

ASC/NZSA 2006 – Statistical Connections!



ASC/NZSA 2006, the international joint conference between the SSAI and the NZSA to be held in Auckland over July 3-7 has attracted substantial registrations to its outstanding scientific program. Features of the program include 4 plenary addresses, a full day focussing on official statistics, over 50 invited talks and over 100 contributed talks as well as poster presentations. Complementing the conference scientific program are workshops on specialist topics and optional tours and social events. For up-to-date and detailed information on the 2006 Program visit:

<http://www.statsnz2006.com/program.asp>.

Apart from its scientific program, the ASC/NZSA 2006 conference will provide a great opportunity for statisticians from Australia and New Zealand to renew or develop links. There will be a Welcome Reception, included in registration, on Monday evening to allow delegates to get to know each other. The optional Gala Dinner on Tuesday night will be held in the Auckland Room at SKY CITY. This will be an ideal opportunity to mingle and meet fellow delegates while enjoying a sumptuous three course meal and New Zealand wines.

The four Plenary Sessions will feature outstanding international speakers. On the first day, Professor David Donoho of Stanford University will deliver his lecture on “*Sparsity in Estimation and Detection*”. On the third day

Professor Peter Hall of the Australian National University will talk on “*PCA For FDA*”. In the closing plenary session on Thursday, Professor Xiao-Li Meng of Harvard University will talk on “*(Data) Size Does Matter, But You Might Be In for a Surprise ...*”.

The remaining plenary session will be the Foreman lecture to be given by Professor Ray Chambers of the University of Wollongong. His talk will address *Models and Auxiliary Information in Survey Sampling*.

The Foreman lecture will kick off a full day theme devoted to Official Statistics. The Foreman lecture will be followed by an invited session featuring Alastair Scott and Kirk Wolter. After lunch there will be an Official Statistics Forum featuring the New Zealand Government Statistician Brian Pink and the Australian Statistician Dennis Trewin, followed by a contributed session. Other contributed sessions on various aspects of official statistics will be held throughout the conference.

The conference has attracted a wide range of leading invited speakers. The 25 Invited Speaker Sessions will feature leading international and local speakers presenting a wide range of the latest research in Statistical and Machine Learning, Statistical Inference, Statistics in Ecology and the Environment, Statistics in Biological Sciences and Medicine, Stochastic Processes, Statistical Inference Econometrics and Finance, Computational Statistics and High Dimensional Data Modelling.

Support for the Contributed program has been overwhelming. Over 30 contributed paper sessions and poster sessions have been organized to complement the Plenary and Invited Sessions. Young Statisticians have organized an invited and contributed sessions and there will be a focal point for them at the Welcome Reception.

Associated workshops on Distance Sampling, Stochastic Processes and the R-Fest Workshops on R Graphics, S programming and Bioconductor. Find out more on the workshops at the following link:

<http://www.statsnz2006.com/workshop.asp>

A series of tours has been organized to allow delegates the opportunity to experience first hand beauty of Auckland and surrounding areas. Please visit the website:

<http://www.statsnz2006.com/sprogram.asp> for more details on these.

We look forward to seeing you in Auckland!

SSAI Review of Statistics at Australian Universities – Follow-up

Since the review report was published in December 2005, it has been distributed to all SSAI members along with a broad range of other stakeholders and interested parties. An electronic copy is readily available from the SSAI website address <http://www.statsoc.org.au/Review0405/ReviewofStatsFinalReport.pdf> Further hard copy booklets are available from Jane Waslin at SSAI.

The SSAI review was discussed by a number of people as part of their representations to the Review of Mathematical Sciences during February this year.

Discussions are planned with senior DEST (Department of Education, Science and Training) officials in conjunction with the advisor (Dr Jade Sharples) to the new Minister of Education, Science and Training (Hon Julie Bishop MP).

There are four main recommendations in the SSAI review report:

- Issues relating to Statistics in Schools
- Issues relating to the Management and Organisation of Universities
- Issues relating to the University/Employer Interface
- Issues relating to Marketing, Communications and Lobbying.

SSAI is setting up two implementation advisory groups with one to focus on the recommendations related to Statistics in Schools and a second group to focus on the second and third main recommendations that concern universities and the university/employer interface. The final recommendation will be looked after by a working party that will focus on the more direct aspects that SSAI and related bodies can do to improve the information available to school students, university students and careers advisors.

The two main advisory groups will be made up of representatives from

the various stakeholders (government departments, universities, professional bodies) as well as statisticians representing a broad cross section of view points. Two main roles are anticipated: firstly, the groups will connect various networks together so that there will be well-informed discussion at various levels and secondly, they will act as forums where ideas and proposals can be debated. As a professional body, SSAI has a responsibility to ensure that there is open constructive debate along with a propensity to use objective methods of validation. There are many 'schools of thought', passions, biases and opinions and these must be heard and considered but ultimately we must be swayed by what 'works'. Of course defining what is meant by this is part of the debate and teasing out the various contexts will only help us all move forward.

Neville Bartlett

In this issue

ASC/NZSA 2006	1
SSAI Review of Statistics at Australian Universities	2
President's Corner	3
ABS Statistical Scholarships awarded in Canberra	3
Editorial	4
Three Doors	5
Notice of AGM	6
Update on AusCan Scholar visit	7
ASC/NZSA Financial Assistance for Young Statisticians	7
ABS Centenary Celebrations Conclude	8
Branch Reports	10

ABS Statistical Scholarships awarded in Canberra

On Thursday 4 May 2006, four students at the Australian National University (ANU) were awarded an Australian Bureau of Statistics (ABS) Statistical Scholarship. These students were Matthew Beale, Taleitha Coon, Adam Franklin and Jessica Twigg. The ceremony and accompanying lunch were hosted by the School of Finance and Applied Statistics (FAS) at the ANU, with well over 100 persons in attendance. After an introduction by Dr Michael Martin of FAS, the awards were presented by Rebecca Cassidy and Eileen Fahey of ABS.

Several ABS Statistical Scholarships are awarded each year by the ABS to students at selected universities who have excelled in

statistical subjects in their first and second years. These universities are the University of Adelaide, the University of Wollongong, the University of Queensland, and the ANU. Each scholarship is to the value of \$4000, and is offered for a full year of study in statistics. Application forms for students who wish to apply for the next round can be obtained from the Careers Office or College/Faculty staff from the relevant universities, or by contacting the ABS, for example by emailing to methodology@abs.gov.au. The closing date for applications in 2006 is early September. The ABS places no restrictions of future employment on winners of the award.

Borek Puza

President's Corner

Reduced student membership rates from 2006

At the end of last year the SSAI Executive Committee of Central Council took the decision to reduce annual membership fees for student members. This means that for \$15 capitation per year to the central body, student members receive copies of every issue of the SSAI Newsletter and electronic access to the Australian and New Zealand Journal of Statistics. I congratulate those branches that only charge this amount as they carry any additional cost at the local level.

SSAI encourages student involvement in its activities and offers an opportunity for networking with experienced statisticians as well as other students and statisticians starting out in their careers. In recent years branches and sections of SSAI have organized workshops, short courses and symposia on a wide range of topics. Members of SSAI are able to take advantage of reduced registration fees for all such events. We also have an active Young Statisticians Section which is planning a conference for 2007. Young Statistician events are held regularly in each Branch – members and prospective members should contact their local branch for details of forthcoming events. There is a link to all branches from the SSAI website – www.statsoc.org.au.

ASC/NZSA 2006

I hope you have all seen the varied and interesting program for the joint SSAI/NZSA Conference in Auckland on 3-6 July. If you haven't already decided to attend, you might like to reconsider that decision. Contrary to my advice in the last newsletter, the SSAI Annual General Meeting will be held on Wednesday, 5 July. Some of the major items for discussion will be the recommendations and resulting outcomes from our Review, the way we elect our office-bearers, capitation fees, and the future of our journal. If you have any issues that you would like addressed, please contact me or our Executive Officer Jane Waslin.



