In this issue

VIC BRANCH

Evidence-based Policy: issues and challenges

For our September 16th function we organised a public lecture given by Dr Adam Bandt (the Federal Member for Melbourne and Deputy Leader of the Australian Greens) on the ideal of having policy based on well informed expert knowledge; not just on political considerations and opinions. The format of this function comprised a presentation by Adam on his perspective of the state of evidence-based policy in Australia followed by a Q & A session moderated by well-known TV and radio broadcaster Gael Jennings. The following is a synthesis of the combined contents of both sections.

Adam started by describing his experience in his first term in trying to get a climate change policy passed through parliament. A committee was formed consisting of both politicians and experts in science, economics and social issues. Adam was very impressed by this structure as when seated around a table, politicians are generally not prepared to dispute the knowledge of experts because they can be "held to account very quickly" if they do.

Adam was pleased with the resultant policy; but not with the opposition to it which he described as one of manufacturing doubt - of attacking the credibility of the science and the experts rather than debating the policy based on the evidence. Adam explained that this practice has been used extensively by the tobacco industry; and it can have a flow on effect by putting scientists under threat (as happened recently in WA where they were stifled in the shark kill debate). He saw it as a major impediment to the greater use of scientific input in political decision making.

Possible ways Adam saw to increase the use of scientific evidence in policy development were:-

• By first defining the boundaries between where scientific evidence should and should not be used,

• By entrenching the role of facts by law. This could be done for example by the establishment of expert advisory committees providing advice based on scientific evidence in combination with legislation mandating that this evidence and advice be taken into account by the appropriate minister in all relevant major decisions,

• Through randomised trials (eg. as has been done in testing anti-smoking policies in Indigenous communities),

EDITORIAL

This is the last issue of the newsletter for 2014, and the last issue for us as the current Editors.

My first issue was August 2001, assisting Editor Eden Brinkley, and when Eden finished in May 2003 I stepped up to be Editor-in-Charge assisted by Michael Adena. I'm the ninth Editor-in-Charge since issue 1 in 1977. I've been involved with 54 newsletters – that's 54 editorials, 54 President's columns to chase down, seven Australian Statistical Conferences to report on and well over 250 Branch reports … with only a handful of printed corrections! In broad terms the newsletter has maintained its page count throughout that period, with an always-changing kaleidoscope of workshop reports, book advertisements and section reports to support the backbone of national and branch activities. I've really noticed the increase in Branch speakers from private industry over the last ten years, and I expect that into the future this is a trend that will continue.

We approved design changes in June 2006, September 2008, March 2009 and March 2011. The 2009 change was probably the most important as it accompanied the change to online dissemination. We did worry about whether members would miss that paper copy of the newsletter to read over coffee, or share around the office, the pdf can still be printed out and the with the number of people you see on public transport reading an electronic device, it's possible readership has even increased with the new availability. We also made the physical move from the Society's offices above the Ainslie shops to the purpose-built ABS House with all the attendant changes in addresses and telephone numbers.

My favourite issue was issue 100 in August 2002. We reprinted the 1st newsletter in its entirety, and the front page featured a photo of the first editor, Dennis Trewin, with Eden Brinkley.

I have very few regrets as Editor, but if pressed to come up with one it would be the difficulty in attracting reports and articles that deviated in any way from the typical meeting reports, conferences, calls for papers, section activities. I really hope that members can think of putting pen to paper for the newsletter on a whole variety of topics of interest to members – we've run a very inclusive editorial policy that more members could have taken advantage of over the years.

I'd like to thank all the contributors to the newsletter in the last 12 years: from the Presidents and other office bearers who provide necessary material for the correct functioning of the Society; to the Branch correspondents (including the Statistical Squirrel of the early 2000s) who have to concentrate hard enough during the Branch meetings in order to write them up for the next newsletter. Thank you also to all our photographers, whose snaps have made the newsletter a much warmer and more personal read over the time I have been Editor. It's great to be able to put faces to names from around the country.

The Editors would like to take this opportunity to wish all Society members a happy Christmas, and a pleasant New Year. Thank you also to the organisations who have supported the activities of the Statistical Society and its members – your input into successful Society activities is much appreciated. In particular, we would like to thank the Australian Bureau of Statistics for hosting the SSAI office in ABS House in Canberra.

Alice Richardson and **Michael Adena**

SECTION CHAIRS

Bayesian Statistics
Chair: Scott Sisson
Scott.Sisson@unsw.edu.au
Assistant Chair: Jannah Baker
Jannah.Baker@qut.edu.au
http://www.statsoc.org.au/bayesian-statistics.htm

Environmental Statistics
Chair: David Clifford
David.Clifford@csiro.au
Assistant Chair: Mayukh Samanta
Mayukh.Samanta@qut.edu.au
http://www.statsoc.org.au/environmental-statistics.htm

Social Sciences
Chair: Michele Haynes
M.Haynes@uq.edu.au
http://www.statsoc.org.au/social-statistics.htm

Statistical Education
Co-Chair: Michael Martin
Michael.Martin@anu.edu.au
Co-Chair: Peter Howley
Peter.Howley@newcastle.edu.au
http://www.statsoc.org.au/statistical-education.htm

Surveys and Management
Co-Chair: Stephen Horn
Stephen.Horn@fahcsia.gov.au
Co-Chair: Ryan Defina
ryan.defina@abs.gov.au
http://www.statsoc.org.au/surveys-and-management.htm

Biostatistics
Co-Chair: Jake Olivier
j.olivier@unsw.edu.au
Co-Chair: Kylie-Ann Mallitt
kmallitt@kirby.unsw.edu.au
Assistant Chair: Nicholas Tierney
Nicholas.Tierney@gmail.com
http://www.statsoc.org.au/medical_statistics

Section for International Engagement
Mark Griffin
m.griffin@adasis-oz.com
http://www.statsoc.org.au/IntEngagementSection

Young Statisticians' Network

(Vacant)
http://www.statsoc.org.au/about-young-stats.htm

Further contact details for Society Secretaries and Section Chairs can be obtained by contacting the Society on (02) 6251 3647

SSAI CENTRAL COUNCIL

Executive Committee
President: John Henstridge
Secretary: Doug Shaw
secretary@statsoc.org.au

Branch Presidents and Branch Secretaries

Canberra
President: Ray Lindsay
Secretary: Warren Müller
secretary.canberrabranch@statsoc.org.au

New South Wales
President: Louise Ryan
Secretary: Ryan Defina
Ryan.Defina@abs.gov.au

Queensland
President: Helen Johnson
Secretary: Tania Patrao
Tania.Patrao@menzies.edu.au

South Australia
President: Lynne Giles
Secretary: Paul Sutcliffe
sutters@bigpond.net.au

Victoria
President: Lyle Gurrin
Secretary: Sandy Clarke
vicbranch@statsoc.org.au

Western Australia
President: Anna Munday
Secretary: Ryan Admiraal
R.Admiraal@murdoch.edu.au

EVENTS

CONFERENCE ON APPLIED STATISTICS AND PUBLIC POLICY ANALYSIS

11th -12th December 2014, Wagga-Wagga

INTERNATIONAL CONFERENCE ON ROBUST STATISTICS 2015

12-16 January 2015, Kolkata, India

YOUNG STATISTICIANS CONFERENCE 2015

5 -6 February 2015, Adelaide

VICBIOSTAT SUMMER SCHOOL 2015

9-13 February 2015, Melbourne

DATA MONITORING COMMITTEES & INTERIM ANALYSES IN CLINICAL TRIALS

26-27 February 2014, Sydney

SPATIAL STATISTICS CONFERENCE (THEME: EMERGING PATTERNS)

9-12 June 2015, Avignon, France

60TH ISI WORLD STATISTICAL CONGRESS

26-31 July 2015, Rio de Janeiro Brazil

JOINT STATISTICAL MEETINGS 2015

8-13 August 2015, Seattle Washington USA

THE 36TH ANNUAL CONFERENCE OF THE INTERNATIONAL SOCIETY FOR CLINICAL BIOSTATISTICS (ISCB 2015)

23-27 August 2015, Utrecht, The Netherlands

2015 RSS CONFERENCE

4-10 September 2015, Exeter UK

XXVIIITH INTERNATIONAL BIOMETRIC CONFERENCE (IBC 2016)

10-15 July 2016, Victoria, BC Canada

PRESIDENT'S COLUMN

Statistics and Education

Recently I have been involved in a number of activities where statistics relates to education.

As President of the Statistical Society of Australia I have recently been involved in Australia's bid for holding the international Congress on Mathematics Education (ICME) in Sydney in 2020. Australia has been shortlisted for this event that, if the bid is successful, will bring several thousand people interested in many different aspects of mathematics education, ranging from the psychology of how children learn mathematical concepts through to how mathematics should be taught in universities. Participants will range from researchers through to active teachers. The Society is supporting this bid, and hopes that a number of members will be involved in its statistical components.

The second activity was my attendance at the Western Australian Junior Mathematical Olympiad, a one day competition for students in Years 7, 8 and 9. My company was a sponsor so my role was to present some of the prizes and to represent mathematical employers. I also visited one school to present a prize at their Year 9 assembly. Seeing hundreds of young students taking a day out of their weekend to participate so enthusiastically was rewarding and made me optimistic about our future.

The third activity is one for which we must credit our very active chairs of the Statistical Education Section of the Society – Peter Howley and Michael Martin. They have put together a trial of a National Secondary Schools Statistical Poster Competition that will reach its peak with an awards night in early December. It is hoped that this will be developed into an Australian component of the International Statistical Literacy Project.

Activities such as these are critical to the future of statistics and mathematics in Australia. At a time when all the sciences are under pressure in Australia, particularly the mathematical sciences, the long term future depends upon how we, the statistical community, nurture young persons' interest in the subject. The Society has achieved a lot with the Young Statisticians activities – the forthcoming conference in Adelaide is an example – but we also need to think of the future Young Statisticians. Improving their exposure to and education in statistics is a challenge for us all.

As a final message for 2014 I thank all members of the Society for their contributions for the year. Special thanks go to our retiring Newsletter editors Alice Richardson and Michael Adena who have done this important job for many more years than might be reasonably expected of anyone! Their work has been critical in helping the Society be a community of enthusiastic individuals with a common passion.

John Henstridge
President
Statistical Society of Australia

John Henstridge

STATISTICAL JOB BOARD NOW AVAILABLE!

Looking for a change? Create a Job Seeker account, upload your resume or simply create a job alert and let the job find you.

Not looking for a job but offering one? Simply create an Employer account and see where it will lead you!

Best of luck finding top talent...or that dream job!

http://careers.statsoc.org.au/home/index.cfm?site_id=18859 (Job Board)

YSC 2015

YOUNG STATISTICIANS CONFERENCE
ADELAIDE FEB 5-6 2015

ADELAIDE
SOUTH AUSTRALIA

- By developing a more scientifically literate population. This could require a greater emphasis to be given to maths and science in schools particularly in primary grades, and

- By giving scientists a more powerful voice. They need to get more political: for example lobby government for an allocation of at least 3% to R & D (as in USA).

The impediments included:-

- The practice of manufacturing doubt as noted above,

- Not having a Minister for Science or a Science Policy,

- Lack of money for research. In Australia money often goes to the big lobby groups- e.g. the miners. In the US, a far greater proportion goes into research,

- The conflict between short term "emergencies" and long term strategies, Adam cited the case of the Brumby government hastily installing air conditioning in schools; instead of solar panels (which had previously been planned) based on the evidence of a single hot day, and

- The possibly apparent need of many politicians "not to know". Politicians don't like changing their mind (breaking promises or making statements that are later found to be wrong).

When asked how we can overcome these impediments, Adam expressed some optimism. First, he felt that there was considerable sympathy and enthusiasm for science and innovation in the general community. Second, he felt that we are running out of alternatives for when the mining boom ends. We have an intelligent educated population; we have good manufacturing facilities and an excellent record in exporting R & D. In fact our medical $ exports have recently been exceeding our automotive $ exports. Greater investment in education and R & D is clearly the best way forward.

Adam was less at ease when questioned on the dichotomy between the use of evidence in policy implementation as distinct from its use in policy development. He stated that both were important. He deeply regretted that in the implementation phase of the climate change policy, the debates switched to credibility of the PM, tax and compensation; and the original aims of the policy (e.g. reduction of global warming) were almost completely forgotten.

There was further discussion on the need to document the scientific evidence behind major decisions. Adam was concerned at the high frequency of ministers citing "commercial in-confidence" as an excuse for withholding this information.

In answer to the question "What to do if parliament's version of the facts is incorrect?", Adam replied that it was very difficult for individuals outside of parliament to respond. However the Senate did have one mechanism which has proved very effective on a number of occasions. This was its power to call in and interrogate senior departmental officers.

Adam concluded the session by reiterating the need for scientists to come together and organize a positive campaign for a greater say in government and for increased spending on R & D.

The session was well received by an appreciative audience of 172 people. Our sincere thanks go to Adam and Gael, and to Damjan Vukcevic for organizing the occasion.

