropriate way to model the phenomenon. If that is the case (even if we pretend otherwise), then any decisions we make are mere opinions, based either on intuition or experience, possibly correct and even valuable. But we ought to be honest enough in such circumstances to acknowledge that we may be wrong, that we have a very limited basis for making inferences and decisions, and that the opinions of others may be more valuable. \textit{Engineers are human}; we are allowed to make mistakes, and

\textsuperscript{1} \textit{Swell}: grow or cause to grow bigger or louder or more intense; expand; increase in force or intensity.
\textsuperscript{2} \textit{Parsimony}: carefulness in the use of money or other resources.
these mistakes will be tolerated if we act with humility, by acknowledging when we don’t really have sound knowledge.

To test the hypothesis that the occurrence of eruptions followed a Poisson process during the period of observation, we need to consider carefully several features of what was being observed and how it was observed. The following questions are relevant:

1. Are the physically relevant factors, as best we understand them, the result of many different influences over time or space (and hence likely to result in random eruptions) or are they apparently deterministic? (If deterministic, we should select and fit a deterministic model rather than a stochastic one.)

2. Were the observations such that detailed analysis is worthwhile? What was the accuracy of measurement, relative to the inferences we wish to draw? What are the likely errors of measurement? What are the sources and nature of those errors? What constitutes an eruption? Might there be errors or misunderstandings in classifying minor eruptions?

3. If satisfied that detailed analysis is worthwhile, were the intervals between eruptions stationary? (What does this mean? Why does it matter?)

4. Were the intervals between eruptions correlated?

5. If satisfied on these counts, is a Poisson process applicable? What test is appropriate for the hypothesis that the process is Poisson? Against what alternative hypotheses? With what power? Do we have the time, skill and resources (tables or software) to find and carry out such a test? Does it matter? (Since engineers often require pragmatic answers, a detailed examination of these issues may not be warranted if the answer isn’t critical to a decision; “near enough” may be good enough, as long as we remember the liberties we took in assumptions and analysis.)

Anyone who has visited the Blowhole more than once knows that the rate and volume of eruptions varies. This variation occurs at several timescales. We might expect that part is explained by the tides, so that eruptions are more frequent and spectacular when the tide is very high, and eruptions obviously depend on the presence of a large ocean swell generated by prolonged strong winds over the ocean well offshore from Kiama. Hence, any stochastic model fitted to data observed over a short period of time is only applicable to that period, and perhaps a few hours either side of the observations. But we might infer from the model fitted to those data that a similar model applies more generally. Observations at other times might then seek to discover whether (a) a Poisson model is generally applicable; and (b) the way in which the single parameter of the Poisson distribution (or of the exponential distribution) is related to the tide level and the ocean swell.

With all these reservations in mind, we can now begin considering the five questions posed above. (There may be others which you consider relevant. Analyse them too.) How should we consider these questions? Some lend themselves to formal statistical

---

3 *Pragmatic:* dealing with matters with regard to their practical requirements or consequences.
tests of hypotheses, while others require an opinion. Even the results of any statistical tests aren’t unambiguous: what does it mean if the data lead to rejection of a hypothesis at a particular level of significance? Well, it depends on the decision we propose to make on the basis of those data! And that is a matter of judgement, NOT a matter of scientific or mathematical ‘fact’.

1. Random or deterministic?

Waves generated in the vicinity\(^4\) of Kiama will have a range of heights and periods, but will have shorter periods than the swell generated at great distances from the coast. The short-period waves will interact with the swell as well as with the waves reflected from the shore. The interaction of the waves from various sources is likely to provide a large measure of randomness in the size and timing of waves entering the Blowhole. Some claim that every tenth wave is larger than others in the group; if credence\(^5\) is given to this or other claims (“hypotheses”) then a test is appropriate. Otherwise, we might consider that the eruption process is largely random rather than deterministic, at least over intervals of an hour or so.

2. Nature of the observations; sources of error

Since a recording device wasn’t used for the arrival times, mistakes could well have been made. The most likely error is of a whole minute; this is a **systematic error**. There are probably random errors of up to a second or so, not because the digital watch is inaccurate but because of the difficulty of observing both the watch and the Blowhole, and deciding on a consistent criterion of **when** an eruption occurs: the beginning, middle, or end? The **resolution** of the observations (to the nearest second) may also limit the accuracy of analyses if a continuous distribution is fitted to data which, in its measurement, is discrete. Finally, uncertainty as to what constitutes an eruption (whether to count a near thing, in which a lot of noise is made and some water vapour is ejected, but no column of water) will affect any analysis, since misclassifications in the data have the potential to severely bias parameter estimates as well as deductions about stationarity etc.

3. Are the intervals **stationary**?

It would be a mistake to fit a stationary model to data which are clearly non-stationary. It is impossible to detect slight non-stationarity in a short time series. There is also the question of what test might be appropriate: a test of the mean or the variance in the first and second halves of the sample? A test for trend? A non-parametric test, such as a CUSUM test?

There is a danger in applying many tests: what inference will be made if all but one or two of the tests indicate that the null hypothesis (stationarity, in this case) shouldn’t be rejected? Different tests are not, in general, independent examinations of the same hypothesis.

---

\(^4\) **Vicinity**: near (to).

\(^5\) **Credence**: belief.
With that caution in mind, two tests will be performed: a comparison of the mean for the first 32 intervals with that for the last 32 intervals, and a test for trend.

Even these tests are problematic: the time intervals are clearly from a population which is very different from a normal distribution; indeed, we suspect that it might be like an exponential distribution. Yet the textbooks only present tests of two sample means and of the slope of a regression line fitted to time series data on the assumption that the population from which the sample was drawn is a normal distribution. How to proceed?

Well, we could ignore the proviso that the population should be approximately normal, perhaps justifying our decision by some remarks about the relative robustness of the test in the face of departures from normality. Or we could do a simulation experiment using randomly generated samples from an exponential population; this might turn out to be more time-consuming than we imagined! Or we could look for a textbook or journal article which deals specifically with samples drawn from exponential populations; this is likely to take some time, both to locate the reference and then to understand how to do the test. Whichever decision we make, we need to report honestly and openly the choices we have made, and the possibility that this has affected the inferences we make from the result of the test.

The result of each test is reported in the Appendix.

4. Are the intervals correlated?

Again, there are many kinds of non-randomness, and in order to undertake a test we need to be specific about which alternative hypothesis we wish to test against. The simplest is to plot each time interval against the previous one and look for any pattern. ‘Eyeball’ tests such as this are suggestive and should be undertaken, particularly because they can be done quickly from within a spreadsheet such as Excel. But who is to say what constitutes a significant pattern? Perhaps we should also test whether successive time intervals are linearly correlated. This still leaves unexplored many other kinds of temporal dependence inconsistent with randomness, but is all that we perhaps have time for, unless some particular form of dependence is suggested by consideration of the physics of the situation or the politics of the decision.

The results of these analyses are shown in the Appendix.

5. If the intervals are stationary and random, is a Poisson model appropriate?

Again, it will take time to find an appropriate test of the hypothesis that the sample is from a Poisson process, and we would need to carefully formulate the alternative hypothesis, which will determine what test is appropriate. What constitutes an unacceptably large deviation from the behaviour which is typical of a Poisson process, which would lead us to reject the hypothesis that the data are, for practical purposes, from such a process?

---

6 Proviso: a clause of stipulation or limitation in a document.
So perhaps we have to settle for a rough-and-ready test, rather than a parametric (and more powerful) test. One such test is to compare the observed number of eruptions in non-overlapping periods of (say) five minutes with the number expected on the assumption that the process is Poisson; this is a chi-squared test. Even quicker (but cruder) is to compare the mean and standard deviation of the sampled 64 time intervals. Recall that the time intervals, if the process is Poisson, have an exponential distribution. The sample may not exactly conform to that distribution, especially in small samples, but approaches the true (‘population’) distribution asymptotically. A property of the exponential distribution is that the population mean is equal to the population standard deviation. We might expect that the sample mean and standard deviation would also be similar. An exact test is available, but would take time to locate; it isn’t discussed in standard textbooks. So instead of the exact test, we might settle for a crude comparison.

The sample mean of the 64 time intervals is 39.8 seconds, and the sample standard deviation is 33.75 seconds. These are not so different to lead us to reject out of hand the hypothesis that a Poisson process may be adequate to represent the data. We can now, if we wish, use one of the strategies previously outlined for a more specific examination of this hypothesis.

Another approach would be to plot the ranked data. If the plot more or less followed the shape expected for an exponential distribution, then we might be satisfied, without a formal test of the hypothesis that an exponential distribution was suitable. Hence we would accept that the process appears to be a Poisson, ie random, one.

The cumulative distribution function for an exponential distribution is $F(x) = 1 - \exp(x/\alpha)$, where $\alpha$ is the mean of the distribution. Hence $\ln(1 - F(x)) = -x/\alpha$. If ranked values on an arithmetic scale are plotted against the rank (a proxy for $1-F(x)$) on a log scale, a linear plot should result for large samples from an exponential distribution.