Kaye Basford
E-mail: k.e.basford@uq.edu.au

The latest in statistics from Australia, New Zealand and the world online.

Published on behalf of the Statistical Society of Australia Inc. and the New Zealand Statistical Association

Your Research Your Journal

Read it online at
www.blackwell-synergy.com/loi/anzs



Managing Editor:
Professor Kerrie Mengersen

Print ISSN: 1369-1473
Online ISSN: 1467-842X
Current Volume: 48
Frequency: 4 issues per volume



**PO Box 5111,
Braddon ACT 2612
Phone (02) 6249 8266
Fax (02) 6249 6558
Email: admin@statsoc.org.au
Society Web Page
<http://www.statsoc.org.au>**

Editors

Alice Richardson, School of ISE,
University of Canberra,
PO Box 1, Belconnen ACT 2616

Michael Adena, Covance Pty Ltd
PO Box 5125, Braddon, ACT 2612

Correspondence

Please direct all editorial
correspondence to Alice Richardson.
Email: newsletter@statsoc.org.au

Disclaimer

The views of contributors to this
Newsletter should not be attributed
to the Statistical Society of Australia,
Inc.

Subscriptions

The Newsletter of the Statistical
Society of Australia is supplied free
to all members of the society. Any
others wishing to subscribe to the
Newsletter may do so at an annual
cost of A\$30.00 for an issue of four
numbers.

Advertising

Advertising will be carried in the
Newsletter on any matters which
the Editors feel are of interest to
the members of the Society. For
details of advertising rates, etc.
contact the SSAI Executive Officer at
newsletter@statsoc.org.au

Printer

National Capital Printing
22 Pirie Street, Fyshwick ACT 2609

**DEADLINE FOR
NEXT ISSUE:
10 August 2006**

Editorial

Have you ever wondered what the process of producing the newsletter involves? After the deadline for copy passes, the Editors meet with the Society's Executive Officer to collate the material sent in, select an item for the front page and edit items so that everything fits within a multiple of four pages. The material is sent to the printers who produce a draft version of the entire newsletter. This is proofread again by the editors, any necessary changes are made and the draft sent back to the printers for correction. This is usually the last stage of the iterative process, and if the second draft looks fine it is cleared for printing. The printing company also has a mailing house that uses mailing address sheets provided by the Executive Officer that are sent with your copy. The plan is always to have the newsletter for a given month posted to members by the middle of the month.

The Editors are grateful to Branch correspondents to supply meeting reports, to committee chairs who report on the work of committees within the Society, and to individuals who supply news of members. We are always on the lookout for more of this sort of material, so don't be shy! If you attended an interesting statistical event that took place in your area, let us know! If you have carried out an interesting or slightly controversial data collection or analysis, let us know! The Newsletter can only print the material that you send in.

We look forward to a deluge of fascinating articles in the months to come.

We also hope you like the new design for the Newsletter, which gives a more modern look and feel.

Looking for a job?

For a listing of current statistical
vacancies in Australia or New Zealand visit:

<http://www.statsci.org/jobs>

With 255 positions listed last year.

**Do you have a job to
advertise on the website?**

Email a position description to
mrobinson@wehi.edu.au. Listing is free!

This service is proudly brought to you by:



Statistical Society of Australia Inc.

Three Doors with Borek Puza (Edition 6)

Welcome to the sixth edition of *Three Doors*. In the last edition I presented The Birthdays Puzzle. Unfortunately, no-one formally emailed a correct solution to newsletter@statsoc.org.au (although several persons had a go at the puzzle and one submitted a correct solution to me personally but deemed it 'not clever enough'). Anyway, the solution to this puzzle is presented below. The next puzzle follows. It has an interesting solution and was brought to my attention by Steve Clarke, who previously submitted a correct solution to The Chess Puzzle. Thanks Steve.

The Birthdays Puzzle

In a room there are 20 persons. (a) Find the probability that exactly 6 of these persons have a shared birthday (i.e. a birthday which is the same as at least one other birthday). (b) Then find and tabulate to 5 decimals the complete probability distribution of the number of persons with a shared birthday. (c) Using (b), or otherwise, find the expected number of persons with a shared birthday. (It may be assumed that the 20 birthdays are distributed independently and uniformly over a year with 365 days.)

Solution to The Birthdays Puzzle

(a) Six persons can have a shared birthday in one of 4 ways:

- (i) all 6 on the same day
- (ii) 4 on one day and 2 on another
- (iii) 3 on each of 2 days
- (iv) 2 on each of 3 days.

Our task is to count the number of sample points corresponding to each way. As regards (i), there are 365 days in the year, $C(20,6) = 20!/(6!14!)$ ways to choose 6 persons from 20, and $P(364,14) = 364!/350!$ ways in which 14 persons can have different birthdays on 364 days. Thus the number of ways that 6 persons can have a shared birthday, with all 6 on the same day, is

$$(i) 365C(20,6)P(364,14).$$

Likewise, the other three numbers of sample points are:

$$(ii) P(365,2)C(20,4)C(16,2)P(363,14)$$

$$(iii) C(365,2)C(20,3)C(17,3)P(363,14)$$

$$(iv) C(365,3)C(20,2)C(18,2)C(16,2)P(362,14).$$

Summing all 4 numbers and dividing by the total number of sample points, 365^{20} , we obtain the required probability, 0.00823.

(b) Repeating the logic in (a) for numbers other than 6, we can obtain the probability mass function (pmf) of

N , the number of persons who have a shared birthday. To 5 decimals, this pmf is given by:

$$f(0) = P(365,20)/365^{20} = 0.58856$$

$$f(1) = 0 \text{ (exactly)}$$

$$f(2) = 0.32320$$

$$f(3) = 0.00559$$

$$f(4) = 0.07132$$

$$f(5) = 0.00218$$

$$f(6) = 0.00823$$

$$f(7) = 0.00033$$

$$f(8) = 0.00054$$

$$f(9) = 0.00002$$

$$f(10) = 0.00002$$

$$f(11) = 0.00000 \text{ (meaning } < 0.000005)$$

$$f(11) = 0.00000$$

$$\dots\dots\dots$$

$$f(20) = 0.00000.$$

(c) The expected number of persons who have a shared birthday is

$$EN = 0f(0) + 1f(1) + 2f(2) + \dots + 20f(20) = 1.0158.$$

Discussion

The above working for N 's distribution in (a) and (b) was rather tedious, and it does not generalise easily. There does exist an explicit formula for N 's pmf, $f(n)$, but it involves Stirling numbers, is fairly complicated, and so will not be presented here. See Puza and O'Neill (2006) for more details. By contrast, N 's expected value can be obtained very easily, as

follows. Let X be the number of persons who share a birthday on 1 January, so that $X \sim \text{Binomial}(20, 1/365)$, and define $Y = XI(X > 1)$, where $I(\cdot)$ is the standard indicator function. Then the expected number of persons who share a birthday on 1 January is

$$\begin{aligned} EY &= EX - E\{XI(X = 0)\} - E\{XI(X = 1)\} \\ &= 20(1/365) - 0 - 20(1/365) \\ &\quad (364/365)^{19} = 0.00278307. \end{aligned}$$

Hence the expected total number of persons who have a shared birthday is

$$EN = 365EY = 1.0158, \text{ as before.}$$

The Passenger Puzzle (courtesy of Steve Clarke)

There is an aeroplane with 100 seats and 100 passengers, each with a seat assignment. The first passenger to enter the plane loses his seat assignment, and so picks a random seat. Each of the following passengers, entering one by one, takes their seat if available; otherwise they pick randomly an available seat. What is the probability the 100th passenger sits in their own seat?

