Contents

1 Chairman’s Foreword 3

2 Abstracts for Abstract-Only Presentations 5

2.1 Statistical Curvature in Two Dimensions – A Teaching Application
   Bernard Ellem, Kylie Thomas ............................... 5

2.2 Expected Shortfall and Value at Risk Via the Asymmetric Laplace Distribution
   Richard Gerlach ............................................... 5

2.3 Level-Correcting Hours Worked Time-Series Estimates
   Noel Hansen, Alan Payne, Mark Zhang, Nick Von Sanden 6

2.4 Applying a Smooth Test of Fit to Poisson Regression Data
   Paul Rippon, John Rayner ................................. 7

2.5 Spatial Small Area Estimation: Comparison of Different Approaches
   George Sofronov ........................................... 7

2.6 Using Social Network Information for Surveys
   Thomas Suesse ............................................. 8

2.7 Semiparametric Regression Models for the Probabilistic Index
   Olivier Thas, Jan De Neve ............................... 8

2.8 Teaching the Concept of Kurtosis in Introductory Statistics Courses Using Mathematica, Platypuses and Kangaroos
   Paul van Staden, M.T. (Theodor) Loots ................ 9

3 Abstracts for Reviewed Papers 11

3.1 Integrated Statistical Inference: The Amalgamation of Conventional and Bayesian Statistical Inference in Introductory Statistics Courses
   James Baglin, Cliff Da Costa ............................ 11
3.2 Improving and Contextualising Statistics Service Teaching: An Online Module-Based Teaching and Assessment Tool for Statistical Laboratory Sessions
   Anthony Bedford, Cliff Da Costa, Sam Glasson . . . . . . . 12
3.3 A Few Moments for Simple Correspondence Analysis
   Eric Beh . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
3.4 Nonparametric Tests for Randomized Block Data with Ties
   John Best, John Rayner, Olivier Thas . . . . . . . . . . . . . . . . . . . . . 13
3.5 Assessing Uncertainty in Environmental Sampling
   Jennifer Brown, Timothy Robinson . . . . . . . . . . . . . . . . . . . . . 13
3.6 Engaging Learners in Statistical Education: Some Whys and Hows
   Mike Camden . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
3.7 Performing a Quantitative Microbial Risk Analysis Using Second-Order Monte Carlo Simulation
   James Dawber, Jennifer Brown . . . . . . . . . . . . . . . . . . . . . . . . . 14
3.8 Versatile Regression: Simple Regression with a Non-Normal Error Distribution
   Ben Dean, Robert A.R. King . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.9 Modelling the Survival of Elite Australian Rules Footballers
   Sam Glasson, Anthony Bedford . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.10 Winning Them Over
    Peter Howley . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.11 Monitoring Clinical Indicators
    Peter Howley, Stephen Hancock, Megan Ford . . . . . . . . . . . . . . . . . 16
3.12 Simple and Multiple Ordered Correspondence Analysis to Evaluate Customer Satisfaction
   Rosaria Lombardo, Eric Beh . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
3.13 Bayesian Inference for an Extreme Value Mixture Model
   Anna MacDonald, Carl Scarrott, Dominic Lee . . . . . . . . . . . . . . . . 17
3.14 Teaching Statistical Consulting: Enabling skills Transfer and Adding Value
    Kaye Marion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
3.15 Testing the Poisson Assumption in Poisson Regression
   John Rayner, Paul Rippon . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
3.16 Retention of Knowledge Between Statistics Courses: Results of a Pilot Study
   Alice Richardson . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
3.17 Scheduling in Australian Rules Football: Does it Affect Performance?
   Richard Ryall, Anthony Bedford ........................................... 19
3.18 Development of Users’ Call Profiles using Unsupervised Random Forest
   Almah Saaid, Robert A.R. King, Darfiana Nur ............................ 19
3.19 Extensions to Poisson Regression for Hospital Admissions Data
   Liz Stojanovski, Ian Robinson .............................................. 19
3.20 Diagnostic Testing: A Simple Approximate Interval for the Positive Predictive Value
   Frank Tuyl ........................................................................... 20
3.21 Pricing Wesfarmers Partially Protected Shares
   Katherine Uylangco, Steve Easton ........................................... 20
3.22 Method of $L$-Moment Estimation for the Generalized Lambda Distribution
   Paul van Staden, M.T. (Theodor) Loots ................................... 21
Organising Committee