Since advanced skill in researching an appropriate test isn’t one of the objectives for this subject, we’ll settle for the straightforward chi-squared test and graphical plot; these are also reported in the Appendix.

**Conclusions**

For practical purposes, an engineer could conclude that the eruptions at Kiama on the afternoon of 12 July 1998 appear to conform to a Poisson process. A much larger data set would be needed to refute this proposition, and then we would run into the problem of non-stationarity mentioned above.

Also, in order to test one hypothesis, we sometimes need to assume the result from a test which we’ve yet to undertake. Model building and hypothesis testing don’t proceed in strictly sequential steps; we reconnoitre, make assumptions, decide which

---

7 *Asymptotically*: modern Latin asymptota (linea ‘line’) from Greek asumptotos ‘not falling together’ (as a-1, sun ‘together’ + ptotos ‘falling’ from pipto ‘fall’).

8 *Reconnoitre*: make a reconnaissance of (an area, enemy position, etc.).
aspect of a problem to investigate first, and then adjust our initial strategy as details emerge. But you are unlikely to need to know much about blowholes to practise your chosen profession! What more general conclusions have we reached from this little problem?

This simple data set required much more thinking about the circumstances in which the observations were made and of the kinds of hypotheses which can usefully be examined than the time taken for a few appropriate statistical tests. As a general rule, it is much more important for an engineer to be clear about what decision (such as the appropriate value of a design parameter, or method of analysis) she or he is seeking to make than to achieve great precision in a model used to assist that decision.

At least three kinds of ‘numeracy’ were called for:

- a skill in considering the nature of the ‘experiment’, the sources of error and their magnitudes, etc;
- a skill in deciding which hypotheses should be considered and in what detail; and
- a skill in finding, applying and reporting the results of any such tests (including any crude ‘tests’);

In other words, to reach a sensible decision.

Engineers need to develop these skills so that, in time, they may develop professional judgement. Professional judgement looks, to many outsiders, just like guessing, but it isn’t. Nor is it an arrogant assertion that one’s opinion is invariably correct, or that there is a single way in which to examine a problem. Rather, it is the ability to make inspired guesses that lead to sensible decisions even when the assumptions may be flawed\(^9\). The true professional has a healthy scepticism about the analyses she or he has undertaken – the chances for mistake, and of erroneous assumptions – but nevertheless has a reliable intuition of when to seek another opinion or to undertake a different analysis, or to take a cautionary approach (to apply a ‘factor of safety’), and when to rely on what will always be a less than complete analysis of the complex problems which engineers deal with daily.

---

\(^9\) **Flaw**: an invalidating defect in a legal matter.
Appendix: Data, hypotheses, and calculations and results of various statistical tests

Data

The following table (from a spreadsheet) shows the data.

**Times of 65 eruptions of the Kiama Blowhole from 1340 hours, 12 July 1998**

<table>
<thead>
<tr>
<th>eruption</th>
<th>time (seconds)</th>
<th>interval (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>83</td>
</tr>
<tr>
<td>2</td>
<td>95</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>146</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>233</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>293</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>321</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>416</td>
<td>95</td>
</tr>
<tr>
<td>8</td>
<td>424</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>451</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>466</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>476</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>494</td>
<td>18</td>
</tr>
<tr>
<td>13</td>
<td>510</td>
<td>16</td>
</tr>
<tr>
<td>14</td>
<td>539</td>
<td>29</td>
</tr>
<tr>
<td>15</td>
<td>593</td>
<td>54</td>
</tr>
<tr>
<td>16</td>
<td>684</td>
<td>91</td>
</tr>
<tr>
<td>17</td>
<td>692</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>709</td>
<td>17</td>
</tr>
<tr>
<td>19</td>
<td>764</td>
<td>55</td>
</tr>
<tr>
<td>20</td>
<td>774</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>809</td>
<td>35</td>
</tr>
<tr>
<td>22</td>
<td>856</td>
<td>47</td>
</tr>
<tr>
<td>23</td>
<td>933</td>
<td>77</td>
</tr>
<tr>
<td>24</td>
<td>969</td>
<td>36</td>
</tr>
<tr>
<td>25</td>
<td>986</td>
<td>17</td>
</tr>
<tr>
<td>26</td>
<td>1007</td>
<td>21</td>
</tr>
<tr>
<td>27</td>
<td>1043</td>
<td>36</td>
</tr>
<tr>
<td>28</td>
<td>1061</td>
<td>18</td>
</tr>
<tr>
<td>29</td>
<td>1101</td>
<td>40</td>
</tr>
<tr>
<td>30</td>
<td>1111</td>
<td>10</td>
</tr>
<tr>
<td>31</td>
<td>1118</td>
<td>7</td>
</tr>
<tr>
<td>32</td>
<td>1152</td>
<td>34</td>
</tr>
<tr>
<td>33</td>
<td>1179</td>
<td>27</td>
</tr>
<tr>
<td>34</td>
<td>1207</td>
<td>28</td>
</tr>
<tr>
<td>35</td>
<td>1263</td>
<td>56</td>
</tr>
<tr>
<td>36</td>
<td>1271</td>
<td>8</td>
</tr>
<tr>
<td>37</td>
<td>1296</td>
<td>25</td>
</tr>
<tr>
<td>38</td>
<td>1364</td>
<td>68</td>
</tr>
<tr>
<td>39</td>
<td>1510</td>
<td>146</td>
</tr>
</tbody>
</table>
40 1599  89 
41 1617  18 
42 1690  73 
43 1759  69 
44 1768   9 
45 1805  37 
46 1815  10 
47 1897  82 
48 1926  29 
49 1934   8 
50 1994  60 
51 2055  61 
52 2116  61 
53 2134  18 
54 2303 169 
55 2328  25 
56 2336   8 
57 2362  26 
58 2373  11 
59 2456  83 
60 2467  11 
61 2509  42 
62 2526  17 
63 2540  14 
64 2549   9 
65 2561  12 

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean  (seconds)</th>
<th>Std. Deviation (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>first half</td>
<td>36.0</td>
<td>26.1</td>
</tr>
<tr>
<td>second half</td>
<td>43.0</td>
<td>39.9</td>
</tr>
</tbody>
</table>

Item 3. Stationarity of the 64 time intervals

The following table shows the mean and standard deviation of the first 32 time intervals and the last 32 time intervals.

It looks as though eruptions occurred less frequently as the sampling period progressed, but, given the small sample, we need to be careful before leaping to conclusions!

For samples drawn from normal populations, the test of the hypothesis that two population means are different employs the *t* distribution. That test is reasonably robust, so we’ll use it, even though the assumptions which underlie the test are not met.
The test statistic is \( t = \left( \frac{\bar{x}_1 - \bar{x}_2}{s_1^2/n_1 + s_2^2/n_2} \right) \), where \((\bar{x}_1, s_1^2)\) are the mean and variance, respectively, of the first half-sample, and \(n_1\) is the size of that sample. By calculation \( t = -0.83 \). The degrees of freedom is calculated from a messy formula; by calculation, the number of degrees of freedom for the \( t \) test is 51. Consulting a table of critical values for the \( t \) distribution, the 5 percent value for 50 degrees of freedom is 1.67; thus, in 5\% of randomly chosen pairs of samples of 32 values from a normal distribution, the test statistic would be greater than 1.67. Since the distribution is symmetrical about \( t = 0 \), 95\% of randomly selected samples of this size would have the test statistic greater than -1.67. Our sample yielded a much less significant test statistic, so it appears that we cannot reject the hypothesis that the sample is stationary.

Alternatively, we could find and employ a test which is specific to the exponential distribution, or find the critical points for the \( t \) statistic for samples of size 32 drawn from an exponential population by carrying out a simulation experiment. The details aren’t reported here.

The following graph shows each time interval plotted against the ordinal number of the eruption; also shown is a trend line fitted by least squares, and the coefficient of determination for that line:

![Graph showing time intervals against ordinal number with trend line and \( R^2 = 0.0029 \)](image)

Because the data aren’t from a bivariate normal distribution, it isn’t appropriate to apply the conventional tests of significance for the slope or the correlation coefficient. But it seems that there is no strong trend; the slope is only -0.097 and the correlation coefficient is only -0.053. Note that the average time interval for the first 32 time intervals was shorter than for the second 32 (perhaps suggesting a slight tendency to less frequent eruptions as time progressed), whereas the slope suggests the opposite. The two tests lead to the same conclusion however. We conclude that the data are consistent, for our purposes, with the hypothesis that the time series was stationary.
**Item 4. Correlation of successive time intervals**

A graph of each time interval against the preceding one follows. It seems that there is no pattern.

![Graph showing time intervals](image)

The correlation between successive time intervals can be calculated to see if a particular kind of non-randomness is evident in the data.

