

Biostatistiikka — työväline ja tutkimusaihe

Sanalla tilastotiede on suorastaan maaginen vaikutus tapaamiini ihmisiin. Kun kerron heille olevani tilastotieteilijä, saan vastaukseksi nyökyttelyä, jäykistyneitä hymyjä ja puheenaiheen vaihtoja. Kun kerron olevani tilastotieteen professori, voisin yhtä hyvin toivottaa näkemisiin. Selvitäksemme kilpailussa vaikkapa parhaista opiskelijoista tai tutkimusrahoituksesta koenkin tarvitsevani uutta sanastoa. On huomattavasti kätevempää kertoa olevan biostatistiikan tai -metrian, numeronmurskauksen tai laskennallisen lääketieteen professori. Kuulija haluaakin tietää lisää. Pienten mutta oleellisten ensimmäisten sanavalintojen jälkeen pääsenkin kertomaan, että biostatistiikalla tarkoitetaan tilastotieteen soveltamista bio- ja lääketieteissä. Mallintamisen keinoin voidaan empiiristä tutkimusaineistoa käyttäen selvittää ilmiöiden syntymekanismeja, syitä ja seurauksia sattuman roolia unohtamatta. Voin viehättää todennäköisyyslaskennan ystäviä riskin käsitteellä tai pyrkimyksillä hallita sattumaa koesuunnittelulla, tai sen osuuden arvioinnilla. Ohjelmoinnista kiinnostuneille haasteita riittää mitattavien aineistojen käsittelyssä ja laskennassa. Biostatistikkona pääsee jatkuvasti käsiksi aineistoihin, joilla on suurta kansanterveydellistä merkitystä. Tyypin 1 diabeteksen syntymekanismien selvittäminen tai eturauhassyövän seulontavälien ja -rajojen optimointi ovat esimerkkejä tutkimuskysymyksistä, joissa aineistot ovat massiivisia, kansanterveydellinen vaikuttavuus suorastaan valtavaa ja tilastollinen mallintami-

nen voi tuoda aivan keskeistä tietoa ja ymmärrystä ilmiöstä. Tämän sanottuani on oikeastaan yllättävää, miten pieni tieteenalamme Suomessa on. Suuntaa antaa se, että yliopistoissamme on tyypillisesti kahdesta neljään tilastotieteen professoria. Esimerkiksi kanadalainen kollegani työskentelee ranskankielisessä kauppakorkeakoulussa Montrealissa. Heillä on yhdeksän tilastotieteen professuuria. Puhumattaakaan brittikollegoistani: London School of Hygiene and Tropical Medicineen Department of Medical Statistics työllistää kahdeksan tilastotieteen professoria. Toki yliopistojen mittakaavat ovat kokonaisuuskin erilaiset, mutta näin suuret yhteisöt luovat aivan toisenlaiset puitteet menetelmätieteiden opetukselle ja tutkimukselle.

Tilastotieteen asema menetelmätieteenä erityisesti kliinisen lääketieteen tutkimuksessa on vahva. Johtavat tieteelliset julkaisukanavat edellyttävät modernien tilastollisten menetelmien käyttöä ja vaatimustaso on yhä korkeampi menetelmällisen vertaisarvioinnin yleistyessä. Oma mielenkiintoni on viime vuosina suuntautunut varsin paljon ravitsemusepidemiologian sovellusalueelle. Tutkimusaihe on mielenkiintoinen paitsi siksi, että ravitsemus on äärimmäisen vaikea ja moniselkoinen mitattava, myös siksi, että aihe herättää niin paljon voimakkaita tunteita. Kaikilla meistä on uskomuksia siitä, mitä on terveellinen ruokavalio. Aihe on jatkuvasti julkisuudessa. Useimmat meistä ovat esimerkiksi motivoituneita uskomaan, että suklaasyönti on terveellistä, vaikka tosiasiaa tutkimuksissa olisikin vain huomattu suklaasyöjien joukon olevan muuta väestöä terveempää. Tämän kaltaisten kansantajuisuuden nimissä tehtyjen loogisten päättelyvirheiden korjaamisen lisäksi tiedonhallinnan

