

Learning Statistics Using Motivational Videos, Real Data and Free Software

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

ABSTRACT

A website and software products with the potential to raise the profile of statistics in society are described. The website has links to case study videos describing contexts, study designs, data files and lessons using new software for data exploration and analysis. Case study videos dealing with current research applying statistics have been selected to motivate discussion in class, and further “hands on” learning can be achieved through use of the software. During the development phase in New Zealand in 2010 the software was trialed and student and teacher experiences are reported. A full day professional development workshop for teachers involving lessons using the software was recorded and these are on the website to assist teachers and students commencing use of the software. The software is free for use by teachers and students in schools, in universities and in homes. The procedure for obtaining a license is outlined.

Key words – online resources, motivational videos, real data, free software

1. INTRODUCTION

The teaching of statistical literacy, statistical practice, design of studies, sampling procedures and data analysis is being extended in New Zealand schools. A learning sequence, Mathematics and Statistics, beginning in the first year of school at age 5 culminates in a subject called Statistics and Modeling in the final year at high school for students aged 17-18 (Ministry of Education, 2007). To assist with teaching this subject, statistical software has been developed to help students explore interesting data and to help teachers prepare lessons and understand unfamiliar topics such as bootstrapping planned for introduction to the final year of the curriculum sequence in 2013. A website (<http://www.maths.otago.ac.nz/videos/statistics>) hosting lesson plans that integrate the recently developed software with videos and accompanying data is described.

The new software, GenStat for Teaching and Learning (GTL Schools and GTL Undergraduate), is menu driven and free in schools, in universities and in homes. GTL covers the statistical techniques in the school curriculum, advanced statistical techniques in undergraduate university subjects, and leads to a commercial version of GenStat for professionals and post graduates. Continuity provided by a comprehensive free package is an innovation beneficial for the transition from school to university and beyond. Every school and university can access GTL upon application for a license to VSNi (short for ViSioN International), at the website <http://www.vsni.co.uk/software/genstat-teaching/> or by emailing info@vsni.co.uk.

The lesson plans accompanying the videos are based on “authentic” data generated from current disciplinary research. Videos of researchers discussing the contexts of their studies are presented along with lessons which are deliberately flexible and relevant for school use or university undergraduate statistics courses. The data can be analyzed independently of GTL using Excel or other programs but the main objective of the website is to provide training in the use of GTL, accomplished by working through lessons and following a set of professional development videos once GTL has been installed. The videos and data are described in Section 2, GTL software in Section 3, five GTL lessons based on the videos in Section 4 with discussion and conclusions about the benefits of the videos and the future of the software in Sections 5 and 6.

2. CASE STUDIES: VIDEOS, DATA AND LESSONS FOR LEARNING

Teachers and students seek access to resources which have interesting applications of statistics. CensusAtSchool (Statistics New Zealand, 2011) provides survey-like data and random samples, GAISE (2010) is a curriculum description with some suggested activities, CAUSE (2011) is a compendium of resources and the video series *Against All Odds* (Moore, 1989) presents entertaining videos intended to provide case studies using statistics. The Data and Story Library (DASL, 2011), is an online library of data files and stories that illustrate the use of basic statistics methods. Data in DASL come from a

variety of topics so that statistics teachers can find real-world examples of interest to their students.

The videos and data files presented in this paper have the same objective as DASL but with the additional feature of showing researchers discussing the context of their own study. Nineteen researchers have been recorded by the Higher Education Development Centre at the University of Otago. Viewed as a community project, free recording of further studies is continuing whenever data with an interesting story is identified. The videos are 10-25 minutes long, with longer videos ideally taken in sections. A teacher can use them to generate discussion about design principles, sampling methods and procedures for data exploration, summary and analysis. The study lessons all use graphical procedures and data summary statistics. Additional techniques required for each video are noted on the website in the introduction to each video, thus providing guidance to analysis options that can be chosen by a teacher from the GTL menu. Tasks are found by selecting links associated with each video to access data files and lessons in both Excel and GTL.

The videos listed below cover applications from biology, ecology, epidemiology, human nutrition, psychology, marketing, tourism, official statistics, health and agriculture:

- 1 Estimation of cockle biomass in a tidal inlet;
- 2 Use of infrared thermography for non-invasive assessment of animal welfare;
- 3 Maui's dolphin: uncovering a new subspecies using museum records;
- 4 Effect of circumcision on herpes (and other sexually transmitted diseases);
- 5 Is iron deficiency common among New Zealand infants and toddlers? ;
- 6 Measuring public opinion on support for public funding of the Otago Stadium;
- 7 Effects of satellite transmitters on titi (a migratory New Zealand bird);
- 8 Exploring data on a rare threatened bird – the rock wren;
- 9 Research trip to the Himalayas: adjustment to high altitude;
- 10 Impact of group rejection from physical activity on student self esteem;
- 11 Effects of investigator disturbance on hatching success in endangered takahe;
- 12 Factors affecting youth motor accidents in New Zealand;
- 13 Trace metal levels in New Zealand dredge oysters;
- 14 Property sales in North Shore City – 1999 to December 2008;
- 15 An exploration of global rotavirus deaths;
- 16 Activity hit parade of international tourists to New Zealand, what is popular;
- 17 Factors affecting aviation accidents and incidents;
- 18 Time series analysis of retail sales;
- 19 Natural resource accounts; aligning economic and environmental performance.

Cockle biomass, infrared thermography, the Maui dolphin study, infant iron deficiency and adjustment to high altitude have been used in first year undergraduate courses at the University of Otago but are also suitable at school level. Studies of the effects of circumcision on sexually transmitted diseases and factors affecting public support for an expensive and controversial stadium in Dunedin city have been used in an advanced undergraduate course on regression modeling. The data on attitudes of international

tourists to attractions in New Zealand have provided project material for an advanced course on multivariate statistics. The level of interest generated by this case study among the students is such that it has resulted in preparation of a joint publication by one student and the researcher presenting the video.

3. THE GENSTAT SOFTWARE

The development of GenStat (General Statistics) began at the Rothamsted Agricultural Research Station at Harpenden in the United Kingdom in 1968 making it one of the first major statistical packages (<http://www.vsni.co.uk/about-vsni/history-of-genstat/>). GenStat was distributed outside Rothamsted during the 1970s and, in 1979, distribution was taken over by the Numerical Algorithms Group (NAG), one of the oldest technical computing companies. VSNi was formed in 2000 from Rothamsted and NAG. GenStat has been developed by a Statistics Department whose members have been responsible for many widely used methods in applied statistics.

