



# SUOMEN TILASTOSEURAN VUOSIKIRJA 2023

ÅRSBOK FÖR STATISTISKA SAMFUNDET  
I FINLAND 2023

THE YEARBOOK OF  
THE FINNISH STATISTICAL SOCIETY 2023

2023

---

# Sisällys

<b>Puheenjohtajan palsta: Vuosi 2023</b>	<b>2</b>
PEKKA PERE	
<b>Toimittajan palsta</b>	<b>5</b>
TOMMI HÄRKÄNEN	
<b>Mahdottomia tuloksia</b>	<b>7</b>
 PEKKA PERE	
<b>100+ -vuotisjuhlat</b>	<b>12</b>
PEKKA PERE	
<b>Kontrafaktuaalit, eventtiivi ja Kilpisenkatu</b>	<b>21</b>
JUHA KARVANEN	
<b>Kaupunkitilastoilla yli satavuotinen historia</b>	<b>24</b>
ASTA MANNINEN	
<b>Leo Törnqvist -palkinto</b>	<b>41</b>
Transformer -neuroverkon robustisuuden parantaminen Gaussisen pro- sessin ja neuroverkon yhdistävillä menetelmillä	41
ANTTI YLÄJÄRVI	
<b>Suomen Tilastoseuran hallitus vuonna 2022</b>	<b>49</b>
<b>Suomen Tilastoseuran hallitus vuonna 2023</b>	<b>51</b>
<b>Gunnar Modeen -minnesmedaljen</b>	<b>52</b>
<b>Scandinavian Journal of Statistics</b>	<b>54</b>
<b>Myönnetyt palkinnot</b>	<b>55</b>
Leo Törnqvist -palkinnot	55
Väitöskirjapalkinnot	57
<b>Suomen Tilastoseuran julkaisuja</b>	<b>58</b>
Tilastotieteellisiä tutkimuksia	59
Suomen Tilastoseuran vuosikirja	61
Muita julkaisuja	62

---

# Puheenjohtajan palsta: Vuosi 2023

PEKKA PERE

PEKKA.J.PERE@AALTO.FI

MATEMATIIKAN JA SYSTEEMIANALYYSIN LAITOS, AALTO-YLIOPISTO

Hallitus ahkeroi vuonna 2023. Aikaansaannosten takana on paljon vapaaehtoistyötä, joka ei tapahtumiin osallistujille näy tai niistä lukijoille ilmene. Kiitän Tilastoseuran hallitusta aktiivisuudesta.

Edellisessä Vuosikirjassa kerroin tekoäly GPT:n näkemyksistä tilasto- ja datatieteestä (Pere 2023).<sup>1</sup> Kysyin GPT:ltä, mitä on tilastotiede. Vastauksessaan GPT painotti aineistoa. GPT ei maininnut satunnaisuutta, todennäköisyyttä, epävarmuutta tai tilastollista päättelyä. GPT heijastelee aikoja vastauksessaan. Datatieteen tapainen fokuointi aineistoon on parrasvaloissa; perinteinen epävarmuuden kvantifiointi ja tilastollinen päättely otoksesta perusjoukkoon ei. Tilastollinen päättely on silti edelleen tilastotieteen ydinteoriaa ja keskeistä empiiristen havaintojen merkitystä arvioitaessa.

Yliopistonlehtori TAPIO NUMMI (Tampereen yliopisto) ehdotti minulle syksyllä 2022, että Tilastoseura ryhtyisi organisoimaan vuosittain valtakunnallista tilastollisen päättelyn jatkokurssia. Ymmärtääkseni Tapion ajatus oli vaalia tilastollisen päättelyn taitoa ja kirkastaa sen kajoa Suomessa. Lisäkannustimeksi oli hyvät kokemukset emeritusprofessorien HANNU OJA (Turun yliopisto) ja JUKKA NYBLÖM (Jyväskylän yliopisto) aikoinaan pitämistä tilastotieteen jatkokoulutuskursseista.

Esittelin Tilastoseuran edustajana idean tilastollisen päättelyn jatkokurssista tilastotieteen professorien valtakunnallisessa kokouksessa Aalto-yliopistossa 13.1.2023. Vuosittainen päättelyn kurssi ei saanut kannatusta. Sen sijaan vuorovuosittaiselle päättelyn kurssille ja vaihtuva-aiheiselle teemakurssille nähtiin tarvetta. Tutkailujen ja tiedustelujen jälkeen Tilastoseura päätti kutsua North Carolina State -yliopiston tilastotieteen professori RYAN MARTININ pitämään valtakunnallisen tilastollisen päättelyn jatkokurssin keväällä 2024. Kiitän yliopistonlehtori HYON-JUNG KIM-OLLILAA (Tampereen yliopisto) avusta luennoitsijan löytämisessä. Luennoitsijan hankkimisen lisäksi rahoituksen hakeminen kurssille oli oleellinen tehtävä loppuvuodelle. Tätä palstaa luettaessa kurssi

---

<sup>1</sup>Viittasin A. Hjeltin artikkeliin "Tilastotieteen alalta", jonka julkaisupaikka ja -ajankohta eivät ole tiedossa. Julkaisupaikaksi paljastui sitemmin Valvoja-lehti ja ajankohdaksi 1.2.1886: <https://digi.kansalliskirjasto.fi/aikakausi/binding/634025?page=35>.

on jo (menestyksellisesti!) pidetty, mistä tarkemmin seuraavassa Vuosikirjassa.

Vuoden suur tapahtuma ja -ponnistus oli tietenkin seuramme 100+-vuosijuhlat 7.–9.6.2023. Kuvaan niitä erillisessä artikkelissa. Sen verran totean juhlita tässä, että merkkipuoden kunniaksi moni – mukaan lukien allekirjoittanut – esitelmöi juhlissa tilastotieteen historiasta Suomessa. Olemme kirjoittajia seuran hankkeessa julkaista kirja tilastotieteen historiasta Suomessa. Etäkokoustimme seitsemän kertaa vuoden aikana, ja työemme jatkuu edelleen.

Iltapäiväseminaareja ei ollut vuonna 2023. Niiden sijaan hallitus keskittyi 100+-vuosijuhlien järjestämiseen.

Finnish Young Statisticians Workshop, eli nuorten tilastotieteilijöiden työpaja, pidettiin 7.11.2023 Jyväskylän yliopistolla Mattilanniemen kampuksella. Tilaisuus järjestettiin totuttuun tapaan yhteistyössä Suomen Biostatistiikan Seuran kanssa. Väitöskirjatutkija TIINA-MARIA PASANEN (Jyväskylän yliopisto) konkreettisesti toteutti tapahtuman. Paikalle tuli kaksitoista nuorta tilastotieteilijää viidestä eri yliopistosta, ja kahdeksan osallistujaa esitelmöi.

Edustin seuraamme ECAS-organisaatiossa (European Courses in Advanced Statistics, <https://ecas.fenstats.eu/>). REIJA HELENIUS (Tilastokeskus) edusti seuraamme FENStatS-organisaatiossa (The Federation of European National Statistical Societies, <https://www.fenstats.eu>).

Reija Helenius palkittiin kattojärjestömme Kansainvälisen tilastoinstituutin ISI:n konferenssissa ISI Service -palkinnolla.<sup>2</sup> Peruste oli Reijan vuodesta 2010 ansiokkaasti johtama ISI:n tilastojen lukutaidon kehittämistä kansainvälisesti edistävä International Statistical Literacy -projekti. Olemme hallituksessa ylpeitä palkinnosta, ja onnittelan Reijaa tässä vielä kerran!

Jäsenemme nauttivat edelleen Cambridge University Pressin, CRC Pressin ja Wileyn kirja-alennuksista. Oxford University Press lopetti alennuksensa. Saimme siltä sopivasti tilalle uuden jäsenedun eli Royal Statistical Societyn, American Statistical Associationin ja Statistical Society of Australian julkaiseman Significance-lehden vuositilauksen edullisesti. Käyttäkää jäsenetujamme, niin niiden tarjoajat haluavat jatkaa niitä. Sihteerimme antaa lisätietoja (sihteerit@tilastoseura.fi).

Tämä on toinen Vuosikirja peräjälkeen, jossa on vertaisarvioitu artikkeli – tänä vuonna ja jatkossa Tieteellisten seurain valtuuskunnan vertaisarviointitunnuksella vahvistettuna. Tilastoseura julkaisee yhdessä muiden pohjoismaisten tilastoseurojen kanssa arvovaltaista aikakauskirjaa Scandinavian Journal of Statistics. Vuosikirja ei kilpaile sen tai muiden englanninkielisten tilastotieteellisten julkaisujen kanssa. Vuosikirjaan voi tarjota englanninkielisiä kirjoituksia. Silti ajattelen Vuosikirjaa ensisijaisesti ainutlaatuisena paikkana julkaista tilastotiedettä käsitteleviä Suomen kansalliskielillä kirjoitettuja artikkeleita, joiden ydin voi olla muukin kuin tutkimus. Tarjotkaa käsikirjoituksia Vuosikirjaan! Toimitus ottaa mielellään tutkailtavakseen teoreettisia tai soveltavia tutkimuksia, aihekatsauksia, kongressiraportteja tai vaikkapa tilastotieteen opetusta tai asemaa yhteiskunnassa käsitteleviä kirjoituksia. Pyrimme kiirehtimään Vuosikirjan aikataulua niin, että se ilmestyisi aina kunkin vuoden alkupuolella kattaen edellisen vuoden.

---

<sup>2</sup><https://stat.fi/uutinen/tilastokeskuksen-reija-helenius-palkittiin-kansainvalisen-tilastoinstituutin-isiin-konferenssissa>

## **Viitteet**

Pere, P. 2023. Mitä on tilastotiede tekoälyn mielestä? *Suomen Tilastoseuran Vuosikirja 2019–2022*. Suomen Tilastoseura.

---

# Toimittajan palsta

TOMMI HÄRKÄNEN

TOMMI.HARKANEN@THL.FI

TERVEYDEN JA HYVINVOINNIN LAITOS (THL)

## Tieteellisten seurain valtuuskunnan vertaisarviointitunnus ja journal.fi

Tilastoseuran tavoitteena on tarjota julkaisufoorumi suomenkielisille, tilastotieteen piiriin kuuluville julkaisuille, jotka täyttävät tieteellisen julkaisun kriteerit. Tärkeä askel otettiin 12.4.2024, jolloin [Tieteellisten seurain valtuuskunnan \(TSV\) vertaisarviointitunnuksen](#) käyttöoikeus myönnettiin Suomen Tilastoseuran Vuosikirjalle. Tässä Vuosikirjassa julkaistaan ensimmäinen TSV:n vertaisarviointitunnuksella merkitty artikkeli alkaen sivulta 7. Vertaisarvioitavaksi lähetettävät käsikirjoitukset lähetetään verkkosivun <https://journal.fi/tilastoseura/> kautta, jolloin TSV:n kriteerit arviointiprosessin ja dokumentoinnin osalta täyttyvät, ja TSV:n tunnus voidaan liittää vertaisarvioituihin artikkeleihin. Tunnus näkyy myös sisällysluettelossa. Myös muut kirjoitukset pyydetään lähettämään tämän sivuston kautta.

## Lähetäkää kirjoituksia!

Kysyin Bing-tekoälyavustajalta, miksi kannattaa julkaista suomeksi.

Julkaiseminen suomeksi voi olla erittäin hyödyllistä useista syistä:

**Kansallinen vaikuttavuus** Suomeksi julkaistut tieteelliset kirjoitukset voivat tavoittaa laajemman kotimaisen yleisön, mukaan lukien päätöksentekijät, opettajat ja opiskelijat.

**Kulttuurinen merkitys** Suomalaisen tieteen ja kulttuurin edistäminen on tärkeää. Julkaiseminen omalla kielellä vahvistaa kansallista identiteettiä ja kulttuuriperintöä.

**Saavutettavuus** Kaikki eivät välttämättä osaa englantia riittävän hyvin ymmärtääkseen tieteellisiä tekstejä. Suomeksi julkaiseminen tekee tiedosta saavutettavampaa.

**Kielitaidon kehittäminen** Tieteellisen terminologian kehittäminen ja ylläpitäminen suomeksi on tärkeää, jotta kieli pysyy elinvoimaisena ja ajantasaisena.

Onko sinulla jokin erityinen aihe, josta haluaisit julkaista?

Mielestäni tämä oli hyvä vastaus! Näitä tavoitteita tukemaan Tilastoseura on julkaissut [Tilastotieteen sanaston](#).

## **L<sup>A</sup>T<sub>E</sub>X-pohja**

Vuosikirjaan voi lähettää käsikirjoituksia sekä L<sup>A</sup>T<sub>E</sub>X- että Word-tiedostoina. Tilastoseura suosittelee L<sup>A</sup>T<sub>E</sub>X-formaattia, joka sujuvoittaa Vuosikirjan taittamista. Seuran verkkosivuilta<sup>1</sup> voi ladata L<sup>A</sup>T<sub>E</sub>X-esimerkkitiedostot, joiden pohjalta on helppo tuottaa oma käsikirjoitus.

Jos et halua asentaa L<sup>A</sup>T<sub>E</sub>X-ohjelmistoa omalle koneellesi, voit käyttää myös verkkopalveluina tarjottavia alustoja, kuten esim. Overleaf,<sup>2</sup> joka tarjoaa tekstin tuottamista varten myös tavallista tekstinkäsittelyohjelmaa muistuttavan *Visual Editor* -tekstinmuokkaimen, jolloin tarve osata L<sup>A</sup>T<sub>E</sub>X-kielen komentoja lähes poistuu – erityisesti, jos et tarvitse matemaattisia kaavoja. Kannattaa tutustua!

## **Tavoitteita**

### **Julkaisufoorumi taso 1**

Vuosikirjan kehittäminen jatkuu edelleen. Tilastoseura hakee Jufo-1-luokitusta<sup>3</sup>, jonka saamisen jälkeen Vuosikirjassa julkaistut, vertaisarvioidut artikkelit ovat hyödyllisiä myös laskettaessa yliopistojen suoritteita. Tiedotamme tästä lisää myöhemmin.

### **Vuosikirjan julkaisemisen kustannusten kattaminen**

Ennen kuin Vuosikirja on julkaistu, tarvitaan monien ihmisten työpanosta. Kirjoittajat ovat käyttäneet aikaansa ja osaamistaan käsikirjoitustensa viimeistelemiseksi, toimittajat ovat käyneet käsikirjoitukset läpi, arvioijat ovat kommentoineet vertaisarvioitaviksi lähetetyt käsikirjoitukset ja taittaja on yhtenäistänyt käsikirjoitusten ulkoasut kirjoitusohjeiden mukaiseksi. Vuosikirjaa painetaan kirjoina, joista osa postitetaan mm. Kansalliskirjastolle ja useille yliopistoille, ja lisäksi sitä jaetaan pdf-tiedostona seuran verkkosivulta. TSV-vertaisarviointitunnuksen ja Jufo-1-luokituksen käyttäminen sekä mahdollistaa Tilastoseuralle apurahojen hakemisen Vuosikirjan julkaisemista varten että kannustaa lähettämään enemmän kirjoituksia julkaistavaksi.

Nyt kannattaa siis suunnitella seuraavien käsikirjoitusten lähettämistä julkaistavaksi Tilastoseuran Vuosikirjassa!

---

<sup>1</sup><https://tilastoseura.fi/fi/pages/vuosikirja.html>

<sup>2</sup><https://www.overleaf.com/about/features-overview>

<sup>3</sup><https://julkaisufoorumi.fi/fi/arvioinnit/luokiteluperusteet>

---

# Mahdottomia tuloksia

PEKKA PERE

PEKKA. J. PERE@AALTO.FI

MATEMATIIKAN JA SYSTEEMIANALYYSIN LAITOS  
AALTO-YLIOPISTO



## Tiivistelmä

Tieteen etiikka ja uskottavuus vaativat, että huolelliseen ja asianmukaiseen menetelmien ja tulosten raportointiin kiinnitetään soveltavassa tutkimuksessa nykyistä enemmän huomiota. Asiaan pureudutaan kahden esimerkin kautta.

## Kaksi tutkimusta

Uutukaisessa artikkelissa raportoitiin logistisia regressiomalleja, joilla pyrittiin selvittämään lapsen sukupuolivähemmistöön kuulumisen ja perheväkivallan kokemisen välistä yhteyttä (Hakala 2023). Aineisto koostui 6 603 kuudes- tai yhdeksäsluokkalaisten peruskoululaisen vastauksista vuoden 2022 kansallisessa lapsiuhritutkimuksessa. Selitettävä muuttuja oli, ovatko vanhemmat kohdelleet lasta väkivaltaisesti (1: kyllä; 0: ei). Henkiselle ja fyysiselle väkivallalle estimoitii erillisiä malleja. Analyysi alla kohdistuu ytimekkyyden ja yksinkertaisuuden vuoksi malliin, jossa selitettiin kokemusta henkisestä väkivallasta yksin sukupuoli-identiteettimuuttujalla ( $x$ ).

Mallin ydin on yhtälö

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x,$$

jossa  $\log(\cdot)$  on luonnollinen logaritmifunktio,  $\beta_0$  on vakiotermi,  $\beta_1$  on regressiokerroin ja  $\pi(x)$  on todennäköisyys ( $0 < \pi(x) < 1$ ), että lapsi on kertonut vanhempansa kohdistaneen häneen henkistä väkivaltaa, kun lapsen sukupuoli-identiteettimuuttujan arvo on  $x$ . Mallin tuloksin ytimessä on ristiuhde (*odds*



*ratio*):

$$\frac{\frac{\pi(x+1)}{1-\pi(x+1)}}{\frac{\pi(x)}{1-\pi(x)}} = \exp(\beta_1), \quad (1)$$

jossa  $\exp(\cdot)$  on luonnollinen eksponenttifunktio. Ristisuhte kertoo, millä luvulla vastasuhte (*odds*)

$$\frac{\pi(x)}{1-\pi(x)}$$

kertaantuu, kun selittävän muuttujan arvo suurenee yksiköllä (Agresti 2019, 90). Vasta- ja ristisuhte ovat Tilastotieteen sanastossa suositeltuja suomennoksia (Alho ym. 2023). Huomion kohteena olevassa artikkelissa suomennokset ovat veto ja vetosuhte.

Artikkelissa kuvattu yhden selittäjän malli ei ole mielekäs, eivätkä sen estimointitulokset ole mahdollisia. Sama pätee muihin artikkelissa esitettäviin malleihin.

Keskeisen selittävän muuttujan kerrottiin olevan kolmiportainen sukupuoli-identiteettimuuttuja, joka sai arvon 0, jos lapsi koki juridisen sukupuolensa omakseen, arvon 1, jos lapsi koki olevansa vastakkaista sukupuolta kuin juridinen sukupuolensa ja arvon 2 muissa tilanteissa (lapsi ei vaikkapa kokenut mitään sukupuolta omakseen). Viiteluokka on mallissa arvoa 0 vastaava sukupuoli-identiteetti. Malli ei ole mielekäs, koska se asettaa sukupuoli-identiteeteille järjestyksen ja numeerisen etäisyyden viiteluokkaan. Jos identiteetit olisi välttämätöntä järjestää tilastollista analyysia varten, moni saattaisi vaihtaa luokkien 1 ja 2 järjestyksen. Artikkelissa asetetut numeeriset etäisyydet sukupuoli-identiteeteille ovat yhtä lailla kyseenalaiset ja mielivaltaiset. Arvolle 2 luokitellut lapset ovat ikään kuin kaksi kertaa niin kaukana viiteluokasta kuin arvolle 1 luokitellut. On luontevaa asettaa lasten pituudet järjestykseen ja määrittellä niiden välille etäisyydet. Sama ei päde lasten sukupuoli-identiteeteille.

Artikkelin taulukon 3 mukaiset ristisuhteen estimaatit eivät voi pitää paikkaansa artikkelissa kuvatulle mallille. Sen mukaan ristisuhteen estimaatti on 1.96 arvolle 1 ja 2.66 arvolle 2 luokitelluille lapsille. Raportoitu malli on mahdoton. Selittäjän  $x$  arvot 0 ja 1 vastaavat artikkelin luokkia 0 ja 1. Artikkelissa raportoidun ristisuhteen ja kaavan (1) mukaan  $1.96 = \exp(\hat{\beta}_1)$ , jossa  $\hat{\beta}_1$  on regressiokerrotoimen estimoitu arvo. Arvo  $\hat{\beta}_1 = \log(1.96) = 0.67$  toteuttaa yhtälön. Luokassa 2 selittäjän arvo  $x$  on yhden verran suurempi kuin luokassa 1. Näin ollen luokassa 2 vasta- ja siten myös ristisuhteen pitäisi kertautua  $\exp(0.67)$ :lla. Tulisi siis päteä  $2.66 = 1.96 \times \exp(0.67)$ , mutta  $1.96 \times \exp(0.67) = 3.84$ . Ristiriidan mallin ja raportoitujen estimointitulosten välillä huomaa niin ikään näin: kahden yksikön muutoksen selittäjässä pitäisi tuottaa ristisuhteeksi  $\exp(2\hat{\beta}_1) = \exp(1.35) = 3.84$  (Agresti 2019, 103–104), mutta raportoitu ristisuhte on 2.66.<sup>1</sup> Artikkelissa raportoidut ristisuhteet ovat mahdottomia siinä kuvatun mallin puitteissa.

Sukupuoli-identiteettien yhteyttä koettuun henkiseen väkivaltaan voitaisiin luontevasti mallittaa korvaamalla artikkelissa kuvattu sukupuoli-identiteettimuuttuja kahdella osoitinmuuttujalla. Ensimmäinen saisi arvon 1, kun lapsi kuuluu luokkaan, johon viitattiin edellä arvolla 1, ja 0 muulloin. Toinen osoitinmuuttuja saisi vastaavasti arvon 1, kun lapsi on luokassa, johon viitattiin edellä arvolla 2, ja 0 muulloin. Viiteluokka olisi edelleen juridisen

<sup>1</sup>Laskut on tehty useamman desimaalin tarkkuudella kuin tekstiin on kirjattu.

sukupuolensa omakseen kokevat lapset. Tällöin sukupuoli-identiteettejä ei asetettaisi järjestykseen, niille ei asetettaisi välimatkoja ja ristisuhteet tulisivat estimoitua ilman artikkelissa kuvatun sukupuoli-identiteettimuuttujan asettamaa kahletta ristisuhteiden välille.

Artikkelin ristisuhteet eivät ole estimoitu artikkelissa kuvatulla tavalla vaan edellisessä kappaleessa kuvatulla tavalla.<sup>2</sup> Silloin arvon 2 määrittelemän luokan ristisuhdetta ei sido lauseke  $\exp(2\hat{\beta}_1)$  ja raportoidut ristisuhteet ovat mahdollisia. Malli on kuvattu artikkelissa väärin.

Sukupuoli-identiteetin ja koetun perheväkivallan yhteyttä nuorilla on mallitettu hyvin samaan tapaan toisessa tutkimuksessa (Kaltiala ja Ellonen 2022). Kolmiluokkaiselle sukupuoli-identiteetille kerrottiin muodostetun yksi selittävä muuttuja viiteluokkana juridisen sukupuolensa omaksi kokevat. Muuten muuttujan rakennetta ei avattu. Estimoidun mallin kerrottiin siis kuitenkin olevan toinen kuin luonteva malli, jossa sukupuoli-identiteetit on osoitettu kahdella osoitinmuuttujalla. Artikkelissa estimoidut ristisuhteet eivät toteuta edellä kuvattuja rajoituksia. Lukija jää aprikoimaan, millainen artikkelissa käytetty sukupuoli-identiteettimuuttuja on ja ounastelemaan, että tässäkin tutkimuksessa on toimittu toisin kuin on esitetty ja todellisuudessa estimoitu kahden osoitinmuuttujan luonteva malli.