ja tilastollisen mallintamisen osalliselle on ravitsemusepidemiologiassa melkoinen kysyntä. Päivittäin syödyn ruuan kirjaaminen päiväkirjoihin on työstä ja siinä on satunnaisia ja systemaattisia mittausrvirheitä. Edelleen tällaisten tietojen purkaminen esimerkiksi raaka-ainetasolle tai ravintoaineisiin on valtava urakka. Olettamat lineaarisista ravintotekijöiden yhteyksistä kliiniseen tautiin ovat epätodennäköisiä. On luultavaa, ettei samanlainen ravinto sovi kaikille, vaan ravinnolla, liikuntatottumuksilla ja genetiikalla on yhdysvaikutuksia. Listaa voisi jatkaa. Ravintotietojen liittäminen taudin syntyprosessiin on siten äärimmäisen haastava ja mielenkiintoinen tehtävä. Se vaatii tilastollisen mallinnuksen huippuosaamista; innovatiivisia ratkaisuja reikäisen ja mittausrvirheistä kärsivän altisteaineiston hallitsemiseksi yhdessä sekoittavien ja muovaavien tekijöiden kanssa. Tarvitaan dimensio- ja riippuvuuden mallintamista. Tarvitaan ilmiöiden rinnakkaista mallintamista. Monimutkaisista ilmiöistä ja malleista huolimatta on samalla välttämätöntä pystyä vastamaan yksinkertaiseen ja klassiseen kysymykseen: voisivatko tulokset johtua sattumasta. Tulosten toistettavuus on informaatiotulvan keskellä yhtä tärkeää, ellei jopa tärkeämpää kuin aiemmin. Tämän kaltainen riskinarviointi on vaikeaa ja parhaillaan likimääräistä, mutta samalla kansanterveydellisesti välttämätöntä. Ravinnolle jokainen meistä altistuu päivittäin ja on varmaa, että sillä on keskeistä merkitystä tautien synnyssä, vaikkakin se on huonosti ymmärrettyä. Monialaisen tutkimusryhmän kanssa tilastotieteen tutkimuskin on mielestäni parhaimmillaan; se on motivoivaa, on innostavaa kehittää uusia räätälöityjä ratkaisuja ja kirjoittaa menetelmäjulkaisuja sovel-

lusjulkaisujen rinnalla. Välillä tutkijan voi kuitenkin vallata epäusko. Julkisuudessa eri alojen asiantuntijoiden vuosien yhteistyön tulokset romuttuvat kovin helposti kun ne rinnastetaan aivan kuin samanarvoisena todisteena ammattiautoilijan kertomukselle siitä, kuinka hän ei huimien D-vitamiinin yliannosten jälkeen ei enää ole kärsinyt lainkaan flunssasta. Periksi ei silti voida antaa.

Nykyaikaiset mittausmenetelmät mahdollistavat massiivisten aineistojen keräämisen, mutta niiden ymmärtäminen on yhä vaikeampaa. Kokeellisissa olosuhteissa voidaan mm. rekisteröidä ihmisen eri elintoimintoja yksittäisen sydämenlyönnin tasolla ja tutkia mitä tapahtuu koehenkilön altistuessa erilaisille ärsykkeille. Empiiriseen aineistoon perustuvat tilastolliset mallit elintoiminnoille auttavat ymmärtämään ihmiskehon toimintaa ja voivat toimia perustana uusien hoitomuotojen kehittämiseksi. Uudenlaisten aineistojen synty jatkuvina signaaleina, pikseleinä tai vokseleina on asettanut tilastotieteen menetelmien kehitykselle haasteen. Peruskurssien lähtökohta normaali-jakaumasta, täydellisestä aineistosta ja riippumattomista havainnoista toteutuu käytännössä hyvin harvoin. Tietokoneiden tallennus- ja laskentakapasiteetin kasvu on ollut huimaa. Näistä syistä tilastotieteen tutkimus on jatkuvassa murroksessa. Muutos on tieteenalallemme uskomaton mahdollisuus, johon tarttuminen vaatii meiltä kaikilta rohkeutta, uudistumishalua ja kovaa työtä.

