The New Zealand Statistical Association

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato



The New Zealand Statistical Association 2008 Conference

University of Waikato
Hamilton, New Zealand
1-2 September 2008

Abstract Booklet

PEARSON Education New Zealand

Statistics New Zealand TATAURANGA AOTEAROA

HRS Hoare Research Software Ltd

the ROYAL SOCIETY of NEW ZEALAND

Bennetts

sas THE POWER TO KNOW®

Department of Statistics
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

J.L. (Judi) McWhirter
BSocSc MSocSc DPhil *Waikato*

Telephone 64-7-838 4774
Facsimile 64-7-838 4155
Email judi@stats.waikato.ac.nz

The University of Waikato
Te Whare Wānanga o Waikato

NZSA2008: The 59th Annual Conference of the New Zealand Statistical Association.

A warm welcome to Hamilton, and in particular, to the University of Waikato, for the 2008 NZSA Conference. We hope that you will enjoy these two days, both professionally and socially.

We have two keynote speakers. On Monday, we have Charmaine Dean (Simon Fraser University, Canada). Her talk is entitled "*Spatial and Mixture Models for the Analysis of Recurrent Events*". On Tuesday, we have Sharleen Forbes (Statistics New Zealand). Her talk is entitled "*Trying to surf the education wave with official statistics*".

We also have invited speakers in **Probability and Stochastic Processes**. These are Estate Khmaladze (Victoria University of Wellington), Jeff Hunter (Massey University), Peter Smith (University of Canterbury) and Mark Holmes (University of Auckland). These talks are scheduled for the morning session on Monday. On Tuesday morning, M. Beatrix Jones (Massey University), Katya Ruggiero (University of Auckland), Ken Dodds (AgResearch) and Sharon Browning (University of Auckland) are invited speakers in **Statistical Genetics**, while David Scott (University of Auckland), Yong Wang (University of Auckland) and Paul Murrell (University of Auckland) are invited speakers for Tuesday afternoon, in **Statistical Computing**.

A special theme for Tuesday, is **Statistical Education**, starting with Sharleen's keynote address, and followed by contributed talks from John Harraway (University of Otago), Maxine Pfannkuch (University of Auckland), Ian Westbrooke (Department of Conservation) and Rolf Turner (STARPath). A series of workshops will be held in the afternoon. We extend a special welcome to the local teachers who are joining us for this day. We trust you will find participation in this event worthwhile.

We are pleased to have eight student presentation, (7 talks and 1 poster), and we are grateful, once again, that Ray Hoare (Hoare Research Software) has agreed to provide sponsorship to the conference by offering prizes for the best student presentations. These will be awarded at the conference dinner, during the presentation session. The NZSA has also provided some funding for students to travel to the conference.

We also thank the following organisations, who have also provided sponsorship for this conference. We are grateful to the Royal Society of New Zealand for hosting the NZSA website where the conference page is located; Statistics New Zealand for sponsoring an event for young statisticians and also providing paper and pens for delegates; Bennetts Campus Bookshop, for conference discounts and vouchers; Pearson Education New Zealand for book vouchers; and SAS for providing conference bags and drink bottles. We are also grateful to the Waikato University, in particular for the use of the Statistics Department's name and some resources to ensure we were able to use the facilities at a discounted rate.

There are many people who have contributed in some way to making this conference a success. Roger Littlejohn and other members of the NZSA Executive Committee and also members of the Department of Statistics at the University of Waikato have provided valuable input at various stages of the planning process. I particularly thank James Curran who was responsible for abstract submission and organising the programme, and Mashewaran Rohan and Maxine Pfannkuch for their work in organising the programme for the Education Session. Roger Littlejohn has also done a wonderful job with the website and Harold Henderson and Catherine Cameron managed the registrations.

My best wishes to you all for a great conference!



Judi McWhirter
Convenor NZSA2008 Local Organising Committee.

For your Information:

Internet Access:

We have provided internet access in the Mathematics and Statistics Senior Computing Laboratory, G.3.12. (This is the building immediately across from the entrance to S-block.) Take the lift or stairs to the 3rd floor. You will find G.3.12, immediately through the door, facing the lift.

There are 10 machines with usernames *stats-nzsa08-##*, where ## is a number from 80 to 89 and corresponds to a particular computer. The computers are named accordingly. Each account uses the same password *conference08*. After you have finished your session, please ensure you have logged off.

There are no printing facilities provided. If you urgently require something printed see me (Judi McWhirter).

Toilet facilities:

Toilets are located on the ground floor of S-block. Go down the main stairs and you will find them on your left. There is also a lift, to the left of the main stairs. Alternatively, there are male and female toilets on each floor of G-Block.

Parking:

On Monday and Tuesday, you may only park in "General Parking" areas (unless, of course, you have a staff permit). These car parks are located through Gate 1 (off Knighton Rd) and through Gate 10 (off Silverdale Rd).

Conference Dinner venue:

The conference dinner is being held at *The Station Café*, which is on Hillcrest Road, within walking distance from the conference venue. From S-Block, head towards Gate 8 and then left down Hillcrest Road, to the pedestrian crossing. The Station is located about 25 metres further along on the other side of the road. All conference participants, even those who are not registered for the dinner, are invited to the pre-dinner drinks session at The Station, following the AGM (in S.1.04). Those attending the pre-dinner drinks session, will be issued with one "free" drink ticket.

Delegates, who have not registered for dinner, but would like to attend, must let me (Judi McWhirter) know (and pay) before midday on Monday, so the caterers can be informed.

Parking for the dinner is available (after 5pm) on the upper level of the staff car park off Gate 7 on Hillcrest Road.

Other points of interest:

You can find Bennetts Bookshop, ASB and National Banks, a Westpac ATM, and a café, underneath the library. Go down the main stairs to ground floor of S-Block and out the doors ahead of you. You will find Bennetts down the outside stairs on your right and the bank and café down another set of stairs to the immediate left of the book shop. There is a pharmacy and more cafés in the Student Union building. Ask for directions, if you need to.

If you have any queries, don't hesitate to ask. Locals, who may be of help, will be wearing blue name badges.

Programme for NZSA2008

Sunday 31 August

5:00 - 7:00	Registration and Reception

Monday 1 September

9:00 - 9:15	Welcome
9:15 - 10:00	Keynote Speaker
10:00 - 10:30	**Morning Tea**
10:30 - 12:30	Invited Session, Contributed Sessions
12:30 - 1:30	**Lunch** Poster Session (Presenters in Attendance)
1:30 - 3:30	Contributed Sessions
3:30 - 4:00	**Afternoon Tea**
4:00 - 5:00	NZSA AGM (S.1.04)
5:00 - 6:00	Young statisticians' Event
5:30 - 6:30	Pre-dinner drinks at The Station
6:30 -	Conference Dinner at The Station

Tuesday 2 September

9:00 - 9:15	Housekeeping
9:15 - 10:00	Keynote Speaker
10:00 - 10:30	**Morning Tea**
10:30 - 12:30	Invited Session, Contributed Sessions
12:30 - 1:30	**Lunch**
1:30 - 3:15	Invited Session, Contributed Sessions, Statistical Education Workshops (1:30-3:30)
3:15 - 3:30	Farewell
3:30 - 5:00	**Afternoon Tea**

NZSA 2008 Conference Timetable

	Monday			
9:00-9:15	Welcome: S 1.04 *Roger Littlejohn (NZSA President), Stephen Joe (Acting Dean for SCMS)*			
9:15 - 10:00	Keynote address: S 1.04 *Charmaine Dean (Simon Fraser University)* Spatial and Mixture Models for the Analysis of Recurrent Events Chair: Murray Jorgensen			
10-10:30	Morning Tea			
10:30-12:30	Invited Session Ia - Probability and Stochastic Processes : S 1.04 Chair: Mark Holmes		Contributed Session Ib: S 1.02 Chair: Bill Bolstad	Contributed Session Ic: S 1.05 Chair: Murray Jorgensen
10:30	Set-valued functions and practical problems of spatial statistics or How to differentiate irregular tubes and why do we need it in statistics? *Estate Khmaladze (VUW)*	10:30	Variances for Petition Estimators *Mike Doherty (SNZ)*	Do emerging markets provide diversification benefits for listed property investors?: An application of Zero Non Patterned Vector Autoregression *Laura Ryan*
11:00	Expected time to coupling in Markov Chains *Jeff Hunter (Massey U. and AUT University)*	10:54	Testing homogeneity with survey data *Alastair Scott, Patricia Metcalf (U. Auckland)*	Autoregressive conditional duration models: Model selection and inference *Gang Xie, Paul Cowpertwait, Daniel Walsh (Massey U.)*
11:30	On the Speed of Random Matrices and their Eigenvalue Decompositions *Peter Smith (U. Canterbury)*	11:18	Longitudinal analysis of business survey data - the potential and the issues *Frances Krsinich (SNZ)*	Model selection and time series asymptotics *Laimonis Kavalieris (U. Otago)*
12:00	Random walk in a random environment - weird and wonderful *Mark Holmes (U. Auckland)*	11:42	Hierarchical Bayes and Polya urn models for creating synthetic microdata. *Patrick Graham (U. Otago)*	Testing Mincer's Linear Regression Model of Earnings and Education using 2002 New Zealand Income Survey data, with special focus on implications for Māori *Vij Kooyela (Te Puni Kokiri)*
		12:06	A New Adaptive Sequential Design for Sampling Rare and Clustered Populations *Jennifer Brown, Mohammad Salehi, Mohammad Moradi, Gavin Bell, David Smith (U. Canterbury)*	Model Averaging in Factorial Data Analysis *David Fletcher, Peter Dillingham (U. Otago)*
12:30-1:30	Lunch Posters (Presenters in attendance)			
1:30-3:30	Contributed session 2a: S 1.04 Chair: Lyn Hunt		Contributed session 2b: S 1.02 Chair: Murray Jorgensen	Contributed Session 2c: S 1.05 Chair: Bill Bolstad
1:30	Advances in Spatial Modelling of Disease *Martin Hazelton, Jonathan Marshall, Sarojinie Fernando, Tilman Davies (Massey U.)*	1:30	The K-Maximum Subarray algorithm as an alternative clustering analysis for the spatial weed aggregation pattern *Kyoko Fukuda, Jennifer Brown, Peter Williams, John Kean (U. Canterbury)*	CANCEIS and its application at Statistics New Zealand *Chang Li (SNZ)*
1:54	Nonparametric computation of survival functions in the presence of interval censoring *Stephen Taylor (U. Auckland)*	1:54	Reducing Respondent Load at Statistics New Zealand *Grace Chiang, Lee Kupferman (SNZ)*	Get a response without direct questioning? - Modelling could be your answer! *Olivia Son, Victoria Wei (SNZ)*
2:18	Adjusting Resource Selection Functions for Imperfect Detection *Miriam Hodge, Jennifer Brown (U. Canterbury)*	2:18	Modelling count and growth data with many zeros *Austina Clark, Eris Zydervelt, Stephen Wing (U. Otago)*	Teaching Bayesian Statistics: What works and what does not *Bill Bolstad (U. Waikato)*
2:42	Graphical Lasso and Shortest Path analysis with gene expression data *Marie Fitch, Beatrix Jones (Massey U.)*	2:42	Hierarchical Modeling of Trace-element Signatures to Determine Animal Movement *Richard Barker, Chikako van Koten, Gerry Closs, Esben Kristensen (U. Otago)*	Measuring the Competitiveness of Domestic Team Sport *Paul Bracewell (Offlode), Don Forbes, Clint Jowett, Heath Kitson*
3:06	Confidentiality: new needs and new solutions *Tim Hawkes, Mike Camden (SNZ)*	3:06	Parallels between mixed and fixed parameter linear models *Steve Haslett (Massey U.)*	New Approach to Robust Estimation *Maheswaran Rohan (DoC)*
3:30-4	Afternoon Tea			
Murray	NZSA AGM: S 1.04			
5-6	Young Statisticians' Event S 1.02			
5:30 6:30	Pre-Dinner Drinks at the Station Conference Dinner at the Station (Presentations)			