George Rennie

TERRY SPEED AMSI-SSAI LECTURE TOUR

Professor Terry Speed has just completed an Australia-wide public lecture tour to showcase and celebrate the work that earned him the premiere prize for science in Australia, the Prime Minister's Prize for Science for 2013. There were thirteen lectures and seminars, each jointly authored with several of Terry's collaborators (see page 18).

This newsletter has several reports on Terry's tour.

The 'new frontier' public lecture, which was given in eight locations, will be reported in the March issue.

Terry's ability to inspire and inform is evident in these three extracts from Graduate Diploma student reports (for a multivariate statistics unit at The University of Canberra) on 'Removing unwanted variation from high-throughput omic data'. Chris Woodbridge summarised the scope of the talk as:

'The talk covered methods for working with large quantities of omic data and removing variation caused by such things as batch effects. The overall aim is to identify and reduce errors that can arise from batch effects from time, space, location etc. This allows the data to be combined without compounding variance. One example used was a study on color blindness in mice. The initial analysis showed variation between plates used when taking measurements. The removal of this variation allowed ... variation between [the] sexes to be shown.'

Terry Speed

whereas Megan Verbakel wrote:

'Noting the potential for batch effects and other unwanted variation in large, generally multivariate, datasets, Terry proposed a framework for removing unwanted variation using relatively simple statistical methods. The framework starts with a linear model, adding unobserved terms to deal with the unwanted variation. Factor analysis is then used to identify positive and negative controls, with the negative controls identifying unwanted factors and the positive controls used to estimate the number of unwanted factors. Utilising regression, Terry then adjusts the model for the unwanted factors. With unwanted variation potentially causing false and missed discoveries, incorrect predictions and artificial clusters, the framework allows multivariate data to be reduced to the variation the researcher is interested in.'

Huanhua Zhang concluded:

'Terry delivered his message in a clear and straightforward style. He always gave the audiences examples to help them catch his thoughts ... In the following Q&A section, Terry shared his opinions about a number of classic questions to a statistical method, such as estimation robustness, replication and resistance to factor analysis ... [I]t is highly impressive how Terry and his colleagues use simple methods to solve a hard problem.'

Abstracts for the two other talks are given on page 18.

Terry has also been honoured with the CSIRO Eureka Prize for Leadership in Science at the 2014 Eureka Science Prizes at the Australian Museum (http://australianmuseum.net.au/media/2014-Eureka-Leadership) and the 2014 Jerome Sacks Award for Cross-Disciplinary Research (page 21).

OUR INSURANCE GIVES YOUR BUSINESS A CRITICAL BIAS

As the appointed Insurance Brokers to the SSAI we are proud to offer their members a truly unique Professional Indemnity Insurance Policy: one that will cover your particular requirements perfectly and save you money.

As the preferred Professional Indemnity Insurance Broker for SSAI and it's Members we have brokered an exclusive arrangement with Chubb Insurance Australia Ltd, who are a market leader in specialty insurance coverages, which are uniquely designed for specific industries.

Benefits to SSAI Members:

1. Competitive premiums -- in a majority of cases, we guarantee to offer a cheaper premium than you paid last year, subject to minimum premium levels.
2. One of the broadest policy wordings available, including:
   a. Full cover for Contractors, not just vicarious
   b. Loss includes punitive and exemplary damages
   c. Built in "contingent" bodily injury/property damage coverage
   d. Worldwide coverage includes "duty to defend" anywhere in the World
   e. "Formal investigations" cover available to the full Policy Limit
   f. 100% of Defence Costs paid within the Limit even if only part of actual Loss if covered
   g. Auto Reinstatement of Limit.

CALL: 0439 562 009

When it comes to finding an insurance broker, do you look for the widest choice of cover, the best value for money or an adviser that really understands your specific needs? Usually finding more than one of these qualities is rare. But IAA is in the unique position of offering you all three.

Our Commitment to You

As your appointed General Insurance Broker, we will:
- Conduct an analysis of your risk/needs and recommend the most appropriate covers to protect your assets and liabilities
- Provide alternatives on your risk retention through various excess options
- Allow you to pay premiums in monthly instalments
- Provide prompt and timely servicing of your accounts to ensure you are adequately informed of issues that may affect your business
- Provide competitive quotations for any other classes of insurance you may have.

In the event of a claim, we will:
- Provide you with the insurers claim form if needed
- Arrange for the loss assessor to be appointed if applicable
- Advise on the completion of the claim form
- Provide advice to ensure you have received your full entitlements.

As one of Australia's biggest General Insurance Brokers, our knowledge of the insurance industry is second to none. Our experience in handling insurance for many thousands of customers enables us to provide the best and most current advice. In other words, at IAA, we're committed to making insurance work for you.

Mike Aurisch
ANZIIF (Snr. Assoc) Dip Fin Serv (Brok) QPIB
Authorised Representative of Insurance Advisernet Australia Pty Limited
ABN 15 003 886 687 | AFSL No 240549 | AR No 403930
Level 31, 100 Miller Street, NORTH SYDNEY NSW 2059
M 0439 562 009 | F (02) 9954 1809 | E maurisch@iaa.net.au

UNDERWRITTEN BY:

SSAI Statistical Society of Australia Inc.

INSURANCE ADVISERNET

REPORT FROM RECIPIENT OF THE
2014 GOLDEN JUBILEE TRAVEL GRANT

I was thrilled to be given the opportunity to present my PhD research at two international conferences this year. In August, I attended the International Society for Clinical Biostatistics conference in Vienna. The following week, the Royal Statistical Society (RSS) International conference was held in Sheffield. The SSAI, through the Golden Jubilee Travel Grant, supported my attendance at the RSS conference.

The program for the RSS conference reflected the important role statistics has in society. Two themes in particular caught my attention: Statistics in sport and Statistics and the law. The relevance of the statistics and the law theme became clear shortly after I arrived in Sheffield - I saw a young offender fleeing the scene of a crime, with the drops of blood from a gash on his face giving away his escape route. I immediately thought about the talks to be given the next day that focused on the communication of probability in the courtroom, such as when considering the likelihood that DNA collected from a crime scene came from the alleged offender. I also saw the Southampton football team in a hotel lobby in London; the players might have been interested in learning more about predictive modelling in football at the statistics in sport session.

The conference was held at Sheffield Town Hall, a grand building in the centre of the city. The organisers had to be creative in order to ensure there were enough rooms for concurrent sessions. I found myself sitting in one of the backstage dressing rooms for one session – the very dressing room Peter Andre would be using a month later to prepare for his performance!

Unfortunately I couldn't attend every session, so I mostly stayed loyal to my research field which is biostatistics. The standard of the talks and research was very high. I had the opportunity to hear more about the current work being done at an international level in my specific area of research, multiple imputation, which was well represented in the sessions.

The plenary speakers presented on a range of interesting topics. Lord Richard Allen of Hallam, director of policy in Europe for Facebook, spoke on the challenges of operating to scale. Richard started by wowing the audience with the number of Facebook users; 829 million users access Facebook daily, with 654 million of these using a mobile device. With these numbers, the main challenges for Facebook are how to use data to provide a good service to the users and advertisers, as well as how to identify bad behaviour or illegal activities being conducted through Facebook. Tim Harford was the second plenary and presented a cautionary tale about the big data trap, with the key message being that we should not forget the important statistical lessons, such as the need to keep our population of interest in mind and to be aware of the potential for bias when using data sourced from administrative systems. The third plenary speaker was our very own Terry Speed. In his engaging and entertaining style, Terry presented a statistical method that he has developed with colleagues in order to remove unwanted variation in high dimensional data.

I presented a poster on the *Comparison of methods for the multiple imputation of limited-range variables*. A wine reception dedicated to the posters was held, which allowed everyone time to browse the posters and chat to the presenters. My poster generated some interesting discussions with other conference delegates.

Laura Rodwell

The young statisticians section of the RSS was active throughout the conference and organised an icebreaker session at lunch on the first day as well as a pub trivia night. Sadly, with trivia rounds that included questions about English counties and American states, my only contribution for the night was our team name: *The Full Monte Carlo*.

I would like to thank the SSAI for providing me with the support to attend the conference through the Golden Jubilee Travel Award, as well as the Murdoch Childrens Research Institute (MCRI) and the Victorian Centre for Biostatistics (VicBiostat) for providing funds to support the rest of my trip. I encourage all eligible students to submit an application for the 2015 travel grant as presenting at an international level is a valuable experience.

Laura Rodwell

PhD Candidate
Clinical Epidemiology and Biostatistics Unit, MCRI and
The University of Melbourne

If you would like to travel as well, please apply for the

SSAI GOLDEN JUBILEE TRAVEL GRANT

It provides overseas travel funds to SSAI student members, who can prove consecutive SSAI membership for a minimum of two years and who wish to attend overseas conferences at which they present a paper or poster.

A complete application will consist of

- Information on the conference and its importance to student's work (2-3 lines)

- Details of the paper/s/poster student wants to present at the conference

- A list of other funds sought or promised, including student's home institution

- Student's out of pocket expenses expected

- Any other supporting material student feels is necessary

- A letter of support SIGNED by one of student's supervisors AND student's Departmental Head

- Student's CV

The application deadline is 31 March 2015.

A maximum of $1000 is available per application, limited to a single trip during the course of the student's studies. Students will not be supported in their first year of study and will have had to be members of the Society for at least 2 years prior to the application deadline. Applications are required to be lodged in advance of travelling. In exceptional circumstances an application can be for post-conference support, but the application will then have to be made within 1 month of returning and the 2 year mandatory membership period prior to departure must still be met. Exceptional circumstances are limited to unforeseeable student out of pocket expenses arising from other funding sources not fulfilling their obligation or changes to the trip that could not have been avoided.

If successful the student member is required to produce original receipts for amounts of equal or greater value than the grant. These receipts will be returned to the student marked with how much has been reimbursed. The student will therefore still be able to use the receipts for proof of attendance or to claim any funding shortfall from other organisations. The student member will also need to supply a report of his or her involvement in the conference to be published in the SSAI newsletter. This report should confirm the actual travel details and papers presented.

Recipients of the grant are asked to acknowledge the SSAI's support in the presentations and in any published version of the paper.

One travel grant is available per year. Assuming that more than one application will be received per year, either the Executive Committee or a special committee would help with the selection process.

For more information or to apply, contact the SSAI Office (eo@statsoc.org).

With this travel grant program the SSAI seeks to underline its objective to further the study, application and good practice of statistical theory and methods in all branches of learning and enterprise. It has been implemented to confirm to members that the SSAI is willing to support student statisticians and their budding careers.

Careers@Singapore: Data Analytics

FACT: Big data and analytics is an important growth area for Singapore.

By the end of 2017, Big Data and Analytics aim to contribute S$1 billion in value-add to the economy.

Being at the heart of Asia, Singapore's location is seen as providing the perfect setting for companies looking to understand Asian consumers. With its culturally diverse population, Singapore offers the ideal location to test new made-for-Asia innovations, and a broad economy provides access to industry-specific knowledge crucial for coming up with the best analytics.

Analytics Who's Who – Multinationals with Analytics Hubs in Singapore

- **Accenture Analytics Innovation Center (AAIC)** will focus on developing innovative approaches for the use of analytics in services related to education, health, procurement, public safety, resource utilisation, smart grids, transportation and workforce effectiveness.

- **EDF-Veolia Centre of Excellence for Sustainable Cities** is in collaboration with Singaporean authorities in the following fields: energy-efficiency and air conditioning systems in buildings, domestic waste collection, the integration of solar panels into buildings, the development of green roofs, as well as local water recycling.

- **NEC Laboratories Singapore (NLS)** aims to produce solutions for urban issues by concentrating on areas that include safety, big data, security and smart energy.

- **NielsenLAB** is the first in Asia outside the US, which develops and leverages cutting-edge technologies and methodologies in neuroscience, shopper technology and measurement science to fuel consumer and market insights for companies working to reach consumers in Asia.

- **SAP Next Business and Technology Research Centre** is the largest research location and headquarters in Asia Pacific focusing on research into social business networks and collaboration, real-time intelligence and analytics.

Here's a quick look at Singapore's Data Analytics Focus Areas.

Data Analytics Landscape in Singapore

Focus areas

Urban Services

- Real-time energy pricing (grid operators)
- Real-time route optimisation (telematics players)
- Smart buildings (energy svcs consultancies, building equipment manufacturers)

Consumer / Retail

- Marketing analytics
- Sentiment analysis
- Product development and strategy (FMCG, E-Commerce, Retailers, Market Research, Telcos)

HOT JOBS

Contact Singapore invites you to embark on a rewarding career in Singapore's Data Analytics industry. Click on the job titles below to know more about each position.

Alternatively, visit our job portal www.contactsingapore.sg/jobs for other career opportunities in Data Analytics.

