

Otantamenetelmät

Seppo Laaksonen

Tentti joulukuun 2007

Vastaa 7 kysymykseen.

1. Kaksiasteista 'aitoa' ryväotantaa käyttäen saatiin keskiarvolle ensimmäisessä ja toisessa asteessa varianssit: $var(psu) = 100$ ja $var(ssu) = 50$. Ryppäiden (psu :iden) määrä perusjoukossa oli 500. Ryppäistä valittiin otokseen 30. Näissä oli yhteensä 300 toisen asteen yksikköä joista siis toisen asteen otos otettiin. Laske totaalin keskivirhe.

2. Olet poiminut tietyllä otanta-asetelmalla otoksen ja laskenut saadusta aineistosta piste-estimaatteja ja näille väliestimaatteja. Haluat toistaa otannon uusilla ja uusilla riippumattomilla satunnaisluvulla. Miten sen teet? Toiseksi ja tärkeämpänä: mitä hyötyä tästä toistamisesta on ja mitä uusia tunnuslukuja/mittareita saat aikaiseksi ja mitä niistä voi hyötyä?

3. Toistetun tasavälipoiminnan idea ja hyöty

4. Esitä kahdelle eri otanta-asetelmalle ensimmäisen ja toisen asteen sisältymistodennäköisyys ja kerro mitä käyttöä kummallakin on.

5. Miten epälineaarinen estimaattori eroaa lineaarisesta ja mitä eroja on kummankin varianssiestimaattoreissa? Voit havainnollistaa tilannetta esimerkillä, jos se sillä tavalla on helpompaa.

6. Ohessa on kaava pps -pohjaisen totaalin varianssiestimaattorille:

$$v(t_{pps}) = \frac{1}{n(n-1)} \sum_k \left(\frac{y_k}{p_k} - t_{pps} \right)^2$$

Selosta ensin miten itse totaali on estimoitu ja jatka kertomalla tuon kaavan filosofia sellaiselle joka ymmärtää tilastotieteen perusteet muttei otantaa syvällisesti.

7. Ohessa on totaalin varianssiestimaattori ositetulla satunnaisotannalla:

$$v(t_{str}) = N^2 v(\bar{y}_h) = \sum_h N_h^2 \left(\frac{N_h - n_h}{N_h} \right) \left(\frac{s_h^2}{n_h} \right) = \sum_h N_h (N_h - n_h) \left(\frac{s_h^2}{n_h} \right)$$

Selosta kunkin kaavan osan luonne sekä muodosta jokin keskiarvon varianssiestimaattori.

8. Selosta lyhyesti vähintään 5 eri otosten allokointimenetelmän idea (kokonaisotoksen allokointi ositteille).

9. Suhde-estimoinnin ja regressioestimoinnin yhtäläisyydet ja eroavuudet.

10. Nettikyselyt: mitä etuja ja mitä kiusallisuuksia niissä kohdataan ja miten kiusallisuuksiin voidaan yrittää vastata?

11. Ohessa on ESS:n otantasuunnitelma Irlannille. Selosta ytimekkään lyhyesti tekijät, joilla tässä on päädytty brutto-otoskokoön 3508. Toiseksi: onkohan suunnitelmassa jotain kritisoitavaa?

Target Population, Population coverage	Persons aged 15 years and over who are resident in private households in the Republic of Ireland
Problems	15-17 year olds will only be interviewed with parental consent. Note, that it is expected that response rates among these age categories will be very low.

Sampling frame	For stratification and selection of addresses: Computer-based National Electoral Register. This register is updated annually. The 2003/2004 register is used with 3,014367 electors.
----------------	---

Sampling design	<p>Three-stage probability sampling</p> <p>PSUs are aggregates of District Electoral Divisions (DEDs). There is a total of 3,440 DEDs in Ireland. The PSUs are ordered geographically (implicit stratification).</p> <p>Stage 1: PPS selection of 250 PSUs (i.e. clusters of individuals) from a national total of 1,475 clusters formed from the Electoral Roll. The minimum cluster size is 1,000 individuals.</p> <p>Stage 2: Systematic sample of 14 addresses within each PSU. The addresses are sorted by Polling Book and Elector reference number - this provides a geographical sort on the data.</p> <p>Stage 3: Random selection of one individual 15 years and over within each household (Next-Birthday-Method)</p>
-----------------	--

Design effects	$DEFF_c = 1 + (9 - 1) * 0.04 = 1.32$ $DEFF_p = 1.04$ $DEFF = 1.37$
----------------	--

Remark	The estimation for DEFF is based on the results of round I. Principally, the design is epsem (equal probability sampling mechanism) for individuals. $DEFF_p$ is because of different inclusion probabilities for 15-17 years old individuals.
--------	--

Target response rate	About 64% (round I)
Problems	The target response rate (70%) will not be reached.