GenStat Discovery (<http://www.vsni.co.uk/software/genstat-discovery/>) based on an early version of the package GenStat, has been free to not-for-profit research organizations, charities and educational institutes in the developing world since 2003 (Andrews, 2009).

During 2010 VSNi statisticians and programmers, in consultation with New Zealand statistics school teachers and lecturers, developed subsets of the menu options of GenStat Version 13 (now Version 14) for use in New Zealand. As a consequence the software packages GTL Schools and GTL Undergraduate evolved providing free data analysis tools for school and undergraduate university use both within the educational institute and at home (Andrews, 2010). Options requested by teachers included all graphical procedures from the GenStat professional version, bivariate data analysis, simple linear and non linear model fitting including residual investigation, estimation and hypothesis testing for means, differences and proportions, chi square tests, time series analysis, bootstrapping and analysis of variance. Design considerations related to ease of use and statistics content at such a non-specialized level of instruction supported menu driven software and Excel-type spread sheets. Ability to transform data easily, ability to upload data from any package such as Excel, and an ability to paste output into a word processing document for editing and report preparation were desirable. GTL was seen as a pathway to the professional use of GenStat 14 which could be acquired at a cost similar to that for other large menu driven packages.

For comparison, IBM SPSS (2011), SAS (2011) and Minitab (2007) are widely available in teaching institutions which have purchased them but they cannot be used on home computers without purchasing a separate license. Fathom (Finzer, 2005) and Tinkerplots (Konold and Miller, 2004) are modestly priced in comparison but do not continue for university use. StatCrunch (2009) is worth considering, but it is not as convenient when moving to more sophisticated analyses, although it is inexpensive provided a text is purchased with the program. R (2011) is open source and a competitor of GTL, although there are complications for school students and the majority of undergraduates in subjects other than statistics when programming and syntax are required. Menu-based programs

for R are available such as Rcommander (Fox, 2005), R Excel (Heiberger and Neuwirth, 2009) and R Deducer (Fellows, 2011). These can be used without the technical programming and syntax of R intruding. Personal experience with R Excel and R Deducer indicates they are complicated for students. Personal experience shows that Rcommander is easy to use but with fewer options than GTL and so offers no advantage over GTL other than a convenience of uploading the syntax for later use of the R programming language. Excel (2003) is not well regarded for statistical analysis intended for the school curriculum and for more advanced analyses. Excel does have benefits for data management and Excel files as well as files from other packages can be uploaded easily into GTL for analysis.

3.1 The Initial Assessment of GenStat Teaching and Learning: Schools (GTL Schools)

In February 2010, GTL Schools was shown to members of the Education Committee of the New Zealand Statistics Association, staff from the New Zealand Ministry of Education and staff from the Wellington College of Education. After a favourable reception it was decided to proceed to the next stage by approaching practicing school teachers to trial the software in the classroom and to receive opinions about its use. A large, typical New Zealand girls' high school, Otago Girls High School, in a slightly above average socio economic region and with the full range of pupil ability was chosen as a pilot school.

The Head of Mathematics at the school was a graduate in mathematics, statistics and computer science and had used R, SPSS, SAS, Minitab and Excel. An ideal person to trial the software, she developed lessons covering regression and time series, sections of the school curriculum which were about to be taught in her school. Teachers and the girls in the school reported they liked GTL Schools. It started to be used at home with work being submitted from home for marking (Chapman et al., 2010). Later in 2010 girls in the junior school aged 13 to 14 were introduced to the software with some finding GTL Schools easier to use than Excel. Biology teachers started to use the software to analyze project data collected in their final school year classes (Chapman et al., 2010). A lesson plan and instructions for Biology Statistics using GTL is now available on the website <http://www.maths.otago.ac.nz/videos/statistics> in a section where experiences and ideas about using GTL Schools can be shared.

3.2 Professional Development for GenStat: Schools (GTL Schools)

The early positive assessments of the educational benefits at Otago Girls' High School were a result of an enthusiastic Head of Department preparing lessons for her colleagues. This identified a need for teacher professional development to extend use of the software to other schools. The Otago Mathematics Teachers' Association held a full day GTL Schools workshop for 35 teachers in mid November 2010 in a computer laboratory at the University of Otago Department of Mathematics and Statistics. The workshop was recorded by the Higher Education Development Centre at the University, resulting in three hours of instructional videos with matching discussion showing the options in GTL

Schools. Details are on the website <http://www.maths.otago.ac.nz/videos/statistics>. Lesson plans with screen shots which can be printed and followed as the videos are viewed provide a starting point for an intending user of GTL Schools. The sessions recorded in the videos relate to the final year school statistics curriculum in New Zealand for students aged 17 to 18 years but the topics included are of general interest. The sessions are:

1. Introduction to free-to-use menu driven GenStat and demonstration describing the spreadsheet, uploading data, graphical work, estimation and pasting results to a word document using the dolphin case study and video (43 minutes);
2. Bivariate data analysis (31 minutes);
3. A follow-up lecture on non-linear models (38 minutes);
4. A lesson on time series analysis reflecting the unit standard at Year 13 in New Zealand (28 minutes);
5. Piecewise linear regression options and bootstrap analysis using the rock wren study and video (41 minutes);
6. Discussion and teacher opinions (12 minutes).

Another 35 teachers attended a shorter presentation and discussion at the College of Education at the University of Canterbury on November 30, 2010.

3.3 Teacher Reaction to GenStat: Schools (GTL Schools)

Five teachers at Otago Girls High School used GTL Schools in the preliminary trials, either in the final school year (ages 17-18) statistics classes or in the middle school (ages 15-16) mathematics classes as a replacement for Excel. In personal interviews they report experiences as follows:

Teacher 1 (first time user with a very low ability class of girls aged 15 to 16):

“GTL was very helpful in the bottom year 11 class. The graphs were easy to produce and we could focus on interpretation rather than graph production. Having the trend line also helped.”

