Sandy Burden
National Insitute for Applied Statistics Research Australia (NIASRA)

Constraint Choice for Spatial Microsimulation

Spatial microsimulation models are increasingly being used to create realistic microdata for geographical areas. The availability of this microdata enables spatial statistical modelling of health, social and economic variables without compromising the confidentiality of the individual records. For a pre-defined set of areas, they combine sample records with benchmark data, typically by sampling, or re-weighting the sample records to fit the bench-marks obtained from a set of constraint variables for each area. The selection of constraint variables and benchmarks is a key factor in the simulation process as the simulated data will reflect the population structure defined by the benchmarks.

In this talk I introduce the use of a within-area homogeneity measure for selecting categorical constraint variables for microsimulation. In particular, I describe the *d*-statistic, which is equivalent to intra-area correlation for areas with equal population. It can be used to identify spatial dependence in the constraint variables, an important feature to reproduce when modelling spatial variation in data. I show how the *d*-statistic can be used to assess the statistical significance of the within-area homogeneity for a given set of constraints; and how it can assist in validating a spatial microsimulation model.