NZSA 2008 Conference Timetable

	Tuesday			
9:00	House Keeping: S.1.04			
9:15 - 10:00	Keynote address: S 1.04 *Sharleen Forbes, Statistics New Zealand* Trying to surf the education wave with official statistics Chair: Ian Westbrooke			
10:00-10:30	Morning Tea			
10:30-12:30	Invited Session 3a - Statistical Genetics: S 1.04 Chair: Sharon Browning	Contributed Session 3b - Statistical Education: S 1.02 Chair: Maheswaran Rohan		Contributed Session 3c: S 1.05 Chair: Lyn Hunt
10:30	Bayesian Parentage Analysis Using Multiple Imputation *Beatrix Jones (Massey U.)*	The development of further teaching resources to motivate statistics learning in years 12 and 13 *John Harraway (U. Otago)*	10:30	Statistical methods for testing equality of survival probability with paired right-censored data *Pei-Fang Su, Yi-De Liao, Yunchan Chi*
11:00	Protein expression experiments using iTRAQ: a unified protocol for design and analysis? *Katya Ruggiero (U. Auckland)*	Informal Statistical Inference *Maxine Pfannkuch (U. Auckland)*	10:54	Estimation in additive hazards model with clustered current status data *Yunchan Chi and Pei-Fang Su*
11:30	Using genetic marker-based estimates of relatedness to detect a heritable trait in unpedigreed populations *Ken Dodds (AgResearch)*	Exploring data on a rare threatened bird - the rock wren *Ian Westbrooke (DoC)*	11:18	Fractional s^n-k Factorial BIB Designs *Mohamed Shamsuddin, Mohammad Albassam (King Abdulaziz U. Jeddah)*
12:00	Detecting Identity-By-Descent For Complex Gene Mapping *Sharon Browning (U. Auckland)*	An Educational Experience *Rolf Turner (STARPath)*	11:42	Multiple Confounded and Orthogonal Replicated Full Factorial BIB Designs *Mohamed Shamsuddin, Mohammad Albassam (King Abdulaziz U. Jeddah)*
			12:06	
12:30-1:30	Lunch			
1:30-3:15	Invited Session 4a - Statistical Computing: S 1.04 Chair: Paul Murrell	Statistical Education Workshops		Contributed Session 4b: 1.02 Chair: Bill Bolstad
1:30	Implementation of Distributions in R *David Scott (U. Auckland)*	1:30-2:30 1. New strengths in the Curriculum's Statistics: a Workshop Mike Camden (SNZ) Room I 1.01 2. Growing Scatterplots Pip Arnold (Team Solutions, U. Auckland) Room I 1.05	1:30	Triangular density approximation for error rates associated with the discriminant function based on absolute deviation from the mean *S. Ganesalingam, Siva Ganesh (Massey U.), A. Nanthakumar (SUNY)*
2:00	Fast Computation for Fitting Non-parametric and Semi-parametric Mixture Models *Yong Wang (U. Auckland)*		1:54	Forming Clusters from Census Areas with Similar Tabular Statistics *Murray Jorgensen (U. Waikato)*
2:30	Reusable Documents *Paul Murrell (U. Auckland)*	2.30-3.30 3. "Making the call" Pip Arnold (Team Solutions, U. Auckland) Room I 1.01 4. Drawing an inference from an experiment: Curriculum Level 8 Maxine Pfannkuch (Department of Statistics, U. Auckland) Room I 1.05	2:18	Developing a Geospatial Land Use Classification for New Zealand: A Review of Current Theory and Practice *Daniel Rutledge, Robbie Price, Sarah Cowell, Craig Briggs (Landcare)*
			2:42	Identification of Higher-Order Factor Analysis Models *Walter Davis (SNZ)*
3:15-3:30	Farewell			

Posters (Presenters in Attendance Monday lunchtime)

Ian Westbrooke (D0C)	If nothing goes wrong, is everything alright?
Rod Ball (Scion)	Case studies in association mapping
Lyn Hunt (U. Waikato), Kaye Basford	Model Selection for the Multimix Class of Mixture Models
Murray Jorgensen (U. Waikato)	Latent variable models for dilution series data
K. C.-T. Chang, K. Ruggiero (U. Auckland)	Protein expression data from an iTRAQ MudPIT
Bill Bolstad (U. Waikato)	Teaching Bayesian Statistics: What works and what does not

Keynote Speaker

Spatial and Mixture Models for the Analysis of Recurrent Events



Charmaine Dean

Department of Statistics and Actuarial Science, Simon Fraser University
British Colombia, Canada
dean@stat.sfu.ca

Studies of recurring infection or chronic disease often collect longitudinal data on the disease status of subjects. Two types of models may be envisioned for the analysis of such data: counting process models such as nonhomogeneous Poisson process models or multi-state transitional models. This talk addresses both scenarios in the specific case where the population consists of mixtures and where spatial correlation is evident in the intensities governing recurrences.

A flexible semiparametric model for analyzing longitudinal panel count data is presented. Panel count data refers here to count data on recurrent events collected as the number of events which have occurred within specific follow up periods. Discrete mixtures of smooth counting process intensity forms are considered. These include mixtures of adaptive splines which permit time-varying covariate effects, with the so-called proportional intensity model as a limiting case. For recurrent events handled in a multi-state transitional model framework, individuals may be said to occupy one of a discrete set of states and interest centers on the transition process between states. This talk examines the use of mixed Markov models for the analysis of such longitudinal data where the processes corresponding to different subjects may be correlated spatially over a region. Both discrete and continuous-time models incorporating spatially correlated random effects are developed.

Several examples will illustrate the methods discussed including an epidemiological study of revascularization intervention in the province of Quebec, a forestry study of recurrent white pine weevil infestation in British Columbia, and a biological study to assess the effectiveness of a pheromone treatment in disturbing the mating habits of the cherry bark tortrix moth.

Keynote Speaker

Surfing the education wave with official statistics



Sharleen Forbes

Statistics New Zealand and School of Government at Victoria University
Wellington, New Zealand
sharleen.forbes@stats.govt.nz

This paper discusses the role of a National Statistics Office in education and why we need to 'pick the wave' when we invest in statistics education. Targeted community groups currently include journalists, small businesses and Maori organisations but one of the major areas of investment is in training for state sector employees. This includes a Certificate of Official Statistics and a half-time professorship within the School of Government at Victoria University.

Another major area of investment is providing resources for schools that fit the new national Mathematics and Statistics Curriculum. This includes part funding of CensusAtSchools. Examples developed for classroom use are presented, using Statistics New Zealand's Census and Survey data and also datasets specially constructed for school teachers (synthetic unit record files - SURFS).

Increased formal academic networks have also been created and the recent Vice-Chancellors' agreement with Statistics New Zealand provides for free academic access to confidentialised official datasets for research and education purposes.

Poster

Case studies in association mapping - frequentist and Bayesian measures of evidence

Rod Ball
Scion
rod.ball@scionresearch.com

Whole genome association studies are extreme examples of the need for Bayesian methods and the failure of frequentist inference. Whole genome association studies aim to exploit *linkage disequilibrium* (i.e. population level non-independence between loci) to find loci associated with variation in complex traits and diseases. Despite initial optimism, many published associations are spurious. For example, 13 published associations for type 2 diabetes were retested and only one confirmed, and similar negative results were found for polymorphisms from 13 genes putatively associated with Alzheimer's disease.

We illustrate these problems and Bayesian statistical analysis and experimental design for association mapping with reference to case studies from Chapters 7 and 8 of the book "Association Mapping in Plants", Springer 2007, and recent results from large scale case-control studies. Case studies include:

- A case-control test for an association between the HbS mutation and malaria.
- Detectability of associations between the APOE locus and Alzheimer's disease from a whole genome scan of SNP.
- Diabetes Genome Initiative and Wellcome Trust Case-control studies.

