HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

SOFTWARE DOMEST FOR SMALL AREA ESTIMATION

Risto Lehtonen (University of Helsinki)
Ari Veijanen (Statistics Finland)

(SAE-kurssi HY 12.4.2011)

B-N-U Workshop, 23-27 August 2010, Vilnius

Outline

- Preliminaries
- Key concepts and definitions
- Domest - Technical documentation
- Domest - Worked examples

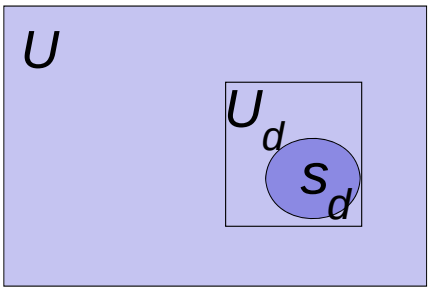
- NOTE: Examples will be worked out during PC session

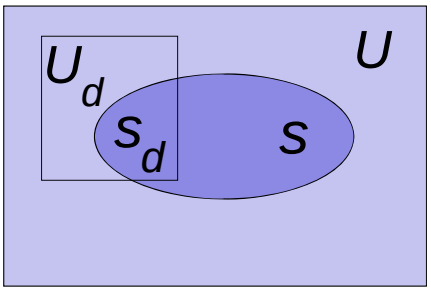
Estimation for domains and SAE

- *Domain estimation*
  - The estimation of population quantities for the desired population subgroups called domains (large or small)
    - Totals , Means, Proportions
    - Medians, Quantiles, Percentiles
    - More complex indicators…
- *Small area estimation*, SAE
  - Estimation for domains whose **sample size** is small or very small (even zero)

Domain estimation design

- Types of domains of interest
  - Planned domains / Unplanned domains
- Type of domain estimator
  - Direct / Indirect
  - Design-based / Model-based
- Availability of auxiliary (population) data
  - Unit-level / Aggregated (area-level)
- Type of model
  - Linear / Non-linear
  - Fixed-effects model / Mixed model
  - Generalized linear mixed models (GLMM)
- Accuracy measures
  - Variance estimators / MSE estimators





Planned domains

U Population

U_d Population domain d

Domains = Strata

$s_d \subset U_d$ Sample in domain d

Sample size n_d in domain d is **fixed**

$d = 1,...,D$

Unplanned domains

U Population

s Sample

U_d Population domain d

$s_d = s \cap U_d$ Sample in domain d

Sample size n_d in domain d is **random**

$d = 1,...,D$

Domain type and estimator type

Domain type	Estimator type	
	Direct	Indirect
Planned	Typical set-up	More rarely
Unplanned	More rarely	Typical set-up

Design-based properties of estimators

	Design-based methods HT, GREG, MC	Model-based methods SYN, EBLUP, EB
Bias	**Design unbiased** (approximately) by the construction principle	**Design biased** Bias does not necessarily approach zero with increasing sample size
Precision (Variance)	**Large variance for small domains** Variance decreases with increasing sample size	**Small variance for small domains** Variance decreases with increasing sample size
Accuracy (Mean Squared Error, MSE)	**MSE = Variance (or nearly so)**	**MSE = Variance + squared Bias** Accuracy can be poor if the bias is substantial
Confidence intervals	Valid design-based CI can be constructed	Valid design-based CI not necessarily obtained