For your chance to win a fabulous mystery prize, please send your solution to newsletter@statsoc.org.au.

References

Puza, B.D., and O'Neill, T.J. (2006). Two birthday problems. *Communications in Statistics - Theory and Methods*, **35**, 415-422.

Conferences

ASC/NZSA 2006 Statistical Connections, Auckland, New Zealand
3-7 July 2006. www.statsnz2006.com

ICOTS July 2006 (even though it clashes with ASC, but then so does Mathsport). URL is <http://www.maths.otago.ac.nz/icots7/icots7.php>.

BioInfoSummer, ANU, Canberra

Dec 4-8. There's no website yet but the email is BioInfoSummer@cbis.anu.edu.au

8th Australasian Conference on Mathematics and Computers in Sport, Coolangatta, Queensland

3-5 July 2006. Contact: jhammond@scu.edu.au

ISI2007, Lisbon, Portugal

<http://www.ine.pt>

Diary note – SSAI Young Statisticians section is planning a conference for April 2007 – see future issues or the SSAI website for updates.

NOTICE

The Annual General Meetings of
The Statistical Society Of Australia Inc and
The Australian Statistical Publishing Association Inc.
will be held on Wednesday 5 July 2006 at 4.15pm
at SKY CITY, Auckland, New Zealand

SSAI Annual General Meeting — Agenda

1. Apologies and Proxies
Proxies must be given in writing as per form inserted in the June 2006 issue of *SSAI Newsletter*. Proxy forms must be received by the SSAI Executive Officer for passing to the Secretary no later than 24 hours before the time of the meeting.
2. Confirmation of the Minutes – Minutes of the meetings as circulated
3. Matters arising
4. Reports
 - 4.1 President
 - 4.2 Treasurer
 - 4.3 Branches
 - 4.4 Sections
5. Accreditation
 - 5.1 Report from Accreditation Committee
6. Conferences
 - 6.1 ASC 2006 (joint with NZSA)
 - 6.2 ASC 2008
7. Election of Section Chairs
Nominations for Section Chairs should be received at the SSAI office no later than 23 June 2006. Nomination Forms have been inserted in each copy of the June issue of *SSAI Newsletter*. All nominations require a seconder and a statement from the nominee that she or he is prepared to stand.
8. Appointment of signatories
9. Special Business
Consider and, if thought fit, approve amendments to the Rules of the Statistical Society of Australia Incorporated. (**)
10. Other business
11. Time and place of next meeting.

ASPAI Annual General Meeting — Agenda

1. Apologies and Proxies
Proxies must be given in writing as per form inserted with June 2006 issue of *SSAI Newsletter*. Proxy forms must be received by the SSAI Executive Officer for passing to the Secretary no later than 24 hours before the time of the meeting.
2. Confirmation of the Minutes – Minutes of the meetings as circulated
3. Matters arising
4. Presentation of the Annual Report by the Editor of the Australian and New Zealand Journal of Statistics
5. Presentation of the Annual Report by the Newsletter Editors
6. Treasurer's Report
7. Appointment of signatories
8. Special Business
Consider and, if thought fit, approve amendments to the Rules of the Australian Statistical Publishing Association Inc. (**)
7. Other business
8. Time and place of next meeting.

(**) Copies of the amendments to the Rules of the Statistical Society of Australia Inc. and of the Australian Statistical Publishing Association Inc. may be obtained from the Society's office, or viewed on the Society's Web site at www.statsoc.org.au/WhatsNew/

Update on AusCan Scholar visit

The inaugural AusCan Scholar, Dr Mu Zhu, is planning on visiting statistics groups in Australia between mid September and early November this year.

An AusCan Scholar liaison group has recently been established in Australia to help communicate and coordinate Dr Zhu's visit to statistical groups in specific cities including arranging opportunities for Dr Zhu to present seminars at branch meetings wherever possible and appropriate.

The group comprises representatives from each branch, see below:

QUEENSLAND

Melissa Dobbie
(Melissa.dobbie@csiro.au)

SOUTH AUSTRALIA

Margaret Swincer
(mswincer@workcover.com)

ACT

Brent Henderson
(brent.henderson@csiro.au)

VICTORIA

Debra Partington
(debra.partington@dpi.vic.gov.au)

WESTERN AUSTRALIA

Aloke Phatak
(aloke.phatak@csiro.au)

NEW SOUTH WALES

Caro Badcock
(caro.badcock@covance.com)

In his most recent proposed itinerary, Dr Zhu expects to visit Sydney, Brisbane, Perth, Melbourne, Adelaide and Canberra. Specific dates of travel are still to be established and these will be communicated to the AusCan Scholar liaison group when available. Any individuals or groups interested in meeting and/or having Dr Zhu spend some time with them during his trip should contact their branch representative (listed above) as soon as possible to register your interest and to help Dr Zhu develop a program that will allow him to fulfil the objectives of the scholarship.

Melissa Dobbie

Melissa.dobbie@csiro.au

CSIRO Mathematical and Information Sciences

ASC/NZSA Financial Assistance for Young Statisticians

The March 2006 issue of the *SSAI Newsletter* contained advertisements for funding available to assist undergraduate students and young statisticians to attend the ASC/NZSA Conference in Auckland in July this year.

Tilman Davies from Western Australia was awarded funds from the CSIRO pool for undergraduate students.

Joesph De Livera, Joanne Potts and Sharon Lau were awarded funds from the Data Analysis Australia pool for recent graduates.

Watch a future issue of the Newsletter for their reports on their experiences.



3-7 July 2006 Auckland, New Zealand

Hosted By

The Statistical Society of Australia (SSAI) and The New Zealand Statistical Association (NZSA).

Contact Details

ASC/NZSA 2006 Conference Managers
GPO Box 128, Sydney NSW 2001
Phone: +61 2 9265 0700 Fax: +61 2 9267 5443
Email: statsnz2006@tourhosts.com.au

Expression of Interest

If you are interested in attending the Conference,
please register your interest
on-line www.statsnz2006.com

Scientific Program

A stimulating and cutting edge Scientific Program is being developed to cover a wide range of topics relevant to all statisticians. The program will provide practical knowledge and insights from prominent international and Australasian speakers and will address the latest developments in statistical research, education and practice.

Workshops

Technical workshops that are of particular interest to practitioners will be included in the Conference Program. The Scientific Program Committee is seeking potential workshop presenters. If you are interested in contributing please contact David Scott at d.scott@auckland.ac.nz.

Australian Bureau of Statistics Centenary Celebrations Conclude

The Australian Bureau of Statistics has just wrapped-up its centenary celebrations, with a finale on December 8. Like all big anniversaries, it was a time to reflect on the past and to look forward.

Since colonial times, Australia's official statistics have been regarded as first class. To a large extent, this has been because sound and contemporary statistical methods have been embraced in the compilation of official statistics. Over the past 50 years, the statistical methods used by the ABS have been heavily influenced by developments in mathematical statistics.

Other Centenary celebrations include the release of the ABS history book, *Informing a Nation*, a National Statistics Day (8 December), the Centenary Ball and the annual Ken Foreman lecture.

In delivering the latter, the Australian Statistician, Dennis Trewin, reviewed the development of statistical methods in official statistics over the past 100 years. He also speculated how mathematical statistical techniques might impact on the future work of the ABS.

He began by taking a look at the legendary figure for whom the lecture is named. The Bureau's long history of commitment to research and methodology is in no small part attributable to the efforts of many statistical officers in pursuit of excellence, and none more so than Ken Foreman. Foreman led the widespread introduction of mathematical statistical methods, especially sample surveys, to the work of the Bureau.

