

rustuu eri julkaisujen artikkeleihin (*Irish Times*, *Guardian*, *New York Review of Books*), reagoi välittömästi brexit-saagan kaikkiin vaiheisiin. Se on vahvaa draamaatiikkaa; asiantuntevaa, oppinut-ta ja ironista. Tässä kirjailija riisuu brutaalisti – raemmin kuin

Lennon – pienen eliitin taistelu purkaa pykäläviidakoksi kutsuttuun asioita: ympäristön, sosiaalisten olojen ja työvoiman oikeuksien turvaa. Tämän saavuttamiseksi Johnson on valmis luomaan uuden kansakunnan, Englannin (ei unionissa Skot-

lannin kanssa, ei unionissa Pohjois-Irlannin kanssa), tietämättä, mitä tämä kansakunta on. Ketterämpi? Singaporelaisempi?

Säälin Englantia, Ergo: En ehkä kaikesta huolimatta olekaan niin huono historiijoitsija. Ymmärrän, että Englantia on sääli.

Paksusti perinteisellä tyylillä

PERTTI TÖTTÖ

Olen käyttänyt Lauri Nummenmaan *Käyttätymistieteiden tilastolliset menetelmät* -teosta oppimateriaalina liki kahden vuosikymmenen verran. Vuosien varrella kirja on painosten myötä paisunut ja lukijoiden palautteesta parantunut. Nyt siinä on 727 sivua, ja se on saanut uuden nimen. Se on kaikin puolin edeltäjänsä parempi.

Valitsin teoksen aikanaan sillä perusteella, että siinä oli kaavoja, kuvioita ja tekstiä sopivassa suhteessa. Menetelmiä ei esitetty liian tilastotieteellisesti. Vaikka uuden version nimi on *Tilastotieteen käsikirja*, tämä piirre on säilynyt.

Siksi hieman ihmettelenkin nimen valintaa. Tilastotieteen oppikirjaksi tästä tuskin on, mutta tilastomenetelmien käsikirjaksi se käy (melkein) alasta riippumatta.

Yhdessä suhteessa nimenmuutoksen voi ymmärtää. Aikaisemmassa versiossa sovel-lusesimerkit esitettiin SPSS:n avulla, mutta nyt on siirrytty käyttämään R-ohjelmistoa. R on nykyään nimenomaan tilastotieteen *lingua franca*, jo-

ta myös osa tilastomenetelmien soveltajista jo käyttää. Siirtymisen tähän ilmaiseen työvälineeseen säästäisi yliopistoilta ja tutkimuslaitoksilta pitkän pennin, mutta ohjelmiston komentokiel-ten omaksuminen ei valikoihin tottuneelle ole helppoa. Kirjassa on johdatus R:ään sekä jokaisen analyysimenetelmän ja tilastollisen testin yhteydessä ohjeet, miten tempu tehdään R:llä, joten kyllä se siitä.

Sivunumerot vilisevät tähän tahtiin:

Lyhyet luvut tilastomenetelmien tarpeesta, tieteellisen tutkimuksen perusteista, mittaamisesta ja R:n perusteista. Ensimmäinen sata sivua käytetty. Aineiston kuvailu, grafiikan käyttö, lyhyt johdatus todennäköisyyslaskentaan. Teoreettisten todennäköisyysjakaumien alkutaipaleella täyttyy toinen satanen. Tilastollisten testien yleisperiaatteiden ja t-testin jälkeen rikkoutuu jo kolmas.

Yksi- ja useampisuuntainen sekä toistomittausten varianssianalyysi ja tiivis luku epäparametrisistä testeistä sekä kaikki mahdolliset korrelaatioker-

Lauri Nummenmaa
Tilastotieteen käsikirja
Tammi, 2021

toimet vievät viidennen satasen puolelle. Yhteensopivuustestien ja χ^2 -riippumattomuustestien jälkeen seuraa yllättävän lyhyt luku lineaarisesta regressiosta, aikasarja-analyysin perusteet, elin-aika-analyysi, ja niin viisisataa on täynnä.