Domest software

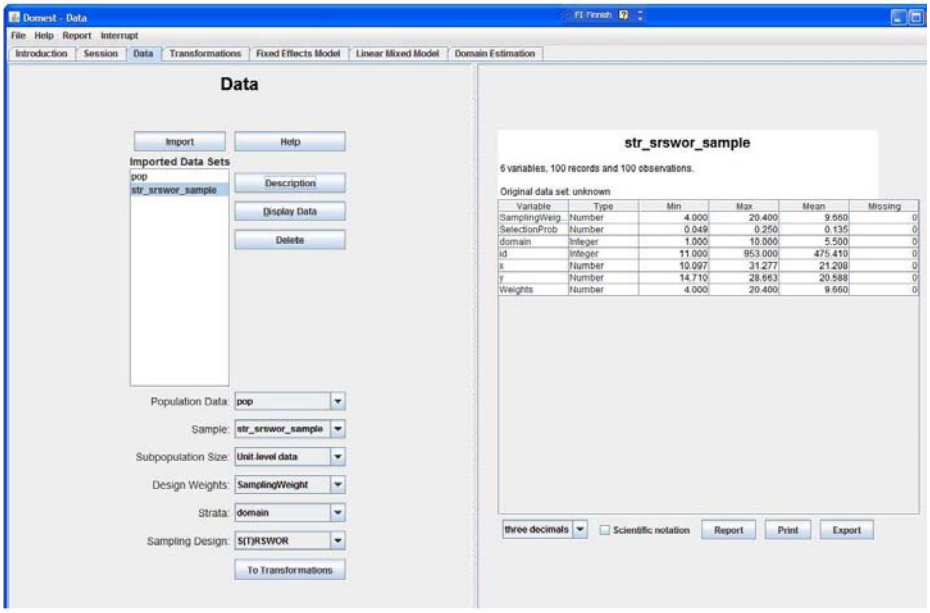
- Stand-alone interactive Java application
- GUI
- Estimation
  - Domain totals and means
  - Variance estimation
  - MSE estimation
  - Random effects estimation
- Domest documentation (2010)
  - Part 1: Technical documentation
  - Part 2: Worked examples

Estimation procedure with Domest

- Domest Pages
  - Introduction
  - Session
  - Data
  - Transformations
  - Fixed Effects Model
  - Linear Mixed Model
  - Domain Estimation

Page Data

- Page Data
  - Import population data set
  - Import sample data set
  - Define level of population data
  - Identify weight variable
  - Define stratum variable (if any)
  - Define sampling design
  - Display data
  - Describe data
- Input data formats
  - Text data set e.g. data.txt
  - SAS data set e.g. data.sas7bdat

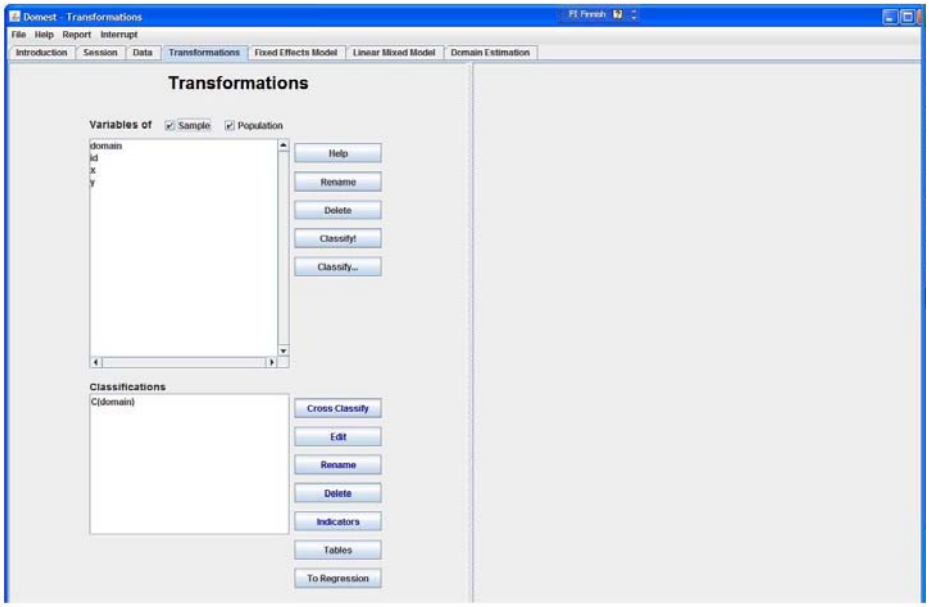


Page Transformations

- Classification of variables
- Renaming variables
- Creating of indicator variables