Teacher 2 (first time user with class of girls aged 15 to 16):

“I found GenStat very easy to learn how to use before doing it with my class for the bivariate data internal. The girls picked up the program very quickly and were confident moving around between the graph, the output window etc. They also had no problems getting their data from an Excel file and putting it into GenStat. I would definitely use it again, and am keen to try it out with the juniors (aged 13-14) when they do their stats project.”

Teacher 3 (first time user with low ability class aged 15 to 16):

“GTL was easy for teachers and students to use. Students enjoyed learning to use a new computer software program, in fact finding it fun. It was great to get students into the computer labs for variation and scatter plot graphs which were very easy. There were some technical problems accessing Excel spreadsheets and log-ons for students but these were not GenStat problems. Overall implementation of the lesson was not too difficult but it was busy to manage.”

Teacher 4 (user in middle and final school years ages 15-18):

“The use of GenStat for bivariate data in the middle school was a huge success in terms of student learning about bivariate data theory. I investigated linear modelling and the meaning of the correlation coefficient in some detail. I think it will be useful background for our final school year bivariate data lesson. I was somewhat reluctant to change initially with my final year class as I was familiar with Excel. However I picked up GTL quickly and the girls found having the lessons beside them as they worked on the computers was good. They submitted their work electronically. With Excel the students achieved gradings but did not really know what they had done. Genstat has enlivened and enriched the teaching in my classes.”

Teacher 5 (user in middle and final school years ages 15-18):

“Students picked it up very easily, we only spent about two hours in the computer room before the assessment and no student had trouble with using GenStat in the assessment though we needed to do some work checking files before importing into GenStat. It does force the students to think more about the statistical process and I am looking forward to doing it again next year. We didn’t do enough practice entering data into GenStat which caused some problems which can be corrected next year.”

Feedback from teachers at the GenStat workshops in Dunedin and Christchurch was positive. Teachers emphasized that professional development workshops were essential for the successful introduction of GTL Schools more widely in New Zealand high schools. Further presentations have been made to the Auckland Mathematics Teachers’ Association and the North Auckland Mathematics Teachers’ Association in early 2011 while in July 2011 two workshops using GTL Schools were scheduled during the biennial New Zealand Association of Mathematics Teachers Conference in Dunedin. A further session was provided at a Schools day at Auckland University in November 2011.

3.4 Student Reaction to GenStat: Schools (GTL Schools)

Two surveys of student opinion of GTL Schools have been carried out at Otago Girls’ High School. In both cases the school students were familiar with Excel and GTL. The information collected is being used for future development of statistics learning programs in the school using the software.

The first survey for a group of 41 students in their final year at school (ages 17-18) was carried out upon completion of a lesson on time series analysis. The five questions produced the following responses:

Question 1: would you have preferred to use a different program?		
Yes		5
No		36
Question 2: how did you find GenStat to use?		
Easy		5
Fairly easy		15
Average		15
Fairly hard		4
Hard		2
Question 3: did you find any part of the analysis using GenStat hard?		
Yes		20
No		21
Question 4: if you answered yes to the last question, select the parts you found most difficult. If you answered no, select not applicable.		
Finding the moving average		0
Graphing the raw data, the centred moving mean and trend		1
Completing linear regression and saving the fitted values		1
Finding the individual seasonal value		7
Finding the average seasonal value		5
Not applicable		27
Question 5: are there any parts you thought could have been improved? If you think they were all fine, select not applicable.		
Finding the moving average		0
Graphing the raw data, the centred moving mean and trend		3
Completing linear regression and saving the fitted values		3
Finding the individual seasonal value		4
Finding the average seasonal value		4
Not applicable		27

The second survey was carried in a class of 19 final school year biology students aged 17-18. The subject biology includes a substantial project which can involve the collection and analysis of data. The head of mathematics at the school prepared a document on statistics for biology students and staff using GTL. In the class of 19 students, nine used GTL for their analyses and ten used Excel. Five of the nine who used GTL had already used it in their statistics classes. The other four said “it was easier for line graphs” and “faster once knowing how to use it” but conceded “GTL has a printout which is harder to understand” and “Excel is easier for some calculations such as a standard deviation”. The ten students who used Excel did so for several reasons such as “data was already in Excel”, “Excel was available at home”, “they were familiar with Excel which they found easy to use” or “they could not get GTL to work”. Two of the ten stated they would have

preferred to use GTL now they know about it. These comments identify modifications for GTL use next year.

3.5 Development of GenStat Teaching and Learning: Undergraduate (GTL Undergraduate)

While GTL Schools was being developed, VSNi decided to extend the free software for use at the undergraduate level in universities using a larger subset of options from the menu in the full commercial package. The menu was extended to include multiple linear regression, multiple logistic regression, log linear models, analysis of variance procedures at least for two factor experiments involving interactions, power calculations, multivariate analyses such as principal components, factor analysis, discrimination, repeated measures and mixed models, multidimensional scaling, cluster analysis, correspondence analysis and canonical correlation analysis.

It is easy to move between GTL Schools mode and GTL Undergraduate mode by first launching GenStat and under **Tools** and **Options** choosing the **General** tab, selecting the GenStat version you wish to use. The instructions tell you to close GenStat but when reopening GenStat the alternative mode is activated.

3.6 Accessing GTL Schools and GTL Undergraduate

GTL Schools and GTL Undergraduate can be installed either in a school or a university or on a home computer once the IT coordinator at the school or university has received the license. The process in place for obtaining a download and license is as follows:

(i) A teacher or lecturer who wants to trial GTL should ask their IT coordinator to contact VSNi at the website <http://www.vsni.co.uk/software/genstat-teaching/> and apply for a download. Alternatively the contact email info@vsni.co.uk can be used or call +44(0)1442 450230;

(ii) After registering interest with VSNi for a trial version, VSNi will send a token (use and time limited) by email for downloading GTL together with a time limit trial license (usually 30 days) to the IT coordinator;

(iii) After a trial, if it is decided that you want to use the software, VSNi will email a EULA (End User License Agreement) for signature to the IT coordinator;

(iv) Once VSNi receives a signed EULA, a token will be emailed to the IT coordinator at the school or university along with the download and a time-limited license key based on academic year. Individual students at the site can now be licensed freely via the network license.

By September 2011, 11 schools had applied for a EULA and 55 additional New Zealand schools had applied for trial licenses (personal communication from Stewart Andrews at VSNi).