A simple hypothesis is that successive values are uncorrelated, against the alternative that they have a correlation coefficient different from zero. The correlation coefficient is readily determined using a spreadsheet. The column of time intervals is copied into the next column, but one row lower. Then the CORREL function is used to compute the (linear) correlation coefficient for the 63 pairs of values. For this sample, the correlation coefficient is 0.057. A value of zero indicates no correlation (strictly, the lack of a linear association between successive time intervals), whereas a value of +1.0 would indicate perfect (linear) association, and a negative value would indicate a tendency for longer-than-average intervals to be followed by short ones, and vice versa. The correlation between time intervals which are two time steps apart etc. could also be calculated, but in the absence of a reason to suspect this kind of association it won’t be undertaken.

An exact test is described in textbooks for the case where the variable is random and from a normal distribution. But a histogram will quickly reveal that the data are from a population which is evidently very different from a normal distribution, so that it may be misleading to report the result of (incorrectly) applying that test. It suffices to say that the linear correlation coefficient is small (having regard to the sample size) and consistent with the sample being from a random process.
**Item 5. Is the process Poisson?**

As noted in the discussion in the text, the sample mean and standard deviation are similar, so we haven’t rejected the hypothesis that the population is an exponential distribution (as it would be if the eruptions occurred in accordance with a Poisson process). Rather than a parametric test (what’s that?), we decided to employ a chi-squared test. The following table compares the expected and observed number of eruptions in succeeding periods of five minutes; observations were continued until precisely 45 minutes had elapsed, by which time 68 eruptions had occurred. The table assumes that the eruptions occurred randomly at the rate of one per 39.8 seconds, on average, or 7.54 per five-minute period, which is our best estimate of the rate.

<table>
<thead>
<tr>
<th>Timer period (minutes)</th>
<th>Observed number of eruptions, $O_i$</th>
<th>$O_iE$</th>
<th>$(O_i - E)^2 / E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 5</td>
<td>5</td>
<td>-2.54</td>
<td>0.86</td>
</tr>
<tr>
<td>5 - 10</td>
<td>10</td>
<td>+2.46</td>
<td>0.80</td>
</tr>
<tr>
<td>10 - 15</td>
<td>7</td>
<td>-0.54</td>
<td>0.04</td>
</tr>
<tr>
<td>15 - 20</td>
<td>11</td>
<td>+3.46</td>
<td>1.59</td>
</tr>
<tr>
<td>20 - 25</td>
<td>5</td>
<td>-2.54</td>
<td>0.86</td>
</tr>
<tr>
<td>25 - 30</td>
<td>6</td>
<td>-1.54</td>
<td>0.31</td>
</tr>
<tr>
<td>30 - 35</td>
<td>7</td>
<td>-0.54</td>
<td>0.04</td>
</tr>
<tr>
<td>35 - 40</td>
<td>7</td>
<td>-0.54</td>
<td>0.04</td>
</tr>
<tr>
<td>40 - 45</td>
<td>10</td>
<td>+2.46</td>
<td>0.80</td>
</tr>
</tbody>
</table>

(The chi-squared test could have been applied in many other ways. Choosing to do it this way was a matter of judgement.) The sum of the values in the final column gives us the value of chi-squared; it is 5.34. There are nine classes, and one parameter was estimated, so that there are eight degrees of freedom. Tables of critical points for the chi-squared distribution with eight degrees of freedom show that 95% of values of chi-squared for such samples will exceed 2.73, and only 5% will exceed 15.5. The expected value (average) of chi-squared calculated for many random samples with eight degrees of freedom is 8.0. There is nothing remarkable about the computed value; it is close to what would be expected if the underlying process is Poisson. Unless the decision is a critical one, there seems little point in finding and applying a more sophisticated test, or model. We do not reject the hypothesis that eruptions follow a Poisson process, at least during the period of observation.
The ranked values are shown on the following graph, plotted on an arithmetic scale against the logarithm of the rank. A linear plot indicates that an exponential distribution is satisfactory. Given the relatively small sample size, the plot is approximately linear, so the hypothesis that eruptions are a Poisson process is accepted.
Test cricket

Preliminary reading

Do the following pre-reading activities before you settle down to read carefully. This will save time and help you get the most out of your reading.

1  Identify the background of the article:
(a) Who is the author?
(b) What is the title?
(c) What is the name of the publication it appears in?
(d) Write out a reference for this article in one of the approved styles.

2  Skim through the article:
(a) Set a timer for 5 minutes and skim the whole article. Get as good an idea as you can of what it is about in the time limit.
(b) Summarise the article in 3-5 dot points.

Now read the whole article carefully and answer the following questions. They will help you identify the reasons why the author has used the particular research design and statistical techniques.

3  Identify the research question:
(a) Write down the research question.
(b) How is it presented?
(c) Why has the author presented it this way?

4  Identify the research methodology used:
(a) Write down the research methodology used.
(b) Is it an observational study or an experimental study, or neither?
(c) What sampling techniques are used (if any)?

5  How does the author deal with data?
(a) What data are given in the article?
(b) How are the data presented or described?

6  What statistical techniques are used and why?
(a) List the statistical techniques used or referred to.
(b) Give a reason for using each of them.
In-depth reading

The preliminary questions considered the techniques used in the original research and then in writing the article. The following set of questions and activities are to help you with thinking about the text and statistical design in more depth.

1 Aim and audience

(a) What is the author’s major aim in writing the article?

(b) What clues do the statistical techniques give you about the aim?

(c) What audience is the author writing for? Give reasons.

2 Content

(a) What is a drawn match in cricket?

(b) Do you think it is reasonable to ignore the drawn matches in the section, “Was Australia The Better Team?” Is your answer influenced by any bias towards or against Australia or England?

3 Statistical analysis

(a) Croucher considers the post-war cricket Test matches played between Australia and England from 1946 to 1975 and asks whether Australia was the better team. Check his analysis using the sign test. What should you do with the drawn matches? What alternative analyses could you carry out if you had the actual scores from these 82 games? Explain the reasons for your choice.

(b) Many sport commentaries include a section from “the statistician”. Listen to such a commentary on radio or television and focus on the statistical aspects. What role does such a person play? What qualifications would you need for such a job? Would it be suitable for someone who had studied statistics as part of a degree in some other area?

4 Activity

Sport, leisure and gambling are important applications of statistics. Consider a sport or hobby that you are interested in. Use the web or other information sources to investigate what statistics are being recorded and which statistical techniques are being used.

Write a letter to the web/information manager. You can either be critical or complimentary about their statistical information.
One of the many sports which seem to provide an endless stream of statistical data is cricket. Such data can provide excellent material for demonstrating statistical techniques and could form the basis for individual project work. Here we examine the post-war cricket matches played between England and Australia from 1946 to 1975 (up to the advent of World Series Cricket) and show how even the most elementary statistics can be used to draw some interesting comparisons between the performances.

Since 1882 England and Australia have played series of ‘Test’ cricket matches with the winner of each series said to have won the ‘Ashes’. Over this period of time thousands of scores and cricketing records have been kept, the most prominent publication of these being the annual Wisden Cricketers’ Almanack.

Elementary statistical techniques may be used to find some surprising results, and also provide an opportunity for students to realise that statistics can be combined with one of their favourite pastimes.

**Was Australia The Better Team?**
The answer to this question depends upon your interpretation of the problem. There were 82 post-war Test matches played between 1946 and 1975 and of these 30 were won by Australia, 16 won by England and 36 drawn. We can use a sign test and consider the problem from two different angles.

(a) If we ignore the drawn matches – Australia has won 30 out of 46 \((p = 0.652)\) of those matches with a result. Under the null hypothesis of no significant difference between the two teams, we can use a \(z\)-test statistic to convert our observed value \(p\) to a \(z\)-score and get \(z = 2.06\) which is significant at \(\alpha = 0.05\).

(b) If we include the drawn matches (and allocate them each as half a win to each team) – Australia has won 48 out of 82 matches. Using the same principles as in (a), this value of \(p = 0.585\) converts to a \(z\)-score of 1.54 which is certainly not significant at \(\alpha = 0.05\).

The reader is left to ponder which of (a) and (b) they choose to believe!

**Are Innings Scores In The Same Game Correlated?**
In some Test matches a country was totally dismissed in both innings (i.e. each batsman was dismissed twice in the match). Could there be a relationship between a team’s first innings total and its second innings total and, if so, can we predict (using regression analysis) what the second innings total might be from the first innings total?
To examine this we consider the most recent 25 Tests (for each country) in which each team is dismissed twice.

**Australia**

If the subscript \(i\) refers to the innings (\(i = 1, 2\)), the following statistics are obtained:

\[
\bar{x}_1 = 247.64 \quad s_1 = 87.94 \quad s_{11} = 185603
\]

\[
\bar{x}_2 = 231.76 \quad s_2 = 97.86 \quad s_{22} = 229838
\]

\[n = 25 \quad s_{12} = -100\]

The correlation coefficient (\(\rho\)) is zero (to the fourth decimal place!) indicating clearly that, for Australia, no relationship exists between the totals of the two innings. No meaningful predictions of second innings scores can therefore be made. (A paired \(\tau\)-test on this data yielded a \(\tau\) statistic of 0.603 with \(v = 24\) degrees of freedom. This is not significant at \(\alpha = 0.05\) and hence no significant difference exists between the two sets of scores.)