Institute for Infocomm Research
▶ Data Scientist

Institute of High Performance Computing
▶ Research Scientist (Bioinformatics)

Singapore Institute of Manufacturing Technology
▶ Research Scientist (Manufacturing Data Analytics) for Planning and Operations Management

accenture
▶ Advance Analytics Predictive Solution Architect
▶ Big Data Consultant/Analyst
▶ Big Data Solution Architect
▶ Data Visualization Consultant/Analyst
▶ Consultant/Manager

Bloomberg
▶ Market Data Analyst

citi
▶ Big Data and Analytics Java Developer

Security

- Modeling of cyber attack profiles (IT security coys)
- Analysis of sensor data to determine threats in public infrastructure (security infrastructure, IT coys)

Healthcare

- Pharmacovigilance (pharma)
- Chronic disease mgmt (pharma, healthcare svcs, insurers)
- Healthcare economics (pharma)
- Clinical research (CROs)

Financial

- Risk modeling & simulation (finance & accounting firms)
- Fraud detection (payment coys)

Supply Chain / Logistics

- Fleet optimisation (airlines)
- Network design & optimisation (3PLs, MNCs)
- Multi-modal city logistics (3PLs)

Manufacturing

- Predictive maintenance (MROs)
- Power-by-the hour (component manufacturers)

Career Opportunities within Singapore's Data Analytics Sector

Over the next five years, Singapore aims to create a pool of 2,500 analytics professionals to support the growth momentum of this exciting sector, leading to an increase in career opportunities in the field.

Check out the **HOT JOBS** on the right and be a part of this exciting growth sector in Singapore.

ABOUT CONTACT SINGAPORE

Contact Singapore is an alliance of the Singapore Economic Development Board and Ministry of Manpower. We engage overseas Singaporeans and global talent to work, invest and live in Singapore.

Contact Singapore actively links Singapore-based employers with professionals to support the growth of our key industries. We work with investors to realise their business and investment interests in Singapore.

For information on working, investing and living in Singapore, please visit www.contactsingapore.sg or contact our worldwide offices.

ABOUT SINGAPORE

In the heart of Asia lies Singapore - a bustling, world-class city state that has made waves around the world for its business excellence, connectivity to the world and its innovation.

Singapore has been consistently acknowledged as a global business hub - one that features developed infrastructure, political stability, open business policies, a skilled workforce, the use of English as the main working language and respect for intellectual property rights. Many multinational companies have chosen to set up their bases in Singapore, next to 154,000 small and medium enterprises. Business owners all over the world regard Singapore as an ideal location to grow their businesses, with many of them using the country as a springboard to tap into other emerging markets in Asia.

Besides being ranked in the World Bank 2013 Report as the world's easiest place to do business (out of 185 countries), Singapore has, among its many accolades, the top ranking in the Global Competitiveness Report 2013-2014 by the World Economic Forum as being the most competitive economy in Asia. Singapore was also ranked by the Mercer 2012 Quality of Living Survey as the top Asian country for its quality of life.

Global talent are attracted to Singapore's meritocratic society that embraces diversity, and a high quality of life in a cosmopolitan environment.

EY

- Data Analytics, Forensic Technology and Discovery Services Associate
- Portfolio Security Consultant

facebook

- Data Insights Analyst

hp

- Data Scientist

iDA

- Data Scientist

Land Transport Authority

- Data Scientist

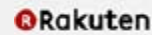### ncs

- Data Scientist

NEC

- Data Scientist

nielsen

- Data Scientist/Statistician

pwc

- Director, Technology (Security)
- Technology Consulting (Cybersecurity) - Manager / Senior Manager

P&G

- R&D Data Scientist/Analyst - Product Vigilance

Rakuten

- Data Scientist at Viki Inc.

SAP

- R&D Engineer/Data Scientist

SingTel

- Senior Data Architect / Data Architect

VISA

- Analytics Delivery Manager

To access this flyer online, please visit: www.contactsingapore.sg/eCareers/Data_Analytics/Oct2014

MATHEMATICIANS IN SCHOOLS

Earlier this year I discovered a fantastic program called Mathematicians in Schools (MiS, http://www.mathematiciansinschools.edu.au/), a project started by the Australian Government Department of Education and managed by the CSIRO. The aim is to pair mathematicians (which has a loose definition, basically means any profession using maths in their work) with schools to provide inspiration and a different perspective to students.

After registering my interest, I was partnered with Middle Harbour Public School in Mosman, New South Wales. Each partnership, according to the MiS website, is different as it is up to the mathematician and the partner school to decide how best to work together. In my case, the School had identified three very bright and motivated students (one Year 6, two year 5) who were already learning at a pace that was well beyond their peers. After meeting with the teacher and the three students in person, we decided that we would communicate through regular e-mails as well as having ad-hoc Skype calls if needed.

Through this partnership, I have so far found myself explaining Pythagoras's Theorem, real numbers vs imaginary numbers, as well as the basic concepts of calculus. It has been a both challenging and rewarding experience. Challenging because it was not an easy task to explain these concepts in an intuitive way via e-mail; rewarding when the students showed that they not only understood but also enjoyed my explanations! For a person who did not attend school in Australia, this also gives me an opportunity to have a better understanding of the Australian syllabus – I'm sure this will come in handy one day!

I was also reminded, through interacting with the students, of some of the discussions on the ANZSTAT mailing list recently about how people from different disciplines (e.g. data miners) use statistics, as well as the perception of the statistics discipline. Some people also expressed concerns that students prefer to study, say, "big data" under the computer science umbrella rather than doing a degree in statistics, because computer science sounds fancier. Such students may then become responsible for analysing important "big data" while not having a grasp of (basic) statistical theories.

It is our responsibility, as professional statisticians, to change people's view on statistics, and there is no better way to do this through educating primary/secondary students that statistics can be fun and cool. It is also our obligation to mentor young people, to show them how they can apply what they learn in schools in the real world, by using real world examples that we have encountered in our professional careers. If, like myself, you want to help shape the future of statistics, I strongly recommend that you participate in this MiS program!

Finally, I would also like to take this opportunity to thank Middle Harbour Public School for giving me the privilege of mentoring the students, as I have learnt at least as much (if not more) as the students. A wise mentor of mine once told me that in a mentoring relationship, the mentor often learns and gains more than the mentee, and now I can see why!

Kevin Wang

Share your love of mathematics by volunteering with Mathematicians in Schools

Feel like making a difference? No fixed or minimum hours, the ability to contribute in a way that suits your schedule, and a way to positively affect the Australian mathematics community of the future! CSIRO Education's Mathematicians in Schools is a skilled volunteering program offering you the opportunity to make a difference to the mathematics education of primary and secondary school students.

Scientists and Mathematicians in Schools is a successful, well-established national program, over 4000 professional partnerships between scientists, mathematicians and teachers have been created since July 2007. Currently more than 1600 scientists and mathematicians across Australia are involved in partnerships.

Interested Mathematicians can register here.

How it works: Individual engineers or mathematicians are partnered with individual teachers in ongoing professional partnerships. There is no cost to participate.

Activities: Each partnership is flexible, unique and voluntary - the mathematician and teacher decide how they will work together taking account of workloads, the mathematician's expertise, and the teacher and class needs. This allows partners to develop their own style and may include hands-on activities, presentations, demonstrations, mentoring, emailing and video conferencing.

Time commitment: No fixed or minimum hours – it's up to the mathematician and teacher to negotiate the frequency of interactions. Mathematicians may visit the school once or twice a year, a couple of times a term, or once a week or month. Other partnerships utilise ICT (email and video conferencing) almost exclusively and have little face to face interaction.

Skill/experience required: The definition of a mathematician for this program includes any professional who is actively engaged in the fields of mathematics and/or technology. It includes economists, accountants, research mathematicians, engineers, cryptographers, surveyors, biometricians and statisticians, amongst others. Generally a Bachelor's degree in a mathematics related field and currently working in a profession where maths is a major component of your work is required.

Available teachers: A map of teachers interested in establishing partnerships is available on our website at www.scientistsinschools.edu.au/scientists/unmatched.htm. You can nominate one of these teachers to be partnered with, but are also welcome to nominate a region/school/teacher of your choosing and we will aim to facilitate a partnership for you.

More information can be found on our website www.mathematiciansinschools.edu.au. You can also register for Mathematicians in Schools from the website under the "Get Involved" Tab. If you have any questions, contact Gill Lunniss, Scientists and Mathematicians in Schools Project Officer in Victoria, on sis.vic@csiro.au or 03 9252 6502.

Gill Lunniss
Project Officer Victoria, Scientists and Mathematicians in Schools
Education
CSIRO
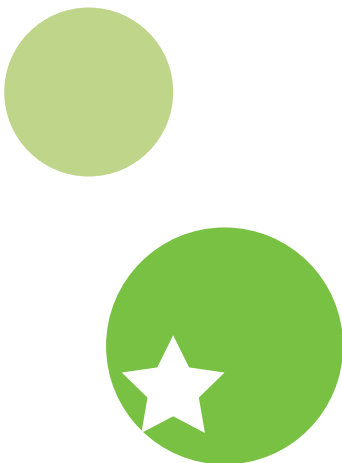
WHAT'S HAPPENING IN STATISTICAL EDUCATION?

Update on a new activity: Secondary Schools Statistical Literacy Poster Competition

In the previous SSAI Newsletter we wrote about the beginning stages of the pilot of a national statistical literacy project competition in the Hunter Region in 2014, being supported by the SSAI and run under the auspices of the International Statistical Literacy Project (ISLP, see http://iase-web.org/islp/).

In brief, the poster competition involves secondary school students undertaking an activity-based learning project, in teams of 2 to 5, and creating a poster presentation (as per conference poster) based on the collection, analysis, interpretation and reporting of data (using principles of statistics and scientific method) towards addressing a practical question/problem of interest to them. The project aims are listed below**.

The competition involves mentors attending schools to help facilitate projects with the school teachers. As part of this Peter has had several discussions with school teachers (phone and email) and attended schools personally, including the one where he school captained; his 4 Unit Mathematics teacher was still at the school. It has been a great experience interacting with teachers and students at all the schools.

We have been buoyed by the level of interest this competition has aroused. In brief:

- 10 schools registered, and are undertaking approx. 50 projects - some have finished, most began late October.

- Years 7 to 11 are represented; primarily Years 9 and 10.

- Project types: fields covered include sciences, mathematics, languages, environmental science, industrial arts, iSTEM classes and a debating team.

- Poster submissions are due 28 November.

- Poster Display and Awards evening Friday 5 December 4:30-7pm at The University of Newcastle.

- Many other schools indicated they were unable to participate this year but keen to do so next year.

- There has been great interest, some schools identified they would like to build the competition into their curriculum

- CSIRO's Scientists-in-Schools and Mathematicians-in-Schools program coordinators have shown support for the competition, as have other similar entities such as The Science and Engineering Challenge.

There will be certificates and prizes (including items such as mobile phone microscopes) for winning teams locally. Following the poster display and awards evening, 5 December at The University of Newcastle, winning posters in each division may compete on the international stage in the International Statistical Literacy Project (judged in March 2015).

So, would you like to know more, possibly be involved as either a mentor or coordinator around Australia as we look to make this competition national in 2015? Perhaps you are near the Hunter Region or know of others who would

be keen to be a mentor, or are teachers or have secondary school students – whatever the connection, please spread the word… there is a national secondary schools poster competition and people from all fields of interest are invited. Please contact peter.howley@newcastle.edu.au or phone 02 49215518 to discuss.

If you have any queries or suggestions, or would like a flyer to send to others outlining the project please contact peter.howley@newcastle.edu.au.

This activity brings additional international advertising for the SSAI, now listed under http://iase-web.org/islp/People.php?p=Country_Coordinators and the ISLP's facebook page. We are in the process of developing a website thanks to CARMA (Computer-Assisted Research Mathematics and its Applications), in the meantime you can obtain some further information from http://numeric.carma.newcastle.edu.au/postercomp/secondary_schools_poster_comp-flyer.pdf.

I'm sure we will continue to learn a lot from the pilot and will keep you informed of progress.

Project Aims:

- Inspire school students' aspirations towards Higher Education and careers in Statistics, Mathematics and Sciences

- Build relationships between University – Schools – Teachers – Students – Professional societies (build community) and increase accessibility of Higher Education

- Inspire school teachers and students with practical applications of curriculum content

- Enable Undergraduate and Postgraduate students in Statistics, Mathematics, Sciences, Psychology to experience project facilitation and mentoring and consider career in teaching mathematics/science

- Provide practical opportunities for pre-service teachers to mentor students and facilitate project management

- Increase awareness of the wide application and need for Statistics

- Inspire and enrich through boundary encounters

Peter Howley and **Michael Martin**
Co-chairs of the Statistical Education Section

TERRY SPEED LECTURE TOUR

The Prime Minister's Prize for Science recipient of 2013 was Society member Professor Terry Speed of the Walter & Eliza Hall Institute of Medical research. Terry embarked on a lecture tour in the middle of 2014.