3.7 Limitations on Data File Size

The schools and undergraduate modes can be accessed freely. Both programs accept 80 variables, the schools mode allows 2500 cases and the university mode 500 cases, these cases either being selected randomly or taken as the first block of cases. These limitations are not restrictive for the analyses and explorations which will be carried out on data at school and undergraduate university. GTL schools uses a simplified menu covering procedures in the school curriculum while GTL undergraduate has a wider range of menu options for undergraduate years. If at a university a student has a data set which has more than 500 cases it is easy to move to schools mode which will accept the greater sample size while if at school and a student wants to explore some of the more advanced techniques it is possible to move to the undergraduate mode. How to move between the two free modes is described in Section 3.5. The greater limitation on sample size in the undergraduate mode is to protect for professional use and post graduate research the commercial package GenStat 14 which follows on naturally from the two free GTL versions.

4. GENSTAT LESSONS FOR CASE STUDY VIDEOS: SOME EXAMPLES

Two honours graduates in Statistics at the University of Otago have helped develop lessons using GTL Schools and GTL Undergraduate to carry out tasks for the first 13 of the case study videos listed in Section 2. The GTL lessons with instructions can be found on the website <http://www.maths.otago.ac.nz/videos/statistics> by selecting a video from the list then clicking on the GenStat.pdf link which appears. It is recommended that in the absence of a GTL manual a copy of a lesson should be printed before working through the lesson plan. The videos describe a study placing data in context. The lessons use the data and are designed as a learning experience for teachers or students to work through by mimicking the analyses themselves once they have GTL installed on their computer. Alternatively, a teacher can use one of the lessons in class and extend the exploration of the problem by asking for more analyses. The following five lessons have been selected to cover many of the options available in the menus. Another eight of the videos have GTL lesson plans which can be followed.

To illustrate the launch of GTL Schools the screen shots of the opening windows of GTL Schools are shown for the example on the Maui dolphins described in video 3. The GenStat start page shown in Figure 1 has information about **Getting Started** including a description of the menu interface, how to manipulate the spreadsheet, how to install a license and how to upload an Excel file. A panel for recently used data sets and files is also shown and underneath this are three options which provide access to data including files used in the videos described in Section 2. Selecting **Example data sets** and scrolling down **Filter by topic** to **NZ School Data sets from J Harraway University of Otago** and selecting **Dolphin.gsh** as shown in Figure 2 is the procedure for loading the dolphin data and all other data files. The spreadsheet for the dolphin data is shown in Figure 3.

