

Otanta-aineistojen analyysi, kevät 2010

Luennoitsija

[Risto Lehtonen](#)

Laajuus

6/8 op.

Tyyppi

78136 Otanta-aineistojen analyysi (aineopinnojen valinnainen erikoiskurssi)

78405 Otanta-aineistojen analyysi (syventävien opintojen 1. erikoiskurssi)

Tavoitteet

Kurssin tavoitteena on perehdyttää opiskelija otanta-aineistojen tilastolliseen analyysiin erityisesti tilanteissa, joissa aineisto on kerätty jollakin mutkikkaalla otanta-asetelmalla. Tällaisia ovat esimerkiksi moniasteiset otanta-asetelmat, joihin sisältyy ositus ja ryvästyminen. Mutkikkaita otanta-asetelmiä käytetään laajasti eri tieteenalojen empiirisessä tutkimuksessa. Hyviä esimerkkejä ovat PISA-tutkimussarja ja Terveys 2000 -tutkimus. Kurssilla käsitellään tilastollisia menetelmiä, joilla otanta-asetelman ominaispiirteitä (ositus, ryvästyminen, painokertoimet) voidaan ottaa huomioon tilastollisen analyysin yhteydessä. Pääpaino on ns. asetelmaperusteisissa (design-based) menetelmissä; lisäksi tarkastellaan malliperusteisia menetelmiä, ml. yleistetyt estimointiyhtälöt (GEE) ja sekamallit. Yhteensopivuus-, homogeenisuus- ja riippumattomuustestien lisäksi käsitellään lineaarisia ja logistisia malleja. Menetelmiä sovelletaan harjoituksissa ja harjoituksissa todellisiin otanta-aineistoihin. Menetelmien soveltamiseen tarkoitettuja tilastollisia ohjelmistoja esitellään (SAS, SPSS, Stata, R). Kurssi on luonteeltaan soveltava. Kurssi voidaan sisällyttää valtiotieteellisen tiedekunnan [menetelmäkoriin](#).

Luentoajat

Luennot tiistaisin 2.2.–2.3.2010 klo 14–18 C323 (Exactum), yhteensä 20 tuntia.

Lisäksi harjoituksia mikroluokassa C128 torstaisin 28.1.–4.3.2010 klo 12–15, yhteensä 15 tuntia.

HUOM: To 25.2. ei ole demoja, viimeiset ovat 4.3.

Suoritustapa

Aineopinnot: Loppukuulustelu (6 op) tai loppukuulustelu ja (vapaaehtoinen) harjoitustyö (8 op)

Syventävät opinnot: Loppukuulustelu ja (pakollinen) harjoitustyö (8 op)

Loppukuulustelu

Tiistai 16.3.2010 klo 14–16 C323 (Exactum),

Oppikirja

[Lehtonen R. and Pahkinen E.](#) (2004). Practical Methods for Design and Analysis of Complex Surveys. Second Edition. Chichester: John Wiley & Sons. Luvut 5, 7, 8.

Web-materiaali

VLISS-virtual laboratory in survey sampling,

<http://mathstat.helsinki.fi/VLISS/>

LINEAARISET JA EPÄLINEAARISET MALLIT SAS-OHJELMISTOSSA

	Lineaariset mallit	Epälineaariset mallit	
		Logistiset	Poisson-mallit
Tulosmuuttuja	Jatkuva	Binäärinen tai moniluokkainen	Lukumäärämuuttuja
Selittäjät diskreettejä	Lineaarinen ANOVA	Logit-ANOVA	Poisson-ANOVA
Selittäjät jatkuvia	Lineaarinen regressio	Logit-regressio	Poisson-regressio
Diskreetit ja jatkuvat selittäjät	Lineaarinen ANCOVA	Logit-ANCOVA	Poisson-ANCOVA
SAS-proseduurit	REG SURVEYREG GENMOD MIXED GLIMMIX	LOGISTIC SURVEYLOGISTIC GENMOD GLIMMIX	GENMOD GLIMMIX
Estimointi	LS WLS WGEE/EXCH ML ja REML	ML PML WGEE/EXCH ML ja REML	ML PML WGEE/EXCH

LS – Pienimmän neliösumman menetelmä (Least Squares)

WLS – Painotettu LS (analyysipainot mukana)

GWLS – Yleistetty painotettu LS (Generalized weighted LS)

ML – Suurimman uskottavuuden menetelmä (Maximum Likelihood)

PML – Pseudo-likelihood (analyysipainot mukana)

GEE – Yleistetyt estimointiyhtälöt (Generalized Estimating Equations)

GEE-EXCH – GEE- Exchangeable-korrelaatorakenne

WGEE – Painotettu GEE (analyysipainot mukana)

REML – Rajoitettu ML (Restricted ML)