## Tilastotieteen etiikka

On oleellista, että sovellettu tilastollinen malli kuvataan oikein. Monen soveltavan tutkijan mielestä tilastotiede on vaikeaa. Tilastollinen malli voi muuttua mahdottomaksi ymmärtää, jos se kuvataan väärin.

Toinen edellä osoitettu ongelmallinen artikkeli on sosiaalityön alalta. Dunleavyn (2024, 12–13) arvio koskee sosiaalityön tutkimuksia mutta käy yleispätevästä summeerauksesta:

Tilastollisten tulosten puutteellinen raportointi vaikeuttaa tutkimustulosten arviointia. Pitkään on tiedetty, että monissa julkaistuissa artikkeleissa ei ole riittäviä tietoja tutkimusmetodeista ja -tuloksista. Tällöin monet tulokset ovat käyttökelvottomia tai mahdottomia tulkita tai molempia.

Tilastotieteen etiikka vaatii, että malli kuvataan huolellisesti ja asianmukaisesti: lukijat hukkaavat aikaansa yrittämällä ymmärtää mallia, joka on mahdoton. Luottamus tieteeseen heikkenee, jos raportoidaan muuta kuin on tehty. Virheellinen kuvaus voi johtaa sekä tekijän että lukijan vääriin johtopäätöksiin. Moni noudattaa julkaistujen analyysien esimerkkiä omissa tutkimuksissaan, joten mallin virheellinen kuvaus tai tulkinta voi monistua eli johdattaa muut tutkijat virheellisiin analyysiin tai tulkintoihin. Yhteiskuntatieteissä seuraus voi olla tosiasioihin pohjaamattomat politiikkasuositukset (esim. Pere 2020; Pere 2021). (Vrt. Altman 1980; American Statistical Association 2022, 2, 11; Vail 2011.) Tilastotieteen etiikan ja uskottavuuden vaaliminen on erityisen tärkeää nykyisinä aikoina, kun tieteen uskottavuutta kyseenalaistetaan monilla tahoilla.

<sup>2</sup>Noora Ellosen välittämä Venla Hakalan henkilökohtainen tiedonanto 13.11.2023. Hakala kertoi saavansa edellä osoitetut estimointitulokset, kun hän estimoi mallin, jossa on kaksi edellä kuvattua osoitinmuuttujaa sukupuoli-identiteeteille (artikkelissa kuvatun sukupuoli-identiteettimuuttujan sijaan).

## Tutkimusten uusittavuus ja toistettavuus

Tieteessä on oleellista, että tutkimukset voidaan uusia ja toistaa. Uusittavuus (*reproducibility*) toteutuu, jos uudet tutkijat päätyvät yhtäpitäviin tuloksiin alkuperäisen tutkimuksen kanssa, kun he soveltavat siinä kuvatulla tavalla menetelmiä alkuperäiseen aineistoon. Toistettavuus (*replicability*) pätee, kun uudet tutkijat päätyvät yhtäpitäviin tuloksiin alkuperäisen tutkimuksen kanssa vastaavassa tutkimuksessa uudella aineistolla.<sup>3</sup> Monilla tieteenaloilla kumpikaan ei usein onnistu ja puhutaan toistettavuuskriisistä. Aiheesta tulvii julkaisuja, ja myös kriminologiassa, jonka alan lehdessä tämän artikkelin lähtökohtana oleva tutkimus julkaistiin, on havahduttu ongelmaan. Yksi syy valitettavaan tilanteeseen on menetelmien ja tulosten puutteellinen raportointi. (JASA 2020; Johnson ym. 2017; McNeeley ja Warner 2015; Monk ja Koziarski 2023; Open Science Collaboration 2015; Pridemore ym. 2018; Savolainen ja VanEseltine 2018; Social Science Data Editors 2024; Xiong ja Cribben 2023; Youyou ym. 2023.)

## Johtopäätös

Tieteen etiikka ja uskottavuus vaativat, että huolelliseen ja asianmukaiseen menetelmien ja tulosten raportointiin kiinnitetään soveltavassa tutkimuksessa nykyistä enemmän huomiota. Tilastotieteen opetuksessa tulee huolehtia, että soveltajat eivät opi vain käyttämään tilastollisia menetelmiä vaan oppivat myös ymmärtämään ja kuvaamaan niitä.

## Viitteet

- Agresti, A. 2019. *Introduction to Categorical Data Analysis*. 3. laitos. Wiley.
- Alho, J., Arjas, E., Läärä, E. & Pere, P. 2023. *Tilastotieteen sanasto*. Suomen Tilastoseuran julkaisuja no. 8. 2. laitos. Suomen Tilastoseura. Helsinki. <https://sanasto.tilastoseura.fi/>
- Altman, D. G. 1980. Statistics and Ethics in Medical Research. Misuse of Statistics Is Unethical. *British Medical Journal*, 281, 1182–1184.
- American Statistical Association. 2022. Ethical Guidelines for Statistical Practice. <http://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx>
- Carriquiry, A. L., Daniels, M. J. & Reid, N. 2023. Editorial: Special Issue on Reproducibility and Replicability. *Statistical Science*, 38(4), 525–526. <https://doi.org/10.1214/23-STS909>
- Dunleavy, D.J. 2024. On the Dearth of Retractions in Social Work: A Preliminary Analysis of Ten Leading Journals. <https://osf.io/preprints/metaarxiv/y9pvm>
- Hakala, V. 2023. Translasten ja -nuorten perheväkivaltakokemukset. *Kriminologia*, 3(2), 73–91. <https://doi.org/10.54332/krim.124999>

---

<sup>3</sup>Termien määritelmät vaihtelevat (National Academies of Sciences, Engineering, and Medicine 2019, 43). Määritelmät tässä ovat mt:ksen (s. 46) ja Carriquiry'n ym:iden (2023) suositusten mukaiset.

- JASA Reproducibility Guide. 2020. <https://jasa-acs.github.io/repro-guide/>
- Johnson, V. E., Payne, R. D., Wang, T., Asher, A. & Mandal, S. 2017. On the Reproducibility of Psychological Science. *Journal of the American Statistical Association*, 112(517), 1–10. <http://dx.doi.org/10.1080/01621459.2016.1240079>
- Kaltiala, R. & Ellonen, N. 2022. Transgender Identity and Experiences of Sexual Harassment in Adolescence. *Child Abuse Review*, e2748. <https://doi.org/10.1002/car.2748>
- McNeeley, S. & Warner, J. J. 2015. Replication in Criminology: A Necessary Practice. *European Journal of Criminology*, 12(5), 581–597. <https://doi.org/10.1177/1477370815578197>
- Monk, K. & Koziarski, J. 2023. Replicating & Reproducing Policing Research. *Police Practice and Research*, 24(5), 519–522. <https://doi.org/10.1080/15614263.2023.2221075>
- National Academies of Sciences, Engineering, and Medicine. 2019. Reproducibility and Replicability in Science. The National Academies Press. <https://doi.org/10.17226/25303>
- Open Science Collaboration. 2015. Estimating the Reproducibility of Psychological Science. *Science*, 349, aac4716. <https://doi.org/10.1126/science.aac4716>
- Pere, P. 2020. Kätevä arvio riskisuhteesta avuksi yhteiskuntatieteilijöille. *Yhteiskuntapolitiikka*, 85, 434–436.
- Pere, P. 2021. Ristisuhde ja väärinkäsitys. *Yliopisto-lehti*, 4/2021, 59.
- Pridemore, W. A., Makel, M. C. & Plucker, J. A. 2018. Replication in Criminology and the Social Sciences. *Annual Review of Criminology*, 1, 19–38.
- Savolainen, J. & VanEseltine, M. 2018. Replication and Research Integrity in Criminology: Introduction to the Special Issue. *Journal of Contemporary Criminal Justice*, 34(3), 236–244. <https://doi.org/10.1177/1043986218777288>
- Social Science Data Editors. 2024. Unofficial Guidance on Various Topics by Social Science Data Editors. Data and Code Guidance by Data Editors. <https://social-science-data-editors.github.io/guidance/>
- Vail, A. 2011. Ethical Review Committees. Teoksessa Everitt, B. S. & Palmer, C. R. (toim.): *Encyclopaedic Companion to Medical Statistics. 2 laitos*. Wiley.
- Xiong, X. & Cribben, I. 2022. The State of Play of Reproducibility in Statistics: An Empirical Analysis. *The American Statistician*, 77(2), 115–126. <https://doi.org/10.1080/00031305.2022.2131625>
- Youyou, W., Yang, Y. & Uzzi, B. 2023. A Discipline-Wide Investigation of the Replicability of Psychology Papers over the Past Two Decades. *PNAS*, 120(6), e2208863120. <https://doi.org/10.1073/pnas.2208863120>

---

# 100+ -vuotisjuhlat

PEKKA PERE

PEKKA.J.PERE@AALTO.FI

MATEMATIIKAN JA SYSTEEMIANALYYSIN LAITOS, AALTO-YLIOPISTO

Tilastoseuran viimeisin käytäntö on ollut järjestää Tilastopäiviä joka toinen vuosi. Edelliset Tilastopäivät olivat Turussa 2017. Seuran hallitus päätti niiden jälkeen lykätä seuraavat Tilastopäivät juhluvuodelle 2020, jolloin seura täyttää 100 vuotta. Sattuma tavalliseen tapansa yllätti: COVID-19-pandemia esti juhlien järjestämisen vuoteen 2023 asti. Tilastopäiville kertyi väliä peräti kuusi vuotta. Viiävästyksen vuoksi juhlimme seuran 100+-vuotista historiaa vasta 7.–9.6.2023.

Kiitän johtamaani hallituksen nimittämää työryhmää, joka suunnitteli juhlat ja hankki niille rahoituksen. Lisäkseni ryhmään kuuluivat REIJA HELENIUS, ARI JAAKOLA, TOMMI HÄRKÄNEN ja JYRKI MÖTTÖNEN. Viimeksi mainittua kiitän erityisesti avusta viime hetken valmisteluissa. Juhlien sujuvaa etenemistä itse päivillä olivat konkreettisesti auttamassa MARI MATTILA, ATTE PIETARINEN, HENRIK TÖRMI ja NATALIA VESSELINOVA. Kiitokseni myös heille.

Raha-avustuksista seura kiittää Magnus Ehrnroothin säätiötä, SAS Institute Oy:tä, Solita Oy:tä, Suomalaista tiedeakatemiaa, Terveyden ja hyvinvoinnin laitosta, Tieteellisten seurain valtuuskuntaa, Tilastokeskusta, Yrjö Jahnssoonin säätiötä sekä CRC Pressiä ja Wileyä. Erityiskiitoksen seura osoittaa Helsingin yliopistolle, joka tarjosi erinomaiset tilat juhlille ja auttoi järjestelyissä.

Päiviin 8.–9.6. oli pakattu paljon ohjelmaa. Seuran satavuotisen historian juhlistamiseksi ohjelmassa oli monia historia-aiheisia esitelmiä. Kuulijoita oli läsnä noin satakaksikymmentä.

## Ohjelma

Juhlien ohjelma on alla. Ohjelma erosi hieman aiemmin julkaistusta muutaman peruutuksen takia. “Varaslähtö” oli opettajien lehtisalissa ja istunnot PII-salissa Helsingin yliopiston Porthaniassa.

### Keskiviikko 7.6.2023

	Tilastotieteen suomalainen historia -keskustelu
17.00–20.00	“Varaslähtö” 100+-vuotisjuhliin ja rekisteröityminen

Torstai 8.6.2023

09.00–10.00	<b>Rekisteröityminen</b>	
09.00–10.00	<b>Istunto 1, puheenjohtajana Pekka Pere</b>	
10.00–10.15	<b>Pekka Pere</b> puheenjohtaja, Suomen Tilastoseura	Vuosijuhlan avaus
10.15–11.00	<b>Deborah Ashby</b> professori, Imperial College, London	Florence Nightingale's legacy: using data to improve public health from the time of the Crimea to the time of the coronavirus
11.00–11.30	<b>Tuomo Nieminen</b> väitöskirjatutkija, VTM, THL	COVID-19 vaccination and adverse events: real-time epidemiological analyses based on national health registers in Finland
11.30–12.00	<b>Tiina Laatikainen</b> professori, Itä-Suomen yliopisto, THL	Do health sciences exist without statistics?
12.00–13.00	<b>Lounastauko ja posterinäyttely</b>	
13.00–15.00	<b>Istunto 2, puheenjohtajana Pekka Pere</b>	
13.00–13.30	<b>Esko Leskinen</b> professori emeritus	Vahvistavat faktorimallit – psykometrian historiaa Suomessa
13.30–14.00	<b>Pekka Pere</b> DPhil, yliopistonlehtori	Tilastotieteen suomalaista historiaa
14.00–14.30	<b>Heikki Mikkonen</b> väitöskirjatutkija, FM	Suomen Tilastoseuran historia
14.30–15.00	<b>Esa Läärä</b> professori emeritus	Uusi Tilastotieteen sanasto ja verkkosanasto
15.00–15.30	<b>Iltapäiväkahvi ja posterinäyttely</b>	
15.30–16.30	<b>Istunto 3, puheenjohtajana Pekka Pere</b>	
15.30–16.00	<b>Markus Sovala</b> pääjohtaja, VTT, Tilastokeskus	Tilastoja ja dataa – joko tai vai sekä että?
16.00–16.30	<b>Risto Lehtonen</b> professori emeritus	Paradigmamuutoksista virallisessa tilastoinnissa kansainvälisesti ja Suomessa
16.30–17.30	<b>Istunto 4, puheenjohtajana Ari Jaakola</b>	
16.30–17.00	<b>Jukka Hoffrén</b> tutkimuspäällikkö, HTT, Tilastokeskus	Luotettavat tilastot demokratian ja yhteiskunnallisen päättöksenteon perustana
17.00–17.30	<b>Asta Manninen</b> johtaja emerita, FM	Kaupunkitilastoista on moneksi
20.00–23.00	<b>Illallinen Katajanokan Kasinolla</b>	

## Perjantai 9.6.2023

09.00–11.15	<b>Istunto 5, puheenjohtajana Pekka Pere</b>	
09.00–9.45	<b>Jouni Kuha</b> professori, London School of Economics and Political Science	Modelling correlation matrices in multivariate dyadic data: Latent variable models for intergenerational exchanges of family support
09.45–10.15	<b>Timo Teräsvirta</b> professori emeritus	Long monthly European temperature series and the North Atlantic oscillation
10.15–10.45	<b>Heikki Kauppi</b> professori	Estimating optimal binary classification rules: A nonparametric method for nonmonotonic density ratios
10.45–11.15	<b>Annika Kangas</b> professori	Statistics in forest sciences in Finland
12.00–13.00	<b>Lounastauko ja posterinäestys päättyy</b>	
12.30–14.15		Sukupolvien ketju: professori emeritus Elja Arjas professori emeritus Hannu Oja professori Pauliina Ilmonen väitöskirjatutkija, DI Jaakko Pere
14.15–14.45	<b>Iltapäiväkahvi ja posterinäyttely</b>	
14.45–16.45		Törnqvist- ja posterikilpailujen palkintojen jako ja voittajien esitykset
16.45–17.00	<b>Pekka Pere</b> puheenjohtaja, Suomen Tilastoseura	Vuosijuhlan päättäminen

## Varaslähtö

Alkuperäinen tarkoitus oli juhlia 8.–9.6., mutta pyynnöstä seura järjesti juhli- le varaslähdön 7.6. Sen teemana oli Tilastotieteen suomalainen historia, koska seuralla on aiheesta tekeillä kirja. Paikalle tuli kolmattakymmentä tilastotieteen ystävää, joista valtaosa oli professoreita (ml. emeritukset ja yksi emerita).<sup>1</sup>

LEIF NORDBERG muisteli LEO TÖRNQVISTIä, ensimmäistä tilastotieteen professoria Helsingin yliopistossa ja Suomessa. Törnqvistille olivat graafiset tarkastelut olleet tärkeitä. Hän ei ollut tavannut juuri valmistella luentoja. Törnqvistiltä oli jäänyt päätösteorian kirjan käsikirjoitus kesken. Nordberg oli ajatellut julkaista käsikirjoituksen kera omien kommenttinsa eläkepäivinä. Eläkkeelle jäädessään Törnqvist oli pyytänyt tilastotieteen laitoksen henkilökuntaa tyhjentämään työhuoneensa todeten, ettei tahdo sieltä mitään. Toivomusta oli kunnioitettu, ja työhuoneessa ollut päätösteorian kirjan keskeneräinen käsikirjoitus oli hävitetty. Käsikirjoitusta ei näin koskaan julkaistu. Esitelmää seuranneessa keskustelussa MATTI HAKAMA kertoi Törnqvistin opastaneen häntä lämpimästi, kun hän oli ollut aloittamassa uraansa Helsingin yliopistolla.

<sup>1</sup>Olen alla paikoin tarkentanut tilaisuudessa puhuttua.



Kuva 4.1: Deborah Ashbyn avausluento 8.6.2023.

TARMO PUKKILA kertoi värikkäistä kokemuksistaan: kannustavasta Tampereen yliopiston ensimmäisestä tilastotieteen professorista EINO HAIKALASTA, vierailustaan tilastotieteen laitoksessa Pittsburgin yliopistossa, jossa kirjoitti C. R. RAON kanssa kaksi artikkelia, JORMA RISSASESTA, ajastaan Tampereen yliopiston rehtorina ja vakuutusmaailmasta. TAPIO NUMMI muisteli Tampereen yliopistolla (lektion jälkeen) noin viisi minuuttia kestänyttä SIMO PUNTASEN väitöstilaisuutta, jossa Haikala oli ollut kustos ja C. R. Rao vastaväittäjä.

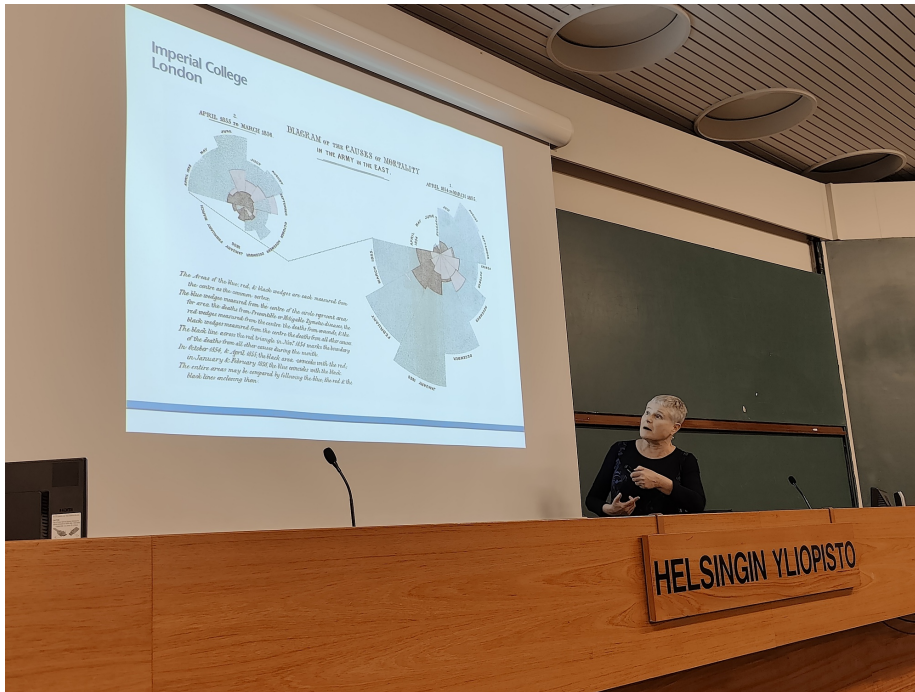
ERKKI PAHKINEN joutui perumaan tulonsa, mutta luin hänen terveisenä läsnäolueille. Niitä olivat muun muassa kuvaus vaaligallupien raportoinnin muuttumisesta. Niiden yhteydessä ei tavattu kertoa kannatusestimaatin keskivirhettä. Pahkinen ja JUHA ALHO puhuivat ja kirjoittivat sen raportoinnin tärkeydestä, ja yleistajuinen virhemarginaali vakiintui gallupien yhteyteen. Siihen myötävaikutti myös Suomen Sanomalehtimiesliiton puheenjohtaja Veikko Löyttyniemi.

## Ensimmäinen päivä

Juhlien avauspuheen piti Royal Statistical Societyn presidenttinä 2019–2020 toiminut DEBORAH ASHBY (Imperial College). Puhetta oli kuulemassa noin satakaksikymmentä tilastotieteen ystävää. (Kuvat 4.1–4.2.) Professori Ashby esitelmöi ”intohimoisena tilastotieteilijänä” ja innovatiivisena tilastografikan kehittäjänä tunnetusta FLORENCE NIGHTINGALEstä. Aihe juontui siitä, että hän oli laatinut puheen vuodelle 2020, jolloin oli kulunut 200 vuotta Nightingalen syntymästä ja 100 vuotta Suomen Tilastoseuran perustamisesta. Aiheen taustalla oli luonnollisesti myös Ashbyn arvostus Nightingalea kohtaan. Nightingale on suurelle yleisölle tutumpi sairaanhoidon kehittäjänä. Ashby oli löytänyt Suomikytköksen Nightingale-koulussa St. Thomasin sairaalassa opiskelleesta SOPHIE MANNERHEIMISTA, CARL GUSTAF EMIL MANNERHEIMIN siskosta. Sophie Mannerheim oli perustamassa Mannerheimin lastensuojeluliittoa ja sai ensimmäisenä suomalaisena Florence Nightingale -mitalin.

Yksi ensimmäisen päivän esitelmöijistä kertoi saaneensa ikäviä yhteydenot-





Kuva 4.2: Deborah Ashby esitelmöi 8.6.2023.

toja koronatutkimuskriitikoilta mutta yrittäneensä keskustella rakentavasti heidän kanssaan. ESA LÄÄRÄN sanastoesitelmä oli erityisen ajankohtainen, sillä Tilastotieteen sanaston 2. laitos oli julkaistu huhtikuussa pdf-tiedostona ja hakukoneena (<https://sanasto.tilastoseura.fi>). Keskustelussa tuotiin esille uuden suomennoksen luotaus (*survey*) idea: luodata eli luo data. Sanasto liittyi kiinteästi myös juhlien satavuotisteemaan. Yksi syy sanaston laatimiselle oli seuran pitkän historian juhlistaminen. Esitelmien lomassa keskusteltiin vilkkaasti.

### Illallinen

Päivä päättyi juhlaillalliseen Kenraalisalissa Katajanokan Kasinolla. Seuran historiaa tällä tavalla juhlistavia oli kahdeksankymmentyhdeksän. Kumolauluyhtye avasi tilaisuuden lauluilla Armaan läheisyys (säv. Toni Edelmann, san. J. W. Goethe, suom. Uno Kailas, sov. Aarre Joutsenvirta) ja Rannalla (san. Saima Harmaja, säv. Tomas Takolander). Laulut olivat seuran perustamisen ajoilta. Jälkimmäinen sopi erityisellä tavalla tilaisuuteen, sillä Saima Harmajan isä oli Tilastoseuran ensimmäinen esimies. Juhlien pääpuhujat Ashby ja JOUNI KUHA saivat muistoksi Aalto-vaasin. Suomen Biostatistiikan Seuran edustajat SAMI HÖKKASEN johdolla onnittelivat Tilastoseuraa. Tunnelma Kenraalisalissa oli sekä juhmallinen että iloinen eikä olisi voinut olla parempi. (Kuvat 4.5–4.6.)

