

Wednesday September 18th, 2013

Last time: Logistic Regression; Classification

Today: Intro to Fellegi-Sunter

Next Time: Continuing Fellegi-Sunter

Announcements:

- Syria HRDAG/Linkage visit, Friday Sept 20th, 12pm
- HW 3 due Friday 9/27
- Start thinking about project ideas

Agreement Patterns:

One-to-One Matching vs Deduplication in One List

Fellegi-Sunter Approach (1969)

Fellegi-Sunter, in pictures:

Simplifying via Conditional Independence:

m-, u- probabilities

How can we actually estimate these things if we don't know the matches and non-matches? More later.

Fellegi-Sunter, Exact Matching:

Using Frequency-Based Matching:

How could we estimate these m-, u-probabilities from the data?