19 Aug	Public Lecture - A New Frontier: understanding epigenetics through mathematics	Swinburne University of Technology
20 Aug	Access Grid Room seminar - Comparing and combining mutation callers	Australian Mathematical Sciences Institute
26 Aug	Seminar - Removing Unwanted Variation from high-throughput omic data	Australian Bureau of Statistics
27 Aug	Public Lecture - A New Frontier: understanding epigenetics through mathematics	Australian National University
22 Sept	Seminar - Removing Unwanted Variation from high-throughput omic data	University of Western Sydney
23 Sept	Public Lecture - A New Frontier: understanding epigenetics through mathematics	UNSW Australia
24 Sept	Seminar - Comparing and combining mutation callers	University of Sydney
25 Sept	Public Lecture - A New Frontier: understanding epigenetics through mathematics	Charles Sturt University, Wagga Wagga
8 Oct	Access Grid Room seminar - Normalization of RNA-Seq Data: Are the ERCC Spike-In Controls Reliable?	University of Western Australia
9 Oct	Public Lecture - A New Frontier: understanding epigenetics through mathematics	University of Western Australia
16 Oct	Public Lecture - A New Frontier: understanding epigenetics through mathematics	University of Tasmania
16 Oct	Seminar - Removing Unwanted Variation from high-throughput omic data	University of Tasmania
30 Oct	Public Lecture - A New Frontier: understanding epigenetics through mathematics	Queensland University of Technology

The 'new frontier' public lecture will be reported in the next newsletter and the 'removing unwanted variation' seminar on page 8. Abstracts for the other seminars are:

Statistics lecture 2. Normalization of RNA-Seq Data: Are the ERCC Spike-In Controls Reliable?

(joint with Sandrine Dudoit, Davide Risso and John Ngai)

The External RNA Control Consortium (ERCC) developed a set of 92 synthetic polyadenylated RNA standards that mimic natural eukaryotic mRNA. The standards are designed to have a wide range of lengths (250-2,000 nucleotides) and GC-contents (5-51%). The ERCC standards can be spiked into RNA at various concentrations prior to the library preparation step and serve as negative and positive controls in RNA-Seq. Ambion commercializes spike-in control mixes, ERCC ExFold RNA Spike-in Control Mix 1 and 2, each containing the same set of 92 standards, but at different concentrations. We investigate the use of the ERCC spike-in controls for two main purposes: (a) Quality assessment/quality control (QA/QC) of RNA-Seq data and benchmarking of normalization and differential expression (DE) methods, and (b) Direct inclusion

in between-sample normalization procedures. We have two RNA-seq data sets which make use of the ERCC controls: a local one concerning treated and untreated zebrafish tissue, and some of the SEQC samples. A variety of normalization methods will be compared, both using and not using the ERCC controls. One of the methods we discuss is a variant on our recently published RUV-2 method, which uses SVD on negative controls.
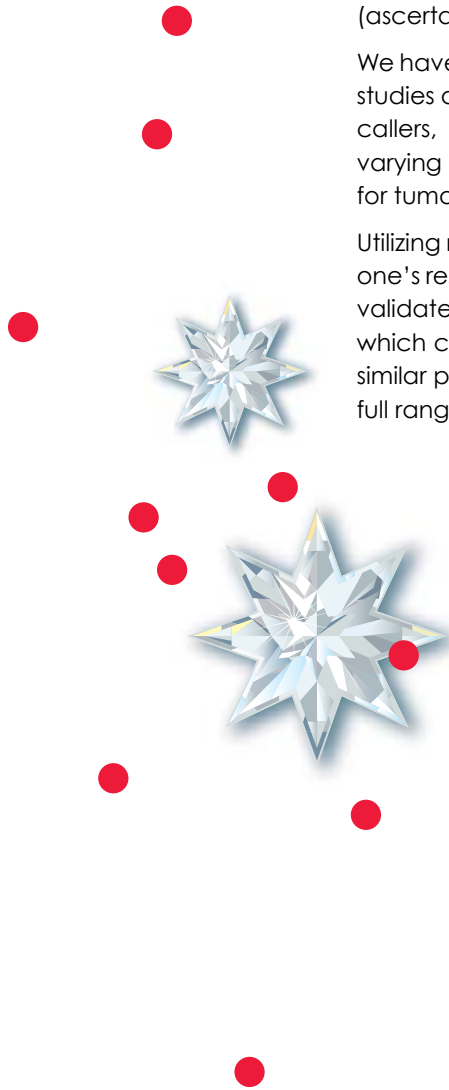
Bioinformatics Lecture. Comparing and combining mutation callers

(joint with Su Yeon Kim and Laurent Jacob)

Somatic mutation-calling based on DNA from matched tumor-normal patient samples is one of the key tasks carried by many cancer genome projects. One such large-scale project is The Cancer Genome Atlas (TCGA), which is now routinely compiling catalogs of somatic mutations from hundreds of paired tumor-normal DNA exome-sequence datasets. Several mutation-callers are publicly available and more are likely to appear. Nonetheless, mutation-calling is still challenging and there is unlikely to be one established caller that systematically outperforms all others. Evaluation of the mutation callers or understanding the sources of discrepancies is not straightforward, since for most tumor studies, validation data based on independent whole exome DNA sequencing is not available, only partial validation data for a selected (ascertained) subset of sites.

We have analyzed several sets of mutation calling data from TCGA benchmark studies and their partial validation data. To assess the performances of multiple callers, we introduce approaches utilizing the external sequence data to varying degrees, ranging from having independent DNA-seq pairs, RNA-seq for tumor samples only, the original exome-seq pairs only, or none of those.

Utilizing multiple callers can be a powerful way to construct a list of final calls for one's research. Using a set of mutations from multiple callers that are impartially validated, we present a statistical approach for building a combined caller, which can be applied to combine calls in a wider dataset generated using a similar protocol. The approach allows us to build a combined caller across the full range of stringency levels, which outperforms all of the individual callers.

2014 CSIRO EUREKA PRIZE FOR LEADERSHIP IN SCIENCE

Winner: Professor Terence Speed, Walter and Eliza Hall Institute of Medical Research - See more at: http://australianmuseum.net.au/media/2014-Eureka-Leadership#sthash.za2nXgUR.dpuf

(Not) curing cancer only part of the story

Professor Terry Speed accepts he's never going to see the headline 'statistician cures cancer'.

However, it's a sure bet that every significant triumph we see in the long fight against the Big C has been won on the back of some serious, high-quality number crunching. And there's a good chance Melburnian Terry Speed helped.

For his superb leadership of the bioinformatics team at the Walter and Eliza Hall Institute of Medical Research, and his other contributions to the science of bioinformatics, Terry has been awarded the CSIRO Eureka Prize for Leadership in Science.

At the Institute, the team that Terry leads uses computational mathematics to help researchers interpret massive amounts of experimental data.

Terry's extraordinary leadership in the field extends well beyond the walls of the Institute. His techniques for improved DNA data analysis are distributed free-of-charge and used by thousands of researchers around the world. He is one of the world's most cited scientists—not only in mathematics, but in computer science, biology and biochemistry.

In the fight against cancer Terry works alongside genetic and molecular biologists, other mathematicians, computer engineers and technicians, and in a 45-year career has also assisted farmers, psychiatrists and diamond miners.

And criminologists too... famously, he was an expert witness at the Ronald Ryan case and the OJ Simpson trial.

"Terry was one of the creators of the field of bioinformatics, and is one of the world's leading statisticians," Australian Museum Director and CEO Kim McKay said. "He also devotes tremendous effort to the people he works with. He has supervised over 60 postgraduate students, and mentored dozens of other researchers, many of whom have gone on to important positions within the academic and scientific community."

This prize, recognising leadership, cements Terry's position as one of Australia's most decorated scientists. He was elected as a Fellow of the Australian Academy of Science in 2001, received an Australia Fellowship in 2009, won the Prime Minister's Prize for Science in 2013, and was elected to Fellowship of the Royal Society that same year.

The Australian Museum Eureka Prizes are the country's most comprehensive national science awards. The Eureka Prizes have been rewarding science since 1990—celebrating 25 years in 2014.

The other finalists were:

- Professor Hugh Durrant-Whyte, CEO of NICTA and a world leader in the field of robotics.
- Professor Michelle Simmons who leads the Quantum Computation and Communications Technology team at University of New South Wales, and led the team that created the world's first single-atom transistor.

Terry Speed

CLICK HERE TO SEE MORE ON THE AUSTRALIAN MUSEUM WEBSITE

YOUNG INVESTIGATOR AWARD FOR SSAI MEMBER DAVID WARTON

David Warton

Congratulations to Australasian IBS member David Warton on being awarded the 2014 Young Investigator Award from the American Statistical Association Section on Statistics and the Environment. Aiming to encourage younger members (under 41) of the environmental statistics community, the award is presented in recognition of outstanding contributions to the development of methods, issues, concepts, applications, and initiatives in environmental statistics.

David is delighted to receive the award, and is "especially pleased to see some international attention for Australian ecological statistics research, as there are some exciting ideas coming out of this corner of the world at the moment". David is grateful to his talented team of students, post-docs and collaborators who "continue to be a pleasure to work with". David Warton is an Associate Professor in the School of Mathematics and Statistics at the University of New South Wales.

David is keen to organise another Eco-Stats Symposium at the University of New South Wales, following on from last year's success. The symposium will be designed as a collaborative forum for researchers with interests in ecology, statistics, or both. If you're interested in organising a session, please contact David at david.warton@unsw.edu.au.

See http://www.eco-stats.unsw.edu.au/symposium.html for more information.

Published with the kind permission of the Australasian Biometrics Society.

NIASRA | **UOW**

**The National Institute for Applied Statistics Research Australia (NIASRA)
at the University of Wollongong will be hosting StatsWeek 2015
from 9 to 13 February on the Wollongong Campus.**

StatsWeek consists of the following three courses:

Sample Design and Analysis using R

A short course with Thomas Lumley (University of Auckland)

Monday and Tuesday, 9 - 10 February

$900, $750 for SSAI members, $400 for students

*This 2-day course covers sample design, weighting and methods for a wide range of
exploratory analysis and modelling of data from complex surveys using the R survey package.*

Spatial Sampling for Agricultural and Environmental Surveys

A short course with Roberto Benedetti (University of Chieti-Pescara, Italy)
and Frederica Piersimoni (Italian National Statistical Institute)

Wednesday, 10 February

$400, $300 for SSAI members, $200 for students

*This 1-day course considers a systematic approach to key sampling algorithms for the selection
of spatial units, with a particular emphasis on agricultural and environmental data collection.*

Modelling Group Heterogeneity via M-quantile Regression

A short course with Nikos Tzavidis (University of Southampton), James Dawber (NIASRA)
and Ray Chambers (NIASRA)

Thursday and Friday, 12 - 13 February

$700, $600 for SSAI members, $300 for students

*This 2-day course provides the fundamental statistical background of necessary for the use of quantile and
M-quantile regression as a semiparametric, robust alternative to standard multilevel modelling.*

You can register for one, two or all three courses.
Please visit the NIASRA webpage and click the StatsWeek link for more information and to register online.
Places are limited and registrations will be processed in the order they arrive.

The City of Wollongong is located just 80km south of Sydney on a spectacular stretch
of the southern New South Wales coast. Wollongong's natural environment offers pleasant
summer weather and includes pristine beaches and a great network of walking/biking trails
and dazzling views of the coastline along the escarpment to the west.

StatsWeek is sponsored by NIASRA and the University of Wollongong.

STATISTICS SEMINAR AT BANGLADESH ACADEMY OF SCIENCES

The Bangladesh Academy of Sciences (BAS) organised a seminar on "The Dominant Role of Statistics in Contemporary Science and Technology" at its auditorium in the Bangladesh Science Museum Buildings in Dhaka on 1 October 2014. The speaker at the seminar was Professor Shahjahan Khan, University of Southern Queensland, Australia and Founding Chief Editor of Journal of Applied Probability and Statistics (JAPS), who was elected as an Expatriate Fellow of BAS last year.

The seminar was attended by a record audience, and the presentation covered various contemporary problems in the area of health, medicine, engineering, climate, environment, big data etc. and highlighted how statistics is playing an increasingly dominant role in the state-of-the-art research in solving complex issues of our time. Some interesting comments emphasising the importance of statistics in medical and physical sciences were made by the audience.

The President of BAS, Professor Mesbahuddin Ahmad, presided over the seminar, and the Secretary of BAS, Prof Khawaja M S Aziz, introduced the speaker. In addition to the members on the executive committee of BAS a large number of scientists, including the immediate past President of BAS, Professor M Shamsher Ali, a former President of Bangladesh Statistical Association, Professor M Kabir, and Professor Liaquat Ali, Vice-Chancellor, Bangladesh University of Health Sciences, along with a number of senior biostatisticians and statisticians, participated in the event.

Later, the President of BAS presented a copy of the Academy's Year Book to Professor Shahjahan Khan along with some memorabilia.

The BAS was established in 1973 with Dr Kudrat-e-Khuda as the Founding President and 13 Fellows to promote research in pure and applied sciences and enhance their practical applications to the national welfare in Bangladesh. The Academy is the leading scientific organisation of the scientists of eminence in Bangladesh and represents the scientific community of the country. Currently there are 112 elected Fellows of the Academy.

Professor S Khan is an elected Member of the International Statistical Institute and Institute of Mathematical Statistics, and a Fellow of the Royal Statistical Society. He served as the President of ISOSS from 2005-2011.

BAS President, Prof M Ahmad (right) and Secretary, Prof K M S Aziz (left) presenting Academy's Year Book to its Fellow Prof Shahjahan Khan (middle).

Seminar audience

Seminar Speaker Prof Shahjahan Khan (standing) and BSA President Prof Mesbahuddin Ahmad

MEETING REPORT: STAT-HAWKERS AT THE JSM-2014, BOSTON, USA

In this forward, we attempt to recall some memories annotated with appropriate photographs from the booth STAT-HAWKERS at the Joint Statistical Meeting (JSM)-2014 held in Boston, USA during August 2–7, 2014.

