

78450 SOSIAALITUTKIMUKSEN TILASTOLLISET MENETELMÄT, kevät 2012

Luentojaksojen 1-6 kuvaukset

1. Tilastotiede ja yhteiskunta

17.–18.1.2012 Hannu Niemi

Jaksolla tarkastellaan tilastojen roolia yhteiskunnan eri ilmiöiden kuvaajana. Tarkastelunäkökulma on Tilastokeskuspainotteinen. Keskeisenä lähtökohtana on tilastojen laatu, ts. tilastojen on oltava mahdollisimman luotettavia ja niiden on annettava oikea kuva yhteiskuntaoloista ja niiden kehityksestä. Tietojen vertailukelpoisuuden varmistamiseksi tilastoja laadittaessa on mahdollisuuksien mukaan käytettävä yhdenmukaisia käsitteitä, määritelmiä ja luokituksia. Aluksi lyhyt historiallinen katsaus virallisen tilastotoimen kehityksestä.

Keskeinen lähde:

Vesa Kuusela: "Paradigms in Statistical Inference for Finite Populations; Up to the 1950s". (Saatavana myös e-thesis -versio: <https://helda.helsinki.fi/handle/10138/27416>.)

Esimerkkinä samaa yhteiskunnallista ilmiötä kuvaavista tilastoista käytetään Suomen EU:hun liittymisen myötä tapahtunutta muutosta työttömyyden tilastoinnissa ja työvoimaviranomaisten tuottamia työttömyystilastoja.

Keskeiset lähteet:

Työttömyys ja työllisyys tilastoissa. Tilastokeskuksen ja Työministeriön tilastojen vertailua. Tilastokeskus 2003. http://www.stat.fi/til/tyti/tyti_2005-05-27_men_002.pdf

2. Surveymetodiikka aineiston keruusta sen puhdistamiseen

24.–25.1.2012 Seppo Laaksonen

Jaksolla käydään läpi surveyaineiston peruskäsitteet alkaen perusjoukoista ja aineiston puutteista. Edelleen käsitellään keruutapoja ja -välineitä sekä lomakkeisiin liittyviä peruskysymyksiä. Jatkossa kerätään aineisto, usein otantaa hyväksi käyttäen. Aineisto ei välttämättä ole tallennuksenkaan jälkeen valmis, vaan tarvitaan laadun parantamista ja muita ns. puhdistavia toimenpiteitä joihin liittyvät myös meta- ja paradatan hyvä käsittely. Jakson lopulla on käsitelty menetelmät ja välineet joiden jälkeen ollaan valmiit aineiston analyysiin.

Kirjallisuus: Seppo Laaksonen, Surveymetodiikka (Bookbon.com 2011), Jelke G. Bethlehem, Applied survey methods: a statistical perspective. John Wiley and Sons, 2009 -375 sivua

Pääaiheet (jokaiseen kohtaan tulee esimerkkejä sekä suomalaisista että monikansallisista surveyistä)

1. Surveyaineiston perusjoukot ja muut peruskäsitteet sekä seuraavia osa-alueita:
 - Poikkileikkaus- ja pitkittäisaineisto
 - Ylipeitto
 - Yksikkövastauskato
 - Erävastauskato
 - Otos- ja muut painot
 - Apumuuttujat ja tulosmuuttujat
 - Metadata ja paradata
2. Keruuvälineet ja lomakkeet
 - Posti-, käynti-, puhelin-, netti- sekä monivälinekyselyt
 - Lomakesuunnittelun periaatteet

3. Otanta-asetelmat ilman tarkempaa tekniikkaa joka tulee enemmän esille jaksolla 3
 - Kokonaisasetelma otannalle
 - Otoksoon määrittely ja efektiivinen otoskoko
 - Innovatiivisiakin otanta-asetelmia voi tehdä eikä vain esimerkiksi yksinkertaista satunnaisotantaa
 - Otantatiedosto ja sen liittäminen varsinaiseen tiedostoon

4. Aineiston puhdistamisen menetelmiä
 - Tilastollinen editointi
 - Korvatako puuttuvaa tietoa imputoimalla vai ei ja jos niin miten voisi imputoida?
 - Perusotospainojen luonti
 - Johdatus parempiin otospainoihin
 - Muut vaiheet puhdistamisen täydentämiseksi (IT ratkaisut ym.)