Chairman
    David Allingham

Technical Chairman, Teaching Sessions
    Peter Howley

Technical Chairman, Modelling Sessions
    John Rayner

Travel and Other Assistance
    Vicki Gumbleton
    Anna-Rosa Gualtieri
    Eric Beh
    Rebecca Smith

Reviewers
    Allingham, David
    Beh, Eric
    Best, John
    Brockoff, Per
    Colyvas, Kim
    Dean, Ben
    Ellem, Bernard
    Gerlach, Richard
    Griffiths, David
    Gulati, Chandra
    Howley, Peter
    King, Robert
    Lombardo, Rosaria
    Marion, Kaye
    Martin, Michael
    McCrae, Michael
    Moffiet, Trevor
    Rayner, John
    Richardson, Alice
    Rippon, Paul
    Russell, Ken
    Smith, Rebecca
    Steel, David
    Stojanovski, Liz
    Thas, Olivier
    Tuyl, Frank
    van Staden, Paul
Chapter 1

Chairman’s Foreword

In December, 2009, the statistics group at the University of Newcastle was proud to host the Third Annual ASEARC Research Conference. Along with members of ASEARC, an equal number of national and international delgates attended, including invited speakers from around the world. The two themes of the conference were Modelling and Statistics Education, and both abstract-only and full paper submissions were considered. In the latter case, each full paper was peer-reviewed by at least two referees. I would like to take the opportunity of thanking our reviewers for their hard work, as well as for promptly meeting deadlines.

This document includes the abstracts of all presentations made at the conference, and the full papers accepted can be found in the “papers” directory on this Proceedings CD-ROM.

David Allingham
Chairman,
Organising Committee,
Third Annual ASEARC Research Conference
Chapter 2

Abstracts for Abstract-Only Presentations

2.1 Statistical curvature in two dimensions – a teaching application

Bernard Ellem, The University of New England, Kylie Thomas, Charles Sturt University

This report presents a novel method of teaching curvature measures via simple 2D examples using the elementary formulae for curvature encountered by undergraduate students. Using this procedure students can verify algebraically and numerically the invariance of intrinsic curvature and corroborate the best parameterisation by examination of parameter effects curvature, again mathematically and empirically. For serious users of curvature measures, this elementary exposition also reconciles the general definition of 'statistical curvature' coined by Efron (1975), with the approach of Amari (1990), albeit for the exponential connection only.

2.2 Expected Shortfall and Value at Risk Via the Asymmetric Laplace Distribution

Richard Gerlach, The University of Sydney

A parametric approach to forecasting Value-at-Risk (VaR) and Expected Shortfall (ES) for a heteroskedastic financial return series is presented. A
GJR-GARCH models the volatility process, capturing heteroscedasticity and the leverage effect. To capture potential skewness and heavy tails, the model assumes an asymmetric Laplace as the conditional distribution of the financial return series. Furthermore, dynamics in higher moments are captured by allowing the shape parameter in this distribution to be time-varying. The model parameters are estimated via an adaptive Markov Chain Monte Carlo (MCMC) sampling scheme, employing the Metropolis-Hastings (MH) algorithm with a mixture of Gaussian proposal distributions. A simulation study highlights accurate estimation and improved inference of parameters compared to a single Gaussian proposal MH method. We illustrate the model by applying it to return series from four international stock market indices, as well as the exchange rate of the AU dollar to US dollar, and generating one-step ahead forecasts of VaR and ES. We apply standard and non-standard tests to these forecasts and find that it is highly competitive with a range of popular alternatives.

2.3 Level-Correcting Hours Worked Time-Series Estimates

Noel Hansen, Alan Payne, Mark Zhang, Nick Von Sanden, The Australian Bureau of Statistics

The Labour Force Survey produces estimates of hours worked by Australian employed persons during just one week of the reference month or quarter. There is a strong user demand for estimates of hours worked during entire years and seasonally adjusted time-series estimates of hours worked during entire months or quarters for labour productivity measures. Users are also interested in these hours worked estimates for detailed categories including industry, sex and type of employment. In this seminar, we’ll present how a range of simple ‘level-correcting’ methodologies were combined to produce a suite of synthetic flow estimates of hours worked for varying durations and categories. We’ll explore some of the problems encountered and how they influenced the methodologies chosen to be applied to each time-series.