We re-analyse previous results using Bayesian methods. Equivalent Bayes factors are derived from published results and posterior probabilities for putative associations assessed. Bayes factors are derived for the common association tests: the chi-squared, Fisher's exact test, and the TDT and S-TDT tests for discrete and continuous traits. Various methods are used including direct integration, MCMC, and the Savage-Dickey density ratio. A common theme is the inadequacy of *p*-values as a measure of evidence for testing scientific hypotheses as noted by Berger and Berry. Higher sample sizes are needed to obtain respectable Bayes factors, and even higher sample sizes are needed to obtain sufficiently high Bayes factors to overcome low prior probabilities for genomic associations. The best study (considering analysis and approach to power of experimental design) is the Wellcome Trust Case Control study (2007), who give Bayes factors in addition to *p*-values. However, even here, around half the reported strongest 25 associations with $P < 5 \times 10^{-7}$ for 7 diseases has posterior odds around 1 or less.

We advocate experimental design with power to achieve a Bayes factor sufficiently high (e.g. $B > 1{,}000{,}000$) to overcome the low prior odds for genomic associations, in combination with other experimental design techniques such as block designs and clonal replication where possible. This requires potentially an order of magnitude greater sample size than previously reported.
An R package, `ldDesign`, implementing the method, is available on CRAN.
http://cran.r-project.org/src/contrib/PACKAGES.html#ldDesign

KEYWORDS: Linkage disequilibrium, association tests, QTL, QTN, Genome scan, Candidate genes, Experimental design, Bayes factor.

Hierarchical Modeling of Trace-element Signatures to Determine Animal Movement

Richard Barker, Chikako van Koten, Gerry Closs and Esben Kristensen
University of Otago
rbarker@maths.otago.ac.nz

The use of trace-element signature analysis to assess movements of animals has become increasingly popular as a method for answering questions that have proven difficult using standard techniques. Applications include studies of fish migration using trace-element signatures from otoliths and bird migration based on stable hydrogen isotope ratios from feathers. To allow probability-based inference about the animal locations we consider models in which location is treated as a latent variable that is informed by trace-element data through a hierarchical model. Bayesian methods based on Markov chain Monte Carlo are described for inference. We present results from a study of fish spawning in the Taieri River.

Talk and Poster

Teaching Bayesian Statistics: What works and what does not

Bill Bolstad
University of Waikato
bolstad@stats.waikato.ac.nz

The author has extensive experience developing and successfully delivering courses in Bayesian Statistics at an earlier stage than most other universities. These include ***Introduction to Bayesian Statistics*** which is aimed at undergraduate students, and ***Computational Bayesian Statistics*** which is aimed at upper level undergraduate students, or first year graduate students. The purpose of this talk is to pass along this experience.

Measuring the Competitiveness of Domestic Team Sport

Paul J. Bracewell, Don G. R. Forbes, Clint A. Jowett and Heath I. J. Kitson
Offlode Ltd
paul@offlode.com

A proprietary methodology for rating the performance of sports teams within and across domestic competitions is briefly outlined. This approach uses an adaptation of the chi-square test of independence to determine how Team A would fare against Team C given the performance of Team A against Team B and how well Team B went against Team C. Such an approach is useful for predicting outcome of matches within tournaments. The output of this method is validated internally using competition points, and externally using finalist probabilities and the standard deviation of win/loss percentages. 22 competitions from a variety of winter sports across a number of countries are compared. The important findings discussed, including the impact of salary caps upon evenness of competitions and a comparison of the competitiveness of northern and southern hemisphere domestic rugby.

A New Adaptive Sequential Design for Sampling Rare and Clustered Populations

Jennifer A. Brown, Mohammad Salehi M., Mohammad Moradi, Gavin Bell and David R. Smith.
University of Canterbury
jennifer.brown@canterbury.ac.nz

Designing an efficient large-area survey is a challenge for ecologists. Adaptive sampling designs can be efficient because it ensures survey effort is targeted to subareas of high interest. In two-stage sampling, higher density primary sample units are usually of more interest than lower density primary units when populations are rare and clustered. Two-stage sequential sampling has been suggested as a method for allocating second stage sample effort among primary units. We present a modification called adaptive two-stage sequential sampling. In this method, the adaptive part of the allocation process means the design is more flexible in how much extra effort can be directed to higher-abundance primary units.

Detecting Identity-By-Descent For Complex Gene Mapping

Sharon Browning
The University of Auckland
s.browning@auckland.ac.nz

Identity-by-descent (IBD) mapping has been proposed as a powerful way to map genes underlying complex diseases. For a disease with a genetic basis, pairs of cases (individuals with the disease) will tend to be more related than pairs of individuals from the population, and in particular will tend to share genetic material inherited IBD from common ancestors at the locations of disease susceptibility variants. Thus, IBD mapping involves testing for excess IBD sharing between cases, and relies on detecting small segments of IBD sharing between individuals. Existing methods for detecting IBD do not adequately account for linkage disequilibrium (LD) correlation between closely spaced genetic markers, such as those in genome-wide association mapping panels. I propose a new method for detecting segments of IBD between pairs of individuals. My method accounts for LD by means of the localized haplotype cluster model. It allows for detection of smaller tracts of IBD than can be found using existing methods, potentially improving power to detect regions harbouring disease-susceptibility variants.

Poster

Protein expression data from an iTRAQ MudPIT: An experimental design for information decomposition

Kevin Chih-Tao Chang and Katya Ruggiero
The University of Auckland
kcha193@aucklanduni.ac.nz

Multidimensional Protein Identification Technology (MudPIT) coupled with isobaric Tags for Relative and Absolute Quantitation (iTRAQ^{TM}) has proved a powerful tool in protein expression profiling. These experiments are two-phase in nature (Jarrett and Ruggiero, 2008). The Phase 1 experiment involves the randomization of animals to treatments and provides the biological material from which protein is extracted and purified for use in the subsequent laboratory-based Phase 2 MudPIT experiment. We will show that consideration must be given to the Phase 1 experimental design in designing the Phase 2 experiment. We will also show how, in addition to estimating biological variation with the maximum available degrees of freedom, our two-phase design also enables us to estimate each of the technical sources of variation associated with such experiments.

Estimation in additive hazards model with clustered current status data

Yunchan, Chi and Pei-Fang Su
National Cheng-Kung University
ycchi@stat.ncku.edu.tw

Current status data arise naturally from tumorigenicity experiment, biomedicine, econometrics, demographic and sociology studies. Moreover, clustered current status data often occur when animals are from same litter in tumorigenicity experiments. The only information extracted from current status data is that the true survival times are before or after the monitoring (censoring) times. Consequently, the nonparametric maximum likelihood estimator of survival function converges at $n^{1/3}$ rate to a complicated limiting distribution (Groeneboon and Wellner, 1992). Hence, semiparametric regressions and linear regression have been extended for independent current status data to derive the test statistics whose distributions converge at the $n^{1/2}$ rate. However, a straightforward application of these statistical methods to clustered current status data is not appropriate. Therefore, the empirical efficient score function derived from the additive hazards model proposed in Martinussen and Scheike (2002) is used as estimating function for current status data. This estimating function can be used for testing the regression coefficient, such as testing treatment effect. A simulation study was carried out to access the performance of the proposed estimating function. The bias from efficient score function is smaller than that from partial score function (Lin et al. (1998)). Moreover, the power of the empirical efficient score function in testing treatment effect is slightly higher than the estimating function based on the partial score function.

Reducing Respondent Load at Statistics New Zealand

Grace Chiang and Lee Kupferman
Statistics New Zealand
lee-kurt.kupferman@stats.govt.nz

Statistics New Zealand is focused on maintaining the relevance and integrity of our surveys, whilst also focusing on the reduction of respondent load. To share the load among businesses with certain characteristics, a sample rotation project was undertaken to investigate the feasibility of introducing regular sample rotation in Statistics New Zealand Business Surveys. Following the success of this feasibility study, a sample rotation was later implemented in the Retail Trade Survey. This presentation will explain the use of the Permanent Random Number line in Statistics New Zealand Business Surveys, and how this facilitates the use of sample rotation within these surveys. We will also look at how rotation is implemented in the Retail Trade Survey, and further work to be undertaken at Statistics New Zealand.

Modelling count and growth data with many zeros

Austina Clark, Erin Zydervelt and Stephen Wing
University of Otago
aclark@maths.otago.ac.nz

We discuss the problem of modelling survival /mortality and growth data that are skewed with excess zeros. This type of data is a common occurrence in biological and environmental studies. The method presented here allows us to utilize both the survival/mortality and growth data when both data sets contain a large proportion of zeros. The method consists of four stages. Firstly the original data is divided into two sets; one contains all the surviving organisms and the other all of the mortalities. Secondly we calculate the actual growth of the surviving organisms and of the mortalities. Thirdly we count the number of surviving organisms for which growth has occurred and the number where no growth occurred, and the same count procedure is carried out on the mortalities. Next we model the survival /mortality data and growth/no growth data using logistic regression, and separately model the growth data using an ordinary regression. Finally we combine the three models to estimate the expected growth for a specific set of values of the explanatory variables. A case study of survival and growth of blue mussels and ribbed mussels trans-located from their natural environment to different depths and sites along the axis of Doubtful Sound, New Zealand, is used for illustration.

Identification of Higher-Order Factor Analysis Models

Walter Davis
Statistics New Zealand
walter.davis@stats.govt.nz

Parameter identification in confirmatory factory analysis (CFA) has long been a thorny issue with a small number of rules developed which cover only particular classes of models. An extension of the existing "2-step rule" of identification for structural equation models (SEM) can be applied to identify higher-order CFAs. A proof and examples of its application will be presented.