Augustus Jayaraj, Sarjinder Singh, Oluseun Odumade, and Cheon-Sig Lee.

Cheon-Sig Lee, Oluseun Odumade and Sarjinder Singh setup the booth on August 2 and 3. Figure 1 shows front side of the booth and a tall tower. It was a challenge to fix the tower, which was 10 feet high. It was also decorated by blinking LED lights. The booth was decorated by two banners from the front and right sides showing the flyer "Thinking Statistically: Elephants Go to School" and the image of a T-shirt which was especially designed this year to promote the journal, "Model Assisted Statistics and Applications." LED lights were incorporated into the decoration around the booth. On August 3, four types of animals were employed to promote the booth. These animals included Elephants, Horses, Monkeys and Zebras. In addition two persons, Supermen and Spidermen were also very active in attracting people to visit the booth. The attendees seemed more inclined towards Monkeys rather than Elephants. We also had flashing lights, goggles, scales, white boards, bouncing balls and pens to distribute as promotional materials, although those are not visible in the picture.

Promotional material on the booth.

Oluseun was very excited to distribute the promotional material very fast, although often reminded him that one professor, Stephen Sedory, in my department suggested that we do not distribute everything in one day. Oluseun could not control his sentiments, and as per my nature, I did not stop him. Ultimately, Oluseun and Cheon distributed all 72 elephants, 50 horses, 50 monkeys, 50 spidermen, 50 supermen and 100 zebras. In Figure 2, both Oluseun and Cheon are shown filling and adjusting animals. It seems that at this time we were already out of Spidermen and Supermen. Figure 2 shows that right hand side of the booth, where three books and two journals by the IOS Press were displayed.

As seen in Figure 2, there was a doll who was singing songs and attracting people. There was also an elephant train running around a track. This year a new item was displayed a handmade "Partapa" which is a kind of balance and can stand on one leg or can dance. It can balance on your one finger. A big poster advertising the journal, "Model Assisted Statistics and Applications" with pictures of Stan Lipovetsky, Stephen Horn and Sarjinder Singh, was also displayed. This year Stan could not make it to attend the conference, but his picture was there. Another big poster of the same size advertising the monograph "Advanced Sampling Theory with Applications", was also posted.

This year, the most attractive promotional materials was T-Shirts bearing the logo "Model Assisted Statistics and Applications" on the front and at the back, "Thinking Statistically: Elephants Go to School." In the mean time, one of Oluseun's friend Augustus Jayaraj from Cornell University also joined the booth and was involved in the activities. A total of 180 Tshirts were distributed, and out of these 60 T-shirts were for kids. The cost of T-shirts distributed at the booth alone was one thousand two hundred and forty five dollars ($1245/-).

On August 4, Oluseun and Cheon tried to find more toys from China Town, Boston, but they could find only 44 toys, which includes inflatable hammers and dora beach balls. In the evening of August 4, we asked to Augustus to drive car for us to a party shop. Next day, we had more promotional material at the booth, Parrots, crabs, monkeys and fish etc. In addition, there was always,

as usual, some chocolate and candies for the booth visitors. The promotional material sent by IOS Press, The Netherlands was also distributed, which included CDs, hard copies of the journals, and sticky pads etc.. In a nutshell, we kept our booth busy for most of the time. Later, I had a special encouraging e-mail from Judy-Anne Chapman, PhD, P.Stat., PStat® (ASA) about praise of the booth activities in that this was the first time she knew of where a commercial group offered a benefit (free t-shirt) for overtly proclaiming to be accredited as an ASA PStat.

One of the major outcome of the booth was that we had a great opportunity to have a picture with the 100th president of the American Statistical Association, Dr. Fritz Scheuren. It was very nice to talk to him. The journal entitled, "Statistical Journal of the IAOS, where he is an Editor-in-Chief, was also promoted at the booth.

Oluseun Odumade, James Alan Fox, Sarjinder Singh, Cheon-Sig Lee and Augustus Jayaraj

Whatever we spent on the booth, it was returned back in one moment as soon as I saw Prof. James Alan Fox, from Northeastern University. He is the lead author of the pioneer monograph on randomized response sampling entitled, "Randomized Response : A Method for Sensitive Surveys" which is one of the SAGE Publications, published during 1986.

Figure 6 provides evidence that five guys in the US are not only cooking burgers; there is another group of five guys who is cooking randomized response models. We had the chance to chat about the future of randomized response techniques. He was so kind that he also attended both my joint presentations of the contributed papers by Cheon-Sig Lee and Augustus Jayaraj. We also showed Prof. Fox that we are collecting data utilizing the crossed model with two-decks of cards.

Figure 7 shows a respondent participating in the survey while responding to a sensitive question after drawing a card from a deck of cards.

Children at the booth, STAT-HAWKERS 2014

This was very much right time to meeting Prof. Fox because we are publishing a special issue celebrating the Golden Jubilee year of the Warner's randomized response model. The Guest Editor for the special issue is Prof. Arijit Chaudhuri, Applied Statistics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, 700108 West Bengal, India who has published three monographs on this topic. His first monograph was published in 1988 jointly with Rahul Mukherjee by Marcel Dekker, the second monograph was published in 2011 by Taylor & Francis, where he is the sole author and the third monograph was published in 2013 jointly with TC Christofides by Springer. This year seems to be so fortunate having all these experts around us when it is very much right time to celebrate the Golden Jubilee Year 2015 of the Warner's randomized response model which he had published during 1965 in the Journal of the American Statistical Association.

One thing certain is that this year not many pictures of children who visited the booth were taken. Figure 8 shows a little girl with pink glasses and a colorful ball who visited the booth Stat-Hawkers. These activities remain in the minds of kids, and should be done at a higher level so that kids could take interest in statistics conferences.

I will be completely mistaken if I will not acknowledge the help from Paul and Rasjel, IOS Press in promoting the journal: Model Assisted Statistics and Applications: An International Journal.

Thank you for your time in reading my little invented story.

Sarjinder Singh
Department of Mathematics, Texas A&M University-Kingsville, USA

Alan Welsh: Model selection in linear mixed models.

Alan Welsh gave a stimulating talk on his joint work with Janice Scealy and lead collaborator Samuel Müller on linear mixed model selection in April at the Canberra Branch. Alan introduced his presentation by remarking on the flexibility, and correspondingly, the popularity, of mixed models in analysing different types of clustered or dependent data. While linear mixed models may be considered as extensions of regression models, Alan reminded the audience that there are important differences between linear regression and linear mixed models that influence the model selection process. These differences include needing to consider the different relative importance that regression vs variance parameters should play in an analysis, and the computational and statistical issues associated with commonly establishing boundaries of the parameter space through the requirement that variance parameters should be non-negative. For his presentation in providing an overview of the methods available for linear mixed model selection, Alan focused on the fixed-dimensional case and discussed three approaches to linear mixed model selection: information criteria, shrinkage methods and fence methods.

Specifying the general form of the linear mixed model with a parameter vector theta that is p regression + q variance parameters long, linear mixed model selection involves the task of choosing a subset of models from a larger countable set of candidate models, by specifying which elements of theta are non-zero. Alan highlighted that linear mixed model selection from the perspective of selecting parameters does not necessarily accord with the perspective of selecting variables. The two perspectives are the same with fixed effects, where a focus on selecting variables by choosing columns of the design matrix corresponds to selecting non-zero parameters in Beta, the vector of fixed effects coefficients. This is not necessarily the case for the random effects component of the model, and this was illustrated by re-writing the conventional product of the random effects design matrix with the random effects vector (Zb), into three different terms.

Some analysts simplify the general model to the independent cluster (Laird-Ware) model with known variance structure, so the focus of model selection is on the fixed effects parameters. Alan remarked that one approach to model selection that analysts might commonly adopt is selecting fixed effects first, and then adding random components after, to reflect study design aspects. Alan and his colleagues are interested in the general model selection problem—the joint selection of both the fixed effects and random effects parameters.

Alan discussed loss functions used to estimate parameters and distinguished between the conditional log-likelihood, the marginal log-likelihood and the restricted maximum likelihood (REML) criteria function before introducing model selection through the minimising of information criterion. The criterion is a sum of two components: a loss function such as a log-likelihood or REML criterion which measures the closeness of the fitted model to the data, and secondly a penalty function, which increases with model complexity. What penalty to use for mixed model selection is not as clear as it is for linear regression where the penalty is often just a function of the number of parameters in the model.

Akaike Information measures the ability of a specified model to predict an independent copy of the observed data. Different versions of the Akaike Information can be used which reflect working with either the (marginal)

log-likelihood, the REML criterion function or the conditional log-likelihood. Alan discussed how the penalty for model complexity when using the Akaike Information Criteria (AIC) can be considered an adjustment to reduce bias and how bias adjustment is related to the calculation of degrees of freedom. Alan summarised several authors' approach to calculating the penalty (including degrees of freedom), for marginal AIC and conditional AIC, including the distinction between effective and generalised degrees of freedom. Alan then went onto summarise work that has been undertaken using Bayesian Information Criteria (BIC), a criterion that is consistent (unlike the AIC), as it corrects for sample size through inclusion of log(n) in the penalty term.

Alan introduced the topic of shrinkage methods for model selection, which incorporate a tuning constant which controls the amount of shrinkage in the coefficients. Shrinkage methods are more computationally feasible than using information criteria when the number of parameters is large. Alan discussed the LASSO (least absolute shrinkage operator) and alternatives like SCAD (smoothly clipped absolute deviation) and the Adaptive LASSO (ALASSO), which help subdue issues of bias associated with excessive shrinkage by the LASSO on the fixed effects parameters. Each of these shrinkage methods can be extended to the Laird-Ware model. In the linear mixed model case, it appears most authors have used shrinkage methods to select on the fixed effects parameters only, but Alan discussed the approaches of three groups of authors that had considered the more general problem of the joint selection of both fixed effects and random effects parameters.

Alan introduced Fence methods mentioning a similar theme to shrinkage methods. Just as applying a shrinkage approach by employing a choice of different tuning constants can be used to obtain a subset of models before further selection, so do Fence methods first identify a group of models within a 'fence', with the least complex (and then if necessary, the best fitting) model being chosen from that subset. The fence is established through estimating the loss for a correct (sufficiently large) model, where candidate models lie inside the fence where their loss is less than the sum of a) the loss of the correct model, and b) the product of a sequence of tuning constants and the standard deviation of the difference in loss between the candidate model and the correct model. The Simplified Adaptive Fence alleviates computational burden by replacing the above standard deviation with a single tuning constant. The Invisible Fence and the Restricted Fence are modifications which seek to further reduce computational effort and involve, respectively, bootstrapping to find the reduced set of models and the partitioning of the initial model space into subsets each of which is then evaluated using the Adaptive Fence. Alan finds the Fence methods very interesting as they are designed for mixed models, and are new and innovative.

Alan concluded his talk by reminding us that there were substantial differences between model selection in regression models compared to linear mixed models, including consideration of marginal and conditional problems. He summarised the basic approaches of information criteria, shrinkage and Fence methods, where he believes using the bootstrap in the latter approach is an important idea. Alan provided a stimulating presentation and left the audience wanting to explore some of these methods in practice. Alan's work with his colleagues is detailed in: Müller, S., Scealy, J.L. and Welsh, A.H. (2013). Model selection in linear mixed models. Statistical Science, 28, 135-167.

Philip Tennant

Modelling the Relationship between Forage Quality and Koala Population Density

On Tuesday, 30 September 2014 the Canberra Branch enjoyed a lively and interesting co-presentation by Kara and Robert. From the talk, we learned a great deal about the statistical analysis, the fieldwork, leaf chemistry and koalas. In addition to very cute pictures of koalas (and some less cute but equally interesting pictures showing some of the travails of fieldwork), we learnt two very interesting facts about koalas. First, while koalas across the whole range have chlamydia, those in the north are succumbing to it while those in the south are not. Second, koalas in the south are physically bigger than those in the north. The reasons why are not currently known.

Kara began by explaining some of the background to the study. She explained that leaf chemistry is important for keeping a balance between plants and folivores in the landscape. Generally, folivores select plants with fewer and lower levels of toxins and more nutrients. Different folivores prefer different eucalypt subgenera. For example, koalas prefer areas where there are a high proportion of symphomyrtle trees; these produce a chemical (FPC) that Koalas have adapted to but other folivores have not. Forage quality varies naturally across the landscape and affects folivore density Koala density is highly variable; in the Otways, koala density is high and koalas are defoliating the forest whereas in the north of their range, koalas are not having the same effect. The purpose of the study was to look at the relationship between koala density and forage quality across the range (from Adelaide to north of Brisbane).

