555 Course Project

Jon Wakefield

January 15, 2016

Course Project

- *The Project:* A paper detailing a spatial analysis of student-chosen data. A preliminary version of the project will be handed in around week 6, and the final version at the end of the quarter.
  The project could take the form of an epidemiology style paper:
  - Introduction: Aims of analysis.
  - Data description: extensive details including sources of data (web addresses).
  - Methods used.
  - Results – not pages and pages of table and figures. Rough guideline: every table and figure in the main body should be discussed. Extra figures and tables can be included as supplementary material.
  - Discussion – Conclusions and limitations.

Course Project

- Usually 3 or 4 files will be submitted – one with the main report, a second with supplementary material, a file containing the data, and a file containing the code used to provide the results/analyses in the project.
- Get started early! I want a brief summary of your intended project in week 2.
- Let me know if you're struggling to find something suitable.
- I can give suggestions on applied projects, or particular methodological investigations.

Course Project

Some questions to ask:

- What is the denominator? In the simplest case this will be the populations at risk hopefully by key confounders such as age and gender.
- In a case-control study, how were the two sets of individuals sampled?
-
  - Is the data a random sample in space? Is some form of weighted analysis required? For example, survey data are often collected via a complex design, with design weights being needed to reflect the sampling scheme.
  - If one is interested in clustering then this usually won't be possible if the data are not gathered "as a simple random sample" (with respect to the population at risk).
  - Disease mapping and spatial regression will be possible, so long as one accounts for the sampling scheme.
- Is the GIS component sorted? i.e. can you draw maps? Early on you should find the appropriate polygon files (say) and read them into R.
- Can you carefully define the model?

Possible Topics

- Disease mapping: Examine SMRs and smoothed rates using data from local cancer registries in a state, at the county level. To take a step further examine the association between cancer risk and income or education levels (available from the census).
- Cluster detection: for the type of data described in 'Disease mapping', apply scan statistic (or Bayesian methods) to the area-level data.
- Small area estimation: DHS survey data are available for many sub-Saharan and South East Asian countries.

Possible Topics

- Reproduce a paper of interest, but there must be real data (which you can get hold of) and spatial vizualization.
- Examine a methodological topic developed for spatial data:
  - INLA with different kind of data, eg survival, categorical,...
  - SPDE.
  - Lumley and Heagerty.