

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Otantamenetelmät (78143) Syksy 2014

TEEMA 3 Erityiskysymyksiä
TEEMA 4 Ohjelmisto

Risto Lehtonen

risto.lehtonen@helsinki.fi





Otannan erityiskysymyksiä

■ Ryväsotanta

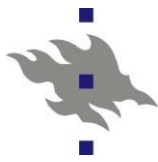
- Survey sampling reference guidelines (2008 edition)
Introduction to sample design and estimation techniques
Section 3.2.5 Cluster sampling

■ Otoskoon määrääminen

- Survey sampling reference guidelines (2008 edition)
Introduction to sample design and estimation techniques
Section 3.3 Sample size determination

■ Vastauskadon hallinnan perusteita

- Laaksonen (2013) *Surveyymetodiikka*
- Lehtonen-Pahkinen (2004) Chapter 3
- VLISS Training Key 114, 117, 123



Sisäkorrelaatio ryväsotannassa

- Samaan rypäeseen kuulumisella on taipumus samankaltaistaa alkioita tutkittavien ilmiöiden suhteen
- Koulututkimukset
 - Rypäänä opetusryhmä
 - Oppimistulokset, esim. [PISA](#)
- Työolotutkimukset
 - Rypäänä työpaikka
 - Työolot, esim. Kelan työterveyshuoltotutkimus
 - OHC data, [VLISS](#)



Sisäkorrelaatio

Sisäkorrelaatio $\hat{\rho}_{\text{int}}$ *Intra - cluster correlation*

Likimääräinen kaava $\hat{\rho}_{\text{int}} = (deff - 1) / (\bar{m} - 1)$

missä \bar{m} on keskimääräinen ryväskoko

Rypäät ovat usein positiivisesti sisäkorreloituneita

eli $\hat{\rho}_{\text{int}} > 0$ kun $deff > 1$

Lisäksi on voimassa: $deff(\hat{\rho}) = 1 + (\bar{m} - 1)\hat{\rho}_{\text{int}}$

missä $deff(\hat{\rho})$ on asetelmakerroin

(*design effect*)



Esimerkkiaineisto: Työterveyshuoltotutkimus OHC

■ Otanta-asetelma

- Ositettu yksi- ja kaksiassteinen ryväotanta
- Toimipaikat rypäinä
- Ositus rypään koon ja toimialan mukaan
 - Pienet toimipaikat: Yksiassteinen otanta
 - Suuret toimipaikat: Kaksiassteinen otanta

- Henkilötasolla likimain **itsepainottuva** (*self-weighting*) otos

■ Demonstraatioaineisto SAS-data OHC

■ Rajaus:

- Toimipaikat, joissa vähintään 10 työntekijää
- $H = 5$ ositetta (*strata*)
- $m = 250$ toimipaikkaa (ryvästä, *clusters*)
- $n = 7841$ henkilöä

- Vaihteleva määrä otosrypäitä per osite



- **Deff-estimaatit OHC**
- **(Lehtonen&Pahkinen 2004)**

Table 5.8

Averages of design-effect estimates of proportion estimates of selected groups of binary response variables in the OHC Survey data set (number of variables in parentheses).

Study variable	Mean deff
Physical working conditions (12)	6.5
Psycho-social working conditions (11)	3.3
Psychosomatic symptoms (8)	2.0
Psychic symptoms (9)	1.8



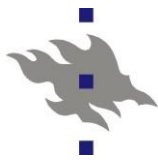
Ohjelmisto

- SAS-ohjelmisto
 - SAS-proseduurit ja SAS-makrot
- SPSS module Complex Samples
- Stata:n svy-ohjelmat
- Erikoisalue: Pienalue-estimointi (kurssi k/2012)
 - SAS-makro EBLUPGREG
 - Ohjelma Domest
- R-kieliset ohjelmat
 - Package [sampling](#) (Tillé and Matei)
 - Thomas Lumley: [Oppikirja ja R-ohjelma](#)
- [Software](#)



SAS-ohjelmisto, kuvailu

- Survey-proseduurit, joilla otanta-asetelma (ositus, ryvästyminen, painotus) voidaan ottaa huomioon estimoinnissa
- SURVEYMEANS
 - Keskiarvot ja kokonaismäärät
- SURVEYREG
 - Regressioestimointi
- SURVEYFREQ
 - Ristiintaulukointi ja perustestit



SAS-ohjelmisto, analyysi

- Survey-proseduurit
- SURVEYFREQ
 - Monipuolinen valikoima tilastollisia testejä
- SURVEYREG
 - Lineaariset mallit
- SURVEYLOGISTIC
 - Logistiset mallit
- SURVEYPHREG
 - Elinaikamallit



SAS-makrot, pienalue-estimointi

- EURAREA Project

- Asetelmaperusteinen estimointi
 - GREG-estimaattori
 - Lineaarinen kiinteiden tekijöiden malli
 - Varianssiestimaattori

- Malliperusteinen estimointi
 - EBLUP-estimaattori
 - Lineaarinen sekamalli
 - MSE:n estimaattori



Kuinka jatkaa eteenpäin?

■ Yhteiskuntatilastotiede

■ Kevät 2015

- Small area estimation (Risto Lehtonen)
- Imputointimenetelmät (Seppo Laaksonen)

■ Pro gradu yhteiskuntatilastotieteen alalta

- Tilastokeskus
- SYKE
- Kelan tutkimusosasto
- Muu valtionhallinnon tutkimuslaitos tai virasto