Using genetic marker-based estimates of relatedness to detect a heritable trait in unpedigreed populations

Ken Dodds
AgResearch
ken.dodds@agresearch.co.nz

When interest arises for a particular trait as a candidate for genetic selection or for investigation of gene effects, it needs to be established whether that trait is actually heritable. The trait may not have been measured widely in pedigreed populations, because the expense had not been justified, or because its expression had narrower variability in breeding populations, or because it is a disease or deformity that has become more prevalent than previously, but is still relatively rare. One approach to detect inheritance would be to screen a commercial population to obtain a sample of 'affecteds' or extremes (the test group) and to also obtain a random control group. These individuals are then genotyped with a set of genetic markers and the relationships between individuals within each group estimated. If the relatedness is higher in the test group than in the control group, this provides initial evidence for the trait being heritable. A power simulation shows that this approach is feasible with moderate resources.

Variances for Petition Estimators

Mike Doherty
Statistics New Zealand
mike.doherty@stats.govt.nz

It is desired to estimate the number of electors who have signed an indicative referendum petition, on the basis of a check of a sample of signatures. This problem is related to estimating the abundance of animal species. Leo Goodman showed there is a unique unbiased estimator. However its properties can be less than ideal. I will describe some results, which may be new, on sample estimates of the variance of the Goodman and related estimators.

Graphical Lasso and Shortest Path analysis with gene expression data

Marie Fitch and Beatrix Jones
Massey University
m.fitch@massey.ac.nz

Gaussian Graphical Models (GGMs) are a popular tool for representing gene association structures. We use yeast (*Saccharomyces Cerevisiae*) gene expression profiles. We propose using estimated partial correlations to create a graph where the length of an edge is inversely related to the partial correlation between the gene pair. Graphical lasso is used to obtain partial correlations. The shortest paths between pairs of genes are found. Where terminal genes have the same biological function intermediate genes on the path are classified as having the same function. The accuracy of the graph is assessed by considering the accuracy of these classifications. We also compare our results with those obtained using a graph constructed using correlations. Using a partial correlation graph we are able to classify approximately twice as many genes to the same level of accuracy as when using a correlation graph. More importantly when we restrict the number of genes classified, by an upper bound on the path length, the use of partial correlations can increase the accuracy of the classifications.

Model Averaging in Factorial Data Analysis

David Fletcher and Peter Dillingham
University of Otago
dfletcher@maths.otago.ac.nz

Suppose we plan to use a standard normal-theory linear model to analyse the results from a study involving a number of explanatory factors. We can perform the usual factorial analysis of variance and assess the evidence for each of the main effects and interactions. In addition, we will usually wish to summarise the results by calculating estimates and confidence intervals one or more combinations of the levels of the different factors. We present results from simulations designed to assess the benefits of using model averaging for calculating such estimates and confidence intervals, compared to using the full model or a reduced model (e.g. one involving only main effects and interactions that are statistically significant).

The K-Maximum Subarray algorithm as an alternative clustering analysis for the spatial weed aggregation pattern

Kyoko Fukuda, Jennifer Brown, Peter Williams and John Kean
University of Canterbury
k.fukuda@math.canterbury.ac.nz

Weed invasions can alter the nature of the environment, the surrounding ecosystem and human activities, and it will cost us to manage and control them. This study introduces the K-Maximum Subarray Algorithm to investigate the spatial distribution patterns of hawthorn, grown over 100 years in Porters Pass. The K-Maximum subarray algorithm (K-MSA) investigates its maximum clustering patterns by identifying K-number of maximum growth patches. Further, the observed and simulated maximum weed growth patchiness were investigated for randomness of the aggregation pattern. Results demonstrated how the weed distribution pattern has changed over four decades, with samples in 1966, 1976, 1986 and 2006.

Triangular density approximation for error rates associated with the discriminant function based on absolute deviation from the mean

S. Ganesalingam, Siva Ganesh and A. Nanthakumar
Massey University
S.Ganesalingam@massey.ac.nz

The choice of a discriminant function is basically determined by the associated error rates and hence the estimation of the error rates is of vital importance. Consider the problem of statistical discrimination involving two multivariate normal distributions with equal means but different covariance matrices. Traditionally, a quadratic discriminant function (QDF) is used to separate two such populations. Ganesalingam and Ganesh (2004) introduced a linear discriminant function called 'Absolute Euclidean Distance Classifier (AEDC)' and compared its performance with that of QDF in terms of their associated misclassification error rates. In this paper a triangular density based approximation for the error rates associated with the function AEDC is derived and computed for the various covariance structures in a simulation exercise which serve as a bench mark for comparison. The approximation we introduced in this paper simplifies the amount of computations involved. Also, this approximation provides a closed form expression for the tail areas of most symmetrical distributions that is very useful in many practical situations such as the misclassification error computation in discriminant analysis.

Hierarchical Bayes and Polya urn models for creating synthetic microdata.

Patrick Graham
Bayesian Research and University of Otago, Christchurch
patrick.graham@xtra.co.nz

The development of methods to enable official statistics agencies to release microdata, while preserving respondent confidentiality, is an active research area within official statistics. Traditional methods, such as adding noise, top coding and data swapping, produce datasets that are not easy to analyse unless the confidentialising process is ignored, an analytical simplification which is, in general, not justified by statistical theory. An alternative approach uses multiple imputation to produce multiple synthetic datasets from which a user can obtain valid inferences by analysing each imputed dataset using standard methods and combining results over the imputed datasets. Construction of the imputation model used to generate the imputed synthetic datasets is a crucial step in producing useable synthetic data because a misspecified imputation model can lead to erroneous inferences from the synthetic data.

In this paper I extend earlier work on hierarchical Bayes imputation methods for categorical data to datasets with a mix of categorical and numerical variables. One approach is to model numerical variables conditionally on categorical variables using a hierarchical multivariate normal regression model. However, this amounts to modeling means, and does not guarantee that other features of the observed data are closely approximated in synthetic data. By borrowing ideas from non-parametric Bayesian theory, an imputation process centred on a hierarchical Bayes model, but allowing closer approximation to the observed data distribution than is possible under the hierarchical Bayes model, can be developed. The simplest approximate implementation of these ideas involves generating imputations from a generalized Polya urn, centred on the hierarchical Bayes model. The performance of synthetic data generated from such a sampling process is compared with the performance of synthetic data generated under a fully parametric hierarchical Bayes model and with a recently proposed perturbation method which preserves sufficient statistics for normal-theory models.

The development of further teaching resources to motivate statistics learning in years 12 and 13

John Harraway
University of Otago
jharraway@maths.otago.ac.nz

Extensive developments are taking place in the teaching of mathematics and statistics in New Zealand schools. To support these developments there is need to provide interesting and relevant sets of data for use in the classroom to motivate students. This is particularly the case for the last two years at high school. CensusAtSchool and the SURF's provided by Statistics New Zealand are widely available. This paper will discuss another source of data and will by way of illustration provide two case studies which can be used in teaching. One study dealing with categorical data investigates the idea of a confounding variable, an important study design idea, while the second involves a large set of data involving both continuous and categorical data. The studies have matching DVD presentations featuring the local University researchers describing the studies, and the data sets are available for analysis using EXCEL or any other statistical package.

Parallels between mixed and fixed parameter linear models

Stephen Haslett
Massey University, Palmerston North
s.j.haslett@massey.ac.nz

Is it necessary to distinguish between fixed and mixed parameter linear models when deriving BLUEs and BLUPs? This question has a rather surprising answer: that irrespective of the ranks of design and covariance matrices, if the mixed model is explicitly rather than implicitly specified, whether parameters are fixed or random is irrelevant to optimisation of the relevant quadratic form. This conclusion can considerably simplify derivation of results for linear mixed models, since by appropriate specification they can be directly linked to their fixed parameter model analogues. For example, it provides a method for finding conditions under which BLUPs in one linear mixed model remain BLUPs in a second; this seems to be an important but previously unsolved problem, with application for example in animal breeding and small area estimation.

Confidentiality: new needs and new solutions

Tim Hawkes and Mike Camden
Statistics New Zealand
Tim.Hawkes@stats.govt.nz

New data collections, and new pressures on existing collections, require us to renew our methods for maintaining confidentiality. We will describe some issues with microdata and aggregate data. For microdata, we will describe our mix of 'data-products', and how this may change. The mix currently includes datasets for use in our three data laboratories, the new AtoM remote access system, the Basic Confidentialised Unit Record File system, and some synthetic files. Ongoing long-term provision of these depends on user demand and uptake. We will compare these modes, and will show how carefully targeted interventions (like top-coding and special uniques modifications) can maintain both utility and safety. For aggregate data, we will describe our thinking about a new social survey. We will conclude by describing the framework in which we make decisions about confidentiality methods.

Advances in Spatial Modelling of Disease

Martin Hazelton, Jonathan Marshall, Sarojinie Fernando and Tilman Davies
Massey University
m.hazelton@massey.ac.nz

Spatial modelling of disease is a problem of fundamental importance in both human and veterinary epidemiology. A popular approach is to use nonparametric smoothing techniques to estimate the relative risk of disease from case-control data over the geographical study area. However, there are a number of technical problems with the implementation of such methodology. First, the population at risk (be it human or animal) is typically distributed in a highly non-uniform manner across space, requiring the use of spatially adaptive smoothers. Second, the boundaries of the study region may highly irregular, making it difficult to obtain effective correction for edge effects. Third, it is hard to provide reliable estimates of precision for the estimated risk surface because of the complexity of the methods required to solve the first two problems. In this talk we describe some recent progress in tackling all these difficulties, and foreshadow extensions to more general spatio-temporal models.

Adjusting Resource Selection Functions for Imperfect Detection

Miriam Hodge and Jennifer Brown
University of Canterbury
mch145@student.canterbury.ac.nz

Resource Selection Functions measure the preference animals have for different available habitat types. One problem in constructing resource selection functions is that some habitat use goes undetected because habitat use is difficult to detect in some habitat types. Resource selection models can be enhanced incorporating detection probabilities.

Random walk in a random environment - weird and wonderful.

Mark Holmes
The University of Auckland
mholmes@stat.auckland.ac.nz

Random walks in random environments are very interesting if you like counterintuitive problems and results. The situation in more than one dimension is relatively poorly understood. We will discuss some of these counterintuitive results, and (once you are feeling sufficiently uneasy) some intuitive ones.

