Biostatistics Capacity Building at SACEMA

Wim Delva - Epidemiologist, SACEMA and Ghent University, Belgium

Gavin Hitchcock - Assistant Director: Training, at SACEMA

Biostatistics training has featured prominently in the Clinics on Meaningful Modelling of Epidemiological Data (MMED) organized annually by SACEMA in partnership with the African Institute for Mathematical Sciences (AIMS) and a consortium of North Americans. In addition, dedicated biostatistics short courses have been offered in June/July over the last three years, in conjunction with the Universities of Ghent (Belgium), KwaZulu-Natal and Cape Town. Very recently, from Monday 28 October to Friday 8 November 2013, SACEMA hosted two more specialised 5-day courses, end to end, in Stellenbosch, presented by acknowledged experts in the respective fields. We asked some participants to write brief reports of these well-received and timely courses.

Joint Modelling of Longitudinal and Survival Data

By Cari van Schalkwyk and Martin Nieuwoudt

What is "joint modelling" and why is it becoming increasingly important? Outcomes that are typically collected in longitudinal studies include repeated measures (such as biomarkers) and time-to-events (such as mortality). These outcomes are often analysed separately, ignoring the association between the two processes. However, if one wishes to account for the effect of the longitudinal outcome when analysing the event outcome, traditional survival analysis approaches are not applicable – these approaches require that time dependent covariates are independent of the event. If interest lies in the longitudinal outcome, bias might arise when ignoring non-random drop-out from the study. To obtain valid inferences for both, the joint distribution of the longitudinal and survival processes requires simultaneous analysis. Although the methods have been in circulation in the theoretical literature since the late 1990s, practical application of these computationally intensive models has only become feasible with progress in computing power.

The course was attended by 23 people from a number of countries and a wide range of institutions, and was presented by Associate Professor Dimitris Rizopoulos, who is based at the Erasmus University Medical Center, Rotterdam, the Netherlands, and published the first book on this topic in 2012.

This was an advanced course, aimed at participants with substantial statistical experience and it was masterfully presented, simplifying the theory and its implementation in R to a level that was manageable for all involved. We are grateful to Dimitris for his efforts and hope for a long and fruitful collaboration.

Updating Scientific Knowledge through Bayesian Learning

By Reshma Kassanjee and Hilmarié Brand

Presented with the same image of a black dog, the *Bayesian* statistician would infer the presence of a black dog, while a *frequentist* statistician would infer the presence of a dog that is black on the side visible in the image.

This is how Prof. Emmanuel Lesaffre (Erasmus University Medical Center, the Netherlands; Catholic University of Leuven, Belgium; Hasselt University, Belgium) contrasted the two schools of thought on statistical inference as he opened his course on *Bayesian Biostatistics*, based on his book with the same name. The course was attended by 15 participants from around Africa – ranging from postgraduate students to department heads, primarily operating in academic institutions in the areas of mathematics or statistics.

A frequentist's view of the world is based on the plausibility of observed data *given a hypothesised underlying process*, where all information is contained in the data. Bayesians have stochastic beliefs about the underlying process, *given the observed data* and any *prior beliefs*, and the updating of these beliefs mimics the process of incremental scientific learning.

In the past, the training of statistics students has been dominated by frequentist inference. Even non-statisticians prioritise the understanding of related concepts (such as p-values, confidence intervals and power) as these occur throughout the biostatistical literature. Consequently, Bayesian models are often overlooked by data analysts. Not only is there a natural tendency to map familiar concepts onto new problems, but mathematical complexity has also hindered the widespread use of Bayesian methods. It is only recently, 200 years after the founding work of Thomas Bayes and Pierre-Simon Laplace, that the application of Bayesian approaches has been revolutionised by the development of algorithms that

sample 'posterior beliefs' when analytical solutions are intractable. Furthermore, software implementing these algorithms is now freely available (for example, WinBUGS, OpenBUGS, JAGS, R packages such as mcmc and R2WinBUGS). This highlights how the utility of theoretical constructs (which may have been previously overlooked) evolves with time as supporting tools and technologies grow. Resources available to analysts are ever-changing, and therefore continued learning is essential to ensure that methods most suited to answer research questions are used.

This course began with an introduction to the Bayesian paradigm, contrasting it to the frequentist one. Selecting prior beliefs, updating these through observed data, and summarising resulting posterior beliefs, were explored. The use of Markov Chain Monte Carlo methods, such as Gibbs sampling and Metropolis-Hastings algorithm, to fit simple to complex Bayesian models was demonstrated in R

and WinBUGS. Practical sessions, interspersed throughout the course, empowered participants to use these tools themselves.

All participants seemed to be frequentist by default on day one of the course, but there were many questioning their beliefs by its end. We hope to see Emmanuel back in the Cape before too long to make more converts.

Wim Delva, Epidemiologist, SACEMA and Ghent University, Belgium. Areas of interest: statistical analysis of sexual behaviour data, stochastic and deterministic modelling of sexual network dynamics and HIV transmission. *wim.delva@ugent.be*

Gavin Hitchcock, Assistant Director: Training, at SACEMA. Areas of interest: topology, history of mathematics, mathematics education, problem-solving, inter-disciplinary discourse and training. *aghitchcock@gmail.com*