### Toinen päivä

Toisen päivän esitelmät olivat yhtäläillä jännittäviä kuin ensimmäisen. Suomalaisen tilastotieteilijöiden neljän sukupolven esiintyminen oli päivän vaikuttava



Kuva 4.3: Illallinen Katajanokan Kasinolla 8.6.2023.

ja mieleenpainuva erikoisuus.

Juhlien toinen pääpuhja, Jouni Kuha (London School of Economics & Political Science), avasi seuraavan päivän esitelmällään. Empiirinen kysymys oli, kuinka sukupolvet auttavat perheissä vastavuoroisesti eli missä määrin auttaminen korreloi sukupolvien välillä. Teoreettinen ongelma oli, kuinka selittää korrelaatioita regressiomalleilla niin, että regressiokertoimet olisi helppo tulkita ja sovitetut korrelaatiomatriisit olisivat positiivisesti definiittejä.

### Palkinnot

Lopuksi oli palkintojen jaon aika. Kansalliseen tilastojulistekilpailuun otti osaa 21 oppilaitosta. Yläkoulusarjan kolme parasta olivat:

1. “Liha kuluttaa ympäristöä”. OSKARI PEKKARINEN, JOONATAN RAJALA ja LEO NIEMINEN, Nöykkiön koulu, Espoo, opettaja PIIA VAININEN (450 e).
2. “Koronan vaikutukset nuorten hyvinvointiin”. LEANNA KESTI ja UNNA ETELÄINEN, Ounasvaaran peruskoulu, Rovaniemi (350 e).
3. “Nuoret ja alkoholi”. EETU HUIKURI, VILJAMI KOPONEN ja NUUTTI ROINE, Pirkkolan koulu, Kangasala, opettaja JUUSO LINNUSMÄKI (200 e).

Ammattikorkeakoulujen opiskelijoiden ja kandidivaiheen yliopisto-opiskelijoiden sarjassa jaettiin vain 1. palkinto: “Mitä suomalaiset tekevät poistotekstiileillä”. WENYING JI, Helsingin yliopisto (1000 e). Voittajat palkittiin paikanpäällä, ja Ji kertoi kuulijoille tekstiilien päättökäytöstä. Raati ei tällä kertaa jakanut lukiosarjassa palkintoja. Seura järjesti kilpailun yhdessä Tilastokeskuksen ja Matemaattisten aineiden opettajien liitto MAOL ry:n kanssa.

Seura on jakanut vuodesta 1978 lähtien Leo Törnqvist -palkintoa parhaasta tilastotieteen pro gradu -tutkielmasta. (Vuodesta 2018 lähtien on tulkittu, että palkinto voidaan myöntää tutkielmasta, joka on tehty nimellisesti toisen oppiaineen kuten datatieteen alla.) Palkinto on jaettu vuodesta 2011 lähtien joka toinen vuosi.



Kuva 4.4: Illallinen Katajanokan Kasinolla 8.6.2023.



Kuva 4.5: Jouni Kuha esitelmöi 9.6.2023.



Kuva 4.6: Kuulijoita 9.6.2023.

Vuosien 2017–2018 Leo Törnqvist -palkinto (1 000 euroa ja CRC-kustantamon 200 euron lahjakortti) oli ollut tarkoitus jakaa juhlallisesti vuoden 2020 satavuotisjuhlissa, mutta juhlimaan päästiin vasta nyt. Vuosien 2021–2022 Leo Törnqvist -palkinnon (1 000 euroa ja CRC-kustantamon 200 dollarin lahjakortti) sai TOPI HALME tutkielmastaan “Quickest Detection under False Discovery Rate and Communication Constraints” Helsingin yliopistolle. Halme oli juhlien aikaan konferenssissa ulkomailla, joten hän ei päässyt paikalle kertomaan tutkielmastaan.

Viimeinen väitöskirjapalkinto (1 000 e) vuosilta 2017–2020 oltiin jaettu ANNA-KAISA YLITALON ja SANTTU TIKAN kesken. Molemmat olivat esittäneet väitöskirjansa (“Statistical inference for eye movement sequences using spatial and spatiotemporal point processes” ja “Improving identification algorithms in causal inference”) Jyväskylän yliopistossa. Tikka esitelmöi väitöskirjastaan ja vastaanotti Wileyn lahjakortin (150 e:n alennus 300 e:n ostosta) lisäpalkintona.

Posterikilpailuun otti osaa kahdeksan ryhmää. Juhlien osanottajat äänestivät parhaaksi posterin “Bayesian species recognition and abundance estimation: Unraveling the mysteries of salmonid migration in the Teno river”. Seura palkitsi tekijät ANTTI RÄDYN, HENNI PULKKISEN, JAAKKO ERKIN-ARON, PANU ORELLIN, SAMU MÄNTYNIEMEN ja MORTEN FALKEGÅRDIN rahapalkinnolla (500 e) ja Wileyn lahjakortilla.

## Eteenpäin!

Tilastoseuran 100+-juhlat olivat iloinen ja ylevöittävä, ikimuistoinen, tilaisuus. Korkeatasoiset esitelmät ja posterit peilasivat oivasti tilastotieteen laaja-alaisuutta, ja esitelmöitsijät olivat arvovaltaisia. Tunnelma oli sekä kansainvälinen että kotoinen. Aamupäivien esitelmät olivat englanninkielisiä; iltapäivien suomenkielisiä.

Seura pyrkii jatkossa järjestämään Tilastopäivät joka vuosi. Saimme monenlaista ja kiintoisaa enimmäkseen kiittävää palautetta ja ideointia juhlista. Palautteesta otamme opiksi, ja ideoinnista laitamme toimeksi. Teistä pyydämme

tekijöitä. Tilastopäivät ovat juhla, jonka teemme itse eikä kukaan muu. Tervetuloa mukaan niin tieteellisen seuran toimintaan hallitukseen kuin tulevien vuosien Tilastopäiville. Tilastotiede, Tilastoseura ja Tilastopäivät muuttuvat tilastotieteilijöiden mukana.

# Kontrafaktuaalit, eventiivi ja Kilpisenkatu

JUHA KARVANEN

JUHA.T.KARVANEN@JYU.FI

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

JYVÄSKYLÄN YLIOPISTO

## Tiivistelmä

Kausaalipäätelyssä käytetty käsite (kontrafaktuaali), suomen kielen harvinainen tapaluokka (eventiivi) ja Jyväskylän keskustan poikki kulkeva katu (Kilpisenkatu) liittyvät yllättävällä tavalla toisiinsa.

## Johdanto

- Kontrafaktuaalilla tarkoitetaan tosiasioden vastaista tilannetta tai väitettä.
- Eventiivi on harvinainen suomen kielen tapaluokka.
- Kilpisenkatu kulkee Jyväskylän keskustan poikki.

Nämä kolme asiaa, jotka ensi silmäyksellä vaikuttavat täysin riippumattomilta, liittyvät toisiinsa yllättävällä tavalla.

## Kontrafaktuaalit

Kontrafaktuaalipäätelyllä (Pearl, 2009; Kopra ja Tikka, 2016) on tärkeä rooli pohdittaessa vastuuseen ja syyllisyyteen liittyviä kysymyksiä. Lähtökohtana kontrafaktuaalipäätelyssä on kausaalimalli, joka kuvaa syitä ja seurauksia. Olemme havainneet joitakin muuttujia, joiden arvot kuvaavat asioiden todellista tilaa. Tarkastelemme kontrafaktuaalista tilannetta, jossa tietyn muuttujan arvon ajatellaan olevan erilainen ja kysymme, millainen jonkin kiinnostuksen kohteena olevan muuttujan jakauma olisi tässä tilanteessa.

Kontrafaktuaalisen jakauman voi esittää muodossa (Karvanen, Tikka ja Vi-hola, 2024)

$$p(\mathbf{W}_{\text{do}(\mathbf{X}=\mathbf{x})} = \mathbf{w} \mid \mathbf{C} = \mathbf{c}) = \sum_{\mathbf{u}} I(f_{\mathbf{W}}^{\circ}(\mathbf{u}) = \mathbf{w})p(\mathbf{U} = \mathbf{u} \mid \mathbf{C} = \mathbf{c}).$$

Tässä esityksessä  $p(\mathbf{W}_{\text{do}(\mathbf{X}=\mathbf{x})} = \mathbf{w} \mid \mathbf{C} = \mathbf{c})$  on kiinnostuksen kohteena oleva muuttujien  $\mathbf{W}$  kontrafaktuaalijakauma. Ehto  $\mathbf{C} = \mathbf{c}$  määrittää asioiden todel-lista tilaa kuvaavien muuttujien arvot. Merkintä  $\text{do}(\mathbf{X} = \mathbf{x})$  tarkoittaa kontrafaktuaalista interventiota, jossa muuttujien  $\mathbf{X}$  arvoksi kiinnitetään  $\mathbf{x}$ . Muuttujajoukkojen  $\mathbf{C}$  ja  $\mathbf{X}$  ei tarvitse olla erillisiä. Kaavan oikea puoli kertoo, kuinka kontrafaktuaalijakauma voidaan määrittää, jos tilanteeseen liittyvä kausaalimalli tunnetaan rakenneyhtälöineen. Summauksessa käydään läpi kaikki mahdolliset arvot  $\mathbf{u}$ , joita havaitsemattomat taustamuuttujat  $\mathbf{U}$  voivat saada. Indikaattorifunktio  $I(f_{\mathbf{W}}^{\circ}(\mathbf{u}) = \mathbf{w})$  tarkastaa, saako muuttujajoukko  $\mathbf{W}$  arvon  $\mathbf{w}$ . Tässä funktio  $f_{\mathbf{W}}^{\circ}$  kuvaa muuttujajoukon  $\mathbf{W}$  määrittäviä rakenneyhtälöitä kausaalimallissa, johon on kohdistettu interventio  $\text{do}(\mathbf{X} = \mathbf{x})$ . Ehdollinen jakauma  $p(\mathbf{U} = \mathbf{u} \mid \mathbf{C} = \mathbf{c})$  antaa todennäköisyyden tai tiheyden jokaiselle taustamuuttujien mahdolliselle arvolle tilanteessa  $\mathbf{C} = \mathbf{c}$ .

Esimerkki kontrafaktuaalisesta kysymyksestä voisi olla ”Millä todennäköisyydellä tentissä reputtanut opiskelija N. N. olisi päässyt läpi, jos hän olisi tehnyt enemmän harjoitustehtäviä?” Arkikielessä tällaisia pohdintoja kutsutaan jossitteluksi, ja niihin liitetään usein kielteisiä mielikuvia. Siksi valveutunut jossittelija kertookin harrastavansa kontrafaktuaalipäättelyä.

## Eventiivi

Suomen kielen harvoin käytetty tapaluokka, eventiivi (Anhava, 2003), on kuin luotu kontrafaktuaalipäättelyä varten. Eventiivi on konditionaalinen ja potentiaalinen yhdistelmä. Eventiiviä käytetään esimerkiksi seuraavissa lauseissa:

- ”Jos minut kutsuttaisiin tapahtumaan, osallistuneisin etänä.”
- ”Jos olisin myöhästynyt bussista, lieneisin kävellyt.”

Tässä siis muoto ”osallistuneisin” on eventiivin preesens ja ”lieneisin kävellyt” on eventiivin perfekt. Eventiivi siis kertoo todennäköisen tapahtumien kulun jossakin oletetussa tilanteessa. Tilastotieteilijälle tällainen ilmaisutapa on luonteva.

Eventiivi esiintyy myös Kalevalassa. Esimerkiksi tästä käy kolmaskolmatta runo, jossa neuvotaan morsianta seuraavasti:

*Kuules, neiti, kuin sanelen, kuin sanelen, kuin puhelen!*

*Elä suihki sutsunatta eläkä räämi rätsinättä,*

*elä liiku liinasetta, elä kengättä kehaja!*

*Tuosta sulho suuttuneisi, mies nuori nuristuneisi.*

Eventiivimuotoja ovat siis ”suuttuneisi” ja ”nuristuneisi”. Katkelmaa on syytä lukea vain kansanrunoutena, ei nykyajan elämänohjeina.

## Kilpisenkatu

1800-luvun lopulla eventiiiviä harkittiin osaksi kirjakieltä (Häkkinen, 1994, s. 268–269). Näistä aikeista kuitenkin luovuttiin, kun ilmeni, että eventiiivi ei olutkaan alkuperäistä kalevalaista runokieltä vaan sen oli keksinyt sanaseppo Wolmar Schildt, jonka tuttavapiiriin Elias Lönnrot kuului. Piirilääkärinä toiminut Wolmar Schildt (1810-1893) vaikutti Jyväskylässä 1800-luvulla (Hirvonen, 1997). Schildt keksi monia uudissanoja ja ajoi yliopiston perustamista Jyväskylään. Kirjoituksissaan hän käytti usein nimeä Wolmari Kilpinen. Eräs Jyväskylän keskustan kaduista on nimetty Schildtin mukaan Kilpisenkaduksi vuoden 1910 asemakaavasta alkaen (Jäppinen, 2005, s. 104–106).

Kilpisenkatu on siis saanut nimensä Schildtistä, joka on keksinyt kontrafaktuaaliseen puheenparteen soveltuvan eventiiivin. Ilman tätä yhteyttä Tilastoseuran Vuosikirjakin lieneisi kolmea sivua lyhyempi.

## Viitteet

- Anhava, J. (2003). ”Lieneisikö mahdollista?” *Kielikello* 1. URL: <https://kielikello.fi/lieneisiko-mahdollista/>.
- Hirvonen, M. (1997). *Schildt, Wolmar. Kansallisbiografia-verkkojulkaisu*. *Studia Biographica* 4. <http://urn.fi/urn:nbn:fi:sks-kbg-002847>, viitattu 31.5.2024. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Häkkinen, K. (1994). *Agricolasta nykykieleen: Suomen kirjakielen historia*. WSOY.
- Jäppinen, J. (2005). *’Oletko koskaan nähnyt kauniin kaupungin?’ : Jyväskylän ruutuasemakaava-alueen vaiheet 1800-luvulta 2000-luvulle*. Väitöskirja, Jyväskylän yliopisto. Minerva Kustannus.
- Karvanen, J., Tikka, S. ja Vihola, M. (2024). ”Simulating counterfactuals”. *Journal of Artificial Intelligence Research* 80, s. 835–857.
- Kopra, J. ja Tikka, S. (2016). *Itseopiskelumateriaalia: Kausaalimallintamisen perusteet tilastotieteessä*. Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos. URL: <http://urn.fi/URN:NBN:fi:jyu-201605242674>.
- Pearl, J. (2009). *Causality*. Cambridge University Press.



---

# Kaupunkitilastoilla yli satavuotinen historia

ASTA MANNINEN

KAUPUNKITIEDON ASIANTUNTIJA JA

HELSINGIN KAUPUNGIN TIETOKESKUKSEN ENTINEN JOHTAJA

## Tiivistelmä

Tämä artikkeli kertoo kaupunkitilastojen taustasta, joka on osa kaupunkihistoriaa. Kaupungistuminen ja kaupunkien kasvu toivat uusia mahdollisuuksia ihmisille ja yrityksille, mutta myös ongelmia ja haasteita, kuten asuntopulaa ja puutteelliset terveyso-  
lot. Kaupunkien kehittäminen ja ongelmiin tarttuminen edellyttivät tietoa, systemaattista ja riittävän tarkkaa tietoa. Näin syntyivät kaupunkitilastot, joille on leimallista kuvata elämää kaupungeissa aina pienalue- ja korttelitasolle, jopa hienojakoisemminkin tilasto- ja paikkatietoaineistoja yhdistellen. Varhaisiin kaupunkitilastoihin lukeutuvat myös kaupunkien väestölaskennat, ns. kaupunkilaskennat, joita toteutettiin Suomessa kymmenen vuoden välein vuosina 1870–1930.

Kaupunkitilastot ovat vuosien saatossa kehittyneet monipuolisen yhteistyön kautta, jota artikkelissa kuvataan monin esimerkein. Kansalliset ja kansainväliset virallisen tilaston tuottajat, kansainväliset organisaatiot, tutkimuslaitokset ja erilaiset kaupunkien yhteistyöprojektit ja verkostot tuottavat säännöllisesti vertailevaa kaupunkitietoa. Kansainväliset vertailut korostuvat globaalissa taloudessa, jossa kaupungit paitsi tekevät yhteistyötä myös kilpailevat keskenään.

## Kaupunkitilastojen kehittymisen taustaa

Mitkä ilmiöt ja trendit ovat kaupunkitilastojen perustamisen ja kehittymisen takana? Vastauksia tarjoavat kaupunkihistorian, kaupunkitalouden, yhteiskuntatieteiden ja muidenkin tieteenalojen tutkimus. Länsi-Euroopassa kaupungistuminen laajassa mitassa otti vauhtia 1850-luvulla teollistumisen myötä ja teollinen vaihe tuotti talouden ja kaupungistumisen kasvua aina 1960- ja 1970-luvuille. Kaupunkien väkiluvun kasvu tarjosi perustan myös palvelusektorin kasvulle. Nykyään kaupunkien elinkeinorakenne on hyvin palveluvaltainen, esimerkiksi Helsingin työpaikoista lähes yhdeksän kymmenestä on palvelualoilla. Kaupungit erikoistuvat ja suurten kaupunkiseutujen eri alueet ja osat voivat erikoistua omien vahvuksiensa mukaan pärjätäkseen ja pysyäkseen vetovoimaisina. Erityispiirteitä on paljon ja niitä kaupungit ja kaupunkiseudut pyrkivät hyödyntämään ja kehittämään.

Kaupunkien koettu merkityksellisyys on vaihdellut eri aikoina. Suomessa *kaupunkien erioikeudet* määrittivät niiden asemaa aina 1970-luvun lopulle, kun taas Ruotsissa kaupunkien erioikeuksista oli luovuttu jo 1800-luvun loppuun mennessä. Kaikkialla kaupungit ovat osa laajempaa kokonaisuutta ja ilmentävät paitsi paikallista, seudullista ja kansallista kehitystä myös kansainvälisiä, yli rajojen esiintyviä ilmiöitä. Kaupungit sekä kilpailevat keskenään että tekevät yhteistyötä. Pienetkin kaupungit ovat aina liittyneet, ainakin välillisesti, muuhun maailmaan osana kaupunkiverkostoja.

Suomessa kaupungistuminen eteni hitaasti verrattuna Länsi-Eurooppaan ja muihin Pohjoismaihin, esimerkiksi maamme suurimman kaupungin eli Helsingin väkiluku oli 21 000 asukasta vuonna 1850. Sittenkin Helsingin väkiluku kasvoi ripeästi ja kohosi 147 000 asukkaaseen 1920 mennessä. Suomen kaupungistuminen lähti kunnolla liikkeelle vasta 1950-luvulla ja nopeutui merkittävästi 1960-luvulla, kun kaupunkien elinkeinorakenne muuttui. Palveluelinkeinot kasvoivat nopeimmin. Kasvatavat työmarkkinat houkuttelivat ihmisiä kaupunkeihin. Kaupunkiväestön osuus ylitti maaseutuväestön osuuden 1970-luvun alussa.

Kaiken kaikkiaan kaupungistuminen, kaupunkien kehitys ja kasvu muodostavat tärkeän perustan kaupunkitilastojen kysynnälle ja myös niiden ymmärtämiselle ja tulkinnalle. Kaupungit ovat muovautuneet ajan kuluessa ja myös suhteessa erilaisiin taloudellisiin, sosiaalisiin, kulttuurisiin, hallinnollisiin ja moniin muihin tekijöihin. Muuttoliike on kautta aikain vahvasti vaikuttanut kaupunkien kehitykseen. Varsinkin vauraat kaupunkialueet, joissa yritystoiminta on monipuolista, työvoiman kysyntä on suurta ja jotka ovat avoimia maahanmuutolle, vetävät puoleensa muuttajia kotimaasta ja kauempaa. Suomessa kansainvälinen muuttoliike on kasvanut 1990-luvulta lähtien.

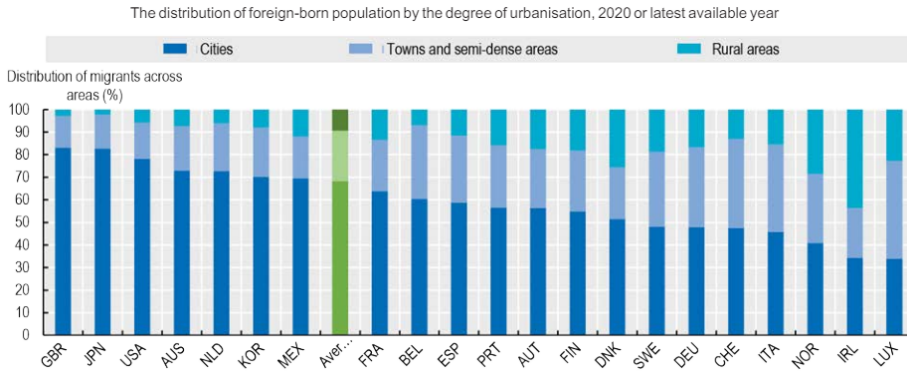
## **Kansainvälinen muuttoliike, suuret kaupungit vetovoimaisia**

Tänä päivänä kansainvälisellä muuttoliikkeellä on suuri merkitys kaupunkien ja maiden kehitykselle. Maailmanlaajuisesti tarkastellen voidaan todeta, että maahanmuuttajat asuvat pääsääntöisesti kaupungeissa (kuva 1, OECD). Enemmän kuin puolet maahanmuuttajista (53 prosenttia) asuu suurilla metropolialueilla, kun vastaavasti vain 40 prosenttia syntyperäisistä asukkaista asuu suurilla metropolialueilla<sup>1</sup>. OECD – maiden kaupungit tarjoavat kodin noin 68 prosentille maahanmuuttajista. Suomessakin Pääkaupunkiseudun suurten kaupunkien (Helsinki, Espoo ja Vantaa) muuttovoitto erityisesti ulkomailta on pysytellyt korkealla tasolla.

## **Kaupungistuminen jatkuu, erityisesti suuret kaupungit kasvavat**

Erityisesti suuret kaupungit kasvavat. Eurostatin vuoteen 2050 saakka ulottuvan väestöennusteen mukaan väestön määrä kasvaa vuosittain Euroopan pääkaupunkien metropolialueilla. Ennuste muille metropolialueille tai suurkaupunkialueille lupaa myös väestökasvua vuosittain aina vuoteen 2040, jonka jälkeen tapahtuisi pientä laskua väestökehityksessä. Asukasluvut muilla kuin metropoli- tai suurkaupunkialueilla on ennustettu vähenevän vuosittain. Kaiken kaikkiaan väestön on ennustettu kasvavan melkein kolmella kaupunkiseudulla viidestä. Kaupunkiseutu tässä Eurostatin laatimassa ennusteessa vastaa alueluokituksen NUTS

<sup>1</sup>OECD Regions and Cities at a Glance 2022. <https://www.oecd.org/cfe/oecd-regions-and-cities-at-a-glance-26173212.htm>



Kuva 6.1: Maahanmuuttajat asuvat pääsääntöisesti kaupungeissa.

