

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Högmander, Harri

Title: Oudokki, verrokki, riski ja moni muu : uusi tilastotieteen sanasto

Year: 2022

Version: Published version

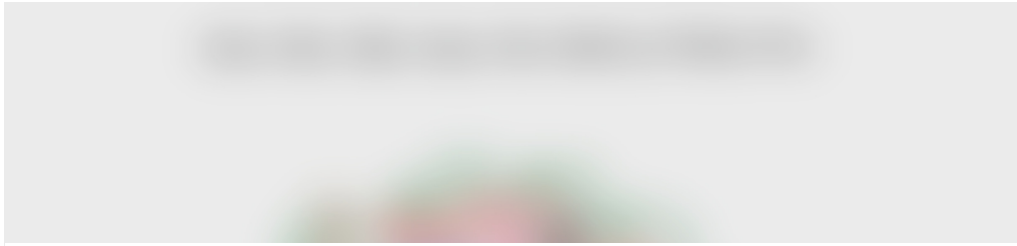
Copyright: © 2022 Tieteellisten seurain valtuuskunta

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Högmander, H. (2022, 40). Oudokki, verrokki, riski ja moni muu : uusi tilastotieteen sanasto. Tieteessä tapahtuu, (2). <https://www.tieteessatapahtuu.fi/numerot/2-2022/oudokki-verrokki-riski-ja-moni-muu-uusi-tilastotieteen-sanasto>

[Etusivu](#) > [Numerot](#) > [Numero 2/2022](#)> [Oudokki, verrokki, riski ja moni muu – Uusi tilastotieteen sanasto](#)

NUMERO 2/2022

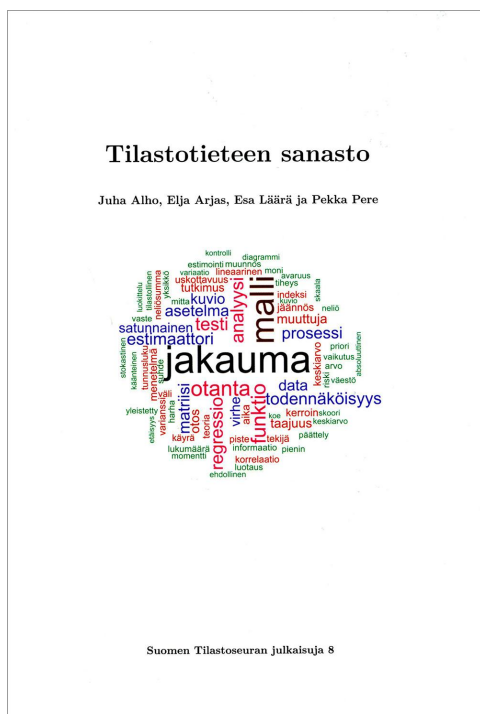
KIRJALLISUUS

Oudokki, verrokki, riski ja moni muu – Uusi tilastotieteen sanasto

Jyväskylän yliopiston tilastotieteen lehtori Harri Högmander arvioi sanaston, johon on koottu suomenkielisiä tilastotermejä. "Uusi tilastosanasto on niin perusteellinen ja perusteltu kooste, että sen termistön soisi tulevan alalle uudeksi standardiksi", hän kirjoittaa.

Harri Högmander

20.4.2023



Juha Alho, Elja Arjas, Esa Läärä ja
Pekka Pere

Tilastotieteen sanasto

Suomen Tilastoseuran julkaisuja 8,
2021



Arvioitu lukuaika: 5 minuuttia

Virhemarginaali, keskiarvo, mediaani, kausivaihtelu, liukuva keskiarvo – monet tilastotieteen käsitteet ovat arkisessa käytössä tiedotusvälineissä, ja jopa yleensä aivan oikein ja ilman termien selittämistä.

Tieteellisissä teksteissä tilastotieteen peruskäsitteiden käyttö on vielä moninaisemmin arkipäiväistä, onhan tilastotiede välttämätön aputiede kaikilla empiiristä numeerista aineistoa käyttävillä tieteenaloidella. Termit, kuten korrelaatio, luottamusväli, regressiosuora, normaalijakauma, otos, toisto, verrokki ja useat muut, esiintyvät minkä tahansa alan tutkimusraporteissa, ja kirjoittajat voivat olettaa lukijoidensa ymmärtävän, mitä niillä tarkoitetaan.

Alalta on aiemmin julkaistu kaksi pohjoismaista tilastosanastoa, joista uudempi on vuodelta 1975. Tilastotieteen käyttäjäkunta on laajentunut jatkuvasti koko tieteenalan historian ajan. On selvää, että puolessa vuosisadassa tilastotiede on tieteenalana laajentunut huomattavasti kuten myös tilastomenetelmiä soveltavien tutkimusalojen joukko.