3. Otanta-asetelmat ja tilastollinen analyysi

31.1.–1.2.2012 **Risto Lehtonen**

Jaksolla annetaan kuva tilastollisen otannan roolista yhteiskuntatieteiden sovellusalojen tutkimusprosessissa, erilaisista otantamenetelmistä ja niiden käytöstä sekä otanta-asetelmalla kerätyn aineiston tilastollisen analyysin perusteista. Menetelmiä valaistaan esimerkeillä, joissa käytetään aitoja tutkimusaineistoja. Pääasiallinen esimerkkiaineisto tulee PISA-tutkimussarjasta. Alan tilastollisia ohjelmistoja (SPSS, SAS, R) esitellään.

Kirjallisuus: Pahkinen-Lehtonen, Otanta-asetelmat ja tilastollinen analyysi (Gaudeamus 1989), Lehtonen-Pahkinen, Practical Methods for Design and Analysis of Complex Surveys (Wiley 2004), Laaksonen, Surveymetodiikka (Bookbon.com 2011) ja erikseen ilmoitettavat verkkomateriaalit (Tilastokeskuksen laatukäsikirja 2007 http://www.stat.fi/meta/gg_2ed.pdf).

Pääotsikot

1. Tutkimusaineistojen lähteitä
 - Rekisteriaineistot ja tietoarkistot
 - Itse kerätyt otosperusteiset kyselyaineistot
 - Eri aineistotyyppien yhdistelmät
 - Esimerkkejä: Tilastokeskuksen työvoimatutkimus ja Eurooppalainen elinolotutkimus (EU-SILC)

2. Otoksen poiminta
 - Otanta tutkimusasetelman osana
 - Otantamenetelmien perusteita
 - Eri otantamenetelmiin liittyvä estimointi
 - Asetelmakerroin ja sisäkorrelaatio
 - Esimerkkejä: PISA-tutkimuksen otanta-asetelmat

3. Aineiston analysointi
 - Analyysivaiheessa huomioon otettavia seikkoja
 - Riippumattomuudesta, t-testi ja otanta-asetelma
 - Regressioanalyysi ja otanta-asetelma
 - Esimerkkejä: PISA-tutkimus

4. Kyselytutkimusaineistojen tiivistäminen ja visualisointi

7.-8.2.2012 **Kimmo Vehkalahti**

Jakson tavoitteena on oppia tiivistämään ja visualisoimaan kyselytutkimusaineistoja tilastollisilla menetelmillä ja ohjelmistoilla. Kyselytutkimusaineistot ovat tyypillisiä yhteiskuntatieteellisessä tutkimuksessa, jossa mielenkiinto kohdistuu mm. asenteisiin, arvoihin ja mielipiteisiin. Näitä moniulotteisia ilmiöitä mitataan kysely- ja haastattelulomakkeilla. Taustalla on yleensä tutkimuskysymyksistä johdettu mittausmalli, joka ohjaa myös aineiston analysointia lähtien tiivistämisestä ja päättyen erilaisiin visualisointeihin ja yksityiskohtaisempiin analyysihin.

Kyselylomakkeen laatimisen tilastollisia näkökohtia kuten muuttujien mittaustasoa sekä validiteettia ja reliabiliteettia on käsitelty Tilastotieteen johdantokurssin osassa 1. Aihepiiriä täydentävät tämän luentokurssin edeltävien jaksosten sisällöt. Tällä jaksolla perehdytään mittausmallia apuna käyttäen aineiston tiivistämiseen faktorianalyysillä. Lisäksi tarkastellaan aineistojen visualisointia tutustumalla korrespondenssianalyysiin ja hierarkkiseen ryhmittelyyn.

