

Statistics opportunities

New Zealand statistics is strong in biosciences such as ecology, biology and genetics. This includes natural populations of animals such as birds; plants and ecologies; fisheries; crop production and animal breeding. It is also strong in graphical displays for statistical information and in earth sciences. In health New Zealand has led the world in aspects of epidemiology, the tracking of health and illness across populations. Survey sampling is strong in certain areas, such as some official statistics, and social issues such as gambling.

The field faces many opportunities. There is a world-wide shortage of survey statisticians, and computing technology is enabling new ways of seeing and animating data, such as animated population pyramids and maps of commuter flows. "The combination of mapping and dynamic graphs is taking off," says Professor Sharleen Forbes, of Victoria University. Statisticians are also challenged by huge increases in the amount of data available.

IMAgEs profiles a few of the many statistical developments and applications in Aotearoa.

Testing for cancer

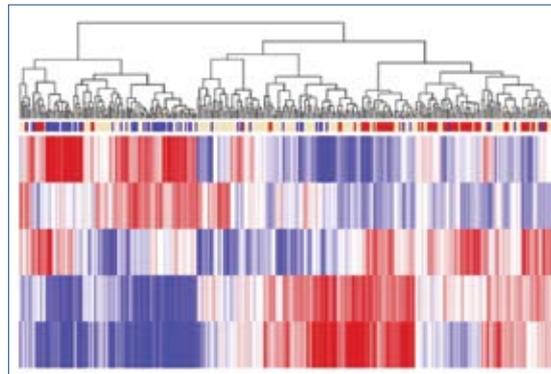
Dr Mik Black works with the Cancer Genetics Laboratory at the University of Otago, which is developing gene expression signatures – patterns of genetic activity – to predict outcomes for people with cancer. Scientists examine tumour samples after surgery, and Black uses standard statistical classification methods to predict whether the cancer is likely to come back or not.

"Three signatures indicate particularly aggressive cancer, and they have been patented by Pacific Edge Biotechnology," he says. Clinical trials are the next, very expensive, step to turn those signatures into diagnostic tests.

These take a long time; "we have to wait for five years to see if the cancer comes back. We started six years ago, patented the

signatures about four years ago, but there are a few years to go to get it into hospitals as a working test."

"Genetic statistics is most rewarding if we can work closely with clinicians, as they are the people actively caring for patients. Anything we can do that they can translate into improvements for patients - that's the real reward."



Identifying blocks of genes with highly correlated activity profiles in breast cancer using the PCOT2 methodology developed by Black's former PhD student Dr Sarah Song.

Census@School

Census@School had its beginnings when Professor Sharleen Forbes convened a group of NZ Statistics Association members in 1990 to run the first children's census in New Zealand schools.

"We ran it without any money, getting government departments to pay in kind with paper and printers – I wouldn't ever attempt it again! It was totally voluntary, on top of our day jobs."

"Students wrote about 60,000 reports and six or seven of us gave up our holidays to analyse and feed it back to schools. It was picked up by the Italians and then the Royal Society in the UK developed an internet version." This became New Zealand's first Census@School in 2003.



World Statistics Day

A free international poster competition for the International Statistical Literacy Project will be announced on World Statistics Day on 20 October. The theme is 'It happens in my neighbourhood' and it offers national and international prizes. It is open to groups of up to three students born in 1995 and younger, and those born in 1992 and younger.

The one-page poster must tell a story about a set of data. Schools can register at www.stat.auckland.ac.nz/~iase/islp/competition-second; registration closes on Friday, 17 December, and a maximum of two posters per age category per school must be submitted by 18 April, 2011. National winners will be announced on 31 May.



A public conference on Women in Statistics will be held on World Statistics Day at Victoria University of Wellington, with the aim of encouraging women's participation in statistics.

Statistics "has long been an area that girls and women are attracted to because it has a real-world orientation," says Professor Sharleen Forbes, of Victoria University.

One of the conference speakers, Associate Professor Megan Clark, says women make up around 60% of applied statisticians, and around 30% of theoretical and mathematical statisticians. Clark heads the School of Mathematics, Statistics and Operations Research at Victoria University.