Sitten palataan mittauksen reliabiliteettiin ja vähän validiteettiinkin, parikymmentä sivua standardoiduista efektikoon estimaateista, hyvä luku meta-analyysista, ja kuusisataa täyttyy otsikolla Tilastollinen luokittelu, jolla tarkoitetaan logistista regressiota (!).

Ryhmittelyanalyysi, moniulotteinen asteikointi, pääkomponenttianalyysi ja faktorianalyysi tuovat maalisuoralle, jossa esitellään lyhyesti rakenneyhtälömallit. Sivulta 685 alkaa lähdeluettelot, joka myös päättyy samalle sivulle, ja sen jälkeen on nippu liitteitä.

”Kattava ja ajantasainen”, todetaan kirjan takakannessa. Onhan se, mutta voisi olla kattavampikin.

Missä on lineaarinen todennäköisyysmalli, ordinaalinen regressio, latentti luokitteluanalyysi, missä kasvukäyrämallit, monitasomallit ja sekamallit? Konfirmatorinen faktorimalli vain vilahtaa R-koodien esittelyssä, vaikka se on hyvin keskeinen väline yhdistettyjen mittareiden psykometrinen ominaisuuksien testaamisessa.

Nummenmaa lähestyy tilastomenetelmiä samalla perinteisellä tyyllillä, jolla minunkin sukupolveni on niitä opiskellut. Menetelmiä tarkastellaan erillään, omissa kappaleissaan ja omilla nimillään.

On olemassa toinenkin lähestymistapa, jossa niitä katsotaan integraation näkökulmasta. Varianssianalyysin ja lineaarisen regression teoriaa ei tarvitsisi esittää sadan sivun erottamana toisistaan, koska matemaattisesti kyse on yhdestä ja samasta lineaarisesta mallista. Ja logistiseen regressioon päästäisiin muuttamalla lineaarisen mallin yhtäläisyysmerkki logit-linkiksi. Faktoriansalyysikin on lineaarinen malli, jossa useita mitattuja muuttujia selitetään yhdellä tai useammalla latentilla muuttujalla. Ja kaikki nämä voidaan esittää rakenneyhtälömallien erikoistapauksina.

Tästä näkökulmasta lineaarisen regression ymmärtäminen on kaiken a ja o. Siksi olenkin hieman pettynyt sen käsittelyyn.

Lineaarisen regression yhteyteen olisin kaivannut muuttujien keskistämistä, interaktiotermin käyttöä sekä monen selittäjän mallin kerrointen tulkin-taan valottavaa tekstiä. Nyt jää vaikutelma, että selityksasteen R^2

kasvattaminen olisi ainut peruste laittaa malliin useampia selittäjiä. Ennustemallissa niin ehkä on, vaikka Daniel Kahneman ja kumppanit kyllä väittävät uudessa kirjassaan *Noise* (2021), että ennustamisessakin yksinkertainen lineaarinen malli toimii yhtä hyvin kuin monimutkainen. Jopa kertoimien estimointi on heidän mukaansa usein turhaa; niiden sijasta voi käyttää yksisiä.

Toisin kuin Nummenmaa kirjoittaa, epälineaaristen riippuvuuksien mallintaminen ei ole lineaarisen regression avulla hankalaa. Toisen asteen termit ja interaktiot hoitavat sen mallikkaasti.

F-testin kuvaaminen ikään kuin se olisi niin sanottu fittiluku (”mallin voidaan katsoa sopivan aineistoon”) on harhaanjohtavaa. Regressiomallissa ei ole vapausasteita, joten se sopii täydellisesti aineistoon. F-testi kertoo vain, onko R^2 tilastollisesti nolasta poikkeava.

