

## Otantamenetelmät (78143), Syksy 2008

### SYVENTÄVÄT OPINNOT

#### Harjoitustyö

Harjoitustyö tehdään henkilökohtaisena työnä. Työn hyväksytty suoritus: 2 op.

Loppuutuotteena on noin 10 sivun raportti liitteineen, esimerkiksi rakenteella:

Kansilehti (otsikko, tekijä, kurssi, päiväys, Syventävät opinnot)

Tekstiosa (jaa soveltuviin lukuihin ja alilukuihin)

Viitteet (kirjallisuus)

Liitteet (valitut ohjelmakoodit ja tulostusotteet).

Työn palautus: **31.1.2009 mennessä** liitetiedostona (PDF) sähköpostitse: [risto.lehtonen@helsinki.fi](mailto:risto.lehtonen@helsinki.fi)  
tai postitse:

Risto Lehtonen, Helsingin yliopisto, Matematiikan ja tilastotieteen laitos, PL 68 (Gustaf Hällströmin katu 2b),  
00014 Helsingin yliopisto

Työn voi myös jättää laitoksen kansliaan (Exactum 3 krs.) kanslian aukioloaikoina.

Estimoitavana parametrina on työttömien kokonaismäärä Keski-Suomen läänissä (tulomuuttuja UE91). Kehikkoperusjoukkona on Province91-populaatio. Tehtävissä voit käyttää ja kehittää edelleen demoissa käsitellyjä SAS-ohjelmakodeja. Voit käyttää myös SPSS:n Complex Samples -modulia tai R-funktioita.

#### Tehtävä 1. Lisäinformaation käyttö otanta-asetelmassa

a) Tutki graafisesti tulomuuttujan UE91 jakaumaa perusjoukossa. Raportoi havaintosi.

b) Poimi perusjoukosta yksinkertainen satunnaisotos palauttamatta (SRSWOR) Bernoulli-otannalla niin, että odotettu otoskoko on 8 kuntaa (valitse oma siemenluku SEED). Estimoi totaali sekä totaaliestimaattorin asetelmavarianssi, keskivirhe ja variaatiokerroin. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset. Mistä syystä yksinkertainen satunnaisotanta ei ole hyvä strategiavalinta tässä perusjoukossa?

c) Poimi perusjoukosta PPS-WOR-otos Poisson-otannalla niin, että odotettu otoskoko on 8 kuntaa (valitse oma siemenluku SEED). Valitse kokomuuttuja poimintaa varten ja perustele valintasi. Estimoi totaali sekä totaaliestimaattorin asetelmavarianssi, keskivirhe ja variaatiokerroin. Estimoi lisäksi totaaliestimaattorin asetelmakerroin. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset.

d) Vertaa kohtien 1b ja 1c strategioiden tehokkuutta.

#### Tehtävä 2. Lisäinformaation käyttö estimointiasetelmassa

a) Poimi perusjoukosta  $n = 8$  kunnan SRSWOR-otos SAS-proseduurilla SURVEYSELECT (valitse oma siemenluku SEED) ja estimoi otoksen perusteella totaali sekä totaaliestimaattorin asetelmavarianssi, keskivirhe ja variaatiokerroin. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset.

b) Suunnittele ja toteuta mahdollisimman tehokas strategia totaalin estimointia varten, kun käytettävissä on kohdassa 2a poimimasi SRSWOR-otos sekä kehikkoperusjoukossa oleva lisäinformaatio. Perustele valintasi. Estimoi laatimasi strategian perusteella totaali sekä totaaliestimaattorin asetelmavarianssi, keskivirhe ja variaatiokerroin. Estimoi lisäksi totaaliestimaattorin asetelmakerroin. Raportoi strategia, käyttämäsi estimaattorit ja numeeriset tulokset.

c) Vertaile tehtävien 1c ja 2b estimointituloksia ja tee perustellut johtopäätökset.

#### Tehtävä 3. Estimaattoreiden ominaisuuksien tutkiminen (Monte Carlo -simulointi)

a) Täydennä demojen SAS-makroa `%MACRO simulate(data,n,k,pps,seed);` regressioestimoinnilla totaalille, yhden z-apumuuttujan tilanne, otanta-asetelmana SRSWOR.

b) Tee simulointikoe (1000 riippumatonta otosta perusjoukosta) ja vertaa SRSWOR-, PPS- ja regressioestimaattoreiden ominaisuuksia: harha (Bias, ARB), keskivirhe (Std\_error) ja tarkkuus (Root\_MSE). Raportoi päätelmät.