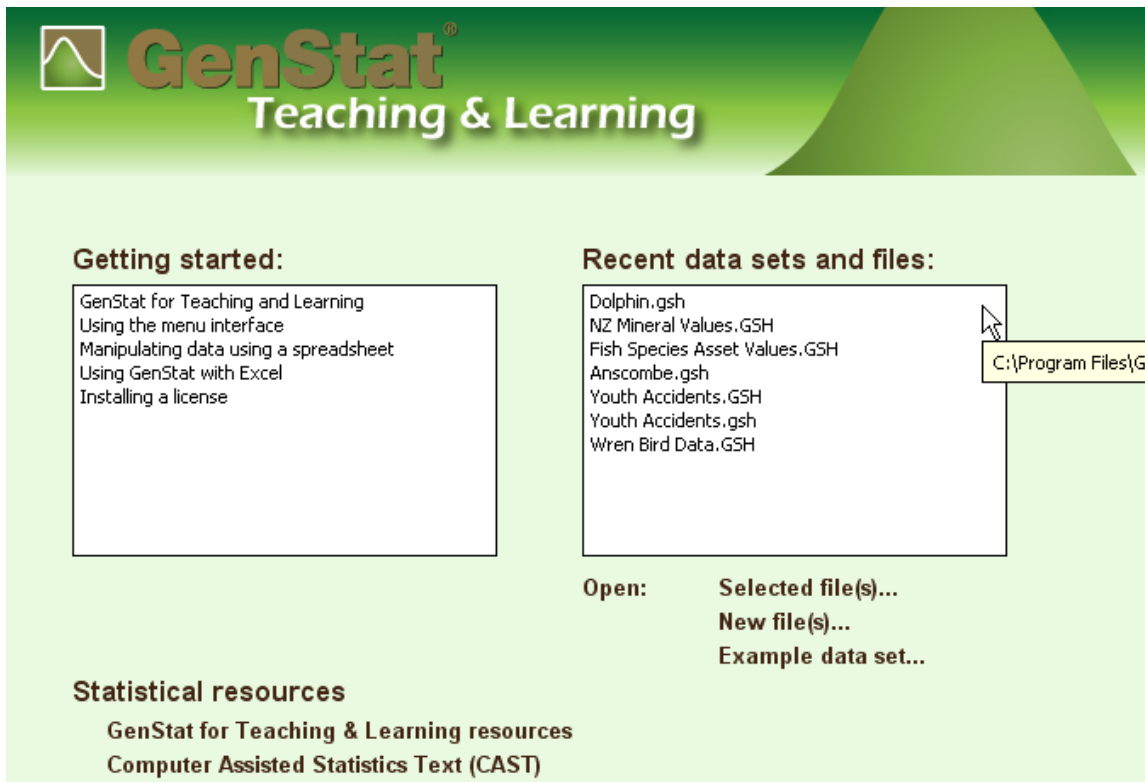


Figure 1 GenStat Start Page

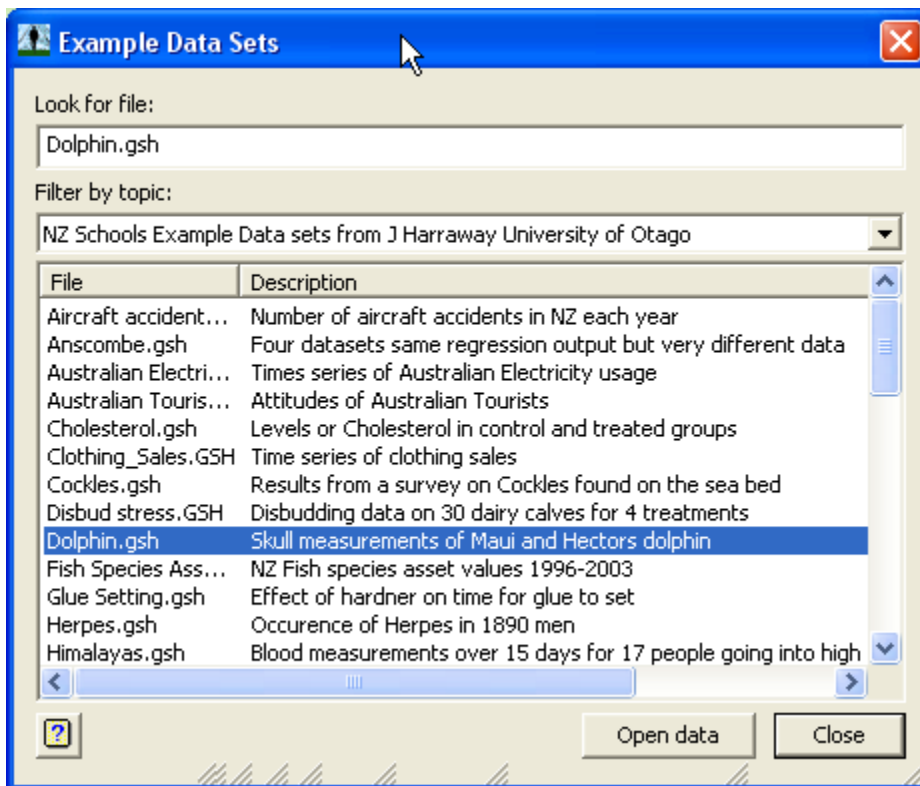


Figure 2 Example Data Sets

4.1 Maui dolphins – uncovering a new sub species

The study video which features the biologist lead researcher, Mr Adam Smith, who collected the data, can be played and [Dolphin-GenStat.pdf](#) selected on the video website to obtain the lesson for GTL. This study also provides the first recorded session for the introductory professional development workshop. It investigates whether the endangered Maui and Hector dolphins are different species even though they are small dolphins with similar appearance; Maui dolphins are found around North Island and Hector dolphins around South Island of New Zealand. Earlier genetic evidence suggests this to be the case (Baker et al., 2002). The video discusses the ecology of the dolphins then describes six measurements recorded on 46 Hector skulls in South Island museums and 13 Maui skulls in North Island museums. A skull diagram in the video shows the six measurements.

Row	Island	X1	X2	X3	X4	X5	X6
1	North	284	57	143	148	232	83
2	North	287	59	147	154	236	84
3	North	296	57	146	153	238	82
4	North	300	61	146	153	246	85
5	North	302	62	153	162	248	86
6	North	302	57	155	153	241	81
7	North	304	58	153	154	246	84
8	North	310	62	154	163	245	90
9	North	310	63	157	159	245	86
10	North	313	63	159	160	255	86
11	North	315	59	153	158	252	89
12	North	317	63	165	167	257	94
13	North	319	64	158	162	255	94
14	South	264	47	128	134	205	71
15	South	265	49	136	137	210	74
16	South	276	49	135	142	220	77
17	South	277	50	135	144	219	78
18	South	280	51	139	145	219	75
19	South	282	52	138	139	225	77
20	South	284	50	141	143	221	80
21	South	284	50	140	137	225	81
22	South	284	48	138	139	222	78
23	South	284	49	134	136	222	76
24	South	289	55	145	146	230	80

Figure 3 Dolphin data spreadsheet

Procedures carried out using the menu options and following the prepared lesson plan involve summary tables and graphs with histograms and box whisker plots, means and confidence intervals for North and South Island comparisons, scatter plots and correlations between the six continuous variables. Students discover clear evidence that North Island Maui skulls are larger than South Island Hector skulls providing additional evidence of a new sub species. The reasons for this can be discussed. One explanation goes back 15000 years when the two main islands of New Zealand separated. The Maui and Hector dolphins are in-shore animals and there was probably no contact between the two animals once the islands separated. It has also been hypothesized that the North Island waters are warmer with greater abundance of food resulting in the larger North Island animals.

Another feature illustrated with this data set is the launch of GTL Undergraduate which allows additional analyses not available in the schools version. By switching from the **Schools Edition** to the **Undergraduate Edition** of GTL, students can carry out a principal component analysis to combine all six measurements. Principal components are not in the school syllabus but are taught in multivariate statistics courses at a university. There are simple interpretations of the first and second principal components representing the classic size and shape measures in this case. The scatter plot of the first and second principal component scores produces further evidence to support different species based on all six measures combined.

4.2 Support for public funding of the Otago Stadium; a survey of city residents

The lead researcher for this study is a market analyst, Dr John Williams. Data and a lesson plan can be found by activating [Stadium-GenStat.pdf](#) on the website and selecting Stadium.gsh which displays 1950 responses to a survey of attitudes in a city to a controversial 200 million dollar covered sports stadium development. The first half of the video discusses question formulation, data collection, data analysis and conclusions reached after considering issues that could bias results. Five thousand questionnaires were mailed to a random sample of the target population, the adult residents of Dunedin City. Replies from 2100 respondents to the questionnaire on this emotive issue were received. The research question was formulated based on conflicting claims by the football union that the new stadium was essential and by the anti-stadium citizens that the existing sports field was adequate for international fixtures. A sampling frame and a sampling method are described. Data gathering used a mail out self-completion questionnaire with 10 questions, one relating to support or no support for the stadium and the others demographic relating to sex, age, financial circumstances, home ownership and education background. Students discuss data cleaning which reduces the cases to 1950 after allowing for missing values. The 1950 cases are analysed using GTL Schools. Tasks accomplished with this lesson include calculating the crude proportion and confidence

interval for the respondents who do not support public funding of the stadium, constructing cross tables of *Sex* against *Support*, *Household income* against *Support* and *Home ownership* against *Support* all with chi square analyses. A sample of 500 needs to be taken if GTL Undergraduate is to be used for making predictions using a logistic or a multinomial logistic regression model.

The second half of the video has discussion on post stratification, threats to validity resulting from unit non-response bias, item non-response bias, instrument bias in relation to the question asked, and data entry errors. These concepts extend beyond the syllabus of a course at school but the underlying ideas discussed in the video can be debated. This study is a relevant exercise illustrating methods for dealing with large sample surveys which frequently end up being reported in the media.

