ESTP course on Small Area Estimation

Computational tools
SAS and related software

Kari Djerf
Statistics Finland
Contents

• Development of SAE software

• Built-in features in SAS

• Specific SAS macros
Development of SAE software

• Traditionally: SAE software were developed for research purposes

• Small markets → large-scale statistical software companies not interested

• Analytical tools exist if modified
Development of SAE software - 2

- Own approaches in statistical agencies depending on type of data (unit or domain level information) and IT environment
- Some typical examples since 1980s:
  - Synthetic estimators
  - Fay-Herriot models
  - GREG-estimators
Development of SAE software - 3

• In Europe development of SAE methods got a lot of support from the EU research programs:
  • SupCom project on SAE (UK, Finland) SAS, SPSS and MLWin
  • FP4 project EURAREA (UK, Finland, Italy, Norway, Poland, Spain, Sweden): SAS macros
  • FP7 project AMELI (Germany, Austria, Estonia, Finland, Slovenia, Switzerland): programs in R
  • FP7 project SAMPLE (Italy, Poland, Spain, UK): programs in R
Development of SAE software - 4

• In addition, a lot of effort in academia, especially US and UK:
  
  • Mixed models

  • Bayesian models

  • etc.
Built-in features in SAS

• SAS/STAT offers analytical tools in for SAE:
  • Survey procedures
  • Generalized Linear models
  • Linear and non-linear mixed models

• Programming
  • SAS language
  • IML
Built-in features in SAS - 2

- Survey procedures which contain both planned and unplanned domains:
  - Direct estimators (Surveyfreq, Surveymeans)
  - Indirect estimators by prediction (Surveyreg, Surveylogistic, Surveyphreg)
Built-in features in SAS - 3

- Indirect estimators

- Generalized linear models: GLM etc.

- Mixed models, i.e. fixed model part and random coefficient part:
  Mixed, Hpmixed, Genmod, Glimmix
  Nlmixed (for really nonlinear models)
Example: Surveymeans

• Program call:
  • Proc Surveymeans data=<data> varmethod=<taylor|brr|jackknife> <statistics> <other options>;
  • Stratum <str>
  • Cluster <clu>
  • Weight <w>
  • Var <Y>
  • Ratio <Y/Z>
  • Domain (d)
  • By <pd>
  • Run;
Example of small area estimation with mixed models

• Article by Mukhopadhyay and McDowell, 2013 (SAS Institute):
  Small Area Estimation for Survey Data Analysis Using SAS® Software

• Unit-level models using Proc Mixed
• Area-level models Proc Mixed and Proc IML
• Un-matched models Proc MCMC
Example of small area estimation with mixed models – 2

• Typical unit-level model is Empirical Best Linear Unbiased Predictor (EBLUP) which contains following parts:

  • Mixed (linear) part, i.e. standard generalized linear model
  • Random intercept part for domains
  • Variance component structure used for MSE (a three-component MSE)
Specific SAS macros

- EURAREA project prepared a set of SAS macros, obtain from the project home page:

- SAS v. 8 or above, with STAT and IML
Specific SAS macros - 2

• Standard estimators:
  • Direct estimator
  • GREG
  • Regression Synthetic Estimators (2)
  • EBLUP using SYN models
Specific SAS macros - 3

- EBLUP estimators with different covariance structures:
  - EBLUP with time-varying area effects
  - EBLUP with spatial correlation structure
  - EBLUP model with unequal sampling weights (Fisher scoring algorithms)

- Structure preserving models (SPREE) and their extensions for cross-classified data
Specific SAS macros - 4

- EBLUPGREG, unit-level models:
  - GREG
  - Regression SYN
  - EBLUP
  - EBLUP with spatial correlation structures
  - EBLUP with many time-series structures
EBLUPGREG macro

- Requirements:

  - SAS software v. 8 or above with STAT and IML

  - Programme runs in RAM and can become slow with huge data sets and complicated models (say, n>20 000, a lot of auxiliary information and complex model)
EBLUPGREG macro – 2

- Requirements:

  - **Sample data set:**
    - One dependent variable \((Y)\)
    - One or more independent variable \((X)\) – continuous or \([0,1]\) indicator
    - Domain indicator \((d)\), normally geographical
    - Optional information for specific model structures:
      - Time identifier \((t)\) for time-dependent models
      - Sampling weight for GREG
EBLUPGREG macro – 3

• Requirements:

  • **Domain-level Population data set:**
    - Domain indicator (d)
    - Population count
    - Sums for auxiliary variables (X)

    Optionally
    - Time indicator (t)
    - And similar population counts and sums of X for each time point
    - Geographical coordinates for spatial correlation
**EBLUPGREG macro – 4**

- **Example:**
  ```sas
  libname a '<DRIVE:\PATH>';
  filename EBLUPLIN '<?DRIVE:\PATH\>EBLUP_LINEAR.SAS';
  %include EBLUPLIN;
  %eblupgreg(sample=a.sample,
  populationSums=a.popsum,
  regionSize=popN,
  y=y1,
  xlist=x1 x2,
  regionIdentifier=domain,
  modules=modules.eurarea,
  greg=1,
  weights=w,
  estimatemeans=1,
  output=a.blups);
  ```
EBLUPGREG macro – 5

- **Output:**
  - **Iterations**
  - **Inverse of information matrix**
  - **Model parameters with some diagnostics**
  - **Correlation structure of estimated \( \beta \)'s**

- **Finally, print or save the results!**
  - **Model prediction for each domain**
  - **Standard errors and/or MSEs with variance components**
Training session

• Some examples will be tried...