



Kirja-arvio: Lääkkeet tilastolliseen lukutaidottomuuteen

Kimmo Vehkalahti

Yhteiskuntatieteiden menetelmäkeskus, Helsingin yliopisto

Tilastoilla on kuulemma helppo valehdella. Kokeillaan:

Roskapostia lähetetään 90 000 000 000 viestiä päivässä.

Uskotko? Vai toivotko? Vai epäiletkö tietoa valeeksi? Miten roskapostien lukumäärä voidaan muka tietää? Ja vielä noin tarkasti? Aika uskomatonta: tasan 10 nollaa! Entä jos oikea määrä onkin 90 123 456 789? *Biljardi?!*

Lääkkeet, tilastot — ja dinosaurukset

”Uusi *Lyhyet johdatukset* -tiedekirjasarja tiivistää noin 150 sivuun keskeisien aiheiden perustiedot – sen, mitä kaikkien pitäisi tietää.” (Art House -tiedote 2.6.2017)

Jos jotain kaikkien pitäisi tietää, niin se, että tilastot eivät valehtele. Sen sijaan tilastot valaisevat tietä kohti totuutta, jota tieteellä tavoitellaan.

Sellaisenaan tilastot ovat kasvottomia lukuja – tyljä tietoja, joihin sisältyy monenmoisia epävarmuuksia. Tietojen keruuta, mittausta, kuvailua, luokittelua, järjestystä tai rekisteröintiä ei pysty tekemään virheittä.

Tilastollisesti lukutaidoton voi sortua tilastoilla valehteluun, mutta on turha syyttää valehtelusta tilastoja.

Tilastolukutaitoinen erottaa faktat höpöhöpöstä, käsittelee suuret ja pienet luvut, päättelee prosenteista, tutkii taulukoita, tulkitsee lukuja ja suhtautuu kiinnostuneen kriittisesti kuvioihin, joilla tilastoja visualisoidaan.



Tilastotieteen professori *David Handin* teos *Statistics: A Very Short Introduction* on suomennetun tiedekirjasarjan numero 2 – ykköseksi pääsi *Lääkkeet*, ja kolmanneksi tulee *Dinosaurukset*. Kova kolmikko kyseessä!

Alkuperäinen Oxfordin yliopistopainon sarja käsittää jo lähes 600 teosta elämän eri aloilta – Aleksanteri Suuri, algebra, arkeologia, islam, Kafka, molekyylibiologia, monikulttuurisuus, musiikki, planeetat, psykologia, termodynamiikan lait, terrorismi, yliopistot jne. On kiinnostavaa nähdä, mitä niistä suomennetaan!

Ensimmäisten joukossa suomennettu *Tilastot* on alan arvostetun professorin sujuvaa ja helposti ymmärrettävää tekstiä, joka esittelee tilastojen ja tilastotieteen maailmaa monista kiinnostavista näkökulmista. Roskaposti ja muut käytännön esimerkit ovat havainnollisia ja auttavat näkemään, miten mielenkiintoista tämä usein kuivaksi ja tylsäksi luultu tieteenala oikeasti on.

Kirja valottaa tilastojen ja tilastotieteen merkitystä maailmassa, datan keräämisen ja kuvailun mahdollisuuksia, todennäköisyyden olemusta, tilastollisen päätelyn perusteita, tilastollisia malleja ja menetelmiä sekä tilastografikkaa ja tilastollista tietojenkäsittelyä. Keräta kaikkiaan kiinnostava kooste!

Suomentaminen voi olla haasteellista

Kirjan kääntämisessä englannista suomeksi on ollut pieniä haasteita, joissa olisi ollut kiva neuvoa kääntäjää. Jo kirjan nimen olisi voinut valita toisin. Valitettavasti englannin kielessä sana ”statistics” on kovin monikäsitteinen, koska se tarkoittaa sekä tilastotiedettä, tilastoja että tilastollisia tunnuslukuja. Parempi nimi kirjalle olisi *Tilastotiede*. Tätä puoltaa sekin, että suomennetuissa alkusanoissa ei mainita kertaakaan tilastoja, mutta sana tilastotiede esiintyy lähes 20 kertaa.