As former Australian Statistician Bill McLennan described him: "Foreman was undoubtedly the father of statistical methods at the ABS. He was the leader of the first methodology unit and continued to foster, expand and lead it until his retirement from the ABS in 1984".

Foreman introduced the Bureau to formal sampling methodology. This included formalising the techniques used for current samples and standardising those to be employed for future sampling work.

He developed comprehensive training for the benefit of those working in the area, as the level of knowledge and

practical ability required for this type of work was not yet obtainable from universities in Australia.

In *Informing a Nation*, a young Ken Foreman is pictured during field work in Papua New Guinea. Wearing a well-worn khaki shirt and broad-rimmed hat, he resembles an Indiana Jones-like figure. Indeed, he saw some real-life adventure. Foreman joined the Royal Australian Air Force in 1943 and became a navigator in a bomber. He was badly injured when he was shot down in a Lancaster in 1943 and had to endure a long convalescence.

He returned to Australia, and graduated with a Bachelor of Economics with Honours at the University of Sydney, but by mid-1952 he went overseas to study in the US Bureau of the Census under Morris Hansen and Bill Hurwitz, two of the most famous names in sampling at the time. He also had the opportunity to meet W. Edwards Deming who provided the inspiration for his work on quality control.

Despite his wartime injuries, Foreman travelled overseas several times to study the latest developments in sampling in the United States. In this way the Bureau's sampling ability remained abreast of developments in theory and technique being made overseas.

As Foreman himself said in 1977: "The application of mathematical statistical techniques to an increasing range of the Bureau's operations has resulted in many notable innovations which, in total, have significantly altered the Bureau's capabilities and economics".

Today Foreman is remembered in the Bureau with respect, awe, gratitude and fondness. He raised the level of methodological competence in the Bureau, and was a key player in four important developments:

- implementation of significant sampling techniques;
- design and implementation of the household survey system;
- pioneering work in development and application of seasonal adjustment; and
- the introduction of data management.

He was also a significant, caring and generous mentor for several generations of young statistical officers in the Bureau.

"He was a legend, he had some god-like qualities", Trewin said.

One of the earliest examples of the Bureau's contribution to international statistical methods was the publication *The Mathematical Theory of Population*. It was released in 1917 as an appendix to the Statistician's Report for the 1911 Census.

"It comprised 450 pages, so it was some appendix", Trewin said.

This publication had a twofold purpose of attempting to establish suitable methods for analysing census and demographic data, and actually using those techniques to analyse the 1911 census results.

Trewin said that elements of this work are still used in demographic work in Australia and elsewhere. "This really is a book that has aged very well. It has stood the test of time", he said.

"Even from those early days, the ABS was interested in methodology ... even though we were a small young country, our statistical services were regarded as world class."

The ABS was involved in a world first when Roland Wilson, then second in charge at the Bureau, developed his balance of payments statistics. Based on his PhD work, it was released as a chapter in the 1933 Year Book.

Developments during the Foreman era included:

1952 – sample to develop income tax statistics (led to creation of methodology unit)

1960s – development of quarterly labour force survey

1969 – introduction of generalised survey systems

1972 – introduction of social surveys
1977 – use of post-enumeration surveys to adjust for the Census undercount.

Trewin pointed out that Australia had pioneered the latter exercise. "The US still has not got around to adjusting their Census undercount", he said.

"I think we were the first country to do it way back in 1977 ... we just thought it

was common sense. There has been no criticism of it since then.”

“But it does mean that the way population statistics are used in Australia is much better than in many other countries. Subsequently, countries such as Canada have copied what we’ve done.”

The key development in the 1980s was the introduction of synchronised sampling (variant of collocated sampling to business surveys).

Trewin said that the reason the ABS had been successful with its methodological developments was that there had been a strong alignment between the ABS’s statistical and methodological areas.

“Ken Foreman was very keen that methodological people interact with the other business areas so that they could be seen and heard; and that they weren’t working in a vacuum – they could actually understand the context of the work they had been doing”, he said.

“The other sense of alignment is that the other business areas ... had a working knowledge of methodology ... so they could have meaningful communications with the methodologists.

“In some statistical offices, that does not happen. The methodologists are poked away in the corner to do their own thing and have a much smaller influence on the work of the office, than what our methodologists have.”

He said that support for methodology by senior ABS managers has been strong. “The past two Australian Statisticians had a methodological background, so it [that support] was probably more straightforward then, but it actually happened right back from Keith Archer’s days in the 1950s and 1960s”, Trewin said.

“There was from the senior ABS management, support for methodology development. They [methodologists] weren’t regarded with suspicion by senior managers, they were regarded as important.”

Turning to the future, Trewin said that as a statistical producer, the Bureau’s methodological challenges included the following:

- Emerging statistical outputs (e.g. environmental statistics, longitudinal analysis) requiring the application of novel sampling methods
- Combining survey and administrative data
- Extracting information from large data bases (e.g. EFTPOS, scanner data bases, tax data bases)
- Use of analytical- or model-based methods to produce official statistics
- Data linking will be increasingly used to provide official statistics
- Confidentiality and microdata access introducing new confidentiality challenges
- Internet and web-based surveys – a number of new challenges in both survey design and form design
- Small area statistics – satisfying the increased data
- International surveys – ensuring the data are really comparable.

He said that the other half of the ABS charter is as a statistical leader. This role would include producing manuals of good practice; training in statistical methods; and providing advisory services.

Despite such a proud history of achievements over a century, the ABS will not rest on reputation. And that includes ensuring the calibre of its staff remains world class. Trewin said he was concerned about the supply of mathematical statistics graduates.

“We feel the supply is dwindling at the same time as the demand is increasing”, he said.

“The ABS has taken a strong interest in seeing what can be done. We want to try and get a more uniform approach to teaching of statistics at schools, so you have a framework for developing resource materials and a framework for the training of teachers of statistics.”

With such an enviable and sustained record of quality, and a focus on the future, another generation should celebrate the ABS’s second century with pride.

This article originally appeared in AMSTAT News.

Society Secretaries

Central Council

President: Prof K. Basford
Secretary: Dr D. Shaw
Email: doug.shaw@csiro.au

New South Wales

President: Ms Caro Badcock
Secretary: Eric Beh
Email: e.beh@uws.edu.au

Canberra

President: Dr Brent Henderson
Secretary: Dr Ray Lindsay
Email: ray.lindsay@netspeed.com.au

Victoria

President: Dr Brian Phillips
Secretary: Dr Ann Maharaj
Email: ann.maharaj@buseco.monash.edu.au

South Australia

President: Ms Margaret Swincer
Secretary: Ms Janine Jones
Email: janine-jones@adelaide.edu.au

Western Australia

President: Dr Brenton Clarke
Secretary: Mr Martin Firth
Email: martin@maths.uwa.edu.au

Queensland

President: Dr Ross Darnell
Secretary: Dr Helen Johnson
Email: h.johnson@qut.edu.au

Section Chairs

Statistics in the Medical Sciences

Mr Peter Howley
stph@hsrg.newcastle.edu.au

Statistics in the Biological Sciences

Dr Simon Barry
Email: simon.barry@brs.gov.au

Survey and Management

Dr Robert Clark
Email: rclark@uow.edu.au

Statistical Education

Dr Michael Martin
michael.martin@anu.edu.au

Statistical Computing

Associate Professor Kuldeep Kumar
Email: kkumar@bond.edu.au

Industrial Statistics

Mr Ross McVinish
Email: r.mcvinish@qut.edu.au

Young Statisticians

Ms Janice Wooton
Email: janice.wooton@abs.gov.au

Bayesian

Professor Kerrie Mengersen
Email: k.mengersen@qut.edu.au

Further contact details for Society Secretaries and Section Chairs can be obtained by contacting the Society on (02) 6249 8266

Branch Reports

SOUTH AUSTRALIA

Young Statisticians Dinner (SA)

The first ever Young Statisticians event in SA was held in March this year. The fantastic turnout exceeded all expectations, with more than 20 people attending. The evening kicked off with pre-dinner drinks at a local pub, which gave everyone an opportunity to introduce themselves to new faces and catch up with old friends in a relaxed environment. From there, people proceeded to Café Buongiorno for dinner. The SA branch of the SSAI kindly sponsored the evening, with a variety of pizza provided for everyone to share. This idea worked well and contributed to the casual atmosphere of the evening.

