Small Area Estimation (78405)**, Spring** 2015

Homework assignment (2-3 ECTS credits)

Intermediate level (Aineopinnot)**: Exercises 1, 2 ja 3** (Voluntary homework)

Advanced level (Syventävät opinnot): **Exercises 1, 2 ja 3 and 4** (Obligatory homework)

Final product: Written report (10-15 pages plus selected annexes)
Tentative structure:
Title page (title, author, course, department, date, student id number, Intermediate level or Advanced level)
Text part (divided into suitable sections and subsections)
References (literature)
Annexes (extracts from output, selected pieces of program code)

Delivery of final product by **31 March 2015** (in pdf format) as email attachment:
risto.lehtonen@helsinki.fi
or as paper copy:
Risto Lehtonen
University of Helsinki, Department of Social Research
P.O. Box 18 (Unioninkatu 35), 00014 Helsingin yliopisto

The population dataset used in PC classes provides the empirical data set (download SAS or .txt version from course webpage).

Population dataset: pop.sas7bdat, $N = 966$ elements
Domains: $D = 10$ population subgroups (domains)
Parameters of interest: Domain totals (or means) of study variable y for the domains
Auxiliary information: Variable x
Computation: SAS and SAS macro EBLUPGREG. Alternatively, R tools can be used if desired.

Exercise 1. Descriptive statistics for the population data

Perform and report selected descriptive statistics for variables y and x in the population domains (incl. graphics).

Exercise 2. Planned domains case (stratified sampling)

a) Draw a 10% sample (samprate=.10) by stratified simple random sampling (STR-SRSWOR) with the domains as the strata ($D = 10$ strata). SAS procedure SURVEYSELECT. Choose your personal SEED.
Perform and report selected descriptive statistics for variables y and x in your sample domains (incl. graphics).

b) Compute HT estimates of domain totals (or means). Report the total (or mean) estimators and their variance estimators. Show program codes. Show results in tables (point estimates, standard errors, coefficients of variation).

c) Compute GREG estimates of domain totals (or means). Use x as a covariate in the model. Report the total (or mean) estimators and their variance estimators. Show program codes. Show results in tables (point estimates, standard errors, coefficients of variation).

d) Compare estimation results from points 2b) and 2c) and draw statistical conclusions.

e) Perform points 2a) and 2b) with sampling rate of .25 (25% sample) by HT estimation and compare results with the results from the 10% sample in point 2b). Draw statistical conclusions.

Exercise 3. Unplanned domains case

a) Draw a $n = 100$ element sample by simple random sampling (SRSWOR, no stratification). NOTE: Set your personal SEED.
Perform and report selected descriptive statistics for variables y and x in your sample domains (incl. graphics).

b) Compute HT estimates of domain totals (or means). Report the total (or mean) estimators and their variance estimators. Show program codes. Show results in tables (point estimates, standard errors, coefficients of variation).

c) Compute GREG estimates of domain totals (or means). Use x as a covariate in the model. Report the total (or mean) estimators and their variance estimators. Show program codes. Show results in tables (point estimates, standard errors, coefficients of variation).

d) Compare estimation results from points 3b) and 3c) and draw statistical conclusions.

For advanced level:

Exercise 4. SAS macro EBLUPGREG

a) Unplanned domains case. Draw SRSWOR samples as follows:
Sample 1: sample size n = 50
Sample 2: sample size n = 100
Sample 3: sample size n = 250

b) Perform HT estimation of domain totals (or means) and their standard errors for the three samples by macro EBLUPGREG. Present the results in tables. Compare results and draw statistical conclusions.

c) Perform GREG estimation of domain totals (or means) and their standard errors for the three samples by macro EBLUPGREG. Present the results in tables. Compare results and draw statistical conclusions.

d) Perform EBLUP estimation of domain totals (or means) and their standard errors (actually, square roots of the MSE approximation) for the three samples by macro EBLUPGREG. Present the results in tables. Compare results and draw statistical conclusions.

e) Compare the HT, GREG and EBLUP estimation results (point and accuracy estimates) and draw statistical conclusions.

ADDITIONAL exercise (extra 1 ECTS credit point, for both intermediate and advanced levels))

Exercise 5. Use the RDomest software for a domain estimation problem of your own choice and report your results and conclusions.

NOTE: Alternative way to perform homework assignment can be agreed for students who have their own data and domain / small area estimation problem (please contact the lecturer).