Sample size	A net sample of size $n_{net} = 2,200$ interviews will be conducted. With 2% of ineligibles (round 1) and a response rate of about 64% the gross sample size must be $n_{gross} = 2,200 / (0.64 * 0.98) = 3,508$. For each cluster about 14 addresses have to be drawn. That results in an average of 9 interviews per cluster.
-------------	--

Otantamenetelmät (78143), Syksy 2007
Loppupentti 18.12.2007

1. Asetelmakerroin (*Design effect*)

- a) Määrittele totaalien regressioestimaattorin \hat{t}_{reg} asetelmakerroin $deff(\hat{t}_{reg})$ strategialle SRS*reg (käytä laajennettua asetelmakertoimen määritelmää).
- b) Mitä päättelet strategian SRS*reg tehokkuudesta yksinkertaiseen satunnaisotantaan (SRSWOR) verrattuna, jos totaaliestimaattorin \hat{t}_{reg} asetelmakerroin on pienempi kuin yksi? Entä jos asetelmakerroin on ykkösen suuruinen?

2. Systemaattinen otanta (*Systematic sampling*)

- a) Selosta systemaattisen otannan poimintaproseduuri ja työvaiheet.
- b) Selosta implisiittisen osittamisen menetelmä systemaattisessa otannassa.

3. Ositettu otanta (*Stratified sampling*)

- a) Selosta alkiotasoisien ositetun otannan tavoite ja työvaiheet. Mitä lisäinformaatiota tarvitaan perusjoukon alkioista ositettua otantaa varten? Millaisissa tilanteissa voidaan odottaa, että ositettu otanta on totaalien estimoinnissa tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR)?
- b) Kuvaile optimaalisen (Neyman) kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa. Mitä lisäinformaatiota tarvitaan perusjoukosta optimaalista kiintiöintiä varten?

4. PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- a) Selosta kumulatiivisen summan menetelmään perustuvan PPS-otannan toimintaperiaate ja työvaiheet.
- b) Selosta toimenpiteet PPS-otannan tilanteessa, jossa yhden tai useamman perusjoukon alkion kokomuuttujan z arvo tuottaa sisällysmistodennäköisyyden, joka on ykköistä suurempi. Anna esimerkki.
- c) Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- a) Selosta regressioestimoinnin tavoite, toimintaperiaate ja työvaiheet totaalien estimoinnin yhteydessä.
- b) Regressioestimaattorin \hat{t}_{reg} varianssiestimaattori voidaan SRSWOR-tilanteessa kirjoittaa muotoon $\hat{v}(\hat{t}_{reg}) = N^2(1 - n/N)(1/n)\hat{s}_e^2$, missä \hat{e} viittaa jäännöksiin. Vastaavan SRSWOR-perusteisen totaaliestimaattorin \hat{t} SRSWOR-varianssiestimaattori on $\hat{v}(\hat{t}) = N^2(1 - n/N)(1/n)\hat{s}_y^2$, missä y viittaa tulosmuuttujaan. Päättele kaavojen perusteella, miksi regressioestimoinnilla on taipumus tuottaa pienempi varianssiestimaatti kuin vastaavalla SRSWOR-perusteisella estimoinnilla. Perustele vastauksesi

6. Regressioestimointi

Tulosmuuttujan y SRSWOR-otoksesta estimoitu kokonaismäärä on $\hat{t} = 26\,440$ ja apumuuttujan z kokonaismäärän estimaatti on $\hat{t}_z = 164\,952$. Otoksesta on lisäksi saatu y :n ja z :n regressiokertoimeksi $\hat{b} = 0.152$. Perusjoukosta tiedetään apumuuttujan z kokonaismäärä; se on $T_z = 91\,753$. Konstruoi tulosmuuttujan y totaalien T regressioestimaattori \hat{t}_{reg} ja laske annetuilla tiedoilla regressioestimaatti.