**England**

The correlation coefficient is \(r = -0.304\). To test this for significance we can use the \(z\) statistic where \(z = r \sqrt{n - 1}\) and the null hypothesis that \(\rho = 0\). Since the value of \(z\) is \(-1.49\) (which is not significant at \(\alpha = 0.05\)), we cannot reject \(H_0\) and again conclude that no significant relationship exists.

(A paired \(t\)-test here yielded a \(t\) statistic of \(-1.41\) with \(v = 24\). This value was not significant at \(\alpha = 0.05\)).

It thus appears that no second innings prediction can be made. A point of interest is that \(\bar{x}_2\) exceeds \(\bar{x}_1\) for England but \(\bar{x}_2\) is less than \(\bar{x}_1\) for Australia. (Is this significant?)

**Are Innings Scores Normally Distributed?**

If we assume that a population is normal with mean \(\mu\) and standard deviation \(\sigma\) we can immediately determine what proportion of that population would lie between any two values. In sampling from that population we would expect the sample values to be distributed similarly to the population and can use a \(\chi^2\) test to see whether the sample pattern is close to that expected.

For Australia we test the null hypothesis that the scores came from a normal distribution with \(\mu = 312.0\) and \(\sigma = 112.9\).

The value of the \(\chi^2\) test statistic for goodness-of-fit in my analysis was 7.14 with \(v=7\) degrees freedom. The critical value at \(\alpha = 0.05\) is 14.07 and so we have no evidence to reject the hypothesis that Australian totals are normally distributed.

For England the corresponding \(\chi^2\) goodness-of-fit test statistic was 13.13 with \(v=7\) degrees of freedom. In this case the value of 13.13 is closer to the \(\alpha = 0.05\) critical value of 14.07, but the normal distribution hypothesis can still not be rejected.
**Are The Distributions Of First Innings Scores The Same?**

To answer this we construct a contingency table and use a $\chi^2$ test statistic to test a null hypothesis that there is no significant difference between the distributions of the two countries. The columns of the contingency table are two teams: the rows are grouped into convenient classes.

In my analysis, the $\chi^2$ test statistic has a value of 5.86 with $v = 4$ degrees of freedom. The critical value for $\alpha = 0.10$ is 6.25 so that at the 10% significance level there appears to be no significant difference between the two distributions.

**Is There Any Significant Difference Between The Two Teams' Scores?**

Since the sample sizes are large and the standard deviations of the distributions appear to be roughly equal, a two-sample $z$ test may be used to test the null hypothesis that no significant difference exists between the two sets of totals. The values of this $z$ test statistic is 1.87 and we cannot reject $H_0$ at $\alpha = 0.05$ (since the critical value in this case is 1.96).

**Are The Standard Deviations Of The Population Of Innings Total Equal?**

Calculations give $s_A = 112.9$ for Australian totals and $s_E = 112.6$ for English totals. We have already shown that it is reasonable to assume that the two populations are normally distributed and so the $F$ distribution may be used to test the null hypothesis that the population variances are equal.

The value of our test statistic is:

$$F = \frac{(112.9)^2}{(112.6)^2} = 1.005$$

which is well within the region of acceptance ($0.68 < F < 1.47$). Hence there is no evidence to reject the null hypothesis at $\alpha = 0.10$.

**Remarks**

With the enormous volume of statistics available from cricket matches all over the world much detailed analysis can be undertaken. The raw data is readily available from a number of sources. Similar analyses to those given here could be carried out on pre-war Tests and comparison made with those for post-war. Another interesting comparison would be that between the performance of players of the now-defunct World Series Cricket with the same players’ efforts in the conventional Test matches.

Further areas of investigation include the shape of the distribution of scores of individual batsmen, and the length of partnerships which immediately follow lengthy partnerships (many commentators claim that when a long partnership is broken another wicket falls soon afterwards). The list is almost endless. Perhaps the best method is to present the data to the students and let them use their own initiative in selecting those analyses which are of most interest.
Conceptions of environment

Preliminary reading

Do the following pre-reading activities before you settle down to read carefully. This will save time and help you get the most out of your reading.

1 Identify the background of the article:
   (a) Who are the authors?
   (b) What is the title?
   (c) What is the name of the publication it appears in?
   (d) Write out a reference for this article in one of the approved styles.

2 Skim through the article:
   (a) Set a timer for 5 minutes and skim the whole article. Get as good an idea as you can of what it is about in the time limit.
   (b) Summarise the article in 3-5 dot points.

Now read the whole article carefully and answer the following questions. They will help you identify the reasons why the authors have used the particular research design and statistical techniques.

3 Identify the research question:
   (a) Write down the research question.
   (a) How is it presented?
   (b) Why have the authors presented it this way?

4 Identify the research methodology used:
   (a) Write down the research methodology used.
   (b) Is it an observational study or an experimental study, or neither?
   (c) What sampling techniques are used (if any)?

5 How do the authors deal with data?
   (a) What data are given in the article?
   (b) How are the data presented or described?

6 What statistical techniques are used and why?
   (a) List the statistical techniques used or referred to.
   (b) Give a reason for using each of them.
In-depth reading

The preliminary questions considered the techniques used in the original research and then in writing the article. The following set of questions and activities are to help you with thinking about the text and statistical design in more depth.

a. **Aim and audience**

(a) What is the authors’ major aim in writing the article?

(b) What clues do the statistical techniques give you about the aim?

(c) What audience are the authors writing for? Give reasons.

2 **Content**

(a) What do the authors mean by ‘object’ and ‘relation’ conceptions?

(b) What are the variables that are not significant in the models? Why is this finding important?

3 **Statistical analysis**

(a) What is the main difference between logistic regression and linear regression? In this study, what was the response variable and what were the predictor or explanatory variables? Which of the explanatory variables were factors (categorical) and which were variates (quantitative)? How does the design of the study (experimental or observational) affect the conclusions?

(b) Work with a partner on this question. Explain the meaning of the terms “odds” and “odds ratio”. Illustrate your answer with examples from the logistic regression carried out by the authors. Write your explanations in a form that would be useful to your colleagues or fellow students.

(c) The most obvious finding of Loughland et al. was that primary school students are much more likely than high school students to view “environment” in terms of a relation rather than as an object. On the basis of your experience with school, can you suggest explanations as to why this is so (maybe starting with the reasons identified by the authors)?

4 **Activity**

Do this activity in small groups.

Do you think that schools are a major influence on young people’s conceptions of the environment? What are other influences? Imagine that your group are environmental officers for the Australian Government. Plan out a research study to find out the main influences on people’s conceptions of the environment, in order to set up a media campaign to encourage people to work towards a sustainable environment.
Factors Influencing Young People’s Conceptions of Environment

Tony Loughland, Charles Sturt University.
Anna Reid, Macquarie University.
Kim Walker, George Washington University.
* Peter Petocz, University of Technology, Sydney,
PO Box 123, Broadway, NSW 2007, AUSTRALIA

(* corresponding author)

[Note: this paper was submitted to *Environmental Education Research* in 2001]

**SUMMARY**  
Environmental education in schools is an important strategy in achieving environmental protection and improvement. However, it needs to be based on children’s understandings of environment rather than on assumptions of what children know and believe. In a previous paper (Loughland, Reid & Petocz, 2002), we reported on a research project where school children’s responses to an open-ended statement ‘I think the term/word environment means …’ were analysed using the qualitative research method of phenomenography. An important qualitative difference was found between conceptions that treat the environment as a relation and those that treat it as an object. In this paper, we examine statistically the factors that incline students to a ‘relation’ rather than an ‘object’ conception of environment. We argue that development of the former would seem to be an important aim of environmental education, and indicate how this may be achieved.

**Introduction**  
Environmental education in schools is seen as an important strategy in achieving environmental improvement. However, relatively little is known about the environmental understandings held by children. Several research programs (for example, Boyes and Stanisstreet, 1993, 1994, 1996, 1997; Stanisstreet and Boyes, 1996) have examined children’s ideas about individual environmental topics. Such programs successfully clarify the way students understand these individual themes. However, this type of research has encouraged the development of environmental education programs in Australia and elsewhere that are based on developing young people’s environmental knowledge and challenging assumptions about what they believe. Rickinson (2001) suggests that this approach represents a ‘passive view of students…who are to be altered through educational programs’ (p. 217). He indicates further that:

Such studies are almost universally ones that seek to evaluate the effects of particular educational treatments (independent variables) on students’ environmental knowledge, attitudes or behaviour (dependent variables) through some kind of quasi-experimental pre-test/post-test design. This means that much of the evidence available on outcomes is quantitative in nature and utilises outcome criteria that are tightly specified prior to investigation.’ (p. 222)

Essentially this implies a dualist paradigm where learner and learning are treated as different entities. An intriguing and pedagogically effective strategy would be to explore young people’s conceptions of their total experience of the environment and the other factors that may contribute to these conceptions. Clearly, more effective programs could be developed if young people’s environmental understandings and beliefs were known and treated in a non-dualist manner.