Keywords: benchmarking, time-series
2.4 Applying a Smooth Test of Fit to Poisson Regression Data

Paul Rippon, J.C.W. Rayner, *The University of Newcastle*

A recently derived smooth goodness of fit test for the Poisson regression model is applied to a cancer study example. Performance of the test using bootstrap p-values is shown to be comparable to the Anderson-Darling and Cramer-von Mises tests. It is also shown that the components of the smooth test statistic offer diagnostic information about the nature of any difference from the hypothesized Poisson distribution.

2.5 Spatial Small Area Estimation: Comparison of Different Approaches

George Sofronov, *The University of Wollongong*

The Australian Bureau of Agricultural and Resource Economics (ABARE) has been conducting farm surveys annually since 1978. The surveys collect detailed financial, physical and socioeconomic information from farm businesses across Australia.

The purpose of our study is how to produce more efficient estimates of means within small regions or domains utilizing spatial covariate information. For testing of small area models unit record farm level survey data from the wheat-sheep zone for the survey years 1978–79 to 1994–95 inclusive were used. Synthetic coordinates (longitude, latitude) for each farm were obtained by using pairwise distances between farms. The variable of interest was average Total Cash Receipts (TCR) within small areas.

We compared the performance of several different methods of small area estimation (SAE) such as

- empirical best linear unbiased prediction (EBLUP),
- spatial empirical best linear unbiased prediction (SEBLUP),
- non-parametric empirical best linear unbiased prediction (NEBLUP),
- M-quantile (MQ) models,
- M-quantile geographically weighted regression (MQGWR) models.

The Root Mean Squared Error (RMSE) was computed as a measure of estimation performance of the predictors.
2.6 Using Social Network Information for Surveys

Thomas Suesse, The University of Wollongong

Standard statistical approaches for social statistics obtained from sample surveys, censuses and administrative sources usually focus on individuals. For example in a study of smoking, drinking and exercise behaviour a survey may collect information on these variables and other social and demographic variables for a sample of individuals. Applying naively a single level analysis assuming independent individuals is inappropriate and can lead to incorrect inferences and conclusions. A more appropriate approach is a multilevel one, with levels corresponding to households and geographic areas.

Here we focus on the inclusion of social networks as another level, since any two individuals having ties in a social network are assumed to behave more similarly than two individuals that do not. In particular, we propose a general linear model where the covariance depends on the social network using different strengths of ties, i.e. direct and indirect links, households and geographical areas.

We consider a simulation study using ML estimation to investigate the importance of the inclusion of social networks for the sampling process. Ideally the inclusion of social networks would lead to more accurate statistical inference. We will present ideas how to tackle these problems.

2.7 Semiparametric Regression Models for the Probabilistic Index

Olivier Thas, Jan De Neve, Ghent University

Many textbooks introduce the Wilcoxon rank sum test, which is equivalent to the Mann-Whitney test, as the nonparametric alternative to the 2-sample t-test for testing equality of means, and some textbooks say that the hypotheses for the test are not formulated in terms of means, but rather in terms of medians. None of these statements is correct, unless some restrictive distributional assumptions are imposed (e.g. location-shift model). Instead the Wilcoxon-Mann-Whitney (WMW) test tests the general two-sample null hypothesis $H_0 : F_1 = F_2$ against the alternative $P(Y_1 \leq Y_2) \neq 1/2$, where $Y_1$ and $Y_2$ are independent and have distribution functions $F_1$ and $F_2$, respec-
tively. The probability \( P_{Y_1 \leq Y_2} \) is known as the probabilistic index, and it is considered as a very informative effect size parameter.

In this presentation I will introduce a family of semiparametric models for the probabilistic index. First we need some notation: let \( Y \) denote the response variable and let \( Y(x) \) denote a random variable which is distributed as \( Y \) conditional on the \( p \)-dimensional covariate pattern \( x \). In particular, we introduce the models

\[
g(Y(x_1) \leq Y^*(x_2)) = x^T \beta, \tag{2.1}
\]

where \( g \) is a link function and \( x \) is a \( q \)-dimensional vector containing elements from \( x_1 \) and \( x_2 \), \( \beta \) is a \( q \)-dimensional parameter vector, and \( Y(x_1) \) and \( Y^*(x_2) \) are two independent random variables with distributions conditional on \( x_1 \) and \( x_2 \), respectively. This is obviously a semiparametric model as it only restricts the conditional distributions of \( Y \) and \( Y^* \) to satisfy (2.1). When \( x_1 = 1 - x_2 \) is a dummy variable that indicates two treatment groups, model (2.1) can be used to test the WMW hypotheses. From this point of view the new semiparametric regression framework is an extension of the WMW test.

