



Tutkijan hallittava otostutkimuksen perusteet

Otos pohjaisia kyselytutkimuksia on Suomessa tehty enenevästi sodan jälkeen. Alkuaikoina, jopa 1980-luvulle asti, mukaan poimitut vastasivat hyvin. Sen jälkeen vastauskato ja muu puuttuneisuus on kehittyneissä maissa lisääntynyt huomattavasti.

Tutkimus on vastannut ongelmaan kehittämällä kyselylomakkeita ja kyselytapoja, joissa muun muassa digitalisointi on auttanut. Silti puuttuneisuutta ei ole saatu aisoihin. Toinen, jälkikäteen keino on käyttää aputietoa sekä otoksesta että vastaajista, ja tiedon avulla oikaista harhaa niin paljon kuin suinkin. Harhattomiin tuloksiin ei silti päästä, mutta kiusaa saadaan pienennettyä.

Viime aikoina tuntuvat sellaiset tutkimukset yleistyvän, joissa otannan ja puuttuneisuuden perusteista ei välitetä. Tämä tuskin on ”tietentahtoista”, vaan kyse on osaamisen puutteista. Ihmeen yleistä on esimerkiksi korostaa vastaajien määrää. Jos se on riittävän iso, lukijalle vakuutetaan tutkimuksen olevan laadukas. Jostakin saattaa löytyä tieto vastaajien suhteellisesta osuudesta. Jos se on matala, vaikkapa 10–30 prosenttia, tämä tieto jää sivurooliin. Pelkään että useimmat lukijat valitettavasti uskovat tuloksiin.

Tuloksen laatu pitäisi sen sijaan todistaa eroilla vastaajien ja vastaamattomien välillä. Tätä varten tarvitaan mainitsemieni aputietoja, jotka tietysti tulee kerätä kaikista mahdollisista lähteistä koko aineiston keruun aikana (ks. suomeksi esim. Laaksonen 2013). Toisaalta eroa voidaan kuroa umpeen noiden tietojen avulla, muttei kokonaan täyttää.

Lukuisissa nykyisin raportoiduissa selvityksissä ei siis eroteta otosta näytteestä, mitä vastaajien joukko on, se ei siis ole otos. Tutkijat suhtautuvat niihin kuin ne olisivat samanlaisia. Useimmiten emme pelkästään raporttia lukemalla tiedä, mitä analysoitu aineisto edustaa. Muutta mutkitta yleistetään näytettä koskevat havainnot ta-

voiteperusjoukkoon, vaikkei tätäkään selkeästi sanota. Tyypillistä on myös, että ongelmasta vaietaan. Tutkijoiden pitäisi raporteissaan keskustella perusteellisesti siitä, kuinka hyvin saadut tulokset edustavat tavoiteperusjoukkoa, kertoen rehellisesti erityisesti edustavuuden ongelmat.

Havainnollistan puuttuneisuus- ja edustavuuskysymystä tässä lehdessä julkaistun kuuden puolueen jäseniä koskevan tutkimuksen avulla (Koiranen ym. 2017). Olen lukenut myös laajemman raportin, mutta se ei ole tältä osin parempi.

Tutkimusraportissa kerrotaan vastaajien määrät, mitkä ovat lähellä laajan raportin liitteessä mainittuja, jonka mukaan vastaajia on yhteensä 12 290. Varsinaisissa tulostaulukoissa esitetään tuloksia puolueiden jäsenistä esimerkiksi iän, sukupuolen ja asuinmaakunnan mukaan, vaikka kyse on vastaajista. Ne voivat poiketa jäsenistöstä hyvinkin paljon, mutta tästä en löydä mitään tietoa.

Tutkimus perustuu ositettuun yksinkertaiseen satunnaisotantaan, jossa kunkin kuuden puolueen jäsenistö on osite. Keskimääräinen otantasuhde on pieni, noin 5,7 prosenttia, mutta vaihteli suuresti puolueittain. Tältä osin laadussa ei ole mitään vikaa, mutta ongelmat havainnollistuvat jo matalasta keskimääräisestä 24,1 prosentin vastausasteesta. Se vaihteli melkoisesti puolueittain. Parhaiten vastasivat Vasemmistoliiton (35,3) ja Perussuomalaisten (32,1) jäsenet, huonoiten Keskustan (17,3) ja Kokoomuksen (19,0). Tämä ei toki kerro suoraan, kuinka hyvin tuloksia puolueiden välillä voidaan verrata.

On mielenkiintoista, että tutkijat ovat havainneet ikäharhan todeten, että ”Iäkkäämmät puoluejäsenten ovat aineistossa hieman aliedustettuina, minkä vuoksi olemme käyttäneet ikäjakaumavirheen korjaavaa painokerroinmuuttujaa”. Tämä voi olla oikeansuuntainen, mutta yksityiskohtia painotuksesta ei ole. Se kertoo kuitenkin että heillä on ollut tiedossa ikätietoa koko jäsenistöstä ja mitä luultavimmin myös sukupuolen mukaan.

Jos olisi haluttu tehdä tuloksiin yksinkertainen perusoikaisu, olisi puolueittaisen ikäryhmä- ja sukupuolitiedon avulla saatu helposti muodostettua aito otospaino. Tämä olisi varmasti oikaissut julkaistuja tuloksia huomattavasti, ikäryhmää ja sukupuolta koskien täydellisesti. Luulen, että muutakin tietoa olisi ollut olemassa ja harhaa saatu pienennettyä lisää.

Tutkijoiden raportti on monilta osin ansiokas. Olen ymmärtänyt, että vaivalla saatua aineistoa aiotaan käyttää jatkotutkimuksessa. Tämä on ihan oikein. Toivottavasti sinne nyt luodaan mahdollisimman paljon harhaa pienentävät otospainot.

Tätäkin kiintoisampi kysymys voisi olla selvittää vastaamattomien poikkeamaa vastaajista. Onko esimerkiksi niin, että vastaajat ovat ikään kuin ”eliittinä” ja vastaamattomien enemmistö sellaisia, jotka eivät ole tyytyväisiä puolueeseen tai politiikkaankaan enää? Miksi tämä joukko on noin iso vaikka jäsenten luulisi olevan aktiivisempia kuin tavallinen kansa? Puoluejäsenten vastausastehan alittaa huomattavasti tason, mihin tavallisissa kyselytutkimuksissa Suomessa päästään, joissa se alimmillaan on 60 prosenttia. Tämä osoittaa, ettei kyselyn tiedonkeruuta toteutettu hyvin.

Lähteet

- Koiranen, I., Koivula, A. ja Saarinen, A. (2017). Puoluejäsenet suurennuslasin alla. Jäsenten sosiaalinen rakenne. *Tieteessä tapahtuu* 2, 11–18.
- Laaksonen, S. (2013). *Surveymetodiikka. Aineiston kokoamisesta puhdistamisen kautta analyysiin*. Ventus Publishing. 230 s. <http://bookboon.com/fi/surveymetodiikka-ebook>

SEPPO LAAKSONEN

Kirjoittaja on Helsingin yliopiston tilastotieteen professori.