

Pienalue-estimointi (78189)

Kevät 2011

Luennoija: Prof. [Risto Lehtonen](#)

Laajuus 6-8 op

Tyyppi

Aineopintojen erikoiskurssi

Syventävien opintojen erikoiskurssi

Luennot

Tiistaisin klo 14–18 Exactumin salissa C323 ajalla 15.3.–12.4.2011, yhteensä 20 tuntia.

Harjoitukset

Torstaisin klo 12–15 mikroluokassa C128 ajalla 17.3.–14.4.2011, yhteensä 12 tuntia.

Loppukuulustelu Tiistai 19.4.2011 klo 14–16 Exactum C323

Suoritustapa

Aineopinnot: Loppukuulustelu (6 op) tai loppukuulustelu ja (vapaaehtoinen) harjoitustyö (8 op)

Syventävät opinnot: Loppukuulustelu ja (pakollinen) harjoitustyö (8 op)

Kirjallisuutta

Rao J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons.

Lehtonen R. and Djerf K. (2008). *Survey sampling reference guidelines*. Luxembourg: Eurostat Methodologies and Working papers. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-003/EN/KS-RA-08-003-EN.PDF

Lehtonen R. and Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Second Edition. Chichester: John Wiley & Sons. Chapter 6.

Lehtonen R. and Djerf K. (eds.) (2001). *Lecture Notes on Estimation for Population Domains and Small Areas*. Statistics Finland: Reviews 2001/15.

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. Vol. 29B. Sample Surveys: Inference and Analysis*. New York: Elsevier, pp. 219-249.

Web-materiaalia

VLISS-virtual laboratory in survey sampling, <http://mathstat.helsinki.fi/VLISS/>

EURAREA Project, <http://www.statistics.gov.uk/eurarea/>

EWORSAE – the European Working Group on Small Area Estimation <http://sae.wzr.pl/>

Tavoitteet

Kurssin tavoitteena on perehdyttää opiskelija perusjoukon osajoukkoja koskevan estimoinnin (*small area estimation*, SAE) lähestymistapoihin, teorioihin, laskentamenetelmiin ja sovelluksiin. Kurssilla käsitellään asetelmaperusteisia malliavusteisia (*design-based model assisted*) menetelmiä, kuten yleistetyt regressioestimaattorit (GREG) ja kalibroitimenetelmät, sekä malliperusteisia (*model-based*) menetelmiä, kuten synteettiset ja EBLUP-estimaattorit. Lisäksi tarkastellaan estimointiin soveltuvia tilastollisia ohjelmistoja. Sovellukset ovat pääasiassa yhteiskuntatieteellisiltä ja terveystieteellisiltä aloilta. Käytännön harjoituksissa käytetään laskentatyökaluina SAS-ohjelmiston makroja ja erikoisohjelmia kuten DOMEST. Kurssin suoritettuaan opiskelijan odotetaan tuntevan alan teorioita, menetelmiä ja laskentaa sekä hallitsevan alan ohjelmistojen soveltamisen tyypillisissä tilanteissa. Kurssin menestyksestä suorittamista edesauttaa, jos opiskelijalla on perustiedot otantamenetelmistä ja tilastollisesta mallinnuksesta. Kurssi soveltuu tilastotieteen aine- tai syventäviä opintoja suorittaville opiskelijoille sekä myös yliopistoissa, korkeakouluissa ja tutkimuslaitoksissa toimiville jatko-opiskelijoille ja tutkijoille.



SOFTWARE DOMEST FOR SMALL AREA ESTIMATION

Risto Lehtonen¹ and Ari Veijanen²

¹University of Helsinki, Finland
e-mail: risto.lehtonen@helsinki.fi

²Statistics Finland
e-mail: ari.veijanen@stbl.fi

Abstract

Statistical software Domest can be used for the estimation of totals and means with design-based and model-based methods for population subgroups or domains and small areas. The basic functions and options of the software are described in the paper. Examples are given.

Keywords: EBLUP estimation, GREG estimation, mixed model.

1 Introduction

Domest is a stand-alone interactive Java application developed for the estimation of totals and means for population subgroups or domains and small areas. The program covers selected methods described in Lehtonen, Särndal and Veijanen (2003), Lehtonen and Veijanen (2009) and Saei and Chambers (2004). Domest provides both design-based and model-based domain estimators with the accompanying variance and MSE estimators.

Domest is developed at Statistics Finland by Dr. Ari Veijanen together with Prof. Risto Lehtonen. Domest is freely available from the authors.

2 Methods

Design-based methods include HT and GREG methods presented in Särndal, Swensson and Wretman (1992) and Lehtonen and Veijanen (2009). GREG estimation is assisted by linear fixed effects regression models or linear mixed models, fitted with or without design weights. Currently, GREG variance estimation allows SKSWOR, Poisson sampling, and π PS with approximated second-order inclusion probabilities (Hájek, 1964; Berger, 2004, 2005). Linear mixed models are incorporated into model-based EBLUP, synthetic estimator and pseudo-EBLUP (Rao, 2003). MSE estimation uses methods described in Rao (2003) and Saei and Chambers (2004).

A linear regression model is fitted by OLS or WLS, and a linear mixed model is fitted by ML or REML. The mixed model can include area and/or time effects. The area effects are assumed independent and time effects have AR(1) correlations.

SAS data or text files can be imported into Domest. Output tables are saved as text files or added incrementally to an HTML file. SAS is only needed if the input data set is in SAS format.

A more complete documentation will be distributed for workshop participants.

References

- Berger, Y. G. (2004). A simple variance estimator for unequal probability sampling without replacement. *Journal of Applied Statistics* **31**, 305-315.
- Berger, Y. G. (2005). Variance estimation with highly stratified sampling designs with unequal probabilities. *Australian & New Zealand Journal of Statistics* **47**, 365-373.
- Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics* **35**, 1491-1523.
- Lehtonen, R., C.-E. Särndal, and A. Veijanen (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology* **29**, 33-44.
- 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.
- Rao, J. N. K. (2003). *Small area estimation*. Wiley: New York.
- 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., B. Swensson, and J. Wretman (1992). *Model assisted survey sampling*. Springer-Verlag, New York.