The data for the study were obtained by collecting and analysing leaves from trees on sites or areas with known koala densities. As Robert pointed out, this required a massive effort both in fieldwork and in the subsequent laboratory analyses of the leaves. First, 16 regions across the range were identified. Then areas within each region were selected to give a mix of koala densities in each region. These areas included sites suitable for koalas but without koalas as well as sites with koalas. One transect was laid out within each area. Robert commented that, when he went into the field, he was impressed by the resolution with which the probability sampling protocol for selecting transects was implemented. Along each transect, the 4 trees nearest the transect every 60 m were selected to give 32 trees per site. A rope throwing device (in fact, a massive slingshot that was brought to the talk) was used to throw a rope over selected branches in these trees so that they could be pulled down and leaves collected from them. Assistance was obtained with tree identification but the field workers were on their own when it came to dealing with swamps, rain, cold, dark, numerous parasites (ticks and leeches) and snakes. The collected leaves were brought back to the laboratory, freeze-dried and ground into a fine powder for chemical analysis in an NIR machine (to measure total nitrogen) as well as in an in vitro digestion system developed by Bill Foley (ANU) to measure available nitrogen.

The study expected to find higher levels of available nitrogen where koala density is higher. However, as Robert noted, there are difficulties in showing this statistically. First, as in a case-control study, the design is based on knowing the response and sampling to measure the covariates. Next, the koala densities are measured at the site level so the leaf data has to be aggregated to the same level, inevitably at the cost of losing some information. Koala density tends to be high when there are more symphomyrtle on the site. Since symphomyrtle trees have higher available nitrogen, it is difficult to separate out the effect of
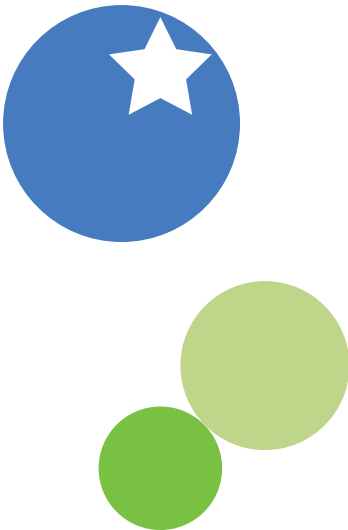
having more available nitrogen and more symphomyrtle trees. As if that is not enough, there are a number of sites with zero density and and the variability in koala density increases markedly when the available nitrogen is greater than one.

Robert described the model he developed for the koala density for area j in region i as a logistic-lognormal model. This is a logistic generalized linear mixed model (GLMM) for the presence/absence of koalas and a lognormal linear mixed model for the density of koalas given that at least one is present. The models included independent normal random effects for region and explanatory variables such as the proportion of symphomyrtle trees, the mean available nitrogen, leaf moisture, dry matter density, density of eucalypts, total nitrogen etc. A key step was to fit available nitrogen and then the effect of symphomyrtle after removing the effect of available nitrogen. This was achieved by doing a simple regression of symphomyrtle on available nitrogen and using the residuals from this regression instead of symphomyrtle in the logistic and lognormal models. Both models had significant components of variance for the random effects; the logistic model has a state effect QLD&VIC vs NSW&SA while the lognormal model has significant available nitrogen, symphomyrtle after removing the effect of available nitrogen, and VIC and SA versus NSW effects. Robert commented that, within a state, whether an area has zero density or not seems random. Some other variables of interest have not been considered yet (e.g. tree density) and their inclusion in the models may change the results.

Following the talk, there were a number of good questions and some very interesting discussion.

Kara Youngentob and **Robert Clark**
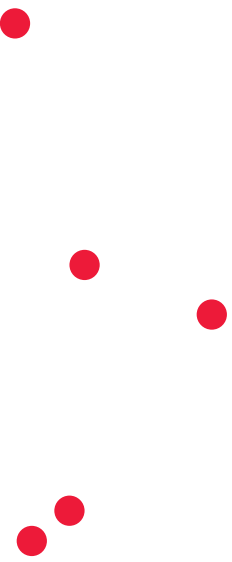
NSW BRANCH

Reproducible Research Using R (and other tools)

The August Meeting of the NSW Branch was held jointly with SURF (Sydney Users of R Forum) and featured a talk by Dr Rich FitzJohn from Macquarie University on 'Reproducible Research Using R (and other tools)'. Rich is a biologist who in the course of his scientific career has become highly proficient at using various computing technologies to not only improve his own productivity but also streamline the collaborative process.

It is becoming more and more the norm that any type of scientific publication makes available everything necessary to reproduce its published results. As such, tools have become available whereby an interested third party can recreate all the results, tables, graphics etc in a paper. As well as increasing transparency in the scientific process, these tools can be used from the genesis of a project to improve organisation and communication between a team of collaborators.

Rich discussed five tools: knitr (an R package for literate programming), git (a version control system), make (a tool for managing compilation dependencies in a software project), travis-CI (a hosted continuous integration service) and packrat (a dependency management system for R). He discussed the various strengths of each in general as well as how they were used on a real biological project he and some collaborators have been working on, seeking an answer to the question: "How many species of plant are woody?"

Knitr is the successor to Sweave. Sweave is a tool which permits mixing of R code and LaTeX code in the same file. The file is processed, any R code is run, and a LaTeX file is created which when compiled will include the R code (nicely formatted), as well as any desired output and graphics in the document. It is also possible to use Sweave to generate HTML output.

Knitr has advanced Sweave in various directions. It incorporates use of markdown, a lightweight text formatting system (where headings, lists, etc look much as they would in a plain text email) and now together with other tools such as pandoc (a converter between different document formats) and RStudio (the fantastic R integrated development environment) facilitates output to many more formats including even Microsoft Word). Rich sung the praises of knitr, also pointing out one or two of its limitations. However the take-home message was that everyone should be using knitr for many things. As well as making it easier for others to reproduce your paper, it actually forces you to be more organised, full automating many aspects of the workflow which could otherwise become "polluted" by ad hoc manual tweaking.

Git can also be used to improve an individual's organisation even though its main purpose is to make it easier for people to collaborate. The key feature is that every time you make a notable change to a file, you also put a comment in git explaining what you have done. At the very least it allows people to more easily seen what you have done. The main power of git though, as Rich explained, was that it allows anyone to go back and recreate the project at any point in time. This is another example of a tool whose primary task is one purpose (here version control), but a happy consequence of using it is that it makes everyone using it more organised, both as individuals and as a team. Rich again believed everyone should try using git and that in particular recent improvements in RStudio make this much easier than it used to be.

On the other hand, the remaining three tools (make, travis-CI and packrat) were classified by Rich as more technical and maybe not for everybody, even though they really do amazing things. Make has been around for a long time and is not so easy to use but when mastered makes complicated projects much simpler to work on. It tracks what files need to be in place for others to run/compile properly so that each time any changes are made to any files, the system only processes the files that need to be processed, not all files in the project. For large projects this can save a lot of time and effort. Some issues with make that Rich touched upon are that it is a command-line tool so many will be put off by that immediately; also there are two popular versions (GNU and BSD) which are slightly different. He dreams of a day where a usable, standardised version of make is available.

Rich then spoke about travis-CI. This is an online system whereby you can specify a full computing environment and travis-CI will "spin" up a virtual machine running this environment and then process your project in it. This forces the project team to set things up in such a way that everything is automated, including a bunch of tests. After a change is made, travis-CI reruns everything in the desired environment. If all the tests pass, nothing more happens until the next change. But if a test fails, the team receives an email with detailed information about what broke and what changes led to the breakage (version control facilitates this). This really makes tracking down bugs very quick and easy. One drawback though is that any sensitive data would be on the travis-CI machines.

Finally, Rich spoke about packrat. This is a system for keeping track of exactly what versions of what packages are needed for an R project to run smoothly. It isolates the project, so the whole system need not be exposed to customised versions of R packages tweaked for that project alone. Furthermore it maintains a record of the current state of all packages so that in the future if packages evolve the project can be aware of this. As Rich summarised: "Will it work later?"

The talk provoked some interesting discussion particularly pertaining to confidentiality: many of these tools require that data is on a non-private machine. Rich mentioned that some of these tools can be installed and run locally. Also, many discussants agreed that once they were exposed to and got used to these tools, they could no longer imagine doing any serious work without them.

This was a rather inspirational talk, in that this reviewer is now motivated to try learning git and make (for the second time!).

Michael Stewart

Use of Health Economics in Decision Making

The guest speaker at the NSW SSAI October 14th meeting was Dr Henry Cutler, a health economist from KPMG. Thousands of decisions, both large and small, are made within the health system every day. For the last decade, pharmaceutical companies have had to demonstrate not only efficacy, safety and quality but also the cost effectiveness (the forth hurdle) of their therapies. The additional scrutiny on value for money is becoming increasingly important in the wider healthcare industry. As with pharmaceutical products, health economists play a central role in gathering data, implementing statistical models, interpreting results and guiding decision making. Three specific examples were made.

Choice modelling in this context is an effort to understand individuals' decision making processes regarding treatment and healthcare decisions in general. A choice experiment often requires participants to choose 1 health related good or service from 2 or more alternatives. This forces them to consider the various trade-offs and benefits of various options in a hypothetical market. An example was given in the setting of dementia care.

Analysis of time-series data is another task familiar to most statisticians and pressed into use by health economists on a wide variety of time series data generated by the healthcare system. An example was given of an error correction model implemented in the setting of a fees charged by pathology labs under The Australian Pathology Funding Agreement.

Finally Markov models are well suited to represent evolving stochastic processes which makes them particularly useful to modelling processes in healthcare, in particular chronic diseases. An example was given on measuring the impacts of reform in mental healthcare in Australia with the requisite transition probabilities and costs elicited from previous research via literature reviews.

The use of statistical models by health economists is helping to formalize decision making in healthcare and shift emphasis towards demonstrating value for money. The barriers associated are all too familiar: gaining access to high quality and timely data, dealing with the overarching complexity of the healthcare system and the ongoing effort required to gain acceptance and widen the scope and range of issues where these methods can be put to use.

Richard Walton

Henry Cutler

Assessing the risk of a rare adverse outcome following rotavirus vaccination: a case study in biostatistical methods and collaborative engagement

The second seminar in October, organised by the SSAI-NSW branch, had Professor John Carlin from Murdoch Children's Research Institute & University of Melbourne as its guest speaker.

Professor Carlin presented a very exciting study on the safety of the rotavirus vaccine. He started by giving an historical perspective of disease management and the serious adverse effects observed with the previous vaccine that lead to its withdrawal from the market. When a new vaccine was introduced in Australia, mid 2007, there was a clear drop in the number of hospitalisations due to rotavirus infection across the entire country. However, there was concern about its safety given the previous experience. Therefore, the Therapeutic Goods Administration (TGA) commissioned a national study to monitor the potential secondary effects of the new vaccine, in particular to determine if the risk of intussusception - an acute bowel obstruction that can be life-threatening if not treated promptly - increased after vaccination.

Using the hospitalization databases, all the cases of intussusception, between 2007 and 2010, were identified and confirmed. The vaccination history of these cases was obtained from the Australian Childhood Immunisation Register. Additionally, an age-matched control group was randomly selected from the same registry.

The analysis had two strategies. The first one followed the classical case-control design and used a conditional logistic regression to estimate the risk (odds) increase of intussusception after vaccination. The second used the Self-Controlled Case Series (SCCS) method. The main idea of this approach is to compare the frequency of the outcome within a pre-specified exposure period, relative to occurrence in unexposed time in the same case. A conditional Poisson regression was used to compare at-risk intervals with other follow-up times, within individuals.

Both analyses showed an increased risk of intussusception in the 21 days after vaccination. However, the benefits of the vaccine clearly surpassed the risk of side effects, with more than 6500 hospitalisations being prevented annually due to rotavirus infection versus expected 14 additional intussusception cases per year due to the vaccination.

Professor Carlin finished his talk with an overview of the VicBiostat initiative and general remarks about the development of Biostatistics in Australia.

Armando Teixeira-Pinto

SA BRANCH

### SA Branch Young Statisticians' Careers Event

On 10 September 2014 a careers event was held for young statisticians at the SA branch of the SSAI. The evening involved short presentations by representatives from several SA organisations that employ statisticians followed by drinks, pizza and networking.

The employers represented were the following:

• Australian Bureau of Statistics

• Centre of Research Excellence (CRE) in Translating Nutritional Science to Good Health, Discipline of Medicine - The University of Adelaide

• Data Management & Analysis Centre (DMAC) – The University of Adelaide

• Flinders Centre for Epidemiology and Biostatistics (FCEB) – Flinders University

• South Australian Research and Development Institute (SARDI)

• National Centre for Vocational Education Research (NCVER)

As well as the different sizes and roles of the organisations represented, those speaking on their behalf were from a variety of experience levels. This gave a diverse and interesting range of perspectives for young statisticians to consider. Useful advice about job applications and interviews was also shared by those with experience from recent interviews and years of experience on interview panels. Over 30 young statisticians attended, and many remarked on the success of the event.