The conference will be chaired by radio presenter Kim Hill, a member of the Minister's advisory committee on official statistics. Other speakers include Associate Professor Jennifer Brown, president of the NZ Statistics Association and head of the Department of Mathematics and Statistics at Canterbury University; Professor Natalie Jackson, director of the Population Studies Centre at Waikato University; and Rachel Milicich, Manager of National Accounts at Statistics New Zealand. Contact Lu.folau@stats.govt.nz





Analysing interactions in natural ecosystems

Statistics is very important in ecology for analysing multiple ecological variables at once, says Professor Marti Anderson of Massey University. "You might walk a transect in a forest, or swim a certain distance underwater and count the individuals of every species you encounter. Each species is a variable and species interact with each other and the environment. One of the challenges is to understand how sets of species change together, either naturally or in response to human-induced changes."

Anderson is pictured with Associate Professor Russell Millar getting ready for a fish biodiversity survey in Northland. She has developed new computer-intensive methods of multivariate analysis for biodiversity and community data.

"The biggest problem with ecological variables is they don't behave like normal Gaussian bell-shaped curves, and almost all classical statistical methods are based on that assumption."

She has worked on communities of Antarctic plankton and bacteria, organisms living in sediments in estuaries, butterfly communities in the tropics, marine fish communities in kelp forests, microalgae in freshwater streams, insects collected in pitfall traps, forest communities, and even suites of behavioural chemicals in birds.

Her software (PERMANOVA+) enables variation in these complex systems to be partitioned, allowing effects of a disturbance to be assessed against natural variation. It is being used around the world in many ecological applications and environmental impact assessments.

How many possums?

Environmental monitoring drove the development of a simple tool to calculate the population of possums in an area.

Associate Professor Jennifer Brown, of Canterbury University, says: "We spend a lot of money trying to control possums, rats and other pests. We want to know the level at which the population impacts on the environment, and whether we have been effective in managing them. If we have reduced the size of pest populations, are we seeing a gain in conservation?"



With Pest Control Research Ltd, her team developed a wax tag that possums can bite and from which they could calculate the surrounding population. "It was revolutionary, compared with the labour-intensive traps we used to use; the tag is now used throughout the country."

She has also been involved in designing survey protocols to find rare species in their environments. "You can waste a lot of time trying to find rare species," says Brown.

"The survey method seems simple but there is a lot of statistics behind it."

The method has been used to find desmans, a very rare river mole that lives in the Pyrenees in France, as well as in Southland to find invasive weeds before they start spreading.

See www.pestcontrolresearch.co.nz/research-monitoring.htm#3
www.pestcontrolresearch.co.nz/monitoring.htm
www.mathsreach.org/Videos

Photo: Malcolm Thomas, Pest Control Research.

R + L = ?

Not satisfied with a statistical package that has "revolutionised the practice of statistics", according to the Royal Society of New Zealand, Ross Ihaka, co-creator of R, is working on the next generation, with the working title L.



R is a free, open-source, extendable model with the highest hit-rate for mathematical publications in the last decade. It is available from more than 75 websites in more than 30 countries.

However, Ihaka (Ngati Kahungunu, Pakeha) says "the world is changing so fast that we desperately need something new now. Data volumes are exploding, and we have no idea as statisticians how to go about analysing petabytes [1,000 terabytes] of data."



His work on L is still theoretical – "you have to get the basics right otherwise you're constrained by your early decisions" – but shows promise of being thousands of times faster than R.

Assessing a vaccine

Roughly 200 cases of Meningococcal B were avoided by the MeNZB vaccine between 2004 to 2008, according to a statistical analysis of the vaccine's

effectiveness. Eighty percent of people under 20 were vaccinated, "a quite remarkable proportion, with the highest coverage in those under five," says Dr Richard Arnold, of Victoria University.

Working with epidemiologists in the Ministry of Health, he used a Poisson regression model to compare vaccinated and unvaccinated populations.

The bacterial infection is spread by airborne droplets and is associated with overcrowded households. Infection varies by age, deprivation and ethnicity, so he also controlled for those factors as well as regional, seasonal and yearly variations.

"The epidemic had peaked in 2001 and was on its way down naturally when the vaccine was introduced, but we found the vaccine was between 70% and 80% effective in avoiding the disease."