Otantamenetelmät (78143), Syksy 2007

Loppuputenti 24.1.2008

Valitse seuraavista neljä tehtävää.

1. Määrittele käsitteet

- Totaaliestimaattorin \hat{t} asetelmakerroin (*Design effect*) $deff(\hat{t})$
- Sisällymistodennäköisyys (*Inclusion probability*)
- Asetelmapaino (*Design weight*).

2. Systemaattinen otanta (*Systematic sampling*)

- Selosta systemaattisen otannan poimintaproseduuri ja työvaiheet.
- Selosta implisiittisen osittamisen menetelmä systemaattisessa otannassa.
- Selosta totaaliestimaattorin \hat{t} asetelmavarianssin $V(\hat{t})$ estimointiin liittyvät ongelmat ja niiden ratkaisuvaihtoehdot systemaattisen otannan yhteydessä.

3 Ositettu otanta (*Stratified sampling*)

- Selosta alkiotasoisesta ositetun otannan tavoite ja työvaiheet. Mitä lisäinformaatiota tarvitaan ositetun otannan yhteydessä? Millaisissa tilanteissa voidaan odottaa, että ositettu otanta on totaalin estimoinnissa tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR)?
- Kuvaile optimaalisen (Neyman) kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa. Mitä lisäinformaatiota tarvitaan perusjoukosta optimaalista kiintiöintiä varten?
- Kuvaile suhteellisen kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa.

4 PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- Selosta systemaattisen PPS-otannan toimintaperiaate ja työvaiheet.
- Selosta toimenpiteet PPS-otannan tilanteessa, jossa yhden tai useamman perusjoukon alkion kokomuuttujan z arvo tuottaa sisällymistodennäköisyyden, joka on ykköstä suurempi. Anna esimerkki.
- Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- Selosta regressioestimoinnin tavoite, toimintaperiaate ja työvaiheet totaalin estimoinnin yhteydessä.
- Regressioestimaattorin \hat{t}_{reg} varianssiestimaattori voidaan SRSWOR-tilanteessa kirjoittaa muotoon $\hat{v}(\hat{t}_{reg}) = N^2(1 - n/N)(1/n)\hat{s}_e^2$, missä \hat{e} viittaa jäännöksiin. Vastaavan SRSWOR-perusteisen totaaliestimaattorin \hat{t} SRSWOR-varianssiestimaattori on $\hat{v}(\hat{t}) = N^2(1 - n/N)(1/n)\hat{s}_y^2$, missä y viittaa tulosmuuttujaan. Päätele kaavojen perusteella, miksi regressioestimoinnilla on taipumus tuottaa pienempi varianssiestimaatti kuin vastaavalla SRSWOR-perusteisella estimoinnilla. Perustele vastauksesi

Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).
Alin hyväksyty pistemäärä on 12 pistettä.

Otantamenetelmät (78143)

Tentti 12.6.2008

Valitse seuraavista neljä tehtävää.

1. Asetelmakerroin (*Design effect*)

- a) Määrittele totaalin regressioestimaattorin \hat{t}_{reg} asetelmakerroin $deff(\hat{t}_{reg})$ strategialle SRS*reg (käytä laajennettua asetelmakertoimen määritelmää).
- b) Mitä päättelet strategian SRS*reg tehokkuudesta yksinkertaiseen satunnaisotantaan (SRSWOR) verrattuna, jos totaaliestimaattorin \hat{t}_{reg} asetelmakerroin on pienempi kuin yksi? Entä jos asetelmakerroin on ykkösen suuruinen?

2. Systemaattinen otanta (*Systematic sampling*)

- a) Selosta systemaattisen otannan poimintaproseduuri ja työvaiheet.
- b) Selosta implisiittisen osittamisen menetelmä systemaattisessa otannassa. Mitä hyötyä implisiittisestä osittamisesta on?

3. Ositettu otanta (*Stratified sampling*)

- a) Selosta alkiotasoisen ositetun otannan tavoite ja työvaiheet. Mitä lisäinformaatiota tarvitaan ositetun otannan yhteydessä? Millaisissa tilanteissa voidaan odottaa, että ositettu otanta on totaalin estimoinnissa tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR)?
- b) Kuvaile optimaalisen (Neyman) kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa. Mitä lisäinformaatiota tarvitaan perusjoukosta optimaalista kiintiöintiä varten?
- c) Kuvaile suhteellisen kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa.

4. PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- a) Selosta kumulatiivisen summan menetelmään perustuvan PPS-otannan toimintaperiaate ja työvaiheet.
- b) Selosta toimenpiteet PPS-otannan tilanteessa, jossa yhden tai useamman perusjoukon alkion kokomuuttujan z arvo tuottaa sisältymistodennäköisyyden, joka on ykköistä suurempi. Anna esimerkki.
- c) Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- a) Selosta regressioestimoinnin tavoite, toimintaperiaate ja työvaiheet totaalin estimoinnin yhteydessä.
- b) Regressioestimaattorin \hat{t}_{reg} varianssiestimaattori voidaan SRSWOR-tilanteessa kirjoittaa muotoon $\hat{v}(\hat{t}_{reg}) = N^2(1 - n/N)(1/n)\hat{s}_e^2$, missä \hat{e} viittaa jäännöksiin. Vastaavan SRSWOR-perusteisen totaaliestimaattorin \hat{t} SRSWOR-varienssiestimaattori on $\hat{v}(\hat{t}) = N^2(1 - n/N)(1/n)\hat{s}_y^2$, missä y viittaa tulosuuttujaan. Päättele kaavojen perusteella, miksi regressioestimoinnilla on taipumus tuottaa pienempi varianssiestimaatti kuin vastaavalla SRSWOR-perusteisella estimoinnilla. Perustele vastauksesi

Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).
Alin hyväksytty pistemäärä on 12 pistettä.

Otantamenetelmät (78143)

Tentti 14.8.2008

Valitse seuraavista neljä tehtävää.

1. Määrittele käsitteet

- a) Totaaliestimaattorin \hat{t} asetelmakerroin (*Design effect*) $deff(\hat{t})$.
- b) Sisällymistodennäköisyys (*Inclusion probability*).
- c) Analyysipaino (*Analysis weight*).

2. Systemaattinen otanta (*Systematic sampling*)

- a) Selosta alkiotasoisien systemaattisen otannan poimintaproseduuri ja työvaiheet.
- b) Selosta implisiittisen osittamisen menetelmä systemaattisessa otannassa. Mitä hyötyä implisiittisestä osittamisesta mahdollisesti on? Perustele vastauksesi.

3. Ositettu otanta (*Stratified sampling*)

- a) Selosta alkiotasoisien ositetun otannan tavoite ja työvaiheet. Mitä lisäinformaatiota tarvitaan ositetun otannan yhteydessä? Millaisissa tilanteissa voidaan odottaa, että ositettu otanta on totaalin estimoinnissa tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR)?
- b) Kuvaile optimaalisen (Neyman) kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa. Mitä lisäinformaatiota tarvitaan perusjoukosta optimaalista kiintiöintiä varten?
- c) Kuvaile suhteellisen kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa.

4. PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- a) Selosta systemaattisen PPS-otannan toimintaperiaate ja työvaiheet.
- b) Selosta toimenpiteet PPS-otannan tilanteessa, jossa yhden tai useamman perusjoukon alkion kokomuuttujan z arvo tuottaa sisällymistodennäköisyyden, joka on ykköstä suurempi. Anna esimerkki.
- c) Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- a) Selosta regressioestimoinnin tavoite, toimintaperiaate ja työvaiheet totaalin estimoinnin yhteydessä.
- b) Konstruoi kalibrointiestimaattorin $\hat{t}_{cal} = \sum_{k=1}^n g_k w_k y_k$ g -paino g_k suhdetehosteisen estimaattorin $\hat{t}_{rat} = \hat{r} \times T_z$ tapauksessa, kun tunnetaan apumuuttujan z perusjoukon kokonaismäärä $T_z = \sum_{k=1}^N z_k$, jonka HT-estimaattori on $\hat{t}_z = \sum_{k=1}^n w_k z_k$, missä $w_k = 1/\pi_k$ on asetelmapaino ja π_k on alkion k sisällymistodennäköisyys.

Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).
Alin hyväksytty pistemäärä on 12 pistettä.