

# Helsingin yliopisto

## Painotusmenetelmät

Loppukoe Seppo Laaksonen

02.04.2009

Merkitse jokaiseen vastauspaperiin nimesi ja opiskelijanumerosi. Kaikkiin muihin paitsi viimeiseen kysymykseen kykenee tyhjentävästi vastaamaan hyvin lyhyesti. Viimeinen kysymys vastaa kahta aikaisempaa kysymystä. Siten painotetusti laskemalla kysymyksiä on yhteensä 14. Vastaa niistä maksimissaan 10:een.

1. Mikroaineiston brutto-otokseen on poimittu kaikki kehikkoperusjoukon yksiköt. Voidaanko analyysissä painot unohtaa ja jos vastauksesi on "ei", niin miksi?
2. Kotitalousaineistosta halutaan tutkia tuloeroja. Hyvän kotitalouden kokonaistulomuuttujan lisäksi aineistossa on hyvin adjustoitu otospaino sekä kunkin kotitalouden kulutusyksikköjen lukumäärä sekä kotitalouden koko. Mitä tulomuuttujaa käytät ja millaisella painolla estimoit tuloeroja?
3. Kurssilla esitettiin seuraava puuttuneisuusmekanismi "*Puuttuneisuus on satunnaista ehdollisesti MAR (Missing At Random)*". Minkälaisen otospainon (voit vastata monikossakin mutta yksikkö riittää) muodostaminen on tämän mekanismin takana ja miten tämä paino muodostetaan? Yleisperiaatteet riittävät.
4. Otannassa on käytetty (esi)ositusta ja on muodostettu tähän sopivat painot. Tulee tarve jälkiositukseen. Miten toimit?
5. Sinulla on hyvät otospainot joiden summa on tavoiteperusjoukon yritysten lukumäärä, myös ositteittain. Se ei tuota oikeita estimaatteja henkilömäärätasolla mutta asiakas haluaa sellaiset. Miten tuotat ne?
6. Tutkija on poiminut otoksen vastaamattomista. Mitä hyötyä tästä on? Konkretisoi tilannetta mahdollisimman uskottavalla esimerkillä.
7. Surveyssä käytettiin kaksiasteista otantaa. Kummassakin asteessa määriteltiin sisällysmistodennäköisyys. Miten saat luoduksi asetelmapainon? Havainnollista tilannetta jollain esimerkillä (siis mitä asteet voisivat olla ja mikä voisi olla otantamenetelmä):
8. Painot analysoitaessa kahden ajanjakson paneliaineistoa.
9. Liitteenä on luentomateriaalista kopioitu sivu jossa on kiintoisa kuvio. Miten se on tuotettu ja miten tulkitset sitä ja mihin sen sisältämää tietoa voi käyttää?
10. Miten vastauskatomalliin ja kalibrointiin perustuvasta 'määräpainosta' saadaan 'analyysipaino?' Miten näiden kahden painon vaihtelu eroaa toisistaan mitattuna vaihtelukertoimella *cv*?
11. Homogeenisten vastaajaryhmien käyttö painojen adjustoinnissa: menetelmän periaatteet ja erot kilpaileviin menetelmiin.

12. Teet logistisia regressiomalleja (tai probit-malleja) adjustoitujen painojen muodostamiseksi. Mitä huomioit mallia rakentaessasi (siis mitkä kelpaavat selittäjiksi, miten selittäjiä voidaan käyttää, mihin mallilla tähtäät, mitä varot)?

13. Alla on kaksi ohjelmaa joissa estimoidaan tuloksia PISA-datasta. Ensin Suomen koululaisista jossa käytössä ovat oppilaspainot:

```
proc surveymeans data=test mean sum; var meanscie non_native
SC09Q11; cluster schoolid; strata stratum; weight w_fstuwt; run;
```

Vastaavasti Suomen koulutason tuloste 'samoista muuttujista' mutta siis ensin aggregoituina koulutasolle, ja siis koulupainoilla:

```
proc means data=schools n mean stderr sum; var meanscie non_native
SC09Q11; weight W_FSCHWT; run;
```

Tulokset ovat alla molemmista (meanscie = luonnontieteellisen osaamisen keskiarvo kullakin oppilaalla ja non\_native tarkoittaa nimensä mukaisesti koululaista jonka taustan ollessa ei-suomalainen arvo=1, muuten=0; aggregointi koulutasolle on tehty summaamalla non\_native-muuttujalle, mutta kahdelle muulle muuttujalle keskiarvoa käyttämällä):

The SURVEYMEANS Procedure  
Data Summary

Number of Strata	12
Number of Clusters	155
Number of Observations	4714
Sum of Weights	61386.9841

Statistics

Variable	Label	Mean	Std Error of Mean	Sum
meanscie		563.322834	2.012347	34580690
non_native		0.024061	0.002922	1477.022800
SC09Q11	Full time teachers in TOTAL Q9a1	35.332474	1.035110	2108397

The MEANS Procedure

Variable	Label	N	Mean	Std Error	Sum
meanscie		155	559.8378691	2.4044177	399970.74
non_native		155	0.6738494	0.0910944	481.4252000
SC09Q11	Full time teachers in TOTAL Q9a1	151	30.3542338	1.1655328	21186.46

Tehtäväsi on tulkita tulosteiden numerot eli mitä mikin tarkoittaa? Jos jollain tuloksella ei ole järkevää tulkintaa, mainitse tämä äläkä yritä tulkita väkisin.

### Vastauskatomalliin perustuvia esimerkkejä 3