Everyone who attended the dinner was asked to complete a short survey, prepared by Penny Bennett, to assist with planning for future Young Statisticians activities in SA. All suggestions were greatly appreciated and the difficulty now is deciding which suggestion should be taken up first!

The whole evening was a great success, with many people commenting on how much they enjoyed themselves. It gave everyone an opportunity to find out where other Young Statisticians in SA are working or studying, the types of methodology being used and some of the statistical problems that people are faced with. As a result of the dinner,

a current honours student in Statistics is now receiving informal support from one of the more senior Young Statisticians with similar interests. This is just one example of the importance of networking, particularly for Young Statisticians. It is hoped that future events for Young Statisticians will result in more collaborations such as this and ultimately contribute to the future of Statistics in Australia.

SA AGM

At the SA Branch Annual General meeting A/Prof Peter Baghurst, epidemiologist from the Public Health Research Unit of the Women's and Children's Hospital in North Adelaide gave an interesting talk about sensitive methods for monitoring the outcomes of health-care.

Some events such as the Bristol Paediatric Cardiac Surgery scandal, and the notorious Dr Harold Shipman, have given impetus to refining statistical techniques to take case-mix into account; and developing control charts to enable shifts in the incidence of adverse events to be detected much more rapidly than ever before.

Peter introduced a monitoring tool called the risk-adjusted cumulative sum or 'cusum' which has been developed particularly for detecting adverse events over time. The cusum is essentially a score over time which begins at zero then accumulates over time by adding

a weight which is calculated differently depending on whether an adverse event has occurred or not. Thus the weight will be negative for a successful outcome and positive (seen as a penalty) for a failure or adverse outcome. The weight forms a log-likelihood ratio which is based on individual's prior risk (explaining the 'risk-adjusted' terminology) which is generally estimated by logistic regression. The log-likelihood ratio is then used for testing whether the observations on consecutive patients are more supportive of a hypothesis of adverse events occurring at the accepted benchmark rate – or at some (pre-specified) higher rate.

Peter went on to discuss the best choice for setting control limits for the cusum given that the traditional alpha levels of 5% are not appropriate for sequential testing. He introduced the Average Run Length (ARL) which can be considered analogous to the type 1 error rate e.g. when the process is in the H_0 state the ARL should be long before a 'signal' occurs (crosses the control limit) and is likely to represent a false alarm and if the process has shifted to the alternative state it should have short run lengths before a 'signal' occurs. Therefore, a control limit far from zero will lead to longer ARLs and what is considered long will depend on the circumstance. The ARL can be approximated by Markov Chain methodology and is calculated based on a pre-specified alternative rate. Peter referenced a paper by Steiner (2000).

Various examples were shown throughout the talk. One example demonstrated the cusum by using fourteen anonymous Paediatric Intensive Care Units across Australia for the adverse event of death after forty-five minutes of admission in PICU. This clearly demonstrated the importance of risk-adjustment given different PICU's had different exposures of risk. The talk was then wrapped up with interesting discussion of the cusum used in practice.

Incorporating LASSO Effects into a Mixed Model for QTL Detection

The speaker for April was Scott Foster, a current PhD student at Biometrics SA. He began his talk by explaining what QTL (quantitative trait loci) are and why they are important. QTL are positions or regions of the genome that are statistically related to a quantitative trait. Usually this trait is of economic importance, such as accelerating breeding in agriculture. QTL



L to R: Amy Glen, Paul Eckermann, Jono Tuke and Sam Cohen at the Young Statisticians Dinner in SA.

can be used to narrow the search for causative genes.

In QTL experiments, populations are created which are genetically diverse. Genetic differences are assessed for given positions of the genome. It is then possible to look for differences in traits based on differences in loci and hence identify QTL. Scott described a motivating data set, where the aim of the experiment was to identify genes for traits of economic importance related to beef production, in particular the birth weight of cattle. The statistical problem arises when trying to identify the genes which are associated with birth weight, while controlling for experimental effects, such as gender and location.

The trait of interest can be modelled as a linear combination of genetic and experimental effects. However, there are many explanatory variables which may be included in the model, creating a model selection problem. The LASSO (least absolute selection and shrinkage operator) may be used to produce estimates of genetic effects from a random model where the random effects are assumed to come from a double exponential distribution so some estimates will be 'shrunk' to zero. The level of shrinkage depends on the dispersion parameter. In order to include fixed and normally distributed random experimental effects in the model, the LASSO effects can be incorporated into a standard linear mixed effects model. This type of model performs well in terms of identifying QTL while maintaining an acceptable number of false positive results, relative to other currently available methods.

After the talk, a farewell dinner was held for Scott at a Thai restaurant. He is moving to Tasmania to take up a position with the Tasmanian Institute of Agricultural Research. Scott has been a valued member of the SA branch of the Statistical Society for a number of years and his presence at meetings will be greatly missed. We wish him all the best for the move and the new job.

Lisa Yelland

WESTERN AUSTRALIA

Statistical Analysis at Western Power

Ross Bowden, a senior analyst at the electricity retailing company Synergy (previously known as the Retail division of Western Power) presented at the

March meeting of the WA branch of the SSAI. His talk was entitled "Statistical Analysis at Western Power", and covered some of the statistical techniques he has used at Western Power. Commercial confidentiality prevented Ross from going into too much detail, but he was able to present some of the methodologies he has used, with five different projects as examples.

The first example involved was calculating the unserved energy estimates from power system simulators. This estimation is important when scheduling maintenance of generators to ensure that there is sufficient generation capacity to meet demand if any of the generators breaks down (a forced outage). This estimation also allows for the prediction of future plant investment. Ross covered two different methods that were used to estimate this unserved energy – firstly, a method of cumulants, using a fixed

annual load profile and set probabilities of outages for each generator. This method was used until advances in computing power allowed for Monte Carlo simulation to simulate the outages of each generator.

The second example of Ross's work was the weather correction of energy and peak load data. Weather correction of peak loads is especially useful, as the peak loads are strongly dependent on the weather, and it can be difficult to determine any long-term trends in the raw peak load data. Two different methods are used for weather-correcting the peak load values. The first uses nominal distribution theory, comparing 40 potential peak loads to the median of the yearly peak load. Similar to the unserved energy methodologies, the second method of bootstrapping only became feasible with the advancement of computing technology. Ross described the bootstrapping of previous year's weather

Thinking Statistically

Elephants Go to School

A UNIQUE TEXTBOOK

By

Sarjinder Singh

St. Cloud State University
Department of Statistics
St. Cloud, MN 56301-4498
U.S.A.

Forewords by
David Robinson
and
Stephen Horn

A new way to learn statistics using pictures, jokes, and tales.

A lot of learning with fun through 676 pages.