Poster

Model Selection for the Multimix Class of Mixture Models

Lyn Hunt and Kaye E. Basford
University of Waikato
lah@stats.waikato.ac.nz

The finite mixture model approach to clustering is a model based approach that requires the specification of both the number of components to be fitted to the model and the form of the component distributions. The Multimix class of mixtures was proposed by Hunt and Jorgensen (1996, 1999) and allows the clustering of data containing both categorical and continuous attributes. In this poster, we show the behaviour of some commonly used model selection criteria when using the mixture model to cluster data containing mixed categorical and continuous attributes. We illustrate the performance of these criteria in selecting both the number of components in the model and the form of the component distributions when fitting the Multimix class of models to simulated data and a medical data set.

Hunt, L. A., and Jorgensen, M. A., (1999). Mixture Model Clustering Using the Multimix Program. *Austral. & New Zealand J. Statist*. 41, 153-171.
Jorgensen, M. A., and Hunt, L. A.,(1996). Mixture Model Clustering of Data Sets with Categorical and Continuous Variables. In *Proceedings of the Conference on Information, Statistics and Induction in Science*, Melbourne, 1996, 375-384.

Expected time to coupling in Markov Chains

Jeffrey Hunter
Massey University, Auckland and AUT University, Auckland
j.hunter@massey.ac.nz

Coupling times in a discrete time Markov chain are discussed and techniques for finding the expected times to coupling are presented. The time to coupling can be used as a measure of time to stationarity. For a general finite state Markov chains it is difficult to derive explicit expressions for the expected time to coupling but it is generally possible to deduce simple upper and lower bounds for such expectations. Extensions to these bounds, when the simple bounds do not apply, are also given with some detailed comparisons provided for general two- and three-state chains.

Bayesian Parentage Analysis Using Multiple Imputation

Beatrix Jones
Massey (Auckland)
m.b.jones@massey.ac.nz

In a parentage analysis, one collects genetic data from a group off offspring, and adults that may be their parents. The genetic data is then used to match up offspring with their parents; questions like "do older adults produce more offspring" can then be answered. When not all offspring can be confidently assigned to a single parent pair, the parent assignments can be treated as latent variables, and demographic parameters estimated by sampling from their joint posterior with the parent assignments. Software performing this Bayesian approach is available, but not as widely used as software that computes a single, best parent assignment for each offspring, with a verdict of 'not enough information' for some offspring. In some cases, this may be because the demographic model fit by the Bayesian procedure is not the one desired by the ecologist. In this presentation, we consider use of multiple imputation for parentage analysis. The statistician provides machinery for sampling several sets of plausible parentage assignments. The ecologist then fits their preferred demographic model for each set of assignments. The between-set parameter variability then contributes a component of the parameter uncertainty. This framework differs from classical multiple imputation in that the models used in the two stages are, by design, different (uncongenial). We compare fitting the full Bayesian hierarchical model with two multiple imputation procedures: one properly accounting for the uncongeniality, one ignoring it.

Forming Clusters from Census Areas with Similar Tabular Statistics

Murray Jorgensen
University of Waikato
maj@waikato.ac.nz

We present a methodology for the clustering of a large number of tables of similar form. Data of this kind is often available from National Statistical Offices as tabulations of a set of variables for each of a large number of geographic units. Model-based clustering is employed, the underlying model being a finite mixture of multinomial distribution. We discuss the choice of the number of clusters and the interpretation of the clusters found.

Poster

Latent variable models for dilution series data

Murray Jorgensen
University of Waikato
maj@waikato.ac.nz

A common approach to modelling microbiological data from dilution assays is to replace the raw data of numbers of positive tubes at each dilution level by a "Most Probable Number" (MPN) estimate of bacterial density and submit the logarithms of these estimates to linear modelling. This approach is not without problems. The linear model assumptions are not met by the estimates and point estimates are not available in the relatively common cases where all tubes are positive or all tubes are negative. In this paper we note the connection of MPN estimates to generalized linear models and develop a family of models appropriate for dilution series data. We assume an underlying linear model for the true but only indirectly observed microbial log densities, which form part of the "complete data" for an application of the EM algorithm. We illustrate our approach using data from a survey of pathogen levels in New Zealand freshwater environments.

Model selection and time series asymptotics

Laimonis Kavalieris
University of Otago
lkavalieris@maths.otago.ac.nz

In any realistic modelling scenario candidate models are chosen to capture an interesting aspect of the data. We cannot assume the existence of a "true" model or even a finitely parameterized data generating process. Thus the model selection criterion should be tuned to the purpose of the statistical analysis, and a mathematical analysis of the criterion must allow increasing model complexity with increasing sample size. I will discuss this model selection framework in the time series context and present some new results for penalty function criteria (such as AIC, amongst others) for model selection.

Set-valued functions and practical problems of spatial statistics
or
How to differentiate irregular tubes and why do we need it in statistics

Estate Khmaladze
Victoria University of Wellington
Estate.Khmaladze@mcs.vuw.ac.nz

The talk aims to bring together two fields, which seemed unconnected and distant from each other. Indeed, we know of no cross-reference up to 2007. We will start with the spatial "change-set problem" - the problem of estimating spatial images like ore deposits, pollution sites, shape of bacteria, etc. (see, e.g., [1] and ref.) We show why and for what reason these problems require an infinitesimal analysis for sets, or, more precisely, set-valued functions. Then we present the recently developed version of such analysis - the notion of the derivative of set-valued functions [2]. Finally we discuss other possible applications in probability and statistics.

[1] E.V.Khmaladze, R.M.Mnatsakanov, N.Toronjadze, "Local covering numbers and change-set problem", *Mathemat. Methods Statist.*, 15 (2006).
[2] E.V.Khmaladze, "Differentiation of sets in measure", *J. Mathemat. Analysis Appl.*, 334 (2007).

Testing Mincer's Linear Regression Model of Earnings and Education using 2002 New Zealand Income Survey data, with special focus on implications for Māori

Vij Kooyela
Te Puni Kōkiri (Ministry of Māori Development)
kooyv@tpk.govt.nz

There is substantial body of evidence that shows that those with higher levels of education are more likely to participate in the labour market, face lower risks of unemployment, have greater access to further training and receive higher earnings on average. Jacob Mincer's model of earnings and schooling establishes the relationship between the logarithm of earnings to years of schooling, years of work experience and years of work experience squared. This paper provides a test to the Mincer's model using the Statistics New Zealand 2002 New Zealand Income Survey CURF individual record. When the Mincer's model was applied to this data, it was found that not all the assumptions of a linear regression model were satisfied. While the Mincer's model could explain roughly 15% of the variability in the data, adding other explanatory variables as extensions to the model, improve the fit to about 34%. Highest qualification, ethnicity, potential experience, the square of potential experience and gender are clearly significant variables contributing to differences between groups. Apart from these variables, it was found that there was a significant regional difference in the earnings of respondents. Across all models fitted, it was found that individuals with a bachelor or higher qualification were more likely to have higher earnings compared to individuals with other qualifications. The amount of experience an individual had also contributed significantly to earnings. We looked at whether it was worthwhile adding an interaction between highest qualification and potential experience. It was found that this interaction term added only slightly to the regression model.

Our special focus was exploring implications for Māori in this paper. There appeared to be clearly a significant difference across ethnic groups in the contribution of highest qualification to earnings. We also found that Māori had lower mean earnings compared to Europeans with regard to the contribution of highest qualification, other things being equal. To investigate whether there was any difference between earnings of individuals of similar qualifications, we looked at a subpopulation of individuals with a bachelor or higher degree. Results showed that for individuals with a bachelor or higher qualification, there was a significant difference between Māori and European earnings. This suggests that there are other factors contributing to the difference in earnings for Māori and Europeans other than highest qualification.

Longitudinal analysis of business survey data - the potential and the issues

Frances Krsinich
Statistics New Zealand
frances.krsinich@stats.govt.nz

We can get more out of much of Statistics New Zealand's business survey data. Although generally designed for cross-sectional and aggregate estimates, highly overlapping samples from period to period mean we can analyse the data longitudinally at the business level. Using the Retail Trade survey as a case study, this talk will outline the types of longitudinal analysis possible, and discuss how we can deal with the issues that arise from using this data in a way it was not originally designed for.

CANCEIS and its application at Statistics New Zealand

Chang Li
Statistics New Zealand
Chang.li@stats.govt.nz

CANCEIS (CANadian Census Edit and Imputation System) is an editing and imputation system developed by Statistics Canada for use in censuses and social surveys. At Statistics New Zealand, CANCEIS was evaluated against our general editing and imputation requirements and has been under implementation in some surveys. My presentation will provide a brief overview of the software – its features, advantages, and limitations. I will also talk about the evaluation result, our progress with its implementation, and the difficulties faced and issues raised.

Reusable Documents

Paul Murrell
The University of Auckland
paul@stat.auckland.ac.nz

This talk looks at the task of producing documents that contain a mixture of normal text with data analysis results. There has been a growing awareness that such documents should be developed as a mixture of text and data analysis **code**, from which data analysis results are generated. This is known as Literate Data Analysis and is supported by software tools such as Sweave and StatWeave. In this talk, I will describe a further extension of this idea, in which a document is developed as a set of chunks, some of which may be normal text, some of which may be computer code, and some of which may be other types of content. This approach will allow the document to be processed in a variety of ways to produce a wider range of results.

Keywords: XML, XPath, XInclude, XSLT

Informal Statistical Inference

Maxine Pfannkuch
The University of Auckland
pfannkuc@math.auckland.ac.nz

Recently, Chris Wild, Matt Regan (The University of Auckland), Nick Horton (USA), and I have been constructing a possible pathway across the curriculum from Year 10 to Year 13 to build up students' concepts of statistical inference. Informal inference includes the drawing of conclusions about populations from samples. Such conclusions mean that students must be aware of the difference between sample distributions and population distributions and have experience of sampling behaviour including sampling variability and the effect of sample size. In particular, I will focus on how sampling reasoning and the drawing of conclusions about populations from samples might be progressed across the curriculum levels using a visual approach.