Ehdottomasti väärin on sanoa, että standardoimattomat regressiokertoimet ”ilmaisevat, kuinka paljon kukin x-muuttuja selittää y-muuttujan vaihtelusta”. Siksi neuvo, että jos jonkin muuttujan kerroin ”on lähellä nollaa (...) se voidaan poistaa mallista”, on virheellinen. Iän kerroin on yleensä lähellä nollaa, ja sukupuolen kerroin siihen nähden moninkertainen, koska edellinen mittaa yhden vuoden ja jälkimmäinen ”sukupuolenvaihdoksen” efektiä y-muuttujaan, kun kaikki muut mallin muuttujat pysyvät muuttumattomina. Silti iän selitysosuus saattaa olla sukupuoleen nähden moninkertainen. Tässä on sotkettu standardoitujen ja standardoimattomien kerrointen tulkinta.

Valitettavaa kohinaa on

myös reliabiliteetin käsittelyssä. Spearman-Brownin kaavan avulla voidaan estimoida, kuinka kyselylomakkeen tai muun monesta osiosta koostuvan mittauksen reliabiliteetti kasvaa, kun mittariin lisätään tai siitä poistetaan itemeitä”. Heti perään tulevasta kaavasta ja laskuesimerkeistä näkee, että osioiden poistaminen ei koskaan kasvata reliabiliteettiä. Tässä on sotkettu reliabiliteetti ja Cronbachin alfa, jonka laskukaava taas sallii sen, että mittarin lyhentäminen voi nostaa kertoimen arvoa.

Tästä seuraa seikka, jota olisi hyvä korostaa: Alfa on huono reliabiliteetin estimaatti, eikä sen nimittäminen ”sisäisen johdonmukaisuuden” mittariksi ole johdonmukaista, koska se voi käyttäytyä reliabiliteetin käsitteeseen nähden epäjohdonmukaisesti.

Perinteiseen tyyliin kuuluu, että tilastollisia testejä kumarretaan syvään. Jokaisen aineistoosi-merkkinsä lopuksi Nummenmaa ohjeistaa, miten analyysin tulos on syytä raportoida. Kaava kuuluu näin: ”Saatiin tilastollisesti merkitsevä” jokin (suluisia testi-suure, vapausasteet ja p-arvo).

On kauan tiedetty ja viime aikoina empiirisesti osoitettu, että tilastollisen merkitsevyyden taivoittelu johtaa julkaisuharhaan ja p-arvohakkerointiin. Koska julkaisut haluavat ”merkitseviä” tuloksia, tutkijat valitsevat analyysistaan ne, joiden tulokset on saatu onnekkaiden tähtien alla (***, **, *). Julkaisuissa raportoitujen, rajan 0,05 alittaneiden p-arvojen jakaumaa tutkimalla on voitu osoittaa, että suuri osa ”merkitsevänä” julkaisuista testituloksista täytyy olla vääriä positiivisia. Biostatistikko John Ioannidixsen aihetta käsittelevän artikkelin (2005) otsik-

ko kuuluikin: ”Why most published research findings are false.”

Vuonna 2016 American Statistical Association antoi suosituksen p-arvojen käytöstä. Siinä sanotaan: ”P-arvo ei itsessään tarjoa hyvää mittapuuta mallia tai hypoteesia koskevalle näytölle”. Vuonna 2019 *Nature* puoles-

taan julkaisi yli 800 tutkijan allekirjoittaman vetoomuksen, ettei p-arvoja käytettäisi ”sen ratkaisemiseen, kumoaako vai tukeeko saatu tulos tieteellistä hypoteesia”.

Kumpikaan kannanotto ei tarkoita p-arvojen hylkäämistä, ainostaan kategorisesta ”tilastol-

lisesta merkitsevyydestä” luopumista (niin ajatuksissa, sanoissa kuin teoissa). Suositusten mukainen raportointikaava kuuluisi ehkä: Saatiin efekti (sulussa testisuure, vapausasteet ja p-arvo), minkä jälkeen efektin suuruus arvioidaan nojaamalla p-arvoon.