This paper emanates from a research project undertaken in New South Wales, Australia, entitled *The Social/Cultural Influences on Environmental Understandings of NSW School Students* that aims to address this problem. The overall aim of the project is to ensure the development of more appropriate and student-centred environmental programs and curricula by a wide range of educational and environmental bodies (Walker *et al.*, 2001; see endnote).
In a previous paper (Loughland et al., 2002), we analysed the responses given by 2249 young people from years 3, 6, 8 and 11 (ages 9, 12, 14 and 17) in NSW to question 8 of the survey, an open-ended question asking students to state what they understand by the word or term ‘environment’. Students’ responses were analysed using a phenomenographic approach. The aim of that paper was to identify an ‘outcome space’, a hierarchical classification of conceptions. Six distinct conceptions were isolated, ranging from the least sophisticated – environment as a place – to the most inclusive and expansive – environment and people in a relationship of mutual care. An important qualitative difference was found between conceptions that focus on the environment as an object and conceptions that see a relation between people and the environment. It is on this distinction that the analysis of the present paper is focused.

We have used the students’ own voices to map out the range of variation in their conceptions of environment, and now we analyse the information in the questionnaire on the basis of this mapping. Using a standard statistical/quantitative methodology, we investigate how the students’ demographics and opinions correlate with their own conception of environment (with ‘object’ and ‘relation’ becoming the dependent variables in contrast to Rickinson’s previous observation). Previously, we were not concerned with identifying which conception of environment may be held by a particular student. Here, in contrast, we use students’ actual answers to categorise them into ‘object’ conceptions, or ‘relation’ conceptions. Of course, this cannot be done without possible error, a problem shared by most investigations using the statistical methodology. One source of error may be in the classification of the students’ answers. For instance a single-word response from a student may have been ‘nature’. This response is limited in range and complexity but other studies (such as Keliher, 1997) confirm our placement of this response in the ‘object’ category. Our classification of individual answers into conceptions was carried out independently by two of the authors, and checked by a third whenever there was disagreement. Another source of error may originate from students not writing enough to allow their views to be accurately ascertained. However, this is certainly less of a problem in the present situation, where we classify students only into two broad groups of conceptions. Without re-surveying students, there is no way of checking this aspect of the reliability. A second phase of the study, consisting of in-depth teaching interactions with schoolchildren, is presently being conducted by Loughland: we anticipate that it will yield further information on this point.

Overall, the combination of methodologies used in this research to explore participants’ conceptions of environment is one of its strengths. Fielding and Schreier (2001) suggest that, as far as the description of lived experience is concerned, qualitative research allows for the complexity of the situation to be uncovered whilst quantitative research is able to distinguish the different extents to which a category of experience may be present. Our complementary methods include a description of the variation found in students’ experience of environment and an analysis of the extent to which these conceptions prevail amongst school students in NSW, and the factors that correlate with them. Many researchers point out that there is strength in the integration of qualitative and quantitative research methods focused on exploring the nature and extent of environmental learning (Rickinson, 2001) and other areas (Onwuegbuzie, 2000; Hartley and Chesworth, 2000).

**Significance of the study**
The most important and unique feature of our analysis is that we have used students’ own voices to map out the range of their conceptions of environment, and then linked other significant factors to these conceptions. As a result, we have shown the significance of a well-developed relation conception to the environment and provided some insights into how relation concepts can be developed.

Lang (1999/2000) has focused on examining changes in teacher understanding of environmental education and how educational interventions may be used to change these understandings. Our research has focused on a related issue, how students understand the environment. We suggest that the key to ‘promoting sustainable practices to save the planet’ (Belgrade Charter, 1976) is through expanding young people’s understanding of the relations between self/people and the environment. Our research suggests that the majority of students...
adhere to an ‘object’ conception while only a few express a ‘relation’ conception. Clearly, the expansion of the ‘object’ view towards a ‘relation’ view will result in a situation where young people expect that the environment contributes to their well-being as they contribute to the environment’s well-being, the essence of the sixth and highest conception that we identified (Loughland et al., 2002).

The importance of a well-developed understanding of the environment is emphasised in the newly-released Environmental Education Policy for Schools (NSW Department of Education and Training, 2001). This document states that: ‘The environment is the aggregate of all the conditions that support living things. In turn, living things, including humans, are all interactive parts of the environment’ (p. 7).

In a recent paper, Stables and Bishop (2001) suggest that the environment can be interpreted as a form of ‘text’. They indicate that the idea of environment can be perceived as the way individuals and groups ‘read’ the environment historically and aesthetically, that ‘there are many correct and different ways of understanding the environment’, and that ‘different cultural and social groups almost inevitably have different views of the environment and of environmental issues’. They conclude that ‘Individuals’ views may therefore change as they move between social settings’. From this point of view, the scope and breadth of meaning embedded in the term ‘environment’ can only be interpreted from within a particular cultural perspective. The phenomenographic analysis and the statistical modelling described in our paper supports this view, as we describe how young people within the physical context of the school attempt to define more broadly what they mean by environment. It may be that they change their views as they move between settings – perhaps more relational in the outdoors – although having made this point we are cautious in our conclusions. The focus group discussions, while in a school setting, asked children to talk about and imagine other settings such as their home. There was no more evidence of a relation concept as they thought about their home life. However, the children did make a strong point that they learn about the environment from what they see. It may be that these are the contexts that are more conducive to forming a relation concept of the environment. Prosser and Trigwell (1999) indicate that the manner in which people report their experience or understanding of a phenomenon is associated with the situation in which they find themselves and on which they are focusing.

Yet another research report (Ballantyne et al., 2001) suggests that the quality of student learning outcome is enhanced through a focus on affective domains of learning such as ‘enjoyment’ and ‘emotion’. This conclusion is reached through a case study analysis of two different learning situations, one a student activity which placed environmental issues within a social science framework, and the other using some basic research methods to engage older students with environmental issues. These case studies describe a difference in attitude between the primary school group and the secondary school group which is also found in our analysis. This focus on affect and attitude suggests that these may be stronger influences on environmental learning than the traditional dimension of ‘knowledge’, and argues for the inclusion of such variables in a study such as the present one.

Method and Scope of Investigation
We look at the factors – both demographic and those based on students’ views expressed in the questionnaire – that correlate with the students expressing a ‘relation’ conception of environment rather than an ‘object’ conception. These factors are examined and selected using the statistical technique of logistic regression, in which the odds of showing a ‘relation’ rather than an ‘object’ conception is modelled in terms of a number of ‘predictor variables’, which are checked for statistical significance in the model. The methods used are described in standard books such as Agresti (1996). The fact that this was an observational study rather than an experiment implies that factors found significant should be interpreted in a correlational rather than a causal manner.

The following demographic factors were included in the statistical investigation, based on information supplied by the school and the individual respondents:
• location (of school, 1 = rural village, 2 = rural town, 3 = regional city, 4 = Newcastle or Wollongong – large cities, 5 = Sydney – state capital)
• population (of school, 1 = 0–100, 2 = 101–200, 3 = 201–500, 4 = 501–1000, 5 = 1001+)
• SES (socio-economic status of school, 1 = high, 2 = medium, 3 = low)
• hsp (1 = high school/secondary school, 2 = primary school)
• year (school year, 3, 6, 8 or 11)
• sex (1 = male, 2 = female)
• nesb (non-English speaking background, 1 = English, 2 = other language)

Additionally, several variables were derived from the students’ responses to the questionnaire. The original groups of questions were examined from the point of view of a subject expert in environmental education, and also using the statistical technique of factor analysis. The first three questions asked students to select three concerns from a list of social and environmental concerns constructed using information from the original focus groups. These responses were used to construct:

• q1env (selecting environmental rather than social concerns in the local area from a list of concerns supplied, 0 = no environmental, 1 = minority environmental, 2 = majority environmental, 3 = all environmental)
• q2env (as above, but for Australian concerns)
• q3env (as above, but global concerns)

Three questions asked students’ opinion of whether the local, Australian and global environment would be better or worse in the future. These opinions were combined into:

optimism (about the future of the local, Australian and global environment, 1 = pessimist, 2 = neutral, 3 = optimist)

A series of questions asked students’ opinions about how much they themselves could help the environment. Factor analysis indicated two dominant dimensions: self and family, government and business. (These dimensions were stable under different methods of extraction and rotation.) The first of these factors accounted for 22% of the variability and was used to construct the derived variable:

• control (extent to which the student feels they can help environmental problems, 1 = low, 2 = medium, 3 = high)

Another series of questions asked students about their sources of environmental knowledge: family and friends, popular media and educational media (including school). A factor analysis revealed that the most important component of the responses to this question was the total number of different sources that a student reported (again, this was unaffected by different methods of extraction and rotation), leading to the derived variable:

• sources (of environmental knowledge selected from a list supplied, 1 = few, 2 = medium, 3 = many)

A series of multiple-choice questions was used to assess students’ objective knowledge of environmental problems. Different, though broadly comparable, questions were used for primary school (6 questions) and high school (9 questions), and the number of correct answers was used to construct the variable:

• mc (knowledge of environmental issues, from a set of multiple choice questions about the environment, values from 0 = lowest to 6 (primary) or 9 (secondary) = highest)

Finally, the response variable was:

• phen (conception of environment, 1 = object, 2 = relation)

and only students who showed one of the six levels of conception identified by the phenomenographic analysis were included in the statistical aspect of the investigation (that is, 1734 of the 2249 surveys, or 77%).