New Approach to Robust Estimation

Maheswaran Rohan
Department of Conservation, Hamilton
mrohan@doc.govt.nz

In the classical model, we have standard methods such as MLE to estimate parameters, but in the robust literature this concept is lacking. We are trying to develop a unified method to estimate parameters for contaminated data. We wish to use maximum likelihood estimation, which is very sensitive to the outliers. In order to reduce the influence on estimates by outliers, we add another parameter-free known density function to the original density function. We employ the EM algorithm to estimate the parameters. In this talk, I will explain a general approach, using the particular problem of robust estimation of a location parameter. The influence function for this robust estimator is computed. From the result, we can see the influence function is bounded, hence the estimator is known as B-robust. Historic speed of light (1878) data will be investigated using our method for this talk. This result is compared with a standard robust estimator, known as the Huber estimator.

Protein expression experiments using iTRAQ: a unified protocol for design and analysis?

Kathy Ruggiero
The University of Auckland
k.ruggiero@auckland.ac.nz

Quantitative proteomics addresses questions beyond protein identification and incorporates questions of differential expression. Technological advances, such as isobaric labelling reagents (iTRAQ^{TM}), enable the simultaneous analysis of up to eight protein complexes in a single MuDPIT run. This goes some way towards addressing the large variability in protein expression measurements from different runs of the mass spectrometer on the same protein sample. Our extensive experience using the iTRAQ technology, however, shows us that challenges still remain in experimental design and data analysis. Parallels can be drawn between iTRAQ^{TM} and two-colour cDNA microarray experiments, with some of the experimental design and data analysis lessons learned from microarrays naturally carrying over to iTRAQ. However, unique challenges remain. I will explore the type of data that is produced by LC-MS/MS experiments using iTRAQ technology and will address how experimental design can be used to control for run-to-run variation, as well as potential differences in labelling efficiencies of the different iTRAQ reagents.

Developing a Geospatial Land Use Classification for New Zealand: A Review of Current Theory and Practice

Daniel Rutledge[1]; Robbie Price[1]; Sarah Cowell[2]; Craig Briggs[1]

[1]Manaaki Whenua Landcare Research, Private Bag 3127, Hamilton, New Zealand
[2]Geography, Regional, Environment, Statistics New Zealand, Private Bag 4741 Christchurch
rutledged@landcareresearch.co.nz

This project examined theoretical and practical approaches to land use classification to support development of a geospatial land use classification for New Zealand. Land use refers to the modification of and related activities on the land by people to sustain human life. We allocate and change land use to produce different combinations of goods and services in response to changing human needs and wants. Some changes are reversible, e.g. crop rotation, while other are not, e.g., urbanisation. Also some consequences of land use change or associated practices are easily observed, e.g., forest harvesting, while others are not, e.g., gradual loss of soil fertility through inappropriate management. Currently New Zealand lacks nationally consistent and comprehensive geospatial land use information. Developing a geospatial land use classification would help fill that gap by 1) meeting a range of needs, 2) helping underpin policy, planning, and resource management, and 3) contributing to reporting on progress towards sustainable development across scales within New Zealand.

Land use classification broadly falls into four types along a gradient from simple to complex. Categorical approaches depict land use as discrete classes with no relationships among classes. Hierarchical approaches depict land use as a nested set of classes. Higher level classes, e.g. urban, become divided into more specific classes, e.g., residential, commercial. Multi-dimensional approaches store multi-attribute information for defined unit areas and allow recombination to generate different land use classes. Descriptive or semantics classifications derive from linguistic theory and describe land use using a rich collection of words and formal grammar from which different interpretations (i.e. classifications) could be derived. International practice tends to favour the use of hierarchical classifications, and many countries have adopted official land use classifications. A few examples (Spain, American Planning Association) take a more multi-dimensional approach. Regardless of the approach most land use classifications internationally remain unimplemented or partially implemented. In New Zealand, four attempts at developing a land use classification have occurred. None of these attempts became officially adopted, although current land valuation practices include an unofficial land use classification system. Our results suggest that New Zealand would benefit from adopting a descriptive/semantic approach to land use classification. Such an approach appears the most promising given its ability to include a diverse range of information, to incorporate new information as it becomes available, and to generate a range of classifications, including reference or official classifications, to meet a variety of needs while retaining the capacity to translate and compare amongst them.

Keywords: categorical, classification, geospatial, hierarchical, land use, land use change, semantic, multi-dimensional, New Zealand

Do emerging markets provide diversification benefits for listed property investors?: An application of Zero Non Patterned Vector Autoregression

Laura Ryan
Australian National University
laura_ryan@fastmail.fm

By applying a Zero Non Zero Patterned Vector Autoregression model we test for Granger causality and instantaneous causal relationships over three different time periods: before the Asian Market Crash (October 1995 to June 1997), during the crisis (July 1997 to June 1998) and after the crises (July 1998 to March 2000). We examine the level of integration between the developed listed property markets of Australia, United States, United Kingdom, Japan, Hong Kong and Singapore and the emerging markets of China, Malaysia, Philippines and Taiwan. My analysis of the period during the Asian Market Crisis challenges the notion that Emerging Markets should be part of a well diversified listed property portfolio and that diversification benefits were in fact eroded at precisely the time that they were intended to provide some protection for the investor.

Testing homogeneity with survey data

Alastair Scott and Patricia Metcalf
The University of Auckland
scott@stat.auckland.ac.nz

Testing homogeneity is a common problem in survey work. We may, for example, want to compare results from different national surveys or different regions in a regionally stratified survey, or perhaps to check the agreement between interviewers based on interpenetrating subsamples. Tests of homogeneity and independence are identical in ordinary multinomial theory, of course, but this is not true for general survey designs. We look at the modifications needed to apply the usual chi-squared test of homogeneity to survey data. In general, these modifications require an estimate of the full covariance matrix of the vector of all cell proportions. In practice, we often only have estimates of the standard errors of the individual cell proportions and hence of the diagonal elements of this covariance matrix. We look at approximations to the distribution of the chi-squared statistic based on this limited information and at improvements to the naïve Pearson test. We apply the results to the Auckland Diabetes, Heart and Health Survey, a dual frame survey with one sample being a stratified random sample based on the electoral roll and the other a two-stage sample based on an area frame.

Implementation of Distributions in R

David Scott
The University of Auckland
d.scott@auckland.ac.nz

Distributions are fundamental to statistics and their implementation is likewise fundamental in statistical software environments and packages. In base there are 20 distributions implemented, at least in part. Besides these there are many distributions implemented in contributed packages. Within the contributed packages there are overlaps in functionality and differences in implementations of the same distribution. In this paper I will discuss the implementation of distributions in R and make some proposals directed towards a standard approach to implementation.

Fractional s^{n-k} Factorial BIB Designs

M. Shamsuddin and M. Albassam
Department of Statistics, King Abdulaziz University, Saudi Arabia
malbassam@hotmail.com

The partially confounded different s^{n-k} fractional factorial designs under complete orthogonal replicates of different groups of factorial effects constitute different BIB designs with detailed but same factorial and BIB result. The method with some examples of different levels and one application are given.

Multiple Confounded and Orthogonal Replicated Full Factorial BIB Designs

M. Shamsuddin and M. Albassam
Department of Statistics, King Abdulaziz University, Saudi Arabia
dr_md_shams2004@yahoo.com

The s^n factorial design, where s is a prime number like 2, 3, 5, etc., generally confounds a set of different orthogonal factorial effects with s^r blocks (n>r>1) under one replicate. Thus there will be a complete set of orthogonal replicates of different group of orthogonal factorial effects for constructing a partially confounded s^n factorial design in s^r blocks. The factorial design, in general, when it partially confounds different orthogonal sets of different groups of orthogonal factorial effects under the complete set of different orthogonal replicates constitutes a BIB design having their individual properties, same result and detailed information of all factorial effects. The procedure is illustrated with some examples and one practical application.

On the Speed of Random Matrices and their Eigenvalue Decompositions

Peter Smith
University of Canterbury
p.smith@elec.canterbury.ac.nz

Consider a complex Gaussian matrix with iid entries that forms a stochastic process over time. The corresponding complex Wishart process has applications in communications engineering and information theory. One important question concerning the Wishart matrix is how quickly it changes as the underlying Gaussian matrix evolves over time. In particular, various functions of the Wishart process, such as the eigenvalues and eigenvectors, are of interest. In this paper we present an analysis of the "speed" at which the processes change. Processes considered include the elements of the Gaussian matrix, the elements of the Wishart matrix and the eigenvalues and eigenvectors of the Wishart matrix. The results demonstrate the potential for extremely rapid changes in the eigenvectors and eigenvalues, relative to the underlying matrices.

Get a response without direct questioning? - Modelling could be your answer!

Olivia Son and Victoria Wei
Statistics New Zealand
olivia.son@stats.govt.nz

Is a survey Statistics New Zealand's only way to get the data it needs, given there are rich pools of data available from other sources such as the New Zealand Tax system? If using this pool of data we can deduce with confidence what a class of respondents' answers to the questionnaire are likely to be, we may not need to explicitly survey them. Statistics New Zealand is redesigning its biggest business survey, the Annual Enterprise Survey (AES). We have investigated using tax data from the Inland Revenue Department as a substitute for directly surveying some subpopulations of businesses. Can we simply copy and paste the corresponding tax value for the questionnaire value? If not, can we build a model to estimate the expected AES questionnaire response from the business tax data? We have evaluated both options and have developed a model for using tax data in the Annual Enterprise Survey.

Statistical methods for testing equality of survival probability with paired right-censored data

Pei-Fang Su, Yi-De Liao and Yunchan Chi
National Cheng-Kung University
spf@stat.ncku.edu.tw

Rather than comparison of entire survival curves, interest is often focused on the comparison at a fixed point in time, especially when the survival curves of the treatments are known to cross. Klein et al. (2007) have studied a number of methods for comparing two independent survival functions at a fixed point in time based on some transformations of the survival function. However, these tests are not appropriate for paired censored data since the intra-pair correlation are ignored. Therefore, their methods are modified to accommodate the possible dependence between matched subjects. Moreover, simulation studies are conducted to evaluate the performance of these tests. Finally, the modified tests are illustrated with an application to a diabetes data in Diabetic Retinopathy Study.

