

# Otantamenetelmät (78143), Syksy 2014

## AINEOPINNOT

### Harjoitustyö

Harjoitustyö tehdään 1-2 hengen ryhmissä. Työn hyväksytty suoritus: 2 op.  
Lopputuotteena on 10-15 sivun raportti liitteineen, esimerkiksi rakenteella:  
Kansilehti (otsikko, tekijä(t), kurssi, päiväys, Aineopinnot)  
Tekstiosa (jaa soveltuviin lukuihin ja alilukuihin)  
Viitteet (kirjallisuus)  
Liitteet (valitut ohjelmakoodit ja valitut tulostusotteet).

Työn palautus: **31.1.2015 mennessä** liitetiedostona (PDF) sähköpostitse: [risto.lehtonen@helsinki.fi](mailto:risto.lehtonen@helsinki.fi)  
tai postitse:

Risto Lehtonen, Helsingin yliopisto, Sosiaalitieteiden laitos, PL 18 (Unioninkatu 35), 00014 Helsingin yliopisto

Estimoitavana parametrina on valitsemasi työvoimaa kuvaavan muuttujan kokonaismäärä (totaali) Keski-Suomen läänissä (esim. UE91 tai LAB91). Kehikkoperusjoukkona on Province91-populaatio. Tehtävissä voit käyttää ja kehittää edelleen harjoituksissa käsiteltyjä SAS-ohjelmakoodeja. Voit käyttää myös SPSS:n Complex Samples -modulia tai R-funktioita.

### Tehtävä 1. Lisäinformaation käyttö otanta-asetelmassa

- Tutki graafisesti valitsemasi tulosuuttujan jakaumaa perusjoukossa. Raportoi havaintosi.
- Poimi perusjoukosta yksinkertainen satunnaisotos palauttamatta (SRSWOR) Bernoulli-otannalla niin, että odotettu otoskoko on 8 kuntaa (valitse oma siemenluku SEED). Estimoi tulosuuttujan totaali sekä totaaliestimaattorin asetelmavarianssi, keskivirhe ja variaatiokerroin. Raportoi käyttämäsi estimaattorit ja tulokset. Mistä syystä yksinkertainen satunnaisotanta ei ole hyvä strategiavalinta tässä perusjoukossa? Perustele vastauksesi.
- Poimi perusjoukosta  $n = 8$  kunnan otos ositetulla otannalla tulosuuttujan totaalin estimointia varten. Valitse itse ositusmuuttuja, kiintiöintimenetelmä ja otantamenetelmä ositteiden sisällä (pyri mahdollisimman tehokkaiseen ositetun otannan asetelmaan). Perustele valintasi. Estimoi laatimasi ositetun otannan asetelman perusteella totaali sekä totaaliestimaattorin asetelmavarianssi, keskivirhe ja variaatiokerroin. Laske myös asetelmakertoimen (deff) estimaatti. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset.
- Vertaa kohtien 1b ja 1c strategioiden tehokkuutta.

### Tehtävä 2. Lisäinformaation käyttö estimointiasetelmassa

- 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 valitsemallesi tulosuuttujalle. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset.
- Suunnittele ja toteuta mahdollisimman tehokas strategia kohdassa 2a valitsemasi tulosuuttujan 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. Laske myös asetelmakertoimen (deff) estimaatti. Raportoi strategia, käyttämäsi estimaattorit ja numeeriset tulokset.
- Vertaile tehtävien 1c ja 2b estimointituloksia ja tee perustellut johtopäätökset.