I will present semiparametric theory that gives a consistent estimator of \( \beta \) and its asymptotic properties, as well as a consistent estimator of its covariance matrix. A simulation study is presented in which the theoretical properties of the estimators are assessed. Finally the new semiparametric regression model is illustrated on two data examples.

### 2.8 Teaching the Concept of Kurtosis in Introductory Statistics Courses Using Mathematica, Platypuses and Kangaroos

Paul J. van Staden, M.T. (Theodor) Loots, The University of Pretoria

The shape of a statistical distribution is described via the skewness and the kurtosis of the distribution. Like location and spread, the concept of skewness is easily taught in introductory statistics courses. Kurtosis however is not a simple characteristic in that it is related to both the tails and the peak of the distribution under consideration. The topic of kurtosis is therefore often avoided in introductory courses and textbooks. We briefly present the history of kurtosis. We then focus on Pearson’s measure of kurtosis and clarify two misconceptions regarding its use and interpretation. Using a
new function in Mathematica, we show how the concept of kurtosis can be illustrated graphically to students and combine our graphical technique with a recently proposed method for calculating the excess tail and peak areas.

Keywords: Gaussian distribution, generalized secant hyperbolic distribution, Kotz-Seier method, kurtosis measures, mesokurtic, platykurtic and leptokurtic distributions, Pearson’s kurtosis moment ratio, statistical education, tail and peak areas, uniform distribution
Chapter 3

Abstracts for Reviewed Papers

3.1 Integrated Statistical Inference: The Amalgamation of Conventional and Bayesian Statistical Inference in Introductory Statistics Courses

James Baglin, Cliff Da Costa, *RMIT University*

Many educators of statistics have considered the idea of introducing Bayesian statistical inference into so called “Bayes for Beginners” courses. There are also many others who have cautioned against doing so, citing the widespread acceptance of conventional statistical methods as a reason to hesitate. A good compromise would see both methods being taught in an integrated fashion to give a student the best of both inferential worlds. This paper will briefly overview Integrated Statistical Inference (ISI), a method for delivering both Bayesian and conventional concepts in an introductory statistics course.

Keywords: Bayesian inference, conventional inference, introductory statistics course
3.2 Improving and Contextualising Statistics Service Teaching: An Online Module-Based Teaching and Assessment Tool for Statistical Laboratory Sessions

Anthony Bedford, Cliff Da Costa, Sam Glasson, RMIT University

The laboratory sessions of service-taught statistics courses are a critical component of achieving statistical capability. The laboratory sessions should challenge the students' ability to apply their knowledge to contextualised data, and equip them with the necessary tools and skills to undertake data analysis in any future course or work-related context. Traditionally, statistics laboratory sessions are taught with a teaching/demonstration focus. Through feedback from courseware evaluations, a teaching via demonstration focus was replaced by a student self-guided learning approach with tutorial support. This approach improved interactive learning and feedback, however necessitated the expenditure of extra time and use of tutors in both supervision and assessment in the lab. In this paper we discuss the features of a completely online laboratory module system, encompassing both self-guided sessions and assessment, within a single framework. This moves the assessment process from the current subjective assessment to objective online assessment. Three years feedback from students indicates that the system provides a clearer teaching and grading process compared to classic lab systems. Results also indicate that the laboratory modules were completed faster than anticipated, with time saved leading to the ability to extend material beyond the current levels.

Keywords: teaching statistics, online learning, SPSS

3.3 A Few Moments for Simple Correspondence Analysis

Eric J. Beh, The University of Newcastle

This paper provides a simple interpretation of the coefficients of skewness and kurtosis of points in a correspondence plot. It helps to provide further information on the configuration of the coordinates and adds to that already provided by the first two moments which have helped to form the foundations
of the mathematical development of correspondence analysis.