Esimerkkiaineistoja ovat mm. Yhteiskuntatieteellisen tietoarkiston tutkimusaineistot, jotka ovat osin tuttuja johdantokursuilta. Luennoilla asioita esitellään eri tilastollisten ohjelmistojen näkökulmista, mutta pääpainona on menetelmien oppiminen.

Pääotsikot

1. Mittaus kyselytutkimuksissa
 - Mittausmalli
 - Mittauksen laatu
2. Aineiston tiivistäminen
 - Oletukset ja rajoitukset
 - Faktorianalyysi ja sen tulkinta
 - Faktoripiste- ja summamuuttujat
3. Aineiston visualisointi
 - Korrespondenssianalyysi
 - Hierarkkinen ryhmittely

Kirjallisuus: Vehkalahti, Kyselytutkimuksen mittarit ja menetelmät (Tammi 2008), Pennings-Keman-Kleinnijenhuis, Doing Research in Political Science: An Introduction to Comparative Methods and Statistics (Sage 2006).

5. Regressioanalyysi ja trendimäisten muuttujien muunnokset

14.–15.2.2012 Pekka Pere

Regressiomalleista:

- Yhden ja monen selittäjän regressiomalli.
- Regressio odotusarvoa kohti.

Aikasarjoista:

- Trendimäisten muuttujien logaritointi ja differenssointi.
- Log-muutokset suhteellisina muutoksina.

Empiirisiä esimerkkejä.

Kirjallisuutta:

- B. Thompson (2006): *Foundations of Behavioral Statistics. An Insight Based Approach*. Guilford Press. (Kaavoja kaih-taville.)
- M.C. Baddeley ja D.V. Barrowclough (2009). *Running Regressions*. Cambridge University Press. (Empiirisesti painot-tunut ja helppolukuinen muttei täysin luotettava.)
- J.M. Woolridge (2012): *Introductory Econometrics. A Modern Approach, 5. laitos*. Southwestern College. (Edellisiä teoreettisempi. Paljon talous- ja yhteiskuntatieteellisiä empiirisiä esimerkkejä.)
- J.D. Angrist ja J.-S. Pischke (2009): *Mostly Harmless Econometrics. An Empiricist's Companion*. Princeton Univer-sity Press. (Täynnä yhteiskuntatieteilijöitä kiinnostavia kysymyksenasetteluja ja analyyseja. Kausaalisuuden poh-dintaa.)

6. Varianssianalyysi ja ei-parametriset menetelmät

21.–22.2.2012 Jyrki Möttönen

Jaksolla opetellaan varianssianalyysin teoriaa ja käytäntöä sekä tulkitaan analyysien tuloksia. Klassisten normaalijakaumaole-tukseen perustuvien varianssianalyysimenetelmien lisäksi tutustutaan niiden ei-parametrisiin vastineisiin. Ei-parametrisissa menetelmissä ei tehdä tiukkoja oletuksia havaintojen jakaumasta, joten niitä voi käyttää myös tilanteissa, joissa klassisten menetelmien jakaumaoletukset eivät päde. Menetelmien ominaisuuksia havainnollistetaan erilaisilla käytännön esimerkeillä. Esimerkkiaineistoina käytetään etupäässä KvantiMOTV:n harjoitusaineistoja.

(<http://www.fsd.uta.fi/menetelmaopetus/SPSS/aineistot.html>)

Kirjallisuus: Hollander, M. & Wolfe, D. A. (1999). *Nonparametric Statistical Methods*. 2. laitos. Wiley: New York, Montgomery, D. C. (2005). *Design and Analysis of Experiments*. 6. laitos. Wiley: New York.

Pääotsikot:

- Kahden riippumattoman otoksen t-testi
- Kahden otoksen Wilcoxonin testi
- Yksisuuntainen varianssianalyysi
- Kruskalin-Wallis testin testi
- Parittaiset vertailut
- Oletusten ja hypoteesien testaus.