Chris Davies
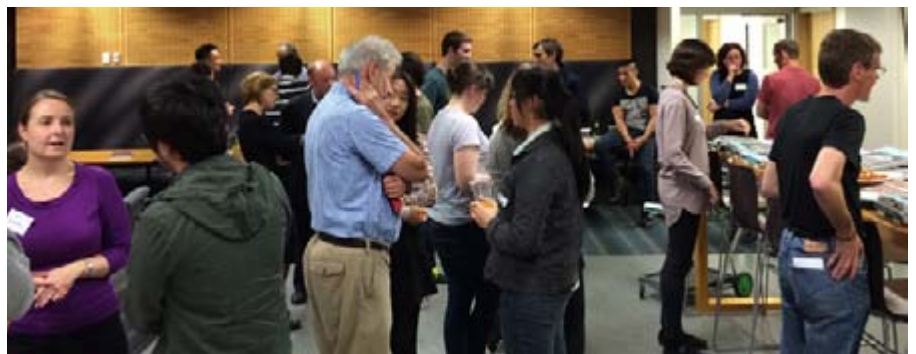
'Optimal Experimental Design for Group Dose-response Challenge Experiments' and 'Data Driven Model Selection with Approximate Bayesian Analysis'

Speakers: David Price & Ben Rohrlach (University of Adelaide)

The presenters at the September meeting of the SSAI South Australian branch were David Price and Ben Rohrlach, both PhD candidates in Mathematical Sciences at University of Adelaide. David is currently in the third year of his PhD, looking into optimal experimental design for Markov Chain models. Ben is currently in the first year of his PhD, looking at informative measures of goodness of fit for models of nucleotide substitution.

David Price began his presentation by describing the dose-response challenge experiments he is considering. A number of chickens are injected with an infectious agent (the bacteria Campylobacter jejuni) and when this occurs there is found to be a positive relationship between the dose level and the probability of infection in the chicken. Because the chickens eat each others' faeces, there is also an oral-faecal transmission from one chicken to the next. David is modelling the dose-response, latent period between exposure and infectiousness, and transmission between infectious and susceptible chickens.

Optimisation is then used on the group dose-response challenge experiments to gain the most information about the dose-response relationship, taking into account group size, dose and time to observe the experiment. A fixed number of chickens were used in two groups with the aim to find optimal group sizes and doses. Optimal experimental design for Markov chains was used within a Bayesian framework, including a 'gold standard' method that uses the true posterior, and an alternative method, which utilises a likelihood-free approach to obtaining the posterior distribution.

David proposed a new method for determining Bayesian optimal experimental designs called ABCDE (Approximate Bayesian Computation Design Efficiently), which utilises a likelihood-free Bayesian method called Approximate Bayesian Computation (Ben's talk covered one of these methods).

David discovered that the ABCDE method has consistent results with the other methods and takes considerably less time to compute, as it can be parallelised, and grossly reduces the required number of simulations while enforcing greater amount of information used in evaluating the utility of each design. His current work is investigating the effect of the latent period and transmission on the resulting optimal design for these group-dose response challenge experiments where the underlying process is modelled by a Markov chain. He is also aiming to publish a paper on the ABCDE method.

Ben Rohrlach began his presentation by considering the Beringian Steppe Bison whose population numbers dropped at some time in the past. Did it happen slowly over time or did it happen abruptly? Can answers be found from DNA in old bones?

Ben compared the frequentist and Bayesian approaches, and described the ABC (Approximate Bayesian Computation) method – a specific, likelihood-free Bayesian analysis used in population genetics. The basic concept behind ABC

is that a large number of simulated data sets are compared to the observed data set and some are retained as the sample for the posterior distribution. Ben described sufficient summary statistics, which rarely exist, and so approximate sufficient summary statistics are used to find the best predicted parameter value.

Ben described model selection in ABC using Bayes Factors, and then introduced an improved model selection method that employs MLR (Multiple Logistic Regression) and hence avoids comparing posterior distributions. MLR was used for each of the 2000 new simulations to find which model, exponential or constant (going back to the bison query), MLR would predict had been used to produce the simulation. This was then compared to the known model of simulation for verification. Ben's MLR classification returned >96% accuracy and did not require ABC to be performed on all possible models (just simulations for training the MLR).

Following the talks, the speakers and branch members adjourned for dinner at nearby Italian Café Amore.

Suzanne Edwards

Brett Bryan

Influence of policy on land use and ecosystem services: sustainability under global change

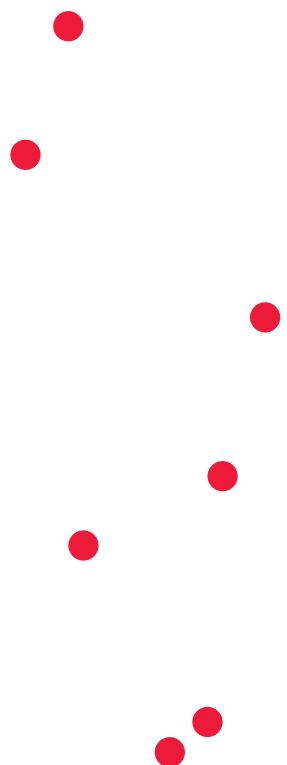### Speaker: Dr Brett A. Bryan (CSIRO)

A small but enthusiastic audience listened to the presentation of Dr Brett Bryan who is a project leader in CSIRO and has research interests in the application and development of computational tools and analytical methods in an environmental context with nearly 70 publications in the international scientific literature.

Brett began by addressing the question, "How do policy instruments influence people on the land?", motivated by increasing pressure on the land due to a variety of factors such as emerging population, competition from other land sources and productivity that once increased by 2% p.a. but has been flat since 1990. In this context Brett noted that the early integrated land use modelling was conducted by way of Global Information Systems that were somewhat limited. Nowadays high performance computing platforms exist, and he is leading the development of the Land Use trade-off (LUTO), a high resolution partial equilibrium model for assessing future scenarios for sustainability of Australia's agricultural regions. The ability of LUTO to integrate a wide variety of biophysical and economic data in a spatio-temporal model of potential land use change in Australia's intensive agricultural sector was demonstrated using several scenarios.

Brett noted that the introduction of climate (e.g. a carbon market) and energy policy may provide significant opportunities for the widespread adoption of new land use and management options for green house mitigation in agricultural landscapes. New land uses may include a range of bioenergy, carbon plantings, bio diverse plantings, and alternative crop and livestock management uses. These changes may generate collateral impacts that are both positive and negative, being both direct and indirect. The implications of policy options, changes in land use and the impact of a range of ecosystem

services such as food and fibre production, energy and biodiversity was modelled using LUTO under four global scenarios: low emissions, two different medium emission pathways and high emissions and the audience was shown, in computer time, the likely outcomes in terms of costs and benefits.

Brett's talk generated a lively discussion at the conclusion with many questions and comments. Following the presentation several members and the speaker adjourned for dinner at a nearly local restaurant.

Steve Quinn

Topic: Upskilling in Data Science in South Australia, for professionals and students.

Speaker: Professor Andy Koronios (University of South Australia)

It is has been reported that the age of Big Data will change our world and generate thousands of jobs and the data scientist may already be one of the most sought-after people in the ICT sector. What are the skills involved in becoming a data scientist? And how can they be acquired to give practitioners the edge in leveraging the benefits of such great possibilities in this field? Professor Andy Koronios of the University of South Australia presented an engaging and thought-provoking lecture which ended with an overview of his new degree course in Big Data.

But the talk began on Big Data began with a Big Bang. An astounding figure that presented as food for thought was that there was more data collected in the last week than previously collected in the whole of recorded history – but less than 1% was analysed!

We now live in an instrumental world. Consider shopping online rather than physical shopping for example - and even in physical shops, checkout cashiers now worry for their livelihoods as automatic checkouts become more and more fully automated.

Andy also made the prediction that in five years the number of connected devices at home would jump from single digits into the hundreds. We'll be connected to the fridge, the washing machine, our plants – everything will have an IP address for us to control when we're away from home. Indeed, will there even be a need for driving licences given that there are now completely automated cars being driven in the United Sates. These are prototypes yes, but still fully functional. All industries are now looking at the consequences of not being e-enabled. Consider the demise of well-known bookshops in Australia such as Borders, and how iTunes has captured the music market. The Commonwealth bank of Australia has spent $1billion on becoming e-enabled.

But this change in technology is at the same time creating huge amounts of data.

Everywhere you look there is now Big Data. Data from an A380 aircraft creates the equivalent of 1.2 billion printed pages/hour of information – so much information that it's impossible to store it all beyond a few days. We now need to think beyond terabytes to petabytes (1000TB) and beyond petabytes to exabytes (1000 PB) and zettabytes (1000 EB)

Big Data issues deal not just with volume but also with velocity and variety of data. And when these exceed the capacity of a organisations capacity for storage and processing power for accurate and timely use you have a "Big Data" problem.

So what exactly is Big Data? In essence Big Data is about predicting buying behaviours, predicting likely failure of critical equipment, improving the customer experience (better recommendations), evaluating the impact of marketing, introducing new products that haven't been identified by your competitors. Some applications of Big Data include:

Marketing, Risk management, Government – tax, social security fraud, terrorism, The Web – analytics and social media, and Logistics – demand forecasting, supply chain analytics.

At the analytic end of things, the innovations in Business Intelligence platforms, data and analytics are the number one area of investment but the amount of innovation themes and paths are causing confusion amongst managers. In short there is Big Data everywhere – there is "Hadoopalooza"! (Hadoop is an open-source software framework for storage and large scale processing of datasets – and there are many more of them available and for free!).

And where then does that leave the data scientist? And what is a data scientist? Data science is a multidisciplinary activity. It involves data capture, data management, data storage, analytics, application and evaluations. A statistician has all the attributes required for the analytics part of Big Data, but a data scientist must also have communication skills, be able to collect and clean, be better at stats than a software engineer and better at software engineering than a statistician. They must have a curiosity about working with data, be able to tell a good story. But, also because the role is also currently freeform, it's also a great time to join!

And yet, although employment growth is far beyond exponential, in 2012 there were no University programs currently offering degrees in Data Science. Now however there are more than 250 – including 4 in Australia (UniSA, Deakin, Macquarie, UTS). Andy finished the talk by detailing the opportunities available at UniSA – namely the UniSA MDSc – its collaboration with IAPA and SAS, its availability as online or face-to face teaching and its emphasis on professional practice, the development of technical skills in data science, project management, communications and visualisation.

Finally the talk was wonderfully illustrated by way of the almost ubiquitous apps. Mobile phone users with wi-fi had the opportunity to download the slides and also join in a discussion board as the talk was in progress. More data, more connectivity and more opportunities for statisticians as well as data scientists!

The evening ended with a meal at Cafe Brunelli, on Rundle Street.

Richard Woodman

Andy Koronios

VIC BRANCH

One of our members, Damjan Vukcevic, blogged about Adam Band's public lecture, discussing evidence-based policy:

On 16 September 2014 the Federal Member for Melbourne, Adam Bandt, gave a public lecture on the role of evidence in public policy in Australia. I helped to organise this talk as one of the monthly events for SSA Vic. Our goal was to hear how evidence is used (or not) by decision makers, in this case politicians.

Adam's covered many topics and fielded a large number of questions from the audience. You can listen to the recording to hear it all (approx. 1 hour). Here, I summarise the points that stood out for me.

Lessons learnt from climate change policy

Climate change featured prominently in both Adam's talk and the audience's questions. As part of his role in the previous government, Adam was frank in describing both their successes and failures. Two of these stuck with me.

Early on, the government put together a committee to develop a set of policies to tackle climate change. It consisted of parliamentarians from multiple parties, and an equal number of experts from a variety of fields. Adam said the presence of the experts changed the dynamic of discussion considerably:

'When you are sitting across the table from an expert…your ability to prosecute crap arguments diminishes drastically. You'll be held to account very, very quickly by someone who'll just tell you that's simply not right.'

Seems like a great idea to me. Getting politicians and experts talking together, surely it's a no brainer? Shouldn't this happen more often?

On the other end, one of their major mistakes started once they had developed their policy and passed the legislation. They presumed there was no longer any need to talk about the problem. The public information campaign that followed concentrated on details of the carbon price and the compensation package, with little mention of global warming or the fact that this legislation is tackling a big social problem.

'The failing to talk about the problem, and just presuming because you have a good technocratic fix to it then that's enough, is part of the problem,' according to Adam. This allowed the Opposition to shift the debate to be about something other than the underlying problem, to a debate about the Government's credibility, without any reference to climate change.

Adam's 3-step plan

Often it's easy to point out problems but much harder to come up with solutions. Adam offered us three.

1. Entrench facts into government decision making, by law

Adam suggested two ways of doing this. Firstly, by setting up a sustainability commissioner in various government departments, whose role is to provide independent scientific advice (for example, about the impacts on biodiversity or energy use). The key point is that the relevant minister would be required, by law, to take that advice into account. Of course, they could chose to 'ignore' any advice but they would need to make a statement to this effect. Adam believes this would change the dynamic of many decisions and make evidence harder to ignore.

Secondly, an increased use of randomised controlled trials (RCTs) as part of policy development. However, Adam was a bit reserved on this point, wanting to see more evidence that these are indeed effective. He mentioned that a large review was underway in the UK to assess the ability of RCTs at measuring the effects of social policy.

2. Increase the scientific literacy of the population through public education

Those who wish to attack evidence-based positions can resort to variety of underhanded tactics. One is to manufacture doubt. Another is to falsely undermine the evidence by blurring the distinction between evidence and moral values.