Virstanpylväs

MATTI VIRTANEN

Maria Forsmanin kirja on erinomainen kuvaus ammatti-ihmisen 43-vuotisesta nousujohteisesta urasta. Selkeä tyyli ja omaehtaisuus, välillä rankkojenkin tunnekokemusten kirjaaminen luo lukijoille vertailu- ja samastumiskohteita, ja tekee tekstistä elävää ja vetävää.

Alkuvuonna 1972 Maria sai yllättävän tarjouksen tulla sijaiseksi Helsingin yliopiston valtiotieteellisen tiedekunnan Franzeniassa sijaisevaan kirjastoon. Silloin sosiologiaa pääaineenaan lukevat opiskelijat koki pääsevän unelmaduuniin, tieteelliseen kirjastoon. Nimikkeeksi tuli ”ylimääräinen tilapäinen järjestelyapulainen”, työaika 28 tuntia viikossa. Palkka ei ollut suuren suuri, mutta se tuli tilille joka kuukausi.

Siitä alkoi Forsmanin pitkä ura kirjastolaitoksen eri tehtävissä: Franzenian jälkeen kuusi vuotta Suomen Akatemian informaattikkona, sitten seitsemän vuotta Tampereen yliopiston kirjastotieteen ja informatiikan laitoksen assistenttina, 13 vuotta samassa yliopistossa keskus-

kirjastonhoitajana, kuusi vuotta samassa yliopistossa Yhteiskuntatieteellisen tietoarkiston informaattikkona ja sitten paluu Helsingin yliopiston valtiotieteellisen tiedekunnan kirjastoon sen johtajaksi vuonna 2005 ja lopulta yhdistyneen Helsingin yliopiston kirjaston johtavaksi tietoasiantuntijaksi uudessa Kaisatalossa. Eläkkeelle Forsman jäi vuonna 2015.

Forsman antaa hyvän kuvan tieteellisten kirjastojen työmenetelmien kehityksestä, varsinkin 1990-luvulla, kun kehittyvä internet alkoi mullistaa tiedonhakuja. 2000-luvulla mukaan tuli kasvava yhteistyö tutkijayhteisöjen kanssa, kun tutkimusten määrän ja laadun arviointi alkoi nousta keskeiseen asemaan yliopistojen rahoituksesta päätettäessä.

Forsmanista kasvoi vähitellen taitava ja monipuolinen informaattikko, mutta toinen keskeinen motiivi oli mukana alusta alkaen. Forsmanin ensimmäisiä työtehtäviä tilapäisenä järjestelyapulaisena Franzenian kirjastossa vuonna 1972 oli koota luette-

Maria Forsman
Kirjavia muistoja
Elämäni kirjastojen ja tieteen
maailmassa
Enostone, 2020

lo kirjastossa olevista painetuista kirjallisuusluetteloista eli bibliografioista ja luokitella niitä aiheen mukaan. ”En ollut ennen tiennyt, että on olemassa kirjoja, joihin on koottu tietoja olemassa olevista kirjoista ja lehtiartikkeleista! Ja nyt minä kokosin niistä luetteloa eli tein bibliografianbibliografiaa.”

Bibliografiat herättivät Forsmanissa uinuvan tutkijan: ne eivät ole vain luetteloja, ne kertovat myös tieteiden kehityksestä ja niiden painopisteiden ajallisista muutoksista. Tämän perusoivalluksen siivittämänä Forsman vuonna 1984 visioi: ”Eilen sain idean kirjastossa tutkiskellessani Pippingin iki-ihanaa klassikkoa. Minä teen väitöskirjaksi tutkimuksen SUOMALAISESTA BIBLIOGRAFIASTA suomalaisessa yhteiskunnassa ja kulttuurissa.” Fred-