Keywords: correspondence analysis, total inertia, skewness, kurtosis

3.4 Nonparametric Tests for Randomized Block Data with Ties

D.J. Best, J.C.W. Rayner, The University of Newcastle, O. Thas, Ghent University

Umbrella, Page and Friedman tests are defined and discussed for randomized block designs. The data may be tied. Two alternative derivations of the Page and umbrella test statistics are given. Two sensory evaluation examples are considered.

Keywords: binary responses, orthogonal contrasts, ranks, sensory evaluation

3.5 Assessing Uncertainty in Environmental Sampling

Jennifer Brown, The University of Canterbury, Timothy Robinson, The University of Wyoming

Uncertainty in estimating a true population parameter from survey data is a result of imperfect detection, imperfect observation, spatial and temporal variation, and sampling error. Some level of uncertainty is inevitable in all surveys. We discuss the sources of uncertainty in surveys of bird counts and show how we have used Monte Carlo simulations to illustrate the effect of uncertainty.

Keywords: survey design, Monte Carlo simulation, variance

3.6 Engaging Learners in Statistical Education: Some Whys and Hows

Mike Camden, Statistics New Zealand

Statistical educators have increasingly vital roles in equipping people for both work and life, and so we have an increasingly vital role in engaging them
as learners of statistics. Luckily, the applied nature of our subject supports learner engagement, and there are many rapidly developing communication methods that we can harness to support it as well. We will view some illustrations of these from school, workplace, and official statistics. The methods may come from outside the tertiary sector, but they have implications for it.

The illustrations will include: data exploration software for 9-year-olds and beyond, media-based activities to engage 12-year-olds, videos and datasets from scientists for 17-year-olds, aims for school curricula, aims for workplace statistical education, methods for data visualisation, informal inference, resampling methods, treatment of hot issues like climate change, and the sea of information available from official statistics agencies.

I hope we will conclude that engaging learners in our subject is vitally important, that there are plenty of smart and developing ways of supporting this engagement, and that many of these methods need careful design, dedicated effort, and large amounts of teamwork.

Keywords: statistics, learning, media, visualisation

3.7 Performing a Quantitative Microbial Risk Analysis Using Second-Order Monte Carlo Simulation

James Dawber, Jennifer Brown, The University of Canterbury, Beverley Horn, Institute of Environmental Science and Research

Quantitative microbial risk analysis modelling is increasingly being used in food safety as a tool to evaluate health risks. Accurately forming such models can be very difficult due to the uncertainty in the available data. Second order Monte Carlo simulation allows for the inclusion of this uncertainty and separates it from the variability incorporated in the model. This modelling process is illustrated by performing a simple risk assessment on the infection of Campylobacter during chicken preparation at a typical New Zealand barbecue.

Keywords: quantitative microbial risk analysis, Campylobacter, variability, uncertainty
3.8 Versatile Regression: Simple Regression with a Non-Normal Error Distribution

Benjamin Dean, Robert A.R. King, *The University of Newcastle*

We present a simple regression technique, called Versatile Regression, where the error distribution is described by the Generalized Lambda Distribution. The flexibility of this distribution allows the error distribution to be heavy-tailed, skewed or approximately normal. Versatile Regression was found to perform well on heavy-tailed and skewed data. Versatile Regression also provided a reasonable approximation to Normal-Error Regression. Simulation studies found that Versatile Regression produced accurate parameter estimates.

Keywords: simple regression, Generalized Lambda distribution, non-normal error distribution

3.9 Modelling the Survival of Élite Australian Rules Footballers

Samuel Glasson, Anthony Bedford, *RMIT University*

The length of player careers is modelled using Survival Analysis. Survival and hazard functions were constructed for the cohort of players who were first drafted between years 1995 and 2008. A Cox regression was run to determine how career length is affected by draft rank. Consequences for drafting strategy and possibilities for future research are discussed.

Keywords: survival analysis, Kaplan-Meier estimator, Cox regression, Australian Rules Football

3.10 Winning Them Over

Peter Howley, *The University of Newcastle*

This paper provides examples of a method to engage and inform a student. The method is based on enabling students to uncover, by themselves, what they did not realise they already knew. The technique involves the students answering a series of questions which ultimately results in the student explaining a concept or method in lieu of simply being told by the lecturer.
This helps to lessen the mystery that otherwise acts as a barrier to students. understanding of, and progression through, the field of Statistics.