Initial Findings
First, in setting up the response variable, we note the numbers of students at each level of conception against their school year. We see that 1866 students wrote enough to determine a particular conception, and the remaining 383 were missing a response. Conception 0 represents an unusable answer, leaving 1734 usable responses.
It is immediately apparent that the majority of usable responses (88%) showed a conception of ‘environment as object’, and only a minority (12%) showed ‘environment as relation’. It is also apparent from the tables that the ‘relation’ conceptions were more frequent in the primary classes (years 3 and 6) than in the secondary classes (years 8 and 11), and that there was no obvious difference between the two primary classes, or between the two secondary classes. It turned out that the variable hsps (high school or primary school) was always the most important factor in the odds of conception of environment as ‘relation’ rather than ‘object’.

Table 1: Frequencies of phenomenographic categories 0 to 6, and object/relation

<table>
<thead>
<tr>
<th>Phenomenographic conception</th>
<th>All responses</th>
<th>Usable responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number</td>
<td>%</td>
</tr>
<tr>
<td>0 unusable</td>
<td>132</td>
<td>5.9</td>
</tr>
<tr>
<td>1 Object</td>
<td>522</td>
<td>23.2</td>
</tr>
<tr>
<td>2</td>
<td>844</td>
<td>37.5</td>
</tr>
<tr>
<td>3</td>
<td>157</td>
<td>7.0</td>
</tr>
<tr>
<td>4 Relation</td>
<td>39</td>
<td>1.7</td>
</tr>
<tr>
<td>5</td>
<td>156</td>
<td>6.9</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>0.7</td>
</tr>
<tr>
<td>missing</td>
<td>383</td>
<td>17.0</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>2249</strong></td>
<td><strong>1734</strong></td>
</tr>
</tbody>
</table>

Table 2: Cross-tabulation of phenomenographic categories by school year

<table>
<thead>
<tr>
<th>Phenomenographic conception</th>
<th>School year</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Object</td>
<td>334</td>
<td>414</td>
</tr>
<tr>
<td>Relation</td>
<td>73</td>
<td>91</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>407</strong></td>
<td><strong>505</strong></td>
</tr>
</tbody>
</table>

Initial examination of models confirmed that hsps (high school or primary school) was a more useful variable than school year, since there was never any significant difference (at the 5% level) between years 3 and 6, and between years 8 and 11. Investigations also indicated that there were interactions between location of school and population of school, and that the use of both these variables and their interaction led to unstable models. A principal components analysis suggested transformation to the variables:

- ‘pop–loc’ (population minus location of school, combine the two lowest values and add 2, values from 0 = small school large place, to 5 = large school small place)
- ‘pop+loc’ (population plus location of school, values from 2 = small school small place, to 10 = large school large place)

The first of these variables is smallest for small schools in large places, and largest for large schools in small places: it is a measure of the discrepancy between the school’s size and the location’s size. The second of these variables is smallest for small schools in small places, and largest for large schools in large places: it is a measure of size, taking school and location in combination.

Results from Modelling

Two models will be presented, one a simplified version of the other, showing those factors that are significant at the (traditional) level of 5% (p < 0.05). The effect of these factors is presented in the form of **odds ratios**, showing the (multiplicative) effect of the factor on the odds of ‘relation’ rather than ‘object’ conception. These models give an interesting insight into the nature of the relationships between the predictor variables and the response variable. Tables 1 and 2 summarise the results from these models; statistical details of both models can be found in the appendix. The modelling process started by examining all the variables that were defined.
earlier; however, some of these variables were never significant in any of the models considered, and so have not been included below.

**Model 1: Binary Logistic Regression 1.** In this model, the variables that are used to explain the choice of ‘relation’ rather than ‘object’ are pop–loc, hsps, sex, q3env and mc within hsps (knowledge score analysed separately for high school and primary school, since the questions were different). The ‘logistic regression table’ shows the relevant information. The table shows that the most significant variable is hsps \((p < 0.001)\): the odds ratio of 6.28 says that primary school children are more than 6 times as likely as high school children to have a ‘relation’ rather than an ‘object’ conception of environment. Moreover, we are 95% sure that the true value of this odds ratio is between 2.48 and 15.89, that is, somewhere between two-and-a-half and 16 times as much.

The next most significant variable is sex. Girls are about 1.5 times as likely as boys to have a ‘relation’ conception \((p=0.010, \text{odds ratio } 1.50, \text{with a } 95\% \text{ confidence interval from 1.10 to 2.04})\). This is closely followed by pop–loc \((p = 0.011)\), which has been treated as a factor rather than assumed *a priori* to represent a continuum. The odds ratios of 1.62, 1.60, 1.91, 1.25 and 4.04 for the higher levels show that compared to the smallest schools in the largest places (level 0), the other schools are all more likely to show ‘relation’ rather than ‘object’ conceptions of environment.

The other odds ratios are interpreted in the same way. The variable q3env shows just-significant odds ratios for a quadratic type of effect \((p = 0.050)\) The odds ratios of 1.61, 1.18 and 0.70 for comparing values of 1, 2 and 3 with 0 show that students who picked a mixture of social and environmental concerns for the global environment (values 1 and 2) were most likely to show a ‘relation’ conception; this was least likely to happen if they picked all environmental and no social concerns (value 3). The same pattern of results was observed for q1env, concerns in the local area, although it was not quite significant. The variable mc, representing environmental knowledge, was not significant overall, or for the high school group, but it was significant for the primary school group. The significant odds ratio of 0.86 \((p = 0.036, 95\% \text{ confidence interval } 0.75 \text{ to } 0.99)\) for primary children implies that an increased score on the multiple choice test results in a lower chance of showing a ‘relation’ rather than ‘object’ conception.

**Table 3: Summary results from Model 1**

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Interpretation</th>
<th>Odds ratio(s)</th>
<th>95% confidence interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsps</td>
<td>primary compared to high school</td>
<td>6.28</td>
<td>(2.48, 15.89)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>pop–loc</td>
<td>larger and larger schools in smaller and smaller places compared with the smallest schools in Sydney</td>
<td>1.62, 1.60, 1.91, 1.25, 4.04</td>
<td></td>
<td>0.011 (overall)</td>
</tr>
<tr>
<td>sex</td>
<td>girls compared to boys</td>
<td>1.50</td>
<td>(1.10, 2.04)</td>
<td>0.012</td>
</tr>
<tr>
<td>q3env</td>
<td>global concerns: minority environmental, … majority environmental, … all environmental, compared to no environmental</td>
<td>1.61, 1.18, 0.70</td>
<td></td>
<td>0.044 (overall)</td>
</tr>
<tr>
<td>mc(hsps)</td>
<td>increasing knowledge (from multiple choice test)</td>
<td>0.97 hs, 0.86 ps</td>
<td>(0.82, 1.15) hs, (0.75, 0.99) ps</td>
<td>0.732 hs, 0.036 ps</td>
</tr>
</tbody>
</table>

**Model 2: Binary Logistic Regression 2.** The similarity of all the odds ratios for the variable pop–loc suggests that the important aspect of this variable is a comparison of the largest schools in small places with all the other schools. The variable can be re-defined by combining values 0, 1, 2, 3 and 4 of pop–loc to be:

- Isca (1 = all schools except … 2 = largest schools in country areas)
The second model highlights the significance of this variable (p = 0.005, odds ratio 2.55, 95% confidence interval 1.33 to 4.89). Compared to the majority of the schools, a few (four) large country schools are about two-and-a-half times as likely to show a ‘relation’ rather than an ‘object’ conception of environment. This change does not seem to affect the effects of the other variables, measured by significance or odds ratios.

Since the large country schools identified in the previous paragraph are in fact all high schools, if the variable lsca is removed from the analysis altogether the differences it represents are included in the differences between high schools and primary schools. This reduces the apparent effect of the variable hsps (new odds ratio 4.99, p = 0.001) while leaving the effects of the other variables essentially unchanged.