Adam believes that increasing scientific literacy can help to blunt both of these attacks, and also lead to increased acceptance for a greater role for evidence in decisions. He would do this by investing more in science and mathematics education in primary and secondary schools.

A byproduct of such an education would be a greater ability by the public to distinguish between the use of evidence versus the use of values to guide decisions. Hopefully, this will lead us to a situation where politicians would be allowed (in fact, compelled) to change their policies in response to new evidence without being falsely accused of 'flip-flopping'.

3. Get scientists & researchers to be more political

Adam's final message was directed squarely at us, the scientists and researchers in the audience. Unless we fight for our slice of the political pie, according to Adam, it will be instead taken by those (of which there are many) who are motivated by self-interest and not necessarily the evidence.

One way to get political is to (like Adam) leave our jobs and stand for election. It would be great to have a few more scientists in Parliament, but that won't be enough nor is it a realistic prospect for most of us.

Instead, Adam urged us to get organised and pool our efforts. Some of us will need to go out in public and advocate on behalf of scientists. We will also need an effective campaigning organisation. (Adam mentioned the Australian Academy of Science but noted that it acts more as an advisory body than as a campaigning organisation.) Comparing our plight with that of the mining industry, which collectively ran a multi-million dollar advertising campaign against the mining tax, Adam asked, 'Where is the alternative, equivalent organisation…[who will] run a TV advertising campaign for science & research?'

The question of money arose. Adam admitted that this is indeed a challenge. However, a surmountable one. He said we need to find 'allies' out there who have an interest in Australia being a well-resourced, research & science community. There are many of them around and they are just waiting to be pulled together.

If you have interesting blogs that you'd like to share, please email the SSAI office at eo@statsoc.org.au.

WA BRANCH

WA Branch joint meeting with the International Biometric Society, Australasian Region

On the 12th August 2014 Dr Karyn Reeves from the Department of Agriculture and Food, Western Australia (DAFWA) spoke at a Special joint meeting of the Statistical Society of Australia, WA Branch and the International Biometric Society, Australasian Region. Her talk was titled "Using linear mixed models and principal components to identify patterns of genotype by environment interaction in pre-breeding trials."

Karyn explained that pre-breeding multi-environment trials (METs) are conducted to identify the best performers amongst newly developed test lines of cereal crops. In these trials the trait of interest is typically yield, but it may involve the ability to maintain yield in constrained circumstances, such as the ability to tolerate salinity, drought or disease. This was the case in the data she was discussing which concerned wheat yields from the MET drought benchmarking study (Merredin site shown below with rain-out shelter), a five-year nationwide project being led by Ben Biddulph from DAFWA which aims to rank named varieties and breeding lines for their ability to withstand drought.

She explained that METs involve the combined analysis of the observations from a series of trials held in different geographic locations and across several years, and that in this context genotype refers to the same kind of idea conveyed by the word cultivar, while environment refers to a year by site combination. Since there is no way to measure the performance of a cultivar free from the confounding influence of the environment in which it is trialled, it is the genotype by environment interaction that is studied. This interaction is manifest as different rankings of the genotypes within each environment. It complicates the selection of superior cultivars and means that breeding lines must be trialled in a wide range of conditions before they can be considered for release.

Linear mixed models are particularly suited to the analysis of agricultural data because they can accommodate the sources of variation induced by designed experiments, but they have additional features which are important in the analysis of MET data, such as the ability to accommodate unbalanced data which occurs because the set of genotypes grown can alter both across years and across locations, and the flexibility to accommodate the modelling of trial-by-trial spatial variation which can be due to a variety of factors such as soil types, fertility gradients, and such things as how the varieties are sown and harvested.

Karyn then explained the ideas of linear mixed model estimation and testing which involve fixed and random effects. This involved briefly discussing REML and the relationships between BLUEs and BLUPs with reference to an example given in "That BLUP is a Good Thing: The Estimation of Random Effects" (G.K.Robinson, Statistical Science 1991,Vol 6, 15-51), and followed with an explanation of the differences between single stage versus two stage MET analysis noting that she felt these titles somewhat misleading as the need for spatial modelling means that in practice both approaches required at least two stages of analysis.

To quote Karyn's notes: "

- Whether you plan to use a single-stage or a two-stage approach, you must first conduct a preliminary analysis of each trial on its own, using genotype as a random effect, in order to control for spatial variability and to look for outliers. The two approaches differ according to what you need to take away from this trial-by-trial analysis.

- If you are planning a single-stage analysis, then you need to take away the spatial models applied to the data from each trial.

- If you are planning a two-stage analysis, then you repeat the preliminary trial-by-trial analysis but now using genotype as a fixed effect and imposing the spatial models you identified through the preliminary analysis in order to generate BLUES, and you also need to capture a set of weights.

- So whereas in a single-stage analysis we have an observation for every plot in every trial, in a two-stage analysis we have an observation for every genotype in every trial."

Karyn then explained the current approach to conducting an MET analysis in Australian agriculture, beginning with the modelling of the spatial variation through the use of autoregressive correlation structures, factors and covariates guided by patterns evident in variograms and residual plots. She additionally explained the idea behind the factor analytic approach which overcomes estimation problems by assuming the genotype by environment interaction can be modelled by a weighted sum of a set of hypothetical or latent factors.

R code was shown which illustrated the use of ASReml-R (http://www.asreml.com) package. A comparison was presented of results from analysing the MEF wheat yield data using mixed-model versions of the approaches of Finlay-Wilkinson and Eberhart-Russell; and of the factor analytic method (single-stage, and two-stage with two choices of weights) developed by Alison Smith and her colleagues. See for example "The analysis of crop variety evaluation data in Australia." Smith AB, Cullis B & Gilmour A, 2001, ANZ J Stats 43:129-145 or recent work, "Factor analytic models for the provision of grower information from national crop variety testing programs." Smith AB, et al Theoretical and Applied Genetics (submitted 2014).

At the end, there was discussion about convergence of the algorithms with specific relation of the code, and the problems induced by agricultural practices such as sowing and harvesting. Following the talk, further discussions were held at a nearby restaurant. Special thanks to SSAI WA Branch and IBS-AR for the pre-dinner drinks and post-dinner dessert/coffee, respectively.

Brenton R Clarke

Members of the IBS-AR and SSAI WA Branch enjoying dinner following a joint meeting of the Societies.

Climate science: Where are the statisticians?

The September meeting of the WA Branch was presented by former Branch President Dr Aloke Phatak, currently the Team Leader of Hydroclimate Statistical Sciences, CSIRO Digital Productivity. He made a plea for statisticians to get involved in both climate science and the public discussion on climate change. Aloke's presentation generated an enthusiastic discussion, continuing over dinner with members.

Aloke's presentation started by noting that the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report, released in September 2013, did not include any statisticians in its list of drafting authors. Aloke explained that as this report was being breathlessly live-blogged around the world, a group of eminent statisticians published a commentary in Nature, pleading for more collaboration between statisticians and climatologists. The commentary pointed out that many of the skills and methods that statisticians take for granted (being able to come up with sensible estimates of uncertainty being chief among them) could help scientists, policymakers, and the public come to a better understanding of the risks of climate change. As the statisticians noted in their commentary, "Much of the disagreement is due to uncertainty" and "Use of modern statistical methods could greatly reduce estimates of uncertainty". Of course, who better to assist than statisticians?

Aloke encouraged statisticians to "join the conversation", be public and be heard. He briefly discussed a blog post article that one of his colleagues had published on the website known as The Conversation. The post, which was based on a recently-published journal article, was written in accessible language and explained why it was very likely that greenhouse gas emissions arising from human activity were responsible for observed mean global temperature increases. The blog post had a lot of hits, and attracted many, many comments, not all of them particularly intelligent. This was a great example of a statistician willing to take a risk and "go public". Be aware though, the headline that a website editor chooses for an article may not always accurately reflect its precise content!

Aloke identified that the statistical community needs to work towards providing the tools, education and improvements necessary for climate scientists use statistics sensibly. Unfortunately it is too common an occurrence that "non-standard" (often poor) statistics get used in climate science, and then gets propagated throughout the field. Aloke implored us to not just complain from the sidelines, but to be constructive and public in providing suggestions and advice for carrying out better scientific analyses. We should be asking ourselves the question "What is the scientific impact of using a more appropriate methodology?", and then promoting the findings in a manner that is not only understood, but taken on board by scientists and policy makers. As statisticians we often err on the side of being too cautious when reporting on our findings, providing too much in the way of technical detail and caveats, thereby losing the key message and outcomes of our analysis. Instead, we need to make an effort to market ourselves and our findings, converting our results into our own 'attention grabbing headlines', but headlines that are statistically and scientifically valid.

Aloke described such an attempt that he and a colleague made in response to a journal article describing trends in tropical cyclones. The article was cited in a newspaper article in the lead-up to Cyclone Yasi in 2011. The analysis in the journal article consisted of fitting a trend line to annual counts of severe

tropical cyclones making landfall in north eastern Australia using ordinary least squares, and finding a linear decrease over time. Of course, the data consist of zeroes and ones, (with the occasional count of 2 or 3), and carrying out the analysis using Poisson regression led to the conclusion that the rate is constant over time. This is not mere statistical pedantry, but shows how using the appropriate statistical method can lead to valid scientific conclusions that may subsequently be used in policy-making and in informing the public about changes in risk over time.

So as a profession, what do we need to do? Aloke proposed the following:

- Help put the 'science' back into climate science by:
  - Working with climate scientists to learn about the subject matter;
  - Convincing our collaborators that statistics isn't just a collection of methods that will automatically lead to the 'correct' scientific conclusion;
  - Reinventing 'classical' methods such as ANOVA for new data types such as spatial output from global and regional climate models; and
  - Not being afraid to stick our heads above the parapet.

Aloke concluded with the sentiment that if he had encouraged just one member of the audience (and now the readership of this article) to take up some of these challenges in statistical climatology, then he would have fulfilled his objective!

Anna Munday
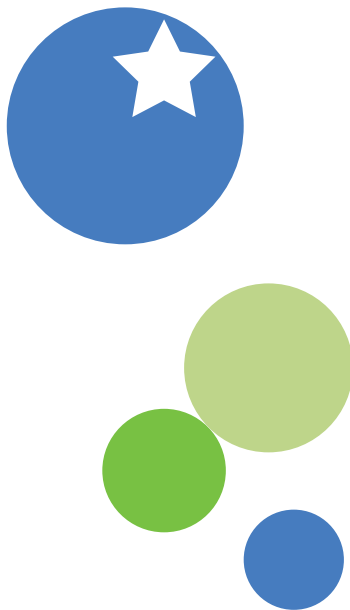
FROM THE OFFICE

As I write this we are celebrating the first anniversary of the SSAI website. It has been a steep learning curve for us at the office, but one year on from its implementation we now feel comfortable navigating our way around the website and online database. Most members have embraced the online facility, renewing their membership with no fuss at all and the number of manual renewals is almost non-existent these days. Obviously there were teething problems but I think we have worked through them quite well and I feel comfortable with where we're at. One thing that could be improved is the number of members getting involved in the "my community" area of the website. This is what I call our internal "pretend-Facebook" page, where members can start discussions and generally interact with each other. To make it an interesting area to visit, we need some people to take the lead and start discussions. Our membership numbers are looking very good, so with close to 730 members there is great potential for very lively online conversations!

I am also very happy with our Jobs Target page and pleased to see that some advertisements have been published there. As a member of SSAI you can post your resume for free and even if you are not actively looking for a new job it might be a good way to let potential employers know what they are missing!

For our webinar program in 2015 I am not only looking for potential speakers but also for enthusiastic young statisticians willing to moderate a webinar. You would only commit to a single webinar and your role would be to welcome the speaker and moderate any questions at the end of the event. If you are interested, please send me an email (eo@statsoc.org.au). If you have thought about volunteering with SSAI, this is an ideal opportunity – a one-off task that is not too onerous and gets your name out there in to the statistical community. I look forward to hearing from you.

With the summer holidays coming up, the business hours at the SSAI will be somewhat erratic. The office will be closed from Friday, 19 December 2014 until Monday, 12 January 2015. However, around 15 January I'll be moving house and be only at the office intermittently until the end of the following week. I will try to check the emails whenever I can, but please bear with me if you don't get a reply to your enquiries straight away. I apologise for any inconvenience.

This is the last "From the office" column that will be edited by our wonderful, hard-working newsletter editors Alice Richardson and Michael Adena. Michael and Alice were already newsletter editors when I started to work for SSAI in 2008 and they have guided me through the administrative side of producing the newsletter with immense patience and a great sense of humour. They have been great to work with and I will miss them very much. All the best!

I'd like to send out another "thank-you" to all the members of SSAI who have kept my job so pleasant and interesting for another year, and I include the members of all the committees I work for in this statement. I can't imagine working anywhere else because I feel so comfortable with you! Those of you who start or finish your emails to me with a friendly one-liner have no idea what it does to me to hear your kind words. It brightens up a day that otherwise might have been a bit dull and it puts a spring in my step when otherwise I might have been a bit down. I really appreciate those little things.

I'll finish off my column by wishing you all the best for the upcoming festive season and I look forward to working with you again in 2015.

Marie-Louise Rankin
Executive Officer

HAPPY HOLIDAYS