Toisaalta voi myös ajatella, että tilastoilla viitataan *dataan*, jonka merkitys on koko ajan rajussa kasvussa. Puhutaanhan myös *datatieteestä*, alasta, joka on viime vuosina tuonut tilastotieteen ja tietojenkäsittelytieteen yhä lähemmäksi toisiaan. Kannattaa myös huomata, että alkuperäinen teos on ilmestynyt jo 2008, juuri ennen sosiaalisen median (Twitter, Facebook jne.) läpimurtoa. Some synnyttää koko ajan valtavasti sellaista dataa, jota möyhennetään datatieteen keinoin.

Tietysti tämäkin osoittaa vain, kuinka tilastotiede on dynaaminen tieteenala, jonka on vastattava aikakautensa haasteisiin. Se ei voi jäädä paikalleen, koska sen kehitys riippuu voimakkaasti reaali maailman ilmiöiden kehityksestä. Tämän viestin professori Handin kirja välittää jo nyt hienosti, mutta on selvää, ettei teos voi kailta osin pysyä ajan tasalla. Hyvät perustiedot alan tähän astisesta kehityksestä se joka tapauksessa tarjoaa.

Joitakin yksittäisiä käsitteitä on myös käännetty niin, etteivät alan asiantuntijat niitä tunnista. Hassuin esimerkki on ”vinoutumaton”, jonka tilalle kuuluisi täysin vakiintunut ”harhaton”. Moiset yksityiskohdat eivät tietysti uuden lukijan kokonaiskuvaa haittaa; enemmän ne pistävät silmään termeihin tottuneelle lukijalle.

Yllä mainituista pienistä huomioista huolimatta kirja on erinomainen valinta uuden tiedekirjasarjan perus-

teokseksi, ja sitä voi helposti suositella kenelle tahansa. Aion ottaa sen oheislukemistoksi myös omalle, kaikille avoimelle yhteiskuntatilastotieteen (verkko)kurssilleni, jonka ensimmäisen osan kantava teema on juuri tilastollinen lukutaito.

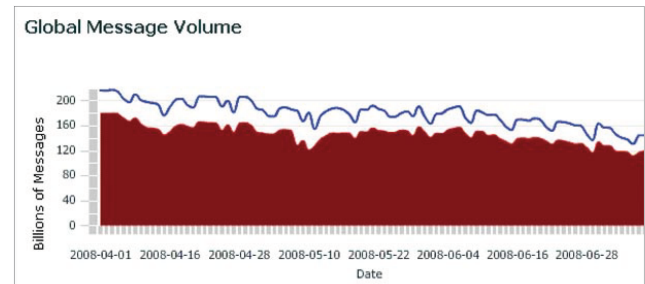
Pienistä huomioista suuriin lukuihin

Miten alussa esitetty iso luku pitäisi lukea? (Aikanaan ehdotettiin (oikeasti!), että se luettaisiin ”90 *biljardia*”, mutta johan siitä vitsaili Aku Ankkakin).

Suurten luonnollisten lukujen, joiden kymmenkantainen logaritmi on jaollinen kolmella tai kuudella, nimet noudattavat Suomessa *pitkää asteikkoa*, jossa jokainen *-iljoona*-loppuinen lukusana on 1 000 000 kertaa suurempi kuin edellinen. Poikkeus on 1 000 miljoonaa tarkoittava *miljardi*. Sekaannuksia on vaikea välttää, kun eräissä muissa maissa on käytössä *lyhyt asteikko*, jossa kerroin on 1 000 ja miljardia vastaakin *biljoona*. Siinä on itse kunkin tilastollinen lukutaito koetuksella.

Paluu roskapostin ääreen

Roskapostin määrä kasvaa kokoajan, joten alussa esitetty *estimaatti* (eli tilastoihin perustuva arvio) ei pidä paikkaansa, kun luet tätä. Ei se pitänyt paikkaansa enää silloinkaan, kun kirjoitin tämän, koska tieto oli 10 vuoden takaa. (*Joten tilasto valehtelee, niinkö?*)



Kuvan lähde: www.securecomputing.com

Viitteitä eteen- ja taaksepäin

arthouse.fi

courses.helsinki.fi/fi/valt-103

fi.wikipedia.org/wiki/Roskaposti

fi.wikipedia.org/wiki/Suurten_lukujen_nimet

veryshortintroductions.com