Table 4: Summary results from Model 2

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Interpretation</th>
<th>Odds ratio(s)</th>
<th>95% confidence interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsps</td>
<td>primary compared to high school</td>
<td>5.89</td>
<td>(2.33, 14.86)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>lsca</td>
<td>large schools in country areas compared to other schools</td>
<td>2.55</td>
<td>(1.33, 4.89)</td>
<td>0.005</td>
</tr>
<tr>
<td>sex</td>
<td>girls compared to boys</td>
<td>1.48</td>
<td>(1.09, 2.01)</td>
<td>0.012</td>
</tr>
<tr>
<td>q3env</td>
<td>global concerns: minority environmental, …</td>
<td>1.58</td>
<td>0.64</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>majority environmental, … all environmental, …</td>
<td>1.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>compared to no environmental</td>
<td>0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mc(hsps)</td>
<td>increasing knowledge (from multiple choice test)</td>
<td>0.96 hs</td>
<td>(0.82, 1.14) hs</td>
<td>0.665 hs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.87 ps</td>
<td>(0.75, 0.99) ps</td>
<td>0.039 ps</td>
</tr>
</tbody>
</table>

Discussion of Models

The first important result to discuss is the interpretation of the high school/primary school differences. Why are primary school students five or six times (the various models gave odds ratios of 6.3, 5.9, 5.0) as likely as high school students to show a ‘relation’ rather than an ‘object’ conception? One plausible explanation is to do with the organisation of the primary school, where children have the one teacher for (almost) all subjects, as opposed to the high school, where they have different teachers for each. In primary school, the level of ‘environmental science’ may be more integrated within the school experience, whereas in high school it becomes more scientific, detailed and objective.

Primary schools practice an integrated environmental education whilst high schools teach environmental education as a separate subject, called ‘Environmental Studies’ or ‘Environmental Science’. From the focus groups conducted before the survey was actually constructed, some students in high school said they had never studied any environmental science. This situation is possible in NSW, since only those students who choose the electives of geography and biology explicitly study any environmental education (although students referred to environmental debates about the ozone layer in their English classes, for instance). Significantly more students in primary school than secondary school reported learning ‘a lot’ about the environment at school (79% as opposed to 62%, p < 0.001). When environmental science is studied in high schools, it may be limited to an examination of a local ecosystem such as a polluted stream. This whole process presents the positivist view of environmental science as just being part of the broader science curriculum. This ‘creates a generation of ecological yahoos without any idea of integrated ecological thinking’ (Orr, 1992, p. 86).

The NSW Environmental Education Policy for Schools (NSW Department of Education and Training, 2001), inspired by the Agenda 21 framework (United Nations, 1993) of ecologically sustainable development, requires schools to adopt a whole school environmental management plan. The policy makes the point that an environmentally active school includes the environment in whole school planning, curriculum, management of resources, and management of school grounds. The recommendation specific to curriculum states that such a school ‘has effective environmental education integrated into all stages and all Key Learning Areas, where appropriate’ (p. 21). High schools will find it harder to integrate environment education across
discipline and faculty boundaries with different teachers for each class: in primary schools it may be easier for one teacher to infuse principles of ecological sustainability into the curriculum.

While high school/primary school (hsps) is the most important variable in the model, school year is never significant beyond the distinction between high school and primary school: years 3 and 6 seem to have the same proportion of students with a ‘relation’ conception of environment, and so do years 8 and 11. This general trend is mirrored in the finding of Hicks and Holden (1995) who found that in a study of children’s’ conceptions of the future, environmental concern peaks with children in year six (ages 11–12) and goes steadily down with age. This is evidence for attributing the differences to organisation factors in high schools vs primary schools, rather than to students’ increasing development with age.

The second important finding is the sex difference. In all models investigated, girls are about one-and-a-half times more likely than boys to show a ‘relation’ conception of the environment rather than an ‘object’ conception. This sex difference could be interpreted by stereotypical sex roles: one explanation is that girls are more socialised than boys to a relational view of life, and hence to a ‘relation’ conception of environment. It could also be argued that girls are socialised to be more caring than boys and consequently display more care about the environment (as suggested by Myers et al., 1999). A review of literature relating to the effects of gender on environmental understandings supports these findings. Research has found that girls are more environmentally conscious and active than boys (Connell et al., 1998; Roper Starch Worldwide, 1994) and that girls have stronger feelings and verbal commitment to the environment while boys have greater knowledge about the environment (Chawla, 1988).

Concern for the environment seems to decrease with age in both sexes, but more markedly in boys (Ainley, 1999).

Another finding concerns the students’ selection of ‘concerns’, constructed from issues that were raised in the initial focus groups. Students who selected a balance of environmental and social global concerns were more likely to have a ‘relation’ conception of environment than those who selected only social concerns and even more so than those who selected only environmental concerns. This was significant for the global environment, and marginally significant for the local environment, but did not reach significance for concerns about the Australian environment. This is an interesting discovery and not one that has emerged in other studies. However, it is not surprising to find that in order to have a well developed relational view of the environment it is important to see that environmental issues and social issues are interrelated. This finding supports the principles of ecological sustainable development espoused in Agenda 21 (United Nations, 1993), the global initiative drawn up at the United Nations Earth Summit in 1992, that environmental improvement involves social, environmental and economic sustainability.

Almost as important are the variables that are not significant in the models. For instance, the population of the school does not seem to have a significant effect on conception of environment, and nor does the location of the school. The two variables showed a significant interaction effect that was stabilised by creating the variable ‘pop–loc’, and finally by looking at large schools in country areas. Thus, the single significant effect of population and location was a group of four large high schools in country areas (these high schools show ‘relation’ conceptions at 10.2% of the total, more than twice that of all other high schools at 4.6%, although this is still lower than the 18.0% for primary schools). A question for further research is: why do students in large country high schools have a greater relational concept of the environment? Our enquiries indicated that these high schools did not have any extra resources for environmental education, for instance, to employ a specialist teacher.

Another variable that appears in the models in an unexpected form is knowledge of environmental issues, obtained from students’ answers to the multiple choice questions. For high school students, knowledge seems to be quite independent of their environmental conceptions. For primary school students, an increased knowledge base is a significant factor in explaining the occurrence of ‘relation’ conceptions of environment. However, it seems to work in the direction that increasing knowledge reduces the odds of the ‘relation’ conception by a small but statistically significant amount! This seems to be in agreement with the primary
school/high school distinction, as students at high school are generally more knowledgeable about environmental ideas, but are significantly less likely to hold ‘relation’ conceptions. It would appear, then, that better knowledge of environmental issues has no correlation (or possibly even a negative correlation) with the development of a relation concept of the environment. We are not suggesting that students do not need environmental knowledge. It would seem, however, that environmental knowledge might be being learned in such a way that relation concepts are not being developed. This finding is not new (Gonzalez-Gaudiano, 2001; Hicks & Bord, 2001) and further supports the need for an integrated approach to environmental education.

Optimism is another variable that is often mentioned in studies of students’ environmental beliefs. In the present study, optimism is significantly negatively correlated with knowledge, as measured by the multiple choice test ($r = -0.25$ for primary students, $-0.12$ for high school students, $-0.18$ for the combined group, $p < 0.001$ in all cases). Previous studies have noted that optimism declines with age (Hutchinson, 1997), and hence with increasing knowledge. The factor ‘optimism’ could be included in the models developed, instead of (but not as well as) knowledge: if included, it shows marginal significance of 0.04 to 0.05 and odds ratios of about 1.25. Interpretation of the influence of optimism seems fairly obvious. Students are 1.25 times as likely to have a ‘relation’ view of the environment for each step up the optimism scale from pessimist to neutral to optimist. This could be an important issue for education if a relational view of the environment is to be nurtured.

Conclusion
Although this investigation focuses on identifying factors that influence conceptions of the environment, the most obvious finding concerns the dominance of the ‘object’ conception and the rarity of the ‘relation’ conception of the environment among young people. In general, the majority of young people see the environment as ‘something out there’ – a place, possibly including living plants and animals, but essentially separated from themselves. Only a minority (about one in eight) see the environment from a relational point of view – something which supports and enhances their living, and which in turn requires their care and support. This single finding could have important implications for environmental education.

In general, it seems that primary rather than secondary students, girls rather than boys, and those students who identify a balance of environmental and social concerns are most likely to have a ‘relation’ conception of environment. Knowledge seems to have little influence in high school, and a small negative influence in primary school. Alternatively, optimism seems to be negatively correlated with knowledge, and hence has a small positive influence on the chance of holding a ‘relation’ conception. A few, large country high schools are more likely than most to have students with a ‘relation’ conception, and this is most likely to be due to specific conditions at those schools.

A possible conclusion from this study is that current environmental education requires a re-orientation. Environmental education as it exists now in Australian high schools, separated into the discrete curriculum areas of Biology and Geography, may not be very effective in creating opportunities for young people to integrate ecological values into their thinking. As it stands, they are subjected to an education system underpinned by the anthropomorphic values of liberalism and modernity: animals are organised into hierarchies, both plants and animals are regarded as pets or pests (Marshall, 1992) and nature is constructed as a ‘scene’ to be viewed from the windows of our cars, or through the television screen (Berry, 1977), or as ‘text’ to be negotiated (Stables and Bishop, 2001). The most dominant message for young people seems to be consumerism (Bowers 1995) fitting as it so neatly does between individualism and the demands of the economy.

As environmental education researchers, it is important to pose further questions, to discuss new theories, and to offer strategies that flow from new ideas. It is equally important to locate such strategies within the reality of people’s lives as they live them. There is a body of environmental education research (for instance, Stables, 2001; Fien et al, 2001) which suggests that the answer to more effective environmental education is an integrated curriculum that cuts across disciplines. This strategy has been recommended for some years without any obvious
changes, particularly in the secondary system: other strategies that will accomplish the same end need to be explored.