Nonparametric computation of survival functions in the presence of interval censoring

Stephen Taylor
The University of Auckland
stay020@aucklanduni.ac.nz

Censoring in time-to-event data poses a challenge in survival analysis. General interval censoring places an interval of doubt around each event time. Even when none of the event times are known exactly, useful analysis techniques are available. The nonparametric maximum likelihood (NPMLE) survival function provides a natural way to demonstrate how survival progresses over time. This talk covers recent research using constrained Newton methods, mixture distributions and non-negative least squares algorithms to efficiently solve the problem of finding the NPMLE survival function.

An Educational Experience

T. Rolf Turner
Starpath Project, University of Auckland
r.turner@auckland.ac.nz

Maori and Pasifika students on average attempt substantially fewer NCEA level 3 standards (particularly standards from the Approved List of subjects) than do Pakeha and Asian students. This fact has a substantial adverse impact on the chances of Maori and Pasifika students achieving University Entrance. A study was undertaken to investigate whether the deficit is due, at least in part, to the lack of availability of standards at schools attended by the majority of Maori and Pasifika students. An analysis of the data obtained tells a rather complicated tale. The number of standards available has “significant” but small impact on the number of standards attempted. Nevertheless there is substantial evidence that student performance is predicted by the availability of standards. Logistic models for success at achieving UE in terms of various possible predictors reveal complex relationships between performance, availability of standards, and ethnicity.

Fast Computation for Fitting Non-parametric and Semi-parametric Mixture Models

Yong Wang
Department of Statistics, The University of Auckland
yongwang@auckland.ac.nz

Non-parametric and semi-parametric mixture models are valuable tools for solving many nasty problems that involve heterogeneity in the population. While the maximum likelihood approach is straightforward, its computation has long been known as being difficult, if not intractable, due to the estimation of a distribution function defined on an infinite-dimensional space. In this talk, I will describe some fast algorithms that I recently developed for fitting these models; present the results of their use in several applications, including the over-dispersion problem, simultaneous hypothesis testing, the Neyman-Scott problem and mixed effects models; and discuss some implementation issues using R.

Exploring data on a rare threatened bird - the rock wren

Ian Westbrooke
Department of Conservation
iwestbrooke@doc.govt.nz

I will introduce a small threatened native bird - the rock wren, and provide resources on the bird itself to set the context. I will then explore a small dataset exactly as it came to me from remote mountains in Fiordland on this fascinating bird's population status. There are 24 numbers - counts from 12 grids of 25 hectares in each of 1984/5 and 2005. This data gives a great opportunity to create tables and graphs, and calculate means, by hand or using a tool like Excel, tasks suitable for both junior and senior classes. For year thirteen students, the data poses the question asked by the person who sent the data - what is a confidence interval for the average change in the population per grid square. A confidence interval can be created using either traditional or resampling techniques.

Poster

If nothing goes wrong is everything alright - confidence bounds for binomial rates when no events occur - an extension to multiple groups of observations

Ian Westbrooke
Department of Conservation
iwestbrooke@doc.govt.nz

The upper confidence limit or bound for the rate of occurrence of a rare event is sometimes a key focus of interest, for example for predicting risk. This issue arises in a number of contexts, such as the incidence of any by-catch on fishing trips, or of medical mishaps. In our case, we have data from monitoring survival of native birds during a poison operation aimed at pest mammals, and want to assess the risk that native birds will be inadvertently killed. If no events are observed out of n Bernoulli trials, for example all the birds observed of a species survive, then, based on a binomial model, standard results give a simple, exact upper confidence bound. We are interested in carrying out a meta-analysis on groups of data that have been collected in separate studies for the same species at varied places and times, all with no mortality. What information about bounds on risks can be gained without assuming the same binomial rate between the various groups? I will show that the upper bound calculated as if they were all in a single homogeneous group case can be interpreted as an upper bound on a weighted average of the binomial rates of mortality across the groups, taken as a fixed set.

Autoregressive conditional duration models: Model selection and inference

Gang Xie, Paul Cowpertwait and Daniel Walsh
Institute of Information and Mathematical Sciences, Massey University
j.xie@massey.ac.nz

The Akaike's Information Criterion (AIC) and Bayes' Information Criterion (BIC) (Schwarz(1978)) are applied to the problem of model selection when fitting a family of autoregressive conditional duration (ACD) models to high-frequency inter-trade time intervals. A range of ACD models are fitted to two data sets: the 1990 IBM New York stock exchange data and 2006 transaction data for the company Sime Darby BHD taken from the Kuala Lumpur stock exchange. Using the method of maximum likelihood, basic ACD models (Engle and Russel (1998)), threshold ACD models (Zhang et al. (2001)), and mixed distribution ACD model (Luca and Gallo (2004)) are fitted to each data set, and the BIC calculated for each fitted model. A mixed distribution ACD model, obtained using lognormal and gamma distributions for the residuals, is found to provide the best fit to the IBM data, which is supported by a quantile plot of the residuals. Based on BIC, the same model also provides the best overall fit to the Darby data, although in this case there was evidence of greater autocorrelation was in the residual series.

Statistical Education Workshops

(Limited to 20 participants per workshop. Priority given to teachers.)

Mike Camden (Statistics New Zealand)
New strengths in the Curriculum's Statistics

The statistics strand in the *Mathematics and Statistics* learning area of the new *Curriculum* is better designed for enjoyment, relevance and challenge than the current one. It has strong structures (like the statistical enquiry cycle, Levels 1 to 8), and fresh features (like multivariate data from Level 3, and re-sampling at Level 8). These structures and features will eventually make teachers' roles more interesting and easier. We will relate new structures and features to data and other resources that students and teachers can access. The contexts may include txt champions, netball players, dolphins, possums, paua, cheese, CO_2 levels, fuel use, population counts, and the new NZ Immigration survey. We will focus on this item from Level 8: 'finding, using and assessing appropriate models including linear regression for bivariate data', and apply John Tukey's statements about the importance of graphs to it.

Pip Arnold (Team Solutions, University of Auckland)
Growing Scatter-plots

This workshop looks at an introductory activity for scatter-plots. Census at school data is used in this activity.

Pip Arnold (Team Solutions, University of Auckland)
"Making the call"

An introductory look at starting to "make the call" from the comparison of box plots in order for students to be able to make informal inferences about populations from samples.

Maxine Pfannkuch (University of Auckland)
Drawing an inference from an experiment: Curriculum Level 8

In this workshop we will use the re-sampling method to assess the strength of evidence for making a claim when comparing two designs.

List of Participants

Edward Abraham	Dragonfly, Wellington	edward@dragonfly.co.nz
Heti Afimeimounga	The University of Auckland	hafi001@stat.auckland.ac.nz
Cherif Aidara	Victoria University of Wellington	cherif.aidara@mcs.vuw.ac.nz
Mohammad Albassam	King Abdulaziz University, Saudi Arabia	malbassam@hotmail.com
Sawsan Al-Shamaa	AIS St Helens, Auckland	sawsana@ais.ac.nz
Muhammad Nafees Anwar	Massey University, Palmerston North	m.n.anwar@massey.ac.nz
Judith Archibald	Statistics New Zealand, Wellington	Judith.Archibald@stats.govt.nz
Greg Arnold	Massey University, Palmerston North	arnoldg@landcareresearch.co.nz
Sweta Baldawa	HortResearch, Auckland	sbaldawa@hortresearch.co.nz
Rod Ball	Scion, Rotorua	rod.ball@scionresearch.com
Richard Barker	University of Otago	rbarker@maths.otago.ac.nz
Bill Bolstad	University of Waikato	bolstad@waikato.ac.nz
Jennifer Brown	University of Canterbury	jennifer.brown@canterbury.ac.nz
Sharon Browning	The University of Auckland	s.browning@auckland.ac.nz
Mike Camden	Statistics New Zealand, Wellington	mike.camden@stats.govt.nz
Catherine Cameron	AgResearch Ruakura	c.cameron@agresearch.co.nz
Kevin Chang	The University of Auckland	kcha193@aucklanduni.ac.nz
Lisa Chen	The University of Auckland	yche285@stat.auckland.ac.nz
Yun-Chan Chi	National Cheng-Kung University, Taiwan	ycchi@stat.ncku.edu.tw
Grace Chiang	Statistics New Zealand, Auckland	Grace.Chiang@stats.govt.nz
Austina Clark	University of Otago	aclark@maths.otago.ac.nz
Neil Cox	AgResearch Ruakura	Neil.Cox@AgResearch.co.nz
Walter Davis	Statistics New Zealand, Wellington	Walter.Davis@stats.govt.nz
Marcus Davy	HortResearch, Hamilton	mdavy@hortresearch.co.nz
Charmaine Dean	Simon Fraser University, Canada	dean@stat.sfu.ca
Ken Dodds	AgResearch Invermay	Ken.Dodds@AgResearch.co.nz
Mike Doherty	Statistics New Zealand, Wellington	Mike.Doherty@stats.govt.nz
Barbara Dow	DairyNZ, Hamilton	Barbara.Dow@DairyNZ.co.nz
Jamas Enright	The Treasury, Wellington	jamas.enright@treasury.govt.nz
Sarojinie Fernando	Massey University, Palmerston North	S.Fernando1@massey.ac.nz
Marie Fitch	Massey University, Auckland	m.fitch@massey.ac.nz
David Fletcher	University of Otago	dfletcher@maths.otago.ac.nz
Christopher Fonnesbeck	University of Otago	fonnesbeck@maths.otago.ac.nz
Sharleen Forbes	Statistics New Zealand and School of Government Victoria University	sharleen.forbes@stats.govt.nz
Kyoko Fukuda	University of Canterbury	k.fukuda@math.canterbury.ac.nz
Ganes Ganesalingam	Massey University, Palmerston North	s.ganesalingam@massey.ac.nz
Patrick Graham	Christchurch School of Medicine	patrick.graham@otago.ac.nz
Peter Green	University of Otago	pgreen@maths.otago.ac.nz
John Harraway	University of Otago	jharraway@maths.otago.ac.nz
Stephen Haslett	Massey University, Palmerston North	s.j.haslett@massey.ac.nz
Tim Hawkes	Statistics New Zealand, Wellington	Tim.Hawkes@stats.govt.nz
Martin Hazelton	Massey University, Palmerston North	m.hazelton@massey.ac.nz
Harold Henderson	AgResearch Ruakura	Harold.Henderson@AgResearch.co.nz
Ray Hoare	Hoare Research Software, Hamilton	ray@hrs.co.nz
Hans Hockey	Biometrics Matters Ltd, Hamilton	hans@biometricsmatters.com
Miriam Hodge	University of Canterbury	mch145@student.canterbury.ac.nz
Mark Holmes	The University of Auckland	mholmes@stat.auckland.ac.nz
Lyn Hunt	University of Waikato	lah@stats.waikato.ac.nz
Jeffrey Hunter	Massey University, Auckland and AUT University, Auckland	jeffrey.hunter@aut.ac.nz
Beatrix Jones	Massey University, Auckland	m.b.jones@massey.ac.nz
Murray Jorgensen	University of Waikato	maj@waikato.ac.nz
Clint Jowett	Offlode Ltd, Wellington	clint@offlode.com
Laimonis Kavalieris	University of Otago	lkavalieris@maths.otago.ac.nz
Estate Khmaladze	Victoria University of Wellington	Estate.Khmaladze@mcs.vuw.ac.nz
Vijayantimala (Vij) Kooyela	Te Puni Kōkiri (Ministry of Māori Development), Wellington	kooyv@tpk.govt.nz
Frances Krsinich	Statistics New Zealand, Wellington	Frances.Krsinich@stats.govt.nz
Lee Kupferman	Statistics New Zealand, Christchurch	lee-kurt.kupferman@stats.govt.nz
Debbie Leader	The University of Auckland	dleader_1@hotmail.com
Chang Li	Statistics New Zealand, Christchurch	Chang.Li@stats.govt.nz
Roger Littlejohn	AgResearch Invermay	Roger.Littlejohn@AgResearch.co.nz
Graham McBride	NIWA, Hamilton	g.mcbride@niwa.co.nz
Judi McWhirter	University of Waikato	judi@stats.waikato.ac.nz
Esther Meenken	Crop & Food Research, Lincoln	meenkene@crop.cri.nz