4.3 Trace metal levels in New Zealand dredge oysters; identifying oyster location

The lead researcher for this study is a marine chemist, Dr Barrie Peake. Data and a lesson plan can be found by activating [Oysters-GenStat.pdf](#) on the website and selecting Oysters.gsh which displays the values of four trace metals, zinc, copper, cadmium and manganese, measured on 111 samples of dredge oysters collected from 14 sites around the South Island of New Zealand. The concentrations of the trace metals depend on oyster habitat and provide scientists with a method of determining an area from which oysters may have been harvested. Oysters purchased in retail outlets can have their origin identified, leading to a prosecution if it is found that the oysters come from a region from which it is illegal to harvest oysters such as the prized Bluff dredge oyster. But more important is origin identification of oysters which may have caused food poisoning.

There are many exploratory tasks that can be carried out on the oysters. One simple GenStat graph which has an interesting associated story is shown for illustration in Figure 4 where the zinc levels in the oysters at the 14 sampled sites are summarised in simple box and whisker plots. Two of the sites, one in Bluff Harbour (BH) and the other in Otago Harbour (SB), have unusually high zinc levels in the samples collected at these sites. Large metropolitan areas are nearby and runoff from the galvanized iron roofs of houses into the harbours raises zinc levels. Similar boxplots for the other three trace metal values can be produced, and one of these displays another unusual pattern which can be interpreted by listening to the story in the video.

A principal component analysis can be performed by switching again to the **Undergraduate Edition** of GTL. There is no need to select a random sample with fewer than 500 cases present. The principal components combine all the information from the four trace metals as can be seen in the plot of the first two principal component scores and the means of these scores at each site shown in the lesson plan printout.

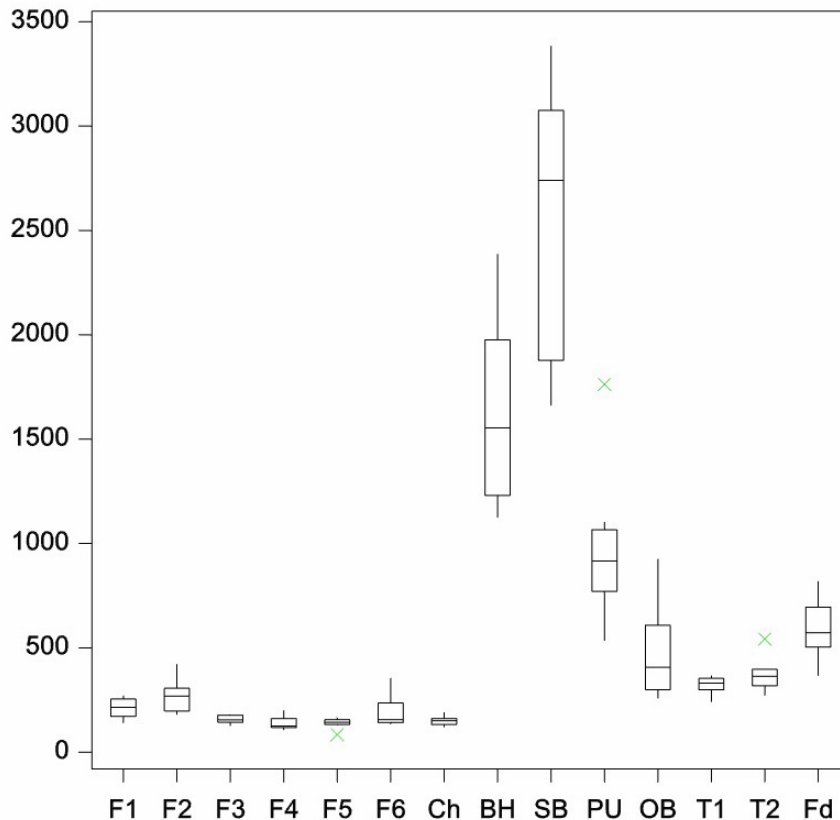


Figure 4 Zinc levels at fourteen South Island sites

Trace element data has many applications especially for identifying food origin. Up to 40 trace elements can be measured and in this circumstance principal component analysis and discriminant function analysis can be used. Two further applications are mentioned at the end of the video. Another involved a court case relating to the origin and therefore ownership of a valuable jade boulder. It turned out that the boulder had been stolen from a region controlled by the rightful owner (Campbell et al., 2009).

4.4 Cockle, where art thou? Investigation of shell fish biomass in a tidal inlet

The lead researcher, Dr Austina Clark, shown in the video, is a statistician using this project as a practical assignment in a post graduate statistics course for marine biologists. A cockle or clam is an edible shell fish living in the sand of tidal inlets. Data can be loaded by selecting Cockles.gsh on the website and the lesson activated at [Cockles-GenStat.pdf](#).

The study design is described in the video. The purpose is to investigate how the growth of cockles is related to proximity to sea water. A stratified sample of sites from different positions in the sand banks of a tidal, sandy inlet is chosen. Counts and weights of cockles at these sites are recorded by the students working in groups. Various hypotheses

are considered such as whether there is a relationship between the density (weight) and the number of cockles, whether cockle biomass is affected by the distance from the sea, whether there are differences between the people collecting the data and whether biomass and density are correlated. Conclusions are drawn from boxplots, confidence intervals for means and independent samples t -tests.

Switching to the **Undergraduate Edition** of GTL allows a multiple regression analysis with biomass as the response and density, water level and their interaction as predictors. Before doing this students must convert water level from a factor to a variate as described in the lesson plan.

4.5 Disappearing rock wren in Fiordland: exploring data on a rare threatened bird

The lead researcher, Mr Ian Westbrooke, is a conservation statistician. For this lesson, students are asked to ignore the tasks for the Excel analysis and instead choose [Wren-GenStat.pdf](#) on the website and Wren Bird Data.gsh to obtain the lessons for GTL.

This study is also part of the fifth session for the professional development workshop described in Section 3.2; it includes t -tests for paired and unpaired data, an introduction to the bootstrap and the use of GTL to carry out the bootstrap procedure on small samples for both paired and unpaired data. The tests first involve matched data, as the 1985 and 2005 measurements were taken at the same grid points; but assuming that the 2005 measurements were instead taken on different grids randomly chosen, the independent samples t -test must then be used. The results from the t -tests and resulting confidence intervals can be compared with the bootstrap confidence intervals in both cases.