Lähde: OECD Regions and Cities at a Glance 2022. OECD 2022. Fig. 3.14 *Migrants live mainly in cities*, s. 75.

Region	Urban population (million)								Percentage urban							
	2015	2020	2025	2030	2035	2040	2045	2050	2015	2020	2025	2030	2035	2040	2045	2050
World	3 981	4 378	4 774	5 167	5 555	5 938	6 312	6 680	53.9	56.2	58.3	60.4	62.5	64.5	66.4	68.4
More developed regions	979	1 003	1 027	1 049	1 070	1 090	1 108	1 124	78.1	79.1	80.2	81.4	82.7	84	85.4	86.6
Less developed regions	3 002	3 375	3 747	4 117	4 485	4 847	5 204	5 556	49	51.7	54.3	56.7	59	61.3	63.4	65.6
Africa	491	587	698	824	966	1 125	1 299	1 489	41.2	43.5	45.9	48.4	50.9	53.6	56.2	58.9
Asia	2 119	2 361	2 589	2 802	2 998	3 176	3 335	3 479	48	51.1	54	56.7	59.2	61.6	63.9	66.2
Europe	547	556	565	572	580	587	593	599	73.9	74.9	76.1	77.5	79	80.6	82.2	83.7
Latin America and the Caribbean	505	539	571	600	626	649	669	685	79.9	81.2	82.4	83.6	84.7	85.8	86.9	87.8
North America	290	304	319	334	349	362	375	386	81.6	82.6	83.6	84.7	85.8	86.9	88	89
Oceania	26	28	30	32	34	36	39	41	68.1	68.2	68.5	68.9	69.4	70.2	71.1	72.1

Kuva 6.2: Kaupunkiväestö ja kaupungistumisaste 2015-2050.

Lähde: World Cities Report 2022. Envisaging the Future of Cities. UN-HABITAT, 2022. Table 1.2 *Urban population and level of urbanization (2015-2050)*, s. 9

3-tasoa (maakunta Suomessa). Todettakoon, että kokoonsa nähden eurooppalainen kaupunki vaihtelee suurista metropoleista, kuten Lontoosta ja Pariisista, suhteellisen pieniin kaupunkeihin. Noin puolet Euroopan kaupungeista olivat pienehköjä, 50 000 – 100 000 asukkaan, kaupunkikeskuksia.

Kaupungistumisaste vaihtelee suuresti maanosasta toiseen ja kaupungistuminen etenee eri tavalla kehittyneissä ja kehittyvissä maissa (kuva 2). Esimerkiksi vuonna 2020 Pohjois-Amerikan kaupungistumisaste oli 82,6 prosenttia, Latinalaisen Amerikan 81,2 prosenttia ja Euroopan 74,9 prosenttia. Suomen kaupungistumisaste (perustuen paikkatietopohjaiseen alueluokitukseen) oli 72,7 prosenttia vuonna 2022 (StatFin tietokanta, Tilastokeskus). Aasian kaupungistumisaste oli 51,1 prosenttia ja Afrikan kaupungistumisaste oli vain 43,5 prosenttia vuonna 2020, mutta kaupungistuminen nopeutuu tulevaisuudessa. YK:n ennusteen mukaan (kuva 2) maailman kaupunkiväestö kasvattaa kauttaaltaan osuuttaan. Kaupungistumisaste nousee voimakkaimmin vähemmän kehittyneissä maissa.

## Euroopan suuriin kaupunkeihin perustetaan tilastotoimistoja

Moninaiset ilmiöt liittyvät kaupunkeihin ja kaupungistumiseen. Taloudellisen toiminnan ja väestön keskittymisen tuomien mahdollisuuksien ohella kaupunkeihin on kertynyt monia ongelmia, kuten ruuhkautuminen, asunnottomuus ja syrjäytyminen. Näitä ongelmia esiintyy yhä, erityisesti suurkaupunkialueilla. Näiden ongelmien ja haasteiden ratkaiseminen edellyttää niiden tuntemista, faktojen hankkimista ongelmista ja niiden syistä. Lisäksi tarvitaan tietoa kaupungin oloista yleensä ja kaupungin toimintaympäristöstä. Tästä asetelmasta kaupunkien päättäjät lähtivät peräänkuuluttamaan systemaattista tietoa kaupungin väestön olojen parantamiseksi ja kaupungin kehittämiseksi.

Euroopan pääkaupungit ja muutamat muut suurkaupungit havahtuivat omien tilastotoimistojen perustamiseen 1800-luvun loppupuolella ja 1900-luvun alussa. Tilastotoimisto perustettiin Berliiniin 1862, Müncheniin 1875, Wieniin 1876, Kööpenhaminaan 1883, Nürnbergiin 1900, Tukholmaan 1905 ja Helsinkiin 1911.

*Helsingin tilastokonttori* aloitti toimintansa 1.1.1911. Viraston perustamiseen vaikutti kaupungin päättäjien havainto siitä, että kaupungin eri alueiden välillä oli isoja eroja kuolleisuuden syissä ja lapsikuolleisuudessa. Helsingin kaupunginvaltuusto edellytti koko kaupungin kattavasti alueittaista tietoa päätöksensä tueksi, jotta se voisi edistää helsinkiläisten terveydentilaa ja hyvinvointia. Niin tilastokonttori ryhtyi systemaattisesti laatimaan tilastoja alueittain erityisesti asukkaiden terveydentilasta, asumisololoista ja kuolleisuuden syistä. Luotettavan ja systemaattisen tiedon saaminen hyvinvointi- ja terveyseroista on yhä ajankohtaista. Sittenkin Helsingin tilastokonttorista kasvoi kattavasti kaupunkitilastoja, kaupunkitutkimusta ja tietopalvelua tuottava että kaupungin historiaa tallentava *Helsingin kaupungin tietokeskus*<sup>2</sup>. Vuonna 2017 toteutetun Helsingin kaupunkiorganisaation kokonaisuudistuksen myötä kaupunkitietopalvelut ovat toimineet osana kaupunginkansliaa<sup>3</sup>.

Todettakoon, että Helsingin kaupungin Rahatoimikamari päätti jo 17.10.1907, että ”Kertomuksiin Helsingin kunnallishallinnosta” liitetyt tilastolliset taulut erotettaisiin erilliseksi kaksikieliseksi ”Statistisk årsbok för Helsingfors stad – Helsingin kaupungin Tilastollinen Vuosikirja” -nimiseksi julkaisuksi. Vuosikirjan sisältö laajeni vuosittain kaupungin kasvun myötä. Vuonna 1936 vuosikirjaan lisättiin pohjoismaiden kaupunkien tilastoja sisältävä luku. Tämä luku on sittemmin laajentunut käsittämään vertailukelpoiset tilastot 16 pohjoismaisesta suurkaupungista ja niiden kaupunkiseuduista. Tilastoja on myös vertailukelpoisesti Baltian maiden pääkaupungeista kaupunkiseutuineen.

Varhaisten kaupunkitilastojen kohdalla on tärkeää mainita myös Tilastollisen päätoimiston 1800-luvun lopun ja 1900-luvun alun vuosikirjat. Nämä vuosikirjat tarjoavat runsaasti relevanttia tilastoa Suomen kaupungeista, suurista ja pienemmistäkin (esim. Vuosikirja 1900). Tilastolliset vuosikirjat ovat yhä keskeisiä tiedon jakajia, joiden erityinen hyve on pitkät aikasarjat. Ne ovat tilastotiedon klassikoita. Tänä päivänä vuosikirjojen ajan saatossa laajentuneet tietosisällöt, niin kaupunkien kuin Tilastokeskuksen laatimien vuosikirjojen, on tehty entistä paremmin saavutettaviksi ja helpommin käytettäviksi.

<sup>2</sup>Helsinki tiedon kohteena. Helsingin kaupungin tietokeskus 100 vuotta. [https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/11\\_08\\_22\\_Helsinki\\_tiedon\\_kohteena.pdf](https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/11_08_22_Helsinki_tiedon_kohteena.pdf)

<sup>3</sup><https://www.hel.fi/fi/paatoksenteoko-ja-hallinto/tietoa-helsingista/tilastot-ja-elamaa-helsingissa/kaupunkitieto-kertoo-helsingin-kehityksesta>

Varhaisiin kaupunkitilastoihin lukeutuvat myös *kaupunkien väestölaskennat*, ns. *kaupunkilaskennat*, joita toteutettiin kymmenen vuoden välein vuosina 1870–1930. Laskentoihin osallistuvien kaupunkien määrä vaihteli vuosikymmenittäin, Helsinki ja Turku osallistuivat kaikkiin laskentoihin. Kaupunkilaskennoissa tiedot kerättiin ruokakunnittain ja julkaistiin kaupunginosittain ja kortteleittain. Helsingin kaupunki suoritti lisäksi oman otospohjaisen laskennan vuonna 1965.

Suomessa monella suurella kaupungilla, kuten Espoolla, Vantaalla, Turulla, Tampereella ja Oululla, on omassa kaupunkiorganisaatiossa kootusti erityinen tilasto-, tutkimus- ja tietopalveluysikkö tai vastaava.

## Kaupunkitilasto kehittyi yhteistyön kautta

Kaupunkitilasto kehittyi yhteistyön avulla ja toimintamalleja on monia. Yhtenäisten ja vertailukelpoisten kaupunkitilastojen kehittäminen edellyttää yhteistä perustaa, yhteisiä käsitteitä, määritelmiä ja luokituksia. Tämän saavuttamiseksi tilastoala on tehnyt pitkäjänteistä yhteistyötä kansallisella ja kansainvälisellä tasolla. Euroopan unionin (EU, Eurostat) lisäksi tärkeitä toimijoita kansainvälisessä tilastoyhteistyössä ovat Yhdistyneet kansakunnat (YK) ja Taloudellisen yhteistyön ja kehityksen järjestö (OECD). Kansalliset tilastokeskukset tekevät tiivistä yhteistyötä mainittujen kansainvälisten organisaatioiden kanssa.

Yhteistyöhön osallistuvia ja tarvittavia tahoja on monia, kuten kaupungit, seudullisia organisaatioita, valtionhallinnon viranomaisia ja organisaatioita, kansalliset tilastokeskukset ja muut virallisen tilaston tuottajat, uudentyypisten aineistojen tuottajat, yliopistot ja korkeakoulut, tutkimuslaitokset, joukko yhteistyöverkostoja, media ja laaja kaupunkitilastojen käyttäjäkunta. Yhteistyö on monella tapaa rajoja ylittävää.

## Kansainväliset kokoukset – uuden esittelyä ja kokemusten vaihtoa

Joka kolmas vuosi järjestettävillä *pohjoismaisilla tilastokokouksilla* (Nordiska statistikermötet) on pitkä perinne<sup>4</sup> ja ne ovat toimineet tilastoalan laajalla kentällä toimivien Pohjoismaiden kollegojen merkittävänä osaamisen ja kokemusten vaihdon foorumina. Tässä yhteistyössä Suomen Tilastoseuralla ja sen sisarseuroilla muissa Pohjoismaissa on keskeinen rooli. Pohjoismaisten tilastokokousten yhteydessä on järjestetty myös kaupunkien tilastokokouksia (Nordiskt kommunalstatistikermöte).

*Kansainvälisen Tilastoinstituutin (ISI)*<sup>5</sup> perheeseen kuului vuonna 1958 perustettu *International Association for Regional and Urban Statistics (IARUS)*. Vuonna 1989 IARUS lakkautettiin ja sen toiminta liitettiin osaksi ISI:n virallisen tilaston järjestöä *International Association for Official Statistics (IAOS)* statuksella pysyvä komitea eli *Standing Committee on Regional and Urban Statistics (SCORUS)*. Kaupunkitilaston kansainvälinen SCORUS-konferenssi pidettiin Helsingissä 1994.

*Tilastoalan maailmankokous* eli ISI Session, myöhemmin nimeltään ISI World Statistics Congress, järjestettiin *Helsingissä 1999*. Kaupunkitilastot (tilastot kaupungeista ja kaupunkiseuduista, pienaluetilastot, jne.) ovat esillä myös ISI maailmankokouksissa. Monet tilastoalan uudet läpimurrot tai kehittämistyön suuntaukset ovat sovellettavissa myös kaupunkitilastoihin, tarjoavathan

---

<sup>4</sup>Pohjoismaisen tilastoyhteistyön satavuotisjuhlaa vietettiin vuonna 1989.

<sup>5</sup>ISI on perustettu 1885 ja käsittelee seitsemän eri tilastotieteen aloihin erikoistunutta alajärjestöä, 21 komiteaa ja jäseniä runsaassa 150 maassa, ks. <https://isi-web.org/about-us>

kaupungit myös oivallisia tilastoalan kehittämistyön kohteita kokonsa, tiheydensä, monipuolisuutensa ja yhdyskuntarakenteensa (perinteiset ja digitaaliset rakenteet) puolesta.

*Tieteellistä yhteistyötä*, yhteistyötä yliopistojen ja korkeakoulujen kanssa on kiittäminen tilastoalan osaamisen varmistajana ja kartuttajana. Kaupunkitilastojen vahvin erityispiirre on päästä lähelle ihmisten arkea ja tämä tarkoittaa tietoa kaupungista ja kaupunkiseudusta (tai työssäkäyntialueesta), tietoa kaupunginosista, pienalueista, kortteleista ja vapaasti rajatuista pieneköistäkin alueista. Tarvitaan tietoa ihmisten liikkumisesta ja muuttamisesta. Silloin on tarve tukeutua moniin tilastollisiin ja muihinkin tieteellisiin menetelmiin. Paikkatietojärjestelmien hyödyntäminen ja spatiaalinen tilastotiede ovat myös kehittäneet kaupunkitilastoja.

Sadan vuoden merkkipaalun vuonna 2020 saavuttanut *Suomen Tilastoseura* on monipuolisessa toiminnassaan, kuten iltapäiväseminaareissa, Tilastopäivissä ja julkaisutoiminnassaan, huomioinut kaupunki- ja aluetilastot. Useat näistä tilaisuuksista ovat tuoneet kansainvälisiä tilastoalan asiantuntijoita paikalle jakamaan uutta tietoa ja tekemään yhteistyötä. Esillä ovat olleet uudet kaupunkiliimiot ja niiden tilastoinnin haasteet, väestöennusteet, ympäristötilastot, pienaluettelut, hyvinvointia kuvaavat tilastot jne. Paikkatietojen hyödyntäminen oli suuresti esillä Tilastoseuran 80 -vuotisjuhlaseminaarissa vuonna 2000<sup>6</sup>.

## Kansallista tilastoyhteistyötä

*Virallisen tilaston* kehitysaskeleet ovat olleet erittäin merkittäviä myös kaupunkitilaston kehittymiselle. Kaupunkien mahdollisuus osallistua virallisen tilaston kehittämiseen sekä tilaston käyttäjinä että myös tilastojen perustietojen tuottajina on ollut palkitsevaa kaikille osapuolille. Hyvä ja pitkäjänteinen yhteistyö on tuottanut tarpeita vastaavia tilastoja, tuotteita ja palveluja. Erityisen merkittävää kaupunkitilastolle ovat monipuoliset ja hienojakoiset aluejaot, sijaintitiedon hyödyntämisen mahdollisuudet, aineistojen yhtenäisyys ja yhdisteltävyys ja tilaston laatu. Laatuun kuuluvat myös virallisen tilaston avoimet tilastoaineistot ja avoimet rajapinnat, samoin kuin se että virallinen tilasto on EU-harmonisoitu. Nämä ominaisuudet antavat lähes rajattomat mahdollisuudet kaupungeille ja muillekin tiedon käyttäjille hyödyntää Tilastokeskuksen ja muidenkin virallisen tilaston tuottajien valtavia tietovarantoja.

*Väestölaskennat* ovat kiistatta yhteiskunnan yksi merkittävimpiä tietolähteitä ja erityinen esimerkki siitä, miten tilastotuotanto laajan yhteistyön tuloksena on kehittynyt raskaasta lomakelaskennasta nopeaan ja tehokkaaseen rekisteripohjaiseen laskentaan. Vuoden 1990 väestölaskenta suoritettiin jo kokonaan rekisteritietoihin perustuvana laskentana. Rekisteripohjaisen väestölaskennan läpimurto mahdollisti sen, että kaikki väestölaskentatiedot voidaan tuottaa vuosittain (eikä joka kymmenes tai mahdollisesti joka viides vuosi).

Tilastokeskukseen perustettiin *Kaupunki- ja seutuindikaattoritietokanta* vuonna 2000 sisäministeriön ja kaupunkien aloitteesta. Valtakunnallisen Kaupunki-indikaattoriprojektin työ ja tulokset<sup>7</sup> toimivat uuden tietokannan

<sup>6</sup>Towards the 3rd Millennium – The 80 Years Anniversary Conference of the Finnish Statistical Society. Spatial Statistics and GIS. Suomen Tilastoseuran Vuosikirja 1999-2000. Helsinki 2000.

<sup>7</sup>Suomalaisia kaupungeja ja kaupunkiseutuja: Kaupunki-indikaattorit 1998. Helsinki 1998. A Portrait of Finnish Cities, Towns and Functional Urban Regions. The Finnish Urban Indicators System. Helsinki 1999. Julkaisija: Kaupunkipoliittikan yhteistyöryhmä, sisäministeriö ja Helsingin kaupungin tietokeskus.

perustana. Kaupunki-indikaattoriprojektiin osallistuivat 38 kaupunkia, sisäministeriö, ympäristöministeriö, Kuntaliitto, STAKES (nykyinen THL, Terveystieteiden tutkimuskeskus ja hyvinvoinnin laitos), Tilastokeskus ja Helsingin kaupungin tietokeskus. Kaupunki- ja seutuindikaattoritietokannan käyttäjäkunta on entisestään laajentunut, sillä Tilastokeskus on lisännyt siihen uuden hyvinvointialueluokituksen 1.1.2022 alkaen. Tietosisältö on laaja ja tarjoaa aikasarjat vuodesta 1990 alkaen, myös kuntatason tiedoista. Näin hyvinvointialueet näkevät omasta alueesta kokonaisuuden ja sen osat ja voivat verrata omaa tilannetta muihin hyvinvointialueisiin. Kaupungit saavat vertailukelpoista tietoa muista maamme kaupungeista seutuineen monenlaisiin käyttötarkoituksiin.

*Elinympäristön tietopalvelu Liiteri*<sup>8</sup> on Suomen ympäristökeskuksen ylläpitämä tieto- ja analyysipalvelu. Se kokoaa yhteen rakennettua ympäristöä ja kaavoitusta koskevia tilasto- ja paikkatietoja, yhteensä yli tuhat karttatasoa ja noin pari tuhatta tilastoa sekä lisäksi kuntien sähköisesti toimittamien asema-kaavojen seurantalomakkeiden tietoja. Tiedontuottajia Liiteriin ovat Suomen ympäristökeskuksen lisäksi muun muassa Tilastokeskus, Digi- ja väestötietovirasto, Maanmittauslaitos, Väylävirasto, Museovirasto, ELY-keskukset, kunnat, ympäristöministeriö, Asumisen rahoitus- ja kehittämiskeskus ARA ja Geologian tutkimuskeskus. Kaupunkien näkökulmasta Liiteri tarjoaa merkityksellistä tietoa moneen keskeiseen tehtävään, kuten maankäyttöön ja maankäytön suunnitteluun, kaavoitukseen, asuinympäristön ja yhdyskuntarakenteen tarkasteluun.

## Seudullisen yhteistyön esimerkkejä

Kaupunkien tilasto- ja tutkimusyhteistyöllä on pitkä perinne Pääkaupunkiseudulla ja laajemmalla Helsingin seudulla. Yhteistyönä kehitetään ja ylläpidetään yhteisiä tuotteita ja palveluja. Yhtenäiset tiedot kaupunkiseudusta ja sen kunnista, ja kuntaa pienemmistäkin alueista, mahdollistaa jo sinänsä monia vertailuja, joista on apua kuntakohtaisten tai yhteisten ongelmien ratkaisemisessa. Kuntien välinen yhteistyö hyötyy myös hyvistä aluetiedoista.

Kaikkia yhteiskunnan sektoreita palvelevana esimerkkinä voidaan mainita *Helsinki Region Infoshare, HRI* ([www.hri.fi](http://www.hri.fi)), joka on Helsingin, Espoon, Vantaan ja Kauniaisten kaupunkien yhteinen avoimen datan palvelu. Palvelu tarjoaa runsaasti pääkaupunkiseudun kaupunkeja ja koko seutua koskevaa tietoa kaikkien vapaasti ja maksutta hyödynnettäväksi. Tarjolla on muun muassa Helsingin seudun väestöä, taloutta ja hyvinvointia kuvaavia tilastoja ja pitkiä aikasarjoja, runsaasti erilaisia paikkatietoaineistoja sekä sovelluskehittäjien suosimia, jopa reaaliaikaista dataa tarjoavia rajapintoja.

Asukkaita on aina kiinnostanut oman asuinalueen vertailu kunnan tai seudun muihin asuinalueisiin esimerkiksi asunnon hankinnan yhteydessä. Asukastoiminta kysyy kaupunginosaa ja sen osa-alueita koskevaa tietoa. *Helsinki alueittain*, ja myöhemmin naapurikunnissa vastaavasti *Espoo alueittain* ja *Vantaa alueittain*, saavutti heti suuren suosion. Yritykset hankkivat perusteelliset ja yksityiskohtaisetkin tiedot alueista esimerkiksi omien sijoitustensa ja suunnitelmiansa pohjaksi. Hyvät alueittaiset tiedot ovat ensiarvoisen tärkeitä kunnille palvelujen suunnittelussa ja toteuttamisessa, kaupunkisuunnittelussa ja kaupunkien johtamisessa ja vuorovaikutuksessa asukkaiden kanssa.

---

<sup>8</sup>[https://www.syke.fi/fi-FI/Palvelut/Elinympariston\\_tietopalvelu\\_Liiteri/Mika\\_Liiteri](https://www.syke.fi/fi-FI/Palvelut/Elinympariston_tietopalvelu_Liiteri/Mika_Liiteri)

*Aluesarjat-tilastotietokannassa*<sup>9</sup>, joka on perustettu 1982 ja uudistettu monta kertaa vuosien varrella, on perustilastoja useista ilmiöistä aikasarjoina pääkaupunkiseudun kunnista alueittain sekä Helsingin seudusta ja Uudelta- maalta kunnittain. Helsingistä, Espoosta ja Vantaalta tietokanta sisältää osal- lueittaista tietoa. Tietokannan tietosisältö jäsenyytensä seitsemän teeman ym- päriin: väestörakenne, väestömuutokset, koulutustaso, asuntokuntien tulot, asuntokunnat ja perheet, rakennukset, asuminen ja rakentaminen, ja työpai- kat. Pisimmät väestösarjat alkavat vuodesta 1962. Aluesarjojen tilastoja voi myös tarkastella suoraan kartalla mm. Maanmittauslaitoksen ylläpitämässä Paikkatietoikkuna-palvelussa.

*Helsingin Seudun Suunnat* (perustettu 1996) on seudun suhdan- nekuvaus, joka on tarjonnut kaikille avoimen ja jatkuvasti ylläpidet- tävän tietovaraston ja julkaisun aina vuoden 2020 loppuun saakka (<https://www.helsinginseudunsuunnat.fi/fi/arkisto>). Vuodesta 2021 eteenpäin Helsingin Seudun Suunnat on neljä kertaa vuodessa verkossa ilmestyvä ajankohtaistarkastus (<https://www.helsinginseudunsuunnat.fi/fi>), joka seuraa seudun kehitystä joukolla indikaattoreita aluetalouden, asunto- ja toimitila- markkinoiden, hyvinvoinnin, liikenteen ja ympäristön, työmarkkinoiden sekä väestön aihealueilta.