Good for all ages

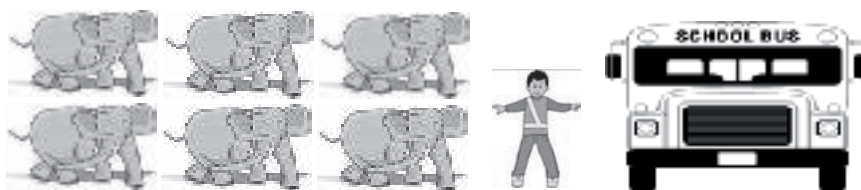
+ Good for all libraries

+ Good for all majors

+ Good for all schools

+ Good for you too

PLEASE HAVE A LOOK



Kendall/Hunt Publishing Company 4050

Westmark Drive, P.O Box 1840 Dubuque

Iowa 52004-1840, U.S.A.

www.kendallhunt.com

results to obtain a distribution of possible peak load values for the day, week or month of interest. From this distribution of peak loads, a central estimate could then be calculated.

Ross's third example of his presentation concerned queuing models and simulation. This work was performed on data from the Call Centre at Western Power relating staffing provisions, expected number of calls and estimated waiting times. Ross described the problems with this simulation method, stating that the method did not reproduce reality.

In the fourth example, Ross covered the use of graphical representations of survey data. Ross explained that he often has to explain statistical results from survey data to staff without a statistical background, and found that describing results graphically was an effective way of explaining his point. He covered such graphical tools as bands on graphics to quickly show differences between categorical variables. Another graphical technique Ross covered was the use of bootstrapping with perceptual maps and correspondence analysis, to represent two-dimensional projections as clouds of boot-strapped results.

The final example that Ross covered was the method to determine the least-cost meter reading cycle, involving a balance between billing, collections, reading and credit costs. Ross explained that this was a typical analysis problem at Western Power, with more emphasis on mathematics than statistics.

Ross finished the presentation with a quick overview of analysis at Western Power. He explained that while most managers are statistically challenged, they do respect sophisticated analysis. For analysts, while analytical skills are very important, so are communication skills. Ross concluded his presentation in his usual manner, with a quick joke about punk statisticians.

A number of questions were raised when the discussion was thrown open to the meeting and an informal discussion continued immediately after the meeting and at the usual post-meeting dinner at a nearby restaurant.

Christopher Milne

NEW SOUTH WALES

Understanding the Illicit Drugs Market (or Learning with Independent Components Analysis) by Stuart Gilmour and Inge Koch.

In the May 2005 meeting of the NSW Branch, Stuart Gilmour from the National Drug and Alcohol Research Centre and Inge Koch from the University of New South Wales presented interesting research using independent components analysis to model the illicit drugs market.

Illicit drugs such as heroin, cocaine and amphetamines are a significant health problem for some young Australians. Globally, researchers have spent considerable energy trying to understand the size of user populations, the spread of these drugs through communities and the effect of changes in the markets. Gilmour and Koch were interested in answering questions such as whether or not drug markets undergo epidemic patterns and whether the markets can be divided into isolated, competing or cooperating partitions. Previous work has used time series analysis to investigate the consequences of the heroin shortage, for example.

The data presented comprised 17 indicator data series collected over 66 months between 1997 and 2002. The series included indicators such as drug use or possession offences, prostitution offences, ambulance attendance to a heroin overdose and break and enter dwellings.

Unlike the time series framework, Gilmour and Koch treated the data as 17 samples with 66 dimensions each. Analysis of such HDLSS (high dimension low sample size) problems requires dimension reduction. Gilmour and Koch proposed a 3 step learning approach (END) consisting of an Extraction step which reduces the dimension, a Non-normalising step, which finds the most non-Gaussian directions in the lower dimensional data, and a Decision step which separates the data in contrasting groups. Principal component analysis is used in the first step. Combined with a careful application of independent component analysis, this leads to a set of non-Gaussian direction vectors spanning the lower dimensional space. A simple

decision rule, based on properties of the independent components (IC) scores, divides the data into two contrasting groups.

Crucial to the success of the proposed learning method is the careful selection of the lower dimensional subspace and the number of IC directions required for the decision making process, as independent component analysis is a very powerful technique which finds non-Gaussian directions even in Gaussian data. They found that a reduction from 66 to 6 dimensions and a search for the 3 most non-Gaussian IC directions lead to a stable partitioning of the data into 2 contrasting and interpretable groups.

Gilmour and Koch noted some interesting results from their two partitions. Partition 1 contained heroin and cocaine possession offences, series relating to overdose and psychoses and serious offences such as robbery with a weapon. These represented direct measures of the illicit drug market. On the other hand partition 2 appeared to represent indirect measures of the illicit drug market. Possession offences for amphetamines appeared in partition 2 and were thought to represent the mixed nature of the amphetamine market, which is only partly connected to street-based illicit drug markets, and partly to a more functional market connected to the party drug scene.

In conclusion Gilmour and Koch found that independent component analysis provides a powerful tool for unsupervised learning with non-Gaussian data. They also found that the method was sensitive to parameter settings and that interpretation of the results depended on good model selection. ICA was able to identify a key subset of series that measured the direct effect of illicit drug markets. It was also able to identify sub-populations of illicit drug users such as amphetamine markets with serious health and social consequences from the "party drug" amphetamine markets.

Petra Graham

November 2005 JB Douglas Postgraduate Awards

November 23rd 2005 saw the NSW Branch hold its 6th annual JB Douglas Postgraduate Awards day and annual dinner. Once again the quality of the



Presenters and special guests for the annual postgraduate awards day. Back row (from left): Daniel Smith, Alex Woolaston, Frank Tuyl and Nathan Pearce. Front row: Maree O'Sullivan, Maureen Morris, Christine Curran, Alun Pope and Jim Douglas.

NSW Branch Annual Dinner: 100 years of official statistics – a review of the development of statistical methods over this time.

After the presentation of the JB Douglas Postgraduate awards the NSW Branch held its annual dinner with guest speaker Dennis Trewin, head of the Australian Bureau of Statistics.

Dennis shared with us the fascinating history of official statistics in Australia which began in 1905 with formation of the Commonwealth Bureau of Census and Statistics. Pre World War II developments included the release of the *Mathematical Theory of Population* (a 450 page appendix to the 1911 Year Book) by George Knibbs, the development of Balance of Payments by Roland Wilson, the creation of research officer positions and a development branch during the 1930's.

Following the Second World War the Bureau entered what is known as the Foreman Era – a period of substantial development of statistical methods for the official data collection. Under the leadership of Ken Foreman, the methodology unit introduced and developed a variety of survey systems including those for quarterly surveys of retail establishments, income tax statistics and social surveys. Other important developments during the Foreman Era included the introduction of statistical quality control techniques for the census, the introduction of seasonal adjustment and experimental design for survey development and a report detailing the need for a data warehousing scheme.

students from around the state was very high and the judges (Dr Alun Pope, Mr Dennis Trewin and Ms Melissa Cassar) were given a difficult job of choosing the best presentation.

The day began with Maree O'Sullivan (Macquarie University, Statistics) presenting her work on whether differing microarray platforms produced concordant results using data on paediatric acute lymphoblastic leukaemia. Next Alex Woolaston, from the University of New England (Mathematics, Statistics and Computer Science) described methods for eliminating the spatial trends found on cDNA microarray slides and Nathan Pearce from the University of New South Wales (Statistics) introduced novel algorithms for training and cross-validating support vector machines.

After a quick afternoon tea Daniel Smith from the University of Sydney (Econometrics and Business Statistics) presented his research comparing binary Markov random fields via applications in image restoration followed by Frank Tuyl from the University of Newcastle (Statistics) who talked about the caution needed in choosing parameters for the Beta distribution in Bayesian intervals. Maureen Morris from the University of Wollongong (Statistics) completed the presentations describing her research in the development of effective assessment

tasks for courses in introductory statistics.

