

## **A/Prof. Robert Clark**

Director, Centre for Sample Survey Methodology  
National Institute for Applied Statistics Research Australia  
School of Mathematics and Applied Statistics  
University of Wollongong

### **Title:**

## **Generalised Linear Models Under Response Dependent Sampling**

### **Abstract:**

Generalised linear models are usually fitted to data on a non-informative sample from a finite or infinite population. However, sometimes the dependent variable may be known in advance of sample selection, for either a population or a very large sample. The independent variables are then collected from a smaller sample. This smaller sample can be selected using the dependent variable values for the population, leading to lower variances of the regression coefficients, but making the analysis more complicated. This situation can occur in biostatistics and ecology, whenever the outcome variable is cheap and easy to measure (e.g. blood pressure, koala abundance) but covariates are expensive (e.g. genome sequencing, diet, leaf nutritional content). Case-control studies are a special case. Three main approaches are reviewed: pseudo-likelihood, conditional likelihood and full likelihood. Efficient sample designs are explored, with some surprising results. Data on koala abundance and leaf nutrition are used to illustrate the methods.