Helsingin ympäristötilasto<sup>10</sup> sisältää keskeiset *Helsingin ympäristön tilaa ja kuormitusta kuvaavat indikaattorit*. Ympäristötilaston tietoja hyödynnetään moniin eri tarkoituksiin, kuten kaupungin ympäristöraportoinnin tueksi. Tiedot painottuvat Helsinkiin, mutta tilastot sisältävät vertailutietoja myös muualta pääkaupunkiseudulta, kuudesta suurimmasta kaupungista ja Suomesta.

Kaupungit tekevät pitkäjänteistä työtä *kestävän kehityksen* eteen ja ovat sitoutuneita YK:n kestävän kehityksen Agenda 2030-ohjelmaan. Iso osa Agen- da 2030-tavoitteista toimeenpannaan paikallistasolla ja kaupungeilla on siinä iso rooli ajatellen maankäyttöä ja rakentamista, liikennettä, hyvinvoinnin ja tervey- den edistämistä, kasvatusta ja koulutusta sekä elinvoimaan ja työllisyyteen liitty- viä toimia. Kaupungit myös seuraavat ja arvioivat kestävän kehityksen onnistu- misiaan ja haasteitaan. Esimerkiksi Helsingin kaupunki, ja myös Espoon ja Van- taan kaupungit, julkaisevat vuodesta 2023 alkaen *kestävän kehityksen vapaaeh- toisen paikallisraportin (Voluntary Local Review)*. Helsingissä SDG-tavoitteita<sup>11</sup> seurataan 60 mittarilla<sup>12</sup>.

## Kansainväliset kaupunkivertailut

Vallalla olevat trendit, kuten kaupungistumisen jatkuminen, globalisaatio, il- mastomuutos ja kansainvälinen muuttoliike, korostavat kaupunkien merkitystä. Digitalisaatio vauhdittaa globaalia taloutta, joka käsittää yhä useampia toimia- loja aina palveluihin asti. Kaupungeista ja erityisesti suurkaupungeista on tullut

<sup>9</sup> Linkki tietokantaan: <https://stat.hel.fi/pxweb/fi/Aluesarjat/>  
Linkki tietokannan rajapintapalveluun: <https://stat.hel.fi/api/v1/fi/>  
<https://www.avoindata.fi/data/fi/dataset/helsingin-seudun-aluesarjat-tilastotietokannan-tiedot-paikkatietona>

<sup>10</sup> <https://kaupunkitieto.hel.fi/fi/helsingin-tilastotietokannat>  
Linkki tietokantaan: <https://stat.hel.fi/pxweb/fi/Ymparistotilasto/>  
Linkki tietokannan rajapintapalveluun: <https://stat.hel.fi/api/v1/fi/>

<sup>11</sup> <https://sdgs.un.org/goals>

<sup>12</sup> <https://kestavyys.hel.fi/yhteenvedo-ja-mittarit/>  
<https://www.helsinginseudunsuunnat.fi/fi/artikkelit/paa-kaupunkiseudun-kaupunkien-kestaavan-kehityksen-arvioissa-eroja-mutta-myos-paljon-samaa>



maailmantalouden keskuksia. Ilmiöiden, yritysten ja organisaatioiden, maiden ja maanosien keskinäisriippuvuudet ovat kasvaneet. Tässä maailman tilanteessa kaupungit tekevät paitsi yhteistyötä käyvät myös kilpailua keskenään erottautuakseen ja menestyäkseen.

Kaupunkien näkökulmasta on välttämätöntä voida seurata ja ymmärtää oman kaupungin ja oman kaupunkiseudun tila ja muutokset, jotka pitkälti ovat seurausta muuttuvasta kansainvälisestä toimintaympäristöstä. Tätä tilannekuvaa on tarpeen päivittää tiheästikin, sillä muutokset voivat olla nopeatempoisia, välillä arvaamattomiakin. Luotettavan tilannekuvan muodostaminen edellyttää laadukkaita ja kansainvälisesti vertailukelpoisia tilastoja. Kaupungin päättäjät haluavat tietää, miten uudet käänteet ja muutokset ilmenevät paitsi omassa kaupungissa niin myös muissa kaupungeissa.

Mahdollisimman luotettavan kokonaiskuvan muodostamista oman kaupungin tilasta ja sijoittumisesta muihin kaupunkeihin nähden edesauttavat mahdollisuus käyttää useampia aineistoja tai lähteitä sekä mahdollisuus seurata mittauksia useammalta ajankohdalta. Vertailevaa kaupunkitilastoa ja vertailevia kaupunki-indikaattoritietoa on melko hyvin saatavilla. Tätä tietoa tuottavat kaupungit ja kaupunkiverkostot, kansalliset tilastoviranomaiset, EU/Eurostat, YK, OECD, Maailmanpankki, korkeakoulut, tutkimuslaitokset, konsulttiyritykset ja muutama muukin taho.

Kaupunkivertailut puhuttavat: miten minun kaupunkini pärjää muiden verokkikaupunkien joukossa tai vielä laajemmassa kaupunkijoukossa? Minkälainen vertailu tuottaa mahdollisimman totuuden mukaista, merkityksellistä uutta tietoa? Mistä näkökulmista tai millä viitekehyksellä vertailua olisi järkevää tehdä? Yksinkertaistaen voitaneen todeta, että kaupunkivertailujen viitekehyksiä on karkeasti ottaen kahta lajia: toisaalta yleisiä ja kattavia, toisaalta erityisiä tai fokuoituja.

Tärkeä ohjenuora vertailutiedon käyttäjälle luotettavan lopputukoksen saamiseksi on, että hän käyttäisi useampaa lähdettä ja että hän tarkastelisi useampaa ajankohtaa, ei ainoastaan kehityksen yhtä poikkileikkausta.

## Esimerkkejä kattavista kaupunkivertailuista

Pohjoismaiden kuudentoista suurimman kaupungin yhteistyönä syntyi vuonna 1990 *tietokanta NORDSTAT*<sup>13</sup>. Tietokannassa on kaupunkien tarpeita ajatellen keskeisiä perustilastoja aikasarjana kaupungeista ja niiden kaupunkiseuduista. Myös koko maan tiedot ovat tietokannassa. Tietosisältö rakentuu kansainvälisten määritelmien mukaan ja tietojen ylläpito perustuu kaupunkien ja kaupunkiseutujen omiin tietovarantoihin sekä kansallisten tilastokeskusten virallisen tilaston tietokantoihin. Ylläpidosta vastaavat kaupungit ja lisäksi pääkaupungit koordinoivat työtä ja tarkastavat oman maansa tiedot. Tietokannan tietoihin perustuen on myös laadittu julkaisuja, ensimmäinen vuonna 1992 sekä pohjoismaisilla kielillä että englanninkielisenä, ”Nordiska storstadsregioner i text och siffror” ja ”Major City Regions of Scandinavia – facts and figures”. Sittenmin on julkaistu vain englanninkielinen vuosijulkaisu ”Nordic Major City Statistics”.

Eurostat käynnisti kaupunkitilastojen tuottamisen 1999 pilottiprojektilla *Urban Audit*. Urban Audit käsitti alkuun 59 kaupunkia ja vuosien saatossa osallistuvien kaupunkien määrä on kasvanut noin 370 kaupunkiin. Alusta lähtien Urban Audit, joka sittemmin alettiin kutsumaan nimellä *City Statistics*<sup>14</sup>, on

<sup>13</sup><https://kaupunkitieto.hel.fi/fi/helsingin-tilastotietokannat/nordstat>

<sup>14</sup><https://ec.europa.eu/eurostat/web/cities/information-data>

perustunut vapaaehtoisuuteen ja siihen ovat osallistuneet Eurostat, Euroopan Komission Alue- ja kaupunkipolitiikan pääosasto, kansalliset tilastokeskukset ja kaupungit. City Statistics tarjoaa laajat tilastotiedot elämästä kaupungeissa ja niiden kaupunkiseuduilla. Eurostat on tuottanut julkaisuja ja artikkeleita tämän kaupunkitilastoaineiston pohjalta, kuten vuonna 2016 erikoisjulkaisun ”*Urban Europe – Statistics on Cities, Towns and Suburbs*”<sup>15</sup> ja esimerkiksi ”*Regional yearbook 2019*”<sup>16</sup>, jossa on yksi luku kaupungeista.

Euroopan Komission Alue- ja kaupunkipolitiikan pääosasto on vuodesta 2007 alkaen vuosittain tuottanut haastatteluihin perustuvan kaupunkitutkimuksen ”*Survey on the Quality of Life in European Cities*”. Uusin tutkimus on vuodelta 2023<sup>17</sup>. Tämä tutkimusten sarja tarjoaa vertailukelpoista tietoa koetusta elämän laadusta kaupungeissa ja myös mahdollisuuden tarkastella ja tutkia muutoksia koetussa elämän laadussa. Kaiken kaikkiaan tämä ”perception survey” tarjoaa merkityksellistä tietoa kaikille kaupunkien kehittämiseen osallistuville.

Luokituksia tarvitaan ilmiöiden alueittaisten vaihteluiden kuvaamiseen ja ymmärtämiseen. Suomessa meillä on alueet seitsemään luokkaan jakava kaupunki-maaseutu -luokitus, EU:ssa on *degree of urbanisation* DEGURBA -luokitus<sup>18</sup>, joka jakaa maiden alueet kolmeen luokkaan väestötiheyden perusteella. Nämä luokitukset soveltuvat esim. asuntojen hintakehityksen alueittaisten erojen kuvaamiseen. DEGURBA -luokituksen kehittämiseen osallistuivat monet kansainväliset organisaatiot, kuten EU, YK, OECD, ILO (International Labour Organization) ja Maailmanpankki.

OECD tuottaa paljon hyvää tietoa kaupungeista ja kaupunkiseuduista maailmanlaajuisesti. Esimerkkeinä voidaan mainita vertailukelpoista tietoa tarjoavat ”*OECD Regions and Cities at a Glance 2022*”<sup>19</sup> ja ”*OECD Regional Outlook 2023*”<sup>20</sup>, jotka ilmestyvät joka toinen vuosi. Lisäksi OECD ylläpitää laajoja tilastotietokantoja, jotka palvelevat laajaa käyttäjäkuntaa ja palvelevat hyvin myös kaupungeja ja kaupunkiseutuja.

YK on merkittävä toimija tilastoalalla monella tapaa. YK edistää tilastoalan käytännössä käyttöönottoa kaikkialla sekä laatii väestötilastoja ja -ennusteita ja laatii väestölaskentojen toteuttamisen suosituksia. Uudempaa tilasto- ja tutkimustoimintaa edustavat kaupungistumisen ja kestävä kehityksen laaja-alainen seuranta sekä kaupunkien mahdollisten tulevaisuuskuvioiden pohdinta. Näistä ajankohtaisista suurista teemoista YK tuottaa säännöllisesti vertailukelpoisia tilastoja ja indikaattoreita sekä ylläpitää laajat tilastotietokannat. Esimerkkeinä uusimmista raporteista voidaan mainita ”*The World Urbanization Prospects: The 2018 Revision*”<sup>21</sup>, ”*World Cities Report 2022. Envisaging the Future of Cities*”<sup>22</sup> ja ”*World Population Prospects 2022*”<sup>23</sup>.

<sup>15</sup><https://ec.europa.eu/eurostat/en/web/products-statistical-books/-/ks-01-16-691>

<sup>16</sup><https://ec.europa.eu/eurostat/en/web/products-flagship-publications/-/ks-ha-19-001>

<sup>17</sup>[https://ec.europa.eu/regional\\_policy/sources/reports/qol2023/2023\\_quality\\_life\\_european\\_cities\\_en.pdf](https://ec.europa.eu/regional_policy/sources/reports/qol2023/2023_quality_life_european_cities_en.pdf)

<sup>18</sup><https://ec.europa.eu/eurostat/web/degree-of-urbanisation/information-data>

<sup>19</sup><https://www.oecd.org/regional/oecd-regions-and-cities-at-a-glance-26173212.htm>

<sup>20</sup><https://www.oecd.org/regional/oecd-regional-outlook-2dafc8cf-en.htm>

<sup>21</sup><https://www.un.org/en/desa/2018-revision-world-urbanization-prospects>

<sup>22</sup>[https://unhabitat.org/sites/default/files/2022/06/wcr\\_2022.pdf](https://unhabitat.org/sites/default/files/2022/06/wcr_2022.pdf)

<sup>23</sup><https://population.un.org/wpp/Publications/>

## Esimerkkejä erityisistä tai fokusoiduista kaupunki-indikaattoreista

Seuraavaksi kaksi esimerkkiä erityisistä tai fokusoiduista kaupunkivertailuista, eurooppalainen ”*Cultural and Creative Cities Monitor*”<sup>24</sup> ja globaali kilpailukykyindeksi ”*Global Power City Index, GPCI*”<sup>25</sup>. Molemmat kaupunkivertailut ovat rakenteeltaan komposiitti-indeksejä ja molempia tuotetaan säännöllisesti, aikasarjaa on jo syntynyt.

*Cultural and Creative Cities Monitor* (C3 Monitor) on Euroopan Komission Yhteisen tutkimuskeskuksen (Joint Research Centre, JRC) kehittämä mittaristo tai työkalu, jolla arvioida kulttuurin ja luovien alojen kehitystä Euroopan kaupungeissa ja jolla kaupungit voivat tehdä vertailuja ja oppia toisiltaan. Mittariston viitekehysten muodostavat *kolme päänäkökulmaa* eli kulttuuri pulssina kaupungeissa (Cultural Vibrancy), luova talous (Creative Economy) ja mahdollistava toimintaympäristö (Enabling Environment), jotka jakautuvat seuraavalla tasolla yhdeksään ulottuvuuteen. Osallistuvia kaupunkeja oli 196 vuonna 2023, niiden joukossa mm. 102 Euroopan kulttuuripääkaupunkia (European Capitals of Culture) ja 29 UNESCO:n luovien kaupunkien -verkoston jäsentä (UNESCO Creative Cities). Suomesta ovat mukana Helsinki, Espoo, Turku ja Tampere.

*Global Power City Index* asettaa suurkaupungit (pääkaupungit ja muut suuret kaupungit maailmassa, yhteensä 48 kaupunkia) paremmuusjärjestykseen, sen mukaan miten ne kokonaisvaltaisen tarkastelun perusteella pärjäävät globaalissa kilpailussa. Indeksii pyrkii myös paljastamaan kaupunkien vahvuudet, haasteet ja heikkoudetkin muuttuvassa maailmassa. Euroopasta on monta kaupunkia mukana, Suomesta Helsinki. Indeksii tarkastelee kustakin kaupungista *kuusi ulottuvuutta* ja samalla tutkitaan, ketkä käyttävät kaupunkia, sen tilaa ja palveluita. Tarkastelun kohteena olevat kuusi ulottuvuutta ovat: talous, tutkimus ja kehittäminen, kulttuurinen vuorovaikutus, elämänlaatu, ympäristö ja saavutettavuus. Kaupungin käyttäjät on jaettu neljään ryhmään: yksi paikallinen toimija eli kaupungin asukkaat, ja kolme globaalia kaupungin käyttäjää tai toimijaa eli yritysjohtajat, korkeasti koulutettu työvoima ja turistit. Indeksii viitekehys ja laajat aineistot tarjoavat mahdollisuuden moniin mielenkiintoisiin ja hyödyllisiin analyysiin.

Muitakin kaupunkien elämänlaatua mittaavia indeksejä on, kuten *Mercer Quality of Living Survey*<sup>26</sup> (241 kaupunkia viidestä maanosasta) ja *the Economist Intelligence Unit's Global Liveability Index*<sup>27</sup> (172 kaupunkia eri maanosista).

## Johtopäätöksiä ja pohdintaa

Kaupungit kiinnostavat, koska kaupunkeihin muutetaan yleensä paremman elämän ja runsaampien mahdollisuuksien vuoksi. Arki kaupungeissa on hyvin moninaista ja monimuotoista. Jokaisella kaupungilla on erityispiirteensä. Kaupunkitilastot kuvaavat elämää kaupungeissa ja sen osissa, kaupungin keskustassa ja

---

<sup>24</sup> <https://composite-indicators.jrc.ec.europa.eu/cultural-creative-cities-monitor> ja <https://publications.jrc.ec.europa.eu/repository/handle/JRC134516>

<sup>25</sup> [https://mori-m-foundation.or.jp/pdf/GPCI2023\\_summary.pdf](https://mori-m-foundation.or.jp/pdf/GPCI2023_summary.pdf) ja [https://mori-m-foundation.or.jp/pdf/GPCI2023\\_release\\_en.pdf](https://mori-m-foundation.or.jp/pdf/GPCI2023_release_en.pdf)

<sup>26</sup> <https://www.mercer.com/insights/total-rewards/talent-mobility-insights/quality-of-living-city-ranking/>

<sup>27</sup> <https://www.economist.com/graphic-detail/2023/06/21/the-worlds-most-liveable-cities-in-2023>

esikaupunkialueella, kaupunginosissa ja asuinalueilla, sekä kaupunkiseudulla ja sen eri osissa. Paikallisten olojen kuvaukseen on helppo samaistua.

Moninaiset ilmiöt liittyvät kaupunkeihin ja kaupungistumiseen. Taloudellisen toiminnan ja väestön keskittymisen ohella kaupunkeihin on kertynyt monia ongelmia, kuten ruuhkautuminen, asuntopula ja syrjäytyminen. Näitä ongelmia esiintyy yhä, erityisesti suurkaupunkialueilla. Ongelmat ovat monimutkaisia ja niiden ratkaiseminen haasteellista, tarvitaan monipuolista kaupunkitietoa. Tarvitaan hyvää perustilastoa, joka on riittävän hienojakoista niin aluejako- kuin sisällöllistenkin luokitusten osalta. Tarvitaan kaupunkitutkimusta, joka auttaa ymmärtämään kaupunki-ilmiöitä syvemmin. Aikasarjat ovat välttämättömiä muutosten seurannan ja tulevan ennakkoinnin mahdollistamiseksi.

Vuosittaiset alueittaiset väestöennusteet<sup>28</sup>, talouskehityksen ennakointi ja työpaikkaprojektit sekä hyvinvointi- ja kestävä kehityksen raportit ovat keskeisiä kaupungin johdon työvälineitä. Koulutus, kulttuuri, matkailu ja monenlaiset palvelut kuuluvat kaupunkielämään ja näistä kaupungit sekä tuottavat että käyttävät runsaasti tilasto- ja tutkimustietoa.

Kaupunkien elinkeinorakenne on hyvin palveluvaltainen, esimerkiksi Helsingissä palvelualat edustivat 89 prosenttia työpaikoista vuonna 2022 (koko Suomen vastaava osuus oli 74 prosenttia). Digitalisaatiokehitys vaikuttaa vahvasti palvelualoihin ja laajemminkin kaupunkien elinkeinorakenteisiin. Kaupunkien palvelusektori on hyvinkin erikoistunut, esimerkiksi Suomen Pääkaupunkiseudulla tukku- ja vähittäiskauppa, hallinto- ja tukipalvelutoiminta sekä informaatio- ja viestintäala ovat palvelusektorin toimialojen suurimpia työllistäjiä. ICT-sektorin kohdalla Pääkaupunkiseudun merkitys on vielä alueen yleistä taloudellista painoarvoa suurempi. ICT-ala sijoittuu yleensä helposti saavutettaviin ja riittävän suuriin ja tiiviisiin kaupunkeihin, joissa palvelutarjonta on monipuolista. Suomen ICT-sektorin henkilöstöstä reilut 56 prosenttia työskenteli Pääkaupunkiseudulla vuonna 2022<sup>29</sup>. Pääkaupunkiseutu ja Uusimaa kokonaisuudessaan sijoittuvat informaatio- ja viestintäalan henkilöstöosuudellaan Euroopan kärkeen yhdessä esimerkiksi Tukholman ja Kööpenhaminan seutujen kanssa.

Tällä hetkellä talous koettelee kaupunkia maailmanlaajuisesti. Kaupungit etsivät uutta kasvua ja samalla ne ovat sitoutuneet edistämään kestävä kehitystä. Kestävä kilpailukyvyyn ja hyvinvoinnin saavuttamiseksi on onneksi monta polkua, mahdollisuudet pitää nähdä. Tässä auttavat luotettavaan tietoon pohjautuva kuva oman kaupungin tilasta ja mahdollisuuksista ja ajantasainen vertaileva tieto muista kaupungeista. Kaupunkikehitys on myös jatkuvaa ideointia, jota hyvä tieto pystyy ruokkimaan. Kaupungit tarjoavat otollisen paikan testata uusien teknologioiden ja innovaatioiden toimivuutta. Nopeat muutokset ja uudet ilmiöt pitävät huolen siitä, että kaupunkitilastot uudistuvat.

## Lähteet ja kirjallisuus

Alberti, V., Caperna, G., Mauri, C., Panella, F., Tacao Moura, C., Banys, K., Symeonidis, K., Dominguez Torreiro, M., Saisana, M. (2023). *Cultural and Creative Cities Monitor*. Report and Annexes 2023. Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/975348, JRC134516.  
<https://composite-indicators.jrc.ec.europa.eu/>

<sup>28</sup> [https://www.hel.fi/static/kanslia/Kaupunkitieto/23\\_12\\_21\\_Tilastoja\\_8.pdf](https://www.hel.fi/static/kanslia/Kaupunkitieto/23_12_21_Tilastoja_8.pdf)

<sup>29</sup> [https://www.hel.fi/static/kanslia/Kaupunkitieto/23\\_11\\_03\\_Tilastoja\\_7\\_Mikkola.pdf](https://www.hel.fi/static/kanslia/Kaupunkitieto/23_11_03_Tilastoja_7_Mikkola.pdf)

cultural-creative-cities-monitor ja <https://composite-indicators.jrc.ec.europa.eu/cultural-creative-cities-monitor/docs-and-data>

*A Portrait of Finnish Cities, Towns and Functional Urban Regions. The Finnish Urban Indicators System.* The Committee for Urban Policy, Ministry of the Interior and City of Helsinki Urban Facts. Editor: Tarja Pyöriä, Ministry of the Interior. Fagepaino Ltd, Helsinki 1999.

Clark, Peter (2009). *European Cities and Towns 400-2000*. Oxford University Press. Inc. New York, 2009.

Clark, Peter (2012). *Helsinki eurooppalaisena pääkaupunkina*. Kvartti 3/2012, s. 26-40. Helsingin kaupungin tietokeskus. Tammerprint Oy, Tampere 2012.

*City of Helsinki Urban Facts 'Centenary Conference, 26-27 May 2011*. Helsinki Quarterly 3/2011, City of Helsinki Urban Facts 1911-2011, s. 6-42. Helsingin kaupungin tietokeskus. Tammerprint Oy, Tampere 2011.

Dijkstra, Lewis and Poelman, Hugo (2011): *Regional Typologies Overview*. European Union, DG Regio, latest update 3 November 2020. Eurostat, Statistics Explained, online publications.