5. DISCUSSION

It has been a challenge developing the two aspects of the website described in this paper. My personal belief, supported by comments received from students over the years, has convinced me that students begin to enjoy statistics classes when they see relevant data in interesting contexts. I therefore identified the research studies described in the website and arranged the recording of the nineteen researchers followed by the editing of the videos. At the same time, as Chair of the Programme Committee of ICOTS8, I was approached by VSN International to present a workshop on proposed new free software. I recognised the potential of the concept immediately. I was pleased, after discussion, to find that VSNi were prepared to include in their menu all procedures relevant to my teaching at undergraduate university level and to the teaching of the new schools statistics curriculum in New Zealand. I arranged presentations of the software for education authorities at Statistics New Zealand early in 2010 and coordinated initial trials in Otago Schools later in 2010. These were followed by a series of professional development workshops which I arranged at the University of Otago. Realising that lessons for classroom use were essential I then employed two honours statistics students to develop, under my supervision, the lessons on the website associated with 13 of the videos.

The website has links to motivational videos, data and lessons based on the data using new software, GenStat Teaching and Learning (GTL) which reflects the New Zealand school curriculum in Mathematics and Statistics (Ministry of Education, 2007). The procedure for installing the software has been outlined and once installed is free to use in a school, in a university for undergraduate study and in the home on a personal computer. It is possible to start training in the use of GTL by mimicking the lessons outlined with each video and the recordings of the sessions at the professional development workshop held at the University of Otago in November 2010. Teacher responses have been positive. A section on the website is reserved for reporting teacher experiences and successful lesson plans.

There are menus for standard analyses: basic statistics, statistical tests, bootstrapping, regression, generalized linear models, nonlinear models, analysis of variance, design of experiments, sample size, REML analysis of linear mixed models, multivariate analysis, six sigma, survival analysis, time series and repeated measurements. GTL can be used in either a Schools mode or an Undergraduate mode. The Schools mode has a simpler menu bar, concealing for simplicity analyses that are not needed until university. But this can facilitate the teaching of extension topics in schools and allow interested pupils to sample undergraduate facilities in advance of attending university. Data from other packages such as Excel, SPSS or R can be uploaded directly into GTL.

The software has been trialed successfully at a large girls' high school in New Zealand. The first phase of the trial was in the final school year subject Statistics and Modeling, with classes of girls aged 17 to 18 years. In a survey of teachers, one teacher reported that GTL had "enlivened and enriched the teaching of statistics in his class". The software was used in class and at home with work either completed in class or submitted from home for marking. Family members could observe the process at home and this provides potential for the wider community to gain an appreciation of the relevance of statistics.

GTL has been introduced into junior classes with girls aged 13 to 14 years with half the girls, after their first experimental use of GTL, reporting that they prefer GTL to Excel. Biology teachers have their students carrying out project work in the final school year. The projects involve data collection, and the biology teachers encouraged their students to use GTL Schools to analyze their project data. Teachers report that most school students are computer literate, often displaying better skills than themselves at entering data, manipulating windows, saving results to word documents and editing word documents.

In a personal communication a Secondary School Mathematics Curriculum and Numeracy Advisor, David Phillipps, who attended the professional development sessions in Dunedin and Christchurch said he was impressed that the new software was free and had been trialed successfully in several classes at different levels at a large high school. He went on to say:

"the software evokes the pedagogy by forcing students to engage in thinking about concepts; it is suited to the aligned statistics NCEA (National Certificate of Educational

Achievement) standards and school syllabus; it has the software support of an international company dealing with real world statistical analysis; it has the enthusiastic backing of the South Island secondary mathematics advisors and it is easy to use”.

The development of GenStat at Rothamsted Agriculture Research Station began in 1968 making it an early statistics package. Its origin is based on good statistics practice following the work of R.A. Fisher, Frank Yates and in particular Professor John Nelder, who followed Frank Yates as Head of Statistics at Rothamsted in 1968. Professor Roger Payne took over leadership of the GenStat activity when Nelder retired in 1985. Now the software is controlled by VSNi, which was formed by a merger of Rothamsted Research and the Numerical Algorithms Group at Oxford University. The current GenStat program developers and statisticians at VSNi have provided for New Zealand schools a simple menu-based system highlighting the techniques that students encounter at school. The free advanced university mode unlocks slightly more detailed and complex statistics in line with the types of analysis studied by undergraduates.

The limitations on use of GTL are not serious. The limit of 2500 cases and 80 variables in Schools mode is enough to allow most data sets at this level to be explored and analyzed. The limit of 500 cases for Undergraduate mode is also not a severe problem for the types of data that arise at the undergraduate university level. These limitations have been imposed to protect the professional version of GenStat as some advanced procedures are included in the free undergraduate version. If a data set exceeds either of these limits, the program selects in the case of Schools mode either the first 2500 cases or, if requested, 2500 cases randomly. For Undergraduate mode the figure 2500 is replaced by 500.

The software is only accessible using a PC platform. This means about 90% of New Zealand schools can access the software and some others if they have a PC emulator for a Macintosh. Consideration is being given by VSNi to providing resources which will allow GTL software to be developed for Mac or Linux users.

There is debate about the use of open source statistical software such as R with many statisticians claiming that open source is the way of the future. While not open source, GTL Schools and GTL Undergraduate come close, at least with respect to price, by being free to educational establishments and their students. GTL Schools and GTL Undergraduate have advantages for students, in that the software is menu driven, rather than requiring a command line structure and syntax, which can be complicated for effective school use and even early use at a university. After all, the majority of students who attend statistics classes at a university are not training to be statisticians. Of the menu versions of R personal experience would indicate that RCommander has potential due to ease of use. But a major advantage of GTL over R is that, in the words of David Phillipps, “it has the software support of an international company dealing with real world statistical analysis”.

Further videos and GTL lessons are being developed and will be found soon on the website <http://www.maths.otago.ac.nz/videos/statistics> as they are completed over the next few months. There are three studies being prepared. The first involves data from

forensic science, the second is a study of the side effects of a cholesterol lowering drug and the third is an investigation into the profiles of women who would consume alcohol when pregnant. The predictor variables in this last study are factors such as ethnicity, education level, current alcohol consumption, household income and smoking status. There is no limit to the addition of further studies and suggestions are welcomed. Presenters can either be flown to Dunedin for recording if resident in New Zealand or a video can be sent for inclusion on both the University of Otago website and the GTL website controlled by VSNi.