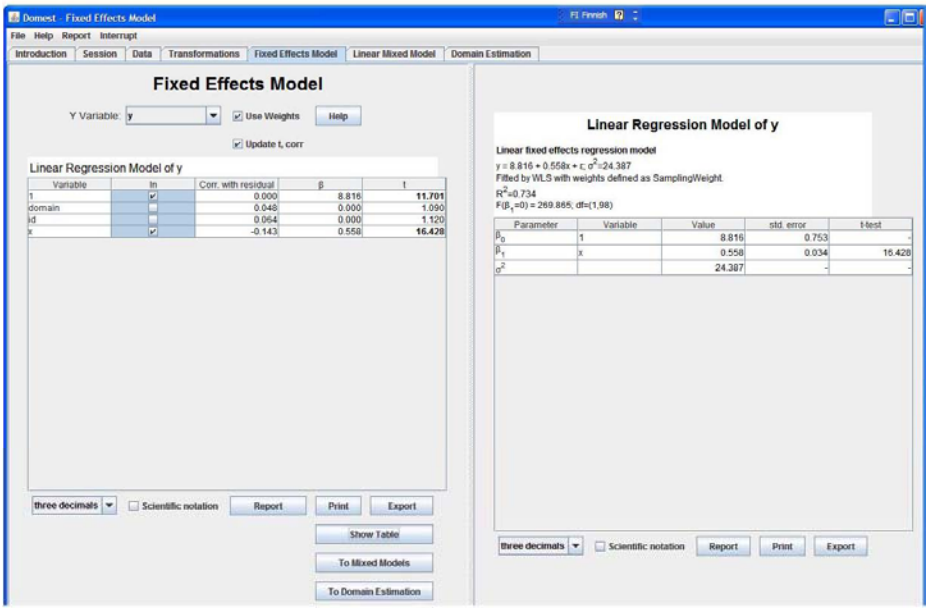
- Function Classify!
  - Creates classified domain variable

- Function Indicators
  - Create domain indicators for inclusion of domain-specific intercepts in the fixed-effects model



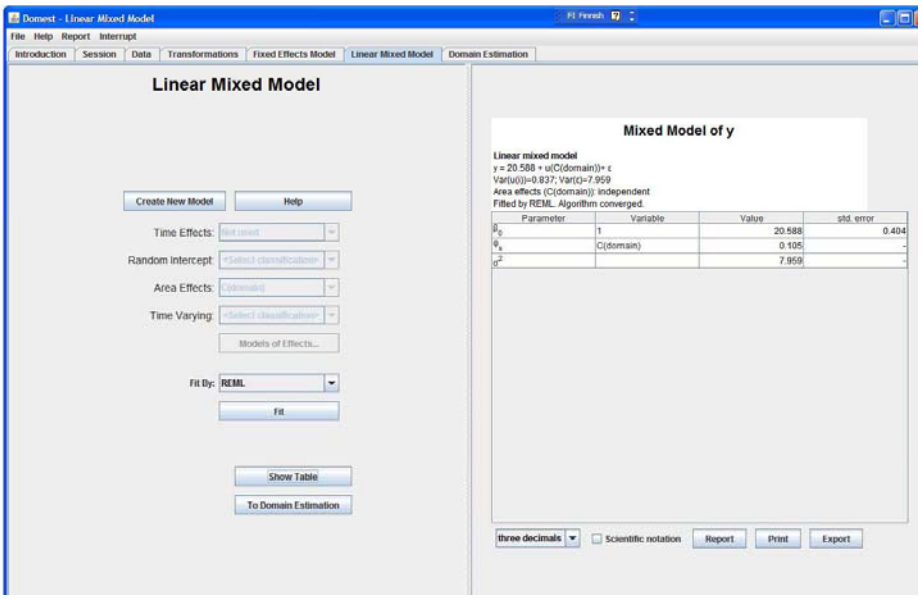
Page Fixed Effects Model

- Interactive model fitting
- Identify y-variable
- Include explanatory x-variables in the model
- Use of weights
  - OLS or WLS estimation
- Fit the model
- Show results



Page Linear Mixed Model

- Specify random effects
  - Random intercepts
  - Area effects
  - Time effects
  - Time varying effects
- Select estimation method
  - REML, REML-W (weighted REML)
  - ML, ML-W (weighted ML)
- Fit the model
- Show results



Page Domain Estimation

- Domains
  - Identify domain variable
- Select Model
  - Linear fixed-effects model
  - Linear mixed model
  - Identify weight variable (if any)
- Select Estimators
  - HT, Hajek, GREG, MGREG, EBLUPs, SYN
- Select Statistics
  - Domain totals
  - Domain means
  - Variance and MSE

Design-based estimators

- Horvitz-Thompson HT estimator
  - Options:
    - Ordinary variance estimator
    - Hájek variance estimator
- Hájek estimator
- GREG estimators
  - Fixed-effects assisting model
  - Options variance estimation
- MGREG estimators
  - Mixed-effects assisting model
  - Options for variance estimation