Keywords: improved learning methods, increasing student confidence, pedagogy, teaching introductory statistics

3.11 Monitoring Clinical Indicators

Peter Howley, Stephen Hancock, Megan Ford, The University of Newcastle

A new control chart for monitoring clinical indicator (CI) data based upon the beta-binomial posterior predictive (BBPP) distribution was compared with the more commonly used Bernoulli cumulative sum (Bernoulli CUSUM) chart. Run lengths were simulated for 3,894 parameter combinations. For the case where the underlying proportion of cases with an event of interest had to be estimated, the BBPP chart was shown to have the desired smaller out-of-control ARL in 71.6% of the simulations. This effect was greatest in the parameter space having the:

- mean proportion across all healthcare providers (HCPs), $\pi$, $< 0.1$;
- percentage change in the underlying proportion (required for Bernoulli CUSUM chart), $\Delta$, $0.15 < \Delta < 0.35$;
- number of admissions at risk of the event of interest at the $i$th HCP, $D_i$, $< 30$;
- proportion of admissions having the event of interest at the $i$th HCP, $\theta_i$, $< 0.1$;

across all values for the standard deviation between HCPs, $\sigma$.

Keywords: Bayesian, Bernoulli CUSUM, beta-binomial model; posterior predictive models

3.12 Simple and Multiple Ordered Correspondence Analysis to Evaluate Customer Satisfaction

Rosaria Lombardo, The Second University of Naples, Eric J. Beh, The University of Newcastle

In marketing research, to evaluate customer satisfaction, individuals are
often required to fill in questionnaires where responses are ordered. To analyze the association of these ordered categorical variables, an alternative method to the usual multiple correspondence analysis has been adopted. It is based not only on the classic singular value decomposition but on a hybrid decomposition.

Keywords: correspondence analysis, ordered categorical variables, singular value decomposition, bivariate moment decomposition, hybrid decomposition

### 3.13 Bayesian Inference for an Extreme Value Mixture Model

Anna MacDonald, Carl Scarrott, Dominic Lee, The University of Canterbury

Extreme value theory is used to derive asymptotically-motivated models for unusual or rare events, e.g., the upper or lower tails of a distribution. A new, flexible extreme value mixture model is proposed combining a nonparametric kernel density estimator with an appropriate tail model, which overcomes the key issue of determining the threshold which defines the distribution tail and accounts for uncertainty due to the threshold choice. Bayesian inference is used to account for all uncertainties and enables inclusion of expert prior information, potentially also overcoming the inherent sparsity of extremal data. An application to estimating the quantiles of daily FTSE log returns is used for demonstration.

Keywords: extreme value analysis, Bayesian inference, mixture model, kernel density

### 3.14 Teaching Statistical Consulting: Enabling skills Transfer and Adding Value

Kaye E. Marion, RMIT University

A statistical consultant needs to understand the scientific process in order to develop models suitable for testing, have the quantitative expertise to apply the statistical test appropriately and communicate the issues involved effectively. A successful consultant will not only address the problem but
will teach the client something about statistics and with this transfer of skills enable the client to make better decisions in the future.

This paper will discuss the skills that need to be acquired by a statistical consultant and cover some techniques developed to teach these skills in the Master’s Degree in Statistics and Operations Research at RMIT University in the first semester of 2009. The author draws on over 40 years as a consultant developing testable models for medical researchers, business analysts, government policy makers and academic researchers with examples where ingenuity was required to handle non-standard decision problems.

Keywords: statistical consulting, graphs, communicating statistical concepts, statistical thinking, problem solving

3.15 Testing the Poisson Assumption in Poisson Regression

J.C.W. Rayner, Paul Rippon, The University of Newcastle

A smooth test of the Poisson assumption in the Poisson regression generalised linear model is derived.

Keywords: generalised linear models, goodness of fit, leverage matrix, smooth tests

3.16 Retention of Knowledge Between Statistics Courses: Results of a Pilot Study

Alice Richardson, The University of Canberra

This paper reports on a pilot study of knowledge retention from the first to subsequent statistics courses. The first aim of the study was to assess the ability of a Statistics Concept Inventory (SCI) to measure the statistical knowledge of students after their first Statistics course. The second aim was to use the Inventory to assess students’ retention of material from an introductory Statistics course into a subsequent statistics course. The usefulness of the SCI was partially supported, as many questions were better answered by the more able students. Retention was found to vary across the topics in the SCI.

