



Surveymetodiikkaa ja työkaluja haasteisiin

Seppo Laaksonen: *Survey Methodology and Missing Data. Tools and Techniques for Practitioners*. Springer 2018.

Emeritusprofessori Seppo Laaksonelta ilmestyi syksyn 2018 aikana kaksi teosta, joista tässä kirja-arvostelussa tarkasteltava on englanninkielinen *Survey Methodology and Missing Data. Tools and Techniques for Practitioners*. Toinen taas on Helsingin yliopiston valtiotieteellisen tiedekunnan julkaisusarjassa ilmestynyt teos *Surveymetodiikan nousu. Henkilöhistoriallinen katsaus vuosilta 1985–2018*. Kummatkin teokset antavat laajan näkymän surveymetodiikkaan tilastotieteen osana. Jälkimmäisen saa ilmaiseksi Helsingin yliopiston Heldasta, edellinen on kaupallinen ja saatavissa myös yliopiston kirjastosta.

Kirjoittaja määrittelee laajasti kohteensa. Surveymetodiikka on tilastotiedettä, mutta myös yhteiskuntatieteitä mukaan lukien taloustiede ja sosiologia, käyttäytymistieteitä ja sovellusmielessä jopa humanistisia. Se on myös tietotekniikkaa, matematiikkaa ja kognitiivista psykologiaa. Tämä määritelmä antaa myös hyvän kuvan kirjoittajan lähestymistavasta. Tarkoituksena ei ole ainoastaan metodinen näkökulma kysely- ja haastattelumenetelmään vaan paljon laajempi lähestymistapa. Tämä näkyy erityisesti kirjan monipuolisissa esimerkeissä, aineistovalinnoissa ja kirjoitustavassa, jolla vaikeita asioita on havainnollistettu lukijalle.

Teoksen voi nähdä jatkona Laaksonen aiemmin vuonna 2010 suomenkielellä julkaisemalle *Surveymetodiikka*-kirjalle ja sen uusintapainokselle vuodelta 2013. Tekijän päätutkimusalueilta kirja sisältää uusinta mahdollista tietoa. Suomenkieliset kirjat ovat olleet käytössä kirjoittajan yliopistolla pitämällä eri surveymetodiikan kursseilla. Nyt julkaistu kirja soveltuu erityisesti surveymetodiikan opiskelun alkutaipaleella oleville, mutta osin se vaatii syvällisempää osaamista. Tätä varten kirjaan on koottu jokaisen luvun lopuksi oma viiteluettelo. Toisaalta jo hieman surveymetodiikkaan perehtyneen tai pidemmälläkin olevalle kirja tarjoaa muun muassa tiiviin paketin tärkeimmistä käsitteistä.

Esimerkit ja aineistot

Kirjassa on paljon esimerkkejä, joissa on käytetty oikeita tilastoaineistoja. Esitystapa ei ole riippuvainen tilasto-ohjelmasta eikä varsinaista ”nappitekniikkaa” esitellä lainkaan. Esimerkeissä käytetään muun muassa European Social Surveyn (ESS), World Values Surveyn (WVS) ja koululaisia koskevia PISA-aineistoja. Lukijan kannalta aidoilla aineistoilla tehdyt esimerkit auttavat ymmärtämään syvällisemmin havainnollistettavia menetelmällisiä asioita. Lisäksi mainitut surveyaineistot ovat vapaasti ladattavissa kirjautumisen jälkeen omaan käyttöön. Tulevaisuudessa kirjoittaja ehkä voisi harkita esimerkkien jakamista esimerkiksi verkkosivun tai Github-tyyppisen ratkaisun kautta.

Haasteellinen puuttuva tieto

Kirjassa, aivan aiheellisesti, korostuu puuttuvan tiedon rooli erityyppisissä aineistoissa. Sillä riippumatta siitä, miten aineisto on kerätty, puuttuva tieto on käytännössä aina läsnä, usein kiusallistavalla tavalla. Laaksonen kirjassa puuttuvan tiedon huomioiva näkökulma on jo mukana, kun aineiston keruuta suunnitellaan. Esimerkiksi esiteltäessä keskeisiä

termejä, on ollut tarpeen määrittellä jopa viisi perusjoukkoa, joista tulokset estimoidaan tavoiteperusjoukon tasolle. Lukijalle korostetaan myös ennakkosuunnittelun ja käytettävien rajoitusten määrittelyn tärkeyttä. Puuttuva tieto (vastaukset) voi olla joko selästä, että ennalta valittu vastaaja ei osallistu ollenkaan eli ei anna yhtään vastausta. Tällaista katoa kutsutaan yksikkövastaukskadoksi. Toisaalta vastaaja voi antaa osaan kysymyksistä vastauksensa, jolloin kyseessä erävastaukset. Kummankinlaiseen katoon voi varautua ennakolta. Kirjassa on käsitelty esimerkiksi lomakkeen suunnittelun yhteydessä keinoja, joilla voidaan auttaa vastaajaa antamaan vastauksensa. Toisaalta puuttuviin tietoihin voi varautua keräämällä hyviä apumuuttujia, joiden arvot ovat tiedossa niin vastanneille kuin vastaamattomille.

Ylipäänsä kirjassa korostuu ennakoinnin tärkeä merkitys eli kaikki se työ, mikä tapahtuu ennen varsinaisen aineiston keräämistä. Kadon huomioivista menetelmistä kirjassa esitellään laajasti painotusta ja imputointia eli puuttuvan tiedon korvaamista yksilötasolla. Aikaisemmin mainitut esimerkkien surveyaineistot eivät sovellu sellaisenaan imputointiin ja tämän vuoksi käytössä on erillinen EU-projektin tuloja koskeva aineisto. Lisäksi on omistettu oma luku tilastolliselle editoinnille, joka on merkittävä toimenpide ennen aineiston varsinaisten analyysien alkua.

Toiseksi viimeinen luku sisältää yhteenvedon aineiston luomisesta ja puhdistamisesta. Viimeisessä luvussa johdatellaan surveyaineiston analyysiin siten, että analyysi tehdään oikein ottaen huomioon erityisesti osituksen, aineiston mahdollisen ryvästyksen ja parhaiden mahdollisten painojen käytön. Tuloksena saadaan laadukkaat piste-estimaatit ja niiden tarkkuutta mittaavat luottamusvälit tai mediassa käytetyt virhemarginaalit. Ala on vaikea eikä laadukkaita tuloksia ole helppo saada.

Kaikkien surveyaineistoja käyttävien olisi hyvä olla perillä kirjassa esitetyistä perusteista.

Kohti yhä monipuolisempia aineistoja

Vaikka tilastotiede on kehittynyt ja erilaiset laskentamahdollisuudet ovat aivan toista luokkaa kuin aiemmin, on surveyaineistojen käyttäjillä edelleen haasteita. Eriytyisesti puuttuneisuus on pahentunut vuosi vuodelta ainakin kehittyneissä maissa. Lisäksi uudet tiedonlähteet ja erilaisten aineistojen yhdistelmät luovat aivan uudenlaisia haasteita, mutta myös mahdollisuuksia. Kirjan korostama ennakoiva tapa reagoida mahdollisiin haasteisiin ennen varsinaisen aineiston keruun alkua mahdollistaa ongelmien ratkaisua tilastotieteen tarjoamilla keinoilla. Kirjan soisi olevan jokaisen surveyaineistoa keruuta suunnittelevan ja käyttävän käsikirjana.

MARIA VALASTE

Kirjoittaja on Helsingin yliopiston yliopistolehtori ja soveltavan tilastotieteen dosentti, joka on erikoistunut surveymetodiikkaan.