*Eurostat Regional Yearbook 2023 edition*. Eurostat, Statistics Explained, Online publication. Last updated 26 March 2024. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Eurostat\\_regional\\_yearbook](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Eurostat_regional_yearbook)

*Eurostat Regional Yearbook 2019 edition*. Statistical Books, Publications Office of the European Union, 2019. <https://ec.europa.eu/eurostat/web/products-flagship-publications/-/KS-HA-19-001>

*Global Power City Index 2023*. Institute for Urban Strategies, The Mori Memorial Foundation. <https://mori-m-foundation.or.jp/english/ius2/gpci2/index.shtml>

*Helsinki alueittain 2019*. Helsingin kaupunki, kaupunginkanslia, kaupunkitietopalvelut, 2020. Otavan kirjapaino Oy, Keuruu 2020. [https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/20\\_06\\_10\\_Hki\\_Aluettain\\_2019\\_Tikkanen.pdf](https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/20_06_10_Hki_Aluettain_2019_Tikkanen.pdf)

*Helsingin kaupungin tilastollinen vuosikirja 100 vuotta. – Helsingfors stads statistiska årsbok 100 år*. 26.11.2006. Paintek Pihlajamäki Oy, Helsinki 2007.

*Helsingin kaupungin tietokeskuksen toimintakertomus 2011. 100 vuotta kaupunkitietoa*. Helsingin kaupungin tietokeskus, 2012. [https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/12\\_04\\_17\\_Toimintakertomus%202011.pdf](https://www.hel.fi/hel2/tietokeskus/julkaisut/pdf/12_04_17_Toimintakertomus%202011.pdf)

*Helsinki tiedon kohteena. Helsingin kaupungin tietokeskus 100 vuotta*. Helsingin kaupungin tietokeskus 2011. Toimittajat: Timo Cantell ja Tero Lahti. Bookwell Oy. Porvoo 2011.

*Helsingin tilastollinen vuosikirja 2023*. (111. vuosikerta) Helsingin kaupunki, kaupunginkanslia, kaupunkitietopalvelut 2023. Otavan kirjapaino Oy, Keuruu 2023.

*Helsingin seudun suunnat. Ajankohtaista tietoa seudun kehityksestä.* <https://www.helsinginseudunsuunnat.fi/fi>

*Helsinki Region Infoshare – avoimen datan palvelu.*  
<https://hri.fi/fi/>

*Helsingin tilastotietokannat.*  
<https://kaupunkitieto.hel.fi/fi/helsingin-tilastotietokannat>

Hietala, Marjatta (1992). *Innovaatioiden ja kansainvälistymisen vuosikymmenet*. Helsinki eurooppalaisessa kehityksessä 1875-1917. Tietoa, taitoa ja asiantuntemusta, osa 1, s. 9-66. Suomen historiallinen seura ja Helsingin kaupungin tietokeskus. Historiallinen Arkisto 99:1/SHS ja Helsingin kaupungin tietokeskuksen Tutkimuksia 1992:5:1. Gummerus Kirjapaino Oy. Jyväskylä 1992.

*Historiallisen tilastotiedon opas. Väestölaskennat 1749-1995 ja Kaupunkien väestölaskennat 1870-1930* (ns. kaupunkilaskennat). Tilastokeskus. <https://guides.stat.fi/historiallisentilastotiedonopas/vaestolaskennat#kaupunkien-vaestolaskennat>

Hoffren, Jukka (2023). *Gloaali kestävän kehityksen indikaattorikokoelma on lähes valmis*. Tieto&Trendit blogi 21.2.2023. Tilastokeskus 2023.  
<https://www2.tilastokeskus.fi/tietotrendit/blogit/2023/gloaali-kestavan-kehityksen-indikaattorikokoelma-on-lahes-valmis/>

IAOS, International Association for Official Statistics, <https://iaos-isi.org/> and SCORUS, Standing Committee for Regional and Urban Statistics, <https://www.scorus.org/>

*Kestävä Helsinki. Yhteenveto ja mittarit*. Helsingin kaupunginkanslia, 2023.  
<https://kestavyys.hel.fi/yhteenveto-ja-mittarit/>

Kaupunki- ja seutuindikaattorit. Tilastokeskus.  
<https://stat.fi/tup/kasit/index.html>

Laakso, Seppo ja Loikkanen, Heikki A. (2004). *Kaupunkitalous. Johdatus kaupungistumiseen, kaupunkien maankäyttöön sekä yritysten ja kotitalouksien sijoittumiseen*. Gaudeamus kirja/Oy Yliopistokustannus. Tammer-Paino, Tampere 2004.

Laakso, Seppo ja Kostiainen, Eeva: *European Metropolises. Recession and Recovery. Statistics 2011:19*. Helsingin kaupungin tietokeskus 2011.

Laakso, Seppo ja Kostiainen, Eeva (2013). *Alueellisesti eriytynyt Eurooppa. Helsinki ja Itämeren alue Euroopan alueiden verkostossa*. Kvartti 2/2013, s. 56-67. Helsingin kaupungin tietokeskus 2013. Tammerprint Oy, Tampere 2013.

Manninen, Asta (2017). *Metropolien merkitys kasvaa – miten pärjää Helsinki?* Tieto&Trendit – Talous- ja hyvinvointikatsaus 2/2017, s. 27-34. Tilastokeskus.

Manninen, Asta (2016). *Mitä kansainväliset vertailut kertovat Helsingistä?* Kvartti 1/2016, s. 6-27. Helsingin kaupungin tietokeskus. Libris Oy, Helsinki 2016.

Marchetti Stefano, Giusti Caterina, Pratesi Monica, Salvati Nicola, Giannotti Fosca, Pedreschi Dino, Rinzivillo Salvatore, Pappalardo Luca ja Gabrielli Lorenzo (2015): *Small Area Model-Based Estimators Using Big Data Sources*. Journal of Official Statistics, Vol. 31, No. 2, 2015, pp. 263–281, <http://dx.doi.org/10.1515/JOS-2015-0017>

*Metropolialueen talous. Näkökulmia kaupunkitalouden ajankohtaisiin aiheisiin.* (Osana Kaupunkitutkimus- ja metropolipolitiikkaohjelmaa 2010-2012.) Helsingin kaupungin tietokeskus 2012. Toimittajat: Heikki A. Loikkanen, Seppo Laakso ja Ilkka Susiluoto.

Mikkola, H. (2023). *Yritysten toimialarakenne pääkaupunkiseudulla*. Tilastoja 2023: 7. Helsingin kaupunginkanslia, kaupunkitietopalvelut. [https://www.hel.fi/static/kanslia/Kaupunkitieto/23\\_11\\_03\\_Tilastoja\\_7\\_Mikkola.pdf](https://www.hel.fi/static/kanslia/Kaupunkitieto/23_11_03_Tilastoja_7_Mikkola.pdf)

*Nordstat- tietokanta.* <https://kaupunkitieto.hel.fi/fi/helsingin-tilastotietokannat/nordstat>

*Nordiska storstadsregioner i text och siffror* 1992. Samnordisk publikation utarbetad av följande städer i samarbete med regionerna København, Aalborg, Odense, Århus, Helsingfors, Tammerfors, Uleåborg, Åbo, Reykjavik, Oslo, Bergen, Stavanger, Trondheim, Stockholm, Göteborg, Malmö samt Nordiska ministerrådets basprojekt. Helsingfors stads faktacentral, Helsingfors 1992.

*OECD Regions and Cities at a Glance 2022.* OECD, published on November 15, 2022. <https://www.oecd.org/cfe/oecd-regions-and-cities-at-a-glance-26173212.htm>

*OECD Regional Outlook 2023. The Longstanding Geography of Inequalities.* OECD, published on October 02, 2023. <https://www.oecd.org/regional/oecd-regional-outlook-2daf82cf-en.htm>

Pasila, Aura (2016): *Enriching census data with the Journey Planner to produce statistics on commuting*. Presentation at Nordiskt Statistiker möte, Stockholm 22-24 August 2016.

Piela, Pasi (2017). *Kohti kestävä työmatkailua – polkisitko puolesta tunnissa töihin?* Tieto& Trendit – Talous- ja hyvinvointikatsaus 2/2017, s. 35-40. Tilastokeskus.

*Population projections at regional level* (data extracted in March 2021, planned update 2024). Eurostat, Statistics Explained, online publications. Eurostat, 20 November 2023.

*Quality of Life in European Cities*, 2023. (Running since 2007, this survey provides a unique insight into city life. It gathers the experiences and opinions of city dwellers across Europe.) European Commission, DG Regional and Urban Policy, 2023.

[https://ec.europa.eu/regional\\_policy/sources/reports/qol2023/2023\\_quality\\_life\\_european\\_cities\\_en.pdf](https://ec.europa.eu/regional_policy/sources/reports/qol2023/2023_quality_life_european_cities_en.pdf) and

<https://ec.europa.eu/eurostat/web/cities/information-data>. For more information on the survey see: [https://ec.europa.eu/regional\\_policy/information-sources/maps/quality-of-life\\_en](https://ec.europa.eu/regional_policy/information-sources/maps/quality-of-life_en)

*Quality of Living City Ranking 2023*. Mercer 2023.

<https://www.mercer.com/insights/total-rewards/talent-mobility-insights/quality-of-living-city-ranking/>

*StatFin- tilastotietokanta*. Tilastokeskus.

<https://stat.fi/tup/statfin/index.html>

*Statistical Journal of the IAOS – Volume 33, issue 3 (2017). Special issue with the theme "Urban, Regional and Small Area Statistics"*.

Editors: Kirsten West, Editor-in-Chief and Asta Mammen Guest Editor, Statistical Journal of the IAOS. IOS Press. <https://content.iiospress.com/journals/statistical-journal-of-the-iaos/33/3>

*Sinetti-arkistotietojärjestelmä*. Helsingin kaupunginarkiston aineistot. <https://hri.fi/data/fi/dataset/helsingin-kaupunginarkiston-aineistot>

<https://hri.fi/data/fi/dataset/helsingin-kaupunginarkiston-aineistot>

Sinkko, Harri (2023). *Helsingin ja Helsingin seudun väestöennuste 2022–2060. Ennuste alueittain 2022–2037*. Tilastoja 2023: 8.

Helsingin kaupunki, kaupunginkanslia, kaupunkitietopalvelut, 2023. [https://www.hel.fi/static/kanslia/Kaupunkitieto/23\\_12\\_21\\_Tilastoja\\_8.pdf](https://www.hel.fi/static/kanslia/Kaupunkitieto/23_12_21_Tilastoja_8.pdf)

*Suomenmaan tilastollinen vuosikirja 1900*, (21. vuosikerta).

Tilastollinen päätoimisto 1900. Suomalaisen Kirjall. seuran kirjapainon osakeyhtiö ja kustantamo. Helsingissä 1900.

<https://www.doria.fi/handle/10024/67164>

*Suomi lukuina 2023 – avainluvut meistä*. Tilastokeskus 2023.

<https://stat.fi/tup/suoluk/index.html>

*Suomen tilastollinen vuosikirja 2023*. Tilastokeskus, 2023.

<https://stat.fi/tup/julkaisut/suomen-tilastollinen-vuosikirja>

*Suomen Tilastoseuran Vuosikirja 1999-2000, Suomen*

*Tilastoseuran juhlakirja. 'Towards the 3rd Millennium – The 80 Years Anniversary Conference of the Finnish Statistical Society. Spatial Statistics and GIS.'* pp. 23-180, and 'Kansainvälisen Tilastoinstituutin 52. maailmankokous Suomessa 1999' by secretary-general Ilkka Mellin, pp.11-20. Edita Oy, Helsinki 2000.

Tammilehto-Luode Marja, Nieminen Jari and Ruotsalainen Kaija (20.9.2016): *Memo on small area statistics in Finland*. Statistics Finland.

*The world's most liveable cities in 2023*. The Economist 2023.

<https://www.economist.com/graphic-detail/2023/06/21/the-worlds-most-liveable-cities-in-2023>

*Tilastotietoja Helsingistä 2023*. Helsingin kaupunki, kaupunginkanslia, kaupunkitietopalvelut.

<https://kaupunkitieto.hel.fi/fi>

Tilastokeskus. <https://stat.fi/>

*Tilastoperiaatteet ja laatu*. Virallisen tilaston periaatteet, Euroopan tilastojen käytännesäännöt ja Kansainvälisen tilastoinstituutin (International Statistical Institute, ISI) ammattietiikkajulistus. Tilastokeskus,

<https://stat.fi/org/tilastotoimi/tilastoperiaatteet-ja-laatu.html>



Tutkimus- ja tilastotietoa Helsingistä – Kaupunkitieto kertoo Helsingin kehityksestä. <https://kaupunkitieto.hel.fi/fi>

*Urban Europe*. Eurostat, Statistics Explained, Online publication. Last updated 27 January 2023. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Urban\\_Europe](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Urban_Europe)

*Urban Europe – statistics on cities, towns and suburbs. 2016 edition*. Eurostat 2016. Statistical books. Publications office of the European Union, 2016.

*Uudenmaan maakunnan vieraskielisen väestön ennuste vuosille 2022-2040*. Helsingin kaupunginkanslia, Kaupunkitietopalvelut; Espoon kaupunki, Tutkimus ja tilastot; Vantaan kaupunki, Strategia- ja tutkimusyksikkö; Uudenmaan liitto. Tilastoja 2023:5, Helsingin kaupunki, kaupunginkanslia, kaupunkitietopalvelut. Paino Next Print Oy, Helsinki 2023.

Vartiainen, Marjukka (2016). *Helsingin historiaa numeroina kolmelta vuosisadalta. Opas Helsingin seudun aluesarjat-tilastotietokannan pitkien aikasarjojen käyttöön. Helsingin kaupungin tietokeskus*, Työpapereita 2016:1. [https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/16\\_01\\_11\\_Tyopapereita\\_1\\_Vartiainen.pdf](https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/16_01_11_Tyopapereita_1_Vartiainen.pdf)

*World Cities report 2022: Envisaging the Future of Cities*. UN-Habitat, 2022.

[https://unhabitat.org/sites/default/files/2022/06/wcr\\_2022.pdf](https://unhabitat.org/sites/default/files/2022/06/wcr_2022.pdf)

World Population Prospects 2022. 2022 Revision. United Nations. Department of Economic and Social Affairs and Population Division. <https://population.un.org/wpp/>

*World Urbanization Prospects. The 2018 Revision*. United Nations Department of Economic and Social Affairs and Population Division. United Nations, New York, 2019. <https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf>

Yamato, N., Hamada, Y., Dustan, P., Jimbo, H. and Isogaya, H. (2023). *Global Power City Index 2023. Summary Report*. Institute for Urban Strategies, The Mori Memorial Foundation. The Mori Memorial Foundation, 2023.

[https://mori-m-foundation.or.jp/pdf/GPCI2023\\_summary.pdf](https://mori-m-foundation.or.jp/pdf/GPCI2023_summary.pdf)

# Leo Törnqvist -palkinto

Tilastoseura on jakanut vuodesta 1978 lähtien Leo Törnqvist -palkintoa parhaasta tilastotieteen pro gradu -tutkielmasta. Palkinto on jaettu vuodesta 2011 lähtien joka toinen vuosi.

## Transformer -neuroverkon robustisuuden parantaminen Gaussisen prosessin ja neuroverkon yhdistävillä menetelmillä

ANTTI YLÄJÄRVI

ANTTI.YLAJARVI@COMBITECH.COM

JYVÄSKYLÄN YLIOPISTO

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

### Tiivistelmä

Neuroverkkoihin perustuvia ennustemalleja käytetään tänä päivänä hyvin monilla sovellusaloilla, joista useilla myös tieto ennusteen luotettavuudesta on tärkeää. Malleja reaali maailmaan sovellettaessa yhtä lailla tärkeää on mallin robusti yleistyminen. Edellisiä tavoitteita ei usein riittävästi saavuteta tavanomaisilla deterministisillä neuroverkoilla. Yksi tutkimussuunta tässä kontekstissa ovat menetelmät, joissa neuroverkkomalli yhdistetään Gaussiseen prosessiin käyttämällä kovarianssifunktiossa neuroverkolla johdettuja piirteitä ja estimoimalla kaikki yhdistetyn mallin sisältämät parametrit stokastisella gradienttimenetelmällä tai jollakin tämän muunnoksella. Tutkielmassa esiteltiin kaksi tällaista menetelmää ja arvioitiin, voidaanko näillä parantaa Transformer -arkkitehtuuriin perustuvan neuroverkkomallin robustisuutta ja ennusteen epävarmuuden kvantifointia luonnollisen kielen käsittelyyn liittyvässä tehtävässä.

## Johdanto

Neuroverkkojen sovellusten yleistyessä on todettu, ettei neuroverkkojen ennustamia ns. *Softmax* -todennäköisyyksiä voida yleisesti pitää luotettavina mittoina ennusteeseen liittyvälle epävarmuudelle (Guo et al., 2017; Hendrycks ja Gimpel, 2016). Koska kerätty aineisto lisäksi harvoin kykenee täysin karakterisoimaan todellista datan generoinutta populaatiojakaumaa (Hendrycks, X. Liu et al., 2020), itsestäänselvänä ei voida myöskään pitää neuroverkkomallien robustia yleistymistä sovitettaessa käytetyn aineiston ulkopuolelle. Hyvän mallin tulisi sekä kyetä ekstrapoloimaan sovitettaessa kohdatun jakauman ulkopuolelle (Hendrycks, X. Liu et al., 2020; Marcus, 2018) että ilmaisemaan, milloin malli todennäköisesti tekee virheen.

Pitkälti edellä kuvattujen tavoitteiden motivoimina paljon tutkimusta on tehty Bayesiläisten menetelmien soveltamiseksi neuroverkkomalleihin. Neuroverkko parametriseoi funktion  $f(\mathbf{x}; \boldsymbol{\theta}); \mathbb{R}^M \rightarrow \mathbb{R}^P$ , millä tyypillisesti edelleen parametriseoidaan jokin ehdollinen todennäköisyysjakauma  $p(y | \mathbf{x}; \boldsymbol{\theta})$  (Goodfellow, Bengio ja Courville, 2016). Bayesiläisissä neuroverkoissa prioriksi asetetaan neuroverkon parametreille  $\boldsymbol{\theta}$ , joita nykyaikaisissa neuroverkoissa voi kuitenkin olla jopa miljardeja. Tämä tekee parametreja koskevasta päätelystä laskennallisesti raskasta ja hidasta (Dusenberry et al., 2020). Parametriavaruudesta voidaan siirtyä funktioavaruuteen Gaussisilla prosesseilla (Rasmussen ja Williams, 2006).

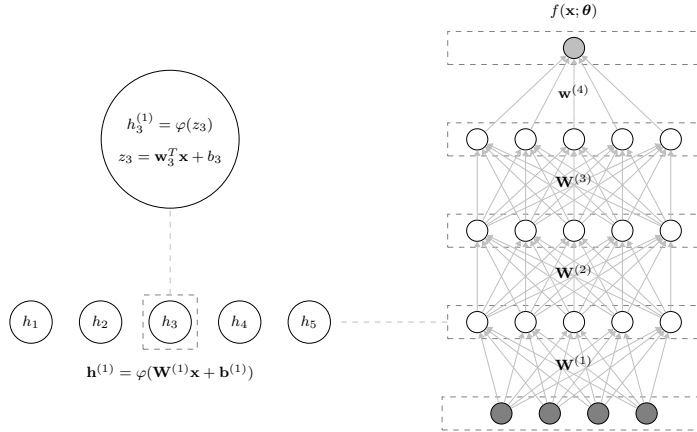
Gaussiset prosessit ovat joustavia parametrittomia Bayesiläisiä malleja (Gelman et al., 2014), joilla priorijakauma voidaan asettaa suoraan funktiolle  $f$ . Gaussinen prosessi kytketään neuroverkkoon siten, että kovarianssifunktiossa käytetään neuroverkolla johdettuja piirteitä ja yhdistetyn mallin kaikki parametrit estimoidaan stokastisella gradienttimenetelmällä tai jollakin tämän muunnoksella. Tutkielmassa esiteltiin kaksi verrattain tuoretta kirjallisuudessa esitettyä Gaussisen prosessin ja neuroverkon yhdistävää menetelmää ja tarkasteltiin, voidaanko näillä parantaa neuroverkkoa käyttävän mallin robustisuutta ja epävarmuuden kvantifiointin luotettavuutta.

## Neuroverkot

Seuraavassa on lyhyesti kuvattu periaate, kuinka neuroverkolla parametriseoitu funktio  $f(\mathbf{x}; \boldsymbol{\theta})$  määräytyy. Esimerkkinä on yksinkertaisuuden vuoksi käytetty täysin kytkettyä eteenpäinsyöttävää neuroverkkoa. Periaate on kuitenkin siirrettävissä esimerkiksi tutkielmassa käytettyyn Transformer -arkkitehtuuria toteuttavaan neuroverkkoon, mikä on kuvattu tutkielman luvussa 1.2.

Neuroverkon ensimmäinen kerros saa syötteenä kovariaattivektorin  $\mathbf{x} = (x_1, \dots, x_M)^T$  ja muodostaa tästä epälineaarilla aktivointifunktiolla  $\varphi$  piirrektorin  $\mathbf{h}^{(1)} = \varphi(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)})$ . Edellä on käytetty merkintää  $\varphi(\mathbf{z}) = (\varphi(z_1), \dots, \varphi(z_M))^T$ , kun  $z_i = \mathbf{w}_i^T \mathbf{x} + b_i$ . Piirrektori  $\mathbf{h}^{(1)}$  toimii edelleen syötteenä seuraavalle kerrokselle. Menettely toistetaan kaikissa neuroverkon kerroksissa eli  $\mathbf{h}^{(l)} = \varphi(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)})$ , missä yläindeksillä  $(l)$  viitataan neuroverkon kerrokseen  $l = 1, \dots, L$ . Merkitsemällä  $\mathbf{h}^{(l)} = f^{(l)}(\mathbf{h}^{(l-1)})$  koko neuroverkon parametriseoitu funktio  $f$  voidaan esittää yhdistettynä funktiona  $f(\mathbf{x}) = f^{(L)}(f^{(L-1)}(\dots(f^{(1)}(\mathbf{x}))))$ .

Kuten edellisessä luvussa todettiin, tyypillisesti neuroverkon määräämä funktio  $f(\mathbf{x}; \boldsymbol{\theta})$  parametriseoi ehdollisen todennäköisyysjakauman  $p(y | \mathbf{x}; \boldsymbol{\theta})$ . Neuroverkon avulla määrätty malli voidaan usein rinnastaa yleistetyksi lineaariseksi malliksi, missä lineaarisessa prediktorissa  $\eta_i = \boldsymbol{\beta} \mathbf{x}_i$  käytetään kovariaattivekto-



Kuva 7.1: Esimerkki täysinkytkestystä eteenpäinsyöttävästä neuroverkosta ja tässä suoritettavasta laskennasta. Neuroverkon kolmessa piilokerroksessa on kussakin viisi laskentayksikköä ja ulostulokerroksessa kolme laskentayksikköä.

rin  $\mathbf{x}_i$  tilalla neuroverkolla johdettuja piirteitä  $\mathbf{h}_i$  ja ulostulokerroksen aktivointifunktio määrätään linkkifunktiota  $g$  vastaten. Neuroverkolla johdetuilla piirteillä tarkoitetaan siis ulostulokerrosta edeltävän kerroksen piirrevektoria, mitä jatkossa merkitään, kuten edellä ilman yläindeksiä.