Keywords: Statistics Concept Inventory, statistics education
3.17 Scheduling in Australian Rules Football: Does it Affect Performance?

Richard Ryall, Anthony Bedford, RMIT University

This paper examines the fairness of the schedule of games in Australian Rules football. Using ordinary least squares regression, the effect of different numbers of days preparation by any team can be quantified, while allowing for the home ground advantage of not having to travel interstate. The results suggest that, for 2007 to 2009, the effect of differing number of days preparation is independent of interstate travel. Furthermore, an in-game (quarter by quarter) analysis shows the effect differing number of days preparation is more crucial earlier in the match, meaning there is no significant fadeout for teams coming off a shorter break than their opposition.

Keywords: performance analysis, scheduling, sport, Australian Rules Football

3.18 Development of Users’ Call Profiles using Unsupervised Random Forest

Fatimah Almah Saaid, Robert A.R. King, Darfiana Nur, The University of Newcastle

The aim of this paper is to detect fraud in telecommunications data which consists of millions of call records generated each day. The fraud detection is implemented via the construction of user call profiles using the calls detail records (CDR) data. This paper attempts to investigate the reliability of the unsupervised Random Forest method in building the profiles using its variable importance measure. Four different simulation scenarios, using different number of variable selection in each node of the tree, are performed.

Keywords: telecommunications fraud, users call profiles, unsupervised random forest, variable importance

3.19 Extensions to Poisson Regression for Hospital Admissions Data

Liz Stojanovski, Ian Robinson, The University of Newcastle
Presented are models for length of hospital stay data. The Hurdle model is an extension to the Poisson model when the data structure can be considered as two separate processes as was evident with the present case study. Hospital data was considered and collected as part of a larger study over a five year period. The Hurdle model appeared the most appropriate in terms of overall goodness of fit.

Keywords: length of stay, Poisson, zero-inflated

3.20 Diagnostic Testing: A Simple Approximate Interval for the Positive Predictive Value

Frank Tuyl, The University of Newcastle

In diagnostic testing, given specificity, sensitivity and prevalence, Bayes’ rule gives the so-called positive predictive value. Generally there are two potential pitfalls that have been identified previously. First, it is incorrect to estimate prevalence from the usual $2 \times 2$ table if the data are from a case control study. Secondly, it would usually seem inadequate to only use point estimates, without considering uncertainty.

In relation to these pitfalls, it is shown that if in fact the prevalence suggested by the $2 \times 2$ table is reasonable, in a Bayesian context a simple beta distribution may be adopted to obtain a credible interval for the positive predictive value, as an approximation of an existing objective Bayesian solution based on Monte Carlo simulation.

Keywords: Bayes-Laplace prior, Jeffreys prior, predictive value, sensitivity, specificity

3.21 Pricing Wesfarmers Partially Protected Shares

Katherine Uylangco, Steve Easton, The University of Newcastle

In April 2007, Wesfarmers announced a proposal to acquire the entire issued capital of Coles. Part of the eventual offer included Coles’ shareholders being given Wesfarmers shares with partial price protection (WESN shares). This paper provides a theoretical valuation for WESN shares and compares these
prices with those observed over the first year of trading. It shows that the modelled price of WESN shares was generally higher than the observed prices. Model assumptions regarding the risk free rate of return, expected dividends, expected volatility and the correlation between Wesfarmers shares and the S&P/ASX 200 All Industrials Index do account for some of this difference, in particular the volatility assumption. An alternative explanation of the difference is that there are several features of the WESN share offer that cannot be incorporated into the model that detract from the value of WESN shares, thus creating a lower market value than would otherwise be expected.

Keywords: option pricing, Monte Carlo simulation, barrier options, rainbow options

3.22 Method of L-Moment Estimation for the Generalized Lambda Distribution

Paul J. van Staden, M.T. (Theodor) Loots, The University of Pretoria

The generalized lambda distribution (GLD) is a flexible distribution for statistical modelling, but existing estimation methodologies for the GLD are computationally difficult, rendering the GLD impractical for many practitioners. We derive a parameterization of the GLD with closed-form expressions for the method of L-moment estimators. A detailed numerical example involving the age of coronary heart disease patients is presented.

Keywords: generalized Pareto distribution, L-moment ratio diagram, quantile function, skew-logistic distribution