## **Record Linkage Project: Report**

Reports due Friday October 18th 5pm to Prof Nugent's mailbox. Email me if there are any issues.

Projects will be generated on some aspect of record linkage. Undergraduate students will work in pairs, master's students on their own (hereafter called a "group"). Each group will present its project (pdf or ppt) on October 9th or 14th. Presentations should be 15 min. A report describing the project and its results is due by Friday Oct 18th, 5pm. Slides should also be turned in then.

- The report should adequately describe the project. There is no page limit or requirement; however, a reasonable writeup will probably be at least six pages.
- Sections should be:
  - *Title/Abstract:* Your abstract should contain around two sentences summarizing each of the below sections. Highlight any exciting results or and lack thereof. End with next steps.
  - *Introduction:* Your project idea and its motivation; what were you interested in exploring? Be explicit about your goals
  - *Methods:* The methods being used; give an overview, include notation, the models, etc. If you're doing EM, describe it here. If you're doing Fellegi-Sunter, logistic regression, k-means it all goes here.
  - Data: the data set you're using: where did it come from; did you generate it? etc
  - *Results:* Try to summarize your results in a table or tables. Define any performance measures you're using (ROC curve, false positive, etc). Don't assume the reader knows your methods. (You can assume a basic knowledge of statistics.) Highlight your interesting results. Why do you think you have them? What "caused" them?
  - *Discussion/Conclusions:* Critical evaluation of the work; what else would you have done? Is there anything else you wish you would have done/had? What worked, what didn't; what would you do next or tell someone to do next?
  - References/Acknowledgements: papers, resources, people, websites, etc

Depending on your specific project, the 3rd and 4th bullets might be switched. If you're looking at a specific application or a specific data set (i.e., if your goal is to analyze that data set), then the data set discussion can come first. If you're exploring the performance of specific methods, the methods part comes first.

This list looks similar to the presentation; however, in the report, you go into much more detail about the approach, data sets, etc. Describe your data. Give the formulas for your models and tell us what they do. What are the consequences for your assumptions? All the info that you can't give in a 15 minute presentation should have a home in the report.