Carolyn Munro	University of Waikato	daphnia@xtra.co.nz
Paul Murrell	The University of Auckland	paul@stat.auckland.ac.nz
Bhaskaran Nair	Ministry of Education, Wellington	bhaskaran.nair@minedu.govt.nz
Maxine Pfannkuch	The University of Auckland	m.pfannkuch@auckland.ac.nz
Nathaniel Pihama	Statistics New Zealand, Wellington	nathaniel.pihama@stats.govt.nz
Eleanor Posadas	Child, Youth and Family, Wellington	Eleanor.GuzmanPosadas001@cyf.govt.nz
Steven Rae	University of Waikato	smr39@students.waikato.ac.nz
Maheswaran Rohan	Department of Conservation, Hamilton	mrohan@doc.govt.nz
Kathy Ruggiero	The University of Auckland	k.ruggiero@auckland.ac.nz
Daniel Rutledge	Landcare Research, Hamilton	rutledged@landcareresearch.co.nz
Laura Ryan	Australian National University	laura_ryan@fastmail.fm
David Scott	The University of Auckland	d.scott@auckland.ac.nz
Alastair Scott	The University of Auckland	scott@stat.auckland.ac.nz
Mohamed Shamsuddin	King Abdulaziz University, Saudi Arabia	dr_md_shams2004@yahoo.com
Peter Smith	University of Canterbury	p.smith@elec.canterbury.ac.nz
Wiremu Solomon	The University of Auckland	solomon@math.auckland.ac.nz
Olivia Son	Statistics New Zealand, Christchurch	Olivia.Son@stats.govt.nz
Pei-Fang Su	National Cheng-Kung University, Taiwan	spf@stat.ncku.edu.tw
Stephen Taylor	The University of Auckland	stay020@aucklanduni.ac.nz
Lisa Teague	Health Waikato	teaguel@waikatodhb.govt.nz
Gloria Teng	University of Canterbury	gat41@student.canterbury.ac.nz
Finlay Thompson	Dragonfly, Wellington	finlay@dragonfly.co.nz
Thomas Tran	The University of Auckland	thtam@stat.auckland.ac.nz
T. Rolf Turner	Starpath Project, University of Auckland	r.turner@auckland.ac.nz
Martin Upsdell	AgResearch Ruakura	Martin.Upsdell@AgResearch.co.nz
John Waller	AgResearch Ruakura	John.Waller@AgResearch.co.nz
Yong Wang	The University of Auckland	yongwang@stat.auckland.ac.nz
Victoria Wei	Statistics New Zealand, Auckland	Victoria.Wei@stats.govt.nz
Ian Westbrooke	Department of Conservation, Christchurch	iwestbrooke@doc.govt.nz
Wa Wrathall	Massey University, Palmerston North	t.wrathall@massey.ac.nz
Gang (John) Xie	Massey University, Auckland	j.xie@massey.ac.nz
Ge Yang	University of Waikato	gy21@waikato.ac.nz
Ilze Ziedins	The University of Auckland	i.ziedins@auckland.ac.nz

Statistical Education Day

Naznin Ali Basha	Ngaruawahia High School	nazninbasha@hotmail.com
Malabika Bhaduri	Papatoetoe High School	mabhaduri@papatoetoehigh.school.nz
Anne Blundell	Auckland Girls' Grammar School	ablundell@aggs.school.nz
Han Cham	Mount Albert Grammar School, Auckland	hcham@mags.school.nz
Monika Cross	Hamilton Girls' High School	crossmf@hghs.school.nz
Bethan De Malmanche	Waikato Diocesan School	bdemalmanche@waikatodiocesan.school.nz
Jason Ellwood	Otumoetai College	jellwood@otc.school.nz
Pamela Fisher	Birkenhead College	fs@birkenhead.school.nz
Kevin Foster	Waikato Diocesan School	kfoster@waikatodiocesan.school.nz
Lewis Hockings	Tauranga Girls' College	lhockings@tgc.school.nz
Jim Hogan	School Support Services	jimhogan@clear.net.nz
Clare Lambert	Waikato Diocesan School	clambert@waikatodiocesan.school.nz
Wendy Lowe	Morrinsville College	wendy-lowe@morrcoll.school.nz
Carl Mawer	Mt Maunganui College	carlm@mmc.school.nz
Jo Mawer	Otumoetai College	jmawer@otc.school.nz
Marina McFarland	Auckland Girls' Grammar School	mmcf002@gmail.com
Johanna McHardy	Mount Albert Grammar School, Auckland	jmchardy@mags.school.nz
Mary McLay	Rotorua Girls' High School	mmclay@rghs.school.nz
Judy Paterson	Senior College of New Zealand	j.paterson@auckland.ac.nz
Alex Randell	Bethlehem College, Tauranga	a.randell@beth.school.nz
Ingrid Rinsma	Hillcrest High School	IRinsma@hillcrest-high.school.nz
Miriam Robinson	Rangitoto College	robinsom@rangitoto.school.nz
Val Rolfe	Hamilton Girls' High School	rolfevm@hghs.school.nz
Jeanette Saunders	St Cuthbert's College	jeanette.saunders@stcuthberts.school.nz
Cynnelle Scully	Auckland Girls' Grammar School	cscully@aggs.school.nz
Michael Shadbolt	Otumoetai College	mshadbolt@otc.school.nz
Lindsay Smith	Epsom Girls Grammar Cchool	sm@eggs.school.nz
Jo Sturme	Rotorua Lakes High School	sturmej@rotorualakes.school.nz
Jean Tafa	Onehunga High School	jtafa@ohs.school.nz
Doug Udy-Bothwell	Rangitoto College	udybothwelld@rangitoto.school.nz

"Data and context in statistics education: towards an evidence-based society"

8th International Conference on Teaching Statistics
Ljubljana, Slovenia, 11–16 July 2010 Website: icots8.org

Preliminary Notice

The 2010 International Conference on Teaching Statistics will be held in the city of Ljubljana, Slovenia, July 11-16. It is being organised by the IASE and the Statistical Society of Slovenia. Types of contribution include **plenary addresses**, **invited papers**, **contributed papers, posters** and **workshops**.

Invited Papers

Invited Paper Sessions are organized within 10 Conference Topics as follows.

Topic Conveners with email contact are on the website.

1. Data and Context in Statistics Education: Towards an Evidence-based Society.
2. Statistics Education at the School Level.
3. Learning to Teach Statistics.
4. Statistics Education at the Post Secondary Level.
5. Assessment in Statistics Education.
6. Statistics Education, Training and the Workplace.
7. Statistics Education and the Wider Society.
8. Research in Statistics Education.
9. Technology in Statistics Education.
10. An International Perspective on Statistics Education.

76 Session themes within these Topics are now set up. The themes and Session organizers with email contact are available on the ICOTS-8 web site icots8.org/, under "Scientific Programme". Those interested in submitting an invited paper should contact the appropriate Session Organiser **before November 30, 2008**.

Contributed Papers

Those interested in submitting a contributed paper should contact the contributed paper organizers **before September 1, 2009.**

Posters

Those interested in submitting a poster should contact the poster organizers **before January 15, 2010.**

General Issues

More information is available from the ICOTS-8 web site at icots8.org/ which will continue to be updated over the next two years. The first circular should be released by early September, 2008.

John Harraway
Chair International Programme Committee ICOTS-8
jharraway@maths.otago.ac.nz