Vastaavasti uusia tilastotieteen haaroja on syntynyt, ja siksi alalle on tarvittu ja luotu uusia käsitteitä. Vaikka englannin kielen käyttö laajenee lähes kaikilla tieteenaloilla, suomenkielisten käsitteistöjen luominen ja ylläpitäminen on arvokasta jo yksin kulttuurisista.

Uusi *Tilastotieteen sanasto* vastaa kehityksen tuomaan haasteeseen erinomaisesti. Kirjoittajina on neljä tilastotieteen konkaria: **Juha Alho**, **Elja Arjas**, **Esa Läärä** ja **Pekka Pere**. Kirjan on kustantanut Suomen Tilastoseura, kuten aiemmatkin tilastosanastot. Sanasto oli tarkoitus julkaista seuran satavuotisjuhlassa 2020, mutta sekin tapahtuma jäi pandemian jalkoihin.

Kun vuoden 1975 tilastosanastossa oli noin 800 termiä, lukumäärä uudessa sanakirjassa on peräti nelinkertainen. Oli siis jo korkea aika yhtenäistää sanastoa ja osin keksiä uuttakin.

Termien evoluutiota

Metoditieteenä tilastotiede sijoittuu monen tieteenalan leikkaukseen. Siten tilastotieteen sanaston on oltava sopusoinnussa matematiikan, erityisesti stokastiikan, ja tietotekniikan – missä datatiede on ollut kasvavan kiinnostuksen kohteena tällä vuosisadalla – termien kanssa. Samoin yhteneväisyys tilastotieteen omien vahvojen haarojen, kuten ekonometrian ja epidemiologian, kielenkäytön suhteen on tärkeää.

Tietenkin me tilastotieteen opettajat ja käyttäjät olemme vuosikymmenten mittaan päätyneet keksimään omia suomennoksia uusille, useimmiten englanninkielisille, termeille. Osa näistä on vähitellen vakiintunut yleiseen käyttöön, osa taas on jäänyt yksittäisten laitosten tai opettajien omiksi yritelmiksi.

Kirjan sanasto-osassa on käännökset englannista suomeen ja suomesta englantiin, ilman termien määrittelyjä. Lisäksi kirjassa on luvut sanastotyön periaatteista, erityisen pulmallisista termeistä, tilastotieteen (sanaston) suomalaisesta historiasta ja vertailusta viron tilastosanastoon. Tämä johdanto-osa on nautinnollista luettavaa kaikille tilastotieteen, kielten ja tieteenhistorian ystäville.

Suomenkielisen tieteellisen kirjoittamisen aamunkoitossa 1800-luvun puolivälissä olisi saatettu kirjoittaa, että ”valdelmassa käytellään todenvaiheita ja vaarinottoja”. Ymmärsitkö? (Nykykielinen käännös kerrotaan lopussa.)

Maltillinen fennomaani, filosofian tohtori **Paavo Tikkanen**, joka oli ensimmäisen suomenkielisen sanomalehden *Suomettaren* avainhahmo, kehitti tilastotiede-sanon vuonna 1859. Hän perusti sen asiain tilaan (*status*) eikä valtion numeerisen aineiston keruuseen (*statistik-johdannaiset*). Tämän valinnan mahdollinen yhteys nousevaan suomalaiseen kansallistuntoon on kiintoisa erityisesti näinä Venäjän varjostamina aikoina.

Ongelmallisesti käännettäviä termejä

Survey on hyvä esimerkki hankalista käännettävistä termeistä. Tästä kertoo sekin, ettei keskeiselle käsitteelle ole aiemmin syntynyt suomennosta vaan englannin sanaa on käytetty suomenkielisissä esityksissä sellaisenaan, vaikka erityisesti sen taivuttaminen on kömpelöä.

Kirjoittajat ehdottavat *surveyn* käännöksi sanoja ”luotaus” (substantiivi) ja ”luodata” (verbi). Valintojen monista perusteluista poimittakoon seuraava yllättävä yhteys: ensimmäiset Kuuta tutkineet avaruusaluukset olivat *avaruusluotaimet Surveyor 1–7*.

sa skoori kuulostaakin ensin oudolta ja enemmän porilaiselta purjehdustermiltä kuin tieteelliseltä käsitteeltä, tilastotieteilijälle se kertoo heti, mistä on kyse.

Toinen vaikea käännettävä tilastotieteen termi on *score*. Sama käännösongelma on ollut monessa muussakin kielessä, ja näissä on usein päädytty ”vääntämään” englantia omankieliseksi sanaksi (esimerkiksi norjassa *skåre* ja virossa *skoor*). Kirjoittajat ovat päätyneet vastaavaan ratkaisuun ja käännökseen ”skoori”. Vaikka sana kuulostaakin ensin oudolta ja enemmän porilaiselta purjehdustermiltä kuin tieteelliseltä käsitteeltä, tilastotieteilijälle se kertoo heti, mistä on kyse.