Tilastotieteen tutkijoiden profiili vaihtelee sovelluslähtöisistä tutkijoista hyvinkin teoreettisiin tutkijoihin ja voi olla kaikkea siltä väliltä. Kumpaakin tutkimusta tarvitaan. Toisaalta tarvitaan sovelluslähtöisesti nopeita ja oikeita vastauksia ajankohtaisiin kysymyksiin, joilla voi olla poliittista, kansanterveydellistä tai taloudellista merki-

tystä. Pelkkä aineistojen kerääminen on voimavarojen väärinkäyttöä, niiden puutteellinen hyödyntäminen keskinkertaisuutta. Meidän tulee pyrkiä erinomaisuuteen ja kohdentamaan suurimmat voimamme ilmiöiden syvälliseen ymmärtämiseen. Ilman kunnianhimoista aineistojen analysointia tämä ei ole mahdollista. Se puolestaan vaatii osaamista ja voimavaroja. Toivottavasti löydämme keinoja ylittää tieteenalojen rajoja menetelmätieteistä sovellusaloille ja päinvastoin saavuttaaksemme tämän tavoitteen. Tilastotieteen näkökulmasta tutkimus ei voi kuitenkaan olla pelkästään soveltavaa. Tarvitaan myös oman tieteenalan identiteetti, joka kokoaa yhteisömme yhteen. Teoreettisen tilastotieteen tutkimusta (menetelmäkehitystä ja menetelmien teoreettisten ominaisuuksien tutkimusta) on aina ohjannut sovellusalojen ongelmat. Tänä päivänä esimerkiksi ns. omics-aineistot ovat saaneet liikkeelle tilastotieteen tutkimusta aineiston mittasuhteiden muuttuessa siten, ettei tehdäkään enää mittauksia suurissa potilasjoukoissa vaan valtavasti mittauksia per potilas. Funktionaalisen data analyysin tutkimusta motivoi yksittäisten havaintopisteiden sijaan signaaliluonteinen mittausteknologia. Laaja-alaisesti sovellettava menetelmä tutkimus voi auttaa meitä ymmärtämään enenevässä määrin kasautuvaa tietoa paremmin ja on keskeisessä roolissa kun halutaan kasvattaa Suomeen eturivin tilastotieteilijöitä hyödyttämään koko tiedeyhteisöä.

Läketieteen ja sen lähialojen tutkimus ja yritystoiminta tarjoaa mielenkiintoisen mutta vaativan työskentelyalan tilastotieteen asiantuntijoille. Näiden asiantuntijoiden koulutuksessa Turun yliopisto on jo nyt maassamme yksi parhaista. Jatkuvana haasteenamme on pieni opiskelijamäärä. Menestyksen on mahdollistanut täsmennetty koulutusohjelma, kansainvälinen opetta-

javaihto ja läheinen yhteistyö tulevien työnantajien kanssa. Kysyntä vastavalmistuneillemme on tarjontaa suurempaa. Opiskelijoiden pro gradut syntyvät useimmin yhteistyöhankkeina yritysten ja tutkimuslaitosten kanssa, ja he työllistyvät jo ennen valmistumistaan. Tästä mallista meidän kannattaa pitää kiinni myös tulevaisuudessa ja jalostaa sitä edelleen. Menestyminen asiantuntijaroolissa vaatii teoreettista ja käytännön tilastollisen mallintamisen osaamista sekä rutkasti itseluottamusta ja kykyä itsenäiseen tiedonhankintaan. Huipulle tähtäävältä tilastotieteilijältä edellytetään ohjelmointitaitoja, vuorovaikutustaitoja ja halukkuutta oppia uutta niin lääketieteellisestä aihealueesta kuin muuttuvasta menetelmäkentästäkin. Tilastotieteen koulutuksen jatkuva uudistaminen ja kehittäminen niin pääaine-, sivuaine- ja jatko-opetuksenkin osalta on tämän vuoksi koko kvantitatiivista tietoa hyödyntävän yhteisön kannalta ensiarvoisen tärkeää. Pienten yksiköiden voimin tämä ei onnistu, vaan tarvitaan elinvoimaisia keskuksia, jollainen Turkuun on syntymässä. Turun yliopiston tilastotieteen keskus tarjoaa tulevaisuudessa kansallisesti poikkeuksellisen hyvät puitteet niin sovelluslähtöiselle kuin teoreettiselle tilastotieteen koulutukselle ja tutkimukselle. Tarvitsemme ennen kaikkea lahjakkuutta, kunnianhimoa ja paloa työntekoon. Tilastotieteen keskus haluaa edustaa tätä kaikkea. Minusta meidän tulee pyrkiä edelläkävijän asemaan tilastotieteen opetuksen ja monialaisen tutkimuksen sarjoilla. Meillä on siihen harvinainen mahdollisuus.

JAAKKO NEVALAINEN
Professori
Turun yliopisto
Yhteiskuntatieteellinen
tiedekunta
Tilastotieteen oppiaine