After some deliberation the judges awarded Daniel Smith the Peter Wright prize for best student presentation with Nathan Pearce and Maree O'Sullivan winning joint second prize. We thank all who attended and our generous sponsors CSIRO, Macquarie University, Roche Pharmaceuticals, SAS Australia, the University of Sydney school of Mathematics and Statistics, the University of Newcastle and the Wollongong Statisticians without whom this event would not have been possible.



Branch members enjoying festivities at the annual dinner.

The mid-1970's saw the Bureau become a statutory agency independent of other government departments and known as the Australian Bureau of Statistics. Key statistical developments of the post Foreman Era included using total survey design that takes account of non-sampling errors as well as sampling errors, forms design, the integration of survey designs to improve cohesion of economic statistics and the further development of methods for time series analysis.

Dennis reflected on how important it was for the Bureau to invest in methodology and that the strong alignment between statistical and methodological areas resulted in innovations that "... significantly altered the Bureau's capabilities and efficiencies".

Finally Dennis described a variety of future challenges and directions for the Australian Bureau of Statistics including the requirement of more graduates in mathematical statistics since current demand outstrips supply! Statistical challenges include the use of analytical or model based methods for producing official statistics, the use for data linkage over time and the combination of survey and administrative data to support official statistics, and the potential use of web-based surveys. As the leader of the National Statistical system the Bureau aims to produce manuals of best practice and standards as well as provide statistical training and advisory services. This recognises that other agencies are going to increasingly be producing statistical outputs for public use.

After thanking Dennis Trewin for his interesting historical perspectives on the Australian Bureau of Statistics, branch members and guests celebrated the end of another year at the annual dinner.

Petra Graham

Professor Anthony Edwards

The NSW branch started the year by looking backwards, at the work and influence of RA Fisher. Professor Anthony Edwards was Fisher's last undergraduate and was attached to the Statistical Laboratory at the University of Cambridge from 1968-70. Anthony not only talked about Fisher's work but also gave us a rare glimpse into the man himself.

Fisher was born in 1890 at East Finchley and moved to Adelaide in 1959 where he remained until his death following

an operation in July 1962. Suffering from extreme myopia he had very poor eyesight, which may have contributed to his exceptional geometric ability and capacity to solve complex mathematical problems entirely within his mind's eye. We heard a story about a lecture when Fisher was asked a hairy problem by one of his students. Rather than fluff an answer using academic double speak, Fisher stopped talking and pondered the question at the front of the lecture hall in complete silence, for 10 minutes! Until he at last said 'yes, yes I think you're right' and continued with his lecture.

Fisher was interested in a diverse range of subjects that included statistical theory, inference and methods, experimental design, evolution and genetics. He was also quite prolific and published several hundred papers and seven books, which included the influential 'Statistical Methods for Research Workers' original published in 1925. This book has been republished at least 14 times and in 5 different languages (French, German, Italian, Japanese and Spanish). Anthony highly recommended it as a good read that introduces the underlying principles of experimental design with a minimal of mathematics.

The importance of this book cannot be understated because it is here that Fisher described his new paradigm for analyzing data. In it Fisher introduced the concepts of randomization, replication, blocking and ANOVA to a mass audience. Concepts that are considered fundamental in modern statistical analysis, but at the time were quite revolutionary.

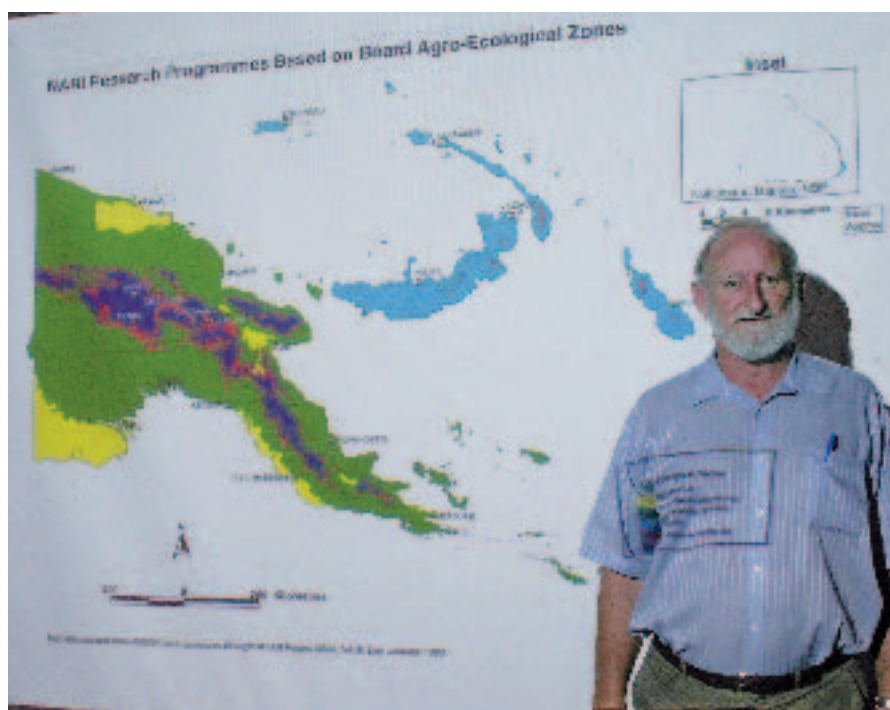
Chris Howden

QUEENSLAND

February Branch meeting

On 7th February Bob Mayer, Principal Biometrician, Department of Primary Industries and Fisheries, Townsville spoke on 'Developing biometrical expertise for PNG agricultural research' at the IMB Biosciences precinct at the University campus in St Lucia.

Bob has worked with the DPI&F biometry unit since graduation, in Brisbane (briefly), Toowoomba and Townsville. Since 1999 a substantial part of his role has been as a member of an AusAID supported overseas aid project team helping develop agricultural research expertise in PNG. This has included reviewing the existing statistical support for agricultural research, several visits to PNG agricultural research stations,



Bob Mayer with a map of the agro-ecological zones of Papua New Guinea.

presenting courses on basic statistics and use of statistical software, guiding the professional development of staff to take on the biometrician role, and statistical consulting to research projects. The talk covered various aspects of his experiences on this project, including background to it and some of the special and on-going challenges associated with it.

Members and guests had a curry at a nearby restaurant after the meeting.

March Branch meeting

The AGM was held on the 21st March at the University of Queensland, and the following officer bearers were nominated and appointed.

President Dr Ross Darnell
Secretary Dr Helen Johnson
Treasurer Dr Melissa Dobbie

COUNCILLORS:

Dr Peter Baker
Prof. John A. Eccleston
(Immediate past president)
Mr James McBroom
Dr Michele Haynes
Dr Miranda Mortlock
Dr Tony Swain
Dr Ian Wood

After the meeting Dr Michael Bulmer, Lecturer, School of Physical Sciences, University of Queensland spoke on 'Teaching Spaces for Learning Statistics in Context', and the talk was held in the new lecture space about which he was talking. Dr Michael Bulmer is a Senior Lecturer at the University of Queensland. He has recently been given an award for excellence in teaching at the university. His main research interests are in symbolic computation, optimisation methods for statistics and operations research and mathematics education.

In 2005 the University of Queensland opened a new "Collaborative Learning Centre". A large statistics service course was taught in this space recently as part of a trial for this Collaborative Learning Centre. Michael treated us to a hands-on session in which we got to do one of the statistical activities that he had developed for this Centre. SSAI members were given tape measures, stop watches and three small furry animals with which to measure the acceleration due to gravity in groups of three. He discussed the pros and cons of this teaching space, and discussed some of the research results from his study of its effectiveness.