## Latentin Gaussisen prosessin malli

Gaussinen prosessi  $F$  voidaan määritellä kokoelmalla satunnaismuuttujia  $F = \{F_{\mathbf{x}}\}_{\mathbf{x} \in \mathbb{R}^M}$ , minkä millä tahansa äärellisellä kokoelmalla indeksejä  $\mathbf{x}_i \in \mathbb{R}^M$  satunnaisvektori  $(F_{\mathbf{x}_1}, \dots, F_{\mathbf{x}_N})^T$  noudattaa moniuolotteista normaalijakaumaa (Rasmussen ja Williams 2006; Kallenberg 1997). Edellä  $F_{\mathbf{x}} : \Omega \rightarrow \mathbb{R}$  eli  $F : (\Omega, \mathbb{R}^M) \rightarrow \mathbb{R}$  ja kiinteällä  $\omega \in \Omega$ ,  $F$  on kuvaus  $\mathbb{R}^M \rightarrow \mathbb{R}$ . Toisin sanoen Gaussisen prosessin realisaatiot ovat satunnaisia funktioita (Gelman et al., 2014). Merkitäänkin jatkossa  $F_{\mathbf{x}} = f(\mathbf{x})$ .

Gaussisen prosessin  $\{F_{\mathbf{x}}\}_{\mathbf{x} \in \mathbb{R}^M}$  määrää yksikäsitteisesti keskiarvofunktio  $m : \mathbb{R}^M \rightarrow \mathbb{R}$  ja kovarianssifunktio  $k : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$ , joille (Rasmussen ja Williams, 2006)

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))].$$

Kovarianssifunktio määrää äärelliselle vektorille  $\mathbf{f} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))^T$ ,  $\mathbf{x}_i \in \mathbb{R}^M$  symmetrisen ja positiivisesti semidefiniitin  $N \times N$  kovarianssimatriisin

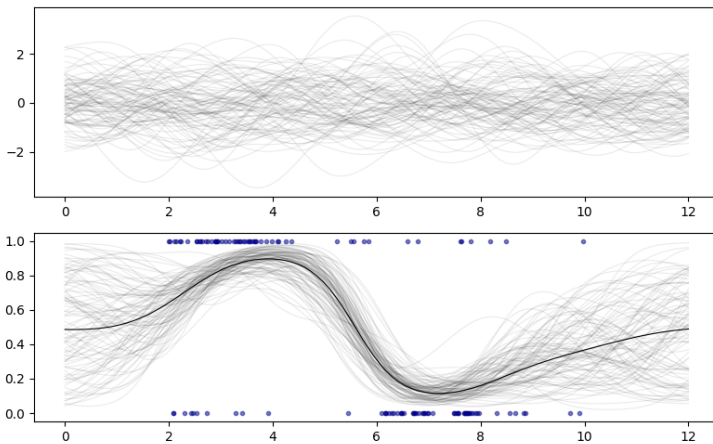
$$\mathbf{K}_{XX} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & k(\mathbf{x}_1, \mathbf{x}_2) & \dots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & k(\mathbf{x}_N, \mathbf{x}_2) & \dots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}.$$

Latentin Gaussisen prosessin mallissa Gaussinen prosessi määrää priorijakauman (GP-priori) latentille funktiolle  $f$ , mikä linkkifunktion  $g$  kautta määrittää uskottavuuden  $p(\mathbf{y} | \mathbf{f}, \mathbf{X})$  (Gelman et al., 2014). Latentin Gaussisen prosessin malleja käytetään, kun uskottavuus ei ole normaalijakauma (Gelman et al., 2014), kuten tutkielmassa dikotomiselle vasteelle, mille  $y_i \sim \text{Bin}(1, \pi_i)$ . Posteriorijakaumaa on ei-konjugaattisen uskottavuuden tapauksessa joko approksimoitava analyttisesti tai simuloitava Monte Carlo Markov Chain -menetelmään perustuvilla menetelmillä (Rasmussen ja Williams, 2006). Päättely perustuu posterioriennustejakaumaan

$$q(f^* | \mathbf{y}) = \int p(f^* | \mathbf{f}, \mathbf{y}) q(\mathbf{f} | \mathbf{y}) d\mathbf{f},$$

missä  $q(\mathbf{f} | \mathbf{y})$  on tässä tarkemmin yksilöimätön approksimaatio posteriorijakaumalle  $p(\mathbf{f} | \mathbf{y})$ .

Menettelyä dikotomisella vasteella on havainnollistettu kuvassa 7.2. Kuvassa ylhäällä on  $n = 100$  otos latentin  $f$  priorista  $GP(0, k(\mathbf{x}, \mathbf{x}'))$ , missä kovarianssifunktiona  $k$  on neliöity eksponenttifunktio<sup>1</sup>. Kuvassa alhaalla on vastaava otos posterioriennustejakaumasta parametrille  $\pi = g^{-1}(f^*)$  sekä mustalla värillä kuvattuna tämän posteriorikeskiarvo.



Kuva 7.2: Otokset ( $n=100$ ) GP-priorista  $f \sim GP(0, k(\mathbf{x}, \mathbf{x}'))$  (ylh.) ja kuvaan lisättyillä havainnoilla parametrille  $\pi = g^{-1}(f^*)$  määrätystä posterioriennustejakaumasta (posteriorikeskiarvo mustalla). Kovarianssifunktiona kuvassa on neliöity eksponenttifunktio.

Gaussista prosessia ei käytännössä kyetä arvioimaan koko määrittelyjoukossa. Sovelluksissa kuitenkin riittää tarkastella indeksijoukon äärellisiä kokoelmia  $\mathbf{x}_i \in \mathcal{D}$ , missä  $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$  eli havaittu aineisto sekä uusia arvoja  $\mathbf{x}_j^*, j = 1, \dots, P$ , joiden ennusteista ollaan kiinnostuneita (Gelman et al., 2014). Kuvan 7.2 kuvaajissa Gaussisen prosessin eli funktion  $f$  arvoja on tarkasteltu

<sup>1</sup> $\exp\left(-\frac{|x-x'|^2}{2}\right)$

sadassa väliltä  $[0, 12]$  tasavälein määrättyssä pisteessä, mutta kuvaajat piirretty jatkuvina.

Eksaktin päättelyn esteeksi Gaussisen prosessin malleissa yleisesti nousee  $N \times N$  kovarianssimatriisin kääntäminen, mikä on vaativuusluokka  $\mathcal{O}(N^3)$ . Ap-proksimaatioita tarvitaan jo, kun havaintojen lukumäärä nousee muutamiin tuhansiin (Hensman, Fusi ja Lawrence, 2013). Seuraavassa luvussa esiteltävissä menetelmissä käytetään approksimaatioita, joilla kovarianssimatriisin kääntämisen vaativuusluokaksi saadaan  $\mathcal{O}(NM^2)$ , missä voidaan valita  $M < N$ .

## Menetelmät

Gaussinen prosessi voidaan yleisesti kuvattuna yhdistää neuroverkkoon käyttämällä Gaussisen prosessin kovarianssifunktiossa  $k(\mathbf{x}, \mathbf{x}')$  syötteinä alkuperäisten piirteiden  $\mathbf{x}_i$  sijaan neuroverkolla johdettuja piirteitä  $\mathbf{h}_i = f^{(L)}(\mathbf{h}_i^{(L-1)})$  (Wilson et al., 2016). Gaussinen prosessi implementoidaan mallin ulostuloskerrokseen, missä kovarianssifunktioksi nyt tulee  $k(\mathbf{h}, \mathbf{h}')$ . Jatkossa tähän kerrokseen viitataan termillä GP-kerros.

Ensimmäisessä tutkielmassa esiteltyssä menetelmässä (J. Liu, Lin et al., 2020; J. Liu, Padhy et al., 2022) latentille funktiolle  $f$  oletetaan priorin  $f \sim N(\mathbf{0}, \mathbf{K})$ . Kovarianssimatriisiin  $\mathbf{K}$  tilalla käytetään approksimaatiota  $\Phi\Phi^T$ , missä

$$\Phi = \begin{bmatrix} \phi_{11}(\mathbf{h}_1) & \cdots & \phi_{1M}(\mathbf{h}_1) \\ \vdots & \ddots & \vdots \\ \phi_{1N}(\mathbf{h}_N) & \cdots & \phi_{NM}(\mathbf{h}_N) \end{bmatrix}$$

ja minkä asteelle  $\text{Rank}(\Phi\Phi^T) \leq \min(\text{Rank}(\Phi), \text{Rank}(\Phi^T)) \leq M$  (Ton et al., 2018). Kovarianssimatriisin kääntämisen vaativuusluokasta edellisen luvun lopussa mainittu seuraa nyt Woodburyn identiteetistä (Woodbury, 1950). Matriisin  $\Phi$  elementtejä

$$\phi_{ij}(\mathbf{h}_i) = \sqrt{2/M} \cos(\mathbf{w}_j^T \mathbf{h}_i + b_j),$$

kutsutaan satunnaisiksi Fourier -piirteiksi (Rahimi ja Recht, 2007). Edellä  $w_j \sim N(0, 1)$  ja  $b_j \sim U(0, 2\pi)$ . Muunnoksen johtamisen osalta tässä viitataan alkuperäiseen julkaisuun Rahimi ja Recht (2007) tai tutkielman lukuun 2.4.1. Nyt siirtyen edellä määritettyä Gaussista prosessia vastaavaan parametriseen esitykseen  $f(\mathbf{h}) = \beta^T \phi(\mathbf{h})$ , missä a priori  $\beta \sim N(\mathbf{0}, \mathbf{I})$ , malli voidaan ajatella kahdesta kerroksesta muodostettuna neuroverkon lohkona, missä ensimmäisen kerroksen kerroinmatriisin alkiot poimitaan otoksina jakaumasta  $N(0, 1)$  aktiivointifunktion ollessa  $\varphi(z) = \sqrt{2/M} \cos(z)$  ja jälkimmäinen kerros käsittää vain yhden yksikön aktiivointifunktion ollessa identiteettifunktio.

Posterijakauman  $p(\beta | \mathbf{y})$  tilalla käytetään Laplace -approksimaatiota  $q(\beta | \mathbf{y})$ . Tällöin  $\beta | \mathbf{y} \sim N(\hat{\beta}, \Sigma_{\hat{\beta}})$ , missä  $\hat{\beta} = \text{argmax}_{\beta} p(\beta | \mathbf{y})$  ja  $\Sigma_{\hat{\beta}} = -\nabla^2 \log p(\hat{\beta} | \mathbf{y})$ . Posteriorimoodi  $\hat{\beta}$  ja neuroverkko-osan parametrit  $\theta$  estimoidaan minimoimalla stokastisella gradienttimenetelmällä normalisoimatonta logaritmistä posteriorijakaumaa

$$-\log p(\beta | \mathbf{y}) \propto -\log p(\mathbf{y} | \beta) + \frac{1}{2} \beta^T \beta.$$

Estimaatti Laplace -approksimaatiota vastaavalle kovarianssimatriisille  $\hat{\Sigma}_\beta$  saadaan tämän käänteismatriisilla, jolle pätee

$$\hat{\Sigma}_\beta^{-1} = \mathbf{I} + \sum_i \hat{\pi}_i (1 - \hat{\pi}_i) \phi(\mathbf{h}_i) \phi(\mathbf{h}_i)^T,$$

missä  $\hat{\pi}_i = g^{-1}(\hat{\beta}^T \phi(\mathbf{h}_i))$ . Kovarianssimatriisin päivitys voidaan siten tehdä osajoukko kerrallaan sovituksen viimeisellä epookilla.

Toisessa tutkielmassa esitellyssä menetelmässä kovarianssimatriisin kääntämisen vaativuusluokkaa pudotetaan indusoiviin muuttujiin perustuvalla approksimaatiolla, joissa perusajatuksena on käsitellä vain osaa latenteista muuttujista eksaktisti ja käyttää muiden kohdalla jotain laskennallisesti kevyempää approksimaatiota (Quinero-Candela ja Rasmussen, 2005). Yhteisenä tekijänä approksimaatioissa on mallin augmentointi *indusoivilla muuttujilla*  $\mathbf{u} = (u_1, \dots, u_M)^T$ , mitkä ovat latentin funktion  $f$  arvoja  $f(\mathbf{z}_1), \dots, f(\mathbf{z}_M)$  *indusoiduissa pisteissä*  $\mathbf{z}_i \in \mathbb{R}^M$ . Indusoiduille muuttujille oletetaan sama GP-priori, kuin muuttujille  $\mathbf{f}$  ja yhteisjakaumalle  $p(\mathbf{f}, \mathbf{u})$  (Hensman, Matthews ja Ghahramani, 2015)

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{u} \end{bmatrix} \sim N \left( \mathbf{0}, \begin{bmatrix} \mathbf{K}_{NN} & \mathbf{K}_{NM} \\ \mathbf{K}_{MN} & \mathbf{K}_{MM} \end{bmatrix} \right),$$

missä lohkomatriisien alaindeksit viittaavat matriisien dimensioihin. Lähes kaikissa approksimaatioissa oletetaan lisäksi, että  $\mathbf{f}$  ja  $f^*$  ovat riippumattomia ehdolla  $\mathbf{u}$  (Quinero-Candela ja Rasmussen, 2005). Tällöin augmentoidussa mallissa  $p(\mathbf{y}, \mathbf{f}, \mathbf{u})$  posterioriennustejakaumalle (Titsias, 2009)

$$\begin{aligned} p(f^* | \mathbf{y}) &= \int p(f^*, \mathbf{f}, \mathbf{u} | \mathbf{y}) d\mathbf{f} d\mathbf{u} \\ &= \int p(f^* | \mathbf{u}) p(\mathbf{f} | \mathbf{u}) p(\mathbf{u} | \mathbf{y}) d\mathbf{f} d\mathbf{u}. \end{aligned}$$

Edellisestä nähdään, kuinka muuttujat  $\mathbf{u}$  *indusoidut* riippuvuuden opetushavaintoja ja uusia havaintoja vastaavien muuttujien  $\mathbf{f}$  ja  $f^*$  välille (Quinero-Candela ja Rasmussen, 2005).

Posteriorijakauman  $p(\mathbf{u} | \mathbf{y})$  tilalla käytetään variationaalista approksimaatiota  $q(\mathbf{u})$ , jolloin indusoivia muuttujia ja pisteitä voidaan käsitellä variationaalisisina parametreina (Titsias, 2009). Jakaumalle  $q$  oletetaan  $q(\mathbf{u}) = N(\mathbf{m}, \mathbf{S})$  (Hensman, Matthews ja Ghahramani, 2015) eli variationaalisisina parametreina on lisäksi odotusarvovektori  $\mathbf{m}$  ja kovarianssimatriisi  $\mathbf{S}$ .

Tutkielmassa sovellettu GP-kerroksen implementaatio jäljitteli Amersfoort et al. (2021) esittelemää menetelmää, missä latentille funktiolle  $f$  oletetaan prior  $f \sim GP(m(\mathbf{h}), k(\mathbf{h}, \mathbf{h}'))$ , missä  $m(\mathbf{h}) = \mu_0$  on estimoitava hyperparametri. Indusoiduille pisteille  $\mathbf{z}_j$  oletetaan, että  $\mathbf{z}_j = \mathbf{h}_j$  eli pisteiden sijainnit estimoidaan neuroverkolla johdettujen piirteiden  $\mathbf{h}$  määräämästä avaruudesta. Variationaaliset parametrit, Gaussisen prosessin hyperparametrit ja neuroverkon parametrit estimoidaan maksimoimalla Hensman, Matthews ja Ghahramani (2015) marginaaliselle uskottavuudelle johtamaa alarajaa

$$\begin{aligned} \log p(\mathbf{y}) &\geq \mathbb{E}_{q(\mathbf{f})} [\log p(\mathbf{y} | \mathbf{f})] - \text{KL}(q(\mathbf{u}) || p(\mathbf{u})) \\ &= \sum_i \mathbb{E}_{q(f_i)} [\log p(y_i | f_i)] - \text{KL}(q(\mathbf{u}) || p(\mathbf{u})), \end{aligned} \tag{1}$$

missä merkintä  $KL(q(\mathbf{u})||p(\mathbf{u}))$  tarkoittaa jakaumien  $q$  ja  $p$  *Kullback-Leibler*-poikkeavuutta. Jakaumalle  $q(\mathbf{f})$  odotusarvossa (1) saadaan (Hensman, Matthews ja Ghahramani, 2015)

$$q(\mathbf{f}) = N(\mathbf{A}\mathbf{m}, \mathbf{K}_{nn} + \mathbf{A}(\mathbf{S} - \mathbf{K}_{mm})\mathbf{A}^T),$$

missä  $\mathbf{A} = \mathbf{K}_{nm}\mathbf{K}_{mm}^{-1}$ . Tarkempien yksityiskohtien osalta tässä joudutaan viittamaan tutkielman lukuun 2.4.2 sekä lähdejulkaisuihin.

## Tuloksista

Tutkielmassa arvioitiin, voidaanko edellisessä luvussa esitellyillä menetelmillä parantaa Transformer -arkkitehtuuriin perustuvan neuroverkon robustisuutta ja epävarmuuden kuvautumisen luotettavuutta tekstin sentimentin ennustamiseen liittyvässä tehtävässä. Mainittuja seikkoja tarkasteltiin asetelmissa, missä mallit oli sovitettu kuluttaja-arvioita käsittävällä aineistolla ja niillä ennustettiin samasta aineistosta erotettua testiaineistoa sekä testiaineistoa, mikä oli muodostettu eri aineistosta, mutta missä tekstin sentimentti edelleen oli mielekäs vasteuuttuja. Tässä mielenkiinnon kohteena oli, kuinka hyvin mallit yleistyvät, kun ennustettaessa kohdattujen näytteiden jakauma poikkeaa sovitettaessa kohdattusta jakaumasta ja edelleen, voidaanko ennusteisiin kuvautuvasta epävarmuudesta päätellä, milloin malli todennäköisesti tekee luokitteluvirheen tai kohtaa poikkeavasta jakaumasta olevan näytteen. Lisäksi arvioitiin mallien epävarmuuden ilmaisua ennustettaessa näytteitä kieliasultaan neutraalista säädösteksteistä muodostetusta aineistosta, missä tekstin sentimentti ei enää ole mielekäs vasteuuttuja. Tässä tarkoituksena oli arvioida mallien kykyä tunnistaa poikkeavia havaintoja, mikä joissakin sovelluksissa voi olla tavoitteena jo itsessään.

Robustin yleistymisen osalta merkittäviä eroja GP-kerroksen implementoitujen mallien ja ilman GP-kerrosta määritetyn, perustasoa kuvanneen mallin välillä ei voitu havaita. Luokitteluvirheitä voitiin tunnistaa kaikilla malleilla sattumaa paremmin. Tässä yleisesti arvioiden parhaiten pärjäsi perustasoa kuvannut malli. Merkittävimmät erot GP-kerroksen implementoitujen mallien eduksi voitiin havaita tunnistettaessa poikkeavia havaintoja, erityisesti kun epävarmuutta kuvaavana mittana käytettiin posterioriennustejakauman varianssille saatua estimaattia.

## Yhteenveto

Robustisuus ja epävarmuuden luotettava kvantifiointi ovat neuroverkkojen käytännön sovelluksissa varsin relevantteja ongelmia ja seikkoja, mitkä ehdottomasti tulisi huomioida myös malleja arvioitaessa ja käytännön sovelluksiin valittaessa. Mikään yksittäinen malli tai menetelmä ei useinkaan ole yksikäsitteisesti suorituskykyisin, mikä voitiin havaita myös tutkielmassa tehdyissä tarkasteluissa. GP-kerroksen implementaatiolla voitiin saavuttaa parempi suorituskyky tunnistettaessa poikkeavia havaintoja, mutta luokitteluvirheitä yleisesti parhaiten ennusti perustasoa kuvannut malli. Mallinvalinnassa voidaankin usein joutua arvottamaan eri suorituskykyjä sen perusteella, minkälaiseen ympäristöön mallia ollaan soveltamassa ja mitkä tekijät tässä ovat merkityksellisimmät. Koska täydellistä mallia ei koskaan tulla saavuttamaan, hyödyllistä on myös eri suorituskykyjen raportointi mallin loppukäyttäjälle, jotta tämä vähintäänkin tulee tietoiseksi ja voi arvioida, miltä osin ja missä määrin mallin ennusteisiin voidaan luottaa. "Kaikki mallit ovat väärinä, mutta jotkut ovat hyödyllisiä"(Box, 1979).



## Viitteet

- Amersfoort, J. van et al. (2021). "On feature collapse and deep kernel learning for single forward pass uncertainty". *arXiv preprint arXiv:2102.11409*.
- Box, G. E. (1979). "Robustness in the strategy of scientific model building". Teoksessa: *Robustness in statistics*. Elsevier, s. 201–236.
- Dusenberry, M. et al. (2020). "Efficient and scalable bayesian neural nets with rank-1 factors". Teoksessa: *International Conference on Machine Learning*. PMLR, s. 2782–2792.
- Gelman, A. et al. (2014). *Bayesian Data Analysis*. Chapman ja Hall/CRC.
- Goodfellow, I., Bengio, Y. ja Courville, A. (2016). *Deep Learning*. MIT Press.
- Guo, C. et al. (2017). "On calibration of modern neural networks". Teoksessa: *International Conference on Machine Learning*. PMLR, s. 1321–1330.
- Hendrycks, D. ja Gimpel, K. (2016). "A baseline for detecting misclassified and Out-of-Distribution examples in neural networks". *arXiv preprint arXiv:1610.02136*.
- Hendrycks, D., Liu, X. et al. (2020). "Pretrained Transformers improve Out-of-Distribution robustness". *arXiv preprint arXiv:2004.06100*.
- Hensman, J., Fusi, N. ja Lawrence, N. D. (2013). "Gaussian processes for big data". *arXiv preprint arXiv:1309.6835*.
- Hensman, J., Matthews, A. ja Ghahramani, Z. (2015). "Scalable variational Gaussian process classification". *Artificial Intelligence and Statistics*, s. 351–360.
- Kallenberg, O. (1997). *Foundations of Modern Probability*. Vol. 2. Springer.
- Liu, J., Lin, Z. et al. (2020). "Simple and principled uncertainty estimation with deterministic deep learning via distance awareness". *Advances in Neural Information Processing Systems* 33, s. 7498–7512.
- Liu, J., Padhy, S. et al. (2022). "A simple approach to improve single-model deep uncertainty via distance-awareness". *Journal of Machine Learning Research* 23, s. 1–63.
- Marcus, G. (2018). "Deep learning: A critical appraisal". *arXiv preprint arXiv:1801.00631*.
- Quinonero-Candela, J. ja Rasmussen, C. E. (2005). "A unifying view of sparse approximate Gaussian process regression". *The Journal of Machine Learning Research* 6, s. 1939–1959.
- Rahimi, A. ja Recht, B. (2007). "Random features for large-scale kernel machines". *Advances in Neural Information Processing Systems* 20.
- Rasmussen, C. E. ja Williams, C. K. (2006). *Gaussian Processes for Machine Learning*. Vol. 2. MIT Press Cambridge, MA.
- Titsias, M. (2009). "Variational learning of inducing variables in sparse Gaussian processes". Teoksessa: *Artificial Intelligence and Statistics*. PMLR, s. 567–574.
- Ton, J.-F. et al. (2018). "Spatial mapping with Gaussian processes and nonstationary Fourier features". *Spatial Statistics* 28, s. 59–78.
- Wilson, A. G. et al. (2016). "Deep kernel learning". Teoksessa: *Artificial intelligence and statistics*. PMLR, s. 370–378.
- Woodbury, M. A. (1950). *Inverting Modified Matrices*. Department of Statistics, Princeton University.