6. CONCLUSIONS

The launch of the powerful, free GenStat software for teaching and learning is an exciting development for improving teaching and motivating the learning of statistics in a school or in a university. The tailored menus and functions covering an extensive range of statistical procedures are easy to use allowing students to focus on statistical ideas without being overwhelmed by syntax details and programming. Resources required are minimal once the hardware is available.

GenStat for Teaching and Learning also has the potential to serve the educational needs of developing countries more directly than the earlier discovery version of GenStat which was free in developing countries.

Continuity provided by the software is beneficial. Introduced at an early age in school, this single package “opens up” to statistical analyses in the final school years, progresses to free use in university undergraduate courses and then provides a pathway to the commercial version, GenStat 14, for professional use and postgraduate research. With a system of further regional professional development workshops for New Zealand high school teachers and support from various teacher organisations, it is possible that many students arriving at a New Zealand university will be familiar with GTL Schools and GTL Undergraduate within the next year.

Because the software can be used freely at home and facilities exist for submitting reports from home to school for marking, other members of a family have the opportunity to see the work being carried out by students and consequently others have good opportunity to learn about statistics and data exploration. For this reason there is potential to raise the profile of statistics more widely in society. Carefully chosen examples in interesting motivational contexts have the ability to accelerate this development. An unpopular technical subject like statistics comes alive when such examples are presented.

Acknowledgements

The following organizations or individuals are acknowledged: the Campbell Fund of the New Zealand Statistics Association which helped fund the recording of the case study videos and for providing support for employing students to develop GTL lessons; the Higher Education Development Centre at the University of Otago for recording and editing the case study videos free of charge and for recording the introductory workshop

sessions; Robert Van der Vyver for the professional job of editing the videos; Greg Trounson of the University of Otago Department of Mathematics and Statistics for building the website <http://www.maths.otago.ac.nz/videos/statistics>; VSNi and their software developers especially for their immediate responses to requests for improvements to software; Jeanette Chapman for her pioneering work introducing the software at Otago Girls High schools, developing lessons and leading professional development; Dr David Baird for his development work associated with the professional development workshops; Ms Freya Broughton-Ansin and Mr Ross Haines, honours statistics graduates who have helped develop and write the lessons associated with each video; the nineteen researchers from the University of Otago and other research institutes around New Zealand who were prepared to be recorded talking about their research.

References

Andrews, S. (2009) GenStat Discovery: Free Sophisticated First World Menu-Driven Analysis Tools Available to the developing World, and in Particular, Africa. *Proceedings 57th Session International Statistical Institute*. Durban: International Statistical Institute. Voorburg, The Netherlands.

Andrews, S. (2010) Statistical software for teaching: relevant, appropriate and affordable. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics*. Ljubljana, Slovenia: International Statistical Institute. Voorburg, The Netherlands.

Baker, A.N., Smith, A.N.H. and Pichler, F.B. (2002) Geographical variation in Hector's dolphin: recognition of new subspecies of *Cephalorhynchus hectori*. *Journal of the Royal Society of New Zealand*. 32: 713-727.

Campbell, G.P., Curran, J.M., Miskelly, G.M., Coulson, S., Yaxley, G.M., Grunsky, E.C. and Cox, S.C. (2009) Compositional data analysis for elemental data in forensic science. *Forensic Science International*. 188: 81-90.

CAUSE Consortium for the Advancement of Undergraduate Statistics Education. (2011) American Statistical Association. Alexandria, VA, USA. [CAUSEweb.org](http://www.causeweb.org)

Chapman, J., Harraway J. and Andrews S. (2010) Using Genstat at school level for free – a New Zealand experience. Special sessions at 8th International Conference on Teaching Statistics. Ljubljana, Slovenia: International Statistical Institute. Voorburg, The Netherlands.

DASL *The Data and Story Library*. Cornell University, Ithaca, NY USA
URL <http://lib.stat.cmu.edu/DASL/> accessed 8 Dec 2011.

Excel. Microsoft® Excel SP3, Microsoft Corporation, Redmond, WA, USA (2003).

Fellows, I. (2011) An R Graphical User Interface (GUI) for Everyone.
URL <http://www.deducer.org/manual.html> accessed 7 Dec 2011.

Finzer, W. (2005) *Fathom Dynamic Data Software* (Version 2) Emeryville, CA: Key Curriculum Press.

Fox, J. (2005) The R Commander: A Basic-Statistics Graphical User Interface to R. *Journal of Statistical Software*. 14: (9).
URL <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/> accessed 7 Dec 2011.

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

Harraway, John. (2009) Learning Statistics in the Senior School using illustrated real data in context. *Proceedings IASE Satellite Conference Next Steps in Statistics Education*. 57th International Statistics Institute, Durban.

Heiberger, R and Neuwirth, E. (2009) *R Through Excel*. Springer Verlag.

IBM SPSS Statistics Release 19(2011) IBM Corporation, Armonk, NY, USA.

Konold, C and Miller, C. (2004) *TinkerPlots Dynamic Data Exploration* (Version 1). Emeryville, CA: Key Curriculum Press.

Ministry of Education (2007) *The New Zealand Curriculum*. Learning Media Limited, Wellington, New Zealand.

Minitab Statistical Software Version 15. Minitab Inc, State College, Philadelphia, PA, USA. (2007).

Moore, D. (1989) *Against All Odds: Inside Statistics*. Video Series Produced by COMAP Inc. Bedford, MA, USA.

R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2011) ISBN 3-900051-07-0,
URL <http://www.R-project.org>, accessed 7 Dec 2011.

SAS System for Windows, Release 9.3, SAS Institute Inc., Cary, NC, USA (2011).

StatCrunch Data analysis on the Web. Pearson. Boston, MA, USA (2009)
URL <http://www.statcrunch.com>, accessed 7 Dec 2011.

Statistics New Zealand. Schools Corner. Wellington, New Zealand. (2011)
URL http://www.stats.govt.nz/tools_and_services/services/schools_corner/census/census-at-school.aspx, accessed 8 Dec 2011.