NIN COLLABORATIVE PROJECT-SMALL AREA ESTIMATION TECHNIQUE

Dafna Kohen and Nancy Ross[1]

Small area estimation procedures have been found to be problematic in the analysis of several national data sets. The problem is that a small number of observations exist per spatial/geographic area such as neighborhoods, enumeration areas, census metropolitan areas etc. The problem with small area estimation is that estimates can be biased due to clustering of observations, due to under sampling or under representativeness, and estimates tend to be unstable over time. These problems arise because of sample size but in addition there is generally a lack of co-incidence between the boundaries of survey designed strata and the small area of interest

The purpose of this study is to examine various methods of analyzing data from small areas, comparing statistical methodologies and applying them to various data sets such as the NLSCY. Examples of methodological techniques could include design based estimation, model based methods, mixed estimation methods, and/or other methods. Results from these methods can be compared with other methods that do not take small areas into account.

Design based estimation-the most common are post-stratified domain estimators. This approach can overcome some of the bias associated with under sampling within a domain.

Model based estimation-these are sometimes referred to as synthetic estimators or structure preserving estimators (SPREE). An example is to allocate the proportion of an outcome to a small area based on a large area by other characteristics such as age/sex based on census counts. This however makes the assumption that there is homogeneity in the population subgroup between the small and large area. If this is not correct there is a large resulting bias with this kind of approach. Linear and log-linear models fall under this category as well.

Mixed models-are a combination of design and model based approaches. For example, use a linear combination of a post-stratified estimate and a synthetic where you post-stratify when the domain is equal to or larger than expected under the survey design, and you use the synthetic when the sample size is less than expected under survey design.

Other approaches that can be used to study small area estimation include kreging, splines and Monte-Carlo Markov Chain methods.

Research Questions

1) What are some of the advantages/disadvantages of using different methods of small area estimation?
2) Using small area estimation techniques how do Canadian cities differ on various child outcome measures (eg. PPVT, behavior problems, math scores, reading scores etc.)?
3) Once city based estimates are established, we can look at measures of income distribution and segregation to examine associations of income inequality with child outcomes

Potential problems

1) we need to examine sample sizes, may need to use school aged children only; may not wish to look at the city as the unit of analysis
2) may be an issue re: releasing data due to small n's
3) we don't want to "stigmatize" or single out certain areas-perhaps this is not a problem at the city level
4) what cycle do we look at?

---

[1] Dafna Kohen and Nancy Ross, Statistics Canada, Tunney's Pasture, Ottawa, Ontario K1A 0T6