

## Otantamenetelmät (78143), Syksy 2008

### AINEOPINNOT Harjoitustyö

Harjoitustyö tehdään 1-2 hengen ryhmissä. Työn hyväksytty suoritus: 2 op.  
Loppuotteena on noin 10 sivun raportti liitteineen, esimerkiksi rakenteella:  
Kansilehti (otsikko, tekijä(t), kurssi, päiväys, Aineopinnot)  
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 (tulosmuuttuja UE91). Kehikkoperusjoukko on Province91-populaatio. Tehtävissä voit käyttää ja kehittää edelleen demoissa käsiteltyjä SAS-ohjelmakoodeja. Voit käyttää myös SPSS:n Complex Samples -modulia tai R-funktioita.

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

a) Tutki graafisesti tulosmuuttujan 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 käyttämäsi estimaattorit ja tulokset. Mistä syystä yksinkertainen satunnaisotanta ei ole hyvä strategiavalinta tässä perusjoukossa?

c) Poimi perusjoukosta  $n = 8$  kunnan otos ositetulla otannalla muuttujan UE91 totaalin estimointia varten. Valitse itse ositusmuuttuja, kiintiöintimenetelmä ja otantamenetelmä ositteiden sisällä (pyri mahdollisimman tehokkaaseen ositetun otannan asetelmaan). Perustele valintasi. Estimoi laatimasi ositetun otannan asetelman perusteella 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.