---

# Suomen Tilastoseuran hallitus vuonna 2022

BOARD MEMBERS OF THE FINNISH STATISTICAL SOCIETY IN 2022

## Hallitus 30.5.2022 asti

Puheenjohtaja Chair	Jyrki Möttönen	Filosofian tohtori PhD
Varapuheenjohtaja Vice Chair	Ari Jaakola	Filosofian maisteri MSc
Sihteeri Secretary	Elisa Falck	Valtiotieteiden kandidaatti BSocSci
Rahastonhoitaja Treasurer	Tapani Linnaluoto	Luonnontieteiden ylioppilas student (Nat sci)
Klubimestari, jäsen Member	Tommi Härkänen	Filosofian tohtori PhD
Jäsen Member	Pekka Pere	Doctor of Philosophy DPhil
Jäsen Member	Reija Helenius	Yhteiskuntatieteiden maisteri MSocSc
Jäsen Member	Paula Bergman	Filosofian maisteri MSc
Jäsen Member	Mika Gissler	Filosofian tohtori PhD
Jäsen Member	Arvi Tolvanen	Luonnontieteiden kandidaatti BSc
Varajäsen Deputy member	Tommi Mäklin	Filosofian maisteri MSc

---

## Hallitus 30.5.2022 jälkeen

Puheenjohtaja Chair	Pekka Pere	Doctor of Philosophy DPhil
Varapuheenjohtaja Vice Chair	Ari Jaakola	Filosofian maisteri MSc
Sihteeri Secretary	Elisa Falck	Valtiotieteiden kandidaatti BSocSci
Rahastonhoitaja Treasurer	Tapani Linnaluoto	Luonnontieteiden ylioppilas
Klubimestari, jäsen Member	Tommi Härkänen	Filosofian tohtori PhD
Jäsen Member	Reija Helenius	Yhteiskuntatieteiden maisteri MSocSc
Jäsen Member	Paula Bergman	Filosofian maisteri MSc
Jäsen Member	Mika Gissler	Filosofian tohtori PhD
Jäsen Member	Arvi Tolvanen	Luonnontieteiden kandidaatti BSc
Jäsen Member	Jyrki Möttönen	Filosofian tohtori PhD
Varajäsen Deputy member	Tommi Mäklin	Filosofian tohtori PhD
Varajäsen Deputy member	Johanna Laiho-Kauranne	Filosofian tohtori PhD

---

# Suomen Tilastoseuran hallitus vuonna 2023

BOARD MEMBERS OF THE FINNISH STATISTICAL SOCIETY IN 2023

## Hallitus 28.2.2023 jälkeen

Puheenjohtaja Chair	Pekka Pere	Doctor of Philosophy DPhil
Varapuheenjohtaja Vice Chair	Ari Jaakola	Filosofian maisteri MSc
Sihteeri Secretary	Roope Rihtamo	Valtiotieteiden maisteri MSocSci
Rahastonhoitaja Treasurer	Tuuli Kauppala	Filosofian maisteri MSc
Jäsen Member	Tommi Härkänen	Filosofian tohtori PhD
Jäsen Member	Reija Helenius	Yhteiskuntatieteiden maisteri MSocSc
Jäsen Member	Mika Gissler	Filosofian tohtori PhD
Jäsen Member	Arvi Tolvanen	Luonnontieteiden kandidaatti BSc
Jäsen Member	Jyrki Möttönen	Filosofian tohtori PhD
Varajäsen Deputy member	Elisa Falck	Valtiotieteiden kandidaatti BSocSci
Varajäsen Deputy member	Tapani Linnaluoto	Luonnontieteiden ylioppilas student (Nat sci)
Varajäsen Deputy member	Tommi Mäklin	Filosofian tohtori PhD
Klubimestari	Henrik Törmi	Filosofian kandidaatti BSc

---

# Gunnar Modeen -minnesmedaljen

JUKKA HOFFRÉN  
TILASTOKESKUS

Statistiska Samfundet i Finland r.f. har i samband med de nordiska statistikdagarna traditionsenligt delat ut Gunnar Modeen -minnesmedaljen till särskilt meriterade statistiker. Praxisen har varit att dela ut medaljen till en representant för det land där statistikdagarna hålls.

Gunnar Modeen -minnesmedaljen beviljas för en betydande livsgärning inom statistikbranschen. Meningen är att den person som belönas är en framstående senior expert inom statistikbranschen, som uttryckligen utmärkt sig i det praktiska statistik-arbetet och som uppskattas av sina kolleger.

Styrelsen för Statistiska Samfundet väljer den person som får medaljen och medaljen överläts i samband med ett nordiskt statistikermöte. Enligt fondens stadga överläts medaljen till en betydande nordisk statistiker från det land som respektive år arrangerar mötet. Den första medaljen överläts vid det nordiska statistikermöte som hölls i Finland år 1989.

## Bakgrunden till och kriterier för GM-minnesmedaljen

Efter Gunnar Modeens bortgång år 1988 grundades en medaljfond till hans minne. Medaljen utarbetades på basis av den medaljong som Gunnar Modeens familj gett konstnären Matti Haupt i uppdrag att utforma till Modeens 70-årsdag år 1965. Mot-tagaren av medaljen väljs av styrelsen för Statistiska Samfundet i Finland och medaljen överläts i samband med ett nordiskt statistikermöte. Enligt fondens stadga överläts medaljen till en betydande nordisk statistiker från det land som respektive år arrangerar mötet. Den första medaljen överläts vid Nordiska Statistikermötet i Finland år 1989. Priset utdelas vart tredje år till en meriterad statistiker från det land där Nordiska Statistikermötet anordnas.

Allmänna kriterier för Gunnar Modeen -minnesmedaljen:

- priset beviljas för en betydande livsgärning inom statistikbranschen.

Den person som tilldelas medaljen:

- 
- är en expert inom statistikbranschen, som uttryckligen utmärkt sig i det praktiska statistikarbetet
  - är en nordisk, framstående senior expert som uppskattas av sina kolleger,
  - har akademisk examen (magister, licentiat eller doktor) och
  - är villig att ta emot GM-medaljen

## Mottagare av GM-minnesmedaljen

Den första medaljen tilldelades Mauno Koivisto, Finlands dåvarande president, som en särskild hedersbetygelse. År 1989 var han beskyddare av Nordiska Statistiker-mötet i Finland som firade 100-årsjubileum för nordisk statistik. Ytterligare en medalj delades ut på mötet och mottagare var professor Eino H. Laurila. Övriga mottagare av medaljen:

År 1992 tilldelades medaljen inte.

År 1995 direktör Poul Jensen, Danmarks Statistik.

År 1998 professor Sven Nordbotten, Universitetet i Bergen.

År 2001 professor Emeritus Gunnar Kulldorf, Umeå universitet.

År 2004 direktör Asta Manninen, Helsingfors stads faktacentral.

År 2007 generaldirektör Hallgrímur Snorrason, Hagstofa, Island.

År 2010 direktör Lars Thygesen, Danmarks Statistik.

År 2013 Liv Hobbeldstad Simpson, pensionerad från Statistisk sentralbyrå (SSB) som Head of National accounts och past chair of IARIW

År 2016 Eva Elvers, PhD, pensionerad från Design and Plan & Build and Test som Process owner

År 2019 professor Risto Lehtonen, Helsingfors universitet

---

# Scandinavian Journal of Statistics

Recognised as a leading journal in its field, the Scandinavian Journal of Statistics is an international publication devoted to reporting significant and innovative original contributions to statistical methodology — both theory and applications. The journal specializes in statistical modelling, showing particular appreciation of the underlying substantive research problems. Scandinavian Journal of Statistics is published on behalf of the Danish Society for Theoretical Statistics, the Finnish Statistical Society, the Norwegian Statistical Society, and the Swedish Statistical Society. The journal is currently edited by professors Peter Dalgaard and Niels Richard Hansen. The national editor for Finland is Jukka Corander (University of Helsinki, Finland), and the other national editors are Jacob von Bornemann Hjelmberg (University of Southern Denmark, Denmark), Geir Olve Storvik (University of Oslo, Norway), and Jimmy Olsson (KTH Royal Institute of Technology, Sweden). The chairman of the board is Thomas Scheike (University of Copenhagen, Denmark) and the board members are Juha Karvanen (University of Jyväskylä, Finland), Hans Karlsen (University of Bergen, Norway), and Sara Sjöstedt de Luna (Umeå University, Sweden).

Scandinavian Journal of Statistics is published quarterly in March, June, September and December by Wiley-Blackwell Publishers, 108, Cowley Road, Oxford OX4, 1JF, UK or 238 Main Street, Cambridge, MA 02142, USA.

Members of the Finnish Statistical Society are entitled to discount prices when ordering the Scandinavian Journal of Statistics. For further information, please see the webpage at <http://www.wiley.com/bw/subs.asp?ref=0303-6898&site=1>

**ISI Journal Citation Reports® Ranking:** 2022: 97/125 (Statistics & Probability).

**Impact Factor:** 1.040 (2021).

**Online ISSN:** 1467-9469.

---

# Myönnetyt palkinnot

## Leo Törnqvist –palkinnot

- 1978 Rene Tigerstedt, Helsingin yliopisto. En modell för valbeteende i trafiken.
- 1979 Pirkko Kirjavainen, Turun kauppakorkeakoulu. Mallin rakentaminen ja ennusteen laatiminen Suomen sähkön kulutukselle kahta aikasarja-analyysimenetelmää käyttäen.
- 1980 Esa Läärä, Helsingin yliopisto. Ikä-, aika- ja kohorttitekijöiden vaikutukset Suomen miesten keuhkosityöpäsairastavuudessa vuosina 1953–76.
- 1981 Arvi Suvanto, Tampereen yliopisto. Kausivaihtelu aikasarjamalleissa.
- 1982 Maija Salo, Helsingin yliopisto. Yritys prioriteeton käytöstä alkoholi-juomien kulutusta selittävän kysyntämallin tukena. Jamel Boucelham, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1983 Vesa Vihriälä, Helsingin yliopisto. Aikasarjojen välisen riippuvuuden mittaaminen ja testaus: sovellus suomalaisiin rahatalouden sarjoihin. Pirkko Welin, Tampereen yliopisto: Tunnustuspalkinto.
- 1984 Jari Palsio, Turun kauppakorkeakoulu. Skenaarioiden rakentaminen ristivaikutusanalyysimallia käyttäen.
- 1985 Kenneth Nordström, Helsingin yliopisto. Gauss-Markov-mallien erikoisongelmista.
- 1986 Tapio Nummi, Tampereen yliopisto. APL-pohjainen ohjelmisto GMANOVA-mallille.
- 1987 Ari Veijanen, Helsingin yliopisto. Pickardin kentän soveltamisesta kuva-analyysissä. Kari Nissinen, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1988 Jaason Haapakoski, Helsingin yliopisto. Binomijakautuneiden muuttujien muutosprosessin malli.
- 1989 Pasi Korhonen, Helsingin yliopisto. Kemometrian tilastollisista menetelmistä.
- 1990 Päivi Partanen, Jyväskylän yliopisto. Suljetun populaation koon estimointi merkintä-takaisinpyynti-menetelmällä: log-lineaarinen lähestymistapa. Markku Nurhonen, Tampereen yliopisto: Tunnustuspalkinto.



- 
- 1991 Elina Järvinen, Helsingin yliopisto. Rajoitettujen, stokastisten ja konveksien estimaattoreiden käytöstä polynomisen viipymämallin parametrien estimoinnissa simulointikokeiden valossa.
- 1992 Jouni Kuha, Helsingin yliopisto. Binääristen regressiomallien selittäjien mittausvirheet ja parametriestimaattien mittausvirhekorjaukset. Juha Heikkinen, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1993 Palkintoa ei jaettu (yhtään ehdotusta ei saatu).
- 1994 Ilkka Taskinen, Jyväskylän yliopisto. Äärelliset Markovin ketjut ja anelointi.
- 1995 Mika Rautakorpi, Teknillinen korkeakoulu. Application of Markov chain techniques in certification of software. Tuija Jäppilä, Jyväskylän yliopisto: Tunnustuspalkinto.
- 1996 Veli-Matti Suppola, Jyväskylän yliopisto. Robustit menetelmät. Jakaumien vinouden vaikutuksesta korrelaatiomatriisiin estimointiin.
- 1997 Albert Höglund, Teknillinen korkeakoulu. An Anomaly Detection System for Computer Networks.
- 1998 Samuli Visuri, Oulun yliopisto. Robustista kovarianssimatriisiin estimoinnista ja sen sovelluksista signaalinkäsittelyssä.
- 1999 Jani Raitanen, Tampereen yliopisto. Jalkapallo-ottelun lopputuloksen tilastollinen mallintaminen.
- 2000 Reijo Sund, Helsingin yliopisto. Tilastollisia menetelmiä dynaamisten potilaspopulaatioiden mallintamiseen. Tapahtumahistoria-analyysia hoitoilmoitusrekisterin skitsofreenikoille.
- 2001 Samu Mäntyniemi, Oulun yliopisto. A Hierarchical Bayes Model for Assessing Salmon (Salmo salar L.) Parr and Smolt Populations.
- 2002 Ilmari Juutilainen, Oulun yliopisto. Teräslevyjen lujuuden ennustaminen regressio- ja neuroverkkomalleilla.
- 2003 Leena Kalliovirta, Helsingin yliopisto. Mar-malli.
- 2004 Mikko Myrskylä, Jyväskylän yliopisto. Estimation of Class Frequencies with Micro Level Auxiliary Information.
- 2005 Antti Liski, Tampereen yliopisto. Lonkkamurtumapotilaiden hoitokustannusten vertailu vastaavuuspistemäärään perustuvalla menetelmällä.
- 2006 Karri Seppä, Oulun yliopisto. Suomalaisen paksusuolisyöpäpotilaiden ennusteen analyysi suhteellisen elossapysymisen ja syykohtaisen kuolleisuuden malleilla käyttämällä suurimman uskottavuuden ja Bayesin menetelmiä.
- 2006 Jukka Siren, Helsingin yliopisto. Populaatioiden geneettisen rakenteen spatiaalinen mallintaminen.
- 2007 Outi Ahti-Miettinen, Helsingin yliopisto. Kaksivaiheisen potenssiikiintiöinnin käyttö otoksen tehostamisessa - Esimerkkinä otoksen suunnittelu työvoimakustannusindeksin tietojen keruulle.

- 
- 2008 Paul Catani, Svenska handelshögskolan. Enhetsrottest och initialvärdet Tillämpning på arbetslösheten i Finland
- 2009 Elina Ahola, Jyväskylän yliopisto. Eksponenttisen perheen tila-  
avaruusmallien sovellus alkoholikuolleisuusaineistoon Matias Leppisaari,  
Aalto yliopiston teknillinen korkeakoulu: Tunnustuspalkinto.
- 2010 Sanna Peltomäki, Tampereen yliopisto. Estimation of Below Threshold  
Intra-EU Trade.
- 2011–2012 Tytti Pasanen, Tampereen yliopisto. Two-Level Structural Equation  
Modeling with Non-Normal Observed Variables for Assessing Poverty  
in Laos.
- 2013–2014 Joni Virta, Turun yliopisto. Some tools for linear dimension reduc-  
tion.
- 2015-2016 Niko Lietzén, Aalto-yliopisto. New Approach to Complex Valued  
ICA: From FOBI to AMUSE
- 2015-2016 Santtu Tikka, Jyväskylän yliopisto. Kausaalivaikutusten identifiointi  
algoritmisesti.
- 2017-2018 Savi Virolainen, Helsingin yliopisto. GMAR- ja StMAR-mallin  
yleistäminen G-StMAR-malliksi.
- 2021-2022 Topi Halme, Quickest Detection under False Discovery Rate and  
Communication Constraints.
- 2023 Antti Yläjärvi, Transformer -neuroverkon robustisuuden parantaminen  
Gaussisen prosessin ja neuroverkon yhdistävillä menetelmillä.

## Väitöskirjapalkinnot

- 2009-2012 Jukka Sirén, Helsingin yliopisto. Statistical models for inferring the  
structure and history of populations from genetic data.
- 2013-2016 Johan Pensar, Åbo Akademi. Structure Learning of Context-Specific  
Graphical Models.
- 2017-2020 Anna-Kaisa Ylätaló, Jyväskylän yliopisto. Statistical inference for  
eye movement sequences using spatial and spatiotemporal point processes.
- 2017-2020 Santtu Tikka, Jyväskylän yliopisto. Improving identification algo-  
rithms in causal inference.

---

# Suomen Tilastoseuran julkaisuja

PUBLIKATIONER UTGIVNA AV STATISTISKA SAMMANFUNDET

PUBLICATIONS ISSUED BY THE FINNISH STATISTICAL SOCIETY

1. *Monikielinen väestötieteen sanakirja, suomenkielinen laitos, Helsinki 1962. Multilingual Demographic Dictionary, Finnish section, Helsinki 1962.*
2. *Suomen Tilastoseura – Statistiska Sammanfundet i Finland 1920-1970, Porvoo - Borgå 1970.*
3. *Pohjoismainen tilastosanasto, toinen tarkistettu laitos. Nordisk statistik nomenklatur, andra reviderade upplagan. Nordic statistical nomenclature, 2nd revised edition. Jyväskylä 1975*
4. *Aikasarja-analyysin menetelmiä, Helsinki 1977.*
5. *Pekka Tavaila: Leo Törnqvist Posti- ja lennätinhallituksen liiketaloudellisen tutkimuslaitoksen esimiehenä 1949–1977, Helsinki 1982.*
6. *Vesa Kuusela ja Leif Nordberg (toim.): Otanta teoriassa ja käytännössä. Helsinki 1986.*
7. *Suomen Tilastoseura 70 vuotta. Statistiska Sammanfundet i Finland 70 år. The Finnish Statistical Society 70 years. Helsinki 1991.*
8.
  - *Juha Alho, Elja Arjas, Esa Läärä ja Pekka Pere: Tilastotieteen sanasto. Suomen Tilastoseuran julkaisuja no. 8. Suomen Tilastoseura. 1. laitos, Helsinki 2021.*
  - *Juha Alho, Elja Arjas, Esa Läärä ja Pekka Pere: Tilastotieteen sanasto. Suomen Tilastoseuran julkaisuja no. 8. Suomen Tilastoseura. 2. laitos, Helsinki 2023.*
  - *Juha Alho, Elja Arjas, Juha Karvanen, Lasse Leskelä, Esa Läärä ja Pekka Pere: Tilastotieteen sanasto. Suomen Tilastoseuran julkaisuja no. 8. Suomen Tilastoseura. Verkossa <https://sanasto.tilastoseura.fi/>.*

---

# Tilastotieteellisiä tutkimuksia

STATISTISKA UNDERSÖKNINGAR

STATISTICAL RESEARCH REPORTS

**ISSN 0356-3499**

1. Pentti Manninen: Puolueiden kannatusosuuksien estimoinnin tarkkuus Demingin vyöhykepoiminnassa. (The Accuracy of Party Support Estimation in Deming Zone Selection. In Finnish with English Summary.) Helsinki 1976.
2. Timo Hakulinen: On Competing Risks of Death. Helsinki 1977.
3. Lars-Erik Öller: Time Series Analysis of Finnish Foreign Trade. Helsinki 1978.
4. Pekka Laippala: The Empirical Bayes Two-Action Rules with Floating Optimal Sample Size and Exponential Conditional Distributions. Helsinki 1980.
5. Markku Nurminen: Some Developments in Quantitative Methods of Epidemiology. Helsinki 1982.
6. Pentti Saikkonen: Comparing Asymptotic Properties of Some Tests Used in the Specification of Time Series Models. Helsinki 1985.
7. Lauri Tarkkonen: On Reliability of Composite Scales. Helsinki 1987.
8. Juni Palmgren: Models for Categorical Data with Errors of Observation. Helsinki 1987.
9. Ari Veijanen: On Estimation of Parameters of Partially Observed Random Fields and Mixing Processes. Helsinki 1989.
10. Ritva Luukkonen: On Linearity Testing and Model Estimation in Non-Linear Time Series Analysis. Helsinki 1990.
11. Hely Salomaa: Factor Analysis of Dichotomous Data. Helsinki 1990.
12. Kenneth Nordström: Contributions to the Comparison of Linear Models and to the Löwner-Ordering Antitonicity of Generalized Inverses. Helsinki 1990.
13. Seppo Laaksonen: Handling Household Survey Nonresponse Data. Helsinki 1992.
14. Mervi Eerola: On Predictive Causality in the Statistical Analysis of a Series of Events. Helsinki 1993.
15. Mikael Linden: Studies in Integrated and Co-Integrated Economic Time Series. Helsinki 1995.
16. Tadeusz Dyba: Precision of Cancer Incidence Predictions Based on Poisson Distributed Observations. Helsinki 2000.
17. Kimmo Vehkalahti: Reliability of Measurement Scales. Helsinki 2000.

- 
18. Sirpa Heinävaara: Modelling survival of patients with multiple cancers. Helsinki 2003.

---

# Suomen Tilastoseuran vuosikirja

ÅRSBOK FÖR STATISTISKA SAMMANFUNDET I FINLAND

THE YEARBOOK OF THE FINNISH STATISTICAL SOCIETY

**ISBN 0355-5941**

1975 Helsinki 1976	1996 Helsinki 1997
1976 Helsinki 1977	1997 Helsinki 1998
1977 Helsinki 1978	1998 Helsinki 1999
1978 Helsinki 1979	1999–2000 Helsinki 2000
1979 Helsinki 1980	2001 Helsinki 2002
1980 Helsinki 1981	2002 Helsinki 2003
1981 Helsinki 1982	2003 Helsinki 2004
1982 Helsinki 1983	2004 Helsinki 2005
1983 Helsinki 1984	2005 Helsinki 2006
1984 Helsinki 1985	2006 Helsinki 2007
1985 Helsinki 1986	2007 Helsinki 2008
1986 Helsinki 1987	2008 Helsinki 2009
1987 Helsinki 1988	2009 Helsinki 2010
1988–1989 Helsinki 1990	2010 Helsinki 2011
1990 Helsinki 1991	2011–2012 Helsinki 2012
1991 Helsinki 1992	2013–2014 Helsinki 2014
1992 Helsinki 1993	2015–2016 Helsinki 2017
1993 Helsinki 1994	2017–2018 Helsinki 2018
1994 Helsinki 1995	2019–2022 Helsinki 2023
1995 Helsinki 1996	2023 Helsinki 2024

Tilastoseuran julkaisuja voi tiedustella sihteeriltä sähköpostitse osoitteesta [suomentilastoseura@gmail.com](mailto:suomentilastoseura@gmail.com). Joidenkin julkaisujen painokset ovat tosin jo loppuneet.

---

# Muita julkaisuja

ANDRA PUBLIKATIONER

OTHER PUBLICATIONS

*Suomen Tilastoseura 1920–1945, Helsinki 1946*

*Statistiska Sammanfundet i Finland 1920–1945, Helsingfors 1946.*

*Pohjoismainen tilastosanasto – Nordisk statistisk nomenklatur, Kööpenhamina 1954.*

*13:e Nordiska statistikermötet i Helsingfors 14–16 juni 1973, Jyväskylä 1974*

*The 13th Joint Meeting of the Nordic Statistical Societies in Helsinki June 1973, Jyväskylä 1974.*

*Det 18:e nordiska statistikmötet i Esbo, Hundraårsjubileum, Helsingfors 1990.*

*The Joint Conference of the Nordic Statisticians in Espoo, Finland 1989, Helsinki 1990.*



ISSN 0355 – 5941 (painettu)

ISSN 2953 – 9307 (verkossa)