

Otantamenetelmät (78143), Syksy 2014

SYVENTÄVÄT OPINNOT

Harjoitustyö

Harjoitustyö tehdään henkilökohtaisena työnä. Työn hyväksytyt suoritus: 2 op
Loppuutotteena 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
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 kehitellä edelleen harjoituksissa käsiteltyjä SAS-ohjelmakoodeja. Voit käyttää myös SPSS:n Complex Samples -modulia tai R-funktioita tai ohjelmia (esim. Thomas Lumleyn Survey-paketti).

Tehtävä 1. Lisäinformaation käyttö otanta-asetelmassa

a) Tutki graafisesti valitsemasi tulosuuttujan 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 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.

c) Poimi perusjoukosta $n = 8$ kunnan otos *ositetulla otannalla* tulosuuttujan 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. Laske myös asetelmakerroimen (deff) estimaatti. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset.

d) Poimi perusjoukosta $n = 8$ kunnan otos *ositetulla PPS-otannalla* tulosuuttujan totaalin estimointia varten. Valitse kokomuuttuja, ositusmuuttuja ja otoksen kiintiöintimenetelmä. Perustele valintasi. Estimoi laatimasi otanta-asetelman perusteella totaali sekä totaaliestimaattorin asetelmavarianssi, keskivirhe ja variaatiokerroin. Laske myös asetelmakerroimen (deff) estimaatti. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset.

d) Vertaa kohtien 1b, 1c ja 1d 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 valitsemallesi tulosuuttujalle. Raportoi otanta-asetelma, käyttämäsi estimaattorit ja numeeriset tulokset.

b) Suunnittele ja toteuta mahdollisimman tehokas *malliavusteinen strategia* regressioestimoinnilla 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 asetelmakerroimen (deff) estimaatti. Raportoi strategia, käyttämäsi estimaattorit ja numeeriset tulokset.

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

Tehtävä 3. Regressioestimointi ohjelmoimalla estimointiyhtälöt

- a) Suorita tehtävän 2b totaaliestimointi ohjelmoimalla regressioestimoinnin estimointiyhtälöt. Raportoi tuottamasi ohjelmakoodi ja tulokset.
- b) Vertaa numeerisia tuloksiasi kohdan 2b estimointituloksiin.

Tehtävä 4. Simulatiivinen tarkastelu

Perehdy harjoitustyön SAS-makroon (ks. kurssin kotisivu), jonka avulla voidaan tarkastella Monte Carlo -simuloinnin avulla totaaliestimaattorin teoreettisia ominaisuuksia SRSWOR-strategiassa. Tässä keskitytään asetelmaharhaan (*Design bias*).

Valitse tulosmuuttuja (UE91 tai LAB91). Tee simulointikoe (500-1000 riippumatonta SRSWOR-otosta perusjoukosta) eri otoskoolla:

- a) $n=4$
b) $n=8$
c) $n=16$

Vertaa kohdissa a-c saatuja totaaliestimaattorin odotusarvon approksimaatioita parametrin T arvoon ja raportoi päätelmät.

Tutki jakaumakuvioiden avulla totaaliestimaattorin jakaumaa eri otoskoon arvoilla ja raportoi päätelmät.