Model-based estimators

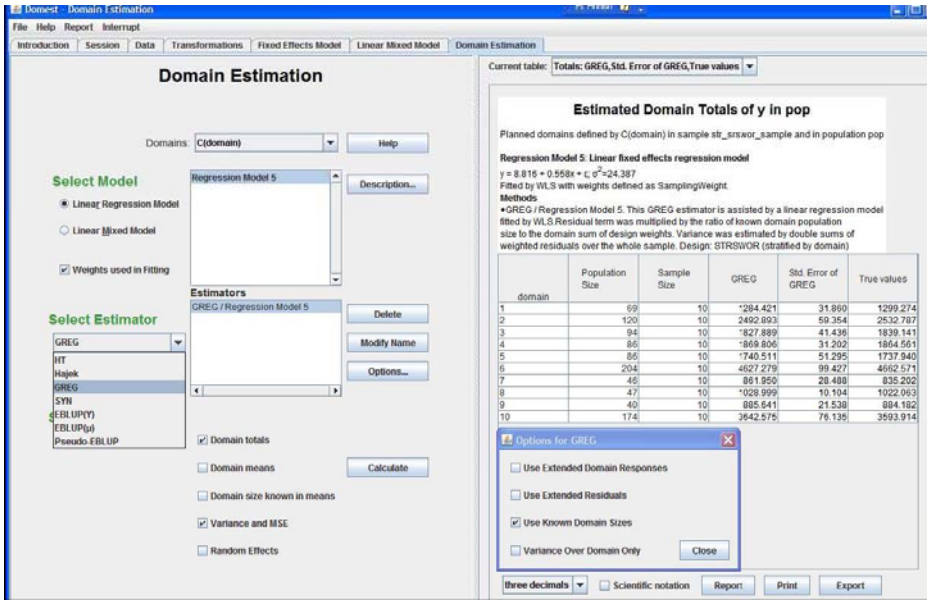
- EBLUP estimators
  - Empirical Best Linear Unbiased Predictor
- Pseudo EBLUP estimator
- MSE estimation
  - Components g1, g2, g3, g4 of Mean Cross Product Error matrix MCPE
- (Synthetic estimator)
- NOTE: Weights can be incorporated into the mixed model estimation procedure

Table 1. Summary of estimators for domains: Estimator and model types and the use of weights in the estimation procedure.

Domain estimator	Weights in domain estimator	Assisting / Underlying model	Weights in model fitting	Model fitting method
Design-based estimators				
HT	Yes	None	-	-
Hájek	Yes	None	-	-
GREG	Yes	Linear fixed-effects model	Yes	WLS
GREG-OLS	Yes	Linear fixed-effects model	No	OLS
MGREGW	Yes	Linear mixed model	Yes	ML-W or REML-W
MGREG	Yes	Linear mixed model	No	ML or REML

Table 1. Summary of estimators for domains: Estimator and model types and the use of weights in the estimation procedure.

Domain estimator	Weights in domain estimator	Assisting / Underlying model	Weights in model fitting	Model fitting method
Model-based estimators				
SYN	No	Linear fixed-effects model	No	OLS
SYNW	No	Linear fixed-effects model	Yes	WLS
EBLUP(Y)	No	Linear mixed model	No	ML of REML
EBLUP(Y)W	No	Linear mixed model	Yes	ML-W or REML-W
EBLUP(μ)	No	Linear mixed model	No	ML or REML
EBLUP(μ)W	No	Linear mixed model	Yes	ML-W or REML-W
Pseudo-EBLUP	Yes	Linear mixed model	No	ML or REML
Pseudo-EBLUPW	Yes	Linear mixed model	Yes	ML-W or REML-W



References

- Lehtonen R. and Veijanen A. (2009). Design-based methods of estimation for domains and small areas. Chapter 31 in Rao C.R. and Pfeffermann D. (Eds.). *Handbook of Statistics. Sample Surveys: Inference and Analysis. Vol. 29B.* New York: Elsevier.
- Lehtonen R., Särndal C.-E. and Veijanen, A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, 29, 33–44.
- Rao J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons.
- Saei, A. and R. Chambers (2004). Small area estimation under linear and generalized linear mixed models with time and area effects. EURAREA Consortium 2004, *Project Reference Volume*, www.statistics.gov.uk/eurarea.
- Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology* 33, 99–119.