Miranda Mortlock

CANBERRA

Talk on bioinformatics by Geoff McLachlan

At the monthly meeting of the Canberra Branch of the SSAI on 21 February 2006, Professor Geoff McLachlan of the Department of Mathematics and Institute for Molecular Bioscience at the University of Queensland gave a talk titled "Some Applications of Statistics in Bioinformatics". Geoff has authored a long list of statistical papers and five books, the most recent one being "Analyzing Microarray Gene Expression Data" (2004; Hoboken, New Jersey: Wiley; co-authored with K.-A. Do and C. Ambrose). He is also a chief investigator in the ARC Centre in Bioinformatics.

Geoff began by describing DNA microarray technology, which was originally conceived to detect expressions of thousands of genes simultaneously. Microarrays present special problems for statistics because the data are very highly dimensional with very little replication, and the challenge is to extract useful information such as gene functions, gene interactions, regulatory pathways, metabolic pathways, etc. That challenge has been increasingly taken up in recent years, and Geoff presented a graph showing how the number of papers published on microarrays skyrocketed from virtually none in 1997 to well over 4000 in 2003. The graph also showed how the proportion of those papers that are statistical grew dramatically to about 90% in 2003.

The output of gene-expression levels from a microarray experiment can be represented by a matrix, with rows representing genes and columns representing samples. Attention was focussed on the case where the tissue samples can be divided into two groups corresponding to two different populations of cells, and the aim is to compare the gene expressions for those two populations. However, separate hypothesis tests conducted to this end may be complicated by the potentially large number of false positives. For example, if there are 10,000 genes and the significance level is 5% then under the null hypothesis of no difference one would expect about 500 false positives.



Geoff McLachlan

There are ways of dealing with this problem which are less restrictive than using the Bonferroni correction, one being the Benjamini-Hochberg procedure which controls the false discovery rate (FDR) using an ordering of the many observed p-values. Another approach involves minimizing the estimated Bayes' risk in the context of a two-component mixture model where the mixing proportions are the probabilities that the genes are, or are not, differentially expressed. With this approach, a local FDR can be provided for each gene. The problem of simultaneous significance testing, especially in the context of microarray data, has recently received a lot of attention from several prominent statistical researchers, including Professor Bradley Efron.

One of the datasets which Geoff used to illustrate the two-component mixture model comprised measurements on about 3,000 genes in breast cancer tissues from women who were carriers of the hereditary BRCA1 or BRCA2 gene mutations, with the aim of finding genes which are differentially expressed between the two different tumour types. Geoff also discussed an application of the theory to HIV data, and briefly discussed a second problem in bioinformatics: clustering of correlated gene profiles, with an application to the yeast cell cycle.

Borek Puza

Talk on deep vein thrombosis by Niels Becker

Following the AGM of the Canberra Branch of the SSAI on 28 March 2006,

Professor Niels Becker of the National Centre for Epidemiology and Population Health (NCEPH) at the Australian National University (ANU) gave a talk titled "Analysis of a case-study, with application to air travel and the risk of DVT". The talk focussed on work done jointly by Niels, Chris Kelman and Agus Salim, as published recently in the British Medical Journal, Biostatistics and the Australian and New Zealand Journal of Public Health. A sumptuous Indian feast concluded the evening's program whose venue was the Faculty Suite and Seminar Room of the recently completed HW Arndt Building at the ANU.

Niels began by describing deep vein thrombosis (DVT) and noting that it is an illness which – whether due to natural causes or due to air travel – has a very small risk of occurrence. For that reason it is difficult to obtain relevant data by following a cohort over time, and one must sample cases disproportionately just to get enough of them. Niels then discussed case-only studies, including case-crossover studies, and their wide range of applications. Difficulties with case-crossover studies include choosing a control period, possible bias if there is a trend in exposure over time, and a potential wastage of information.

One of the studies on which Niels has worked involves all cases of VTE (venous thromboembolism; which includes DVT) in the Western Australia hospital system from 1981 to 1999. During this period there were 16,205 admissions by 13,184 distinct patients. These records were linked to international travel data from DIMIA (the Department of Immigration and Multicultural and Indigenous Affairs) in the form of 2,217 cases of VTE with a flying history among holders of an Australian passport and 2,979 cases with a flying history with non-Australian passports. The resulting dataset was used to address the question of how the outcome variable – defined as the time from the start of the observation period until hospitalisation due to VTE – is related to exposure to international flight, to calendar time, and to personal characteristics such as sex and age.

Several models were fitted to the data and it was found, for example, that sex and age are not significant predictors of the outcome variable. One of the main conclusions of the study is that on an international flight a passenger is about 30 times more likely to initiate a

VTE than on a flight-free day. This is a serious concern for frequent flyers, but the elevated risk for infrequent flyers is only slight. Ideally one would also like to estimate the relevant 'dose-response curve', but data on duration of flight are elusive.

Talk on estimating the distribution of hourly pay by Ray Chambers

At the monthly meeting of the Canberra Branch of the SSAI on 2 May 2006 Professor Ray Chambers of the Centre for Statistical and Survey Methodology at the University of Wollongong gave a talk titled "Estimation vs Imputation for the Distribution of Hourly Pay". The talk was a summary of research which Ray carried out whilst at the University of Southampton, where he was Director of the Southampton Statistical Sciences Research Institute from 1995 to 2005. Ray recently gave similar lectures on this topic in Texas and Brazil.

The focus of Ray's talk was estimation of the distribution of hourly pay rate, denoted Y , for employees in the United Kingdom. This problem is of interest because: employment law in the UK dictates that all employees must be paid an hourly wage greater than or equal to a set minimum value; because this 'minimum wage' is subject to change over time due to inflation; because the government must be able to assess the impact of proposed changes to the minimal wage on the national wage bill; and, in turn, because such assessment requires knowledge of the distribution of wage rates.

Unfortunately, Y cannot usually be obtained for all sampled employees, because many are not paid hourly. In such cases it is possible to calculate a derived hourly rate, denoted X , based on total earnings and hours worked, which are available for all employees. However, X and Y are typically not the same, even when both are available. There are at least three options for dealing with missing Y values. The first option is to ignore those values, meaning to base inference only on the data for which Y is available. This leads to what Ray calls the 'response-based estimator' of the distribution function. The second option is to take the missing Y values as exactly equal to the available X values so as to construct the 'substitution estimator' of this function. And the third option is to construct an 'imputation estimator',

which involves replacing each missing Y value by a random draw from the set of Y values having the same X value.

After dismissing the first two options as inadequate (for example, the first leads to highly biased predictions), Ray detailed the imputation option, some variants of it, such as nearest neighbour imputation (NNI), and multiple imputation (which is better than single imputation). He then discussed the strong links between imputation and the technique of 'weighting'; and this, in turn, led him to consider a fourth option involving estimation as an alternative to imputation.

There are at least two ways to estimate an unknown population distribution function: the model-based prediction approach, as in Chambers and Dunstan (1986); and the model-assisted calibration approach, as recently suggested by Harms and Duchesne (2004). After describing these two approaches, Ray produced results of a Monte Carlo study of various estimators, using data from the 2002 UK New Earnings Survey (NES), as provided by the UK Office of National Statistics (ONS). In terms of bias, Ray found the estimator based on NNI to be the best amongst those considered. However, this approach was quite inefficient from a mean squared error (MSE) perspective, with a nonparametric version of the model-based prediction approach producing the best MSE performance. This finding was under the fairly reasonable assumption that the missing Y values are missing at random (MAR). Ray concluded by discussing variance and confidence interval estimation. This is difficult in the very unbalanced situations explored in the simulation study, and Ray recommended it as a worthy topic for further research.

Borek Puza

MASA

**Model Assisted Statistics
and Applications**
An International Journal
ISSN: 1574-1699

www.iospress.com
Editor-in-Chief: Sarjinder Singh
Managing Editor: Stephen Horn
Treasurer: Sylvia R. Valdes