Bootstrap-menetelmät ja Bayes-tilastotiede ovat tulleet merkittäväksi osaksi tilastotieteilijöiden työkalupakkia sitten edellisen tilastosanaston julkaisemisen. Näihin metodeihin liittyvä suomalainen terminologia on osin ”itsestään” vakiintunutta (esimerkiksi priori- ja posteriorijakaumat), osin kovin villiä sanailua. Uudessa sanastossa *bootstrap*-johdannaiset käännetään uusio-otannaksi ynnä muiksi uusioalkuisiksi termeiksi. Sitähän se on, saman aineiston kierrätystä.

Ilahduin myös esimerkiksi termistä posterioriväli, joka selittää itsensä: väliestimointia, jossa alkuliite posteriori liittyy sanan Bayes-tilastotieteeseen. Posterioriväli on siis perinteisen luottamusvälin Bayes-vastine.

Betajakauma vai beetajakauma?

Sanastotyössä hyvä periaate on säilyttää vanhat ja vakiintuneet termit, vaikkeivat ne olisikaan parhaita kuviteltavissa olevia. Tuskin regressio- tai

varianssianalyysiä nimettäisiin enää samoin, jos metodeille pitäisi nyt keksiä nimet, luottamusvälistä puhumattakaan. Uudessa sanastossa tämä periaate on omaksuttu hyvin.

Kirjoittajat ovat myös tukeutuneet suomen kielen ammattilaisiin, ollaanhan osin luomassa uusia sanoja suomen kieleen. Yhdessä kohdassa kirjoittajat ovat jättäneet käyttämättä mahdollisuuden uudistaa sanastoa yleisten kirjoitusohjeiden mukaisesti. He pitäytyvät vanhoissa termeissä betajakauma, fi-kerroin, khi-neliöjakauma, vaikka kyseiset kreikkalaiset aakkoset neuvotaan kirjoittamaan pitkin vokaalein beeta, fii ja khii – aivan kuten ne meillä lausutaankin. Muutos olisi ollut niin pieni, ettei se olisi aiheuttanut ymmärtämättömyyttä tai sekaannusta.

Sellaista nimistötyö on; yksikään lukija ei ole tyytyväinen jokaiseen sanavalintaan. Kokonaisuus kuitenkin ratkaisee tässäkin.

Sanasto vakiintuu sitä käyttämällä

Uusi tilastosanasto on niin perusteellinen ja perusteltu kooste, että sen termistön soisi tulevan alalle uudeksi standardiksi. Oma suosikkini uusista sanoista on erityisen poikkeavaa havaintoa tarkoittavan termin *outlier* suomennos ”oudokki”. Uuden mainion sanan omaksumista edistää vielä sen alkusointuisuus alkukielisen sanan kanssa.

Meidän tilastotieteen käyttäjien, opettajien ja kirjoittajien ulla ottaa uusi sanasto käyttöön.

Kuultuani uutuudesta jo kirjan käsikirjoitusvaiheessa, otin oudokin käyttöön luennoidessani tilastotieteen perusteita uusille opiskelijoille toissa syksynä. Ilokseni sana tarttui kerrasta, sillä opiskelijani alkoivat heti ja luontevasti käyttää sitä laskuharjoituksissa. Ymmärsin, että näin edistys voi tapahtua ja näin sen pitääkin mennä: uusi sukupolvi oppii uudet termit siinä kuin vakiintuneetkin.

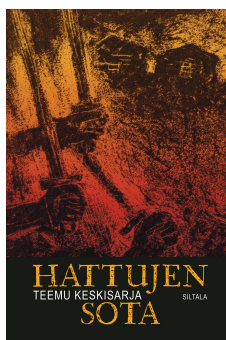
On meidän tilastotieteen käyttäjien, opettajien ja kirjoittajien vastuulla ottaa uusi sanasto käyttöön. Toivottavasti uudet termit vakiintuvat myös akatemisen maailman ulkopuolella nyt kun vaikkapa yritysmaailmassa työskentelevä tilastotieteilijä – *Data scientist* tai *Big data engineer* – pystyy entistä paremmin raportoimaan suomenkielisin termein.

Niin, se mainittu 1800-luvun lause on nykysuomeksi ”tilastotieteessä käytetään todennäköisyyksiä ja havaintoja”.

Harri Högmander on Jyväskylän yliopiston tilastotieteen lehtori. Hän on myös BirdLife Suomen nimistötoimikunnan puheenjohtaja.

[PALAA ALKUUN ↑](#)

KIRJALLISUUTTA



AKI ALANKO

**Ruotsalaisten
rähmäinen
revanssi –
Teemu
Keskisarjan
persoonallinen
teos hattujen
sodasta**



ANNE VIRKKI

**Tähtitieteen
naiset
näkyviksi**



MERJA LEPPÄLAHTI

**Japanilaisen
kirjallisuuden
kääntämisestä
– Mikään ei ole
itsestäänselvä
ä**



JARI LYYTIMÄKI

**Valeyön vaarat
– Valosaaste on
biologinen
kaikensotkija**