Of course, schools are only one source for learning about the environment: perhaps it is time to focus on other sources. The ‘industry’ partners in this study – a museum and government departments connected with environment – all produce environmental programs for schools and all have an integrated policy on the environment. Maybe they could take a lead in environmental education. Likewise, family and other social groups are in a position to take a greater role in the education of young people about the environment. For instance, the rise of ecotourism, and an awareness of the idea that maintaining and enjoying the environment is a family-oriented leisure activity, may also lead to the development of concerned and active social groups (Welford et al., 1999; Bjork, 2000). It may be more effective to take environmental education out of the formal school system and locate it in the community in which young people live.

The traditional distinction made in philosophy between people and nature as object (Marshall, 1992) has become blurred. Ihde (1990) speaks of our basic existence as being technologically textured. He describes the ‘technosphere’ – the relationship between technological artefacts and humans – as a cocoon that gives a sense of familiarity difficult to escape. Young people in the early twenty-first century live not only in landscapes, but in ‘technoscapes’ created by global multimedia (Payne, 1998). This cultural reality is far from the consciousness of most curriculum designers. It constitutes a blind spot for researchers in environmental education that needs to be explored and examined (Gough, 2001). Studies such as the present one will increase our knowledge of the factors influencing young people’s social construction of the environment.

References


[1]

**End Note:**
This paper is based on results from a research project undertaken in New South Wales, Australia, entitled *The Social/Cultural Influences on Environmental Understandings of NSW School Students*. The research project has as industry partners important stakeholders in environmental education: the NSW Environment Protection Authority, the NSW Department of Education and Training, the National Parks and Wildlife Service, the Powerhouse Museum and the Department of Land and Water Conservation. The aim of the project is to ensure the development of more appropriate and student-centred environmental programs and curricula, particularly those developed by the industry partners. The long-term anticipated outcome of the research is to provide information that will assist those who have responsibilities for the protection, restoration and enhancement of the quality of the environment of NSW in the context of ecologically sustainable development.

**APPENDIX**

**Model 1: Binary Logistic Regression 1**

Link Function: Logit

**Response Information**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>phen</td>
<td>2</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1509</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1719</td>
</tr>
</tbody>
</table>

**Factor Information**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>pop-loc</td>
<td>6</td>
<td>0 1 2 3 4 5</td>
</tr>
<tr>
<td>hsp</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>sex</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>q3env</td>
<td>4</td>
<td>0 1 2 3</td>
</tr>
</tbody>
</table>

1719 cases were used
15 cases contained missing values

**Logistic Regression Table**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>Z</th>
<th>P</th>
<th>Odds Ratio</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3.8463</td>
<td>0.5219</td>
<td>-7.37</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pop-loc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.4815</td>
<td>0.2690</td>
<td>1.79</td>
<td>0.073</td>
<td>1.62</td>
<td>0.96</td>
<td>2.74</td>
</tr>
<tr>
<td>2</td>
<td>0.4711</td>
<td>0.3170</td>
<td>1.49</td>
<td>0.137</td>
<td>1.60</td>
<td>0.86</td>
<td>2.98</td>
</tr>
<tr>
<td>3</td>
<td>0.6484</td>
<td>0.2622</td>
<td>2.47</td>
<td>0.013</td>
<td>1.91</td>
<td>1.14</td>
<td>3.20</td>
</tr>
<tr>
<td>4</td>
<td>0.2214</td>
<td>0.3478</td>
<td>0.64</td>
<td>0.524</td>
<td>1.25</td>
<td>0.63</td>
<td>2.47</td>
</tr>
<tr>
<td>5</td>
<td>1.3968</td>
<td>0.4076</td>
<td>3.43</td>
<td>0.001</td>
<td>4.04</td>
<td>1.82</td>
<td>8.99</td>
</tr>
<tr>
<td>hsp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.8379</td>
<td>0.4734</td>
<td>3.88</td>
<td>0.000</td>
<td>6.28</td>
<td>2.48</td>
<td>15.89</td>
</tr>
<tr>
<td>2</td>
<td>0.4032</td>
<td>0.1569</td>
<td>2.57</td>
<td>0.010</td>
<td>1.50</td>
<td>1.10</td>
<td>2.04</td>
</tr>
<tr>
<td>sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.4791</td>
<td>0.2353</td>
<td>2.04</td>
<td>0.042</td>
<td>1.61</td>
<td>1.02</td>
<td>2.56</td>
</tr>
<tr>
<td>2</td>
<td>0.1688</td>
<td>0.2509</td>
<td>0.67</td>
<td>0.501</td>
<td>1.18</td>
<td>0.72</td>
<td>1.94</td>
</tr>
<tr>
<td>3</td>
<td>-0.3570</td>
<td>0.5169</td>
<td>-0.69</td>
<td>0.490</td>
<td>0.70</td>
<td>0.25</td>
<td>1.93</td>
</tr>
</tbody>
</table>
mc(hsps)
(1)  -0.02884  0.08420  -0.34 0.732     0.97     0.82     1.15
(2)  -0.14902  0.07089  -2.10 0.036     0.86     0.75     0.99

Tests for terms with more than 1 degree of freedom

<table>
<thead>
<tr>
<th>Term</th>
<th>Chi-Square</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>pop-loc</td>
<td>14.753</td>
<td>5</td>
<td>0.011</td>
</tr>
<tr>
<td>q3env</td>
<td>7.808</td>
<td>3</td>
<td>0.050</td>
</tr>
<tr>
<td>mc(hsps)</td>
<td>4.538</td>
<td>2</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Log-Likelihood = -588.799
Test that all slopes are zero: G = 98.638, DF = 12, P-Value = 0.000

Goodness-of-Fit Tests

<table>
<thead>
<tr>
<th>Method</th>
<th>Chi-Square</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>421.762</td>
<td>420</td>
<td>0.467</td>
</tr>
<tr>
<td>Deviance</td>
<td>335.699</td>
<td>420</td>
<td>0.999</td>
</tr>
<tr>
<td>Hosmer-Lemeshow</td>
<td>9.061</td>
<td>8</td>
<td>0.337</td>
</tr>
</tbody>
</table>

Model 2: Binary Logistic Regression 2

Link Function: Logit

Response Information

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>phen</td>
<td>2</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1509</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1719</td>
</tr>
</tbody>
</table>

Factor Information

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>lsca</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>hsps</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>sex</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>q3env</td>
<td>4</td>
<td>0 1 2 3</td>
</tr>
</tbody>
</table>

1719 cases were used
15 cases contained missing values

Logistic Regression Table

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>Z</th>
<th>P</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3.3233</td>
<td>0.4612</td>
<td>-7.20</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lsca</td>
<td>0.9367</td>
<td>0.3322</td>
<td>2.82</td>
<td>0.005</td>
<td>2.55</td>
<td>1.33</td>
</tr>
<tr>
<td>hsps</td>
<td>1.7728</td>
<td>0.4724</td>
<td>3.75</td>
<td>0.000</td>
<td>5.89</td>
<td>2.33</td>
</tr>
<tr>
<td>sex</td>
<td>0.3904</td>
<td>0.1559</td>
<td>2.50</td>
<td>0.012</td>
<td>1.48</td>
<td>1.09</td>
</tr>
<tr>
<td>q3env</td>
<td>0.4545</td>
<td>0.2344</td>
<td>1.94</td>
<td>0.052</td>
<td>1.58</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.1355</td>
<td>0.2498</td>
<td>0.54</td>
<td>0.588</td>
<td>1.15</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>-0.4474</td>
<td>0.5141</td>
<td>-0.87</td>
<td>0.384</td>
<td>0.64</td>
<td>0.23</td>
</tr>
<tr>
<td>mc(hsps)</td>
<td>(1)</td>
<td>-0.03655</td>
<td>0.08441</td>
<td>0.43</td>
<td>0.665</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-0.14492</td>
<td>0.07016</td>
<td>2.07</td>
<td>0.039</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Tests for terms with more than 1 degree of freedom

<table>
<thead>
<tr>
<th>Term</th>
<th>Chi-Square</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>q3env</td>
<td>8.093</td>
<td>3</td>
<td>0.044</td>
</tr>
<tr>
<td>mc(hsps)</td>
<td>4.456</td>
<td>2</td>
<td>0.108</td>
</tr>
</tbody>
</table>

Log-Likelihood = -592.510
Test that all slopes are zero: G = 91.215, DF = 8, P-Value = 0.000
Goodness-of-Fit Tests

<table>
<thead>
<tr>
<th>Method</th>
<th>Chi-Square</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>166.612</td>
<td>152</td>
<td>0.198</td>
</tr>
<tr>
<td>Deviance</td>
<td>152.524</td>
<td>152</td>
<td>0.473</td>
</tr>
<tr>
<td>Hosmer-Lemeshow</td>
<td>7.455</td>
<td>8</td>
<td>0.488</td>
</tr>
</tbody>
</table>