Monday September 30, 2013

Last time: E-M; Fellegi-Sunter with Logistic Regression

Today: Blocking; Capture-Recapture; Work on Projects

Next Time: Clustering Records; Colombia Homicide Record Linkage

Blocking:

Choosing Blocking variables:

Winkler Case Study:

What are the pros/cons of blocking? What are the consequences?

Finding Field Similarity Score:

Supervised Models:

Fellegi-Sunter/E-M:

Using Blocking to Identify Duplicate List Entries:

Winkler Case Study:

Capture-Recapture: estimating the population size when we don't observe a subgroup

Capture-Recapture with Matches:

Capture-Recapture, cont: