

Tilastotieteen sanasto

Juha Alho, Elja Arjas, Esa Läärä ja Pekka Pere



Tilastotieteen sanasto

Juha Alho, Elja Arjas, Esa Läärä ja Pekka Pere

Suomen Tilastoseuran julkaisuja 8

2. laitos
Suomen Tilastoseura ry
Helsinki 2023

© Suomen Tilastoseura ry

2. laitos

ISSN 2954-2286 (verkkojulkaisu)

ISBN 978-952-9795-14-7 (pdf)

Kansien sanapilvet: Juha Alho 2020

Helsinki 2023

Sisällys

Alkusanat	v
1 Sanaston tarve, käyttötarkoitus ja periaatteita	1
1.1 Maailma muuttuu	1
1.2 Pohdintoja hyvistä termeistä	2
1.3 Käytännön periaatteita sanastotyössä	3
1.4 Monesta moneen -kuvaus	5
2 Esimerkkejä pulmallisista termeistä	7
2.1 <i>Significance</i>	7
2.2 <i>Effect size</i> ja <i>effect magnitude</i>	9
2.3 <i>Accuracy</i> ja <i>precision</i>	10
2.4 <i>Bayes</i>	11
2.5 <i>Gauss</i> ja <i>Markov</i>	12
2.6 <i>Kurtosis</i> ja <i>peakedness</i>	13
2.7 <i>Frequency</i>	15
2.8 <i>Rate</i>	15
2.9 <i>Odds ratio</i> ja <i>odds</i>	17
2.10 <i>Score</i>	20
2.11 <i>Survey</i>	21
3 Tilastotieteen termien suomentamisen historiaa	23
3.1 Valtioon ja väestöön liittyvät termit	23
3.2 Luonnontieteellinen havainnointi	26
3.3 Tilastotieteen teoria, oppikirjat ja oppaat	30
3.4 Tilastotiede vakiintuu itsenäiseksi oppiaineeksi	32
3.5 Väestötiede, psykometriikka, yhteiskuntatieteet ja laadun- valvonta	34
3.6 Tilastotieteen sanastot	35
4 Lähivertailuja viron tilastosanaston kanssa	37
4.1 Suomi, viro ja tilastotieteen termit	37
4.2 Perustermit: <i>statistika</i> ja <i>andmed</i>	38
4.3 Sanastojen kontrastiivinen nelikenttä	39
4.4 Onnistuneita omakielisiä termejä puolin ja toisin	40
4.5 Petollisia ystäviä	41
5 Ohjeita käyttäjälle – Instructions for the user	43
5.1 Englanti-suomi-englanti-sanasto	43
5.2 Finnish-English-Finnish dictionary	43

6 Englanti–suomi	45
7 Suomi–englanti	118
Lähteet	182

Alkusanat 1. laitokseen

Suomen Tilastoseuran (1954, 1975) sanastot ovat olleet arvovaltaisia ja käytettyjä. Syksyllä 2017 Suomen Tilastoseuran hallitus päätti tuottaa ajantasaisen englanninkielisten termien sanaston juhlistamaan 1920 perustetun seuran satavuotista taivalta. Työryhmän alunperin runsaan kahden vuoden työ koostui puolesta tusinasta tapaamisesta Helsingissä, kerran Taivalkoskella ja yli 50 etäkokouksesta. Ne kestivät kolmesta yhdeksään tuntiin.

Työn lähtökohtana oli Kansainvälisen tilastoinstituutin (*International Statistical Institute*, ISI) monikielinen sanakirja sekä useat muut koti- ja ulkomaiset tilastotieteen sanastot. Sanalistoja täydennettiin kunkin kirjahyllystä löytyvistä lähteistä ja, mikä tärkeintä, netistä. Netistä löytyneet lukumäärätiedot eri termien käytöstä olivat usein ratkaisevia, kun pohdimme, mikä englannin termi ylipäättään otetaan mukaan.

Vuoden 2019 loppupuoliskolla sekä alkutalvella 2020 pyysimme runsaalta tusinalta tilastotieteen tutkijalta ja soveltajalta kommentteja koostamamme sanalistasta. Ilahduttavasti lähes kaikki vastasivat. Se oli arvokasta. Saimme tietoa sanaston aukoista, ison joukon tarpeellisia lisäyksiä sekä monia tarkennuksia ja tärkeitä korjauksia. Kirjallista palautetta antoivat Timo Alanko, Kari Djerf, Mervi Eerola ja hänen turkulaiset kollegansa, Lasse Holmström, Antti Honkela, Pauliina Ilmonen, Juha Karvanen, Jukka Kohonen, Lasse Leskelä, Jukka Nyblom, Jaakko Peltonen, Paavo Salminen ja Matti Vihola. Esitämme heille parhaat kiitoksemme. Kiitämme myös Matti Hakamaa, Seppo Honkapohjaa, Pentti Huuhtasta, Seppo Laaksosta, Risto Lehtosta, Timo Mäkeläistä, Leif Nordbergia, Erkki Pahkista, Pentti Saikkosta, Pekka Tavailaa, Timo Teräsvirtaa sekä Anna Torvaldsia valaisevista keskusteluista.

Kielikysymyksiin saamistamme hyödyllisistä kommentteista kiitämme Irja Alhoa, Annekatrin Kaivapalua, Harri Mantilaa ja Tiina Onikki-Rantajääsköä. Kaisa Häkkistä ja Petri Lauermaa kiitämme suuresta avusta historialluvussa. Eduskunnan kirjaston ja Kansalliskirjaston avuliaille henkilökunnille osoitamme myös kiitoksemme. Jarno Niemelää ja Tampereen yliopistoa kiitämme taittoavusta.

Uusi sanasto kuroo kiinni vuoden 1975 jälkeisen kehityksen. Termien lukumäärä on kasvanut 1950-luvun sanaston noin 200:sta ja 1970-luvun sanaston noin 800:sta (Suomen Tilastoseura 1954, 1975) tämän sanaston yli 3000:een.

Lopullisista valinnoista vastaamme ryhmänä. Sanaston alkutekstit ovat niinkään yhteisen harkinnan tulosta. Luku 3 on kuitenkin olennaisilta osin Pekka Peren ja luku 4 Esa Läärän kirjoittama.

Päiväsimme sanaston alunperin Helsingissä Agricolankadulla Kalevalan päivänä 28.2.2020. Toukokuulle 2020 aiottu Suomen Tilastoseuran sa-

tavuotisjuhla ja sanaston julkaiseminen lykkääntyivät koronapandemian takia. Teimme lisääjalla tarkennuksia ja lisäyksiä sekä muokkasimme historia-lukua.

Aleksis Kiven päivänä 10.10.2021,

Tekijät

Alkusanat 2. laitokseen

Kun *Tilastotieteen sanasto* ilmestyi vuoden 2021 lopussa, se otettiin vastaan innostuneesti ja noteerattiin myönteisesti myös mediassa. Suomen Tilastoseuran piirissä pidettiin tärkeänä, että sanastoa ylläpidetään ja jatkuvasti kehitetään organisoidusti, ja että termit ja suomennokset ovat yleisesti saatavilla verkosta. Sanastotyöryhmässä tapahtui osittainen vahdinvaihto, kun alkuperäisistä jäsenistä Juha Alho ja Elja Arjas siirsivät viestikapulan Juha Karvaselle ja Lasse Leskelälle.

Olemme täydentäneet *Tilastotieteen sanastoa* niin, että 2. laitoksessa on yli sata uutta termiä. Lisäksi olemme muokanneet joidenkin vanhojen termien suomennoksia, laatineet selventäviä alaviitteitä ja korjanneet löytämämme painovirheet. Kiitämme Hannu Ojaa palautteesta, jonka perusteella olemme muokanneet alalukua 2.6. Vastaaotamme edelleen mielellämme palautetta.

Sanaston 2. laitos julkaistaan pdf-muodossa. Sanaston termit julkaistaan samalla Internetissä Tilastotieteen verkkosanastona eli hakukoneena. Verkkosanaston teknisen toiminnallisuuden on Tilastoseuran toimeksiantosta toteuttanut Niklas Laxström. Verkkosanastosta löytää kätevästi termien käännökset. Hakualgoritmi esimerkiksi palauttaa *statistical graphics* -termin ja sen suomennoksen, jos hakee termin *graphics* suomennosta.

Mikael Agricolan ja suomen kielen päivänä 9.4.2023,
Juha Karvanen, Lasse Leskelä, Esa Läärä ja Pekka Pere

1 Sanaston tarve, käyttötarkoitus ja periaatteita

1.1 Maailma muuttuu

Tilastotieteen suomenkielisten termien keskeisimpiä käyttäjiä ovat alan opettajat ja opiskelijat yliopistoissa, ammattikorkeakouluissa, lukioissa ja muissa toisen asteen oppilaitoksissa. Ministeriöt ja tutkimuslaitokset julkaisevat tilastollisia tutkimusraportteja, joissa edellytetään ajanmukaisen ja täsmällisen kielen käyttöä. Lehdissä, radiossa, TV:ssä ja enenevässä määrin netissä esitetään päivittäin tilastollisen tutkimuksen tuloksia yleis-tajuisessa muodossa laajalle lukijakunnalle.

Tilastotieteen perustermistö on vakiintunutta, ja pääosin se on onnis-tuneesti käännetty suomen kielelle. Tiede kehittyy kuitenkin nopeasti sekä käytännön tarpeiden että teoreettisten innovaatioiden myötä. Suomessa tilastotieteen eri osa-alueilla toimii vain kourallinen aktiivisia tutkijoita, ja he kirjoittavat tieteelliset artikkelinsa pääosin englanniksi. Uusia suomen kielen termejä otetaan käyttöön esimerkiksi luento-opetuksessa lähinnä ”pakon edessä”, jolloin yksittäisten tutkijoiden tekemät valinnat voivat poi-keta toisistaan. Alan opettajat yliopistoissa ovat olleet hyvin tietoisia siitä, että tämä ei aina johda hyvään kieleen. Tästä aiheutuvat sivistykselliset ongelmat ovat olleet yleinen huolen aihe.

Datatieteen nopea kehittyminen korostaa tätä ilmiötä, kun sen alan tutkijat kohtaavat tilastotieteen termistöä lähes yksinomaan englanninkie-listen tekstien kautta. Soveltajat muokkaavat tilastotieteen kieltä oman alansa julkaisuissa ja opetuksessa. Tilastotieteen kieli on kuin englantia: si-tä käyttävät kaikki, niin sitä taitavat kuin taitamattomat. Yhdessä nämä ilmiöt uhkaavat pirstouttaa tilastotieteen suomenkielistä sanastoa.

Tilastotieteen opetus on Suomen yliopistoissa enenevässä määrin englan-ninkielistä. Tähän ovat johtaneet sinänsä myönteiset pyrkimykset lisätä yliopistolaitoksen kansainvälisyyttä. Englanninkielisen ammattitermistön tuntemus on toki pitkään ollut käytännön ammattitaidon osa. Ammattikie-len säilymisen ongelma syntyy, jos suomenkieliset opinnäytteet harvinais-tuvat. Erityisesti kandidaatin- ja pro gradu -tutkielmilla, diplomitoilla sekä ammattikorkeakoulujen lopputöillä on tässä keskeinen asema. Niissä opis-keija on perinteisesti joutunut esittämään synteessin lähdekirjallisuudesta äidinkielellään, mikä on ylläpitänyt suomenkielisen termistön ajantasaista kehitystä. Tätä näkökulmaa ei kenties ole riittävästi otettu huomioon, kun on päätetty opinnäytteiden kielivaatimuksista.

Tämän sanaston tarkoitus on olla osa tilastotieteellistä sivistysprojek-tia Suomessa. Olemme käyneet läpi tilastotieteen keskeistä termistöä, mu-

kaan lukien viime vuosien kehitysvirtauksia. Pyrkimyksenämme on tarjota lyhyitä, suomen kielen mukaisia vastineita englanninkielisille tilastotieteen termeille. Monet englanninkieliset termit on luontevasti mukautettu suomeen; esimerkiksi lineaarinen ja varianssi. Sen sijaan joidenkin raakalainojen kuten *odds ratio* ja *survey* luontevuutta ei voi kiittää (alaluvut 2.9 ja 2.11).¹

1.2 Pohdintoja hyvistä termeistä

Tieteellinen ajattelu perustuu käsitteisiin, jotka määritellään. Sanat, joilla käsitteisiin viitataan, ovat termejä. Termit ovat sopimuksenvaraisia, ja siksi yksittäinen termi ei voi olla sen enempää oikea kuin vääräkään. Asiatyhteydessä termin käyttö, niin kuin muidenkin sanojen, voi tietysti olla oikeaa tai väärää.

Tieteen termistöltä odotetaan yksikäsitteisyyttä. Yksittäisen termin pitäisi mielellään viitata vain yhteen käsitteeseen. Vaikka yksikäsitteisyydestä ei pystytä aina pitämään kiinni eri tieteiden välillä, on se ihannetila, johon pyritään. Toiseen suuntaan yksikäsitteisyys toimii huonommin. Tilastotieteessäkin on runsaasti esimerkkejä siitä, että samaan käsitteeseen viitataan useilla eri termeillä.

Tilastotieteen teoria on matemaattista, ja käytetyt käsitteet ovat universaaleja, kansallisesta kulttuurista riippumattomia. Erilaisista käyttötarpeista ja historiaperinteistä johtuen tilastotieteen sovellukset voivat kuitenkin poiketa maasta toiseen. Myöskään virallisen tilastotoimen käsitteistöt eivät kaikin osin ole vastanneet toisiaan.

Kansainvälisten tilastojen vertailtavuuden kohentamiseksi muun muassa YK ja OECD ovat pyrkineet yhdenmukaistamaan tiedon keruun tapoja ja ja tilastoinnissa käytettävää termistöä. Tämän tuloksena tilastotieteen käsitteistö on ainakin teollistuneissa maissa pääosin yhtenevää. Kansainvälinen tilastoinstituutti ISI on tuottanut tilastotieteen sanaston 29:llä eri kielellä (*International Statistical Institute* 2011), ja siinä mukana on myös suomi. Se on ollut tärkein lähde sanastollemme.

Tilastotieteen kielessä on paljon sanoja, joita esiintyy arkikielessä tai toisten erikoisalojen termistössä. Useimmiten tällaisten sanojen sivumerkitykset eivät aiheuta sekaannuksia. Esimerkiksi termi regressioanalyysi (*regression analysis*) tilastollisessa tekstissä tuskin johtaa mielikuvaan, että kyse on psyykkisestä regressiosta. Termissä logistinen regressio määritte logistinen ei myöskään vie ajatuksia tavaroiden kuljetusta ja varastointia koskeviin ratkaisuihin. Sana *harha* (*bias*) johdannaisineen esiintyy monenlaisissa ei-tilastollisissa yhteyksissä kuten aistiharha, harhaluulo, harhau-

¹Huomion kohteena ovat suomenkieliset sanat osoitamme sans serif -kirjasinlaajilla; vieraskieliset *italics* -kirjasinlaajilla. Luvussa 3 suomenkieliset esitetyssä merkityksessä vanhentuneiksi tulkitut termit merkitsemme pääsääntöisesti ristillä †.

tuminen, jne. Nämäkään sivumerkitykset johtavat harvoin harhaan, kun puhutaan estimaattorin harhasta² tai harhattomasta otannasta.

Kielten ja kulttuurien eroista johtuen tilastotieteen termien arkikieliset sivumerkitykset voivat olla hyvinkin erilaisia eri kielissä. Onnellisessa tapauksessa sivumerkitykset olisivat samansuuntaisia termin tieteellisen sisällön kanssa. Tämä helpottaisi tieteen tutkimustulosten esittämistä julkisessa keskustelussa.

Valitettavasti merkitysten samansuuntaisuuteen ei aina päästä, ja eroavat sivumerkitykset häiritsevät ymmärtämistä. Tuonnempana (esim. alaluvuissa 2.3 ja 2.8) käsitellään yksityiskohtaisesti esimerkkejä, joilla on englannissa suuri määrä sivumerkityksiä ilman, että nämä liittyvät tilastotieteeseen millään tavalla.

Toisenlainen vastaavuusongelma syntyy, kun termin tekninen merkitys jää katveeseen, ja sanan arkikielinen, mahdollisesti latautunut merkitys värittää tilastollisen analyysin tulkintaa. Ongelmallisia ovat esimerkiksi termit, jotka on johdettu englannin sanasta *significance*, suomeksi ”merkittävyys”, ja joiden tekninen sisältö on riippumaton mahdollisesta sisällöllisestä merkittävyydestä (ks. alaluku 2.1). Toinen virhetulkinta on antaa termille *odds ratio*, suomeksi *ristisuhde*, merkitys ”riskisuhde” sen vuoksi, että englannin sanaa *odds* usein käytetään kuvaamaan tapahtuman mahdollisuutta tai riskiä (ks. alaluku 2.9).

Tieteen kehittyessä uutta käsitteistöä syntyy jatkuvasti. Vaikka vakiintuneeseen termistöön liittyy edellä kuvattuja ongelmia, uusia käsitteitä vastaavien uusien termien soisi olevan niistä vapaita. Tämä on yksi tieteellisen kielenhuollon tehtävä. Yksikäsitteisyyden ideaali johtaa usein siihen, että termit otetaan käyttöön muodossa, joka on mutkikas. Käytännössä saatetaan sitten havaita, ettei lyhyempikään muotoilu johtaisi käsitteellisiin ongelmiin. Kielenhuoltoon kuuluu siten sekin, että aika ajoin pohditaan, voisiko termistöä napakoittaa.

Lopulta on vielä sellaisia termien kivettymiä, jotka ovat syystä tai toisesta kielen rakenteisiin sopimattomia mutta joita voi olla vaikea muuttaa. Kieleen paremmin sopivien vaihtoehtojen tarjoaminen on kuitenkin kielestä kiinnostuneen tutkijayhteisön eettinen velvollisuus, jota olemme parhaamme mukaan pyrkineet täyttämään.

1.3 Käytännön periaatteita sanastotyössä

Englannin kielessä sama sana voi usein olla substantiivi tai verbi. On myös tapauksia, joissa sama sana voi olla substantiivi ja adjektiivi. Näissä tapauksissa sanastoon on yleensä otettu substantiivi. Esimerkiksi termille

²Tilastotieteen opetuksessa joudutaan toki perustelemaan, miksi harhainen estimaattori on usein parempi kuin harhaton.

sample on annettu vain kaksi merkitystä: otos ja näyte, vaikka molemmat toimivat myös verbeinä: ottaa otos ja ottaa näyte. Vastaava tekijänimi on *sampler*, jonka olemme kääntäneet sanalla poimija, joten verbin *sample* voi myös kääntää sanalla poimia.

Englanninkielisissä tilastotieteen teksteissä esiintyy adjektiiveja tai partitiippimuotoisia sanoja, joilla sellaisenaan voi olla monia merkityksiä myös yleiskielessä, mutta vasta niiden yhdistäminen yhden tai useamman sanan kanssa sanaliitoksi tuottaa täsmällisen termin. Pääsääntöisesti tällaisia määresanoja ei yksinään ole otettu sanastoon. Sanastossa on esimerkinomaisesti sanaliittotyyppisiä termejä, joissa kyseinen määre esiintyy. Niinpä sana *adjusted* ei yksinään ole mukana, mutta termit *adjusted estimator* (vakioitu estimaattori) ja *adjusted profile likelihood* (muokattu profiiliuskottavuus) ovat. Toisaalta termi *admissible* (käypä, hyväksyttävä) on mukana, koska sillä on itsenäinen tekninen merkitys, mutta mukana on myös esimerkiksi *admissible decision function* (käypä päätösfunktio).

Tilastotieteessä on suuri joukko matemaattisia tuloksia (lauseita ja apulauseita), jotka erotetaan toisistaan keksijänsä mukaan. Samoin on valtava määrä jakaumia, prosesseja ja testejä, joihin viitataan henkilön nimellä. Nämä termit on pääosin jätetty pois. Mukaan on otettu kuitenkin joukko kaikkein keskeisimpiä. Perusteena on ollut klassisuus ja se, että suomenkieliseen kirjoitusasuun liittyy jokin valinta. Niinpä mukana on *Poisson distribution* (Poisson-jakauma), *Gaussian distribution* (Gauszin jakauma), *Markov chain* (Markovin ketju) ja *Neyman–Pearson lemma* (Neymanin–Pearsonin apulause). Valinta yhdyssana- ja genetiivimuodossa olevan termin välillä tehdään myös englannissa. Esimerkiksi *Fisher information* (Fisher-informaatio) ja *Fisher’s distribution* (Fisherin jakauma) ovat tavallisimmin käytettyjä muotoiluja, mutta varsinaista syytä erolle ei ole.

Lisäongelma syntyy siitä, että vierasperäisen nimen äänneasu voi olla vaikeasti lausuttava tai tuntematon, tai siitä että nimeä on vaikea taivuttaa sijamuotoihin. Nimestä voi myös olla hankalaa muodostaa adjektiivimuotoa. Näihin aiheisiin palataan alaluvuissa 2.4 ja 2.5.

Suomen kielen joustava piirre on yhdyssanojen vapaa muodostus. Olemme pyrkineet valitsemaan yhdyssanojen osat niin, että lopputulos on kohutuullisen lyhyt – siitäkin huolimatta, että sananmukainen käänнос vastaisi yksityiskohtaisemmin alkuperäistä termiä. Niinpä *leave-one-out cross-validation* on käännetty yksi-pois-ristiinvalidointi. Vastaava pyrkimys on havaittavissa englannissakin, jossa *box plot* on syrjäyttämässä alkuperäisen termin *box-and-whisker plot*. Molempien käänнос on laatikkokuvio. Näitä hankalampi tapaus on toksikologisiin eläinkokeisiin liittyvä termi *median lethal dose*, jonka vakiintunut lyhenne on *LD50*. Tämän käänsimme termillä puolet tappava annos.

Lyhenteitä käytetään tilastotieteessä runsaasti, ehkä liiaksikin, sillä ne ovat usein spesifejä jollekin osa-alueelle. Niitä on otettu mukaan arvioi-

dun yleisyyden mukaan. Ongelmia syntyy, kun esimerkiksi ekonometriassa GMM viittaa termiin *generalized method of moments* mutta toisaalla termiin *Gaussian mixture model*. ABC voi puolestaan viitata termiin *approximate Bayesian computation* tai termiin *approximate bootstrap confidence interval*.

Sanastoa laatiessamme olemme pyrkineet kääntämään samanlaiset tapaukset yhdellä tavalla. Periaatetta ei voi kuitenkaan aina noudattaa johdonmukaisesti. Termin *nonparametric* olemme kääntäneet parametrin tai epäparametrinen. Melko yleisesti käytettyä muotoa ei-parametrinen pidämme tarpeettomana, koska parempia vaihtoehtoja on. Kuitenkaan termiä *non-normal* ei voi kääntää ”normaaliton” tai ”epänormaalinen”, ja ei-normaalinen näyttää sopivimmalta.

Edeltävästä esimerkistä käy myös ilmi se, että englannissa yhdysmerkin käytöstä ei ole yksikäsitteisiä sääntöjä. Olemme valinneet sanastoon yleisempänä pitämämme muodon.

Omanlaisensa ongelma ovat sellaiset englannin termit, jotka on alun perin huonosti muodostettu. Näitä olemme harkintamme mukaan jättäneet pois. Epätyytyttäviä termejä on silti sanastossa.

Amerikanenglanti ja brittienglanti poikkeavat eräissä tapauksissa toisistaan. Olemme valinneet johdonmukaisesti amerikanenglannin kirjoitus-asun.

Tilastotieteessä on tapana käyttää kreikkalaisia kirjaimia, matematiikan tapaan, parametrien symboleina. Myös joissain termeissä on kreikkalaisia kirjaimia. Näitä ovat esimerkiksi χ^2 -distribution, Cronbach's α ja δ -method. Yksinkertaisuuden vuoksi olemme kirjoittaneet ne englanniksi latinalaisin kirjaimin *chi-squared distribution*, *Cronbach's alpha* ja *delta method* ja suomeksi Cronbachin alfa ja deltamenetelmä. Suomen kielen ohje on kirjoittaa kreikkalaiset kirjaimet niiden ääntämyksen mukaan suosien kaksoisvokaaleja: β = ”beeta”, ϕ = ”fii” ja χ = ”khii”. Pidämme luontevampana kirjoittaa lyhyin vokaalein: betajakauma, fi-kerroin ja khi-neliöjakauma. Tällaista kirjoitusasua käytetään suomenkielisissä matemaattisissa teksteissä paljon.

Foneettista kirjoitusta on tarvittu selventämään valintoja esimerkiksi alaluvussa 2.4. Olemme valinneet yksinkertaisimpana pitämämme kirjoitustavan, jolla olemme saaneet ilmaistua tarkoituksemme.

1.4 Monesta moneen -kuvaus

Käsitteiden ja niitä vastaavien sanojen suhde vaihtelee kielestä toiseen. Esimerkiksi sekä englannin *expectation* että *mean* voivat viitata odotusarvoon. On myös mahdollista, että yhdellä sanalla viitataan asiayhteydestä riippuen eri käsitteisiin. Tästä on esimerkkinä alaluvussa 2.8 tarkemmin käsitelty englannin sana *rate*. Kummassakin tapauksessa suomen kielen

käytäntö poikkeaa englannista.

Tässä sanastossa monesta moneen -kuvaus on ratkaistu seuraavasti. Silloin kun lähtökohtana olevalla englannin sanalla on useita käsitteellisiä vastineita, ne on erotettu alaindeksillä (esim. *rate*₁, *rate*₂ ja *rate*₃). Näin lähtökohtana on aina yhtä käsitettä vastaava englannin sana. Silloin kun samaa käsitettä vastaa useita sanoja (esim. *expectation*, *expected value* ja *mean*), niistä on harkinnanvaraisesti valittu suositeltavin (esimerkin tapauksessa *expectation*) ja muiden kohdalla mainitaan, että ne ovat tämän sanan synonyymejä. Suositeltavimmalle sanalle on esitetty korkeintaan kolme suomenkielistä vastinetta (esimerkin tapauksessa vain yksi: *odotusarvo*). Tässä tapauksessa englannin *mean*-sanalla on kaksi käsitteellistä vastinetta: keskiarvo ja odotusarvo. Ne on merkitty sanastoon muodossa *mean*₁ ja *mean*₂. Edellisen synonyymi on *average*, suomeksi keskiarvo, ja jälkimmäisen *expectation*, suomeksi odotusarvo.

Pääosa monesta moneen -tapauksista ratkeaa edellä kuvatulla heuristiikalla, etenkin kun huonoina pitämiämme suomenkielisiä termejä tai englannin yleiskielisiä ilmauksia on jätetty pois. Alaviitteissä on esitetty lisätarkennuksia. Esimerkiksi sananmukainen käännös ”standardipoikkeama” saksan, englannin tai ruotsin kielistä ei mielestämme ole tarpeen, kun nasevampi *keskihajonta* on alalla vakiintunut termi.

Koska tilastotieteen käsitteet ovat Suomessa ja anglosaksisissa maissa samat, suomi-englanti-sanasto voidaan olennaisilta osiltaan koostaa samoilla periaatteilla kuin englanti-suomi-sanastokin. Yksityiskohdissa on kuitenkin eroja, jotka syntyvät siitä, että suomessa ei ole kaikille käsitteille yhtä vakiintuneita termejä kuin englannissa. Näissä tapauksissa yhtä englannin termiä on joskus valittu vastaamaan kaksi tai kolmekin suomenkielistä vastinetta, joista jotkin voivat olla uudissanoja. Tämän seurauksena suomenkielisiä käännöksiä on suomi-englanti-sanastossa enemmän kuin alkuperäisiä englanninkielisiä sanoja.

Sanastossa ei ole pyritty täydelliseen kattavuuteen. Englannissa on runsaasti harvinaisia tilastotieteen termejä, joita ei ole otettu mukaan sanastoon siksi, ettei niitä suomen kielessä juuri ole käytetty. Lisäksi englanti-suomi-sanastoon mukaan otetun harvinaisen sanan tai uudissananan englannin-nos voi tästä asiayhteydestä erotettuna tuntua oudolta. Näistä syistä kaikkia englanti-suomi-sanastossa esiintyviä suomen sanoja ei ole otettu mukaan suomi-englanti-sanastoon.

2 Esimerkkejä pulmallisista termeistä

2.1 *Significance*

Merkitsevyystestit (*significance test*) ovat olleet tilastollisen analyysin keskeisiä työkaluja yli sadan vuoden ajan. Testeihin liittyy monenlaisia käsitteellisiä, menetelmällisiä ja termistöllisiä ongelmia. Ne ovat yhteydessä koulukuntaeroihin, peruskäsitteiden ymmärtämisen ja niiden opettamisen haasteisiin – ja sanavalintoihin.

Merkitsevyystestit toi laajan soveltajajoukon tietoisuuteen R. A. Fisher vaikutusvaltaisessa oppikirjassaan (Fisher 1925). Termin merkitsevyystaso (*level of significance, significance level*) sisällön Fisher esitti yhtäpitäväksi *p*-arvon (*p value*, Fisherin teksteissä “*value of P*”) kanssa. Samassa hengessä merkitsevyystaso on määritelty osassa oppikirjoja (esim. Cox ja Hinkley 1974). Monissa kirjoissa (esim. Lehmann ja Romano 2006) merkitsevyystaso on kuitenkin määritelty samoin kuin hylkäysvirheen (*false rejection*) todennäköisyys.

Merkitsevyystason keskenään ristiriitaiset määritelmät ovat aiheuttaneet käsitteellistä sekaannusta (ks. esim. Greenland 2019). Hylkäysvirheen todennäköisyys on ennen testaamista asetettu kiinteä luku α , eikä havaintojen funktio kuten *p*, jota ei pidä tulkita kiinteänä virhetodennäköisyytenä. Monissa lähteissä erotellaan α ja *p* toisistaan siten, että edellistä kutsutaan nimellä *pre-specified significance level* ja jälkimmäinen on *observed significance level*. Ne voidaan suomentaa vastaavasti ennalta asetettu merkitsevyystaso sekä havaittu merkitsevyystaso.

Riippumatta siitä, miten *significance level* määritellään, sillä on tilastotieteessä neutraali tekninen sisältö. Sanan *significance* arkikielisen merkityksen luoman mielikuvan vaikutuksesta moni soveltaja saattaa tulkita “tilastollisesti merkitsevän” (*statistically significant*) eron tai riippuvuuden sisällöllisesti merkittävänä tai tärkeänä. Samassa hengessä “ei-merkitsevä” (*non-significant*) ero usein koetaan aivan kuin havaittu ero olisi käytännössä merkityksetön. Kummassakin tapauksessa tuloksiin liittyvä epävarmuus usein sivuutetaan, ja johtopäätös esitetään varmempana kuin on aihetta. Sanan *significant* yleiskieliset sivumerkitykset ovat osaltaan myötävaikuttaneet siihen, että tilastollisen testauksen käsitteisiin ja testitulosten tulkintaan liittyvät monet kognitiiviset harhat ovat hyvin yleisiä (ks. Greenland 2017).

Tilastollisen päättelyn pelkistämisestä siihen, onko tulos “merkitsevä” vai “ei-merkitsevä” (riippuen siitä, onko $p < 0.05$ vai ei), sekä *p*-arvon virhetulkinnoista ja valikoivasta raportoinnista on vuosikymmenten kuluessa kirjoitettu lukuisia kriittisiä artikkeleita niin tilastotieteen kuin myös

sovellusalojen lehdissä. Siitä huolimatta “merkitsevän” p -arvon kalastelu (p -hacking) ja testituloksista tehdyt yksioikoiset päätelmät ovat edelleen yleisiä esimerkiksi biologian, terveystieteiden, psykologian ja yhteiskuntatieteiden tutkimuskirjallisuudessa (lääketieteen osalta ks. esim. Ioannidis 2019).

Tilastotieteilijöiden keskuudessa koettu huoli aiheesta kulminoitui viimeksi *The American Statistician* -lehden teemanumerossa *Statistical Inference in the 21st Century: A World beyond $p < 0.05$* , joka ilmestyi 2019. Siinä käsiteltiin monipuolisesti merkitsevyystestien käyttöön liittyviä ongelmia, joiden yhtenä perussyynä tuotiin esiin adjektiivin *significant* latautuneisuus yleiskielessä. Monet kirjoittajat suosittelevat sen välttämistä tutkimustulosten raportoinnissa. Radikaaleimmillaan ehdotettiin koko sanan *significance* hylkäämistä tilastotieteestä (Wasserstein ym. 2019).

Kriittisten kirjoitusten kärki ei enimmäkseen ole kohdistunut tilastolliseen testaamiseen eikä p -arvoihin sinänsä vaan siihen, että p -arvoja tulkitaan ja päätelmiä niiden pohjalta tehdään mekaanisen kaksijakoisesti, ja että tutkijoilla on vahva taipumus tuoda esiin vain “merkitsevät” tulokset. Valtavirtaa edustavan näkemyksen mukaan ajatellaan, että “*significance tests have a valuable but limited role in the analysis of data*” (Cox ja Snell 1981).

Sanastoon on otettu myös mukaan neutraalimpi termi *statistically discernible*, suomeksi tilastollisesti erottuva. Sitä on suositeltu ilmauksen *statistically significant* sijaan (Kelson 2013) haluttaessa kommentoida p -arvoa testaustilanteessa, jossa “signaali erottuu kohinasta” riittävässä määrin – kuitenkin sitoutumatta aina samaan kynnysarvoon kuten 0.05.

Testitulosten tulkinnassa on myös syytä pitää mielessä, että yksittäisen testin tuottama p -arvo on toteuma satunnaismuuttujasta P , jonka nimellinen otantajakauma riippuu havaintoasetelmasta ja käytettyä mallia koskevista oletuksista.¹ Pieni p -arvo ilmaiseekin havaintojen yhteensopimattomuutta (*incompatibility*) joko nollahypoteesin tai tehtyjen mallioletusten tai samanaikaisesti molempien kanssa (Greenland 2019). P -arvojen vaihtelevuus sekä havaintojen mahdollinen yhteensopimattomuus mallioletusten kanssa jäävät valitettavan usein katveeseen, kun tilastollisen testauksen periaatteita opetetaan.

Termi p -value on alun perin lyhenne ilmaukselle *probability value*, jota ei ole otettu sanastoon sen sivumerkitysten takia. Tämä ilmaus ylläpitää sitä yleistä virhekesitystä, jonka mukaan p -arvo olisi “todennäköisyys, että H_0 on tosi”. Tästä mielikuvasta erossa pysymisen kannalta onkin kiinnostava ehdotus skaalata p -arvot informaatioteoriassa käytetyllä muunnoksella $s = -\log_2(p)$. Tuloksena saadaan s -arvo (s value) eli yllätysarvo (*surprisal value*), jolla voi olla p -arvoa puhuttelevampi tulkinta (Greenland 2019).

¹Jos mallioletukset pätevät, niin nollahypoteesin vallitessa satunnaisella P -arvolla on yksikköväliille rajattu tasajakauma.

Nämä uudet termit on myös otettu sanastoon mukaan.

Ilmauksen *statistically significant* korvaaminen esimerkiksi luonnehdinnalla *statistically discernible* tai siirtyminen *p*-arvosta *s*-arvoihin ei sinänsä takaa, että testitulosten mekaaninen jako jollakin mielivaltaisella kriteerillä loppuisi, tai että niitä raportoitaisiin vähemmän valikoivasti (ks. esim. Gelman ja Carlin 2017). Termien muutoksista voi olla haittaakin, jos seurauksena on entistä suurempi sekaannus. Ongelmat ovat kuitenkin todellisia ja osoittavat termistön huolellisuuden ja täsmällisyyden tarpeen.

Korostettakoon, että kritiikki sanan “merkitsevä” käyttöä sekä *p*-arvojen virhetulkintoja ja valikoivaa raportointia kohtaan ei suuntaudu frekventististä tilastollista päättelyä vastaan yleisesti. Myös bayesiläisiä menetelmiä voidaan käyttää väärin mm. siten, että havainnoille oletettu malli ei ole tarpeeksi realistinen, tai että oletetulle priorijakaumalle ei ole asiallisia perusteita.

2.2 *Effect size ja effect magnitude*

Merkitsevyydestit ovat hallinneet monien tieteenalojen tilastollista käytäntöä ainakin 1930-luvulta alkaen. Viime vuosikymmenien aikana on enenevästi vaadittu sellaisten suureiden kuin *effect size* tai *effect magnitude* niin piste-estimaattien kuin luottamusvälien laskentaa ja raportointia merkitsevyyksien ja *p*-arvojen rinnalla.²

Nämä termit ovat kuitenkin monimielisiä. Etusana *effect* luo mielikuvan, että kohteena olisi kausaalinen parametri. Tulkinta toimii, jos *effect size* tarkoittaa esimerkiksi vastemuuttujan odotusarvojen erotusta tai riskisuhdetta vertailtavien käsittelyjen välillä, kun nämä estimoidaan satunnaistetusta kokeesta, tai kun sekoittavat tekijät ovat muuten hallinnassa. Näiden termien alle mahtuu kuitenkin mitä moninaisimpia vertailuparametreja ja riippuvuusmittoja (Kirk 2007). Ympäripyöreimmän luonnehdinnan mukaan “... *an effect size (ES) is the amount of anything that’s of research interest...*” (Cumming ja Calin-Jageman 2017).

Sekaannusta lisää että meta-analyysissä termin *effect size* merkitys on usein rajattu kuvaamaan odotusarvojen skaalattua erotusta käsittelyjen välillä, jossa alkuperäinen erotus jaetaan muuttujan keskihajonnalla (Hedges ja Olkin 1987, s. xvi). Tällaisia parametreja ovat mm. *Cohen’s d* ja *Hedges’ g*. Näitä käytetään, kun meta-analyysiin sisällytetyissä alkuperäis-tutkimuksissa samaa teoreettista vastetta mitataan erilaisilla asteikoilla tai mittareilla. Keskihajonnalla skaalattua kontrastia käyttämällä toivotaan päästävän yhteismitallisuuteen eri tutkimusten kesken.

²Psykologian piirissä on jopa alettu kutsua “uudeksi tilastotieteeksi” (*New Statistics*, ks. esim. Cumming ja Calin-Jageman 2017) piste- ja väliestimoinnin perinteisiä menetelmiä.

Huomattakoon, että Hedges ja Olkin (1987) tekevät eron termien *effect size* ja *effect magnitude* välillä. Heidän termistössään edellinen tarkoittaa nimenomaan odotusarvojen skaalattua erotusta. Jälkimmäinen puolestaan kattaa kaikenlaiset vertailuja ja riippuvuuksia kuvaavat parametrit, jotka siis mahtuvat esimerkiksi Cummingin ja Calin-Jagemanin (2017) termin *effect size* alle.

Tiedostaen näiden termien monimielisyyden ja niiden käytön kirjavuuden sanastossa päädyttiin suoriin käännöksiin: *effect size* on vaikutuskoko ja *effect magnitude* on vaikutuksen suuruus jättäen niin lukijan kuin näitä termejä käyttävän vastuulle selvittää, millaisesta suureesta kulloinkin on kysymys.

2.3 Accuracy ja precision

Substantiiveja *accuracy* ja *precision* käytetään englannin arkikielessä synonyymisesti. Sanakirjoissa ne määritellään usein kehämäisesti toistensa avulla. Esimerkiksi *Oxford English Dictionary* -sanakirjan mukaan *accuracy* on “*The quality or state of being correct or precise*”, ja *precision* vastaavasti “*The quality, condition, or fact of being exact and accurate*”. Pohjoismaisista kielistä tuttua on, että kun norjalainen sanoo “*Akkurat!*”, niin ruotsalainen sanoo “*Precis!*” tarkoittaen samaa.³

Englanti-suomi-sanakirjat tarjoavat tyypillisesti kummallekin suomen-nokset “tarkkuus” ja “täsmällisyys” samanarvoisin. Esimerkiksi MOT-sanakirjaston Pro englanti -sanakirjan mukaan *accuracy* on “virheettömyys, tarkkuus, täsmällisyys” ja *precision* on “tarkkuus, täsmällisyys”.

Tilastotieteen samoin kuin metrologian eli mittauksia ja mittayksiköitä käsittelevän opinalan termistössä sanalla *precision* on ollut tapana viitata mittauksen tai estimoinnin satunnaisvirheen pienuuteen ilman, että välttämättä täsmennetään, millä suureella (varianssi, keskihajonta, keski-
virhe, luottamusväli) tätä virhekomponenttia mitataan. *Precision* esiintyy myös teknisenä terminä tarkoittaen kohdemuuttujan jakauman varianssin käänteislukua.

Accuracy on aiemmin tarkoittanut usein samaa kuin *unbiasedness*; suomeksi harhattomuus eli systemaattisen virheen puuttuminen estimoinnissa tai mittauksessa. Nykyään kuitenkin ISO:n suosituksen mukaan *accuracy* kuvaa mieluummin kokonaisvirheen pienuutta, jossa yhdistyvät systemaattinen virhe ja satunnaisvirhe, ja jota voidaan mitata esim. muuttujan keskineliövirheellä tai sen neliöjuurella. Lisäksi ISO suosittelee termin *bias* ja sen johdannaisten välttämistä näissä yhteyksissä ja tarjoaa sanaa *trueness* korvaamaan termin *unbiasedness*. Tämä ratkaisu on sikäli hankala, että

³Suomessa on viime vuosina yleistynyt sanonta “Aivan!” kutakuinkin samassa merkityksessä.

vastaava adjektiivi olisi *true*, mikä on liian vahvasti sanottu mittauksen tai estimoinnin lopputuloksesta.

Pohjoismainen tilastosanasto (1975) tarjosi aikanaan seuraavat yksikäsitteiset suomennot: *accuracy* on tarkkuus ja *precision* on täsmällisyys. Myös metrologian suomenkielisessä sanastossa nämä ovat nykyisin suositeltuja (Hiltunen ym. 2011). Lisäksi tällä alalla *trueness* on suomennettu todenmukaisuus. Todettakoon, että sanan *precision* vastineena on ollut myös toistotarkkuus, joka Hiltusen ym. (2011) mukaan on metrologian sanastosta ”poistuva termi”.

Suomenkielisissä tilastollisissa teksteissä ”tarkkuutta” ja ”täsmällisyyttä” näyttää käytetyn vaihtelevasti. Viime aikoina on yhä enemmän näkynyt termin *precision* kääntämistä ”tarkkuudeksi”. Niinpä *precision matrix*, joka on kovarianssimatriisin käänteismatriisi, on usein suomennettu ”tarkkuusmatriisiksi”.

Ottaen huomioon sanojen ”tarkkuus” ja ”täsmällisyys” synonyymisuuden arkikielessä, on ongelmallista liittää ”täsmällisyyttä” yksikäsitteisesti satunnaisvirheen pienuuteen. Jos esimerkiksi huonosti kalibroitu mittari, jonka satunnaisvirhe on pieni, tuottaa suuresti harhaisia tuloksia, ei mittaus kuulosta kovin täsmälliseltä. Pidämmekin metrologiasta poistuvaa termiä toistotarkkuus parempana yleisterminä kuvaamaan mittauksen tai estimoinnin vähähajontaisuutta, joten suosittelemme sitä termin *precision* suomennotukseksi. Samalla *accuracy* saa edelleen olla tarkkuus. Koska *precision matrix* tunnetaan myös nimellä *concentration matrix*, sille luonteva suomennotus on keskittyneisyysmatriisi. Tällä linjalla edeten *precision parameter* voidaan myös suoraan kääntää keskittyneisyysparametriksi, mikä hyvin kuvaakin varianssin käänteisluvun olemusta.

Koneoppimisyhteyksissä luokittelualgoritmeja arvioidaan suurella *precision*, joka tilastotieteessä tunnetaan nimellä *positive predictive value*, suomeksi positiivinen ennustearvo. Koneoppimiseen *precision* lienee omaksuttu tiedonhankinnan (*information retrieval*) termistöstä, jossa se määrittää tiedonhakujärjestelmästä tulostettujen relevanttien tietueiden osuudeksi kaikista tulostetuista tietueista (Tietohuollon sanasto 1993). Tällä alalla termin synonyymeinä esiintyvät mm. *precision ratio*, *relevance ratio* – ja jopa *accuracy*. Suomenkielisiä vastineita ovat tarkkuus ja tarkkuusarvo. On siis monta syytä olla tarkkana termien *accuracy* ja *precision* kanssa.

2.4 Bayes

Erisnimi Bayes esiintyy sanastossa mm. termien *Bayesian*, *Bayes factor*, *Bayes network*, *Bayes risk*, *Bayes’ formula*, *Bayesian inference*, *Bayesian information criterion (BIC)* ja *Bayesian updating* yhteydessä. Nimi Bayes voi esiintyä siten perusmuodossa ilman yhdistävää yhdysmerkkiä termin toiseen osaan, genetiivimuodossa tai nimestä johdettuna adjektiivina. Kai-

kista kolmesta erisnimen käyttötavasta tarvitaan vastineet myös suomeen.

Termien *Bayes factor*, *Bayes network* ja *Bayes risk* kohdalla olemme lisänneet erisnimen jälkeen vain yhdysmerkin. Siten esimerkiksi *Bayes factor* saa sanastossa vastineet Bayes-kerroin ja Bayes-tekijä. Genetiivimuoto *Bayes' formula* on suomennettu termiksi Bayesin kaava. Tästä termistä käytetään usein myös synonyymimuotoa *Bayes' theorem*. Se on edellä luetelluista Bayes-termeistä ainoa, joka voidaan suoraan liittää Thomas Bayesin omaan tieteelliseen kontribuutioon, postuumina julkaistuu kirjoitukseen vuodelta 1763.

Vierasperäisten termien kirjoitusasuun lisäksi on tärkeää sopia, miten nimi tulisi suomen kielessä lausua. Erisnimen Bayes kirjaimellinen ääntäminen muodossa /bayes/ on hankalaa, etenkin taivutettuna. Nimen alkuperäinen englantilainen äänneasu /beiz/ olisi sitä vastoin suhteellisen helposti lausuttavissa myös suomessa. Vielä helpommin, mutta edelleen ymmärrettävästi, se voidaan ääntää sanan *base* tapaan muodossa /beis/. Alkukieltä kokonaan vastaamattomia, pitempiä äänneasuja /beies/ ja /baies/ emme pidä hyvinä.

Englannin kielessä adjektiivien muodostaminen substantiivista on suomea helpompaa. Edellä mainituista seitsemästä termistä kolmessa on adjektiiviattribuutti *Bayesian*, lausuttuna /beiziän/. Näiden vastineina ovat suomenkielisissä tekstissä esiintyneet ainakin “bayesiläinen” ja “bayesilainen”. Katsomme, että monesti olisi parempi käyttää yhdysviivaa: *Bayesian inference* on Bayes-päätely, *Bayesian information criterion* on Bayes-informaatiokriteeri ja *Bayesian updating* on Bayes-päivitys. Toisissa yhteyksissä on luontevampaa käyttää adjektiivia kuten luonnehdittaessa yksittäisen henkilön suhtautumista tilastotieteen eri paradigmoihin: “*She is a Bayesian*”. Ehdotamme, että tällöin käytettäisiin adjektiivia bayesiläinen, luettuna /beisiläinen/, tai hieman lyhyemmin bayesläinen, luettuna /beisläinen/.

2.5 Gauss ja Markov

Toisin kuin sanan *Bayes* kohdalla, erisnimien *Gauss* ja *Markov* kirjaimellinen suomalainen ääntämys ei poikkea suuresti alkukielisestä (saksalaisesta ja venäläisestä) ääntämyksestä. Nimen *Gauss* kohdalla ongelmia syntyy enemmänkin englannin kielen käytännöistä. Genetiivimuoto *Gauss'* on ilmeisesti koettu hankalaksi etenkin esiintyessään tieteellisten termien alkuosana. Genetiivin sijaan käytetään yleisesti adjektiivia *Gaussian*. Esimerkiksi termin *sub-Gaussian* käänneämme muodossa *aligaussinen*.

Useimmissa tapauksissa suomen genetiivimuoto on vaikuttanut meistä adjektiivia luontevammalta ratkaisulta. Esimerkiksi termit *Gaussian distribution* ja *Gaussian process* ovat sanastossa muodossa *Gaussin jakauma* ja *Gaussin prosessi*. Nimen Bayes yhteydessä suosimaamme kirjoitustapaa olemme tässä yhteydessä vierastaneet, eli emme kirjoita “Gauss-prosessi”.

Englannissa *Markov* kirjoitetaan termin osana useimmiten perusmuodossaan, esimerkkinä *Markov chain* ja *Markov process*. Suomessa käytetään sekä genetiivimuotoa että yhdysmerkkiä ilman selvästi tunnistettavaa logiikkaa. Emme ole tässä asiassa pyrkineet oppimestareiksi vaan olemme tarjonneet sanastossa molemmat vaihtoehdot asettaen ensimmäiseksi vaihtoehdon, jota olemme pitäneet yleisempänä. Niinpä termin *Markov chain* ykkösvaihtoehto on Markovin ketju, mutta termin *Markov process* ykkösvaihtoehto on Markov-prosessi. Sen sijaan *Markov property* on sanastossa vain Markov-ominaisuus. Sekä genetiiviä että yhdysmerkkiä käytetään, kun *Gauss-Markov theorem* kirjoitetaan suomessa muodossa *Gaussin-Markovin lause*.

Analogisesti *Gaussian*-adjektiivin kanssa myös nimestä *Markov* käytetään englannissa adjektiivimuotoa, kuten “*The dependence is Markovian*”. Tämän voi kääntää esimerkiksi muodossa “*Riippuvuus on markovinen*”. Eriytyisesti satunnaisprosesseista puhuttaessa englannissa saatetaan käyttää sukunimeä sellaisenaan adjektiivisesti: “*This process is Markov*”, jonka ytimekäs käännös olisi “Tämä prosessi on markov”. Aika näyttää, tuleeko tämä lyhyt muoto suomen kielessä käyttöön.

2.6 Kurtosis ja peakedness

Tarkastellaan satunnaismuuttujaa X , jolla on odotusarvo $\mu = E(X)$ sekä keskusmomentit $\mu_k = E(X - \mu)^k$, $k = 2, 3, 4$, keskihajonnan ollessa $\sigma = \sqrt{\mu_2}$. Standardoitua 4. momenttia $\gamma = \mu_4/\sigma^4$ tai vaihtoehtoisesti muunnosta $\kappa = \gamma - 3$ kutsutaan englanninkielisessä kirjallisuudessa yleisesti nimellä *kurtosis*. Sanan alkuperä on kreikan $\kappa\rho\tau\omega\sigma\eta\varsigma$, suomeksi “kaarevuus”. Myös termejä *excess* ja *excess kurtosis* käytetään erityisesti suurelle κ . Jos X noudattaa normaalijakaumaa, niin $\gamma = 3$ ja $\kappa = 0$.

Termin *kurtosis* otti käyttöön Karl Pearson (1905). Hänen kehittämässään jatkuvien yksihiippuisten jakaumien järjestelmässä kunkin tiheysfunktion muotoon vaikuttavat parametrit ovat esitettävissä edellä mainittujen momenttien avulla. Suurella κ Pearson mittasi tiheysfunktion “huipun laakeutta” (“*flat-toppedness*”) verrattuna normaalijakaumaan. Hän kohdisti huomionsa parametrin κ poikkeamiin nolasta lähinnä omaan järjestelmänsä sisältämien symmetristen jakaumien osajoukossa luonnehtien adjektiivein *platykurtic*, *mesokurtic* ja *leptokurtic* jakaumia, joilla vastavasti $\kappa < 0$, $\kappa = 0$ ja $\kappa > 0$. Kun $\kappa < 0$, jakauman tiheys μ :n ympäristössä on normaalijakaumaa laakeampi eli vähemmän kaareva, mutta kun $\kappa > 0$, huippu on vähemmän laakea (“*less flat-topped*”) eli tiheys on keskialueella kaarevampi kuin normaalijakaumalla.

Pearsonin jälkeen termin *kurtosis* synonyymiksi on laajalti omaksuttu *peakedness*, suomeksi huipukkuus (Suomen Tilastoseura 1975). Lukuisat oppikirjat tarjoavat edelleenkin vain parametria κ tai γ näin luonnehdi-

tun jakaumapiirteiden mitaksi ilman varauksia sen yleispätevästä mielekkyydestä. Monet ohjelmistot tulostavat rutiininomaisesti muiden kuvailevien tunnuslukujen rinnalla estimaatin γ :lle tai κ :lle.

Jo 1940-luvulta lähtien on julkaistu esimerkkejä symmetrisistä yksi-huippuisista jakaumista, joissa (a) selvästi laakeatiheyksisellä jakaumalla $\kappa > 0$, (b) normaalijakaumaa jyrkkäharjaisemmalla jakaumalla $\kappa < 0$, ja (c) on sekä jyrkkäharjaisia että laakeita jakamia, joilla $\kappa = 0$. Nämä samoin kuin muutkin paradoksaaliselta näyttävät havainnot κ :n arvoista erityyppisissä jakaumissa ovat herättäneet kriittisiä kysymyksiä kuten, mitä ominaisuutta κ oikeastaan mittaa, onko huipukkuus ylipäättään hyvin määriteltävissä oleva ominaisuus, ja jos on, niin miten sitä on mielekäästi mitata. Tällaista syistä on ehdotettu, että käsitteestä *peakedness* luovutaan, ja että termi *kurtosis* – erityisesti parametri κ – varataan sen sijaan luonnehtimaan jakauman “häntäpainoa” (*tailweight*).

Alan johtavat teoreetikot pitävät edelleen kiinni termistä *peakedness* käyttäen sitä synonyymisesti termin *kurtosis* kanssa. Universaalin mitta-asteikon sijaan lähtökohdaksi otetaan yksittäisten jakaumien *kurtosis*-ominaisuuden eli huipukkuuden keskinäinen vertailu tähän mielekkäästi soveltuvan järjestysrelaation avulla. Järjestys perustuu vertailtavien jakaumien kertymäfunktioiden F ja G pohjalta määriteltyyn kriteeriin. Tällainen on esimerkiksi van Zwetinin kriteeri symmetrisille jakaumille: yhdistetyn funktion $G^{-1} \circ F$ konkaavisuus-konveksisuus (ks. esim. Oja 1981). Kaikkia jakaumia ei voida asettaa yksikäsitteiseen järjestykseen, vaan on tyydyttävä huipukkuuksien vertailuihin ja mittaamiseen rajatun jakaumaluokan sisällä. Todettakoon, että sellaisessa van Zwetinin ehdon täyttävien jakaumien luokassa, joilla 4. momentti on olemassa, parametrin κ arvot noudattavat asianomaista järjestystä, jolloin κ :a voidaan käyttää huipukkuuden mittana tämän luokan jakaumille. Huipukkuudelle on esitetty κ :n rinnalle monia vaihtoehtoisia mittoja, jotka tyypillisesti perustuvat jakauman sopivasti valittuihin kvantileihin.

Huipukkuuden mielekäs tarkastelu ei rajoitu vain symmetrisiin jakaumiin. Jakauman vinoutta (*skewness*) koskevilla lisäehdoilla varustetun ja määrättyä järjestysrelaatiota noudattavan jakaumaluokan sisällä voidaan huipukkuutta vertailla ja mitata myös epäsymmetristen jakaumien kesken (ks. Eberl ja Klar 2023).

Ehdotuksemme on, että termi *kurtosis* suomennetaan sanalla *kurtoosi*. Synonyymeiksi soveltuvat *peakedness* ja *huipukkuus* silloin, kun puhutaan jakauman muodosta yleisellä tasolla ao. ominaisuuden tarkastelun perustuessa asianmukaisesti määriteltyyn järjestysrelaatioon ja sitä noudattavaan huipukkuusmittaan. Sen sijaan viitattaessa pelkästään standardoituun 4. momenttiin γ , sen muunnokseen κ ja/tai ao. suureiden arvoihin, emme suosittele käyttämään synonyymejä *peakedness* ja *huipukkuus*. Jos halutaan luonnehtia parametrin κ arvoa verrattuna 0:aan, niin voidaan

käyttää suomennoksia platykurttinen, mesokurttinen ja leptokurttinen aiemmin mainituille adjektiiveille.⁴

2.7 Frequency

Sanalla *frequency* on englannissa yleensä viitattu jonkin toistuvan tapahtuman esiintymistiheyteen ajassa. Termi *frequency* ja sen suomennos *taajuus* onkin varattu fysiikassa suurelle, jonka SI-mittayksikkö on hertsi = 1/s. Englannin kielessä tämän synonyymina käytetään usein sanaa *rate* (ks. alaluku 2.8).

Muissa yhteyksissä *frequency* voi tarkoittaa jonkin ilmiön yleisyyttä tai useutta ilman täsmällistä määrä- tai aikaviitettä. Epidemiologian kirjallisuudessa *frequency* esiintyi aiemmin siinä tehtävässä, jossa nykyään käytetään sanaa *occurrence*. Tällä viitataan yleisesti sairauksien esiintyvyyteen ilman että täsmennetään, onko kyse sairauden ilmaantuvuudesta (*incidence*) vai vallitsevuudesta (*prevalence*, ks. alaluku 2.8).

Tilastotieteen kielessä termi *frequency* on puolestaan enimmäkseen irti fysikaalisista taajuuksista. Niinpä *frequency distribution* eli frekvenssijakauma kuvaa jonkin muuttujan arvojen empiiristä jakautumista eri luokkiin joko lukumäärinä (*count*) eli absoluuttisina frekvensseinä (*absolute frequency*) tai osuuksina (*proportion*) eli suhteellisina frekvensseinä (*relative frequency*). Jakaumaa voi edelleen kuvata esimerkiksi frekvenssikertymien (*cumulative frequency*) avulla. Sanastossa on useita termejä, joissa *frequency* esiintyy yhdessä jonkin täsmennyksen määrittämisen kanssa. Monien kohdalla frekvenssi on mielestämme edelleen kelpo suomennos. Näin on esimerkiksi genetiikan termin *allele frequency* eli alleelifrekvenssi kohdalla.

2.8 Rate

Sana *rate* on englannissa vielä monikäyttöisempi kuin *frequency*, kuten *Oxford English Dictionary* -sanakirjan sille substantiivina antamasta määritelmästä käy ilmi: “A measure, quantity, or frequency, typically one measured against another quantity or measure”. www.sanakirja.org tarjoaakin sille jopa 23 eri suomennosta! Lisäksi *rate* on verbi, jolla myös on useampi merkitys. Tämä moninaisuus heijastuu väistämättä sanan tulkintaan tilastollisissa yhteyksissä.

Substantiivina *rate* voi ilman täsmennystä viitata ainakin seuraavantyyppisiin suureisiin: (1) *taajuus* (*frequency*, ks. alaluku 2.7) (2) *osuus*

⁴Englanninkielisissä teksteissä näkee joskus myös pitempiä muotoja, esim. *platykurtotic*. Tähän muotoon päädytään, jos substantiivista *kurtosis* muodostetaan adjektiivi *kurtotic*. Kuitenkin alkukielen sana *κύρτωσης* on johdettu adjektiivista *κυρτός*, äännetään /kyr-tos/ (suomeksi “kaareva”), jonka suora englanninno on *kurtic*. Kiitämme Lauri Thurénia tästä tiedosta.

(*proportion*) sekä (3) suhdeluku tai osamäärä (*ratio, quotient*). Kontekstia valaisevan määreen sisällyttäminen käytettyyn termiin auttaa yksikäsitteisyyden saavuttamisessa. Osuva suomenkielinen vastine on silloin usein yhdyssana, jossa alkuosa on ao. määreen suomennos ja loppuosa riippuu kyseisen *rate*-käsitteen erityisluonteesta.

Suuretyyppejä (1)–(3) edustavia termejä ovat esimerkiksi epidemiologisen tutkimuksen piirissä *incidence rate*, *participation rate* sekä *infant mortality rate*.

Incidence rate, suomeksi ilmaantumistiheys, kuvaa uusien tautitapausten esiintymistaaajuutta tai -tiheyttä ajassa. Ilmaantumistiheydet ilmaistaan usein yksinkertaisesti lukumäärinä valittua aikayksikköä kohti, esim. “tapauksia vuodessa” tai “per 100 000 henkilövuotta”. Vastaava suure luotettavuustekniikassa on *failure rate*, suomeksi vikataajuus, joka ilmaisee tietyn laitteistoon tai laitteistoryhmään ilmaantuvien vikojen lukumäärän aikayksikössä. Näiden empiirisesti määritettyjen suureiden teoreettinen vastine on laskuriprosessin $\{N(t); t \geq 0\}$ intensiteetti (*intensity*) $\lambda(t)$, jolle pätee todennäköisyystulkinta $\lambda(t)dt \approx E\{N(t+dt) - N(t)\}$. Tätä yleisempi intensiteettikäsite syntyy, kun oikean puolen odotusarvo määritellään ehdollisena suhteessa laskuriprosessin hetkeä t edeltävään historiaan.

Näille termeille läheistä sukua ovat *hazard rate*, suomeksi uhkatiheys, ja sen synonyymi *hazard function* eli uhkafunktio. Niillä kuvataan yksittäisen päätetapahtuman kuten kuoleman riskiä tilanteessa, jossa tapahtuma ei ole vielä sattunut. Yksinkertaisimmassa tapauksessa, jos T on elinajan pituutta kuvaava jatkuvasti jakautunut satunnaismuuttuja, $F(t)$ sitä vastaava kertymäfunktio ja $f(t) = F'(t)$ tiheysfunktio, niin uhkatiheys $h(t)$ määritellään kaavalla $h(t) = f(t)/[1 - F(t)]$. Uhkatiheydellä on siten todennäköisyystulkinta $h(t)dt \approx P(t < T \leq t + dt \mid T > t)$. Termin *hazard rate* synonyyminä käytetään joskus myös jo edellä esiintynyttä termiä *failure rate*. Saman termin käyttäminen vastaamaan kahta toisistaan hieman poikkeavaa käsitettä johtaa kuitenkin helposti sekaannuksiin. Tästä syystä suosittelemme termin *failure rate* suomentamista sen merkityksen mukaan joko sanalla vikataajuus tai uhkatiheys.

Joskus *rate* viittaa spatiaaliseen tiheyteen. Näin on esimerkiksi silloin, kun *traffic-related death rate* eli liikennekuolevuus tarkoittaa liikenneonnettomuuksissa kuolleiden lukumäärää jaettuna henkilökilometrien määrällä tietyllä alueella jonakin ajanjaksona.

Participation rate eli osallistumisosuus tai osallistumisaste on puolestaan suhteellinen osuus eli *proportion*, joka on laaduton luku 0:n ja 1:n välillä (tai välillä 0–100 %). Osuuksia ovat myös kansantaloutta kuvaavat *unemployment rate* eli työttömyysosuus tai työttömyysaste sekä *tax rate* eli veroaste. Vanhemmassa epidemiologisessa kirjallisuudessa paljon käytettyjä termejä olivat *prevalence rate* ja *survival rate*, mutta nyttemmin näitä osuuksia enimmäkseen kutsutaan *prevalence proportion* tai pelkkä *preva-*

lence, suomeksi vallitsevuus tai vallitsemisosuus, sekä *survival proportion* eli elossaolo-osuus.

Infant mortality rate eli imeväiskuolleisuusluku on suure, jonka osoittajassa on kohdeväestössä määrättyinä kalenterivuonna alle vuoden ikäisenä kuolleiden lukumäärä ja nimittäjässä samana vuonna elävänä syntyneiden määrä. Kyseessä ei ole osuus, koska osoittajassa oleva lukumäärään sisältyy lapsia, jotka ovat syntyneet edellisenä kalenterivuonna. Täsmällisempi termi olisikin *infant mortality ratio* eli imeväiskuolevuussuhde. Kansainvälisessä rahaliikenteessä tarvitaan suhdelukua nimeltä *exchange rate*, joka on yksittaisen valuutan vaihtokurssi suhteessa viitevaluuttaan.

Sanastossamme on lisäksi yleistermi *rate of change*, suomeksi muutosvauhti tai muutostahti, joka matemaattisesti ilmaistaan derivaattana, ja joka on monikäyttöinen erilaisten dynaamisten ilmiöiden kuvaamisessa. Esimerkiksi väestötieteessä viime vuosikymmeninä suuren kiinnostuksen kohteena on ollut *rate of decline in mortality*, eli kuolevuuden laskuvauhti. Muita ovat mm. kemiassa tarpeellinen *reaction rate* eli reaktionopeus sekä kansantalouden keskeinen tunnusluku *inflation rate* eli inflaatiovauhti.

Todettakoon lopuksi, että verbinä *rate* tarkoittaa usein jonkin asian arvointia tai arvottamista. Suomenkielisissä asiateksteissä näkee puhuttavan “reittauksesta”, esimerkiksi kun kyse on yritysten luottokelpoisuudesta.

2.9 Odds ratio ja odds

Näille keskeisille termeille ei ole vielä syntynyt vakiintuneita suomennoksia, vaikka tarve on ilmeinen. Ehdotamme termin *odds ratio* käännöksiksi ristisuhde ja vetosuhde. Termin *odds* suomennoksina tarjoamme yhdyssanoja vastasuhde ja vetokerroin. Taustana on, että näillä termeillä viitatuilla käsitteillä on matemaattisen perustan ohella myös pitkä historia vedonlyönnissä, erityisesti anglosaksisissa maissa.

Tarkastellaan rinnakkaisten ryhmien asetelmaa, jossa on n_1 altistunutta yksilöä ja n_0 altistumatonta, ja kaksiarvoinen $(1/0)$ vastemuuttuja Y kuvaa sairastumista johonkin tautiin. Olkoot p_1 ja p_0 sairastumisen todennäköisyydet eli riskit näissä joukoissa: $p_x = P(Y = 1 \mid X = x)$, $x = 1, 0$, jossa X on kaksiarvoinen altistusmuuttuja. Kummassakin ryhmässä sairastuvien lukumäärän oletetaan noudattavan binomijakaumaa $\text{Bin}(n_x, p_x)$. Tavanomaisia todennäköisyyksien vertailuparametreja ovat niiden erotus $p_1 - p_0$ ja suhde p_1/p_0 .

Esitettäessä binomijakauma eksponenttiperheen erikoistapauksena, sen kanoninen parametri on todennäköisyyden logit-muunnos, joka määritellään kaavalla $\text{logit}(p_x) = \log[p_x/(1 - p_x)]$, $x = 1, 0$. Näille saadaan yleistetyn lineaarisen mallin mukainen esitys $\text{logit}(p_x) = \alpha + \beta x$, joka on erikoistapaus logistisesta regressiosta. Parametri β saadaan erotuksena $\text{logit}(p_1) - \text{logit}(p_0)$,

ja tulos voidaan kirjoittaa muodossa

$$\beta = \log \left\{ \frac{p_1}{1-p_1} / \frac{p_0}{1-p_0} \right\} \Leftrightarrow e^\beta = \frac{p_1/(1-p_1)}{p_0/(1-p_0)}.$$

Kaavarivin oikealla puolella oleva osamäärien $p_1/(1-p_1)$ ja $p_0/(1-p_0)$ suhde e^β on englanniksi *odds ratio*, jota käytetään paljon todennäköisyyksien vertailussa. Englanniksi voidaan sanoa esimerkiksi: “*The odds of disease among the exposed equal $p_1/(1-p_1)$, and among the non-exposed $p_0/(1-p_0)$* ”. Käännös tälle on “Sairastumisen vastasuhde altistuneille on $p_1/(1-p_1)$ ja altistumattomille $p_0/(1-p_0)$ ”. Peruste uudissanalle **vastasuhde** on, että kyseessä on kummassakin tapauksessa sairastumisen ja sen vastatapausten, eli ei-sairastumisen, todennäköisyyksien suhde.

Tästä vertailuasetelmasta peräisin olevat empiiriset lukumäärät esitetään alkeisoppikirjoissa usein 2×2 -taulukossa seuraavasti

	$Y = 1$	$Y = 0$	Yhteensä
$X = 1$	a	b	n_1
$X = 0$	c	d	n_0
Yhteensä	m_1	m_0	n

Taulukko voi syntyä paitsi asetelmasta, jossa altistusryhmien koot n_1 ja n_0 on kiinnitetty, niin myös vastaavasta **tapaus-verrokkiasetelmasta**, jossa tapausten ja verrokkien lukumäärät m_1 ja m_0 on kiinnitetty.

Kun todennäköisyydet estimoidaan havaituilla suhteellisilla frekvensseillä: $\hat{p}_1 = a/n_1$ ja $\hat{p}_0 = c/n_0$, saadaan parametrin e^β estimaatiksi seuraava suhdeluku:

$$\frac{a/b}{c/d} = \frac{ad}{bc},$$

Odds ratio -termin synonyymi *cross-product ratio* tulee yhtälön oikeasta puolesta. Sen suora käännös on **ristitulosuhde**, mikä on kelpo suomea. Termin voi ilman sekaannuksen vaaraa lyhentää muotoon **ristisuhde**⁵. Tätä napakkaa muotoa suositamme kumppaniksi termin *odds* käännökselle **vastasuhde**.

Huomataan myös, että tapaus-verrokkiasetelmasta laskettu empiirinen **ristisuhde** altistumiselle $\frac{a/b}{c/d} = \frac{ad}{bc}$ on arvoltaan sama kuin yllä. Parametri e^β on siis estimoituva myös tästä asetelmasta, vaikka itse todennäköisyyksiä p_1 ja p_0 ei onnistuttaisi suoraan estimoimaan.

Matemaattis-tilastollisia tarkasteluja varhaisempi lähtökohta näille termeille on vedonlyönti. Olkoon kohteena tapahtuman A todennäköisyys. Oletetaan, että pelaaja on valmis asettamaan $a > 0$ euroa A :n tapahtumisen puolesta, kun vastapelaaja asettaa $b > 0$ euroa. Toisin sanoen, jos A

⁵Tosin geometriassa termillä **ristisuhde** (*cross ratio*) on toinen merkitys.

tapahtuu, pelaaja saa pitää peliin panostamansa a euroa ja saa voitokseen vastapelaajan b euroa, muutoin hän häviää panoksensa a euroa. Tämä on pelaajan kannalta järkevää odotusarvon mielessä, mikäli hän uskoo, että A :n todennäköisyys on $P(A) \geq a/(a+b)$, sillä tällöin vedonlyönnin tuloksen odotettu arvo pelaajalle on vähintään $b \times a/(a+b) - a \times (1 - a/(a+b)) = 0$. Nähdään, että valmius lyödä vetoa a euroa vastaan b euroa, eli suhteessa $a : b$, vastaa suoraan pelaajan arviota tapauksen A ja sen vastatapausten todennäköisyyksistä. Tätä vastaavuutta on pitkään pidetty yhtenä intuitiivisena tapana sanoa, mitä tapahtuman A todennäköisyys ”oikeasti tarkoittaa”, ja sitä käytetään yleisesti subjektiivisten todennäköisyyksien haarukointiin käytännön tilanteissa. Edellä kuvatusa näkökulmasta tapahtuman A vetokerroin olisi pelaajalle $a : b$. Englanniksi voidaan sanoa: “*The odds of A happening are (at least) a : b*”, ja tämän voi kääntää: “Tapauksen A vetokerroin on (ainakin) $a : b$ ”.⁶

Epidemiologisissa ja sosiaalitieteellisissä yhteyksissä *odds ratio* on yleisempi kuin termi *odds*. Vedonlyönnissä tilanne on päinvastainen. Kuitenkin, koska vedonlyönnin panoksilla on edellä kuvattu vastaavuus tapausten arvioituihin todennäköisyyksiin, on mahdollista verrata tapahtumien todennäköisyyksiä pelaajan kannalta suhteuttamalla vetokertoimia toisiinsa. Jos tapauksen A vetokerroin on $a : b$ ja tapauksen B vetokerroin on $c : d$, niin näiden tapausten todennäköisyyksien suhdetta kuvaa vetokertoimien suhde, eli vetosuhde, joka on $(a : b)/(c : d) = \frac{ad}{bc}$. Termi vetosuhde tuo vedonlyönnin mieleen samaan tapaan kuin *odds ratio* ja sitä on jonkin verran käytetty esimerkiksi epidemiologiassa termin *odds ratio* käännöksenä. Tästä syystä pidämme myös näitä vedonlyönnin maailmasta peräisin olevia käännöksiä kelvollisina ja moniin sovelluksiin sopivina.

Kuten edellä olevasta käy ilmi, perussyyt ristisuhteen käytölle ovat teoreettisia. Alkeisopetuksessa sen käyttöä on joskus perusteltu sanomalla, että kun p_1 ja p_0 ovat pieniä, niin ristisuhde on $\approx p_1/p_0$, ts. se olisi suunnilleen sama kuin riskisuhde. Näin onkin. Esimerkiksi jos $p_1 = 0.02$ ja $p_0 = 0.01$, niin riskisuhde on 2 ja ristisuhde on 2.02. Mutta kun $p_1 = 0.6$ ja $p_0 = 0.3$, niin riskisuhde on edelleen 2 mutta ristisuhde 3.5 on siitä jo kaukana. Tuloksen perustana oleva oletus tapausten harvinaisuudesta tuntuu julkisuudessa aina silloin tällöin unohtuvan, kun esimerkiksi logistisen regression tuloksia esitellään.

Sanaa *odds* käytetään englannin arkikielessä hyvin yleisesti myös viittamaan kvalitatiivisessa mielessä jonkin tapahtuman todennäköisyyteen. Esimerkiksi lause “*The odds are high that candidate X wins*” on luontevaa kääntää “On suuri todennäköisyys, että ehdokas X voittaa”. Vastaavasti “*Against all odds, candidate X won*” käännetään “Vastoin kaikkia odotuksia ehdokas X voitti”.

⁶Vetokertoimen $P(A)/[1 - P(A)]$ asemesta Suomessa urheiluvedonlyönnin perussuure on nimeltään ”kerroin”, joka on todennäköisyyden käänteisluku $1/P(A)$.

2.10 Score

Laajassa amerikanenglannin sanakirjassa *Webster's Unabridged Dictionary* annetaan sanalle *score* 12 merkitystä, kun sitä käytetään substantiivina, 13 merkitystä, kun sitä käytetään transitiiiviverbinä, ja vielä viisi lisää, kun sitä käytetään intransitiiviverbinä. Mikään näistä ei vastaa termin *score* tilastotieteellistä päämerkitystä.

Lienee niin, että tilastotieteen ja lähialojen ulkopuolella termin *score* tilastotieteellistä merkitystä ei tunneta, eikä kenties tarvitsekaan. Termin suomentamisen kannalta tämä on helpotus: ei ole sivumerkityksiä, jotka tulisi saada käännökseen mukaan. Suomennoksen käyttäjäkuntakin jäänee varsin rajatuksi. Hiukan paradoksaalisesti tämä merkitsee toisaalta sitä, että on vaikea löytää vahvaa perustetta millekään yksittäiselle käännös-vaihtoehdolle.

Kuvatkoon vektori $\mathbf{y}$ saatuja havaintoja, joiden oletetaan olevan otos tiheysfunktioista $f(\mathbf{y}; \boldsymbol{\theta})$, jossa $\boldsymbol{\theta}$ on parametrivektori. Havaintojen määrä $\boldsymbol{\theta}$:n log-uskottavuusfunktio (*log-likelihood function*) on $\ell(\boldsymbol{\theta}; \mathbf{y}) = \log f(\mathbf{y}; \boldsymbol{\theta})$. Sen gradientti eli osittaisderivaattojen vektori $s(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ on englanniksi *score* tai *score function*. Sopivien oletusten vallitessa sillä on asymptoottinen k -ulotteinen normaalijakauma. Tämän perusteella log-uskottavuuden gradientille voidaan perustaa testejä, joihin englanniksi viitataan termillä *score test*.

ISI:n sanaston mukaan monissa germaanisissa ja romaanisissa kielissä englannin termi on otettu sellaisenaan tilastosanastoon. Ääntämyksen mukainen versio on otettu käyttöön virossa (*skoor*) ja kreikassa ($\sigma\kappa\omicron\rho$, eli /skor/). Ääntämystä muokkaamalla saatu versio on käytössä norjassa (*skå-re*). Ehdotamme, että seuraamme norjan käytäntöä, ja otamme suomeen yhdeksi vaihtoehdoksi sanan *skoori*. Vastaavasti kirjoitamme *skooritesti* ja tehokas *skoori* (*score test*, *efficient score*).

Suurimman uskottavuuden estimaattorin (*maximum likelihood estimator*) arvot ovat usein yhtälön $s(\boldsymbol{\theta}) = \mathbf{0}$ nollakohtia. Yhtälöt ovat pääsääntöisesti epälineaarisia,⁷ ja ne ratkaistaan numeerisin menetelmin. Lähtien alkuarvosta $\boldsymbol{\theta}^{(0)}$ rekursiivinen Newtonin-Raphsonin algoritmi tuottaa askeleella $t + 1$ päivitetyn arvon $\boldsymbol{\theta}^{(t+1)}$ kaavalla

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - [\mathbf{H}(\boldsymbol{\theta}^{(t)})]^{-1} s(\boldsymbol{\theta}^{(t)}),$$

jossa $k \times k$ -matriisi $\mathbf{H}(\boldsymbol{\theta}^{(t)})$ eli Hessen matriisi (*Hessian*) sisältää log-uskottavuusfunktion 2. kertaluvun osittaisderivaatat pisteessä $\boldsymbol{\theta}^{(t)}$. Termi *Fisher's method of scoring* tarkoittaa sitä, että tämä matriisi korvataan sen odotusarvolla $E[\mathbf{H}(\boldsymbol{\theta}^{(t)})]$. Tässä viitataan siis siihen tapaan, jolla päivitysaskel numeerisesti tuotetaan, eikä siihen, että kyseessä on *skoorin*

⁷Normaalijakaumaan perustuvat lineaariset mallit ovat tästä huomattava poikkeus.

nollakohdan etsiminen. Siksi käännämme termin muodossa Fisherin estimointimenetelmä, emmekä sotke termiä skoori tähän yhteyteen.

Tilastotieteen kirjallisuudessa matriisi $\mathbf{J}(\boldsymbol{\theta}) = -\mathbf{H}(\boldsymbol{\theta})$ tunnetaan nimellä havaittu informaatiomatriisi (*observed information matrix*) ja sen odotusarvo $\mathbf{I}(\boldsymbol{\theta}) = E[\mathbf{J}(\boldsymbol{\theta})]$ on Fisher-informaatiomatriisi (*Fisher information matrix*).⁸

Termillä *score* on tilastotieteessä muitakin merkityksiä kuin se, jota on edellä tarkasteltu. Kuten *Webster's* mainitsee, *score* voi olla lukumäärä, pistemäärä, arvosana jne. Olemme poimineet näistä termin pistemäärä toiseksi merkitykseksi. Suomalaisessa kielenkäytössä tämä sana on joskus yhdistetty edellä käsittelemäämme päämerkitykseen. Sillä ei kuitenkaan näytä olevan tekemistä pistemäärien kanssa.

Kokonaan toinen asia on, että pisteytys, englanniksi *scoring*, on sekin tärkeä osa monenlaisia tutkimusasetelmia. Esimerkiksi psyykkisten sairauksien tilastollisessa tutkimuksessa tyydytään usein, mittausvaikeuksien vuoksi, harkinnanvaraiseen tapausten pisteytykseen. Kyseessä voi olla saman oireyhtymän asteiden erottelu tai erilaisten oireiden yhdistäminen yhdeksi tunnusluvuksi. Käytäntö on yleinen myös psykologiassa ja kasvatustieteissä. Samoin menetellään esimerkiksi monitavoitteisessa päätöksenteossa käytetyssä AHP-menetelmässä.

Kolmantena tyyppinä *score*-sanaan liittyvistä tilastotieteellisistä merkityksistä otamme esiin termin *normal score* eli normaaliarvo. Tällä viitataan siihen, että muuttujan todelliset havaitut arvot korvataan muilla arvoilla sen tähden, että näin saadun testisuureen uskotaan olevan jossain suhteessa parempi kuin itse havainnoista laskettu suure. Se, että havainnot korvataan niitä vastaavilla normaaliarvoilla voi esimerkiksi olennaisesti vähentää oudokkien (*outlier*) vaikutusta ja tuottaa pätevämpiä arvioita tarkastelun kohteena olevien ilmiöiden eroista. Normaaliarvoja, jotka ovat standardinormaalijakautuksen kvantiileja, käytetään myös esimerkiksi normaali-jakaumakuvioissa.

2.11 Survey

Suomen kielen termiksi “survey” on huono. Se ei taivu luontevasti kieleemme sääntöjen mukaan. Esimerkiksi monikon inessiivi “surveyissä”, on kömpelö. On lisäksi epäselvää, pitäisikö sana huolellisessa puheessa ääntää kirjaimellisesti /survey/ vai englannin mukaisesti /söövei/.⁹ Monikon inessiivit /surveyissä/ ja /sööveissä/ tuntuvat erityisen kömpelöiltä. Asia yhteydessä termin *survey* voi kääntää eri tavoin, mutta yleistermeiksi ehdotamme substantiivia luotaus ja verbiä luodata.

⁸Normaalijakaumaan perustuvan lineaarisen mallin tilanteessa nämä ovat samat.

⁹Amerikanenglannissa mukaan tulee vielä r-äänne: /söörvei/. Jos kyse on verbistä, painotus on toisella tavulla: /sö-véi/ tai /sör-véi/.

Kun englannin sanaa *survey* käytetään yhteiskuntatieteellisessä yhteydessä, kyse voi olla satunnaisotantaan perustuvasta kyselytutkimuksesta, rekistereistä keräystä kokonaistutkimuksesta tai muusta laajahkosta kar-toituksesta. Näihin voi yhdistyä jonkin osajoukon tarkempi tutkimus.

Luonnontieteiden puolella *survey* voi viitata esimerkiksi maantieteelli-sen tiedon keruuseen. Englannissa ja Yhdysvalloissa tämän tyyppisiä tie-toja kerää *Geological Survey*. Maantiekartat, merikartat tai hallinnolliset kartat ovat näiden laitosten tuottamia. Suomessa näitä instituutioita vas-taa Maanmittauslaitos, joka huolehtii esimerkiksi paikkatietojärjestelmäs-tämme. – Luonnontieteisiin liittyy sekin, että kun ensimmäiset avaruusa-lukset lähetettiin tutkimaan kuun pintaa, niiden nimi oli *Surveyor 1–7*. Ehkä tästä on jäänyt suomeenkin sana “avaruusluotain” tarkoittamaan avaruusalusta, joka on lähetetty tiedonkeruuta varten.

Englannin termi *survey* näyttäisi tulevan vanhahtavasta ranskan sa-nasta *surveier*, joka tarkoittaa “päältä katsomista”. Tästä etymologisesta yhteydestä huolimatta ranskassa ei käytetä *survey*-sanaa tai sen johdannai-sia. Useimmiten *survey*-sanan paras käännös ranskaan lienee substantiivi *sondage* (suomeksi *luotaus*) ja verbi *sonder* (suomeksi *luodata*).¹⁰ Nämä ovat tulleet ruotsin *sondera*-verbin kautta suomen slangiinkin, kun puhu-taan “sondeeraamisesta”. Tätä sanaa saatetaan käyttää esimerkiksi silloin, kun halutaan alustavasti tiedustella jonkun mielipidettä.

Pidämme sanaa *survey* suomeen sopimattomana, vaikka sitä jonkin verran käytetään. Ehdottamamme *luotaus* ei ole aiemmin esiintynyt tilas-totieteen oppikirjoissa. Silti lyhytkin etsintä netissä tuottaa suuren määrän käytännön esimerkkejä, joissa vastaavaa verbiä on jo käytetty nimenomaan tarkoittamaan enemmän tai vähemmän systemaattista tiedonkeruuta, oli sitten kyseessä yhteiskuntatieteellinen tai luonnontieteellinen aihepiiri.¹¹

Luotaus ei tietenkään ole jokaisessa asiayhteydessä paras käännös. Niin-pä, vaikka käännämme esimerkiksi englannin *survey methods* termillä *luo-tausmenetelmät*, esimerkiksi *opinion survey* kääntyy luontevammin termillä mielipidetiedustelu.

¹⁰Toinen sana, joka tietyissä yhteyksissä olisi sopiva käännös, on *enquête*, joka on tullut ruotsiin muodossa *enkät*. Nämä tarkoittavat “tiedustelua”, “tutkimusta”, tai “selvitystä”.

¹¹Kolme esimerkkiä Google-hausta: “Tutkimuksessa *luodataan* mm. yrityksen suuntaa, johtajuutta, kyvykkyyksiä ja käytäntöjä, jotka näkyvät työntekijäkokemuksessa – – ”; “Suomalaiset autokaupoilla -tutkimuksessa *luodattiin* vastaajilta myös niitä tekijöitä, jotka puoltaisivat sähköauton hankintaa tulevaisuudessa.”; “Tutkimuksessa *luodattiin* myös ympäristöystävällisyyden merkitystä. Ympäristöystävällisyys korostuu erityisesti niiden arvoissa, joilla on jo – – ”

3 Tilastotieteen termien suomentamisen historiaa

Termi *statistik* ja sen erikieliset vastineet tarkoittivat 1700- ja 1800-luvuilla yhteiskunnallisten olojen sanallista ja numeerista kuvausta (Harmaja 1939, Luther 1993). Tilastotiede kehittyi voimallisesti 1800-luvulla. Alalla on kolme juurta: valtion ja sen väestön kuvaaminen, tähtitieteen ja geodesian tarpeet sekä todennäköisyyslaskenta.

Ensimmäiset tilastotieteen seurat perustettiin 1830-luvulla Iso-Britanniassa ja Yhdysvalloissa. Ruotsin tilastoseura (*Statistiska föreningen*) sai alkunsa 1901 ja Suomen Tilastoseura 1920.

Ruotsin ja Venäjän vallan aikoina ruotsi oli keskeinen hallinnon kieli. Venäjän vallan aikana ajoittain tuettiin suomen kieltä siteen Ruotsiin heikentämiseksi; ajoittain heikennettiin sananvapauden rajoittamiseksi. Vähitellen kieliasetuksilla 1863 ja 1902, keisarin julistuksella 1881 ja lopulta hallitusmuodon säännöksellä 1919 suomi saatettiin tasavertaiseksi ruotsin kanssa.

Suomen Tiedeseura (*Finska Vetenskaps-Societen*) perustettiin 1838 ja Suomalainen tiedeakatemia 1908. Tiedeseura julkaisi tutkimuksia alkuvuosisikymmeninä ruotsiksi, ranskaksi, saksaksi ja latinaksi (Arppe 1888). Tiedeakatemian yhtenä ajatuksena oli nostaa suomen kieli sivistyskieleksi.

Turun Akatemian väitöskirjat olivat alkuun yleensä latinankielisiä. Vuodesta 1752 lähtien väitöskirja saattoi olla ruotsinkielinen tietyissä oppiaineissa kuten matematiikassa. (Vallinkoski 1966.) Latinan kausi suomalaisessa tieteessä oleellisesti päättyi Aleksanterin yliopiston statuutteihin 1852. Väitöskirja sai jatkossa olla ruotsinkielinen kaikilla aloilla. Keisari määräsi 1858, että väitöskirja voi olla myös suomenkielinen. Aleksanterin yliopisto siirtyi käyttämään suomea pääpöytäkirjakielenään 1911. (Heikel 1940.)

Nykysuomi muotoutui 1800-luvun loppupuoliskolla. Tilastotiede oli lääketieteen ohella ala, jolla suomi murtautui tieteen kieleksi 1800-luvun puolivälissä.

3.1 Valtioon ja väestöön liittyvät termit

Matematiikan soveltuvuudesta yhteiskunnallisten ilmiöiden kuvaamiseen kiisteltiin 1800-luvulla. Adolphe Quetelet (1796–1874) loi 1830-luvulla keskiarvoimisen (*l’homme moyen*) käsitteen ryhmien kuvaamiseen. Monet kysyivät, millaisista ryhmistä keskiarvoja on mielekästä laskea ja miten ryhmien sisäistä ja välistä vaihtelua tulisi kuvata. Jotkut olivat sitä mieltä, ettei keskiarvoja voi käyttää sosiaalisten ilmiöiden kuvaamiseen. Kes-

kiarvoja ja muita kuvailevia tunnuslukuja laskettiin silti valtionhallinnon tarpeisiin. (Stigler 1986, Desrozières 2010.)

Ruotsi – ja Suomi valtakunnan osana – aloitti valtakunnallisten väestötietojen säännöllisen keräämisen 1749 ensimmäisenä maailmassa (Teräsvirta 1987, Luther 1993). Suomessa tilastotieteellä oli erityinen tehtävä 1800-luvun lopulla: kansallisuusaatteeseen herännyt sivistyneistö piti välttämättömänä, että maamme ja väestömme tutkitaan systemaattisesti (Elfving 1985).

Haparoiva alku

Wolmar Schildt (nimimerkki W. Kilpinen) esitti *Helsingfors' Morgonblad* -lehdessä 28.7.1842 valtiota sivuavat sanat †valde (*stat*), †valteinen (*hörande till stat*), †valdelma (*statistik*) ja †valdelmainen (*statistisk*). †Valdelmaa Schildt perusteli aiemmalla †geografi-tieteen suomennoksella †maidelma. Sanan *politik* Schildt suomensi 1848 †valtineeksi (Rapola 1943). †Valdelma ja †valdelmainen jäivät vähälle käytölle.

Reinhold von Becker luennoi 1820-luvulla “tilastoasioita” (Luther 1993). †Statistiikki (*statistik*) on von Beckerin suomennos Gymnaasi- ja Koulu-järjestys -asetuksessa 1843 (Westzyntius 1844, Metsikkö 1951, Kotimaisten kielten keskus 2019):

†Statistiikki eli tieto Euroopan waltakunnista ja erinomaiten Keisarillisesta Wenäjän waltakunnasta.

Paavo Tikkasen tilasto ja tilastotiede

Suomettaressa 11.5.1847 uutisoitiin raha-avustuksesta Paavo Tikkaselle, jotta hän voisi kerätä “Maatieteen ja †Waltawara-tieteen (*Statistik*) ainetta” Savosta (Suomalaisen kirjallisuuden Seura 2019). Edelleen Suomettaressa 9.11.1847 tiedotettiin, että Suomalaisen Kirjallisuuden Seura aikoo kustantaa “alkeis- ja rahwas-kouluille” oppikirjan “Suomen historia ynnä Maatiede ja †Olollinen Näyte (*Statistik*)” ja kehottaa kirjoittamaan “†Olonnäyttäviä ja maatieteellisiä Pitäjä-kertoelmia, jotta saataisiin täydellisiä tietoja maamme kansallisesta ja †olollisesta tilasta (*statistiska uppgifter*)”. †Olollinen on tulkittavissa ensin osoitetussa yhteydessä tilastolliseksi (Suomalaisen kirjallisuuden Seura 2019). Tulkinta lainauksen lopussa on epäselvempi, mutta †tilaa oltiin jollain tapaa selvittämässä.

Ajoittain †luvunlaskulle on annettu merkitys tilasto. Varhaisin esimerkki on Maamiehen ystävä -lehdessä 17.1.1846 (Kotimaisten kielten keskus 2019). Paavo Tikkanen kertoo †luvunlaskuista ja †waltiollisista luvunlaskuista Suomen maatiede -kirjassaan (Tikkanen 1846). Kirjan yhteyksissä sanan

voi ajatella tarkoittavan tilastoa tai kirjanpitoa; jälkimmäiselle Tikkanen lienee kuitenkin varannut sanan †*luwunpitämys* (“Senaatin menoista”).

Tilasto painettiin ensi kerran Tikkasen (1848) kirjassa Suomen Suuriruhtinanmaan Nykynen Tilasto: Yritös Alkeis- ja Rahwaan-kouluin tarpeeksi. Tikkanen johti tilaston sanasta *tila* esikuvanaan latinan valtiooon tai tilaan viittaava sana *status* (Luther 1993). On esitetty, että Tikkanen pyrki tila-johdoksella välttämään *statistik*-tyyppisiä sanoja, koska niillä olisi ollut vahvempi kytkös valtioon, joka silloisessa Suomessa viittasi Venäjän keisariin (Tilastokeskus 2021). Pian Tikkasen toimittamassa Suometar-lehdessä 9.6.1848 kerrottiin Tikkasen aikeista “tutkistella – – †*tilatieteellisiä*”, nykytermillä tilastotieteellisiä, asioita.

Von Becker käänsi Tikasta seuraten Gymnaasi- ja Koulu-järjestys-asetuksessa 1856 *statistik*-oppiaineen ensisijaisesti tilastoksi ja toissijaisesti †*statistikiksi* (Wænerberg 1856, Metsikkö 1951):

Euroopan ja erinomattain Suomen, ja Venäjän waltakunnan tilasto (†*statistikki*) ja Lainoppineiksi ja kameraali-wirkoihin aikowille täydellinen kurssi yhteisessä tilastossa. – – Waltakuntain tilasto eli †*statistikki*.

Tikkasen väitöskirja 1859 oli kolmas suomenkielinen (Pihkala 2000). Tikkanen muotoili siinä sanat tilastotiede, tilastollinen ja †*tilastontutkija* sekä kirjoitti tieteistä †*Suomen tilasto* ja †*Suomen tilastotiede* (Tikkanen 1859).

Pian Ferdinand Ahlmanin ruotsi-suomi-sanakirjassa (1865) oli termit *statistik*, tilasto-tiede, tilasto, *statistiker*, †*tilaston-tutkija*, †*tilaston-taitaja* ja *statistisk*, tilastollinen, tilasto-tieteellinen. Elias Lönnrotin suomi-ruotsi-sanakirjaan (1874) päättyi rivi “Tilasto – – (tila) *statistik* – Tilastollinen – – *statistisk*”.

Tikkanen myös loi tilastosta täysin eroavan sanan valtio (Itkonen 1990, Hakulinen 1979, Rapola 1943, Tikkanen 1847).¹ Matti Rapola (1943) arveli, että Tikkanen pyrki poimimaan sanaan †*waltiollinen* hallinnollisen merkityksen.

Tikkasen tapaan Tilastollisen Viraston esimies Karl Ferdinand Ignatius kirjoitti 1872 tilastotieteestä ja †*Suomen tilastotieteestä* (Ignatius toim. 1872). Niillä hän tarkoitti yhteiskunnan numeerista kuvaamista.

Tilastotiede ujuttautui 1800- ja 1900-lukujen vaihteessa nuorten mieliin Zachris Topeliuksen Maamme Kirjan kautta (Topelius 1876, 1878, 1905). Ensimmäisessä Johan Bäckvallin suomentamassa painoksessa 1876 kirjoitettiin tilastosta:

Kuinka maamme kasvaa varallisuudessa uuden ajan virkistävillä voimilla, sen näemme silmäimme edessä, ja sen sanoo nykyajan tilasto

¹Sana painettiin ensi kerran Rietrikki Polénin matkakertomuksessa mutta on silti Tikkasen luomus. Kaisa Häkkisen tiedonanto.

eli [†]valtiotieto. Niinkuin historia näyttää meille kuluneen ja jo loppuneen ajan, niin kirjoittaa tilasto nykyajan suhteet, jotka vielä ovat tekeillä, ja vertaa ne toistensa kanssa, tahi entisten aikain kanssa.

Seuraavassa painoksessa Paavo Cajanderin muokkaamassa suomennoksessa 1878 jälkimmäinen ja 17. Paavo Cajanderin suomentamassa painoksessa 1905 myös ensimmäinen tilasto-sana korvattiin tilastotiede-sanalla. Topelius kuvasi tilastotieteen valtion tiedonkeruuksi ja suomennoksissa sitä kutsuttiin [†]valtiotiedoksi tai valtiotieteeksi.

3.2 Luonnontieteellinen havainnointi

Tieteen kuuluisimpia prioriteettikiistoja on, keksikö Carl Friedrich Gauss (1777–1855) vai Adrien-Marie Legendre (1752–1833) pienimmän neliösumman menetelmän. Kiistattomasti Legendre julkaisi menetelmän ensimmäisenä ja sovelsi sitä maapallon muodon selvittämiseen 1805. Suomessa Gustaf Hällström (1755–1844) ja Henrik Walbeck (1793–1822) omaksuivat nopeasti menetelmän ja sovelsivat sitä menestyksekkäästi 1815 alkaen muun muassa maapallon litistyneisyyden arviointiin (Pere 2019, Pere ja Nyblom 2020). Walbeckin estimaattia maapallon litistyneisyydestä käytettiin kansainvälisesti, ja itse Gauss hyödynsi sitä (Mattila 2017). Cambridgen yliopiston tähtitieteen professori George Biddell Airy (1835) arvioi Walbeckin saavuttaneen “huomattavan etevyyden” (*considerable eminence*).

Ilmatiede

Ernst Biese suomensi saksan kielestä norjalaisen Henrik Mohnin oppikirjan Ilmatieteen pääpiirteet (1892). Kirjassa laajennetaan tilasto yhteiskuntaa kuvaavasta tieteestä yleisemmäksi numeeriseksi tieteeksi:

Klimatologia [*sic*] on ilmallisten alkeitten tilasto ([†]statistiikka), joka luvuilla ilmaisee keskimääräiset, monivuotisista havaintosarjoista johdetut – – arvot eri aikoina – – ja ajastaikaiset muutokset.

Vastaava laajennus tehtiin suurelle yleisölle kalenterissa 1890 (Kansanvalistus-Seura 1889):

Jonkun paikan ilmanalan selvittäminen yleensä on maan tawallista tilastoa, jota meidän aikana tehdään muistakin suhteista (wäen lu’usta, taloudesta j. n. e.). Tilaston tärkeys nykyisissä oloissa alkaa jo olla suureltakin yleisöltä tunnustettu. Sen aikaansaaminen luonnonilmiöistäkin on ilmatieteen tärkeimpiä tehtäviä.

Johan Eklöf murtaa jään

Johan Eklöf kirjoitti ensimmäisenä akateemista matematiikkaa suomeksi (Anonyymi 1860, Heikinheimo 1955). Suomen Tiedeseura kieltäytyi julkaisemasta Eklöfin suomenkielistä tutkimusta (Huumo 2005), joten Eklöf julkaisi sen Suomi-lehdessä (Eklöf 1850). Artikkelista tuli suomenkielisen tilastotieteen ja suomalaisen tieteen virstanpylväs.

Eklöfin artikkeli ja Erik Alexander Ingmanin lääketieteellinen artikkeli (1850) johtivat kiistaan, voiko suomea käyttää tieteen kielenä (Huumo 2005).² Ennen Eklöfin artikkelia ei “ollut käsitelty suomeksi yhtään puhtaasti tieteellistä aihetta” (Elfving 1938). Viimeksi mainittu pitäneek tulkita viittauksena tieteelliseen artikkeliin, sillä Eklöf oli jo aiemmin julkaissut akateemisen matemaattisen oppikirjasen Kolmiomittaviivain – Merkillisimmät Yhtisoudet ynnä Tasannes-Kolmiomitanto (1848). Kirjasessa oli lyhyt matematiikan suomi–ruotsi-sanasto.

Eklöf (1850) tutki, onko jäidenlähdön ajankohta muuttunut Kokemäenjoessa vuosina 1801–1849. Hän lainasi Suometarta 19.11.1849:³

Niille, jotka rakastavat [†]luvunlaskuja, liitämme tähän 49 vuotiset [†]katsomat niistä ajoista, milloin aina se osa Kokemäen virtaa, joka juoksee sivu kaupunkiamme, on luonut jäätänsä – – .

Ajankohtaa hän selitti yhtälöllä

$$x = p + s \times (v - 1800).$$

Selitettävä muuttuja x on päivien lukumäärä maaliskuun 1. päivästä alkaen jäidenlähtöön (esim. 2.4. $\rightarrow$ 33), p on vakiotermi ja s on regressiokerroin vuodelle $v - 1800$, joka saa arvot $1, \dots, 49$. Yhtälön piti päteä jokaiselle 49 vuodelle. Tunte mattomia parametreja oli kaksi, joten yhtälöryhmä ei ollut matemaattisesti ratkaistavissa. Eklöf kirjoitti:

Tästä näkyy meillä olevan 49 [†]yhtisoutta, vaan ainoastaan 2 [†]tiety-mätöntä; nämät pitää meiltä sentähden erinomaisella keinolla määrättämän. Ja juuri tähän tarvitsemme [†]Todenvaihe-laskua.

Eklöf ratkaisi parametrien arvot pienimmän neliösumman menetelmällä⁵, laski taulukon luvuista jäännösneliösumman (S) ja päätteli sen avulla tilastollisesti moderniin tapaan edelleen ajankohtaisen tuloksen kevään aikaisumisesta:

²Kiitämme Kaisa Häkkistä Eklöfin ja Ingmanin artikkelien painovuosien täsmentämisestä (ks. myös Rapola 1960).

³Teksti alla on Eklöfin tutkimuksesta ja eroaa hieman alkuperäisestä Suomettaressa.

⁴Kuva 3.1: Kansalliskirjaston digitoidut aineistot. <https://digi.kansalliskirjasto.fi/aikakausi/binding/498473> (haettu 21.2.2020).

⁵Eklöf oli huolellinen: R-ohjelmisto (R Core Team 2019) laskee täsmälleen Eklöfin raportoidut estimaatit.

Jäänlähtö-alaat Kokemäen virrassa

vuosina 1801—1849,

todenvaihe-laskulla määräsi

Johan Henrik Eklöf.

Suomen Tiede-Seuran toimituksissa löytyy jo ennen sen ensimmäisessä osassa jäänlähtö-ajaat Todenvaihe-laskulla määrättyt Herra Proffessori vainaja Hällströmiltä Turuun-, Porvoon- ja Pohjan-Kyrön joissa Suomessa sekä Nevajoessa Venäjällä että Maälärin järvessä Vesterääsin tienoilla Ruotsissa; ja kuin sanomalehden Suomettareen 19:ssä N:ossa tänä vuonna seuraavat vaarinotot ilmestyivät, emme milloinkaan epäilleet niitä Todenvaihe-laskun mukaan käytellä ja mitä sen kautta voisimme löytää ja ilmaaada mainitulle Seuralle nöyryydessä tarjoa, semmenkin kun tällaiset kokeet aina, miten voivat, näyttelevät paikkakuntain sekä säät että ilmanlaadut. Suomettareen sanat ovat seuraavat: "Niille, jotka rakastavat luvun-laskuja, liitämme tähän 49 vuotiset katsomat niistä ajoista, milloin aina se osa Kokemäen virtaa, joka juoksee silu kaupunkiamme, on luonut jäätänsä: 17 p. Maaliskuuta 1822; 2 p. Huhtikuuta 1832; 8 p. Huhtik. 1802, 1848; 9 p. Huhtik. 1836; 10 p. Huhtik. 1803; 13 p. Huhtik. 1815; 14 p. Huhtik. 1826, 1827, 1834; 15 p. Huhtik. 1813; 16 p. Huhtik. 1846; 17 p. Huhtik. 1842; 18 p. Huhtik. 1801, 1841; 19 p. Huhtik. 1824,

Kuva 3.1: Ensimmäinen suomenkielinen pienimmän neliösumman menetelmän sovellus (Eklöf 1850).⁴

Tästä nyt rinta rinnallen pannen sekä [†]vaarinotetun että laskulla määrätyn arvon ja niitten eron saamme seuraavat jäänlähtöajaat – – tällaisissa laskuissa on myös hyvä tietää kuinka vähäin rajain sisällä nämät arvot ovat saatu määrätyiksi. – – hakekaamme kuinka suuret – – [†]tapulassa [†]Eroilla nimitettyjen lukujen [†]Neljut ovat ja kutsukaamme niitten summan S – – josta [†]todenvaiheinen haitta jokaisen vuoden [†]vaarinotossa on $E = 0.6744897\sqrt{\frac{S}{47}} = \pm 7.3$ päivää, jotka niin muodoin määräävät minkä rajain sisällä kaikkien vuotten [†]vaarinotot näyttää itsensä pitävän; niin että jos jonkun vuoden jäänlähtö tapahtuu varhaimmin eli myöhemmin kuin 7.3 päivää mitä [†]yhtisou-temme $x = 56.311 - 0.0851(v - 1800)$ määrää on se vuosi luettava hairahtavaksi. – – jäänlähdon näin aivoin olevan $4\frac{1}{4}$ päivää varhemman kun 50 vuotta ennen, jossa määräyksessä toki taitaa olla puolen välin [†]haitta, eli erehdys, joko hän on varhemmalle eli myöhemmälle puolellen.

[†]Todenvaiheisella haitalla Eklöf tarkoitti Friedrich Besselin nimeämää suuretta *wahrscheinlicher Fehler* (Bessel 1816). Tätä suuretta, joka sittemmin on suomennettu todennäköiseksi virheeksi, käytettiin yleisesti vielä 1900-luvun alkupuolella kuvaamaan estimoinnin satunnaisvirhettä (esim. Student 1908), kunnes keskivirhe korvasi sen.

Eklöfin sepittämät ja käyttämät tilastotieteelliset ja matemaattiset termit ovat huomionarvoisia: [†]ero = erotus, [†]haitta = virhe, [†]luvunlasku =

tilasto, †nelju = neliö, †tapula = taulukko (Eklöf 1848: †tapula = *tabell*), †tietymätön = tuntematon, †todenvaihe = todennäköisyys, †todenvaiheinen haitta = todennäköinen virhe, †yhtisous eli †yht'isous (Eklöf 1856) = yhtälö, †vaarinotettu arvo = havaittu arvo ja †vaarinotto = havainto.

†Yhtisous päättyi melkein heti David Emanuel Daniel Europaeuksen ruotsi-suomi-sanakirjaan (1852–1853) muodossa †yhtä-isous (*equation*). †Katsomilla Eklöf viittasi päivämääriin, mutta merkitys lienee yleisempi kuten kirjaukset tai tiedot.

†Vaiheen tulkinnassa artikkelin ydinkohdat ovat (Eklöf 1850):

– – tahtoisimme tietää minä päivänä †todenvaiheellisemmasti virta jäänsä vuoteisesti purkaa – – . – – †todenvaiheinen haitta $x:n$ arvossa on $- \pm 1.04$, joka näyttää $x:n$ olevan tällä laskulla määrätyn noin päivän †vaiheella.

Lainauksen yllä ensimmäisessä virkkeessä †todenvaihe saattaisi olla “todenvälisyyttä” tai “todenläheisyyttä”.⁶ Jälkimmäisessä virkkeessä Eklöf viittaa †vaiheella tarkkuuteen tai varmuuteen. †Todenvaihe voisi olla myös “todentarkkuutta” tai “todenvarmuutta”.

Eklöf ei suomentanut †todenvaihetta ja †todenvaiheista haittaa välttämättä saksan kielestä. Hällström (1823) oli aiemmin soveltanut todennäköistä virhettä kirjoittaen *sannolika fel*, *sannolik* oli muutenkin käytetty termi ja Eklöfin artikkeli juonsi Hällströmin tutkimuksista.

†Tapula oli painettu aiemmin edellä osoitetuissa Suomettaressa 9.11. 1847 ja Eklöfin kirjasessa (1848). Eklöf (1852, 1856) siirtyi pian kirjoittamaan taulu †tapulan sijaan. Taulukko tuli kirjakieleen 1903 Maamies-lehdessä (Häkkinen 2020).

†Nelju oli Schildtin (Saima Kallavesi 9.10.1845) suomennos. Rapolan mukaan Kilpinen (1847) olisi †neljun viime-esiintymä (Kotimaisten kielten keskus 2019), mutta se lienee siis Eklöfin artikkelissa. Lönnrotin 1845 ehdottama neliö (Saima Kallavesi 9.10.1845, Salminen 1906, Rapola 1960, Kotimaisten kielten keskus 2019) vakiintui sittemmin.

Tilastotieteessä tarvittavia trigonometrisia funktioita Eklöf ei tietoisesti suomentanut (1848):

Ett' en ole suomentanut nimiä Sinus, Cosinus j. n. e., ei mahda kellekään kummalta näkyä, koska samat sanat jokaisessa minulle tutussa kielessä aina merkitään samoin. Ainoastaan Saksalaiset ovat koittaneet heitä kielellensä kääntää, esim. Sinus = Träger j. n. e., mutta ei siinä onnistuneet.

Eklöfin termejä ei enää käytetä, mutta ne olivat aikanaan uraauurtavia. Eklöfin käännösten omakielisyys korostuu, jos niitä vertaa muutamaa

⁶†Vaihe on vanha sana, jolla on monesti viitattu väliin (Häkkinen 2020). “Todenläheisyys” on Kaisa Häkkinen ehdotus.

vuotta aiempaan Gymnaasi- ja Koulu-järjestys -asetukseen 1843. Siinä kehoitettiin opettamaan gymnaaseissa “trigonometria plana’a” ja “toisenkin graadin equationeja selittäin” (Westzynthius 1844).

Yhtälö-sana painettiin 1865 (Rapola 1960), mutta sen vakiintuminen vei pitkään. Sakari Levänen oli ensimmäinen suomeksi luennoinut matemaatikko (Elfving 1985). Vielä muistiinpanoissaan 1885–1893 hän kysyi: “Minkä ekvationin juuret ovat – – ?” (Kansallisarkiston Sakari Leväsen II arkisto). Aleksanterin yliopistossa luennoitiin yhä 1890 ja 1894 “differentiaali-ekvationien” teoriasta. “Differentiaaliyhtälöiden” teoriasta luennoitiin 1895. (Aleksanterin yliopisto 1890, 1894, 1895.)

Eklöf on suomenkielisen tilastotieteen edelläkävijä – tilastotieteen Aleksis Kivi. Tai pikemminkin Kivi on kaunokirjallisuuden Eklöf, sillä Eklöfin suomenkieliset julkaisut edelsivät Kiven suomenkielisiä teoksia. Kiven tapaan Eklöf kuoli nuorena, 35-vuotiaana, tiettävästi kuluttaviin elämäntapoihin (Suomen historiallinen seura 1883).

3.3 Tilastotieteen teoria, oppikirjat ja oppaat

Matemaattinen tilastotiede kulkeutui vähitellen Suomeen. Jarl Waldemar Lindebergin 1922 johtama välttämätön ja riittävä ehto Keskeiselle raja-arvauseelle on matemaattisen tilastotieteen suomalainen merkkipaalu (Paulauskas 1999). Suomenkielisiä termejä ja oppikirjoja tarvittiin vakiinnuttamaan uudet käsitteet.

Matemaattinen tilastotiede

Suomen Tilastoseura perustettiin 4.10.1920. Seura pyysi Helsingin yliopiston matematiikan apulaista Jarl Waldemar Lindebergiä laatimaan [†]matemaattista tilastoa käsittelevän oppikirjan ja muita tieteellisiä yhdistyksiä mukaan hankkeeseen. Hän teki kirjan (Lindeberg 1927) ja esitelmöi siitä seurassa. (Tudeer 1946.)

Kirjan, Todennäköisyyslasku ja sen käytäntö tilastotieteessä – Alkeellinen esitys, sanasto elää osin edelleen. Lindeberg esitteli matemaattisen tilastotieteen jo esipuheessa. Hän käytti tilastoa ja tilastollista aineistoa aineiston synonyymeinä. Havainto, keskiarvo, keskivirhe, regressiokerroin, regressiosuora ja systemaattinen virhe ovat kirjan vakiintuneita termejä kuten myös tunnusluku ja hajonta. Viimeksi mainittujen synonyymeistä [†]karakteristika ei käytetä ja [†]dispersiota käytetään harvoin. Ne juontanevat saksan tai ranskan kielistä, joilla Lindeberg julkaisi tutkimuksiaan. Näyte on edelleen ymmärrettävä kuten [†]umpimähkäinen näytekin, mutta jälkimmäisen on syrjäyttänyt täsmällisempi yksinkertainen satunnaisotos. [†]Keskimomentti ja [†]korrelatiokerroin ovat lähes sellaisinaan edelleen käytössä, mutta Lindebergin itsenäisesti kehittämä tunnusluku [†]korrelatioprosentti ja sen määrittä-

vät †positiiviset ja †negatiiviset parit (konkordantit ja diskordantit parit) eivät ole. †Frekvenssiluku eli †lukuisuusluku, †jakaantuminen, †jakaantumistaulu, †keskimääräinen poikkeus ja †punnittu keskiarvo ovat ymmärrettäviä mutteiivät käytettyjä. †Todennäköisyyslaki on vanhentunut nimitys jakaumalle. †Gaussin laki, †normaalinen laki ja †normaalikäyrä ovat niinkään normaali-jakauman vanhentuneita nimityksiä. †Yli- ja †alinormaalinen hajonta eivät ole myöskään jääneet elämään. †Todennäköisestä arvosta on (syystä!) siirrytty käyttämään englanninkielisen termin lähes kirjaimellista käännöstä odotusarvo. Satunnaissuure on harvinaisempi mutta edelleen oiva termi ja löytyy käsillä olevasta sanastosta.

Väestö-, vakuutus-, metsä-, yhteiskunta- ja kauppatieteet 1900-luvun alkupuoliskolla

Aleksanterin yliopiston tilastotieteen dosentti Edvard Gyllingin ja Jaakko Forsmanin Tilasto- ja Tilastotiede-artikkelit Tietosanakirja-osakeyhtiön Tietosanakirjassa edelsivät oppikirjoja (Gylling 1917, Gylling ja Forsman 1917). Gylling (mt.) osoitti †statistiikan tilaston vaihtoehtoiseksi nimitykseksi. Hän kertoi, että havainnoilla luodaan kuva “joukkoilmiön tilasta” ja että sitä varten on “toimitettava lasku” kuten †väenlasku tai †viljavarojenlasku “aika-ajoin”. Gylling kutsui otantaa †edustavaksi tilastolliseksi havaintomenetelmäksi ja aineistoa †ainehistoksi. Jälkimmäisessä artikkelissa Gylling ja Forsman (mt.) merkitsivät †statistiikan myös tilastotieteen synonyymiksi, osoittivat †statistiikalle kolmannen aiemman merkityksen †valtiotiedon sekä nimesivät ehkä ensimmäisinä ammatin tilastotieteilijä.

Ernst Bonsdorffin kirjassa Vakuutustieteen matemaattiset alkeet (1919) oli suomalaisittain uudenlaista tilastotieteen soveltamista. Havainto, keskiarvo, keskivirhe, †pienimmän neliösumman laskutapa, suurten lukujen laki, tasoittaminen, tasoitus, tilastotiede (nykymerkityksessä) ja todennäköisyys ovat nykylukijan tuttuja. Poikkeamiskerroin ja todennäköinen virhe ovat nekin asianmukaisia termejä joskin vähän käytettyjä. Hylättyjä suomennoksia ovat †dispersio-oppi eli †hajaantumisoppi, †divergenssikerroin, †poikkeusten neliösumma, †tilastollinen aines, †tode[n]mukainen virhe (myös todennäköinen virhe), †todenmukaisin arvo ja †virheoppi.

Vilho Pesola, tuleva Maatalouden tutkimuskeskuksen professori, opasti Kenttäkoeoppaassa (1924) tilastotieteen käyttöön, tulevan metsäekonomin professorin Eino Saaren avustuksella. Oppaassa neuvottiin laskemaan keskiarvo, sen †keskimääräinen virhe, †variatio-koeffisientti eli †vaihtelukerroin, todennäköinen virhe ja havainnon keskivirhe eli †dispersio. Termien yhteydessä Pesola esitti niiden saksankieliset vastineet.

Väinö Fritiof Johanson käytti Lindebergin termejä mutta esitti myös omia luentomonisteissaan Tilastotieteen perusteet ja pääkohtia Suomen tilastosta (1935). †Vuorosuhdeluku (korrelaatiokerroin), †tihein normaaliar-

vo (moodi) ja [†]keskeinen arvo (mediaani) ovat omintakeisia suomennoksia (mt.). Johanson ei selitä, miksi hän ei tyytynyt Lindebergin suomennokseen [†]korrelatiokerroin.

Johansonin kirja (1940) Suomen tilaston pääkohdat perustui monisteeseensa. Nimi viittaa kirjassa selostettaviin virallisiin tilastoihin. III luku on “Tilaston teoriasta”, jossa Johanson selittää tilastotieteellisiä käsitteitä. Tilastolla oli kaksi merkitystä: viralliset tilastot tai tilastotiede nykymerkityksessä. Johanson kirjoitti monisteessaan samaan tapaan tilastosta molemmissa merkityksissä. Kirjan nimestä hän pudotti tuntemattomasta syystä tilastotieteen pois. Johanson kutsui moodia nyt [†]yleisimmäksi arvoksi eli [†]tavallisimmaksi arvoksi eli [†]modiksi eli (Harmajan 1939 mukaisesti) [†]ordinaariarvoksi. [†]Frekvenssilukuja hän nimitti nyt [†]taajuusluvuiksi. [†]Painollinen keskiarvo ja [†]liikkuva keskiarvo ovat selkeitä suomennoksia. Korrelaatio ja korrelatiokerroin ensiesiintyvät kirjassa.

Leo Harmajakaan ei tyytynyt Lindebergin [†]korrelatiokertoimeen. Hän kirjoitti oppikirjassaan (1939) [†]korrelaatioluvusta. (Törnqvist 1946 nimesi [†]korrelaatioluvun eri merkityksessä.) Harmaja esitteli ensimmäisenä termit moodi ja mediaani. Jälkimmäiselle hän esitti rinnakkaissuomennokset [†]keskusluku, [†]sentraaliarvo ja [†]ordinaariarvo. Myös standardipoikkeama sai ensiiltansa. Sille Harmaja osoitti synonyymit [†]hajaanto ja Lindebergin hajonta. Viimeksi mainitun Harmaja totesi jo aiemmin esitetyksi muttei pitänyt sitä kielellisesti “täysin hyväksyttävänä”.

[†]Keskiarvo ja keskiluku olivat synonyymejä Johansonille ja Harmajalle. Molemmissa kirjoissa on estimointivirhettä käsittelevä alaluku [†]virheteoria. (Johanson 1940, Harmaja 1939.)

Helsingin kauppakorkeakoulun liiketaloustieteen professori Ilmari Kaitila sovelsi tilastotiedettä tarmokkaasti kirjassaan Liiketaloustilasto (1945). Aikasarjat kiinnostivat Kaitilaa. Kaitilan muut termit kuten [†]korrelaatiokoeffikientti, [†]trend, [†]useuskäyrä tai [†]varmuusaste ovat unohtuneita. Kirja painottui kuvailuun ja estimointiin. Niihin liittyen Kaitila pohti otantaa ja syy-seuraussuhteita. Kaitilalle moodi oli [†]tavallisin eli [†]tihein arvo ja mediaani oli [†]keskusarvo. (Aikalaisen Paul Bremerin 1945 suomennos oli [†]keskeisarvo.) Kirjasta löytyy normaalijakauman yleiskielinen nimitys kellokäyrä. Kirja perustui valtaosin saksalaisiin lähteisiin.

Johansonin (1935, 1940) [†]edustava metodi, Harmajan (1939) [†]edustava menetelmä ja Kaitilan (1945) [†]edustava tilastometodi ovat joissain muodoissa jääneet kummittelemaan soveltajien teksteissä.

3.4 Tilastotiede vakiintuu itsenäiseksi oppiaineeksi

Ensimmäinen tilastotieteen laitos perustettiin University Collegeen Lontooseen 1911. Suomessa ensimmäinen tilastotieteen professuuri perustettiin 1945 Helsingin yliopiston valtiotieteelliseen tiedekuntaan. Matemaat-

tista tilastotiedettä opetettiin Helsingin yliopistossa myös matematiikan oppiaineessa.

Gustav Elfving

Gustav Elfvingistä tuli matematiikan dosentti 1937 ja professori 1948 Helsingin yliopistoon (Draper ym. 1999). Elfvingin suomennokset näkynevät jo Kaarlo Pihkalan, tulevan maatalouspolitiikan professorin, väitöskirjassa (Pihkala 1941). Pihkala kiitti Elfvingiä kaavojen johtamisesta ja matemaattisia menetelmiä sisältävien kohtien tarkastamisesta. Uudelta suomennokselta vaikuttaa satunnaisvaihtelu. Pihkala kirjoittaa Bonsdorffia ja Lindebergiä seuraten tutkimusaineistosta, havainnoista ja systemaattisesta virheestä. Keskiarvo oli Pihkalalle keskiluku Johansonin ja Harmajan tapaan; myös keskipoikkeaman merkitys erosi nykyisestä (mt.). [†]Havaintoluku oli vanhahtava. [†]Vuorosuhdekerroin (mt.) oli yhdistelmä Johansonin (1935) [†]vuorosuhdelukua ja Lindebergin (1927) [†]korrelatiokerrointa.⁷

Elfvingin opetusmoniste Todennäköisyyslaskenta (1946) on nimeä myöten edelleen pätevä. Monisteessa analysoitiin tilastotieteellisesti teoreettisia aineistoja, vaikka vain yhden alaluvun otsikko on “Matemaattisen tilastotieteen alkeet”.

Havainto oli analyysien keskiössä. Elfving kutsui jakaumaa [†]jakautumaksi. Elfving ja akateemikko Rolf Nevanlinna olivat päättäneet [†]jakautuma-suomennoksesta kävelyllä Erottajalta Helsingin yliopistolle.⁸ Termi ei sisälly enää tähän sanastoon. Lindebergiä mukaillen Elfving kirjoitti [†]Gaussin jakautumasta ja [†]normaalisesta jakautumasta.

Elfving piti keskiarvoa, [†]todennäköistä arvoa ja odotusarvoa synonyymeinä. Merkittävä lienee kirjoitettu ensi kertaa monisteessa. [†]Hajontaneliöstä, [†]näyteteoriasta tai [†]presumtiiviarvosta (estimaatti) ei enää puhuta. [†]Stokastinen keskiarvo (odotusarvo) antoi ehkä väärän mielikuvan. Emäfunktio, tapahtuma ja todennäköisyystiheys ovat monisteen uudissanoja. [†]Maksimalisointiperiaate (suurimman uskottavuuden menetelmä) ja [†]konfidenssiväli kertovat Elfvingin tieteellisestä edistyneisyydestä. Moniste päättyy lyhyeen ruotsi-suomi-sanastoon. Moniste johti kirjaan Todennäköisyyslaskenta (Elfving 1956). Se ja sen suomi-ruotsi-englanti-sanasto epäilemättä vakiinnuttivat termejä.

Leo Törnqvist

Leo Törnqvist valittiin valtiotieteellisen tiedekunnan tilastotieteen professorin virkaan 1950 (Nordberg 1999). Tarve oli erityisesti teoreettisimpien

⁷Väitöskirja on ensimmäinen ekonometrian alaan kuuluva, vaikka tätä termiä ei väitöskirjassa ole.

⁸Timo Mäkeläisen tiedonanto.

käsitteiden suomentamiselle englannin kielestä. Törnqvist suomensi monia termejä kekseliäästi: harha, ryväsotanta, tarkentuvuus (*consistency*), virhe ja tyhjentävä tunnusluku (*sufficient statistic*).⁹ Tarkentuvuus ja tyhjentävä tunnusluku ovat englanninkielisiä vastineitaan osuvampia viestien suomentajansa syvällistä tilastotieteen ymmärrystä.

Törnqvistin opetusmoniste Tilastollisten tunnuslukujen laskukaavoja – Tilastotieteen teorian peruskurssi (1946) oli tarkoitettu jatkeeksi todennäköisyyslaskennan kurssille. †Empiirinen summafunktio eli †empiirinen jakautumafunktio ja †frekvenssitiheysfunktio ja †eksessi (normaalijakaumasta poikkeava kurtoosi) ovat jääneet käytöstä. Törnqvist erotti keskiarvon ja odotusarvon ja käytti molempia termejä nykyisissä merkityksissä. Varianssi on monisteen oleellinen uutuus.

Kokonainen sukupolvi tilastotieteilijöitä sai oppinsa Törnqvistiltä (Nordberg 1999). Tätäkin kautta Törnqvist on vaikuttanut tilastotieteen suomenkieliseen termistöön.

3.5 Väestötiede, psykometriikka, yhteiskuntatieteet ja laadunvalvonta

Harmajan oppilaan Väinö Kanniston kilpailevien kuolemansyiden teoriaa koskevaa väitöskirjaa (1947) on pidetty ensimmäisenä tilastotieteen väitöskirjana tilastotieteen professuurin perustamisen 1945 jälkeen (Niitamo 1991). Tilastollisesti merkitsevä ensiesiintynee väitöskirjassa. Kanniston †vapaa-aste on melkein nykyinen vapausaste.

Helsingin yliopiston psykometriikan dosentiksi 1950 nimitetyn Toivo Vahervuon Johdatus matemaattis-tilastollisiin menettelytapoihin (Vahervuo 1946) oli kattava tilastotieteen esitys. Järjestyskorrelaatiokerroin, kvartiilipoikkeama, osittaiskorrelaatio, osittaisregressiokerroin, pienimmän neliösumman menetelmä, suhteellinen frekvenssi ja vinous ovat kirjan edelleen päteviä termejä. Termeistä moni on toisaalta hylätty kuten †jakaantumiskäyrä, †normaalisuusalue (hyväksymisalue) tai †tilastollinen varmuus. Vahervuolle mediaani oli myös †keskeisarvo; moodia hän nimitti †todennäköisimmäksi arvoksi tai †valta-arvoksi. Hajonta, †dispersio, †standardipoikkeama ja †deviaatio olivat myös synonyymejä. Vahervuo kutsui †tilaston tunnusluvuiksi “lukuja, joista tilaston pääominaisuudet selviävät”. Vahervuolle tilasto oli aineiston synonyymi. Termistö oli vielä kirjavaa.

Sanastoa lienee muokannut myös Törnqvistin assistenttien Rolf Alameren ja Pentti Pöyhösen yhteiskuntatieteilijöille suunnattu oppikirja Johdatus tilastolliseen tutkimukseen (1953). He esittelivät tilastollisen päätelyn käsitteet †arviovälin, †konfidenssivälin, luottamusvälin (synonyymejä), jaottelun †melkein merkitsevä, merkitsevä ja †erittäin merkitsevä sekä termin

⁹Matti Hakaman tiedonanto.

vapausaste. Törnqvist on vaikuttanut kirjaan, sillä Alameri ja Pöyhönen (mt.) kiittävät häntä huomioista.¹⁰

Olli Lokki nimitettiin sovelletun matematiikan professoriksi 1962 Teknilliseen korkeakouluun (nykyisin osa Aalto-yliopistoa). Lokin opetusmonisteet Tilastollisen laaduntarkkailun perusteet ja Tilastomatematiikan perusteet (1956, 1959) tutustuttivat insinöörejä tilastotieteeseen. Tilastomatematiikka on lukiossa ja insinööritieteissä edelleen käytetty termi. Sanastossa se on matemaattinen tilastotiede. Lokki (1956) mainitsi Elfvingin (1956) oppikirjan lähteenään ja käytti monia sen termejä. Lokki suomen-si erityisesti laadunvalvonnan ja [†]laaduntarkkailun termejä: valvontakortti, ala- ja ylävalvontaraja, kaksinäytetarkastus, peräkkäisnäytetarkastus eli sekvenssianalyysi ja todennäköisyysmassa. Hiipuneita suomennoksia ovat [†]havaintomateriaali, [†]summafunktio eli [†]kumulatiivinen todennäköisyysfunktio ja [†]signifikantti. (Lokki 1956.)

3.6 Tilastotieteen sanastot

Monet Pohjoismainen tilastosanasto -teoksen ensimmäisen painoksen (Suomen Tilastoseura 1954) termit ovat edelleen käytössä kuten harhaton, kertymäfunktio, odotusarvo, suurimman uskottavuuden menetelmä, tarkentuvuus, tyhjentävä, uskottavuus ja vaihteluväli. Raakalainoja ovat estimaatti, estimaattori, estimoida, mediaani, momentti, regressio, testi, varianssi ja varianssianalyysi. *Sample* käännettiin näytteeksi tai otokseksi. [†]Hylkäämis- ja [†]hyväksymiserhe sekä [†]kriittinen alue eivät asettuneet tilastotieteen sanastoon. Monet termeistä löytyvät Elfvingin (1956) oppikirjasta. Sanastossa on Kaitilan (1945) [†]keskusarvo. Se on poistunut käytöstä, mikä on ehkä menetys. Kolmikko keski-, [†]keskus- ja tyyppi-arvo olisi iskevä vastine kolmikolle mean, median ja mode. Vahervuon [†]valta-arvo (tyyppi-arvo) on jäänyt käytöstä — ehkä liian värikkäänä ilmaisuna. Syrjään jääneitä ovat myös [†]jakautuma, [†]konfidenssiväli ja [†]varmuusväli (luottamusväli) sekä [†]täystehoinen (tehokas). Osa suomennoksista ei ole yleistynyt kenties siksi, ettei niistä voi päätellä kantasanaa: [†]paikallistamisluku (*average*), [†]lukuisuus (*frequency*), [†]bruttovarianssi (*mean square error*) ja [†]varianssijuuri (*standard deviation*). Viimeksi mainitun vaihtoehto keskihajonta on kiinnittynyt tilastotieteen ydinsanastoon. *Observation* ja *data* eivät yllättäen ole hakusanoja. Niitä ei ole ehkä mielletty termeiksi. Sanaston saavutus oli keskeisten teoreettisen käsitteiden suomentaminen ja kokoaminen.

Monikielisen väestötieteen sanakirjan (Fougstedt ym. 1962) hakemistosta löytyy havainto, joista leipätekstissä kerrotaan [†]perusaineiston koostuvan. Näyte kuvataan yleisnimityksenä; otos rajataan moderniin tapaan

¹⁰Pöyhönen väitteli ekonometrisella tutkimuksella (Pöyhönen 1955). Hänestä tuli 1958 ekonometrian dosentti Helsingin kauppakorkeakouluun ja kansantaloustieteen professori Helsingin yliopistoon 1959 (Sullström 2000).

†todennäköisyysvalintaan eli todennäköisyysotantaan. †Varmuusvälin ja †konfidenssivälin rinnalla on jo Alameren ja Pöyhösen (1953) kirjaama luottamusväli. Vanhentuneita termejä ovat muun muassa †tabulointi ja †tilastollinen käyttö.

Georg Lutherin johdolla laaditun Pohjoismainen tilastosanaston toinen painos (Suomen Tilastoseura 1975) on nykytilastotieteilijälle enimmäkseen tuttua. Esimerkiksi luottamusvälin rinnalla ei ole enää vaihtoehtoisia suomennoksia. Termejä haettiin (mt.) muun muassa Kendallin ja Bucklandin (1971) sanastosta. Monikielisen väestötieteen sanakirjan tapaan *sample* suomennetaan yleispätevästi näytteeksi mutta otokseksi, jos näyte on saatu todennäköisyysotannalla. Suomentamisessa hyödynnettiin Elfvingin todennäköisyyslaskennan opetusmonistetta¹¹, mutta myös uudissanoja luotiin (Suomen tilastoseura 1975). Piste- ja väliestimointi ovat keskeisiä uutuuksia. Sanaston termit kuvaavat yhä teoreettisempia käsitteitä. Monte Carlo -menetelmä ja linkkuveitsi ennakoivat tietokoneiden mahdollistamaa mullistusta tilastotieteessä. Syy- ja välinemuuttuja enteilevät syy-seuraussuhteiden mallittamisen valtavirtaistumista. †Keinomuuttujasta (välinemuuttuja) ei tullut pysyvää termiä, mahdollisesti koska sillä on muu matemaattinen merkitys. †Yksilöparametri (tilanneparametri) on hankala suomennettava, jota edelleen käännellään.

Suomen Tilastoseuran (1954, 1975) sanastot tehtiin yhteistyössä muiden Pohjoismaiden tilastoseurojen kanssa. Kansainvälisen tilastoinstituutin monikielisen tilastotieteen sanakirjan (*International Statistical Institute* 2011) suomenkielinen osuus toteutettiin 2004 Kari Djerfin johdolla. Sen jälkeen on julkaistu keskeisen sovellusalan, epidemiologian, englanti-suomi-englanti-sanasto (Läärä ym. 2008). Muita tilastotieteen sanastoja on laadittu opetustarpeisiin. Hyviä esimerkkejä ovat Petri Koistisen (2013) ja Erkki Pahkisen (2012) sanastot. Ne ovat kaikki olleet tarpeellisia ja ohjanneet tilastotieteen käyttäjiä täsmälliseen ja ymmärrettävään ilmaisuun.

¹¹Seppo Honkapohjan tiedonanto.

4 Lähivertailuja viron tilastosanaston kanssa

4.1 Suomi, viro ja tilastotieteen termit

Tilastosanaston kehittämisen yksi keskeinen haaste juontuu siitä, että suomi poikkeaa rakenteeltaan monella tavalla englannista, joka kuuluu indo-eurooppalaiseen kieliperheeseen. Siksi on valaisevaa tarkastella, millaisia ratkaisuja on tehty suomalais-ugrilaisiin kieliin kuuluvassa virossa. Suomen ja unkarin ohella viro on ainoa tämän ryhmän kieli, jolla annetaan korkeampaa opetusta tilastotieteessä. Parhaimmillaan naapurikielessä käytössä olevat termit antavat virikkeitä oman sanastomme uudistustyöhön.

Suomi ja viro ovat kehittyneet yhteisestä itämerensuomalaisten kielten kantamuodosta, jota puhuttiin Suomenlahden molemmin puolin ennen ajanlaskumme alkua. Suomi on arkaaisempi ja on näihin päiviin asti säilyttänyt enemmän kantakielen piirteitä. Nykyisen Viron alueella puhutut kielimuodot alkoivat eriytyä noin 1000 vuotta sitten. Sanat ovat lyhentyneet ja eräät kieliopin rakenteet ovat yksinkertaistuneet. Kumpaankin kieleen ovat vaikuttaneet entisten hallitsijoiden kielet: ruotsi, saksa ja venäjä. Eri alojen kreikasta ja latinasta peräisin olevat vierassanat ovat suodattuneet suomeen enimmäkseen ruotsin kautta ja viroon saksan kautta.

Vironkielistä termistöä on viime aikoina dokumentoitu useassa lähteessä. *Statistikasõnastik* (Tooding, 1997) sisältää viron termien rinnalla niiden vastineet paitsi englanniksi niin myös saksaksi ja venäjäksi. *Eesti Keele Instituutin* verkkosivuilta löytyvien erikoisalakohthaisten sanastojen joukossa on myös *Andmeanalüüsi ja statistika oskussõnastik* (2020). Viron kieli on mukana myös ISI:n monikielisessä sanastossa (*International Statistical Institute* 2011). Tilastotieteen selittävä sanakirja *Statistikaleksikon* (Tiit ja Tooding 2019) sisältää vironkielisten termien selitykset samoin kuin vastineet englanniksi ja venäjäksi. Termien lukumäärä on siinä suppeampi kuin em. muissa lähteissä samoin kuin tässä englanti–suomi–englanti-sanastossa, mutta termien selitykset ovat paikoin hyvin yksityiskohtaisia.

Virolaiset kollegamme näyttävät tähän asti olleen meitä aktiivisempia omakielisen tilastosanaston systemaattisessa päivittämisessä. Toisaalta vaikuttaa siltä, että edellä mainitut lähteet kattavat lähinnä vain perinteisen tilastotieteen termistön. Niistä puuttuu tilastotieteen ja sen uusien lähialojen kuten koneoppiminen (*machine learning*) yhteiselle leikkausalueelle kuuluvia termejä. Koneoppimista toki harjoitetaan myös Virossa nimen *masinõppimine* alla, ehkä enimmäkseen tietotekniikan ja tietojenkäsittelytieteiden piirissä.

Tilastosanastojen systemaattinen vertailu suomen ja viron välillä an-

saitsisi perusteellisemmän kielitieteellisen tutkimuksen. Tässä yhteydessä tuomme lyhyesti esiin vain joitakin yksittäisiä havaintoja yhtäläisyyksistä ja eroista.

4.2 Perustermit: *statistika* ja *andmed*

Kuten luvussa 3 jo tuotiin esille, termille *statistics* tarjosi Paavo Tikkanen (1848, 1859) edelleen pätevät suomennokset *tilasto* ja *tilastotiede*. Virossa puolestaan on vakiintunut lainasana *statistika*, joka viittaa sekä lukutaulukkoon että tieteenalaan. Kuitenkin 1800-luvun lopulla lukuisia uudis-sanoja luoneen Ado Grenzsteinin sanakirjaan *Eesti sõnaraamat (sisaldab 1600 uut ja vöörsõna)* (Grenzstein 1884) sisältyi *tilastoa* tarkoittava oma-kielinen termi [†]*arwustik*. Se on johdettu kantasanasta [†]*arw*, nykyiseltä kirjoitusasultaan *arv*, suomeksi ”luku”, joten suora suomennos olisi ”lurvusto” (vrt. suomennos [†]*luwunlasku* alaluvussa 3.1). Tämä sana samoin kuin [†]*arwustikuteadus* eli ”lurvustotiede” esiintyivät vironkielisissä teksteissä yleisesti sanan *statistika* ohella 1900-luvun alussa, jälkimmäinen paljolti samassa hengessä kuin *tilastotiede* Topeliuksen Maamme Kirjassa (ks. luku 3). Niiden käyttö on sittemmin vähentynyt. Viron itsenäistyttyä ja viron tultua Tarton yliopiston opetuskieleksi 1919 tilastotieteen opetuksessa päätermiksi omaksuttiin *statistika*. Vuonna 1921 perustettu Viron Valtion Tilastollinen Keskusvirasto sai nimekseen *Riigi Statistika Keskbüroo*. Siitä huolimatta sana [†]*arvustik* on esiintynyt harvakseltaan vielä 2000-luvulla, ja yllättäen vironkielisessä Wikipedia-artikkelissa *Statistika* otsikkotermin synonyymina mainitaan edelleenkin [†]*arvustikuteadus*.

Termin *data*, joka on monikko latinan sanasta *datum* eli ”annettu”, viroksi luovasti uudissanalla *andmed* jo 1900-luvun alussa viron sanaston merkittävä kehittäjä Johannes Voldemar Veski. Sana on johdettu verbistä *andma*, suomeksi ”antaa”. Latinan ja englannin esikuvansa tapaan *andmed* on monikollinen, eikä sillä itse asiassa ole mitään yksikkömuotoa! Lähisanoja ovat *and*, monikko *annid*, suomeksi ”anti, antimet”, sekä *anne*, monikossa *anded* eli ”lahja(kkuus), kyky”. Yhdyssanojen muodostuksesta olkoon esimerkkinä *andmeanalüüs*, joka on termin *data analysis* vastine; suomeksi *data-analyysi*.

Huomattakoon, että sanaa ”anne” tai ”anteet” esitettiin termin *data* suomennokseksi *Virittäjän* 1971 järjestämässä uudissanakilpailussa. Sanat ”tieto” ja ”tiedot” vakiintuivat kuitenkin kieleemme samalla, kun *computer* eli ”laskija” suomennettiin ”tietokoneeksi”. Tämänkin termin virolaiset käänsivät osuvasti sanalla *arvuti*, suomeksi ”laskin”.

4.3 Sanastojen kontrastiivinen nelikenttä

Jäsennämme sanastojen vertailua yksinkertaisen 2×2 -ristitaulukon avulla. Siinä suomen ja viron samaa käsitettä vastaavat termit luokitellaan sen mukaan, onko niille valittu ”puhtaasti” omakielinen vastine vai vierassana. Jälkimmäinen on tyypillisesti enemmän tai vähemmän suora laina alkukielen (englannin) termistä. Taulukossa kullekin termille löytyy paikka nelikentässä, jonka jokaiseen soluun on poimittu neljä esimerkkitermiä, joista kustakin esitetään alekkain englannin-, suomen- ja vironkielinen muoto.¹

Tällainen nelikenttä esiintyy kaltaistettujen parien (*matched pairs*) tilastollisissa analyysissä. Nelikentän päälävistäjältä löytyvät samanlaiset parit (*concordant pairs*) ja lävistäjän ulkopuolen soluissa ovat erilaiset parit (*discordant pairs*). Kontrastiiviselta kannalta kiinnostavaa olisi erityisesti selvittää erilaisten parien lukumäärien suhde suomen ja viron välillä.

Viro	Suomi			
	Oma		Vieras	
Oma	<i>dispersion</i>	<i>expectation</i>	<i>estimation</i>	<i>deviance</i>
	hajonta	odotusarvo	estimointi	devianssi
	<i>hajuvus</i>	<i>keskväärtus</i>	<i>hindamine</i>	<i>hällbimus</i>
	<i>power</i>	<i>likelihood</i>	<i>spatial</i>	<i>covariate</i>
	voima	uskottavuus	spatiaalinen	kovariaatti
	<i>võimsus</i>	<i>tõepära</i>	<i>ruumiline</i>	<i>ühismuutuja</i>
Vieras	<i>statistic</i>	<i>cluster</i>	<i>median</i>	<i>regression</i>
	tunnusluku	ryväs	mediaani	regressio
	<i>statistik</i>	<i>klaster</i>	<i>mediaan</i>	<i>regressioon</i>
	<i>prediction</i>	<i>efficiency</i>	<i>censoring</i>	<i>variance</i>
	ennuste	tehokkuus	sensurointi	varianssi
	<i>prognoos</i>	<i>efektiivsus</i>	<i>tsenseerimine</i>	<i>dispersioon</i>

Taulukon vasemman ylänurkan sanoja voidaan jaotella edelleen sen mukaan, perustuvatko molempien kielten termit samaan itämerensuomalaiseen kantamuotoon. Tätä edustaa termin *dispersion* suomennos *hajonta* ja vironnos *hajuvus*, jotka kumpikin ovat peräisin samasta hajoamiseen viittaavasta kantaverbistä. Samoin esimerkiksi viron *lähima naabri meetod* on todellakin lähinaapurimenetelmä (*nearest neighbor method*). Näyttää kuitenkin siltä, että enemmistö vasemman ylänurkan termeistä pohjautuu

¹Monet termit koostuvat kahden tai useamman erillisen sanan muodostamasta yhdyssanasta tai sanaliitosta. Näiden joukossa on paljon sellaisia, joissa osa termin sisältämisestä sanoista on omakielisiä ja loput vierasperäisiä. Hienojakoisempi analyysi edellyttäisikin, että oma/vieras-akselille lisättäisiin kolmas kategoria, joka kattaa molempia elementtejä sisältävät sekamuotoiset termit.

suomessa ja virossa erilaisiin kantasanoihin. Esimerkkinä olkoon termipari *sampling* ja *sample*. Suomessa ne ovat *otanta* ja *otos*, mutta vironkieliset vastineet *valik* ja *valim* on johdettu verbistä *valima*, joka on suomeksi “valita”. Aikoinaan *sampling* ja *sample* on käännetty viroksi sanoilla [†]*väljavõtt* ja [†]*võend*, jotka on johdettu verbistä *võtma*, suomeksi “ottaa”, mutta näitä termejä ei enää käytetä.

Analogisesti voidaan jaotella myös oikean alanurkan solun sisältämät vierassanat sen mukaan, ollaanko samalla lähteellä vai ei. Esimerkiksi termin *median* kohdalla lähde on sama; samoin termeillä *correlation* ja *regression*. Sen sijaan termi *variance*, joka on suomeksi *varianssi*, onkin viroksi *dispersioon*, mikä saattaa olla yhteydessä siihen, että mm. saksan kielessä *varianssia* kutsuttiin aiemmin nimellä *Dispersion*². Tällaiset tapaukset ovat kuitenkin varsin harvinaisia.

4.4 Onnistuneita omakielisiä termejä puolin ja toisin

Suomalaisen sanaston kehittäjän kannalta nelikentän oikea ylänurkka saattaa tarjota virikkeitä siitä, miten meillä käytössä oleva vierassana voitaisiin suomentaa. Esimerkki hyvästä omakielisestä viron tilastosanasta on termin *range* vastine *haare*, joka voisi kelvata sellaisenaan suomeenkin pitkänpuoleista oppisanaa vaihteluvälin pituus paljon lyhyempänä.

Myös *neelav olek* on termin *absorbing state* hyvä käännös, joka on yhdistelmä verbistä *neelama* eli “niellä” ja substantiivista *olek* eli “tila”. Tämän innoittamana olemme ehdottamassa uutta suomennosta *nielevä tila* meillä vakiintuneen *absorboivan tilan* rinnalle.

Ongelmatermi *odds* on onnistuneesti vironnettu vieraslainalla *šansid*, joka on sanan *šanss* monikkomuoto, ja *odds ratio* on puolestaan *šanssisuhe*. Vaikka “sanssit” ja “sanssisuhde” ovatkin houkuttelevia, ehdotamme ensin mainitulle suomennoksia *vastasuhde* ja jälkimmäiselle *ristisuhde* ja *vetosuhde* (ks. alaluku 2.9). Sanastossa termille *odds ratio* on ajateltu riittävän nämä kaksi suomennosta, mutta pitempi synonyymi *ristitulosuhde* tulee myös vastaan termin *cross-product ratio* yhtenä suomenkielisenä vastineena. *Homoskedasticity* ja *heteroskedasticity* on vironnettu napakasti *püsihajuvs* ja *muuthajuvs*. Ehdottamamme *samavarianssinen* ja *erivarianssinen* ovat merkitykseltään näitä lähellä. Termi *deviance* on lainattu suomeen suoraan muodossa *devianssi*, mutta virossa on omakielinen *hälbimus*, suomeksi “poikkeama”.

Termin *sufficiency* suomennosta *tyhjentävyys* rohkenemme puolestaan pitää jopa alkukielistä termiä parempana (ks. alaluku 3.4). Vironkielinen vastine *piisavus* on sananmukaisesti täysin piisaava käännös, mutta suomen

²Nykyään saksassa käytetään pääsääntöisesti termiä *Varianz*, joskin joissakin oppikirjoissa vieläkin tarjotaan synonyymiksi *Dispersion*.

esimerkkiä noudattava *ammendavus* voisi olla hyvä vaihtoehto, koska viron adjektiivi *ammendav* on suomeksi tyhjentävä. Myös termi tarkentuvuus on mielestämme parempi kuin englannin *consistency*, jolla on monenlaisia sivumerkityksiä. Termi on viroksi käännetty *mõjus*, jolle sanakirja tarjoaa suomennokset “tehokkuus” ja “vaikuttavuus”.³ Tämä näyttää menevän rishtiin estimaattorin kolmannen tärkeän ominaisuuden eli *efficiency* kanssa, joka on suomeksi suoraan tehokkuus, mutta virossa turvaudutaan vieraslainaan *efektiivsus*. Viroksi tarkentuvuus voisi esimerkiksi olla *täpsustuvus*, koska verbi “tarkentua” on viroksi *täpsustuma*.

4.5 Petollisia ystäviä

Erityisen kiinnostavia, usein suorastaan huvittavia (viron *huvitav* tarkoittaa “mielenkiintoinen”) ovat *faux amis* -sanat eli petolliset ystävät, jotka muistuttavat erehdyttävästi toisen kielen sanoja, mutta joiden merkitys on pahimmillaan hyvin erilainen. Tällaisia viron tilastosanoja ovat esimerkiksi *pimekatse*, suomeksi sokkokoe (*blinded experiment*), samoin kuin *kihtvalim*, joka ei suinkaan ole “kihtivalimo” vaan ositettu otos (*stratified sample*). Yhdyssana *vabaliige* (äännetään melkein kuin “vapaliike”) ei ole ongen heilahdus vaan on regressiomalleista tutun termin *intercept* vironkielinen vastine; *vaba* on “vapaa” ja *liige* tarkoittaa yleiskielessä “jäsen” mutta matematiikassa termiä merkityksessä “summan jäsen”. *Arvutusstatistika* on kaukana arvoituksista ja arvuuttelusta; kyse on erikoisalasta nimellä laskennallinen tilastotiede eli *computational statistics*. *Ruumiline statistika* ei liity vainajien eikä kehollisen aktiviteetin tilastointiin vaan on termin *spatial statistics* vironnos, jossa alkusanan kantana oleva *ruum* tarkoittaa sanakirjan mukaan seuraavia asioita: “tila, sija, huone, avaruus”, kuten pohjana oleva saksan *Raum* tai ruotsin *rum*. Sana *ümardamisviga* ei tarkoita puutetta käsityskyvyssä vaan kyseessä on pyöristysvirhe.

Suomalaisista termeistä välimatka-asteikko (*interval scale*) kuulostaa virolaisen korvaan melkein kuin puhuttaisiin “kenttäretken järjestysluvusta”, koska *väli* on “kenttä”, *matk* on “retki” ja *astak* on järjestysluku, sijaluku (*rank*). Mittaamisen tai estimoinnin tarkkuudesta (*accuracy*) puhuminen puolestaan saattaa johdatella virolaisen kuulijan suorastaan metatilastollisiin ajatuksiin, koska viron *tarkus* tarkoittaa “viisautta”.

Todettakoon lopuksi, että virolaisen kollegan kuullessa suomalaisen suusta sanan tunnusluku (*statistic*), hän saattaa mielessään tulkita sen “muuttujatarinaksi”. Viron tilastosanaston *tunnus* tarkoittaa nimittäin muuttujaa (*variable*), mutta *lugu*, joka äännetään kuten suomen “luku”, onkin suomeksi “juttu”, “kertomus” tai “tarina”. Tulkinta ei ole aivan väärä, kos-

³Vastaava adjektiivi *mõjus* voidaan joissakin yhteyksissä kääntää suomeksi samasta kantasanasta peräisin olevalla sanalla “mojova”.

ka voidaan hyvin ajatella, että tunnusluku kertoo muuttujan ja sen arvojen sisältämää tarinaa tiivistetyssä muodossa.

5 Ohjeita käyttäjälle – Instructions for the user

5.1 Englanti–suomi–englanti-sanasto

Sanasto ei sisällä termien määritelmiä. Oletuksena on, että lukija tuntee alan käsitteet ja tehtävänä on englannissa ja suomessa käytettyjen sanojen vastaavuuden löytäminen. Englanninkielisten termien asianmukaiset määritelmät löytyvät lähes poikkeuksetta Wikipediasta. Toki poikkeuksia on.

Englanti–suomi-sanasto alkaa englanninkielisellä termillä. Seuraavana on sen synonyymi suluissa, mikäli yhtä hyvä tai parempi englanninkielinen termi on olemassa. Tämän jälkeen tulee *suositeltava* suomenkielinen vastine. Useassa tapauksessa on olemassa englannin kielen mukainen *vaihtoehto*, joka sopii suomen kieleen ja joka on yleisessä käytössä. Viimeisenä voi tulla *mahdollisena* pitämämme kolmaskin käännös, erityisesti silloin kun vakiintunutta käytäntöä ei ole.

Suomi–englanti-sanasto on rakenteeltaan muuten vastaava, mutta emme ole merkinneet synonyymejä siksi, että sanastossa on koko joukko uudissanoja, joiden synonyymisyyden aste ei ole lukijalle ilmeinen. Kaikkia englanti–suomi-sanastossa esiintyviä suomenkielisiä sanoja ei ole otettu mukaan sellaisissa tapauksissa, joissa ehdottamamme uudissanon merkitys vaatisi lisäselityksiä.

Sanat on aakkostettu sivuuttaen sanavälit, yhdysmerkit ja numerot. Näin esimerkiksi rivi *cross-correlation* on ennen riviä *cross tabulation* ja H_0 on ensimmäinen H-alkuinen “sana”. Ajatus on nopeuttaa sanojen löytämistä.

5.2 Finnish-English-Finnish dictionary

The English-Finnish-English list of statistical terms has primarily been compiled with a Finnish reader in mind. However, there is an increasing number of professionals from abroad in Finland’s institutions of higher learning, and in the media. A brief explanation of the contents follows.

The English-Finnish word list contains statistical terms in English, and their possible synonyms in parentheses. The decision on synonymy is a judgment call on the part of the authors, intended to provide a clue as to possible alternative wordings. Moreover, words that are synonymous to one person may mean slightly different things to another. In particular, no attempt has been made to display the numerous equivalent names of statistical distributions.

Then comes our preferred Finnish translation, an alternative, and a possible third version. The ordering is a compromise and represents the view of the authors. We have attempted to provide an expression of Finnish origin whenever possible, even when a Finnicized version of the English term exists. Thus, for example, we have translated *adjusted*, depending on the context, into either *muokattu* or *vakioitu*, thereby avoiding “adjustoitu”, in spite of its common use among professionals. When the Finnicized version is deeply ingrained in the Finnish terminology (e.g., *median* → *mediaani*), it has been given as the first translation, in many cases as the only translation. However, when the Finnicization has not been used with high frequency, and a natural neologism exists, we have given preference to the neologism.

The English-Finnish list contains words with subscripts. These terms correspond to different concepts. No attempt has been made to explain to an English reader the semantic distinctions. On the other hand, in most cases the Finnish translations should convey to a Finnish reader the semantic content. A fair degree of substantive knowledge is needed to fully comprehend each term. In practice, Wikipedia provides a good place to start disentangling alternative meanings.

The Finnish-English list is based on the Finnish translations of the English-Finnish list. In the latter, the number of Finnish terms exceeds that of the English terms for two reasons. First, we have deliberately introduced neologisms as alternatives to earlier usage. Second, compounding words is the preferred way to form expressions for new concepts in Finnish. In English, the compound expressions typically consist of separate words. For these reasons the list of Finnish words is a subset of those appearing in the English-Finnish list.

6 Englanti–suomi

A

ABC₁ (approximate Bayesian computation)	acceptable quality limit hyväksyttävä laatutaso
ABC₂ (approximate bootstrap confidence interval)	acceptance boundary (acceptance line) hyväksymisraja
absolute deviation poikkeaman itseisarvo, itseispoikkeama	acceptance control chart hyväksymisvalvontakaavio
absolute difference erotuksen itseisarvo, itseiserotus	acceptance error (false acceptance) hyväksymisvirhe
absolute error virheen itseisarvo, itseisvirhe	acceptance line (acceptance boundary) hyväksymisraja
absolute frequency lukumäärä, absoluuttinen frekvenssi	acceptance quality level (acceptable quality level) hyväksyttävä laatutaso
absolute moment itseismomentti	acceptance region hyväksymisalue
absolute risk absoluuttinen riski	acceptance sampling hyväksymisotanta
absolute scale absoluuttinen asteikko	accrual rate kertymisnopeus
absolute value itseisarvo	accuracy tarkkuus ₁
absolutely continuous absoluuttisesti jatkuva	ACE₁ (alternating conditional expectations regression)
absolutely convergent itseisesti suppeneva	ACE₂ (average causal effect)
absolutely summable itseisesti summautuva	ACF (autocorrelation function)
absorbing barrier absorboiva reuna, nielevä reuna	acquiescence bias myöntyväisyys-harha
absorbing region absorboiva alue, nielevä alue	activation function aktivointifunktio
absorbing state absorboiva tila, nielevä tila	active learning aktiivinen oppiminen
absorption distribution imeytymisjakauma, absorptiojakauma	actuarial estimator aktuaariestimattori
A/B-test A/B-testi	actuarial statistics vakuutustilastotiede
accelerated failure time model (AFT-model) nopeutetun vikaantumisajan malli, AFT-malli	AdaBoost (adaptive boosting)
accelerated life testing nopeutettu elinaikatestaus	adapted process mukautettu prosessi
	adaptive boosting (AdaBoost) mukautuva tehostus
	adaptive design mukautuva ase-

telma	maattori
adaptive inference mukautuva päättely	admissible test käypä testi
adaptive regression mukautuva regressio	a.e. (almost everywhere) m.k.
added variable plot (partial regression plot) lisäselittäjäkuvio, osittaisregressiokuvio	AFT (accelerated failure time)
addition rule yhteenlaskusääntö	age-adjusted ikävakioitu
additive hazards model additiivisten uhkien malli, summautuvien uhkien malli	age-dependent process ikäriippuva prosessi
additive model additiivinen malli, summautuva malli	age heaping ikäsuosio, ikäkasautuminen
additive process additiivinen prosessi, summautuva prosessi	agent-based model agenttipohjainen malli
additive seasonal model additiivinen kausivaihtelumalli, summautuva kausivaihtelumalli	age-period-cohort model ikä-periodi-kohorttimalli
adequate model kelpoinen malli	age pyramid ikäpyramidi
adherence to treatment (compliance with treatment) hoitositoutuneisuus, hoitomyöntyvyys	age-specific ikäryhmittäinen, i'ittäinen
adjacency matrix vierusmatriisi	age-specific death rate (age-specific mortality rate) ikäryhmittäinen kuolevuus, i'ittäinen kuolevuus
adjacent categories model vierusluokkamalli	age-specific mortality rate (age-specific death rate) ikäryhmittäinen kuolevuus, i'ittäinen kuolevuus
adjugate matrix liittomatriisi	agglomerative hierarchical clustering kokoava hierarkkinen ryhmittely
adjusted estimator vakioitu estimaattori, rukattu estimaattori	aggregate imputation ryhmäpaikkaus, aggregaatti-imputointi
adjusted profile likelihood muokattu profiiliuskottavuus, rukattu profiiliuskottavuus	aggregate-level study ryhmätason tutkimus
adjusted R^2 muokattu R^2 , rukattu R^2	aggregation aggregointi
adjusted sample weight muokattu otospaino, rukattu otospaino	AHP (analytic hierarchy process)
adjustment vakiointi, muokkaus, rukkaus	AI (artificial intelligence)
admissible käypä, hyväksyttävä	AIC (Akaike information criterion)
admissible decision function käypä päätösfunktio	AID (automatic interaction detection)
admissible estimator käypä esti-	Akaike information criterion (AIC) Akaiken informaatiokriteeri, AIC
	algorithm algoritmi
	aliasing₁ täydellinen kollineaari-

suus	tu jakauma
aliasing₂ laskostuminen	annual data vuosiaineisto
allele frequency alleelifrekvenssi	annuity₁ elinkorko, annuiteetti
allocation₁ kiintiöinti	annuity₂ tasaerä, annuiteetti
allocation₂ kohdentaminen	anomaly anomalia
allometry allometria	ANOVA (analysis of variance)
almost everywhere (a.e.) melkein kaikkialla, m.k.	antecedent of a rule säännön edeltäjäosa
almost surely (a.s.) melkein varmasti, m.v.	antedependence riippuvuus edeltävästä
alternating conditional expectations regression (ACE ₁) vuorottelevien ehdollistusten regressio	anticipative control ennakoiva säätö
alternative hypothesis vastahypoteesi, vaihtoehtoinen hypoteesi	antithetic vastakkaismerkkinen, antiteettinen
American option amerikkalainen optio	antithetic variate vastamuuttuja
amplitude amplitudi, värähdyslaajuus	antitonic regression antitoninen regressio
analysis of covariance (ANCOVA) kovarianssianalyysi	aperiodic jaksoton
analysis of deviance devianssianalyysi	applied statistics sovellettu tilastotiede
analysis of dispersion hajontanalyysi	approximate likiarvoinen, likimääräinen
analysis of variance (ANOVA) varianssianalyysi	approximate Bayesian computation (ABC ₁) likimääräinen Bayes-laskenta
analytic hierarchy process (AHP) analyyttinen hierarkia-prosessi	approximate bootstrap confidence interval (ABC ₂) likimääräinen uusioluottamusväli
analytics analytiikka	approximation₁ likiarvoistus, approksimointi
ancillary information apuinformaatio	approximation₂ likiarvo, approksimaatio
ancillary statistic aputunnusluku	arbitrage arbitraasi
ANCOVA (analysis of covariance)	ARCH model (autoregressive conditional heteroskedastic model) ARCH-malli
angular transformation kulmuunnos	arcsine distribution arkussinijakauma
anisotropic anisotrooppinen	arcsine transformation arkussinimuunnos
annealed distribution hehkutet-	ARE (asymptotic relative efficiency)
	area sampling alueotanta

area under curve (AUC) käyrän- alainen pinta-ala	(ARE) asymptoottinen suhteelli- nen tehokkuus
ARFIMA model (autoregressive fractionally integrated moving average model) ARFIMA-malli	asymptotics asymptotiikka
ARIMA model (autoregressive integrated moving average model) ARIMA-malli	ATE (average treatment effect)
arithmetic average (arithmetic mean) aritmeettinen keskiarvo, keskiarvo	atom atomi
arithmetic distribution arit- meettinen jakauma	ATT (average treatment effect on the treated)
arithmetic mean (arithmetic average) aritmeettinen keskiarvo, keskiarvo	attack rate sairastuneiden osuus
ARMA model (autoregressive moving average malli) ARMA- malli	attainable portfolio saavutetta- va salkku
AR model (autoregressive model) AR-malli	attenuation vaimeneminen
array (table) taulukko	attractor attraktori
arrival process saapumisprosessi	attributable fraction syyksiluku- osuus
artificial intelligence (AI) tekoä- ly, keinoäly	attribute ominaisuus, piirre
a.s. (almost sure) m.v.	attrition harventuminen, kato
as-randomized analysis satun- naistuksen mukainen analyysi	ATU (average treatment effect on the untreated)
ascertainment₁ jäljitys	AUC (area under curve)
ascertainment₂ varmistus	augmented Dickey-Fuller test täydennetty Dickeyn–Fullerin tes- ti
ascertainment bias jäljitysharha	autocorrelation autokorrelaatio
assay määrittäminen, määrittäminen	autocorrelation function (ACF) autokorrelaatiofunktio
asset arvopaperi	autocovariance autokovarianssi
association yhteys, assosiaatio	autocovariance function auto- kovarianssifunktio
association analysis yhteysanalyysi, assosiaatioana- lyysi	autocovariance generating function autokovarianssiemä- funktio, autokovarianssit generoi- va funktio
astrostatistics tähtitilastotiede	automatic interaction detec- tion (AID) automaattinen inte- raktiohaku, automaattinen yhdys- vaikutushaku, AID
asymptotic asymptoottinen	autoregression autoregressio
asymptotic distribution (limiting distribution) rajajakau- ma	autoregressive conditional het- eroskedastic model (ARCH- model) autoregressiivinen ehdol- lisen heteroskedastisuuden malli, ARCH-malli
asymptotic relative efficiency	autoregressive distributed-lag

model autoregressiivinen ekso-
geenisten viipeiden malli
**autoregressive fractionally
integrated moving average
model** (ARFIMA model) auto-
regressiivinen fraktionaalisesti
integroitu liukuvan keskiarvon
malli, ARFIMA-malli
**autoregressive integrated mo-
ving average model** (ARIMA
model) autoregressiivinen integroi-
tu liukuvan keskiarvon malli,
ARIMA-malli
autoregressive model (AR-
model) autoregressiivinen malli,
AR-malli
**autoregressive moving avera-
ge model** (ARMA-model) auto-
regressiivinen liukuvan keskiarvon
malli, ARMA-malli
auxiliary information lisäinfor-
maatio
auxiliary variable apumuuttuja,
lisämuuttuja
available case analysis saatavil-
la olevien havaintojen analyysi
average (mean_1) keskiarvo
average causal effect (ACE_2)
keskimääräinen kausaali vaikutus
average run length keskimääräi-
nen merkkijakson pituus
average treatment effect (ATE)
keskimääräinen käsittelyvaikutus
**average treatment effect on
the treated** (ATT) keskimääräi-
nen käsittelyvaikutus altistuneilla
**average treatment effect on
the untreated** (ATU) keskimää-
räinen käsittelyvaikutus altistu-
mattomilla
average Value at Risk
(expected shortfall) odotustap-
pioarvo

axial distribution akselijakauma

B

backcalculation taakselaskenta
backcasting (backprediction)
menneistäminen
backdoor adjustment takaoviva-
kiointi
backdoor path takaovipolku
backfitting taaksesovitus
backprediction (backcasting)
menneistäminen
backprojection taakseprojisointi
backpropagation algorithm
vastavirta-algoritmi
backshift operator (lag operator)
viivästysoperaattori, viiveoperaat-
tori
backward elimination
(backward selection) poistovalinta
backward equation taakseyhtälö
backward selection (backward
elimination) poistovalinta
bagging (bootstrap aggregating)
laukutis
bag-of-words model sanapussi-
malli
bagplot (starburst plot) tähtisade-
kuvio
balance equation tasapainoyhtä-
lö
**balanced incomplete block de-
sign** (BIBD) tasapainoinen vajaa-
lohkoasetelma
BAN (best asymptotically normal
estimator)
bandwidth₁ kaistanleveys, taa-
juusalue
bandwidth₂ (window width) ikku-
nanleveys
bar chart (bar plot) pylväskuvio,

palkkikuvio ¹	belief network uskomusverkko
bar plot (bar chart) pylväskuvio, palkkikuvio	Bernoulli distribution Bernoulli-jakauma
baseline lähtötaso, perustaso	Bernoulli experiment (Bernoulli trial) Bernoulli-koe
baseline characteristics lähtöarvot	Bernoulli trial (Bernoulli experiment) Bernoulli-koe
baseline hazard perusuhka	best asymptotically normal estimator (BAN) paras asymptootisesti normaalin estimaattori
base period (reference period) perusjakso, viitejakso	best linear unbiased estimator (BLUE) paras lineaarinen harhaton estimaattori
base rate bias (base rate fallacy) perustasoharha	best linear unbiased prediction (BLUP) paras lineaarinen harhaton ennustin
base rate fallacy (base rate bias) perustasoharha	best practice life expectancy parhaan olosuhteen elinajanodote
basic reproduction number (R_0) perustartuttavuusluku, perusuusiutumisluku, R_0	beta-binomial distribution beta-binomijakauma
basic sample weight perusotospaino	beta distribution betajakauma
basis kanta	between-groups estimator ryhmien välinen estimaattori
batch variation erävaihtelu	between-groups sum of squares ryhmien välinen neliösumma
bathtub curve ammekäyrä	between-groups variance ryhmien välinen varianssi
battery of tests testipatteri	bias (systematic error) harha, systemaattinen virhe
Bayes factor Bayes-kerroin, Bayes-tekijä	bias-corrected and accelerated interval (BCa) nopea harhakorjattu luottamusväli
Bayes' formula Bayesin kaava	biased estimator harhainen estimaattori
Bayesian bayesiläinen, bayesläinen	biased sampling harhainen otanta
Bayesian inference Bayes-päätely	biased test harhainen testi
Bayesian information criterion (BIC) Bayes-informaatiokriteeri, BIC	bias-variance tradeoff harha-varianssikompromissi
Bayesian nonparametrics parametrittomat Bayes-menetelmät	BIBD (balanced incomplete block design)
Bayesian updating Bayes-päivitys	
Bayes network Bayes-verkko	
Bayes risk Bayes-riski	
BCa (bias-corrected and accelerated interval)	
belief function uskomusfunktio	

¹Pylväskuviossa pylväät ovat pystysuorassa; palkkikuviossa vaakasuorassa.

BIC (Bayesian information criterion)	birth cohort syntymäkohortti
bidegree distribution kaksiastejakauma	birth-death process syntymä-kuolemaprosessi
big data suuraineisto, suurdata, massadata	birth process syntymäprosessi
bilinear model bilineaarinen malli	birth rate syntyvyys
bimodal distribution kaksihuippuinen jakauma	biserial correlation biseriaalinen korrelaatio
bin luokkaväli	bispectrum bispektri
binary variable (dichotomous variable) kaksiarvoinen muuttuja, binäärimuuttuja, dikotominen muuttuja	bit bitti
binned scatter plot siilosirontakuvio	bivariate distribution kaksiulotteinen jakauma, kahden muuttujan jakauma
binomial coefficient binomikerroin	bivariate normal distribution kaksiulotteinen normaalijakauma, kaksinormaalijakauma
binomial distribution binomijakauma	black box musta laatikko
binomial experiment (binomial trial) binomikoe	Black–Scholes model Blackin–Scholesin malli
binomial test binomitesti	blinded experiment sokkokoe
binomial trial (binomial experiment) binomikoe	blinding sokkoutus
bioassay biomääritys	block lohko
bioequivalence test bioekvivalenssitesti	block diagonal matrix lohkolävistäjämatriisi, lohkodeagonaalimatriisi
bioinformatics bioinformatiikka	block diagram lohkokaavio
biometric identification (biometrics) biometrinen tunnistaminen, biometriikka	block randomization lohkosatunnaistaminen
biometrics₁ (biometry) biometria	BLUE (best linear unbiased estimator)
biometrics₂ (biometric identification) biometrinen tunnistaminen, biometriikka	BLUP (best linear unbiased predictor)
biometry (biometrics ₁) biometria	BMI (body mass index)
biostatistics biotilastotiede, biostatistiikka	body mass index (BMI) painoindeksi
bipartite kaksijakoinen	body of a rule säännön vartalo
biplot bikuvio, yksilö-muuttujakuvi	bond velkakirja
	Bonferroni inequality Bonferronin epäyhtälö
	boosting tehostus
	bootstrap aggregating (bagging) uusioaggregointi, laukutus
	bootstrap method uusiomenetelmä, uusio-otantamenetelmä

bootstrap confidence interval uusioluottamusväli
bootstrap estimator uusioestimaattori
bootstrapping (bootstrap sampling), uusio-otanta
bootstrap p -value uusio- p -arvo
bootstrap sampling (bootstrapping), uusio-otanta
bootstrap testing uusiotestaus
borrowing of strength voiman lainaaminen
bounded in probability stokastisesti rajoitettu
bounded variation function rajoitetusti heilahteleva funktio
box-and-whisker plot (box plot) laatikkokuvio
box plot laatikkokuvio
branching process haarautumisprosessi
breakdown point murtumispiste
Brownian bridge Brownin silta
Brownian motion Brownin liike
bubble plot kuplakuvio
burn-in period (warm-up period) sisäänajojakso₁
business cycle suhdanne
busy period kiirejakso

C

càdlàg (continue à droite, limit à gauche) cadlag
calculus of probability (probability calculus) todennäköisyyslaskenta
calibrated weight kalibroitu paino
calibration kalibointi
caliper matching välikaltaistus
call-back toistokutsu, toistokäynti, karhunta

call option osto-oikeus, osto-optio
candidate distribution (proposal distribution) ehdokasjakauma, ehdotusjakauma
canonical analysis kanoninen analyysi
canonical correlation kanoninen korrelaatio
canonical link function kanoninen linkkifunktio
canonical parameter (natural parameter) kanoninen parametri, luonnollinen parametri
CACE₁ (complier average causal effect)
CACE₂ (conditional average causal effect)
CAPI (computer-assisted personal interview)
capture-recapture sampling (mark-recapture sampling) merkintä-takaisinpyyntiotanta, toistopyyntiotanta
CAR (conditional autoregression)
cardioid distribution (cosine distribution) kosinijakauma
carrier of disease taudinkantaja
carrier variable kantajamuuttuja
carry-over effect jälkivaikutus
CART (classification and regression tree)
cartogram kartogrammi
cascade process ryöppyprosessi, kaskadiprosessi
case tapaus₁
case-base study tapaus-pohjatutkimus, tapaus-alustatutkimus
case-cohort study tapaus-kohorttitutkimus
case-control study (case-referent study) tapaus-verrokkitutkimus
case fatality ratio (CFR) tapauskuolleisuus, CFR

case-referent study (case-control study) tapaus-verrokkitutkimus	terveys)
CASI (computer-assisted self-interview)	CDE (controlled direct effect)
catastrophe theory katastrofi-teoria	CDF (cumulative distribution function)
CATE₁ (complier average treatment effect)	ceiling effect kattovaikutus
CATE₂ (conditional average treatment effect)	cell frequency solulukumäärä, solufrekvenssi
categorical distribution luokkajakauma	censored observation sensuroitu havainto
categorical variable luokkamuuttuja, luokitteluaasteikon muuttuja	censoring sensurointi
CATI (computer-assisted telephone interview)	census₁ (complete enumeration) kokonaistutkimus, kokonaislaskenta
cauchit transformation cauchitmuunnos	census₂ väestölaskenta
Cauchy distribution Cauchy-jakauma	census survival rate laskentojen välinen elossaolosuhde
causal chain syyketju, kausaaliketju	centering keskistäminen
causal diagram (causal graph) kausaalikaavio, kausaaliverkko	centile sentiili
causal discovery kausaalirakenteen löytäminen	central limit theorem keskeinen raja-arvolause
causal graph (causal diagram) kausaalikaavio, kausaaliverkko	central moment keskusmomentti
causal inference syy-seurauspäätely, kausaalipäätely	centroid method sentroidimenetelmä
causality syy-seuraussuhde, kausaalisuus	chain ketju
causal model kausaalimalli, syy-seurausmalli	chain binomial model ketjubinomimalli
cause variable syytuuttuja	chain graph ketjukaavio
cause-specific hazard (cause-specific hazard rate) syykohtainen uhka	chain index ketjuindeksi
CAWI (computer-assisted web-interview)	chance variation (random variation) satunnaisvaihtelu
	changeover design (cross-over design) vaihtovuoroasetelma, riskikäsasetelma
	change point muutospiste
	change score muutosarvo
	characteristic equation ominaisarvoyhtälö, karakteristinen yhtälö
	characteristic function ² karakteristinen funktio
	characteristic value (eigenvalue)

²Todennäköisyyslaskennassa karakteristinen funktio on satunnaismuuttujan X jakauman määrittävä kompleksiarvoinen funktio $\phi(t) = E[e^{itX}]$, mutta muualla matematiikassa termillä tarkoitetaan jonkin joukon A indikaattorifunktiota.

ominaisarvo	tö
Chebyshev's inequality ³ Tshe- byshevin epäyhtälö	classification table luokittelutau- lukko
chemometrics kemometria	classification statistic luokittelu- tunnusluku
Chernoff's faces Chernoffin kas- vokuviot	classifier luokitin, luokittelija
chi distribution khi-jakauma	class interval luokkaväli
Chinese restaurant process kiin- nalainen ravintolaprosessi	class midpoint luokkakeskus
chi-plot khi-kuvio	clinical trial kliininen koe
chi-squared distance khi-neliö- etäisyys	cloglog-transformation (complementary loglog transfor- mation) cloglog-muunnos
chi-squared distribution khi- neliöjakauma	closed population suljettu väes- tö
chi-squared statistic khi-neliö- suure, khi-neliötunnusluku	cluster ryväs ⁴
chi-squared test khi-neliötesti	cluster analysis ryhmittelyana- lyysi
chi-statistic khi-suure, khi-tun- nusluku	clustering ryvästys
choice-based sampling valinta- perusteinen otanta	cluster process ryväsprosessi
CIF (cumulative incidence func- tion)	cluster randomization ryvässa- tunnaistus
circular data kehäaineisto, kehä- data	cluster sampling ryväсотanta
circular distribution kehäjakau- ma	coalescence sulautuminen, koale- senssi
city block distance (Manhattan distance) korttelietäisyys, manhat- tanetäisyys	coarse data karkea data
class frequency luokkafrekvenssi	coarsening karkeistus
classification luokittelu, luokitus	COD (coefficient of divergence)
classification and regression tree (CART) luokittelu- ja regres- siopuu	CoDA (compositional data analy- sis)
classification attribute luokitte- luattribuutti, luokittelumääre	coefficient kerroin
classification error luokitteluvir- he	coefficient of association assosi- aatiokerroin, yhteyskerroin
classification rule luokittelusään-	coefficient of concentration kes- kittymiskerroin
	coefficient of concordance kon- kordanssikerroin, yhteensopivuus- kerroin
	coefficient of determination (R^2) selityssaste, determinaatioker- roin

³Nimestä on vaihtelevia kirjoitusasuja niin englannin- kuin suomenkielisissäkin teksteissä.

⁴Taivutetaan "rypään".

coefficient of disarray	group) vertailuryhmä
(Kendall's tau) Kendallin tau	compartment model lokeromalli
coefficient of divergence (COD)	compatibility interval
poikkeamiskerroin	(confidence interval) yhteensopi-
coefficient of variation	vuusväli, luottamusväli
variaatiokerroin	competing event kilpaileva ta-
cofactor kofaktori	pahtuma
cognitive bias kognitiivinen har-	competing risk kilpaileva riski
ha	complementary error function
coherence johdonmukaisuus, ko-	(erfc) komplementtinen virhefunk-
herenssi	tio
cohort kohortti	complementary loglog
cohort study kohorttitutkimus	transformation (cloglog-
cointegration	transformation) komplementtinen
yhteisintegroituvuus	loglog-muunnos, cloglog-muunnos
cold deck imputation kylmä-	complement event
paikkaus, kylmäimputointi	vastatapahtuma, komplementti-
collapsibility koontuvuus, yhdis-	tapahtuma
tettävyys	complete case analysis koko-
collider törmäytin	naisten havaintojen analyysi
collider bias törmäytinharha	complete enumeration (census ₁)
collinearity kollineaarisuus	kokonaistutkimus, kokonaislasken-
COMFAC model (common fac-	ta
tor model ₂)	completely randomized design
commonality kommunaliteetti	täyssatunnaistettu asetelma
common factor yhteisfaktori	complete separation täydellinen
common factor model₁	erottelu
yhteisfaktorimalli	complete statistic erotteleva
common factor model₂	tunnusluku
(COMFAC model) yhteisten te-	compliance with treatment
kijöiden malli	(adherence to treatment) hoito-
communicating class	myöntövyys, hoitositoutuneisuus
yhteysluokka	complier average causal effect
community detection	(CACE ₁) keskimääräinen kausaali-
yhteisötunnistaminen	vaikutus sitoutuneilla
companion form kumppanimuo-	complier average treatment
to	effect (CATE ₁) keskimääräinen
comparative experiment	käsittelyvaikutus sitoutuneilla
(controlled experiment) vertaile-	component and compo-
va koe, vertailukoe	ment-plus-residual plot (partial
comparative parameter	residual plot) osittaisjäännöskuvio
vertailuparametri	composite hypothesis
comparison group (control	yhdistetty hypoteesi

compositional data analysis	conditional autoregression
(CoDA) rakenneosa-aineiston analyysi	(CAR) ehdollinen autoregressio
compound interest korkoa korkolle	conditional average causal effect (CACE₂) ehdollinen keskimääräinen kausaalivaikutus
compound Poisson distribution yhdistetty Poisson-jakauma	conditional average treatment effect (CATE₂) ehdollinen keskimääräinen käsittelyvaikutus
compound Poisson process yhdistetty Poisson-prosessi	conditional density function ehdollinen tiheysfunktio
computationally intensive method laskentaintensiivinen menetelmä	conditional dependence ehdollinen riippuvuus
computational statistics laskennallinen tilastotiede	conditional distribution ehdollinen jakauma
computer-assisted personal interview (CAPI) tietokoneavusteen henkilökohtainen haastattelu	conditional expectation ehdollinen odotusarvo
computer-assisted self-interview (CASI) tietokoneavusteinen itsehaastattelu	conditional independence ehdollinen riippumattomuus
computer-assisted telephone interview (CATI) tietokoneavusteen puhelinhaastattelu	conditional likelihood ehdollinen uskottavuus
computer-assisted web-interview (CAWI) tietokoneavusteen verkkohaastattelu	conditional probability ehdollinen todennäköisyys
computer vision konenäkö	conditional variance ehdollinen varianssi
concavity konkaavisuus, kupeisuus	conditioning plot (coplot) ehdollistettu sirontakuvio
concentration curve keskittyneisyyskäyrä	confidence band (confidence envelope) luottamusvyöhyke
concentration matrix (precision matrix) keskittyneisyysmatriisi	confidence envelope (confidence band) luottamusvyöhyke
concordance yhteensopivuus, konkordanssi	confidence interval (compatibility interval) luottamusväli, yhteensopivuusväli
concordance index yhteensopivuusindeksi	confidence level luottamustaso
concordant pairs konkordantit, samantyyppiset parit	confidence region luottamusalue
concurrent study tosiaikainen tutkimus	confidence set (confidence region) luottamusjoukko, luottamusalue
conditional association ehdollinen yhteys	configuration model konfiguraatiomalli
	confirmatory data analysis vahvistava data-analyysi
	confounder (confounding vari-

able) sekoittaja, sekoittava muuttuja	continuous-time process jatkuva-aikainen prosessi
confounding factor (confounder) sekoittaja, sekoittava tekijä	contour line tasa-arvokäyrä
confounding variable (confounder) sekoittaja, sekoittava muuttuja	contrast kontrasti
confusion matrix ⁵ (classification table) luokittelutaulukko	control verrokki
conjugate distribution liittojakauma	control chart valvontakortti
conjugate prior distribution liittopriorijakauma	control group (comparison group) vertailuryhmä, verrokki-ryhmä
consequent of a rule säännön seuraajaosa	controlled direct effect (CDE) kontrolloitu suora vaikutus
conservative test konservatiivinen testi	controlled trial (comparative trial) vertaileva koe, vertailukoe
consistency ₁ tarkentuvuus	control theory säätöteoria
consistency ₂ yhdenmukaisuus	control variable ₁ säätömuuttuja
consistency of classification luokittelun yhdenmukaisuus	control variable ₂ vakioiva muuttuja
consistent estimator tarkentuva estimaattori	convenience sample mukavuusotos
content analysis sisällönanalyysi	convergence suppeneminen, konvergenssi
contingency coefficient kontingenssikerroin	convergence in distribution jakaumasuppeneminen, jakaumakonvergenssi
contingency table ristitaulukko, kontingenssitaulukko	convergence in probability stokastinen suppeneminen, stokastinen konvergenssi
continuation ratio model jatkosuhtemalli	convex hull konvekksi peite
continue à droite, limite à gauche ⁶ (càdlàg) oikealle jatkuva vasemmalta raja-arvot, cadlag	convexity konveksisuus, kuperaus
continuity jatkuvuus	convolution konvoluutio
continuity correction jatkuvuuskorjaus	co-occurrence yhteisesiintyvyys
continuous distribution jatkuva jakauma	Cook's D (Cook's distance) Cookin etäisyys
	Cook's distance (Cook's D) Cookin etäisyys
	coplot (conditioning plot) ehdollistettu sirontakuvio
	copula kopula

⁵Tätä termiä ja suomennosta "sekaannusmatriisi" käytetään erityisesti koneopimisen kirjallisuudessa.

⁶Tämän ranskankielisen termin vastine englannissa on *right continuous with left limits*. Silti ranskankielistä termiä ja sen lyhennettä *càdlàg* käytetään paljon englanninkielisessä kirjallisuudessa.

correlated	korreloiva	coverage probability	peittotodennäköisyys
correlation	korrelaatio	Cox model	(proportional hazards model) Coxin malli, verrannollisten uhkien malli
correlation coefficient	korrelaatiokerroin	Cramér–Rao lower bound	Cramérin–Raon alaraja
correlation matrix	korrelaatiomatriisi	credibility theory	luottoteoria
correspondence analysis	(dual scaling) korrespondenssianalyysi	credible interval ⁷	(posterior interval) luottoväli, posterioriväli
cosine distribution	(cardioid distribution) kosinijakauma	credible region	luottoalue
cost-benefit analysis	kustannus-hyötyanalyysi	credit	luotto
cost-effectiveness	kustannus-vai-kuttavuus	critical region	(rejection region) hylkäysalue, kriittinen alue
cost-efficiency	kustannus-tehokkuus	critical value	hylkäysraja, kriittinen arvo
count	(frequency ₁) lukumäärä, frekvenssi	Cronbach’s alpha	Cronbachin alfa
countable	numeroituva	cross-correlation	ristikorrelaatio
count data	(frequency data) lukumääräaineisto	cross-covariance	ristikovarianssi
counterfactual	kontrafaktuaalinen	cross-over design	(changeover design) vaihtovuoroasetelma, ristikkäisasetelma
counter-matching	vastakaltaistus	cross-product ratio	(odds ratio) ristisuhde, vetosuhde, ristitulosuhde
counting measure	laskurimitta	cross-sectional study	poikittais-tutkimus
counting process	laskuriprosessi	cross-spectral density	ristispektiritiheys
count time series	lukumääräaikasarja	cross spectrum	ristispektri
coupling	kytkentä	cross table	(contingency table) ristitaulukko, kontingenssitaulukko
coupling of processes	prosessien kytkentä	cross tabulation	ristiintaulukointi
covariance	kovarianssi	cross-validation	ristiinvalidointi, ristinvahvistus
covariance analysis	(ANCOVA) kovarianssianalyysi	crude estimator	karkea estimaattori
covariance matrix	kovarianssimatriisi	cubic lattice	kuutiohila
covariance stationary	(weakly stationary) kovarianssistationaarinen		
covariate	kovariaatti		

⁷Kirjoitusasua *credibility interval* käytetään myös.

cubic spline kuutiosplini
cumulant kumulantti
cumulant generating function kumulanttiemäfunktio, kumulantit generoiva funktio
cumulative cost kustannuskertymä
cumulative distribution function (CDF) kertymäfunktio
cumulative error virhekertymä, kumulatiivinen virhe
cumulative frequency frekvenssikertymä, kumulatiivinen frekvenssi
cumulative hazard (integrated hazard) uhkakertymä
cumulative incidence ilmaantumisosuus, kumulatiivinen ilmaantuvuus
cumulative incidence function (CIF) kumulatiivinen ilmaantuvuusfunktio
cumulative intensity (integrated intensity) intensiteettikertymä
cumulative model kertymämalli, kumulatiivinen malli
cumulative odds ratio kertymäristisuhde
cumulative risk riskikertymä
cumulative sum kertymäsumma, kumulatiivinen summa
cumulative sum chart (CUSUM chart) summakaavio, CUSUM-kaavio
curse of dimensionality dimensiokirous
curtailed inspection keskeytetty tarkastus
curtate kokonaisosa
curved exponential family kaa-reva eksponenttiperhe

curve fitting käyrän sovitus
curvilinear käyräviivainen
CUSUM chart (cumulative sum chart) summakaavio, CUSUM-kaavio
cut-off point kynnysarvo, katkaisupiste, rajapiste
cybernetics kybernetiikka
cycle jakso, sykli, kierros
cyclic model jaksollinen malli, syklinen malli

D

DAG (directed acyclic graph)
DALY (disability adjusted life year)
Dantzig selector Dantzig-valitsin
data⁸ data, aineisto, havaintoaineisto
data analysis data-analyysi, aineistoanalyysi
data augmentation muuttujalavennus
data bank (database) tietokanta
database tietokanta
data cleaning aineiston siivoaminen, datan siivoaminen
data cube havaintokuutio, datakuutio
data depth datan syvyys, havainnon sisäisyys
data dredging (fishing expedition) datan ruoppaus, kalastelu
data envelopment analysis (DEA) verhoikäyräanalyysi, DEA
data driven datavetoinen, aineistovetoinen
data frame datakehikko
data generating process havainnot tuottava prosessi

⁸Sana on monikollinen. Huolessisessa englannissa kirjoitetaan “*the data are*” eikä “*the data is*”. Yksikkömuotoa *datum* ei juuri koskaan käytetä.

data linkage	datan linkittäminen	deep learning	syväoppiminen
data matrix	havaintomatriisi, datamatriisi	de facto population	de facto -väestö, hetkellinen väestö
data mining	datanlouhinta, tiedonlouhinta	deff	(design effect)
data peeking	datan kurkistelu, aineiston kurkistelu	deflation	deflaatio
data pipeline	dataputki	degenerate distribution	degeneroitunut jakauma
data poisoning	datan myrkyttäminen	degree distribution	astejakauma
data preprocessing	aineiston esikäsittely, datan esikäsittely	degree of belief	uskomusaste
data reduction	datan tiivistäminen	degrees of freedom	vapausasteet, vapausasteluku
data science	datatiede	de jure population	de jure -väestö, vakinainen väestö
data screening	aineiston esitarkastus	delayed entry	viivästynyt aloitus
data set	datajoukko, tiedosto	delta method	deltamenetelmä
data wrangling	aineiston jalostaminen, datan jalostaminen	demography	väestötiede, demografia
DEA	(data envelopment analysis)	dendrogram	ryhmittelypuu
death process	kuolemaprosessi	denoised	kohinasta puhdistettu
death rate	(mortality rate) kuolevuus	denoising	kohinanpoisto
decile	desiili, kymmenys ⁹	denominator	nimittäjä
decision function	päätössääntö, päätösfunktio	density estimation	tiheysestimointi
decision space	päätösavaruus	density function	tiheysfunktio
decision theory	päätösteoria	density histogram	tiheyshistogrammi
decision tree	päätöspuu	density ratio model	tiheyssuhdemalli
decomposable model	osittuva malli, dekomponoituva malli	density sampling	tiheysotanta, tiheyspoiminta
decomposition ₁	osittaminen, dekomponointi	denumerable	numeroituvasti ääretön
decomposition ₂	hajotelma	dependence	riippuvuus
deconvolution	dekonvoluutio	dependent variable ¹⁰	(response variable) vastemuuttuja, selitettävä muuttuja
decorrelated	korrelaatiopuhdistettu, korrelaatiosta puhdistettu	derivative ₁	derivaatta
		derivative ₂	johdannainen

⁹Oppikirjojen mukaan desiilit ovat fraktiileja, jotka jakavat aineiston kymmeneen yhtä suureen osajoukkoon. Joissakin yhteyksissä, esim. tulojakaumia kuvattaessa, desiileillä viitataan näihin osajoukkoihin eli kymmenyksiin. Vastaavaa rinnakkaistulkintaa näkee käytettävän myös kvartiileille ja kvintiileille.

¹⁰Parempia termejä ovat *outcome variable* ja *response variable*.

descriptive analytics kuvaileva analytiikka	DIC (deviance information criterion)
descriptive statistic kuvaileva tunnusluku	dichotomous variable (binary variable) kaksiarvoinen muuttuja, dikotominen muuttuja, binäärimuuttuja
descriptive statistics kuvaileva tilastotiede	dichotomy kahtiajako, dikotomia
descriptive study kuvaileva tutkimus	DID (difference in differences)
design asetelma	difference erotus
design-based inference asetelmaperusteinen päättely	difference equation differenssiyhtälö
design effect (deff) asetelmakeroin	difference in betas (DFBETA) kerroinero
design matrix asetelmamatriisi	difference in differences (DID) erotusten erotus
design of experiments koesuunnittelu	difference in fits (DFFIT) soviteero
design optimality asetelmaoptimaalisuus	difference operator erotusoperaattori
design space asetelma-avaruus	differencing erotusten laskeminen, differenssointi, differointi
detailed balance condition tilaparikohtainen tasapainoehto	differentiability derivoituvuus, differentioituvuus
detection bias löytymisharha	differential equation differentiaaliyhtälö
determinant determinantti	differential misclassification eriävä virheluokittelu
deterministic deterministinen	diffuse prior diffuusi prior
detrending trendin poisto	diffusion index diffuusiaindeksi
deviance devianssi	diffusion process diffuusioprosessi
deviance information criterion (DIC) devianssi-informaatiokriteeri	digamma function digamma-funktio
deviance residual devianssijäännös	digit preference numerosuosio
deviation poikkeama	dimension ulottuvuus, dimensio
DFBETA (difference in betas)	dimensionality reduction (dimension reduction) dimension pienennys, ulottuvuuksien vähentäminen
DFFIT (difference in fits)	dimension reduction (dimensionality reduction) dimension pienennys, ulottuvuuksien vähentäminen
diagnostics diagnostiikka	
diagnostic statistic diagnostinen tunnusluku, diagnostinen suure, vikasuure	
diagnostic test diagnostinen testi	
diagonal matrix lävistäjämatrisi, diagonaalimatriisi	
diagram kuvio, diagrammi, kaavio	

Dirac measure	Diracin mitta	dissimilarity coefficient	erilaisuuskerroin
Dirac's delta function	Diracin deltafunktio	dissimilarity matrix	erilaisuusmatriisi
direct effect	suora vaikutus	distance measure	etäisyysmitta
directed acyclic graph (DAG)	suunnattu syklitön verkko, suunnattu silmukaton verkko, DAG	distance sampling	etäisyysotanta
direct likelihood	suora uskottavuus	distributed-lag model	eksogeenisten viipeiden malli
direct standardization	suora vakiointi	distribution	jakauma
directional data	suunta-aineisto	distribution-free	jakaumasta riippumaton
Dirichlet distribution	Dirichlet-jakauma	distribution function	(cumulative distribution function) kertymäfunktio
Dirichlet process	Dirichlet-prosessi	disturbance₁	(perturbation ₁) häiriö
disability adjusted life years (DALY)	toimintakykyiset elinvuodet, DALY	disturbance₂	(error term) virhetermi
disclosure control	tietosuoja	dithering	(jittering) ravistaminen
discontinuous	epäjatkuva	divergence₁	hajaantuminen, divergenssi
discordance	diskordanssi	divergence₂	poikkeama ¹¹
discordant pairs	diskordantit	do-calculus	do-laskenta
discounting	diskonttaus	domain	osajoukko ¹²
discrete	diskreetti	domain adaptation	kohdesopeutus
discrete distribution	diskreetti jakauma	domain of attraction	attraktioalue
discrete-time	diskreettiaikainen	donor imputation	luovuttajapaikkaus
discrete variable	diskreetti muuttuja	do-operator	do-operaattori
discriminant	diskriminantti	D-optimality	D-optimaalisuus
discriminant analysis	erotteluanalyysi	dose metameter	annosmuunnos, annosmetametri
discriminant function	erottelufunktio	dose-response curve	annos-vasetekäyrä
discriminatory power	erottelukyky	dot chart	pistekuvio
dispersion	hajonta	dot plot	pistekuvio, pistediagrammi
dispersion matrix	hajontamatriisi	dot product	(inner product) pis-

¹¹Esim. *Kullback-Leibler divergence* on suomeksi Kullbackin–Leiblerin poikkeama.

¹²Otantateoriassa *domain* viittaa perusjoukon kiinnostavaan osajoukkoon.

tetulo, sisätulo
double-blind experiment kaksoissoikkokoe
double exponential distribution (Laplace distribution) kaksoiseksponenttijakauma, Laplace-jakauma
double sampling (two-phase sampling) kaksivaiheotanta
doubly robust estimation kaksoisvakaa estimointi
doubly stochastic model kahdesti satunnainen malli
downward bias negatiivinen harha, alasharha
draftsman's plot kuviomatriisi
drift₁ vieto, suunta
drift₂ ajautuminen
dropout kato
d-separation *d*-separaatio
duality duaalisuus
dual scaling (correspondence analysis) korrepondenssianalyysi
dual systems estimation kaksoisrekisteriestimointi
dual theorem duaalilause
dummy variable (indicator variable) osoitin, indikaattorimuuttuja, ilmaisain
duration kesto
dynamic optimization (dynamic programming) dynaaminen optimointi, dynaaminen ohjelmointi
dynamic population (open population) dynaaminen populaatio, avoin väestö
dynamic programming (dynamic optimization) dynaaminen ohjelmointi, dynaaminen optimointi
dynamic treatment allocation dynaaminen kiintiöinti

E

ECM (error correction model)
ECM-algorithm ECM-algoritmi
ECME-algorithm ECME-algoritmi
ecological bias ekologinen harha
ecological correlation (geographical correlation) ekologinen korrelaatio
ecological fallacy (ecological bias) ekologinen harha, ekologisen virhepäätelmä
ecological study ekologinen tutkimus
econometrician ekonometrikko
econometrics ekonometria
ED50 (median effective dose)
EDA (exploratory data analysis)
edit rule editointisääntö
effect vaikutus₁, seuraus
effective degrees of freedom tehollinen vapausasteluku
effective sample size tehollinen otoskoko
effectiveness vaikuttavuus
effect magnitude vaikutuksen suuruus
effect modifier vaikutuksen muovaaja
effect of treatment on the treated (ETT) käsittelyvaikutus altistuneilla
effect size vaikutuskoko
effect variable (outcome variable) vastemuuttuja, tulosmuuttuja
efficacy teho
efficiency tehokkuus
efficient estimator tehokas estimaattori
efficient market hypothesis tehokkaiden markkinoiden hypoteesi

efficient score (score_1)	skoori
eigenspace	ominaisavaruus
eigenvalue	ominaisarvo
eigenvector	ominaisvektori
elasticity	jousto, elastisuus
elastic net	joustava verkko
ELBO (evidence lower bound)	
element	alkio
elementary contrast	alkeiskont-rasti
elementary event	alkeistapaus, alkeistapahtuma
elementary unit	alkeisyksikkö
elicitation of prior	priorin koos-taminen, esijakauman koostami-nen
elimination	poisto, eliminointi
elliptical distribution	elliptinen jakauma
EM-algorithm (expectation-maximization algorithm)	EM-algorit-mi
embedded process	upotettu pro-ssessi
emigration	maastamuutto
empirical	empiirinen, kokemuspe-räinen
empirical Bayes estimation	em-piirinen Bayes-estimointi
empirical coverage	empiirinen peittotodennäköisyys
empirical distribution function	empiirinen kertymäfunktio
empirical likelihood	empiirinen uskottavuus
empirical measure	empiirinen mitta
empirical size of test	testin em-piirinen koko
empty set	tyhjä joukko
encompassing model	kattava malli
endogeneity	endogeenisuus, sisä-määräytyneisyys
endogenous variable	(endogenous variate) endogeeni-nen muuttuja, sisämääräytynyt muuttuja
endpoint (outcome event)	pääte-tapahtuma, vastetapahtuma
engineering statistics	insinööri-tilastotiede
ensemble (statistical ensemble)	kooste, tilastollinen kooste
ensemble average	koostekeskia-rvo
ensemble learning	koosteoppimi-nen
ensemble prediction	kooste-en-nuste
entrance time (first passage ti-me)	saavutushetki, ensikulun het-ki
entropy	entropia, haje
enumeration	laskenta, luettelemi-nen
enumerator	väestönlaskija
envelope function	verhofunktio
environmetrics	ympäristötilastotiede
<i>E</i>-optimality	<i>E</i> -optimaalisuus
epidemic model	epidemiamalli
epidemiology	epidemiologia
equilibrium	tasapaino
equilibrium distribution	tasa-painojakauma
equivalence class	ekvivalenssi-luokka
equivalence testing	ekvivalenssi-testaus
equivariant estimator	ekviva-riantti estimaattori
erf (error function)	
erfc (complementary error func-tion)	
ergodic	ergodininen, keskiarvosup-

peneva	etiologic fraction etiologinen osuus
ergodicity ergodisuus	ETT (effect of treatment on the treated)
error virhe, erhe	EUBO (evidence upper bound)
error components model virhekomponenttimalli	Euclidean distance euklidinen etäisyys
error correction model (ECM) virheenkorjausmalli	European option eurooppalainen optio
error function (erf) virhefunktio	evaluation metric (operating characteristic) toimivuusmitta
error margin (margin of error) virhemarginaali	event tapahtuma
error matrix (classification table) luokittelutaulukko	event history analysis (life history analysis) tapahtumahistoria-analyysi
error mean square (residual mean square) virhekeskineliö, jäännöskeskineliö	event space tapahtuma-avaruus
error propagation (propagation of error) virheen kasautuminen, virheen leviäminen	evidence lower bound (ELBO) näytön alaraja, variaatioalaraja
error rate virheosuus	evidence upper bound (EUBO) näytön yläraja, variaatioyläraja
errors in variables model (measurement error model ¹³) virheet selittäjissä -malli, mittausvirhemalli	exact test eksakti testi
error sum of squares (residual sum of squares) jäännösneliösumma	exceedance curve ylitysarvokäyrä, ylityslukemakäyrä
error term virhetermi	exceedance test ylitysarvotesti, ylityslukematesti
error variance virhevarianssi	excess hazards model lisäuhkamalli
estimable estimoituva, arvioituva	excess risk lisäriski
estimand estimoitava suure, estimandi	exchangeability vaihdettavuus, vaihtokelpoisuus
estimate estimaatti, arvio	exchangeable variables keskenään vaihtokelpoiset muuttujat
estimating equation estimointiyhtälö, arviointiyhtälö	exclusion criterion poissulkukriteeri
estimating function estimointifunktio	exercise date lunastuspäivä
estimation estimointi, arviointi	exercise price lunastushinta
estimation error estimointivirhe, arviointivirhe	exercise value lunastusarvo
estimator estimaattori, arviofunktio	exogeneity eksogeenisuus, ulkomääräytyneisyys
	exogenous variable eksogeeni-

¹³Esim. monografiassa Fuller (1987) tämä käsite on laajempi.

nen muuttuja, ulkomääräytynyt muuttuja	explosive process räjähtävä prosessi
expectation (expected value) odotusarvo	exponential curve eksponenttikäyrä
expectation-maximization algorithm (EM-algorithm) EM-algoritmi	exponential dispersion family eksponenttinen hajontaperhe
expected frequency (expected number) odotettu lukumäärä	exponential dispersion model eksponenttinen hajontamalli
expected loss odotustappio	exponential distribution eksponenttijakauma
expected number (expected frequency) odotettu lukumäärä	exponential family eksponenttiperhe
expected shortfall (average Value at Risk) odotustappioarvo	exponential growth eksponenttinen kasvu, eksponentiaalinen kasvu
expected utility theory odotetun hyödyn teoria	exponential smoothing eksponenttinen tasoitus, eksponenttisen silotus
expected value (expectation) odotusarvo	exponential trend eksponenttinen trendi, eksponentiaalinen trendi
experiment (trial) koe	exposure altistus
experimental design koeasetelma	exposure time altistusaika
experimental error koevirhe	external validity ulkoinen validiteetti, ulkoinen pätevyys
experimental plot koeala, koe ruutu	extinction probability sukupuuton todennäköisyys
experimental study kokeellinen tutkimus	extra-Poisson variation ylipoissonvaihtelu
experimental unit koeyksikkö	extrapolation ekstrapolointi
expert judgment asiantuntija-arvio	extremal intensity ääriarvointensiteetti, äärilukemaintensiteetti
expert overconfidence asiantuntijaliikavarmuus	extremal quotient vaihtelusuhte
expert system asiantuntijajärjestelmä	extreme value ääriarvo, äärilukema
explanatory trial selittävä koe	extreme value distribution ääriarvojakautuma, äärilukemajakautuma
explanatory variable (regressor) selittävä muuttuja, selittäjä	
exploration vs. exploitation trade-off etsintä-hyödynnyskompromissi	
exploratory data analysis (EDA) etsivä data-analyysi	
exploratory survey esitutkimus, alustava kartoitus, esiluotaus	

F

face-to-face interview käyntihaastattelu

factor ₁ tekijä, selittävä luokkamuuttuja	false acceptance (acceptance error) hyväksymisvirhe
factor ₂ faktori	false discovery rate (FDR) väärin löydösten osuus
factor analysis faktorianalyysi	false negative väärä negatiivinen
factorial kertoma	false positive väärä positiivinen
factorial cumulant kertomakumulantti	false rejection (rejection error) hylkäysvirhe
factorial design tekijäkoeasetelma	familial correlation sukurkorrelaatio
factorial experiment tekijäkoe	family-wise error rate perhekohtainen virhetodennäköisyys
factorial moment kertomamomentti	fast Fourier transform (FFT) nopea Fourier-muunnos
factorial moment generating function kertomamomenttien emäfunktio, kertomamomentit generoiva funktio	father wavelet isäaaloke, isäväre
factorization tekijöihinjako	fatigue model väsymismalli
factorization criterion tekijöihinjakokriteeri, faktorointikriteeri	fat-tailed distribution ¹⁵ paksuhäntäinen jakauma
factorization theorem tekijöihinjakolause, faktorointilause	F-distribution (Fisher's distribution) <i>F</i> -jakauma, Fisherin jakauma
factor loading faktorilataus	FDR (false discovery rate)
factor matrix faktorimatriisi	feasibility study (pilot study) toteutettavuustutkimus
factor pattern faktorirakenne	feature piirre
factor reversal criterion tekijänvaihtokriteeri	feature selection piirrevalinta
factor rotation faktorirotaatio, faktorikierto	feature space piirreavaruus
factor score faktoripistemäärä	feed-forward neural network myötävirtaneuroverkko
failure vikaantuminen, vioittuminen	fertility hedelmällisyys
failure function vikakertymä	fertility rate hedelmällisyysluku
failure rate ₁ ¹⁴ vikataajuus	FFT (fast Fourier transform)
failure rate ₂ (hazard rate) uhkatilays	fiducial inference fidusiaalinen päättely
failure time vikaantumisaika	field experiment kenttäkoe
fair game reilu peli	field work kenttätö
	filter suodin, suodatin

¹⁴Ks. alaluku 2.8.

¹⁵Englannin kielen termillä "fat-tailed" ei ole vakiintunutta täsmällistä merkitystä. Useimmiten sitä käytetään termin "heavy-tailed" synonyyminä, toisinaan sillä viitataan potenssilain mukaiseen jakauman häntään ja joissain yhteyksissä jakaumiin, joilla on normaalijakaumaa suurempi kurtoosi. Suomen kielessä "paksuhäntäinen" on vakiintunut tarkoittamaan samaa kuin englannin kielen "heavy-tailed".

filtering suodatus	fixed sample kiinteä otos
filtration (history) filtraatio, historia	fixed variable kiinteä muuttuja
fine ordering hieno järjestys	floor effect lattiavaikutus
finite mixture distribution äärellinen jakaumasekoite	flow chart vuokaavio
finite population correction (fpc) äärellisyyskerroin, äärellisen perusjoukon korjauskerroin	<i>F</i> measure (<i>F</i> score) <i>F</i> -mitta
first digit distribution ensinumeron jakauma	folded contingency table taitettu kontingenssitaulukko, taitettu ristitaulukko
first-order inclusion probability yksikön sisältymistodennäköisyys	folded distribution taitettu jakauma
first order model ensimmäisen kertaluvun malli	folded median taitettu mediaani
first passage time (entrance time) saavutushetki, ensikulun hetki	folding taittaminen
first return time ensipaluun hetki	follow-up seuranta, seuruu
first stage sampling unit (primary sampling unit) ensias-teen otantayksikkö	force of mortality (hazard of death) kuolinhaka
Fisher consistency Fisher-tarkentuvuus	forecast ennuste ₁
Fisher information Fisher-informaatio	forecasting ennustaminen
Fisher information matrix Fisher-informaatiomatriisi	forensic statistics rikostekninen tilastotiede
Fisher's distribution (<i>F</i> -distribution) Fisherin jakauma	forest plot metsikkökuvio
Fisher's scoring method Fisherin estimointialgoritmi	forward contract termiinisopimus
fitted model sovitettu malli	forward equation eteenyhtälö
fitted value sovite, sovitettu arvo	forward selection etenevä valinta
fitting sovittaminen, sovitus	fourfold plot nelikenttäkuvio
five-number summary viiden luvun tiivistys	fourfold table (two-by-two table) nelikenttä, 2×2 -taulukko
fixed base index kiinteäkantainen indeksi	Fourier analysis Fourier-analyysi
fixed-effects model kiinteävaikutusmalli	Fourier transform Fourier-muunnos
	fpc (finite population correction)
	fractal fraktaali
	fractile (quantile) fraktiili, kvantili
	fractional Bayes factor fraktionaalinen Bayes-tekijä
	fractional Brownian motion fraktionaalinen Brownin liike
	fractional factorial experiment osittainen tekijäkoe
	fractional polynomial murtopolynomi

- fractional replication** osittainen toisto
- fraction defective** virhekappaleiden osuus
- frailty model** haurausmalli
- frame** kehikko, kehys
- frame population** kehikkoperusjoukko
- frequency**₁ (count) lukumäärä
- frequency**₂ (rate)₁ taajuus
- frequency**₃ (occurrence) esiintyvyys
- frequency curve** frekvenssikäyrä
- frequency data** (count data) lukumääräaineisto
- frequency distribution** lukumääräjakauma, frekvenssijakauma
- frequency domain** taajuusalue
- frequency histogram** frekvenssihistogrammi
- frequency matching** (group matching) ryhmäkaltaistus
- frequency response function** taajuusvastefunktio
- frequency surface** frekvenssipinta, taajuuspinta
- frequency table** lukumäärätaulukko, frekvenssitaulukko
- frequentist** frekventistinen
- frequentist inference** frekventistinen päättely
- frontdoor adjustment** etuoviva kiointi
- F score** (F measure) F -mitta
- F statistic** F -suure, F -tunnusluku
- F-test** F -testi
- full factorial design** täysitekijäkoe
- full information method** täyden informaation menetelmä
- fully conditional distribution**¹⁶ täysehdoollistettu jakauma
- functional central limit theorem** funktionaalinen keskeinen raja-arvolause
- functional data analysis** funktionaalinen data-analyysi
- functional determinant** (Jacobian determinant) funktionaalideterminantti, Jacobin determinantti
- functional relationship** funktionaalinen yhteys
- fund** rahasto
- funnel plot** suppilokuvio
- futures contract** futuurisopimus
- fuzzy logic** sumea logiikka
- fuzzy set theory** sumea joukkooppi

G

- gain**₁ vahvistus₁
- gain**₂ hyöty
- Galton–Watson process** Galtonin–Watsonin prosessi
- GAM** (generalized additive model)
- gambler’s fallacy** pelurin virhepäätelmä
- gambler’s ruin** pelurin vararikko
- game theory** peliteoria
- GAMM** (generalized additive mixed model)
- gamma coefficient** gammakeroin
- gamma distribution** gammajakauma
- gamma-frailty model** gamma-haurausmalli
- gamma function** gammafunktio

¹⁶Myös muotoa *full conditional distribution* käytetään.

gap statistic aukkosuure, aukko-tunnusluku

GARCH model (generalized autoregressive conditional heteroskedasticity model) GARCH-malli

Gaussian distribution (normal distribution) Gaussin jakauma, normaalijakauma

Gaussian elimination Gaussin eliminointimenetelmä

Gaussian mixture model (GMM_1) Gaussin sekoitemalli, GMM_1

Gaussian process Gaussin prosessi

Gaussian quadrature Gaussin kvadratuuri

Gauss–Markov theorem Gaussin–Markovin lause

GCV (generalized cross-validation)

GEE (generalized estimating equation)

gene mapping geenikartoitus

general factor yleisfaktori

generalizability yleistettävyyys

generalized additive mixed model (GAMM) yleistetty additiivinen sekamalli, yleistetty summautuva sekamalli, GAMM

generalized additive model (GAM) yleistetty additiivinen malli, yleistetty summautuva malli, GAM

generalized autoregressive conditional heteroskedasticity model (GARCH model) yleistetty autoregressiivinen ehdollisen heteroskedastisuuden malli, GARCH-malli

generalized cross-validation (GCV) yleistetty ristiinvalidointi, yleistetty ristiinvahvistus, GCV

generalized estimating equation (GEE) yleistetty estimointiyhtälö, GEE

generalized extreme-value distribution yleistetty ääriarvojakaua

generalized gamma distribution yleistetty gammajakauma

generalized inverse yleistetty käänteismatriisi

generalized least squares (GLS) yleistetty pienin neliösumma, GLS

generalized linear mixed model (GLMM) yleistetty lineaarinen sekamalli, GLMM

generalized linear model (GLM) yleistetty lineaarinen malli, GLM^{17}

generalized method of moments (GMM_2) yleistetty momenttimenetelmä, GMM_2

generalized normal distribution yleistetty normaalijakauma

generalized regression estimation (GREG estimation) yleistetty regressioestimointi, GREG-estimointi

generalized stochastic process yleistetty stokastinen prosessi

general linear model yleinen lineaarinen malli

generating function emäfunktio, generoiva funktio

generator of process prosessin generaattori, prosessin kehitin

genetic algorithm geneettinen algoritmi

genetic heterogeneity geneetti-

¹⁷Joissakin yhteyksissä tällä lyhenteellä viitataan termiin *general linear model*.

nen heterogeenisuus

genomewide association study (GWAS) genomilaajuinen assosiaatiokartoitus, GWAS

geographical information system (GIS) paikkatietojärjestelmä

geometric distribution geometrinen jakauma

geometric mean geometrinen keskiarvo

geometric probability geometrisen todennäköisyys

geostatistics geotilastotiede

g-estimation g-estimointi

g-formula g-kaava

Gibbs sampler Gibbsin poimija, Gibbsin algoritmi

Gini index Gini-indeksi, Gini-kerroin

GIS (geographical information system)

GLM (generalized linear model)

GLMM (generalized linear mixed model)

GLS (generalized least squares)

GMM₁ (Gaussian mixture model)

GMM₂ (generalized method of moments)

goodness of fit yhteensopivuus, soviteen hyvyys

graduation curve tasoituskäyrä

Graeco-Latin square kreikkalaislatinalainen neliö

Gramian Gramin matriisi

grand mean kokonaiskeskiarvo, yleiskeskisarvo

graph₁ kuvio

graph₂ kuvaaja

graph₃ verkko, graafi

graphical estimator graafinen estimaattori

graphical model verkkomalli, graafinen malli

GREG estimation (generalized regression estimation) GREG-estimointi

grey box model harmaan laatikon malli

grid hila, ruudukko

grid sampling ruutuotanta

gross migration bruttomuutto

gross reproduction rate (GRR) bruttouusiutumisluku

group divisible design ryhmäjaollinen asetelma

grouped data ryhmitelty aineisto

group factor ryhmäfaktori

grouping lattice luokitushila

group matching (frequency matching) ryhmäkaltaistus

group screening design ryhmäseulonta-asetelma

group-sequential design ryhmäperäkkäisasetelma

group-sequential experiment ryhmäperäkkäinen koe

growth curve kasvukäyrä

growth rate kasvunopeus

GRR (gross reproduction rate)

Guttman scale Guttmanin asteikko

GWAS (genomewide association study)

H

H₀ (null hypothesis)

H₁ (alternative hypothesis)

half-normal distribution (folded normal distribution) puolinormaalijakauma, taitettu normaalijakauma

halo effect halovaikutus

Hamming code Hamming-koodi

Hamming distance Hamming-etäisyys

hanging rootogram	roikkukuva	sekakoosteinen
histogrammi		
Hann window ¹⁸	Hann-ikkuna	heteroskedastic heteroskedastinen, erivarianssinen
hard-core model	kovaydinmalli	heteroskedasticity-consistent
HARKing	jälkihypotisointi	t-ratio heteroskedastisuus-tarkentuva <i>t</i> -suhde
harmonic distribution	harmonin jakauma	hidden Markov model (HMM) piilo-Markov-malli
harmonic mean	harmoninen keskiarvo	hierarchical classification (nested classification) hierarkkinen luokittelu, sisäkkäinen luokittelu
harmonic regression	harmonin regressio	hierarchical clustering hierarkkinen ryhmittely
hat matrix	hattumatriisi	hierarchical model (multilevel model) hierarkkinen malli, monitasomalli
hazard	uhka, hasardi	high-dimensional data suuriulotteinen aineisto
hazard function	(hazard rate) uhkafunktio, uhkatiheys	higher-order asymptotics korkeamman kertaluvun asymptotiikka
hazard of death	(force of mortality) kuolintuho	highest posterior density interval (HPD interval) suurimman posterioritiheyden väli, HPD-väli
hazard rate	(hazard function) uhkatiheys, uhkafunktio	highest posterior density region (HPD region) suurimman posterioritiheyden alue, HPD-alue
hazard ratio	(HR) uhkasuhde, HR	high-throughput data isovuodatta
heads	kruuna	high-throughput screening isovuoseulonta, sarjaseulonta
headship rate method	perheenpääkohtainen menetelmä	hill-climbing algorithm vuorikiipeilyalgoritmi
heat map	lämpökartta	hinge (quartile) kvartiili
Heaviside function	Heavisiden funktio	histogram histogrammi
heavy-tailed distribution	pak-suhäntäinen jakauma	history (filtration) historia
hedging strategy	suojautumisstrategia	hitting probability osumatodennäköisyys
hedonic regression	hedoninen regressio	hitting time osuma-aika, osumahetki
Helmert contrast	Helmert-kontrasti	
herd immunity	laumasuoja, laumaimmuneetti	
heritability	periytyvyys, heritabiliteetti	
Hessian	Hessen matriisi	
heterogeneous	heterogeeninen,	

¹⁸Termiä *Hanning window* käytetään myös, mutta se ei ole suositeltava.

HMM (hidden Markov model)
homogeneous homogeeninen, tasakoostainen
homoskedastic homoskedastinen, samavarianssinen
horseshoe effect hevosenkenkäälmio
hot deck imputation kuuma-paikkaus, kuumaimputointi
household kotitalous
HPD interval (highest posterior density interval) HPD-väli
HR (hazard ratio)
hurdle model (zero-altered model) rimamalli
hyperexponential distribution hypereksponttijakauma
hypergeometric distribution hypergeometrinen jakauma
hyperparameter hyperparametri
hypothesis hypoteesi, oletus
hypothesis testing hypoteesin testaus

I

ICA (independent component analysis)
ideal index ideaali-indeksi, ihanneindeksi
idempotent idempotentti
identifiability identifioituvuus
identifier tunnistin
identity function identiteettifunktio
identity matrix yksikkömatriisi
idle period joutojakso, lepojakso
IFR (infection fatality ratio)
ignorable sivuutettava
i.i.d. (independent and identically distributed)
ill-conditioned matrix lähes vajaa-asteinen matriisi, häiriöaltis

matriisi
illness-death model sairaus-kuolemamalli
ill-posed problem huonosti asetettu ongelma
IMA model (integrated moving average model) IMA-malli
image analysis kuva-analyysi
image processing kuvankäsittely
immigration maahanmuutto
implicit stratification implisiittinen ositus
importance sampling painotusotanta, tärkeysotanta
improper distribution epäaito jakauma
improper imputation epäkelpo imputointi
improper prior epäaito prior, epäaito esijakauma
impulse response function impulssivastefunktio, sysäysvaste-funktio
imputation paikkaus, imputointi
imputation model paikkausmalli, imputointimalli
inadmissible estimator epäkel-po estimaattori
incidence ilmaantuvuus
incidence matrix₁ indikaattorimatriisi, osoitinmatriisi
incidence matrix₂ kytkentämatriisi
incidence proportion (cumulative incidence) ilmaantumisosuus, kumulatiivinen ilmaantuvuus
incidence rate ilmaantumistiheys
incidental parameters tilanneparametrit
inclusion criterion mukaanotto-kriteeri
inclusion probability sisältymis-

todennäköisyys	indirect sampling epäsuora otanta
incomplete block design vajaalohkoasetelma	indirect standardization epäsuora vakiointi
incomplete Latin square vajaa latinalainen neliö	induction induktio
inconsistent estimator tarkentumaton estimaattori	inductive behaviour induktiivinen käyttäytyminen
independence riippumattomuus	inductive inference induktiivinen päättely
independent and identically distributed (i.i.d.) riippumattomat ja samoin jakautuneet, i.i.d.	inefficient estimator vajaatehoinen estimaattori
independent censoring riippumaton sensurointi	inefficient markets tehottomat markkinat
independent component analysis (ICA) riippumattomien komponenttien analyysi, ICA	inequality coefficient eriarvoisuuskero
independent increments riippumattomat lisäykset	infection fatality ratio (IFR) infektiokuolleisuus
independent variable ¹⁹ (explanatory variable) selittävä muuttuja, selittäjä	inference päättely
index (index number) indeksi, indeksiluku	infinitely decomposable äärettömästi osittuva
index number (index) indeksiluku, indeksi	infinitely divisible rajatta jakautuva, äärettömästi jakautuva
index of concentration keskittymisindeksi	inflation inflaatio
index of dispersion (dispersion index) hajontaindeksi	influence vaikutus ₂
index of dissimilarity erilaisuusindeksi	influence function vaikutusfunktio
index of diversity monimuotoisuusindeksi, diversiteetti-indeksi	influence statistic vaikutustunnusluku
index plot indeksikuvio	influential observation vaikuttava havainto
Indian buffet process intialainen noutopöytäprosessi	information informaatio
indicator variable osoitin, indikaattorimuuttuja, ilmaisun	information criterion informaatiokriteeri
indirect effect epäsuora vaikutus	information function informaatiotiofunktio
indirect least squares epäsuorapienin neliösumma	information matrix informaatiomatriisi
	information theory informaatioteoria
	informative censoring informatiivinen sensurointi

¹⁹Suosittelava termi on *explanatory variable*.

informative missingness	informatiivinen puuttuvuus	intelligence quotient (IQ)	älykkyysosamäärä
informative prior	informatiivinen prior, informatiivinen esija-kauma	intensity (rate_1)	intensiteetti, taajuus
informed consent	tietoinen suostumus	intensity function	intensiteettifunktio
initial distribution	alkujakauma	intention-to-treat analysis	hoitoaikeen mukainen analyysi, ITT-analyysi
inlier	piilokki	interaction	yhdyisvaikutus, interaktio
in-migration	sisäänmuutto	interarrival time	(interevent time) saapumisväliaika
inner fence	sisäaita	interblock	lohkojen välinen
inner product (dot product)	sisätulo, pistetulo	intercept	vakiotermin
innovation gain	innovaatiohyöty	inter-class similarity	luokkien välinen samankaltaisuus
innovation term	satunnaistermi, innovaatiotermi	interest rate	korkotaso
input-output matrix (IO matrix)	panos-tuotostmatriisi	interevent time	(interarrival time) tapahtumaväliaika
input-output model	panos-tuotostmalli	interim analysis	välialalyysi
inquiry	tutkimus, selvitys, kysely	internal validity	sisäinen validiteetti
instance	tapaus2	internal variance	sisäinen varianssi
instance space	tapausavaruus	internet survey	verkkokysely, nettikysely
instantaneous death-rate (force of mortality)	kuolemanuhka	interpenetrating samples	(network of samples) kytketyt otokset, otosverkosto
instantaneous state	hetkellinen tila	interpolating spline	interpoloiva splini
instrumental variable	välinemuuttuja, instrumenttimuuttuja	interpolation	interpolaatio
insurance	vakuutus	interquartile range (IQR)	kvartiliväli
integrable	integroituva	intersection	leikkaus
integral	integraali	interval censoring	välisensurointi
integrated hazard (cumulative hazard)	uhkakertymä	interval estimation	väliestimointi
integrated intensity (cumulative intensity)	intensiteettikertymä	interval scale	välimatka-asteikko, intervalliasteikko
integrated moving average model (IMA model)	integroitu liukuvan keskiarvon malli, IMA-malli	intervention analysis	interven-
integrated spectrum	spektrikerätymäfunktio, integroitu spektri		

tioanalyysi, toimenpideanalyysi
intervention study interventio-
 tutkimus
interviewer bias haastattelijahar-
 ha
intra-block lohkonsisäinen
intraclass correlation sisäkorre-
 laatio
intra-class similarity luokkien
 sisäinen samankaltaisuus
intraclass variance sisävarianssi
intrinsic growth rate sisäinen
 kasvuvauhti
invariance invarianssi
invariant invariantti
invariant distribution invariant-
 ti jakauma
inverse Bernoulli sampling
 käänteinen Bernoulli-poiminta
inverse chi-square distribution
 käänteinen khi-neliöjakauma
inverse gamma distribution
 käänteinen gammajakauma
inverse Gaussian distribution
 käänteinen normaalijakauma,
 käänteinen Gaussin jakauma
inverse matrix käänteismatriisi
inverse multinomial sampling
 käänteinen multinomipoiminta
inverse probability (posterior
 probability) käänteistodennäköi-
 syys, posterioritodennäköisyys
inverse probability weighting
 käänteistodennäköisyyspainotus
inverse problem inversio-
 ongelma, käänteisongelma
inverse sampling käänteisotanta
inverse transformation method
 käänteismuunnosmenetelmä
inversion inversio, käänteismuun-
 nos
inverted beta distribution
 käänteinen betajakauma
inverted Dirichlet distribution
 käänteinen Dirichlet-jakauma
invertible kääntyvä
invertible process kääntyvä pro-
 sessi
IO matrix (input-output matrix)
IO model IO-malli
IPF (iterative proportional fitting)
IQ intelligence quotient
IQR (interquartile range)
irreducible chain pelkistymätön
 ketju
IRT (item-response theory)
isometry isometria
isotonic regression isotoninen
 regressio
isotropic isotrooppinen
item analysis osioanalyysi
item non-response eräkato,
 muuttujakato
item-response theory (IRT)
 osiovasteteoria, IRT
iterated bootstrap iteroitu
 uusio-otanta
iteration iteraatio, iterointi
iterative proportional fitting
 (IPF) iteratiivinen verrannollinen
 sovitus
**iteratively weighted least squa-
 res** (IWLS) iteratiivisesti paino-
 tettu pienin neliösumma
Itô integral Itô-integraali
ITT analysis (intention-to-treat
 analysis) hoitoaikeen mukainen
 analyysi, ITT-analyysi
IWLS (iteratively weighted least
 squares)

J

jackknife estimator linkkuveit-
 siestimaattori, jackknife-estimaat-
 tori

jackknife method linkkuveitsimenetelmä
Jacobian determinant (functional determinant) Jacobin determinantti, funktionaalideterminantti
Jacobian matrix Jacobin matriisi
James–Stein estimator Jamesin–Steinin estimaattori
Jeffreys prior Jeffreysin prior
Jensen’s inequality Jensenin epäyhtälö
jittered sampling ravistettu otanta
jittering (dithering) ravistaminen
joint confidence region yhteisluottamusalue
joint cumulative distribution function yhteiskertymäfunktio
joint density function yhteistiheysfunktio
joint distribution yhteisjakauma
jointly sufficient statistics yhteisesti tyhjjentävät tunnusluvut
joint moment yhteismomentti
joinpoint regression liitospiste-regressio
joint prediction region yhteisennustealue
joint probability yhteistodennäköisyys
joint regression yhteisregressio
J-shaped distribution J:n muotoinen jakauma
judgment sample harkintanäyte
jump probability hyppytodennäköisyys
jump rate hyppyaajuus
jumping distribution (proposal distribution) ehdotusjakauma, ehdokasjakauma

jurimetrics oikeustilastotiede, jurimetria
just identified täsmälleen identifioituva

K

Kalman filter Kalman-suodin
Kalman gain Kalman-hyöty
Kaplan–Meier estimator (product limit estimator) Kaplanin–Meierin estimaattori, rajatuloestimaattori
kappa coefficient kappakerroin
k-class estimator k -luokan estimaattori
Kendall’s tau (coefficient of dissimilarity) Kendallin tau
kernel density estimator ydintiheysestimaattori
kernel estimator ydineestimaattori
kernel function ydinfunktio
kernel mean function ydinodotusarvofunktio
kernel method ydinmenetelmä
kernel regression estimator ydinregressioestimaattori
kernel smoothing ydintasointus, ydinsiloutus
k-means clustering k :n keskiarvon ryvästys
knot solmupiste
knowledge technology (knowledge engineering) tietämystekniikka
kriging method kriging-menetelmä
Kronecker delta Kroneckerin delta
Kronecker product Kroneckerin tulo
k-statistic k -tunnusluku

Kullback–Leibler divergence

Kullbackin–Leiblerin poikkeama

kurtosis₁ kurtoosi²⁰

kurtosis₂ (peakedness) huipukkuus

L

L^1 metric L^1 -etäisyys

L^1 norm L^1 -normi

L^2 distance L^2 -etäisyys

L^2 norm L^2 -normi

labeled data nimiöity data

labeling rule nimeämissääntö

labor-force survey työvoimatutkimus

lack of memory property muistittomuus, unohtavaisuus

LAD (least absolute deviation)

ladder index tikasindeksi

ladder variable tikasmuuttuja

lag viive²¹

lag operator (backshift operator) viiveoperaattori

Lagrange multiplier test (score test) Lagrangen kerrointesti

lag window viiveikkuna

landmark analysis virstanpylväsanalyysi

Laplace distribution (double exponential distribution) Laplace-jakauma

Laplace transform Laplace-muunnos

large deviation theory suurten poikkeamien teoria

large sample theory suurten otosten teoria

Laspeyres' index Laspeyresin indeksi

LASSO (least absolute shrinkage

and selection operator)

last observation carried forward (LOCF) viimeisen havainnon toisto, LOCF-menetelmä

LATE (local average treatment effect)

latent class analysis piiloluokkanalyysi

latent structure piilorakenne

latent trait analysis piilopiirreanalyysi

latent variable piilomuuttuja

Latin cube latinalainen kuutio

Latin hypercube sampling latinalainen hyperkuutio-otanta

Latin rectangle latinalainen suorakulmio

Latin square latinalainen neliö

lattice hila

lattice design hila-asetelma

lattice distribution hilajakauma

lattice sampling hilaotanta

lattice square hilaneliö

law of large numbers suurten lukujen laki

law of small numbers pienten lukujen laki

law of the iterated expectations iteroitujen odotusarvojen laki

law of the iterated logarithm (LIL) iteroidun logaritmin laki

law of truly large numbers todella suurten lukujen laki

linear discriminant analysis (LDA) lineaarinen erotteluanalyysi

LD (linkage disequilibrium)

LD50 (median lethal dose)

LDA (linear discriminant analysis)

leading indicator ennakoiva indi-

²⁰Kun *kurtosis* viittaa nimenomaisesti κ -parametriin (alaluku 2.6).

²¹Taivutetaan “viipeen” tai “viiveen”.

kaattori, ennakoiva osoitin	lyysi
leading question johdatteleva kysymys	life cycle utility elinkaarihyöty
leaps-and-bounds algorithm harpaten valinta	life expectancy elinajanodote
least absolute deviation method (LAD) pienimmän itseispoikkeaman menetelmä	life insurance henkivakuutus
least absolute shrinkage and selection operator (LASSO) LASSO, pienimmän itseiskutistamisen ja valinnan operaattori	life table elinajantaulukko
least favorable distribution epäsuotuisin jakauma	lifetime elinaika
least significant difference test (LSD test) vähiten merkitsevän erotuksen testi, LSD-testi	lifetime data analysis (survival analysis) elinaika-analyysi
least squares estimator (LSE) pienimmän neliösumman estimaattori, PNS-estimaattori	lifetime distribution elinaikajakauma, elinajan jakauma
least squares method (method of least squares) pienimmän neliösumman menetelmä, PNS-menetelmä	light-tailed distribution ohut-häntäinen jakauma
least variance difference method pienimmän varianssin erotusmenetelmä	likelihood uskottavuus
leave-one-out cross-validation (LOOCV) yksi-pois-ristiinvalidointi, yksittäisristiinvalidointi	likelihood function uskottavuusfunktio
left censoring sensurointi vasemmalta	likelihood principle uskottavuusperiaate
left truncation katkaisu vasemmalta, typistys vasemmalta	likelihood ratio uskottavuusosamäärä
length-biased sampling pituus-harhainen otanta	likelihood ratio statistic uskottavuusosamääräsuure, uskottavuusosamäärätunnusluku
level map tasokartta	likelihood ratio test (LRT) uskottavuusosamäärätesti
level of factor tekijän taso	likely (probable) todennäköinen
level of significance (significance level) merkitsevyystaso	Likert scale Likertin asteikko
leverage vipuvoima	LIL (law of the iterated logarithm)
Lexis diagram Lexis-kuvio	l.i.m. (limit in the mean)
life course analysis elinkaariana-	limit in probability (plim) stokastinen raja-arvo
	limit in the mean (l.i.m.) keskineliöraja-arvo
	limited dependent variable rajoitettu vaste
	limited information estimator rajoitetun informaation estimaattori
	limiting distribution rajajakautuma
	linear lineaarinen
	linear constraint lineaarinen rajoite

linear dependence	lineaarinen riippuvuus	(complier average treatment effect) keskimääräinen käsittelyvaikutus sitoutuneilla
linear estimator	lineaarinen estimaattori	
linear filter	lineaarinen suodin	
linearized	linearisoitu	
linear model	lineaarinen malli	
linear optimization	(linear programming) lineaarinen optimointi, lineaarinen ohjelmointi	
linear predictor	lineaarinen ennustin	
linear probability model	lineaarinen todennäköisyysmalli	
linear process	lineaarinen prosessi	
linear programming	(linear optimization) lineaarinen optimointi, lineaarinen ohjelmointi	
linear regression	lineaarinen regressio	
linear trend	lineaarinen trendi	
line chart	(line plot) viivakuvi	
line-intersect sampling	linjaleikkausotanta	
line of equal distribution	(line of equidistribution) tasaisen jakautumisen suora, tasa-arvosuora	
line plot	(line chart) viivakuvi	
line sampling	linjaotanta	
line spectrum	viivaspektri	
linkage analysis	kytkentäanalyysi	
linkage disequilibrium	(LD) kytkentäepätasapaino, LD	
linked blocks	kytketyt lohkot	
linked samples	kytketyt otokset	
link function	linkkifunktio	
liquidity	likviditeetti	
local asymptotic efficiency	paikallinen asymptoottinen tehokkuus	
local average treatment effect		
		(complier average treatment effect) keskimääräinen käsittelyvaikutus sitoutuneilla
local likelihood	paikallisuuskottavuus	
locally asymptotically most powerful test	paikallisesti asymptoottisesti voimakkain testi	
locally estimated scatterplot smoothing	(LOESS) paikallisesti estimoitu sirontakuviosilotus, LOESS	
locally weighted regression	paikallisesti painotettu regressio	
locally weighted scatterplot smoothing	(LOWESS) paikallisesti painotettu sirontakuviosilotus, LOWESS	
local regression	paikallinen regressio	
location family	sijaintiperhe	
location parameter	sijaintiparametri	
location-scale family	sijaintiskaalaperhe	
location shift	sijainnin siirto	
LOCF	(last observation carried forward)	
LOESS	(locally estimated scatterplot smoothing)	
logarithmic scale	logaritminen asteikko	
logarithmic series distribution	logaritmisen sarjan jakauma	
logarithmic transformation	logaritmimuunnos	
logistic curve	logistinen käyrä	
logistic distribution	logistinen jakauma	
logistic regression	logistinen regressio	
logit model	logit-malli	
logit transformation	logit-muun-	

nos
log-likelihood function (support function) log-uskottavuusfunktio, tukifunktio
loglinear model log-lineaarinen malli
loglog-transformation loglog-muunnos
lognormal distribution log-normaali jakauma
log-odds log-vastasuhde, log-veto-kerroin
logrank test logrank-testi
logworth log-arvo²²
longitudinal data pitkittäisaineisto, pitkittäisdata
long-memory process pitkämuistinen prosessi
long-range dependence pitkän kantaman riippuvuus
long-tailed distribution pitkähäntäinen jakauma
LOOCV (leave-one-out cross-validation)
lookup table hakutaulu
loop silmukka
Lord's paradox Lordin paradoksi
Lorenz curve Lorenzin käyrä
loss tappio
loss aversion tappion karttaminen
loss function tappiofunktio
loss matrix tappiomatriisi
loss triangle tappiokolmio
lot erä
lower control limit alavalvontaraja
lower quartile alakvartiili
lower triangular matrix alakolmiomatriisi
LOWESS (locally weighted scat-

terplot smoothing)
LRT (likelihood ratio test)
LSD test (least significant difference test) LSD-testi
LSE (least squares estimator)

M

machine intelligence koneäly
machine learning koneoppiminen
MAD (mean absolute deviation)
MAE (mean absolute error)
magic square design taikaneliöasetelma
Mahalanobis distance Mahalanobisin etäisyys
main effect päävaikutus
Mallows' C_p statistic Mallowsin C_p -tunnusluku
MA model (moving average model) MA-malli
Manhattan distance (city block distance) korttelietäisyys, manhattanetäisyys
manifest variable ilmimuuttuja
manifold monisto
manifold learning monisto-oppiminen
Mann-Whitney test Mannin-Whitneyn testi
MANOVA (multivariate analysis of variance)
Mantel-Haenszel test
 Mantelin-Haenszelin testi
MAP estimator (maximum a posteriori estimator) MAP-estimaattori
MAR (missing at random)
marginal association reunayhteys

²²Tämä on p -arvon muunnos $-\log_{10}(p)$.

marginal causal effect	reuna- kausaalivaikutus	Markov process	Markov-prosessi, Markovin prosessi
marginal density function	reunatiheysfunktio	Markov property	Markov-ominaisuus
marginal dependence	reunariippuvuus	Markov random field	Markovin satunnaiskenttä
marginal distribution	reunajakauma, marginaalijakauma	Markov switching model	Markovin vaihtomalli
marginal effect	reunavaikutus	mark-recapture sampling	(capture-recapture sampling) merkintä-takaisinpyyntiotanta, toistopyyntiotanta
marginal frequency	reunalukumäärä, reunafrekvenssi	MARS	(multivariate adaptive regression splines)
marginal homogeneity	reunahomogeenisuus, marginaalihomogeenisuus	martingale	martingaali
marginal independence	reunariippumattomuus	martingale measure	martingaalimitta
marginal likelihood	reunauskottavuus, marginaalioskottavuus	martingale residual	martingaali-jäännös, martingaaliresiduaali
marginal model	reunamalli, marginaalimalli	masking effect	peittovaikutus
marginal probability	reunatodennäköisyys	master sample	perusotos
marginal structural model	(MSM) reunarakennemalli, marginaalirakennemalli	matched case-control study	kaltaistettu tapaus-verrokki-tutkimus
marginal table	reunataulukko	matched pair	kaltaistettu pari
marginal total	reunasumma	matched samples	(paired samples) kaltaistetut otokset
margin of error	(error margin) virhemarginaali	matched set	kaltaistusjoukko
Marimekko plot	(mosaic plot) marimekkokuvio, mosaiikkikuvio	matching	kaltaistus
marked point process	merkkipiste prosessi	mathematical statistics	matemattinen tilastotiede
Markov chain	Markovin ketju, Markov-ketju	matrix algebra	matriisilaskenta, matriisialgebra
Markov chain Monte Carlo method	(MCMC method) Markovin Monte Carlo -menetelmä, MCMC-menetelmä	matrix decomposition	matriisihajotelma
Markovian	markovinen, markov ²³	matrix factorization	matriisin tekijöihinjako
		matrix normal distribution	matriisinormaalijakauma
		matrix sampling	matriisiotanta
		matrix variate regression	mat-

²³Ks. alaluku 2.5.

riisivasteregressio	kineliövirhe
maturity date eräpäivä	mean time to failure (MTTF) keskimääräinen vikaantumisaika
maximum a posteriori estimator (MAP estimator) suurimman posterioritiheyden estimaattori, MAP-estimaattori	mean time between failures (MTBF) keskimääräinen vikaantumisväli
maximum entropy method suurimman entropian menetelmä	mean vector keskiarvovektori
maximum likelihood estimator (MLE) suurimman uskottavuuden estimaattori, SU-estimaattori, ML	measurable mitallinen
matched case-control study kaltaistettu tapaus-verrokkitutkimus-estimaattori	measure mitta
maximum likelihood method suurimman uskottavuuden menetelmä, SU-menetelmä, ML-menetelmä	measurement error mittausvirhe
maximum tolerated dose (MTD) suurin siedetty annos	measurement error model mittausvirhemalli
MCAR (missing completely at random)	measurement scale mitta-asteikko
MCMC (Markov Chain Monte Carlo)	measure of association yhteysmitta, assosiaatiomitta
MDL (minimum description length)	measure of central tendency (measure of location) keskiluku, sijaintimitta, sijaintiluku
MDS (multidimensional scaling)	measure of dispersion (measure of variation) hajontamitta, hajontaluku, vaihtelumitta
mean ₁ keskiarvo	measure of location (measure of central tendency) sijaintimitta, sijaintiluku, keskiluku
mean ₂ (expected value) odotusarvo	measure of variation (measure of dispersion) vaihtelumitta, hajontamitta, hajontaluku
mean absolute deviation (MAD) keskipoikkeama	measure theory mittateoria
mean absolute error (MAE) keski-itseisvirhe	median mediaani
mean integrated squared error (MISE) integroitu keskineliövirhe	median effective dose (ED50) puoleen tehoava annos, ED50
mean residual life jäljellä olevan elinajan odote	median effective level puoleen tehoava taso
mean reversion odotusarvoon palautuva	median lethal dose (LD50) puolet tappava annos, LD50
mean square continuity keskineliöjatkuvuus	median lethal level puolet tappava taso
mean squared error (MSE) kes-	median regression curve mediaaniregressiokäyrä
	median survival time mediaanielinaika

median test	mediaanitesti	minimax criterion	minimax-kriteeri
median unbiasedness	mediaaniharhattomuus	minimax estimation	minimax-estimointi
mediation	välitys, mediaatio	minimum chi-square	pienin khineliö
mediation analysis	välitysanalyysi, mediaatioanalyysi	minimum description length principle	(MDL) lyhimmän kuvauspitäyden periaate
mediator	välittävä tekijä	minimum norm quadratic unbiased estimator	(MINQUE) kvadraattinen, harhaton miniminormiestimaattori, MINQUE-estimaattori
medical statistics	lääketieteellinen tilastotiede	minimum variance unbiased estimator	(MVUE) pienimmän varianssin harhaton estimaattori
memorylessness property	muistamattomuusominaisuus	MINQUE	(minimum norm quadratic unbiased estimator)
<i>M</i>-estimator	<i>M</i> -estimaattori	misclassification	väärinluokittelu, virheluokittelu
meta-analysis	meta-analyysi	MISE	(mean integrated squared error)
metameter	muunnosarvo, metametri	missing at random	(MAR) satunnaisesti puuttuva, MAR
meta regression	metaregressio	missing by design	suunnitellusti puuttuva
method of least squares	(least squares method) pienimmän neliösumman menetelmä	missing completely at random	(MCAR) täysin satunnaisesti puuttuva, MCAR
method of maximum likelihood	(maximum likelihood method) suurimman uskottavuuden menetelmä	missing data	puuttuva data
method of moments	momenttimenetelmä	missing data pattern	puuttuvuusrakenne
methodology	metodologia	missing information principle	puuttuvan informaation periaate
Metropolis–Hastings algorithm	Metropolisin–Hastingsin algoritmi	missing not at random	(MNAR) ei-satunnaisesti puuttuva
MGF	(moment generating function)	misspecification	väärinspesifointi
mid-mean	keskipuolikkaan keskiarvo	misspecified model	väärin spesioitu malli
mid-<i>p</i>-value	keski- <i>p</i> -arvo	mixed distribution	seka jakauma
mid-range	vaihteluvälin keskipiste, kesko		
mid-rank	keskisijaluku		
Mills ratio	Millsin suhde		
minimally connected design	minimaalisesti kytketty asetelma		
minimal sufficient statistic	minimaalinen tyhjentävä tunnusluku, karkein tyhjentävä tunnusluku		

mixed-effects model (mixed model) sekamalli	monotone monotoninen
mixed model (mixed-effects model) sekamalli	monotone likelihood ratio monotoninen uskottavuusosamäärä
mixed spectrum sekaspektri, yhdistetty spektri	Monte Carlo method Monte Carlo -menetelmä
mixed strategy sekastrategia	monthly average kuukausikeskiarvo
mixing sekoittava, sekoittavuus	moralization of graph verkon moralisointi
mixing distribution sekoittava jakauma	morphometrics muotoanalyysi, morfometria
mixture distribution jakauma-seos, jakaumien sekoitus	mortality rate (death rate) kuolevuus
mixture experiment sekoitekoe	mosaic plot (Marimekko plot) mosaiikkikuvio, marimekkokuvio
mixture model sekoitemalli	most accurate confidence interval lyhyin luottamusväli
MLE (maximum likelihood estimator)	most powerful test voimakkain testi
MNAR (missing not at random)	most stringent test tiukin testi
mode moodi, tyyppi-arvo	mother wavelet äitiaalloke, äitiväre
model malli	mover-stayer model liikkuja-pysjämalli
model-assisted malliavusteinen	moving average liukuva keskiarvo
model averaging mallikeskiarvoistus	moving average model (MA model) liukuvan keskiarvon malli, MA-malli
model-based inference malliperusteinen päättely	moving range liukuva vaihteluväli
model building mallinrakentaminen	MSE (mean squared error)
model diagnostics mallidiagnostiikka	MSM (marginal structural model)
model error mallivirhe	MTBF (mean time between failures)
model fit mallin sopivuus	MTD (maximum tolerated dose)
modeling mallinnus, mallitus	MTTF (mean time to failure)
model matrix mallimatriisi	multiarmed bandit monikätkäinen rosvo
modified profile likelihood muokattu profiiluskottavuus	multicategory logit model (multinomial logit model) multinomilogit-malli
moment momentti	multicenter study monikeskus-
moment estimator momenttimestimaattori	
moment generating function (MGF) momenttiemäfunktio, momentit generoiva funktio	
moment matrix momenttimatriisi	
moment ratio momenttisuhte	

tutkimus	väskuvio, ositettu palkkikuvio
multicenter trial monikeskuskoe	multiple comparison moniver-
multicollinearity (collinearity)	tailu
kollineaarisuus, multikollineaari-	multiple correlation coefficient
suus	yhteiskorrelaatiokerroin
multidimensional moniulottei-	multiple criteria decision ma-
nen	king monitavoitteinen päätöksen-
multidimensional scaling	teko
(MDS) moniulotteinen skaalaus	multiple criteria optimization
multiepisode model monijakso-	monitavoiteoptimointi
malli	multiple decrement model mo-
multiequational model moniyh-	nipoistumamalli
tälömalli	multiple factor analysis (factor
multifactorial design moniteki-	analysis) faktorianalyysi
jäkoe	multiple-frame survey monike-
multigraph moniverkko, multi-	hikkoluotaus
verkko	multiple imputation monipaik-
multihit model moniosumamalli	kaus, moni-imputointi
multilevel model monitasomalli,	multiple range test usean vaih-
hierarkkinen malli	teluvälin testi
multilinear process multiline-	multiple recapture design tois-
aarinen prosessi, monilineaarinen	toppyyntien asetelma, sarjatoisto-
prosessi	pyynti
multimodal distribution moni-	multiple regression monen selit-
huippuinen jakauma	täjän regressio
multinomial coefficient multino-	multiple response analysis mo-
mikerroin	nivastausanalyysi
multinomial distribution multi-	multiple smoothing method
nomijakauma	monisilotusmenetelmä, monitasoi-
multinomial logistic regression	tusmenetelmä
multinomilogistinen regressio	multiple stratification moniosi-
multinomial logit model	tus
(multicategory logit model) multi-	multiple testing monitestaus
nomilogit-malli	multiple time series (vector ti-
multinormal distribution	me series) moniulotteinen aikasar-
(multivariate normal distribution)	ja, vektoriaikasarja
multinormaalijakauma, moninor-	multiplicative model multiplika-
maalijakauma	tiivinen malli, kerrannainen malli
multiphase sampling monivai-	multiplicative seasonal model
heinen otanta, monivaiheotanta	multiplikatiivinen kausivaihtelu-
multiphase screening monivai-	malli, kerrannainen kausivaihtelu-
heseulonta	malli
multiple bar chart ositettu pyl-	multistage estimation monias-

teinen estimointi
multistage sampling moniasteinen otanta, moniasteotanta
multistate life table monitilainen elinajantaulukko
multistate model monitilamalli
multi-task learning monitehtävöppiminen
multivalued decision diagram monivaihepäättöskaavio
multivariable regression (multiple regression) monen selittäjän regressio
multivariate monimuuttujainen
multivariate adaptive regression spline (MARS) monimuuttujainen mukautuva regressiosplini
multivariate analysis monimuuttuja-analyysi
multivariate analysis of variance (MANOVA) monen vasteen varianssianalyysi, MANOVA
multivariate counting process monimuuttujainen laskuriprosessi
multivariate distribution (joint distribution) moniulotteinen jakauma
multivariate Gaussian distribution (multinormal distribution) moniulotteinen Gaussin jakauma, multinormaalijakauma, moninormaalijakauma
multivariate imputation by chained equations (MICE) ketjuimputointi
multivariate moment moniulotteinen momentti
multivariate normal distribution (multivariate Gaussian distribution) multinormaalijakauma,

moninormaalijakauma
multivariate regression²⁴ monen vasteen regressio
mutual fund sijoitusrahasto
mutually exclusive events erilliset tapahtumat, toisensa poissulkevat tapahtumat
MVUE (minimum variance unbiased estimator)

N

Nadaraya–Watson estimator Nadarayan–Watsonin estimaattori
naive forecast naiivi ennuste
natural conjugate prior luonnollinen liittopriori
natural direct effect (NDE) luonnollinen suora vaikutus
natural experiment luonnon koe
natural indirect effect (NIE) luonnollinen epäsuora vaikutus
natural language processing (NLP) luonnollisen kielen prosessointi, käyttökielen prosessointi
natural parameter (canonical parameter) luonnollinen parametri, kanoninen parametri
NBU (new better than used)
NDE (natural direct effect)
nearest-neighbour clustering lähinaapuriryhmittely
nearest-neighbour method lähinaapurimenetelmä
near unit root yksikköjuurta lähestyvä
negative binomial distribution negatiivinen binomijakauma
negative exponential distribution (exponential distribution) eksponenttijakauma

²⁴Ei ole synonyymi termeille *multiple regression* ja *multivariable regression*, vaikka monet soveltajat siinä merkityksessä tätä termiä käyttävät.

negative hypergeometric distribution	negatiivinen hypergeometrinen jakauma
negative predictive value	negatiivinen ennustearvo
negative skewness	vasemmalle vinous
neighbor embedding	naapurustoupotus
nested case-control study	pesitetty tapaus-verrokkitutkimus, upotettu tapaus-verrokkitutkimus
nested design	sisäkkäinen asetelma
nested hypotheses	sisäkkäiset hypoteesit
nested models	sisäkkäiset mallit
nested sampling	sisäkkäinen otanta
net migration	nettomuutto
net reproduction rate (NRR)	nettouusiutumisluku
network information criterion (NIC)	verkkoinformaatiokriteeri
network of samples	(interpenetrating samples) otosverkosto, päällekkäiset otokset
network sampling	verkko-otanta
neural computing	neurolaskenta
neural network	neuroverkko
new better than used (NBU)	uusi vanhaa parempi, NBU
Newton–Raphson method	Newtonin–Raphsonin menetelmä
Neyman allocation	Neymanin kiintiöinti
Neyman–Pearson lemma	Neymanin–Pearsonin apulause, Neymanin–Pearsonin lemma
NIC (network information criterion)	
NID (normally and independently distributed)	
NIE (natural indirect effect)	
NLP (natural language processing)	
node (vertex)	solmu
noise	kohina, häly
nominal confidence level	nimellinen luottamustaso
nominal coverage probability	nimellinen peittotodennäköisyys
nominal scale	luokka-asteikko, luokitteluasteikko
nominal significance level	nimellinen merkitsevyystaso
nominal size	nimellinen koko
nomogram	nomogrammi
non-central	epäkeskinen
non-centrality parameter	epäkeskisyyssparametri
noncoverage bias	peittoharha
non-differential misclassification	yhtäläinen virheluokittelu
non-experimental study	epäkokeellinen tutkimus
nonignorable	sivuuttamaton
noninformative prior	epäinformatiivinen prior, epäinformatiivinen esijakauma
non-life insurance	vahinkovakuutus
nonlinear model	epälineaarinen malli
nonlinear optimization	epälineaarinen optimointi
nonlinear regression	epälineaarinen regressio
nonnegative definite	(positive semidefinite) ei-negatiivisesti definitti, positiivisesti semidefiniitti
nonnested classification	epäsisäkkäinen luokitus
non-normal	ei-normaalinen
non-orthogonal	epäortogonaalinen

nonparametric parametrin, epäparametrinen
non-probability sampling epäai-
to otanta
non-random ei-satunnainen
non-randomized satunnaistama-
ton
nonresponse vastauskato
non-sampling error otannasta
johtumaton virhe
nonsense correlation hölynpöly-
korrelaatio, puppukorrelaatio
nonsignificant ei-merkitsevä
nonsingular epäsingulaarinen
non-terminal event
välitapahtuma
normal approximation normaali-
liapproksimaatio
normal distribution (Gaussian
distribution) normaalijakauma,
Gaussin jakauma
normal equations normaaliyhtä-
löt
normality normaalisuus
normalizing transformation
normalisoiva muunnos
**normally and independently
distributed** (NID) riippumatto-
mat ja normaalijakautuneet
normal probability plot nor-
maalijakaumakuviio
normal score normaaliarvo
nowcasting nykyistäminen
NR (net reproduction rate)
nuisance parameter kiusapara-
metri, häittäparametri
null deviance nolladevianssi
null distribution nollajakauma
null hypothesis nollahypoteesi
null model nollamalli
null set nollajoukko
null space nolla-avaruus
numeraire numeraari

numerator osoittaja
numerical optimization numee-
rinen optimointi
nuptiality avioliittoisuus

O

objective function kohdefunktio
oblimax rotation oblimax-rotaa-
tio, oblimax-kierto
oblimin rotation oblimin-rotaa-
tio, oblimin-kierto
oblique vino
oblique projection vinoprojektio
oblique reflection vinoheijastus
observable havaittava
observation havainto
observational error havaintovir-
he
observational study havainnoiva
tutkimus
observational unit havaintoyk-
sikkö, tilastoyksikkö
observed frequency (observed
number) havaittu lukumäärä
observed information havaittu
informaatio
observed information matrix
havaittu informaatiomatriisi
observed number (observed
frequency) havaittu lukumäärä
observed significance level (p -
value) havaittu merkitsevyystaso,
 p -arvo
occupation time viipymisaika
occurrence esiintyvyys
occurrence-exposure rate ta-
pahtuma-alttiussuhde
octile oktiili
odds vastasuhde, vetokerroin
odds ratio (OR) ristisuhde, veto-
suhde, OR
official statistics virallinen tilas-

to	lu
offset term offset-termi	optimal allocation optimaalinen kiintiöinti, optimikiintiöinti
offspring distribution jälkeläisjakauma, lisääntymisjakauma	optimal design optimiasetelma
oldest old vanhoista vanhimmat	optimality optimaalisuus
OLS (ordinary least squares)	optimal scaling optimiskaalaus
omnibus test kattava testi, yhteistesti	optimization optimointi
one-armed bandit yksikätkäinen rosvo	optimization method optimointimenetelmä
one-compartment model yksilokeromalli	optional time (stopping time) valintahetki, pysäytyshetki
one-hit model yksiosumamalli	OR (odds ratio)
one-sample t-test yhden otoksen t-testi	ordered categorical variable (ordinal categorical variable) järjestetty luokkamuuttuja
one-sided alternative hypothesis yksisuuntainen vastahypoteesi, yksisuuntainen vaihtoehtoinen hypoteesi	ordered sample järjestetty otos
one-sided test (one-tailed test) yksisuuntainen testi, yksitahoinen testi	order of interaction yhdysvaikutuksen kertaluku
one-tailed test (one-sided test) yksisuuntainen testi, yksitahoinen testi	order of magnitude suuruusluokka
one-way analysis of variance yksisuuntainen varianssianalyysi	order statistic järjestystunnusluku
one-way classification yksisuuntainen luokittelu	ordinal categorical variable (ordered categorical variable) järjestetty luokkamuuttuja
open data avoin data	ordinal data järjestysasteikollinen aineisto
open-ended class avoin päätyluokka	ordinal regression järjestysregressio, ordinaaliregressio
open population avoin väestö	ordinal scale järjestysasteikko, ordinaaliasteikko
operating characteristic (evaluation metric) toimivuusmitta	ordinal scale variable järjestysasteikollinen muuttuja
operations research operaatiotutkimus	ordinal variable järjestysmuuttuja
opinion poll (opinion survey) mielipidekysely, mielipidetiedustelu	ordinary least squares estimator (OLS) pienimmän neliösumman estimaattori
opinion survey (opinion poll) mielipidekysely, mielipidetiedustelu	ordination of species lajien ordinaatio
	orthogonal (perpendicular) ortogonaalinen, kohtisuora
	orthogonal polynomial ortogono-

naalipolynomi
orthogonal projection ortogonaaliprojektio, kohtisuora projektiio
orthonormal ortonormaallinen
oscillation heilahtelu, värähtely
oscillatory process värähtelevä prosessi
outcome (response) vaste, tulos
outcome variable (response variable) vastemuuttuja, selitettävä muuttuja
outer fence ulkoaita
outer product ulkotulo
outlier oudokki, vierashavainto
out-migration poismuutto
out-of-bag error laukutusvirhe
overadjustment liikavakiointi
overcount (overenumeration) laskennan ylipeitto
overcoverage ylipeitto
overenumeration (overcount) laskennan ylipeitto
overdispersion ylihajonta
overfitting likasovitus, ylisovitus
overidentification yliidentifiointi
overmatching liikakaltaistus, ylikaltaistus
overparameterized liikaparametroituu, yliparametroituu
own children method omien lasten menetelmä

P

paired comparison parivertailu
paired samples pariotokset
paired *t*-test parittainen *t*-testi
pairwise independence parittaisen riippumattomuus
panel data paneeliaineisto, paneelidata

panel study paneelitutkimus
panel wave paneeliaalto
parallel coordinate plot rinnakkaisten koordinaattien kuvio
parallel groups design rinnakkaisten ryhmien asetus
parameter parametri
parameter-driven model parametrivetoinen malli
parameter space parametriavaruus
parametric distribution parametrinen jakauma
parametric inference parametrinen päättely
parametric model parametrinen malli
parametrization parametointi
Pareto distribution Pareto-jakauma
parity₁ lapsiluku
parity₂ pariteetti
parity-progression ratio pariteettisuhde
parsimony vähäparametrisuus, säästeliäisyys
partial association osittaisyyhteys, osittaisassosiaatio
partial autocorrelation osittaisautokorrelaatio
partial correlation osittaiskorrelaatio
partial least squares (PLS) osittainen pienin neliösumma
partial likelihood osittaisuskottavuus
partially balanced array osittain tasapainotettu asetus
partial rank correlation osittaisjärjestyskorrelaatio
partial regression coefficient osittaisregressiokerroin
partial regression plot (added

variable plot) osittaisregressiokuva, lisäselittäjäkuvio	penalty function (penalty term) sakkofunktio
partial replacement osittainen palautus	penalty term (penalty function) sakkotermi
partial residual plot osittaisjäännöskuvio	penetrance penetranssi, ilmene- misyleisyys, läpäisy
partial table osataulukko ²⁵	pension fund eläkerahasto
participation rate osallistumis- suus, osallistumisaste	pentad criterion viisikkokriteeri
partition ositus	percentage distribution pro- senttijakauma
partitioned matrix ositettu mat- riisi	percentage point prosenttiyksik- kö
passage time kuluaika	percentile prosentiili, prosentti- piste
path analysis polkuanalyysi	perceptron päättelin
path coefficient polkukeroin	percolation theory perkolaatio- teoria
path dependency polkuriippu- vuus	periodic process jaksollinen pro- sessi
pattern mixture model puuttu- vuuden sekoitemalli	periodic survey (repeated sur- vey) toistoluotaus
pattern recognition hahmontun- nistus	period of seasonality kausijakso
pay-off matrix tuottomatriisi	periodogram jaksokuva, perio- dogrammi
PCA (principal components analy- sis)	permissible estimator hyväksyt- tävä estimaattori
PDF (probability density func- tion)	permutation permutaatio
peakedness (kurtosis ₂) huipuk- kuus	permutation test permutaatio- testi
Pearson residual Pearson-jään- nös	perpendicular (orthogonal) kohti- suora, ortogonaalinen
Pearson's chi-squared statistic Pearsonin khi-neliösuure, Pearso- nin khi-neliötunnusluku	perpetuity pysyvä elinkorko
Pearson's chi-squared test Pearsonin khi-neliötesti	per-protocol analysis protokol- lan mukainen analyysi
penalized estimation sakotettu estimointi	personal probability omato- dennäköisyys, henkilökohtainen todennäköisyys
penalized least squares sakotet- tu pienin neliösumma	person-time henkilöaika
penalized likelihood sakotettu uskottavuus	person-year henkilövuosi
	perspective plot perspektiiviku- vio

²⁵Tietojenkäsittelytieteessä käytetään tietokantojen yhteydessä suomennosta "osit-
tainen taulu".

perturbation₁ (disturbance ₁) häiriö	vio
perturbation₂ sysäys	PLS (partial least squares)
perturbation method häiriömenetelmä, perturbaatiomenetelmä	plug-in estimator pistoke-estimaattori
p-hacking <i>p</i> -arvon kalastelu	PNDE (pure natural direct effect)
phase vaihe, vaihekulma	PNIE (pure natural indirect effect)
phase diagram vaihediagrammi, vaihekaavio	point biserial correlation piste-biseriaalin korrelaatio
phase spectrum vaihespektri	point density pistetiheys
phase transition vaihemuutos	point estimation piste-estimointi
phenotype fenotyyppi, ilmiö	point pattern pistekuvio
phi-coefficient fi-kerroin	point probability pistetodennäköisyys
Phillips curve Phillipsin käyrä	point process pisteprosessi
pictogram piktogrammi	point sampling pisteotanta
piecewise constant paloittain vakio	pointwise confidence band (pointwise confidence envelope) pisteittäinen luottamusvyöhyke
piecewise linear paloittain lineaarinen	pointwise confidence envelope (pointwise confidence band) pisteittäinen luottamusvyöhyke
piecewise regression (segmented regression) paloittainen regressio	Poisson distribution Poisson-jakauma
pie chart (pie diagram) sektorikuvio, piirakkakuvio	Poisson process Poisson-prosessi
pie diagram (pie chart) piirakkakuvio, sektorikuvio	Poisson sampling Poisson-otanta
pilot study esitutkimus, pilotti	Poisson variation Poisson-vaihtelu
pink noise vaaleanpunainen kohina	polar coordinate napakoordinaatti
Pitman efficiency Pitman-tehokkuus	Pólya urn model Pólyan uurnamalli
pivot (pivotal quantity) napasuure	polychotomous (polytomous) moniluokkainen
pivotal quantity (pivot) napasuure	polynomial regression polynomiregressio
placebo lume	polynomial trend polynominen trendi
placebo effect lumevaikutus	polyspectrum polyspektri
planned comparison suunniteltu vertailu	polytomous (polychotomous) moniluokkainen
play-the-winner rule suosi voitattajaa -sääntö	pooled estimator yhdistetty estimaattori
plim (limit in probability)	
plot₁ kuvio	
plot₂ koeala	
plot dendrogram puu-pistekaa-	

population₁ perusjoukko, populaatio	vertailu, post hoc -vertailu
population₂ väestö, populaatio	post-stratification jälkiositus
population-averaged model populaatiokeskiarvoistettu malli	potential potentiaali
population genetics populaatio-genetiikka	potential outcome potentiaalin vaste
population growth model populaation kasvumalli, väestönkasvumalli	power₁ voima, hylkäystodennäköisyys
population pyramid populaatiopyramidi, väestöpyramidi	power₂ potenssi
population structure väestörakenne	power function voimafunktio
portemanteau test portemanteau-testi, naulakkotesti	power law potenssilaki
portemanteau theorem portemanteau-lause, naulakkolause	power spectrum spektritiheys
portfolio optimization arvopaperisalkun optimointi	power transformation potenssimuunnos
positive definite positiividefiniitti, positiivisesti definiitti	P-P plot P-P-kuvio, todennäköisyyskuvio
positive predictive value positiivinen ennustearvo	pps (probability proportional to size)
positive semidefinite (non-negative definite) positiivisesti semidefiniitti	pragmatic trial (effectiveness trial) pragmaattinen koe, vaikutavuuskoe
positive skewness oikealle vinous	precision₁ toistotarkkuus
postal survey postikysely	precision₂ keskittyneisyys
post-clustering jälkiryvästys	precision₃ (positive predictive value) positiivinen ennustearvo ²⁶
posterior distribution posteriorijakauma, jälkijakauma	precision matrix (concentration matrix) keskittyneisyysmatriisi
posterior interval posterioriväli	precision parameter keskittyneisyysparametri
posterior predictive distribution posterioriennustejakauma, jälkiennustejakauma	predictable process ennakoituva prosessi
posterior probability posterioritodennäköisyys, jälkitodennäköisyys	predicted value ennustettu arvo, ennuste ₃
post hoc comparison jälkikäteis-	prediction ennuste ₂ , ennustaminen
	prediction error ennustevirhe
	prediction interval ennusteväli
	prediction residual ennustejäännös
	prediction residual sum of squares (PRESS) ennustejäännösten neliösumma

²⁶Koneoppimisen alalla sanaa *precision* käytetään tässä merkityksessä.

predictive analytics	ennustava analytiikka		esiennustejakauma
predictive distribution	ennustejakauma		prior probability (a priori probability) prioritodennäköisyys, esitodennäköisyys
predictive inference	ennustava päättely, prediktiivinen päättely		probabilistic inference todennäköisyyspäättely
predictive mean matching	ennustekeskiaRVokaltaistus		probability todennäköisyys
predictive value	ennustearvo		probability calculus todennäköisyyslaskenta
predictor	ennustin, ennustava muuttuja		probability density function (PDF) tiheysfunktio, todennäköisyystiheys
preferential attachment process	valikoivan kasvun prosessi		probability distribution todennäköisyysjakauma
preregistration	ennakkorekisteröinti		probability generating function todennäköisyysmäfunktio, todennäköisyydet generoiva funktio
prescriptive analytics	ohjaava analytiikka		probability integral transformation kertymäfunktioMuunnos
pre-specified significance level	ennalta asetettu merkitsevyystaso		probability judgement todennäköisyysarvio
PRESS (prediction residual sum of squares)			probability kernel todennäköisyysydin
prevalence	vallitsevuus		probability limits todennäköisyysrajat
prevention trial	ehkäisykoe		probability mass function (point probability function) pistetodennäköisyysfunktio
prewhitening	esivalkaisu		probability measure todennäköisyysmitta
price index	hintaindeksi		probability model (stochastic model) todennäköisyysmalli
primary sampling unit (PSU)	ensiasteen otantayksikkö		probability plot todennäköisyyskuvio
principal axes decomposition	pääakselihajotelma		probability proportional to size (pps) kokoon verrannollinen todennäköisyys
principal components analysis (PCA)	pääkomponenttianalyysi		probability sampling (random sampling) todennäköisyysotanta, satunnaisotanta
principal components regression	pääkomponenttiregressio		probability space todennäköisyysavaruus
principal coordinate	pääkoordinaatti		
principal stratification	pääositus		
prior distribution	priorijakauma, esijakauma		
prior elicitation (elicitation of prior)	priorin koostaminen, esijakauman koostaminen		
prior predictive distribution			

probability theory todennäköisysteoria

probability weighting todennäköisyyspainotus

probable (likely) todennäköinen

probable error todennäköinen virhe

probit model probit-malli

probit transformation probit-muunnos

procedural bias menettelyharha

processing error käsittelyvirhe

Procrustes analysis Prokrustes-analyysi

product integral tulointegraali

product limit estimator (Kaplan–Meier estimator) rajatuloestimaattori, Kaplanin–Meierin estimaattori

product measure tulomitta

product moment tulomomentti

product moment correlation tulomomenttikorrelaatio

product space tuloavaruus

profile analysis profilianalyysi

profile likelihood profiiliuskottavuus

profile log-likelihood profili-loguskottavuus

profile plot profilikuvio

profiling profilointi

prognostic factor ennustetekijä

prognostic model ennustemalli

projection projektio

projection plot projektiokuvio

projection pursuit projektion haku

projection pursuit regression optimiprojektio regressio

propagation of error (error propagation) virheen kasautuminen, virheen leviäminen

propensity propensiteetti, alttius

propensity matching propensiteettikaltaistus, alttiuskaltaistus

propensity score propensiteettiluku, alttiusluku

propensity weighting propensiteettipainotus, alttiuspainotus

proper imputation kelvollinen imputointi

proper prior aito prior

proportion osuus

proportional allocation suhteellinen kiintiöinti

proportional hazards model (Cox model) verrannollisten uhkien malli, Coxin malli

proportional odds model verrannollisten vastasuhteiden malli

proportional sampling suhteellinen otanta

proposal distribution (candidate distribution) ehdotusjakauma, ehdokasjakauma

prospect theory prospektiteoria, näkymäteoria

prospective study prospektiivinen tutkimus

protected data suojattu tieto

protocol protokolla

protocol violation protokollariike

proximity matrix läheisyysmatriisi

pruning karsinta

pseudo-likelihood pseudouskottavuus

pseudo-random number pseudosatunnaisluku, näennäissatunnaisluku

PSU (primary sampling unit)

psychometrics psykometriikka

publication bias julkaisuharha

pure natural direct effect (PNDE) puhdas luonnollinen suo-

ra vaikutus
pure natural indirect effect
 (PNIE) puhdas luonnollinen epäsuora vaikutus
purposeful selection harkintavallinta
purposive sample harkintaotos
put option myyntioptio
p-value *p*-arvo

Q

QALY (quality adjusted life years)
QCA (qualitative comparative analysis)
Q-Q plot (quantile-quantile plot)
 Q-Q-kuvio
quadrant dependence kvadranttiriippuvuus
quadrant sampling kvadranttio-
 tanta
quadratic kvadraattinen, neliöllinen
quadratic discriminant function kvadraattinen erottelufunk-
 tio, neliöerottelufunktio
quadratic estimator kvadraatti-
 nen estimaattori, neliöestimaatto-
 ri
quadratic form neliömuoto
quadratic loss function neliö-
 tappiofunktio
quadratic variation kvadraatti-
 nen vaihtelu, neliövaihtelu
qualitative laadullinen, kvalitatiivinen
qualitative comparative analysis (QCA) laadullinen vertaileva
 analyysi
quality adjusted life years
 (QALY) laatuelinvuosi
quality assurance laadunvarmis-

tus
quality control laadunvalvonta
quality control chart laadunval-
 vontakortti
quantal dose-response kaksiar-
 voinen annos-vaste
quantile (fractile) kvantiili, fraktii-
 li
quantile regression kvantiili-
 regressio
quantitative määrällinen, kvanti-
 tatiivinen
quantum probability kvanttito-
 dennäköisyys
quarterly data neljännesvuosiai-
 neisto
quartile kvartiili
quartile deviation kvartiilipoik-
 keama
quartimax rotation kvartimax-
 rotaatio, kvartimax-kierto
quasi experiment kvasikoe
quasi independence kvasiriippu-
 mattomuus
quasi likelihood kvasiuskotta-
 vuus
quasi symmetry kvasisymmetria
quenched distribution karkais-
 tu jakauma
questionnaire kyselylomake
questionnaire survey kyselytut-
 kimus
queue jono
queueing theory jonoteoria
quintile kvintiili
quota sampling kiintiöotanta
quotient (ratio) osamäärä, suhde,
 suhdeluku

R

R_0 (basic reproduction number)
 R_0

R² (coefficient of determination)	tettu vastaus
R²	randomized test satunnaistettu testi
radial function radiaalifunktio	randomized trial (randomized experiment) satunnaistettu koe
radial plot sädekuvio	random matrix satunnaismatriisi
Radon–Nikodym derivative	random measure satunnaismitta
Radonin–Nikodymin derivaatta	randomness satunnaisuus
raking (iterative proportional fitting) haravointi	random number satunnaisluku
Rand index Rand-indeksi	random number generator satunnaislukugeneraattori, satunnaislukukehitin
random satunnainen	random phenomenon satunnaisilmiö
random allocation satunnaisjako	random process (stochastic process) satunnaisprosessi, stokastinen prosessi
random coefficient satunnaiskerroin	random sample satunnaisotos
random component satunnaiskomponentti	random sampling satunnaisotanta
random digit dialing puhelinnumeron satunnaisvalinta	random selection satunnaisvalinta
random effect satunnaisvaikutus	random sequence satunnaisjono
random effects model satunnaisvaikutusmalli	random set satunnaisjoukko
random error satunnaisvirhe	random tessellation satunnais-tessellaatio, satunnaismosaiikki
random event satunnaistapahtuma	random variable (random variate) satunnaismuuttuja, satunnais-suure
random field satunnaiskenttä	random variation (chance variation) satunnaisvaihtelu
random forest satunnaismetsä	random vector satunnaisvektori
random function satunnaisfunktio	random walk satunnaiskulku
random graph satunnaisverkko	random walk with drift suunnallinen satunnaiskulku
random-effects model satunnaisvaikutusmalli	range₁ vaihteluväli
randomization satunnaistus	range₂ vaihteluvälin pituus
randomization test satunnais-tustesti	range chart vaihteluvälikortti
randomized blocks design satunnaistettujen lohkojen asetelma	rank₁ sijaluku, järjestysluku
randomized clinical trial (RCT) satunnaistettu kliininen koe	rank₂ aste ₁
randomized decision rule satunnaistettu päätössääntö	rank correlation järjestyskorrelaatio, sijalukukorrelaatio
randomized experiment (randomized trial) satunnaistettu koe	
randomized response satunnais-	

ranking	sijavertailu, sijakisa	record₁	tietue, kirjaus, merkintä
rankit transformation	rankit-muunnos	record₂	ennätys
rank order statistic	sijatunnus-luku	record linkage	tietuekytkentä
rank-randomization test	satun-naistettu järjestyslukutesti	rectangular distribution	(uniform distribution) tasajakau-ma
rank regression	sijalukuregressio	rectangular lattice	suorakaidehi-la
rank transformation	järjestys-muunnos	rectifying inspection	korjaava tarkastus, korjaileva laadunvalvon-ta
Rao–Blackwellization	Raon–Blackwellin tehostus	recurrence risk	uusiutumisriski
rate₁ (intensity)	taajuus, tiheys	recurrence time₁	paluu aika
rate₂ (proportion)	osuus, aste ₂	recurrence time₂	uusiutumisaika
rate₃ (ratio)	suhde	recurrent₁	palautuva
rate of change	muutosvauhti, muutostahti	recurrent₂	uusiutuva
ratio	suhde, suhdeluku, osamäärä	recursive	toistava, rekursiivinen
ratio estimator	suhde-estimaat-tori	reduced form	pelkistetty muoto, redusoitu muoto
rational expectations	rationaali-set odotukset	reduced inspection	supistettu tarkastus
ratio scale	suhdeasteikko	reduced maximum likelihood	(restricted maximum likelihood) rajoitettu suurin uskottavuus
raw moment	raakamomentti	reference interval	viiteväli
raw score	raaka-arvo	reference limit	viiteraja
RCT (randomized clinical trial)		reference period	viitejakso
RDS (respondent-driven sam-pling)		reference population	viitepopulaatio, viiteväestö
realization	toteuma, realisaatio, toteutuma	reference prior	viitepriori
recall (sensitivity ₁)	herkkyys, sen-sitiivisyys ²⁷	reference value	viitearvo
recall bias	muistiharha	reflecting barrier	heijastava reu-na
receiver operating characteris-tic (ROC)	erottelukykysoitin	reflection principle	heijastuspe-riaate
receiver operating characteris-tic curve (ROC curve)	erottelu-kykykäyrä, ROC-käyrä	refusal rate	kieltäytymisosuus
reciprocal transformation	kään-teislukumuunnos	regenerative	uudistuva
		regressand (response variable)	vastemuuttuja, selitettävä muuttu-ja

²⁷Koneoppimisen alalla termiä *recall* käytetään samassa merkityksessä kuin *sensitivity* kuvattaessa kaksiarvoisen luokittelun erottelukykyä. Tiedonhankinnassa *recall* ja sen synonyymi *recall ratio* ovat merkitykseltään analogisia suomennoksenaan “saanti”.

regression	regressio	reinforcement learning	vahvistusoppiminen
regression analysis	regressioanalyysi	reinsurance	jälleenvakuutus
regression attenuation	(regression dilution) regression vaimeneminen	rejectable quality level	hylättävä laatutaso
regression calibration	regressiokalibrointi	rejection error	hylkäysvirhe
regression coefficient	regressiokerroin	rejection region	(critical region) hylkäysalue, kriittinen alue
regression curve	regressiokäyrä	rejection sampling	hylkäysotanta
regression diagnostics	regressiodiagnostiikka	rejective sampling	hylkäävä otanta
regression dilution	(regression attenuation) regression laimeneminen	relational database	relaatiotietokanta
regression discontinuity design	(RDD) regressioepäjatkuvuusasetelma	relative efficiency	suhteellinen tehokkuus
regression estimator	regressioestimaattori	relative frequency	(proportion) osuus, suhteellinen frekvenssi
regression function	regressiofunktio	relative likelihood	suhteellinen uskottavuus
regression line	regressiosuora	relative risk	(risk ratio) riskisuhde, suhteellinen riski, RR
regression model	regressiomalli	relative survival	suhteellinen elossaolo
regression modeling	regressiomallinnus, regressiomallitus	relative variance	(relvariance) suhteellinen varianssi
regression specification error test	(RESET) RESET-testi	reliability₁	luotettavuus
regression spline	regressiosplini	reliability₂	reliabiliteetti
regression sum of squares	regressioneliösumma	reliability coefficient	reliabiliteettikerroin
regression surface	regressiopinta	reliability function	luotettavuusfunktio
regression through the origin	regressio origon kautta	reliability theory	luotettavuusteoria
regression to the mean	regressio keskiarvoa kohti	relvariance	(relative variance) suhteellinen varianssi
regressor	(explanatory variable) selittävä muuttuja, selittäjä	remaining life expectancy	jäljellä oleva elinajanodote
regret	menetyt	REML	(restricted maximum likelihood)
regularization	sääntely	Rényi divergence	Rényin poikkeama
regularly varying	säännöllisesti vaihteleva	Rényi entropy	Rényin entropia

remote sensing	kaukokartoitus	RESET	(regression specification error test)
removal method	poistomenetelmä	residual	jäännös, residuaali
renewal density	uusiutumistiheys, uusiutumistaajuus	residual deviance	jäännösdevianssi
renewal distribution	uusiutumisaikajakauma	residual diagnostics	jäännösdiaagnostiikka
renewal function	uusiutumiskäyrä	residual effect	jäännösvaikutus
renewal process	uusiutumiskäyrä	residual lifetime	jäljellä oleva elinikä
renewal theory	uusiutumisteoria	residual maximum likelihood	(restricted maximum likelihood)
reparametrization	uusparametrisointi, uudelleenparametrisointi		rajoitettu suurimman uskottavuuden menetelmä, REML
repeatability	(replicability) toistettavuus	residual mean square	jäännöskeskinneliö
repeated measurements design	(repeated measures design) toistomittausasetelma	residual plot	jäännöskuvio
repeated measures design	(repeated measurements design) toistomittausasetelma	residual standard deviation	jäännöshajonta
repeated survey	(periodic survey) toistoluotaus	residual sum of squares	jäännösneliösumma
repetition	toisto, uusinta, kertaus	residual variance	jäännösvarianssi
replicability	(repeatability) toistettavuus	residual waiting time	jäljellä oleva odotusaika
replicate	toisinne, replikaatti	resistant estimator	resistentti estimaattori
replication	toistaminen, replikaatio	resolvable block design	ratkeava lohkoasetelma
replication study	tutkimuksen toisto	respondent-driven sampling	(RDS) vastaajavetoinen otanta
reporting error	raportointivirhe	response₁	(outcome) vaste, tulos
representative sample	edustava otos	response₂	(answer) vastaus
representativeness	edustavuus	response error	vastausvirhe
reproducibility	uusittavuus	response group	vastaajaryhmä
reproducing kernel Hilbert space	(RKHS) toisintavaytiminen Hilbertin avaruus	response indicator	vastausilmaisimen, vastausindikaattori
resampling	toisto-otanta	response mechanism	vastausmekanismi
rescaled sample weight	skaalattu otospaino	response metameter	vasteen muunnos
		response propensity weight	

vastausalttiuspaino	raali mitta
response rate vastausaste	risk prone person riskihakuinen henkilö
response surface vastepinta	risk ratio (relative risk) riskisuhde, suhteellinen riski, RR
response surface methodology (RSM) vastepintamethodologia	risk set riskijoukko
response time distribution vasteaikajakauma	risk set sampling riskijoukkopointiminta
response variable (outcome variable) vastemuuttuja, selitettävä muuttuja	risk theory riskiteoria
restricted maximum likelihood (REML) rajoitettu suurimman uskottavuuden menetelmä, REML	RJMCMC (reversible jump Markov chain Monte Carlo)
restricted randomization rajoitettu satunnaistus	RKHS (reproducing kernel Hilbert space)
retrospective study retrospektiivinen tutkimus	robust estimator vakaa estimaattori, robusti estimaattori
return level paluukvantiili	robustness vakaus, robustisuus
return period paluujakso	robust statistic vakaa tunnusluku, robusti tunnusluku
return time paluuaika	ROC (receiver operating characteristic) erottelukykysoitin
reversal design käänteisasetelma	ROC curve (receiver operating characteristic curve) erottelukykykäyrä
reversible jump Markov chain Monte Carlo (RJMCMC) vaihtelevaulotteinen MCMC, RJMCMC	rooted graph juurellinen verkko
reversible process aikakääntyvä prosessi	root mean squared error (RMSE) jäännösvirhehajonta, RMSE
reweighting uudelleenpainotus, uuspainotus	rose diagram ruusukaavio
ridge regression harjanneregressio, harjaregressio	rotatable design kiertoasetelma, rotaatioasetelma
ridit analysis ridit-analyysi	rotation kierto, rotaatio
right censoring sensurointi oikealta	rotation sampling kierto-otanta
right truncation katkaisu oikealta, typistys oikealta	roughness penalty rososakko
risk riski	Round Robin design kiertovuoroasetelma, round robin -asetelma
risk averse person riskiä kaihtava henkilö	rounding pyöristys
risk factor riskitekijä	route sampling reittiotanta
risk function riskifunktio	RR (relative risk)
riskless asset riskitön arvopaperi	RSM (response surface methodology)
risk neutral measure riskineut-	rug plot hapsukuvio
	ruin probability vararikkotodennäköisyys

ruin problem vararikko-ongelma
rule of three kolmen sääntö
run₁ merkkijakso
run₂ ajo
run-in period sisäänajojakso₂
runs test merkkijaksotesti
RV-coefficient RV-kerroin

S

saddle point approximation
 satulapisteapproksimaatio
sample₁ otos
sample₂ näyte
sample covariance matrix otos-
 kovarianssimatriisi
sample design (sampling design)
 otosasetelma, otanta-asetelma
sample distribution otosjakau-
 ma
sample distribution function
 otoskertymäfunktio
sample mean otoskeskiarvo
sample moment otosmomentti
sample path otospolku, realisaat-
 io
sample plan (sampling plan)
 otossuunnitelma, otantasuunni-
 telma
sample point otospiste
sampler poimija
sample size otoskoko
sample space otosavaruus
sample statistic otostunnusluku,
 otossuure
sample survey otosuotaus, otos-
 tutkimus
sample unit otosyksikkö
sample variance otosvarianssi
sample weight otospaino
sampling otanta
sampling design (sample design)
 otanta-asetelma, otosasetelma

sampling distribution otantajaja-
 kauma
sampling error otantavirhe
sampling fraction otantaosuus
sampling frame otantakehikko
sampling inspection otantatar-
 kastus
sampling interval poimintaväli
sampling plan otantasuunnitel-
 ma
sampling unit otantayksikkö
sampling without replacement
 otanta palauttamatta, pito-otanta
sampling with replacement
 otanta palauttaen, palautusotanta
sandwich estimator kerrosemi-
 maattori, sandwich-estimaattori
SAR (simultaneous autoregres-
 sion)
SARMA (seasonal autoregressive
 moving average process)
saturated model kyllästetty mal-
 li, saturoitu malli
scalability skaalautuvuus
scalar skalaari
scale₁ skaala
scale₂ mittakaava
scaled deviance skaalattu de-
 vianssi
scale factor skaalakerroin, skaala-
 tekijä
scale parameter skaalaparametri
scaling skaalaus
scan statistic liukuva tunnusluku
scatter (dispersion) sironta, hajon-
 ta
scatter diagram (scatter plot)
 sirontakuvio, hajontakuvio
scatter plot (scatter diagram) si-
 rontakuvio, hajontakuvio
scatterplot matrix sirontakuvio-
 matriisi, hajontakuviomatriisi
scenario skenaario

sceptical prior	skeptinen priori, skeptinen esijakauma	ma
SCM (structural causal model)		secular trend trendi
score₁ (efficient score)	skoori	seemingly unrelated regressions (SUR) näennäisesti riippumattomat regressiot
score₂	pistemäärä	segmented regression (piecewise regression) paloittainen regressio
score function	skoorifunktio	segregation₁ eriytyminen, segregatio
score residual	skoorijäännös	segregation₂ lohkeaminen, segregatio
score statistic	skoorisuure, skooritunnusluku	segregation analysis segregatioanalyysi, eriytymisanalyysi
score test (Lagrange multiplier test)	skooritesti, Lagrangen kerointesti	SEIR (susceptible-exposed-infectious-recovered)
score vector	skoorivektori	SEIR model SEIR-malli
scoring	pisteytys	selection bias valikoitumisharha, valintaharha
screening	seulonta	selection model valikoitumismalli
screening design	seulonta-asetelma	selection prior valikoitumispriori, valintapriori
screening test	seulontatesti	self-administered survey oma-toimiluotaus
scree plot	taluskuvio, vyörykuvio	self-exciting process itsekehkeytyvä prosessi
scree test	talustesti, vyörytesti	self-financing itsensä rahoittava
SD (standard deviation)		self-fulfilling prophecy itsensä toteuttava ennuste
SE (standard error)		self-negating prophecy itsensä estävä ennuste
series of experiments	koesarja	self-organizing map (SOM) itseorganisoituva kartta, SOM
seasonal adjustment	kausipuhdistus	self-selected sample itsevalikoitunut otos
seasonal autoregressive moving average process (SARMA)	kausi-ARMA-malli	self-similar process itsesimilaarinen prosessi
seasonal root	kausijuuri	self-weighting sample itsepainottuva otos
seasonal variation	kausivaihtelu	semi-invariant (cumulant) kumulantti
secant method	sekanttimenetelmä	semi-Markov process semimarkovprosessi
secondary sampling unit (SSU)	toisen asteen otantayksikkö	
second order efficiency	toisen kertaluvun tehokkuus	
second-order inclusion probability	kahden yksikön sisältymistodennäköisyys	
second order stationary (covariance stationary)	kovariansistationaarinen	
secretary problem	sihteeriongelma	

semi-martingale	semimartingaa- li, puolimartingaali	informaatio
semi-parametric	semiparametri- nen, osaparametrinen	shape parameter muotoparamet- ri
sensitivity ₁	herkkyys, sensitiivi- syys	share of a stock osake
sensitivity ₂	arkaluonteisuus	shock sokki
sensitivity analysis	herkkyysana- lyysi	shock model sokkimalli
sentiment analysis	tunneanaly- ysi	short-memory process lyhyt- muistinen prosessi
separable space	separoituva ava- ruus	short sale lyhyeksi myynti
separating hyperplane	erottava hypertaso	short-tailed distribution lyhyt- häntäinen jakauma
sequence model	sekvenssimalli	shot-noise pistekohina
sequential allocation	peräkkäi- sallokointi	shrinkage estimator kutistuses- timaattori
sequential analysis	peräkkäis- analyysi	sieve estimator seulaestimaattori
sequential model	peräkkäismalli;	sigma-algebra (sigma-field) sigma-algebra, sigmakenttä
sequential probability ratio test (SPRT)	peräkkäinen toden- näköisyysuhdetesti	sigma-field (sigma-algebra) sigma- kenttä, sigma-algebra
sequential design	peräkkäisase- telma	sigmoid curve sigmoidikäyrä, S- käyrä
sequential experiment	peräk- käiskoe	signal processing signaalinkäsit- tely
sequential sampling	peräkkäis- otanta, peräkkäispoiminta	signal to noise ratio signaali-ko- hinasuhde
sequential test	peräkkäistesti	signed rank test merkkijärjestys- testi
sequential updating	vaiheittainen päivittäminen	signed root-transformation etumerkkinen neliöjuurimuunnos
serial correlation	(autocorrelation) autokorrelaatio, sarjakorrelaatio	significance merkitsevyys
serial dilution assay	sarjalai- mennuskoe	significance level (level of signi- ficance) merkitsevyystaso
serial measurements	mittaussar- ja	significance test (test of signi- ficance) merkitsevyystesti
serial variation	sarjavaihtelu	sign test merkkitest
set	joukko	silhouette plot siluettikuvio
Shannon information	Shannon-	similarity coefficient samanlai- suuskerroin
		similarity matrix samanlaisuus- matriisi
		similarity measure samanlai- suusmitta
		similar matrices similaarit mat-

riisit	yksiasteotanta
similar regions similaariset alueet	singular distribution singulaarinen jakauma
similar test similaarinen testi	singular matrix singulaarinen matriisi, vajaa-asteinen matriisi
simple hypothesis yksinkertainen hypoteesi, piste-hypoteesi	singular value decomposition (SVD) singulaariarvohajotelma
simple linear regression yhden selittäjän regressio	SIR (susceptible-infectious-recovered)
simple random sampling (SRS) yksinkertainen satunnaisotanta	SIR model SIR-malli
simplex algorithm simpleksialgoritmi	six sigma principle kuuden sigma periaate
Simpson's paradox Simpsonin paradoksi	size-biased kokovalikoitunut
simulated annealing simuloitu hehkutus	size of test testin koko
simulated envelope simuloitu verho	skew distribution vino jakauma
simulation simulaatio, simulointi	skewness vinous
simulator simulaattori	skew-normal distribution vinonormaali-jakauma
simultaneous autoregression (SAR) simultaaninen autoregressio	skew-symmetric matrix vinosymmetrinen matriisi
simultaneous confidence intervals simultaaniluottamusvälit, samanaikaiset luottamusvälit	sliced inverse regression viipaloitu käänteinen regressio
simultaneous equations bias simultaaniyhtälöharha	slice sampler viipalepaimija
simultaneous equations model simultaaniyhtälömalli	sliding square plot liukuneliökuvio
simultaneous estimation yhteisestimointi, samanaikainen estimointi	slope kulmakerroin
simultaneous inference samanaikainen päättely	slope ratio assay jyrkkyyssuhdemääritys
single-blind experiment yksöissokkokoe	slowly varying function hitaasti vaihteleva funktio
single-index model yhden indeksin malli	small area estimation pienalueestimointi
single-linkage clustering yksikytkösryhmittely	smooth transition autoregressive model (STAR) tasaisen siirtymän autoregressiivinen malli, STAR
single-stage sampling	smoother silotin, tasoitin
	smoothing silotus, tasoitus
	smoothing spline silottava splini
	SMR (standardized mortality ratio)
	snowball sampling lumipallootanta

snowflake diagram	lumihiutale-kuvio	spectral distribution function	spektrikertymäfunktio
social data science	yhteiskuntadatatieide	spectral radius	spektrisäde
social statistics	yhteiskuntatilastotiede	spectrum	spektri
soft computing	pehmo-laskenta	speech recognition	puheentunnistus
sojourn time	viipymisaika	spherical coordinate	pallokoordinaatti
solvency	solvenssi	spherical distribution	pallopintajakauma
SOM (self-organizing map)		spherical variable	pallopintamuuttuja
space	avaruus	sphericity test	pallopintatesti
space-time cluster	tila-aikaryvä	spike and slab	piikki ja laatta
spaghetti plot	spagettikuvio	spike and slab distribution	piikki ja laatta -jakauma
sparse matrix	harva matriisi	spline	splini
spatial data	spatiaalinen data, aluedata, paikkadata	split-half method	puolitusmenetelmä
spatial experiment	spatiaalinen koe	split-plot design	osaruutuasetelma
spatial median	spatiaalinen mediaani	split-unit design	osayksikköasetelma
spatial pair-potential	spatiaalinen paripotentiaali	spread (dispersion)	hajonta
spatial point pattern	tilapistekuvio	spreadsheet	laskentataulukko
spatial point process	spatiaalinen pisteprosessi, tilapisteprosessi	spurious correlation	näennäiskorrelaatio
spatial statistics	spatiaalinen tilastotiede, aluetilastotiede, paikkatilastotiede	spurious precision	näennäistarkkuus
speech recognition	puheentunnistus	square lattice	neliöhila
species abundance	lajien runsaus	square matrix	neliömatriisi
specification bias	määrittelyharha	square root rule	neliöjuurisääntö
specificity	tarkkuus ₂ , spesifisyys	square root transformation	neliöjuurimuunnos
spectral analysis	spektrianalyysi	square summable	neliöllisesti summautuva
spectral decomposition	(eigendecomposition) spektriha-jotelma, ominaisarvohajotelma ²⁸	SRS (simple random sampling)	YSO
spectral density	spektritiheys	SSU (secondary sampling unit)	
		stable distribution	stabiili ja-

²⁸Symmetriselle matriisille pääakselihajotelma.

kauma	re
stable population stabiili väestö	statistical₁ tilastollinen
stable state stabiili tila	statistical₂ tilastotieteellinen
staggered entry porrasteinen	statistical authority tilastoviran-
sisääntulo	omainen
staircase design porrasetelma	statistical ensemble (ensemble)
staircase method (up-and-down	tilastollinen kooste, kooste
method) ylös-alasmenetelmä, tik-	statistical expert system tilas-
kausmenetelmä	tollinen asiantuntijajärjestelmä
standard deviation keskihajonta	statistical graphics tilastografiik-
standard error keskivirhe	ka
standardization₁ vakiointi	statistical hypothesis tilastolli-
standardization₂ standardointi,	nen hypoteesi
normitus	statistical inference tilastollinen
standardized coefficient stan-	päättely
dardoitu kerroin	statistical learning tilastollinen
standardized mortality ratio	oppiminen
(SMR) vakioitu kuolevuussuhde,	statistical literacy tilastolukutai-
SMR	to
standardized residual standar-	statistically discernible tilastol-
doitu jäännös	lisesti erottuva
standardized variable standar-	statistically significant tilastolli-
doitu muuttuja, normitettu muut-	sesti merkitsevä
tuja	statistical machine learning
standard normal distribution	tilastollinen koneoppiminen
standardinormaalijakauma, normi-	statistical model tilastollinen
tettu normaalijakauma	malli
standard population	statistical reasoning tilastolli-
vakioväestö	nen järkeily, tilastollinen peruste-
standard rate vakioilmaantuvuus	lu
STAR (smooth transition auto-	statistical significance tilastolli-
regressive model)	nen merkitsevyys
starburst plot (bagplot) tähtisa-	statistical science (statistics ₂)
dekuvio	tilastotiede
state space tilajoukko, tila-	statistical service tilastotoimi
avaruus	statistical space tilastollinen ava-
stationary distribution statio-	ruus
naarinen jakauma	statistical test tilastollinen testi
stationary population stationaa-	statistician tilastotieteilijä
rinen väestö	statistics₁ tilasto
stationary process stationaari-	statistics₂ (statistical science) ti-
nen prosessi	lastotiede
statistic tunnusluku, havaintosuu-	statistics anxiety tilastoahdistus

steady-state distribution	tasapainotilajakauma	stochastic volatility	stokastinen volatiliteetti, satunnaisheilunta
stem-and-leaf diagram	varsilehtikuvio, lehtikuvio	stopping rule	pysäytyssääntö
step-down procedure	alasaskeltava menettely	stopping time	(optional time) pysäytyshetki, valintahetki
stepped wedge design	porraskiila-asetelma	stopword	yleissana, stop-sana
stepwise regression	askeltava regressio	St. Petersburg paradox	Pietarin paradoksi
stereology	stereologia	stratification	ositus
stick breaking process	tikunkatkaisuprosessi	stratified model	ositettu malli
stochastic	stokastinen, satunnainen	stratified sampling	ositettu otanta
stochastic approximation	stokastinen approksimaatio	stratum ²⁹	osite
stochastic block model	stokastinen lohkomalli	strict exogeneity	aito eksogeenisuus
stochastic convergence	stokastinen suppeneminen	strictly stationary process	aidosti stationaarinen prosessi, vahvasti stationaarinen prosessi
stochastic dependence	stokastinen riippuvuus	strike price	rajahinta
stochastic imputation	satunnaispaikkaus, stokastinen imputointi	stripes plot	viivakoodikuvio, juovakuvio
stochastic integral	stokastinen integraali	strong consistency	vahva tarkentuvuus
stochastic matrix	stokastinen matriisi	strong convergence	(almost sure convergence) vahva suppeneminen, melkein varma suppeneminen
stochastic model	(probabilistic model) todennäköisyysmalli, stokastinen malli	strong exogeneity	vahva eksogeenisuus
stochastic optimization	stokastinen optimointi	strong law of large numbers	vahva suurten lukujen laki
stochastic ordering	stokastinen järjestys	structural equation model	rakenneyhtälömalli
stochastic process	(random process) satunnaisprosessi, stokastinen prosessi	structural change	rakennemuutos
stochastics	stokastiikka	structural causal model	(SCM) rakenteellinen kausaalimalli
		structural time series model	rakenneaikasarjamalli
		structural vector autoregression	rakenteellinen vektoriauto-

²⁹Monikko on *strata*.

regressio	sufficiency tyhjentävyys
structural zero rakenteellinen nolla	sufficient dimension reduction tyhjentävä dimension pienennys, tyhjentävä ulottuvuuksien vähentäminen
studentized difference in betas <i>t</i> -kerroinero	sufficient statistic tyhjentävä tunnusluku
studentized difference in fits <i>t</i> -sovite-ero	sum contrast summakontrasti
studentized range statistic <i>t</i> -vaihteluvälisuure, <i>t</i> -vaihteluväli-tunnusluku	sunflower plot auringonkukkakuvio
studentized residual <i>t</i> -jäännös	superexogeneity supereksogeneisuus
Student's distribution (<i>t</i> -distribution) Studentin jakauma, <i>t</i> -jakauma	superefficient ylitehokas
study tutkimus	superiority trial paremmuuskoe
study population tutkimuspopulaatio, tutkimusväestö	super learner superoppija
study unit tutkimusyksikkö	superleverage supervipuvoima
stylized facts tyylitellyt ominaispiirteet, olennaiset piirteet	supermartingale supermartingaali, ylimartingaali
stylometry stylometria, tyylitilastotiede	supernormality supernormaalisuus
subdensity function alitiheysfunktio	superpopulation model superpopulaatiomalli
subdistribution alijakauma	superposition of processes prosessien summaus, prosessien kerrostaminen
subdistribution function alikertymäfunktio	supersaturated design ylikyllästetty asetelma
sub-Gaussian distribution aligaussinen jakauma	supervised classification ohjattu luokittelu
subgroup analysis alaryhmäanalyysi	supervised learning ohjattu oppiminen
subjective probability subjektiivinen todennäköisyys	support kantaja
subject-specific model yksilökohtainen malli	support function (log-likelihood function) tukifunktio, log-uskottavuusfunktio
submartingale submartingaali, alimartingaali	support vector machine (SVM) tukivektorikone
subexponential aliekspONENTTinen	suppressor variable alistaja, alistava muuttuja
subsample osaotos, aliotos	SUR (seemingly unrelated regressions)
subset osajoukko	surprisal value (<i>s</i> -value) yllätysarvo, <i>s</i> -arvo
subspace aliavaruus	
substantive model ilmiömalli	

surrogate outcome sijaisvaste
surrogate variable sijaismuuttuja
survey³⁰ luotaus, kartoitus, kyselytutkimus
survey instrument luotausinstrumentti
survey methodology luotausmetodologia
survey mode luotautapa
survey round kyselykierros
survey sampling luotausotanta
survival analysis (lifetime data analysis) elin aika-analyysi, elossaoloanalyysi, kestoanalyysi
survival curve elossaolokäyrä, kestokäyrä
survival function elossaolofunktio, kestofunktio
susceptible-exposed-infectious-recovered³¹ (SEIR) altis-tartunnan saanut-tartuttava-toipunut, SEIR
susceptible-infectious-recovered³¹ (SIR) altis-tartuttava-toipunut, SIR
s-value (surprisal value) *s*-arvo, yllätysarvo
SVD (singular value decomposition)
switch-back design vaihtoasetelma
switching regression model vaihtoregressiomalli
symmetric distribution symmetrinen jakauma
symmetric sampling symmetrisen otanta

synthetic data synteettinen aineisto
synthetic estimator synteettinen estimaattori
systematic allocation systemaattinen jako, systemaattinen allokointi
systematic error (bias) systemaattinen virhe, harha
systematic sampling systemaattinen otanta, tasaväliotanta

T

table taulukko³²
tag cloud (word cloud) sanapilvi
tail exponent (tail index) häntäeksponentti
tail index (tail exponent) häntäeksponentti
tail of a distribution jakauman häntä
tails klaava
tapering kartiosilotus
TAR (threshold autoregression) kynnysautoregressio
targeted maximum likelihood estimation kohdistettu suurimman uskottavuuden menetelmä
target population kohdeperusjoukko, kohdepopulaatio, kohdeväestö
target trial tavoitekoe
target trial emulation tavoitekokeen jäljittely
taxonomy taksonomia
t-distributed stochastic neighborhood embedding (*t*-SNE)

³⁰Ks. alaluku 2.11.

³¹Adjektiivien *infectious* ja *recovered* paikalla käytetään myös adjektiiveja *infected* ja *removed*.

³²Tietojenkäsittelytieteessä on tietokantojen yhteydessä termille vakiintunut suomenos "taulu".

<i>t</i> -ympäristöupotus, <i>t</i> -jakaumaan perustuva stokastinen naapurustoupotus, <i>t</i> -naapurustoupotus	tu tarkastus
<i>t</i>-distribution (Student's distribution) <i>t</i> -jakauma, Studentin jakauma	time-dependent aikariippuva
temporal aggregation aika-aggregointi	time-dependent covariate aikariippuva selittäjä
temporal homogeneity (time-homogeneity) aikahomogeenisuus	time-dependent distribution aikariippuva jakauma
tensor tensori	time domain aika-alue
terminal event (endpoint) pääte-tapahtuma	time-frequency distribution aika-taajuusjakauma
test testi	time-homogeneity (temporal homogeneity) aikahomogeenisuus
test data (test set) testiaineisto	time-independent aikariippumaton, ajasta riippumaton
test of significance (significance test) merkitsevyystesti	time series aikasarja
test series testisarja	time-series analysis aikasarja-analyysi
test set (test data) testiaineisto	time-series econometrics aikasarjaekonometria
test statistic testisuure, testitunnusluku	time-to-event tapahtuma-aika
tetrachoric correlation tetrakorinen korrelaatio	time-to-event analysis tapahtuma-aika-analyysi
tetrad difference nelierotus	time-varying aikamuuttuva
text mining tekstinlouhinta	TNDE (total natural direct effect)
theorem lause, teoreema	TNIE (total natural indirect effect)
thinning harvennus	tobit-model tobit-malli
thin plate spline ohutlevysplini	Toeplitz matrix Toeplitz-matriisi
thin-tailed distribution ohut-häntäinen jakauma	tolerance toleranssi
three series theorem kolmen sarjan lause	tolerance interval toleranssiväli
three-stage least squares kolmivaiheinen pienin neliösumma	tontine tontiini
three-stage sampling kolmias-teotanta	topic modeling aihemallitus
threshold autoregression (TAR) kynnysautoregressio, TAR	total kokonaismäärä
threshold model kynnysmalli	total correlation kokonaiskorrelaatio
threshold value kynnysarvo	total natural direct effect (TNDE) luonnollinen suora kokonaisvaikutus
tie sidos, tasahavainto	total natural indirect effect (TNIE) luonnollinen epäsuora kokonaisvaikutus
tight ordering tiukka järjestys	total probability kokonaistodennäköisyys
tightened inspection tiukennet-	total sum of squares kokonais-

neliösumma	kontrasti
total variation distance (variational distance) kokonais- vaihteluetäisyys	treatment group hoitoryhmä, käsittelyryhmä
trace jälki	tree diagram puukaavio, puudia- grammi
training data (training set) ope- tusaineisto	tree model puumalli
training set (training data) ope- tusaineisto	trend trendi, kehityssuunta, pitkä- aikaismuutos
trait piirre	trial (experiment) koe
trajectory kehityspolku, trajekto- ri	triangular distribution kolmio- jakauma
transfer function siirtofunktio	triangular matrix kolmiomatriisi
transfer learning siirto-oppimi- nen	triangular table kolmiotaulukko
transformation muunnos	triangular test kolmiotesti
transient state väistyvä tila	trimmed mean leikattu keskiar- vo, trimmattu keskiarvo
transition intensity (transition rate) siirtymäintensiteetti, siirty- mätaajuus	trimodal kolmihiippuinen
transition matrix siirtymämatrii- si	triples test kolmikkotesti
transition probability siirtymä- todennäköisyys	Trojan square troijalainen neliö
transition rate (transition inten- sity) siirtymätaajuus, siirtymäin- tensiteetti	truncated distribution katkais- tu jakauma, typistetty jakauma
translation translaatio, siirto	truncated observation katkais- tu havainto, typistetty havainto
transmission probability tar- tuntatodennäköisyys	truncation kokonaisuosa
transmission rate ₁ tarttuvuus	t-SNE (<i>t</i> -distributed stochastic neighborhood embedding)
transmission rate ₂ tiedonsiirto- nopeus	t-statistic <i>t</i> -suure, <i>t</i> -tunnusluku
transportability siirrettävyys	t-test <i>t</i> -testi
transpose transpoosi	turning point taitekohta, käänne- kohta
trapdoor variable ansaluukku- muuttuja	twin study kaksostutkimus
traveling salesman problem kauppamatkustajan ongelma	two-armed bandit kaksikäätinen rosvo
treatment käsittely, hoito	two-by-two table (fourfold table) 2×2 -taulukko, nelikenttä
treatment arm hoitolinja, käsit- telyhaara	two-phase sampling (double sampling) kaksivaiheotanta
treatment contrast käsittely-	two-sided alternative hypothe- sis kaksisuuntainen vastahypotee- si, kaksisuuntainen vaihtoehtoinen hypoteesi
	two-sample t test kahden otok- sen <i>t</i> -testi

two-sided test (two-tailed test)
kaksisuuntainen testi, kaksitahoinen testi

two-stage least squares kaksivaiheinen pienin neliösumma

two-stage sampling kaksiasteotanta

two-tailed test (two-sided test)
kaksisuuntainen testi, kaksitahoinen testi

two-way analysis of variance
kaksisuuntainen varianssianalyysi

two-way classification kaksisuuntainen luokittelu

two-way contingency table kaksisuuntainen ristitaulukko, kaksuotteinen ristitaulukko

type I error (rejection error) hylkäysvirhe

type II error (acceptance error)
hyväksymisvirhe

U

ultrametric tree ultrametrisen puu

umbrella problem sateenvarjoongelma

UMP (uniformly most powerful)

UMPU (uniformly most powerful unbiased)

unadjusted estimator
vakioimaton estimaattori

unbalanced data epätasapainoinen aineisto

unbalanced design epätasapainoinen asetelma

unbiased harhaton

uncertainty epävarmuus

uncertainty quantification epävarmuuden määrittäminen

uncorrelated korreloimaton

unemployment rate työttömyysosuus, työttömyysaste

underadjustment vajaavakiointi

undercount (underenumeration)
laskennan alipeitto

undercoverage alipeitto

underdispersion alihajonta

underenumeration (undercount)
laskennan alipeitto

underfitting vajaasovitus

underidentification ali-identifioituvuus

undirected graph suuntaamaton verkko

unequal probability sampling
otanta vaihtelevin todennäköisyyksin

unidentifiability identifioimattomuus

uniform convergence tasainen suppeneminen

uniform correlation tasakorrelaatio

uniform distribution tasajakautuma

uniformly integrable tasaisesti integroituva

uniformly minimum risk kauttaaltaan pienin riski

uniformly most accurate interval kauttaaltaan tarkin väli

uniformly most powerful
(UMP) kauttaaltaan voimakkain, UMP

uniformly most powerful unbiased (UMPU) kauttaaltaan voimakkain harhaton

uniform sampling fraction
vakio-otantaosuus

unimodal yksihuippuinen

union yhdiste, unioni

union-intersection test yhdiste-leikkaustesti

uniqueness yksikäsitteisyys

unit yksikkö
unit circle yksikköympyrä
unit interval yksikköväli
unit non-response
 yksikkövastauskato
unit root yksikköjuuri
unit-root test yksikköjuuritest
univariate distribution
 yksiulotteinen jakauma, yhden
 muuttujan jakauma
univariate function yhden muut-
 tujan funktio
universe (population) perusjouk-
 ko, populaatio
unlabeled data nimiöimätön da-
 ta
unobserved confounder piilose-
 koittaja
unordered set järjestämätön
 joukko
unreduced design pelkistymätön
 asetelma
unsupervised classification oh-
 jaamaton luokittelu
unsupervised learning ohjaama-
 ton oppiminen
up-and-down method ylös-alas-
 menetelmä, tikkausmenetelmä
upper control limit
 ylävalvontaraja
upper quartile yläkvartiili
upper triangular matrix
 yläkolmiomatriisi
upward bias positiivinen harha,
 ylösharha
urn model uurnamalli
U-shaped distribution U:n muo-
 toinen jakauma
U-statistic *U*-tunnusluku
utility hyöty, utiliteetti
utility function hyötyfunktio
utility theory hyötyteoria, utili-
 teettiteoria

V

vague convergence löyhä suppe-
 neminen
vague prior heikko prior, laakea
 prior
validation validointi, vahvistus
validation data (validation set)
 validointiaineisto, vahvistusaineis-
 to
validation set (validation data)
 validointiaineisto, vahvistusaineis-
 to
validity validiteetti, pätevyys
Value at Risk (VaR) tappioarvo,
 VaR
value index arvoindeksi
value of information informaatio-
 arvo
VaR (Value at Risk)
VAR (vector autoregression)
variability (variation) vaihtelu
variable (variate) muuttuja
variable selection muuttujavalin-
 ta
variance varianssi
variance analysis (analysis of
 variance) varianssianalyysi
variance component
 varianssikomponentti
variance-covariance matrix
 (covariance matrix) kovarianssi-
 matriisi
variance inflation factor (VIF)
 varianssin inflaatiokerroin, VIF
**variance stabilizing trans-
 formation** varianssin vakioiva
 muunnos
variate (variable) muuttuja
variation (variability) vaihtelu
variational distance (total varia-
 tion distance) variaatioetäisyys
variational inference

variaatiopäätely
varimax rotation varimax-rotatio, varimax-kierto
variogrammi variogrammi
vector autoregression (VAR) vektoriautoregressio
vector time series vektori aikasarja
Venn diagram Venn-diagrammi
veracity totuudenmukaisuus
vertex (node) solmu
VIF (variance inflation factor)
violin plot viulukuvio
visualisation visualisointi
vital statistics väestönmuutostiedot, väestönmuutostilasto
volatility volatiliteetti, heilunta
volume, variety, velocity volyymi-vaihtelu-vauhti
volunteer bias vapaaehtoisharha
voting theory äänestysteoria

W

waiting time odotusaika
Wald statistic Waldin suure, Waldin tunnusluku
Wald test Waldin testi
Walsh averages Walshin keskiarvot
warm-up period (burn-in period) lämmittelyjakso
wash-out period puhdistumisjakso
wavelet aallocke, väre
weak convergence heikko suppeneminen, jakaumasuppeneminen
weakest-link model heikoimman lenkin malli
weak exogeneity heikko eksogeisuus
weakly stationary (covariance

stationary) kovarianssistationaarinen, heikosti stationaarinen
weight paino
weighted average (weighted mean) painotettu keskiarvo
weighted least squares method painotettu pienimmän neliösumman menetelmä
weighted likelihood painotettu uskottavuus
weighted mean (weighted average) painotettu keskiarvo
weight of evidence näytön paino
white noise valkoinen kohina
wide sense stationary (covariance stationary) kovarianssistationaarinen, laajassa mielessä stationaarinen
Wiener process Wiener-prosessi
Wilcoxon test Wilcoxonin testi
Wilks phenomenon Wilks-ilmiö
window ikkuna
window width ikkunanleveys
winsorized mean winsorkeskiarvo
Wishart distribution Wishartin jakauma
within-groups estimator ryhmien sisäinen estimaattori
within-groups sum of squares ryhmien sisäinen neliösumma
within-groups variance ryhmien sisäinen varianssi
word cloud (tag cloud) sanapilvi
word embedding sanan esitys, sanan upotus
working variate työmuuttuja
worm plot matokuvio
wrapped distribution kääritty jakauma

Y

Yates' correction Yatesin korjaus

Yule–Walker equations Yulen–Walkerin yhtälöt

Z

zero-altered model³³ (hurdle model) rimamalli

zero-inflated distribution nolla-

paisunut jakauma

zero-inflation nollapaisunta

zero matrix nollamatriisi

zero-sum game nollasummapeli

zero-truncated distribution
nollakatkaistu jakauma, nollatypistetty jakauma

zeta distribution zetajakauma

z-score *z*-arvo

z-test *z*-testi

z-transformation *z*-muunnos

³³Zero-altered model esitetään joskus synonyymiksi termille hurdle model, suomeksi rimamalli. Toisissa lähteissä zero-altered model tarkoittaa yleisemmän rimamallin (hurdle model) erikoistapausta, jossa nolla"-arvolle rakennetun mallin ja nollasta poikkeavien arvojen mallin selittäjä rakenne on täsmälleen samanlainen.

7 Suomi–englanti

A

- aallope** wavelet
absoluuttinen asteikko absolute scale
absoluuttinen riski absolute risk
absoluuttisesti jatkuva absolutely continuous
absorboiva alue absorbing region
absorboiva reuna absorbing barrier
absorboiva tila absorbing state
absorptiojakauma absorption distribution
A/B-testi A/B-test
additiivinen kausivaihtelumalli additive seasonal model
additiivinen malli additive model
additiivinen prosessi additive process
additiivisten uhkien malli additive hazards model
agenttipohjainen malli agent-based model
aggregaatti-imputointi aggregate imputation
aggregointi aggregation
aidosti stationaarinen prosessi strictly stationary process
aihemallitus topic modeling
aika-aggregointi temporal aggregation
aika-alue time domain
aikahomogeenisuus time-homogeneity, temporal homogeneity
aikakääntyvä prosessi reversible process
aikamuuttuva time-varying
aikariippumaton time-independent
aikariippuva time-dependent
aikariippuva jakauma time-dependent distribution
aikariippuva selittäjä time-dependent covariate
aikasarja time series
aikasarja-analyysi time-series analysis
aikasarjaekonometria time-series econometrics
aika-taajuusjakauma time-frequency distribution
aineisto data
aineistoanalyysi data analysis
aineiston esikäsittely data preprocessing
aineiston esitarkastus data screening
aineiston jalostaminen data wrangling
aineiston kurkistelu data peeking
aineiston siivoaminen data cleaning
aineistovetoinen data driven
aito eksogeenisuus strict exogeneity
aito priori proper prior
ajasta riippumaton time-independent
ajautuminen drift₂
Akaiken informaatiokriteeri Akaike information criterion
akselijakauma axial distribution
aktiivinen oppiminen active learning
aktivointifunktio activation func-

tion	exposed-infectious-recovered
aktuaariestimaattori actuarial estimator	altis-tartuttava-toipunut susceptible-infectious-recovered
alakolmiomatriisi lower triangular matrix	altistus exposure
alakvartiili lower quartile	altistusaika exposure time
alaryhmäanalyysi subgroup analysis	alttius propensity
alasaskeltava menettely step-down procedure	alttiuskaltaistus propensity matching
alavaltontaraja lower control limit	alttiusluku propensity score
algoritmi algorithm	alttiuspainotus propensity weighting
aliavaruus subspace	aluedata spatial data
alieksponenttinen subexponential	alueotanta area sampling
aligaussinen jakauma sub-Gaussian distribution	aluetilastotiede spatial statistics
alihajonta underdispersion	alustava kartoitus exploratory survey
ali-identifioituvuus underidentification	amerikkalainen optio American option
alijakauma subdistribution	ammekäyrä bathtub curve
alikeräysfunktio subdistribution function	amplitudi amplitude
alimartingaali submartingale	analytiikka analytics
aliotos subsample	analyttinen hierarkiaprosessi analytic hierarchy process
alipeitto undercoverage	anisotrooppinen anisotropic
alistava muuttuja suppressor variable	annometometri dose meter
alitiheysfunktio subdensity function	annomuunnos dose meter
alkeiskontrasti elementary contrast	annos-vastekäyrä dose-response curve
alkeistapahtuma elementary event	annuiteetti annuity
alkeistapaus elementary event	anomalia anomaly
alkeisyksikkö elementary unit	ansaluukkumuuttuja trapdoor variable
alkio element	antiteettinen antithetic
alkujakauma initial distribution	antitoninen regressio antitonic regression
alleelifrekvenssi allele frequency	approksimaatio approximation ₂
allometria allometry	approksimointi approximation ₁
altis-tartunnan saanut-tartuttava-toipunut susceptible-	apuinformaatio ancillary information
	apumuuttuja auxiliary variable
	aputunnusluku ancillary statistic
	arbitraasi arbitrage
	aritmeettinen jakauma arith-

metic distribution
aritmeettinen keskiarvo arithmetic mean
arkaluonteisuus sensitivity₂
arkussinijakauma arcsine distribution
arkussinimuunnos arcsine transformation
arvio estimate
arviofunktio estimator
arviointi estimation
arviointivirhe estimation error
arviointiyhtälö estimating equation
arvioituva estimable
arvoindeksi value index
arvopaperi asset
arvopaperisalkun optimointi portfolio optimization
asetelma design
asetelma-avaruus design space
asetelmakerroin design effect
asetelmamatriisi design matrix
asetelmaoptimaalisuus design optimality
asetelmaperusteinen päättely design-based inference
asiantuntija-arvio expert judgment
asiantuntijajärjestelmä expert system
asiantuntijaliikavarmuus expert overconfidence
askeltava regressio stepwise regression
assosiaatio association
assosiaatioanalyysi association analysis
assosiaatiokerroin coefficient of association
assosiaatiomitta measure of association
aste₁ rank₂
aste₂ proportion, rate₂
astejakauma degree distribution
asymptoottinen asymptotic
asymptoottinen suhteellinen tehokkuus asymptotic relative efficiency
asymptotiikka asymptotics
atomi atom
attraktioalue domain of attraction
attraktori attractor
aukkosuure gap statistic
auringonkukkakuvio sunflower plot
autokorrelaatio autocorrelation, serial correlation
autokorrelaatiofunktio autocorrelation function
autokovarianssi autocovariance
autokovarianssiemäfunktio autocovariance generating function
autokovarianssifunktio autocovariance function
autokovarianssit generoiva funktio autocovariance generating function
automaattinen interaktiohaku automatic interaction detection
automaattinen yhdysvaikutushaku automatic interaction detection
autoregressiivinen ehdollisen heteroskedastisuuden malli autoregressive conditional heteroskedastic model
autoregressiivinen eksogeenisten viipeiden malli autoregressive distributed-lag model
autoregressiivinen fraktionaalisesti integroitu liukuvan keskiarvon malli autoregressive fractionally integrated moving average model

autoregressiivinen integroitu liukuvan keskiarvon malli autoregressive integrated moving average model
autoregressiivinen liukuvan keskiarvon malli autoregressive moving average model
autoregressiivinen malli autoregressive model
autoregressio autoregression
avaruus space
avioliittoisuus nuptiality
avoin data open data
avoin päätyluokka open-ended class
avoin väestö open population

B

bayesiläinen Bayesian
Bayes-informaatiokriteeri Bayesian information criterion
Bayesin kaava Bayes' formula
Bayes-kerroin Bayes factor
bayesläinen Bayesian
Bayes-päivitys Bayesian updating
Bayes-päätely Bayesian inference
Bayes-riski Bayes risk
Bayes-tekijä Bayes factor
Bayes-verkko Bayes network
Bernoulli-jakauma Bernoulli distribution
Bernoulli-koe Bernoulli experiment, Bernoulli trial
beta-binomijakauma beta-binomial distribution
betajakauma beta distribution
bikuvio biplot
bilineaarinen malli bilinear model
binomijakauma binomial distri-

bution
binomikerroin binomial coefficient
binomikoe binomial experiment, binomial trial
binomitesti binomial test
binäärimuuttuja binary variable, dichotomous variable
bioekvivalenssitesti bioequivalence test
bioinformatiikka bioinformatics
biometria biometry, biometrics₁
biometriikka biometric identification, biometrics₂
biometrinen tunnistaminen biometrics₂
biomääritys bioassay
biostatistiikka biostatistics
biotilastotiede biostatistics
biseriaalinen korrelaatio biserial correlation
bispektri bispectrum
bitti bit
Blackin–Scholesin malli Black–Scholes model
Bonferronin epäyhtälö Bonferroni inequality
Brownin liike Brownian motion
Brownin silta Brownian bridge
bruttomuutto gross migration
bruttouusiutumisluku gross reproduction rate

C

cauchit-muunnos cauchit transformation
Cauchy-jakauma Cauchy distribution
Chernoffin kasvokuviot Chernoff's faces
cloglog-muunnos cloglog-transformation, complementary loglog

transformation
Cookin etäisyys Cook's distance
Coxin malli Cox model, proportional hazards model
Cramérin–Raon alaraja
 Cramér–Rao lower bound
Cronbachin alfa Cronbach's alpha
CUSUM-kaavio CUSUM chart, cumulative sum chart

D

Dantzig-valitsin Dantzig selector
data data
data-analyysi data analysis
datajoukko data set
datakehikko data frame
datakuutio data cube
datamatriisi data matrix
datan esikäsittely data pre-processing
datan jalostaminen data wrangling
datan kurkistelu data peeking
datan linkittäminen data linkage
datanlouhinta data mining
datan myrkyttäminen data poisoning
datan ruoppaus data dredging
datan siivoaminen data cleaning
datan syvyys data depth
datan tiivistäminen data reduction
dataputki data pipeline
datatiede data science
datavetoinen data driven
de facto -väestö de facto population
deflaatio deflation
degeneroitunut jakauma degenerate distribution
de jure -väestö de jure population
dekomponointi decomposition₁
dekomponoituva malli decomposable model
dekonvoluutio deconvolution
deltamenetelmä delta method
demografia demography
derivaatta derivative₁
derivoituvuus differentiability
desiili decile
determinaatiokerroin coefficient of determination
determinantti determinant
deterministinen deterministic
devianssi deviance
devianssianalyysi analysis of deviance
devianssi-informaatiokriteeri
 deviance information criterion
devianssijäännös deviance residual
diagnostiikka diagnostics
diagnostinen suure diagnostic statistic
diagnostinen testi diagnostic test
diagnostinen tunnusluku diagnostic statistic
diagonaalimatriisi diagonal matrix
diagrammi diagram
differenssiyhtälö difference equation
differensointi differencing
differentiaaliyhtälö differential equation
differentioituvuus differentiability
differointi differencing
diffuusiaindeksi diffusion index
diffuusioprosessi diffusion process

diffuusi priori diffuse prior
digammafunktio digamma function
dikotomia dichotomy
dikotominen muuttuja dichotomous variable, binary variable
dimensiokirous curse of dimensionality
dimension pienennys dimension reduction
Diracin deltafunktio Dirac's delta function
Diracin mitta Dirac measure
Dirichlet-jakauma Dirichlet distribution
Dirichlet-prosessi Dirichlet process
diskonttaus discounting
diskordanssi discordance
diskordantit parit discordant pairs
diskreetti discrete
diskreettiaikainen discrete-time
diskreetti jakauma discrete distribution
diskreetti muuttuja discrete variable
diskriminantti discriminant
divergenssi divergence₁
diversiteetti-indeksi index of diversity
do-laskenta do-calculus
do-operaattori do-operator
***D*-optimaalisuus** *D*-optimality
***d*-separaatio** *d*-separation
duaalilause dual theorem
duaalisuus duality
dynaaminen kiintiöinti dynamic treatment allocation
dynaaminen ohjelmointi dynamic programming, dynamic optimization
dynaaminen optimointi dy-

namic optimization
dynaaminen populaatio dynamic population, open population

E

editointisääntö edit rule
edustava otos representative sample
edustavuus representativeness
ehdokasjakauma candidate distribution, proposal distribution
ehdollinen autoregressio conditional autoregression
ehdollinen jakauma conditional distribution
ehdollinen keskimääräinen
 kausaalivaikutus conditional average causal effect
ehdollinen keskimääräinen
 käsittelyvaikutus conditional average treatment effect
ehdollinen odotusarvo conditional expectation
ehdollinen riippumattomuus conditional independence
ehdollinen riippuvuus conditional dependence
ehdollinen tiheysfunktio conditional density function
ehdollinen todennäköisyys conditional probability
ehdollinen uskottavuus conditional likelihood
ehdollinen varianssi conditional variance
ehdollinen yhteys conditional association
ehdollistettu sirontakuvio conditioning plot, coplot
ehdotusjakauma proposal distribution, candidate distribution

ehkäisykoe	prevention trial	tial trend	
ei-merkitsevä	nonsignificant	eksponenttiperhe	exponential family
ei-negatiivisesti definiitti	non-negative definite	ekstrapolointi	extrapolation
ei-normaali	non-normal	ekvivalenssiluokka	equivalence class
ei-satunnainen	non-random	ekvivalenssitestaus	equivalence testing
ei-satunnaisesti puuttuva	missing not at random	ekvivariantti estimaattori	equivariant estimator
ekologinen harha	ecological fallacy, ecological bias	elastisuus	elasticity
ekologinen korrelaatio	ecological correlation	eliminointi	elimination
ekologinen tutkimus	ecological study	elinaika	lifetime
ekologinen virhepäätelmä	ecological fallacy	elinaika-analyysi	lifetime data analysis, survival analysis
ekonometria	econometrics	elinaikajakauma	lifetime distribution
ekonometrikko	econometrician	elinajan jakauma	lifetime distribution
eksakti testi	exact test	elinajanodote	life expectancy
eksogeeninen muuttuja	exogenous variable	elinajantaulukko	life table
eksogeenisten viipeiden malli	distributed-lag model	elinkaarianalyysi	life course analysis
eksogeenisuus	exogeneity	elinkaarihyöty	life cycle utility
eksponentiaalinen kasvu	exponential growth	elinkorko	annuity ₁
eksponentiaalinen trendi	exponential trend	elliptinen jakauma	elliptical distribution
eksponenttijakauma	exponential distribution	elossaoloanalyysi	survival analysis
eksponenttikäyrä	exponential curve	elossaolofunktio	survival function
eksponenttinen hajontamalli	exponential dispersion model	elossaolokäyrä	survival curve
eksponenttinen hajontaperhe	exponential dispersion family	eläkerahasto	pension fund
eksponenttinen kasvu	exponential growth	EM-algoritmi	EM-algorithm, expectation-maximization algorithm
eksponenttinen silotus	exponential smoothing	empiirinen	empirical
eksponenttinen tasoitus	exponential smoothing	empiirinen Bayes-estimointi	empirical Bayes estimation
eksponenttinen trendi	exponential trend	empiirinen kertymäfunktio	empirical distribution function
		empiirinen mitta	empirical measure
		empiirinen peittotoden-	

näköisyys	empirical coverage	ennustin	predictor
empiirinen uskottavuus	empirical likelihood	ennätys	record ₂
emäfunktio	generating function	ensiasteen otantayksikkö	first stage sampling unit, primary sampling unit
endogeeninen muuttuja	endogenous variable	ensimmäisen kertaluvun malli	first order model
endogeenisuus	endogeneity	ensinumeron jakauma	first digit distribution
ennakkorekisteröinti	preregistration	ensipaluun hetki	first return time
ennakoituva prosessi	predictable process	ensikulun hetki	first passage time
ennakoiva indikaattori	leading indicator	entropia	entropy
ennakoiva säätö	anticipative control	<i>E</i>-optimaalisuus	<i>E</i> -optimality
ennalta asetettu merkitsevyystaso	pre-specified significance level	epidemiamalli	epidemic model
ennustaminen	forecasting, prediction	epidemiologia	epidemiology
ennustava analytiikka	predictive analytics	epäaito esijakauma	improper prior
ennustava muuttuja	predictor	epäaito jakauma	improper distribution
ennustava päättely	predictive inference	epäaito otanta	non-probability sampling
ennuste₁	forecast	epäaito prior	improper prior
ennuste₂	prediction	epäinformatiivinen prior	non-informative prior
ennuste₃	predicted value	epäjatkuva	discontinuous
ennustearvo	predictive value	epäkelpo estimaattori	inadmissible estimator
ennustejakauma	predictive distribution	epäkelpo imputointi	improper imputation
ennustejäännös	prediction residual	epäkeskinen	non-central
ennustejäännösten neliösumma	prediction residual sum of squares	epäkeskisyyssparametri	non-centrality parameter
ennustekeskiarvokaltaistus	predictive mean matching	epäkoikeellinen tutkimus	non-experimental study
ennustemalli	prognostic model	epälineaarinen	nonlinear
ennustetekijä	prognostic factor	epälineaarinen malli	nonlinear model
ennustettu arvo	predicted value	epälineaarinen optimointi	nonlinear optimization
ennustevirhe	prediction error	epälineaarinen regressio	nonlinear regression
ennusteväli	prediction interval		

epäortogonaalinen	non-orthogonal	hyperplane	
epäparametrinen	nonparametric	erotteleva tunnusluku	complete statistic
epäsingulaarinen	nonsingular	erotteluanalyysi	discriminant analysis
epäsisäkkäinen luokitus	nonnested classification	erottelufunktio	discriminant function
epäsuora otanta	indirect sampling	erottelukyky	discriminatory power
epäsuora pienin neliösumma	indirect least squares	erottelukykysoitin	receiver operating characteristic
epäsuora vaikutus	indirect effect	erottelukykykäyrä	receiver operating characteristic curve, ROC curve
epäsuora vakiointi	indirect standardization	erotuksen itseisarvo	absolute difference
epäsuotuisin jakauma	least favorable distribution	erotus	difference
epätasapainoinen aineisto	unbalanced data	erotusoperaattori	difference operator
epätasapainoinen asetelma	unbalanced design	erotusten erotus	difference in differences
epävarmuuden määrittäminen	uncertainty quantification	erotusten laskeminen	differencing
epävarmuus	uncertainty	erä	lot
ergodinen	ergodic	eräkato	item non-response
ergodisuus	ergodicity	eräpäivä	maturity date
erhe	error	erävaihtelu	batch variation
eriarvoisuuskerroin	inequality coefficient	esiennustejakauma	prior predictive distribution
erilaiset parit	discordant pairs	esiintyvyys	occurrence
erilaisuusindeksi	index of dissimilarity	esijakauma	prior distribution
erilaisuuskerroin	dissimilarity coefficient	esijakauman koostaminen	elicitation of prior
erilaisuusmatriisi	dissimilarity matrix	esiluotaus	exploratory survey
erilliset tapahtumat	mutually exclusive events	esitodennäköisyys	prior probability
erivarianssinen	heteroskedastic	esitutkimus	exploratory survey, pilot study
eriytyminen	segregation ₁	esivalkaisu	prewhitening
eriytymisanalyysi	segregation analysis	estimaatti	estimate
eriävä virheluokittelu	differential misclassification	estimaattori	estimator
erottava hypertaso	separating	estimandi	estimand
		estimointi	estimation

estimointifunktio estimating function
estimointivirhe estimation error
estimointiyhtälö estimating equation
estimoitava suure estimand
estimoituva estimable
eteenyhtälö forward equation
etenevä valinta forward selection
etiologinen osuus etiologic fraction
etsintä-hyödynnyskompromissi exploration vs. exploitation trade-off
etsivä data-analyysi exploratory data analysis, EDA
etumerkkinen neliöjuurimuunnos signed root-transformation
etuovivakiointi frontdoor adjustment
etäisyysmitta distance measure
etäisyysotanta distance sampling
euklidinen etäisyys Euclidean distance
eurooppalainen optio European option

F

faktori factor₂
faktorianalyysi factor analysis
faktorikierto factor rotation
faktorilataus factor loading
faktorimatriisi factor matrix
faktoripistemäärä factor score
faktorirakenne factor pattern
faktorirotaatio factor rotation
faktorointilause factorization theorem
fenotyyppi phenotype
fidusiaalinen päättely fiducial inference
fi-kerroin phi-coefficient
filtraatio filtration
Fisherin estimointialgoritmi Fisher's scoring method
Fisher-informaatio Fisher information
Fisher-informaatiomatriisi Fisher information matrix
Fisherin jakauma Fisher's distribution, F -distribution
Fisher-tarkentuvuus Fisher consistency
 F -jakauma F -distribution
 F -mitta F score
Fourier-analyysi Fourier analysis
Fourier-muunnos Fourier transform
fraktaali fractal
fraktiili fractile, quantile
fraktionaalinen Bayes-tekijä fractional Bayes factor
fraktionaalinen Brownin liike fractional Brownian motion
frekvenssi frequency
frekvenssihistogrammi frequency histogram
frekvenssijakauma frequency distribution
frekvenssikertymä cumulative frequency
frekvenssikäyrä frequency curve
frekvenssipinta frequency surface
frekvenssitaulukko frequency table
frekventistinen frequentist
frekventistinen päättely frequentist inference
 F -suure F statistic
 F -testi F -test
funktionaalideterminantti functional determinant, Jacobian determinant
funktionaalinen data-analyysi

functional data analysis
funktionaalinen keskeinen raja-arvolause functional central limit theorem
funktionaalinen yhteys functional relationship
futuurisopimus futures contract

G

Galtonin–Watsonin prosessi Galton–Watson process
gammafunktio gamma function
gammahaurausmalli gamma-frailty model
gammajakauma gamma distribution
gammakerroin gamma coefficient
Gaussin eliminointimenetelmä Gaussian elimination
Gaussin kvadratuuri Gaussian quadrature
Gaussin jakauma Gaussian distribution, normal distribution
Gaussin–Markovin lause Gauss–Markov theorem
Gaussin prosessi Gaussian process
Gaussin sekoitemalli Gaussian mixture model
geenikartoitus gene mapping
geneettinen algoritmi genetic algorithm
geneettinen heterogeenisuus genetic heterogeneity
generoiva funktio generating function
genominlaajuinen assosiaatiokartoitus genomewide association study
geometrinen jakauma geometric distribution
geometrinen keskiarvo geomet-

ric mean
geometrinen todennäköisyys geometric probability
geotilastotiede geostatistics
g-estimointi g-estimation
Gibbsin algoritmi Gibbs sampler
Gibbsin poimija Gibbs sampler
Gini-indeksi Gini index
Gini-kerroin Gini index
g-kaava g-kaava
graafi graph₃
graafinen estimaattori graphical estimator
graafinen malli graphical model
Gramin matriisi Gramian
GREG-estimointi GREG estimation, generalized regression estimation
Guttmanin asteikko Guttman scale

H

haarautumisprosessi branching process
haastattelijaharha interviewer bias
hahmontunnistus pattern recognition
haittaparametri nuisance parameter
hajaantuminen divergence
haje entropy
hajonta dispersion, spread, scatter
hajonta-analyysi analysis of dispersion
hajontaindeksi index of dispersion
hajontakuvio scatter plot, scatter diagram
hajontakuviomatriisi scatterplot

matrix	hasardi hazard
hajontaluku measure of dispersion, measure of variation	hattumatriisi hat matrix
hajontamatriisi dispersion matrix	haurausmalli frailty model
hajontamitta measure of dispersion, measure of variation	havainnoiva tutkimus observational study
hajotelma decomposition ₂	havainnon sisäisyys data depth
hakutaulu lookup table	havainnot tuottava prosessi data generating process
halovaikutus halo effect	havainto observation
Hamming-etäisyys Hamming distance	havaintoaineisto data
Hamming-koodi Hamming code	havaintokuutio data cube
Hann-ikkuna Hann-window	havaintomatriisi data matrix
hapsukuvio rug plot	havaintosuure statistic
haravointi raking	havaintovirhe observational error
harha bias, systematic error	havaintoyksikkö observational unit
harhainen estimaattori biased estimator	havaittava observable
harhainen otanta biased sampling	havaittu informaatio observed information
harhainen testi biased test	havaittu informaatiomatriisi observed information matrix
harhaton unbiased	havaittu lukumäärä observed frequency, observed number
harha-varianssikompromissi bias-variance tradeoff	havaittu merkitsevyystaso <i>p</i> -value, observed significance level
harjaregressio ridge regression	Heavisiden funktio Heaviside function
harjanneregressio ridge regression	hedelmällisyys fertility
harkintänäyte judgment sample, purposive sample	hedelmällisyysluku fertility rate
harkintavalinta purposeful selection	hedoninen regressio hedonic regression
harmaan laatikon malli grey box model	hehkutettu jakauma annealed distribution
harmoninen jakauma harmonic distribution	heijastava reuna reflecting barrier
harmoninen keskiarvo harmonic mean	heijastusperiaate reflection principle
harmoninen regressio harmonic regression	heikko eksogeenisuus weak exogeneity
harva matriisi sparse matrix	heikko priori vague prior
harvennus thinning	heikko suppeneminen weak convergence
harventuminen attrition	heikoimman lenkin malli weak-

est-link model	hintaindeksi price index
heikosti stationaarinen weakly stationary	histogrammi histogram
heilahteleva prosessi oscillatory process	historia history, filtration
heilahtelu oscillation	hitaasti vaihteleva funktio slowly varying function
heilunta volatility	hoito treatment
Helmert-kontrasti Helmert contrast	hoitoaikeen mukainen analyysi intention-to-treat analysis
henkilöaika person-time	hoitolinja treatment arm
henkilökohtainen todennäköisyys personal probability	hoitomyöntyvyys compliance with treatment, adherence to treatment
henkilövuosi person-year	hoitoryhmä treatment group
henkivakuutus life insurance	hoitositoutuneisuus compliance with treatment, adherence to treatment
heritabiliteetti heritability	homogeeninen homogeneous
herkkyys sensitivity ₁	homoskedastinen homoskedastic
herkkyysanalyysi sensitivity analysis	HPD-alue HPD region, highest posterior density region
Hessen matriisi Hessian	HPD-väli HPD interval, highest posterior density interval
heterogeeninen heterogeneous	huipukkuus peakedness
heteroskedastinen heteroskedastic	huonosti asetettu ongelma ill-posed problem
heteroskedastisuus-tarkentuva t-suhde heteroskedasticity-consistent <i>t</i> -ratio	hylkäysalue rejection region, critical region
hetkellinen tila instantaneous state	hylkäysotanta rejection sampling
hetkellinen väestö de facto population	hylkäysraja critical value
hevosenkenkäilmiö horseshoe effect	hylkäystodennäköisyys power ₁
hieno järjestys fine ordering	hylkäysvirhe rejection error, false rejection
hierarkkinen luokittelu hierarchical classification	hylkäävä otanta rejective sampling
hierarkkinen malli hierarchical model, multilevel model	hylättävä laatutaso rejectable quality limit
hierarkkinen ryhmittely hierarchical clustering	hypereksponenttijakauma hyperexponential distribution
hila grid, lattice	hypergeometrinen jakauma hypergeometric distribution
hila-asetelma lattice design	hyperparametri hyperparameter
hilajakauma lattice distribution	hypoteesi hypothesis
hilaneliö lattice square	hypoteesin testaus hypothesis
hilaotanta lattice sampling	

testing
hyppytaajuus jump rate
hyppytodennäköisyys jump probability
hyväksymisalue acceptance region
hyväksymisotanta acceptance sampling
hyväksymisraja acceptance boundary, acceptance line
hyväksymisvalvontakaavio acceptance control chart
hyväksymisvirhe acceptance error, false acceptance
hyväksyttävä admissible
hyväksyttävä estimaattori permissible estimator
hyväksyttävä laatutaso acceptable quality limit
hyöty gain₂
hyöty utility
hyötyfunktio utility function
hyötyteoria utility theory
häiriö perturbation₁
häiriöaltis matriisi ill-conditioned matrix
häiriömenetelmä perturbation method
häly noise
hätäeksponentti tail exponent
hölynpölykorrelaatio nonsense correlation

I

ideaali-indeksi ideal index
idempotentti idempotent
identifioimattomuus unidentifiability
identifioitavuus identifiability
identiteettifunktio identity function
ihanneindeksi ideal index

i.i.d. independent and identically distributed
i'ittäinen age-specific
i'ittäinen kuolevuus age-specific mortality rate, age-specific death rate
ikkuna window
ikkunanleveys window width, bandwidth₂
ikäkasautuminen age heaping
ikä-periodi-kohorttimalli age-period-cohort model
ikäpyramidi age pyramid
ikäriippuva prosessi age-dependent process
ikäryhmittäinen age-specific
ikäryhmittäinen kuolevuus age-specific death rate, age-specific mortality rate
ikäsuosio age heaping
ikävakioitu age-adjusted, age-standardized
ilmaantumisosuus incidence proportion, cumulative incidence
ilmaantumistiheys incidence rate
ilmaantuvuus incidence
ilmaisín indicator variable
ilmiasu phenotype
ilmimuuttuja manifest variable
ilmiömalli substantive model
implisiittinen ositus implicit stratification
impulssivastefunktio impulse response function
imputointi imputation
imputointimalli imputation model
indeksi index, index number
indeksikuvio index plot
indeksiluku index number, index
indikaattorifunktio characteristic function₂

indikaattorimatriisi	incidence matrix ₁	integroitu spektri	integrated spectrum
indikaattorimuuttuja	indicator variable	integroituva	integrable
induktiivinen käyttäytyminen	inductive behaviour	intensiiteetti	intensity, rate ₁
induktiivinen päättely	inductive inference	intensiiteettifunktio	intensity function
induktio	induction	intensiiteettikertymä	cumulative intensity, integrated intensity
infektiokuolleisuus	infection fatality ratio	interaktio	interaction
inflaatio	inflation	interpolaatio	interpolation
informaatio	information	interpolaiva splini	interpolating spline
informaatioarvo	value of information	intervalliasasteikko	interval scale
informaatiofunktio	information function	interventioanalyysi	intervention analysis
informaatiokriteeri	information criterion	interventiotutkimus	intervention study
informaatiomatriisi	information matrix	intialainen noutopöytäprosessi	Indian buffet process
informaatioteoria	information theory	invarianssi	invariance
informatiivinen esijakauma	informative prior	invariantti	invariant
informatiivinen prior	informative prior	invariantti jakauma	invariant distribution
informatiivinen puuttuvuus	informative missingness	inversio	inversion
informatiivinen sensurointi	informative censoring	inversio-ongelma	inverse problem
innovaatiohyöty	innovation gain	isometria	isometry
innovaatiotermi	innovation term	isotoninen regressio	isotonic regression
insinööritilastotiede	engineering statistics	isotrooppinen	isotropic
instrumenttimuuttuja	instrumental variable	isovuodata	high-throughput data
integraali	integral	isovuoseulonta	high-throughput screening
integroitu keskineliövirhe	mean integrated squared error	isäaalloke	father wavelet
integroitu liukuvan keskiarvon malli	integrated moving average model	isäväre	father wavelet
		iteraatio	iteration
		iteratiivinen verrannollinen sovitus	iterative proportional fitting
		iteratiivisesti painotettu pienin neliösumma	iteratively weighted least squares
		iteroidun logaritmin laki	law of

the iterated logarithm
iterointi iteration
iteroitujen odotusarvojen laki law of the iterated expectations
iteroitu uusio-otanta iterated bootstrap
Itô-integraali Itô integral
itseisarvo absolute value
itseiserotus absolute difference
itseisesti summautuva absolutely summable
itseisesti suppeneva absolutely convergent
itseismomentti absolute moment
itseispoikkeama absolute deviation
itseisvirhe absolute error
itsekehkeytyvä prosessi self-exciting process
itsensä estävä ennuste self-negating prophecy
itsensä rahoittava self-financing
itsensä toteuttava ennuste self-fulfilling prophecy
itseorganisoituva kartta self-organizing map
itsepainottuva otos self-weighting sample
itsesimilaarinen prosessi self-similar process
itsevalikoitunut otos self-selected sample
ITT-analyysi intention-to-treat analysis

J

jackknife-estimaattori jackknife estimator
Jacobin determinantti Jacobian determinant, functional determinant
Jacobin matriisi Jacobian ma-

trix
jakauma distribution
jakaumakonvergenssi convergence in distribution
jakauman häntä tail of a distribution
jakaumaseos mixture distribution
jakaumasta riippumaton distribution-free
jakaumasuppeneminen convergence in distribution, weak convergence
jakaumien sekoitus mixture distribution
jakso cycle
jaksokuvio periodogram
jaksollinen malli cyclic model
jaksollinen prosessi periodic process
jaksoton aperiodic
Jamesin–Steinin estimaattori James–Stein estimator
jatkosuhdemalli continuation ratio model
jatkuva-aikainen prosessi continuous-time process
jatkuva jakauma continuous distribution
jatkuvuus continuity
jatkuvuuskorjaus continuity correction
Jeffreysin priori Jeffreys prior
Jensenin epäyhtälö Jensen’s inequality
J:n muotoinen jakauma J-shaped distribution
johdannainen derivative₂
johdatteleva kysymys leading question
johdonmukaisuus coherence
jono queue
jonoteoria queueing theory

joukko set
joustava verkko elastic net
jousto elasticity
joutojakso idle period
julkaisuharha publication bias
juovakuvi stripes plot
jurimetria jurimetrics
juurellinen verkko rooted graph
jäljellä oleva elinaika residual lifetime
jäljellä oleva elinajanodote remaining life expectancy
jäljellä oleva odotusaika residual waiting time
jäljitysharha ascertainment bias
jälkeläisjakauma offspring distribution
jälki trace
jälkiennustejakauma posterior predictive distribution
jälkihypotisointi HARKing
jälkijakauma posterior distribution
jälkikäteisvertailu post hoc comparison
jälkiositus post-stratification
jälkiryvästys post-clustering
jälkitodennäköisyys posterior probability
jälkivaikutus carry-over effect
jälleenvakuutus reinsurance
järjestetty luokkamuuttuja ordered categorical variable, ordinal categorical variable
järjestetty otos ordered sample
järjestysasteikko ordinal scale
järjestysasteikollinen aineisto ordinal data
järjestysasteikollinen muuttuja ordinal scale variable
järjestyskorrelaatio rank correlation
järjestysluku rank₁

järjestysmuunnos rank transformation
järjestysmuuttuja ordinal variable
järjestysregressio ordinal regression
järjestystunnusluku order statistic
järjestämätön joukko unordered set
jäännös residual
jäännösdevianssi residual deviance
jäännösdiagnostiikka residual diagnostics
jäännöshajonta residual standard deviation
jäännöskeskineliö residual mean square, error mean square
jäännöskuvio residual plot
jäännösneliösumma residual sum of squares, error sum of squares
jäännösvaikutus residual effect
jäännösvarianssi residual variance
jäännösvirrehajonta root mean squared error

K

kaareva eksponenttiperhe curved exponential family
kaavio diagram
kahden muuttujan jakauma bivariate distribution
kahden otoksen t-testi two-sample t test
kahden yksikön sisältymistodennäköisyys second-order inclusion probability
kahdesti satunnainen malli doubly stochastic model

kahtiajako	dichotomy	kaksoiseksponenttijakauma	double exponential distribution
kaistanleveys	bandwidth ₁	kaksoisrekisteriestimointi	dual systems estimation
kaksiarvoinen annos-vaste	quantal dose-response	kaksoissokkokoe	double-blind experiment
kaksiarvoinen muuttuja	binary variable, dichotomous variable	kaksoisvakaa estimointi	doubly robust estimation
kaksiasteinen pienin neliösumma	two-stage least squares	kaksostutkimus	twin study
kaksiastejakauma	bidegree distribution	kalibrointi	calibration
kaksiasteotanta	two-stage sampling	kalibroitu paino	calibrated weight
kaksihuippuinen jakauma	bimodal distribution	Kalman-hyöty	Kalman gain
kaksijakoinen	bipartite	Kalman-suodin	Kalman filter
kaksikätkäinen rosvo	two-armed bandit	kaltaistettu pari	matched pair
kaksisuuntainen luokittelu	two-way classification	kaltaistettu tapaus-verrokki-tutkimus	matched case-control study
kaksisuuntainen ristitaulukko	two-way contingency table	kaltaistetut otokset	matched samples
kaksisuuntainen testi	two-sided test, two-tailed test	kaltaistus	matching
kaksisuuntainen vaihtoehtoinen hypoteesi	two-sided alternative hypothesis	kaltaistusjoukko	matched set
kaksisuuntainen varianssianalyysi	two-way analysis of variance	kanoninen analyysi	canonical analysis
kaksisuuntainen vastahypoteesi	two-sided alternative hypothesis	kanoninen korrelaatio	canonical correlation
kaksitahoinen testi	two-sided test, two-tailed test	kanoninen linkkifunktio	canonical link function
kaksiulotteinen jakauma	bivariate distribution	kanoninen parametri	canonical parameter, natural parameter
kaksiulotteinen normaali-jakauma	bivariate normal distribution	kanta	basis
kaksiulotteinen ristitaulukko	two-way contingency table	kantaja	support
kaksivaiheotanta	two-phase sampling, double sampling	kantajamuuttuja	carrier variable
		Kaplanin–Meierin estimaattori	Kaplan–Meier estimator, product limit estimator
		kappakerroin	kappa coefficient
		karakteristinen funktio	characteristic function
		karakteristinen yhtälö	characteristic equation
		karhunta	call-back
		karkaistu jakauma	quenched

distribution	kausipuhdistus seasonal adjustment
karkea data coarse data	kausivaihtelu seasonal variation
karkea estimaattori crude estimator	kauttaaltaan pienin riski uniformly minimum risk
karkein tyhjentävä tunnusluku minimal sufficient statistic	kauttaaltaan tarkin väli uniformly most accurate interval
karkeistus coarsening	kauttaaltaan voimakkain uniformly most powerful
karsinta pruning	kehikko frame
kartiosilotus tapering	kehikkoperusjoukko frame population
kartogrammi cartogram	kehityspolku trajectory
kartoitus survey	kehityssuunta trend
kaskadiprosessi cascade process	kehys frame
kasvukäyrä growth curve	kehäaineisto circular data
kasvunopeus growth rate	kehädata circular data
katastrofiteoria catastrophe theory	kehäjakauma circular distribution
katkaistu jakauma truncated distribution	keinoäly artificial intelligence
katkaisu oikealta right-truncation	kelpoinen malli adequate model
katkaisupiste cut-off point	kelvollinen imputointi proper imputation
kato dropout, attrition, loss	kemometria chemometrics
kattava malli encompassing model	Kendallin tau Kendall's tau
kattovaikutus ceiling effect	kenttäkoe field experiment
kaukokartoitus remote sensing	kenttätö field work
kauppamatkustajan ongelma traveling salesman problem	kerrannainen kausivaihtelumalli multiplicative seasonal model
kausaalikaavio causal diagram, causal graph	kerrannainen malli multiplicative model
kausaaliketju causal chain	kerroin coefficient
kausaalimalli causal model	kerroinero difference in betas
kausaalipäättely causal inference	kerrosetimaattori sandwich estimator
kausaalirakenteen löytäminen causal discovery	kertoma factorial
kausaalisuus causality	kertomakumulantti factorial cumulant
kausaaliverkko causal graph, causal diagram	kertomamomentit generoiva funktio factorial moment generating function
kausi-ARMA-malli seasonal autoregressive moving average process	kertomamomentti factorial mo-
kausijakso period of seasonality	
kausijuuri seasonal root	

ment	age treatment effect
kertomamomenttien emäfunktionio factorial moment generating function	keskimääräinen merkkijakson pituus average run length
kertymisnopeus accrual rate	keskimääräinen vikaantumisaika mean time to failure
kertymäfunktio distribution function, cumulative distribution function	keskimääräinen vikaantumisväli mean time between failures
kertymäfunktiomuunnos probability integral transformation	keskineliöjatkuvuus mean square continuity
kertymämalli cumulative model	keskineliöraja-arvo limit in the mean
kertymäristisuhde cumulative odds ratio	keskineliövirhe mean squared error
kertymäsumma cumulative sum	keski-p-arvo mid- p value
keskeinen raja-arvolause central limit theorem	keskipoikkeama mean absolute deviation
keskenään vaihtokelpoiset muuttujat exchangeable variables	keskipuolikkaan keskiarvo mid-mean
keskeytetty tarkastus curtailed inspection	keskisijaluku mid-rank
keskiarvo mean ₁ , average	keskistäminen centering
keskiarvosuppeneva ergodic	keskittymisindeksi index of concentration
keskiarvovektori mean vector	keskittymiskerroin coefficient of concentration
keskihajonta standard deviation	keskittyneisyys precision ₂
keski-itseisvirhe mean absolute error	keskittyneisyyskäyrä concentration curve
keskiluku measure of location	keskittyneisyysmatriisi precision matrix, concentration matrix
keskimääräinen kausaalivaikutus average causal effect	keskittyneisyysparametri precision parameter
keskimääräinen kausaalivaikutus sitoutuneilla complier average causal effect	keskivirhe standard error
keskimääräinen käsittelyvaikutus average treatment effect	keskusmomentti central moment
keskimääräinen käsittelyvaikutus altistumattomilla average treatment effect on the untreated	kesko mid-range
keskimääräinen käsittelyvaikutus altistuneilla average treatment effect on the treated	kesto duration
keskimääräinen käsittelyvaikutus sitoutuneilla complier average	kestoanalyysi survival analysis
	kestofunktio survival function
	kestokäyrä survival curve
	ketju chain
	ketjubinomimalli chain binomial model
	ketjuimputointi multivariate imputation by chained equations

(MICE)	koeasetelma experimental design
ketjuindeksi chain index	koeruutu experimental plot
ketjukaavio chain graph	koesarja series of experiments
khi-jakauma chi distribution	koesuunnittelu design of experiments
khi-kuvio chi-plot	koevirhe experimental error
khi-neliöetäisyys chi-squared distance	koeyksikkö experimental unit
khi-neliöjakauma chi-squared distribution	kofaktori cofactor
khi-neliösuure chi-squared statistic	kognitiivinen harha cognitive bias
khi-neliötesti chi-squared test	kohdefunktio objective function
khi-suure chi-statistic	kohdentaminen allocation ₂
kieltäytymisosuus refusal rate	kohdepopulaatio target population
kierto rotation	kohdesopeutus domain adaptation
kierto-otanta rotation sampling	kohdeväestö target population
kiertoasetelma rotatable design	kohdistettu suurimman uskottavuuden menetelmä targeted maximum likelihood estimation
kiertovuoroasetelma Round Robin design	koherenssi coherence
kiinalainen ravintolaprosessi Chinese restaurant process	kohina noise
kiinteävaikutusmalli fixed-effects model	kohinanpoisto denoising
kiinteäkantainen indeksi fixed base index	kohinasta puhdistettu denoised
kiinteä otos fixed sample	kohortti cohort
kiintiöinti allocation ₁	kohorttitutkimus cohort study
kiintiöotanta quota sampling	kohtisuora perpendicular, orthogonal
kiirejakso busy period	kohtisuora projektio orthogonal projection
kilpaileva riski competing risk	kokeellinen tutkimus experimental study
kilpaileva tapahtuma competing event	kokemusperäinen empirical
kiusaparametri nuisance parameter	kokoava hierarkkinen ryhmitely agglomerative hierarchical clustering
klaava tails	kokonaiskeskiarvo grand mean
kliininen koe clinical trial	kokonaiskorrelaatio total correlation
k-luokan estimaattori <i>k</i> -class estimator	kokonaislaskenta complete enumeration
k:n keskiarvon ryvästys <i>k</i> -means clustering	kokonaismäärä total
koalesenssi coalescence	kokonaisneliösumma total sum
koe experiment, trial	
koeala experimental plot	

of squares	konkordanssi concordance
kokonaisosa curtate, truncation	konkordanssikerroin coefficient of concordance
kokonaisten havaintojen analyysi complete case analysis	konkordantit parit concordant pairs
kokonaistodennäköisyys total probability	konservatiivinen testi conservative test
kokonaistutkimus census ₁ , complete enumeration	kontingenssikerroin contingency coefficient
kokonaisvaihteluetäisyys total variation distance	kontingenssitaulukko contingency table
kokoon verrannollinen todennäköisyys probability proportional to size	kontrafaktuaalinen counterfactual
kokovalikoitunut size-biased	kontrasti contrast
kollineaarisuus collinearity	kontrolloitu suora vaikutus controlled direct effect
kolmen sarjan lause three series theorem	konvekksi peite convex hull
kolmen sääntö rule of three	konveksisuus convexity
kolmiasteotanta three-stage sampling	konvergenssi convergence
kolmihiippuinen trimodal	konvoluutio convolution
kolmikkotesti triples test	koontuvuus collapsibility
kolmiojakauma triangular distribution	kooste ensemble
kolmiomatriisi triangular matrix	kooste-ennuste ensemble prediction
kolmiotaulukko triangular table	koostekeskiarvo ensemble average
kolmiotesti triangular test	koosteoppiminen ensemble learning
kolmivaiheinen pienin neliösumma three-stage least squares	kopula copula
kommunaliteetti commonality	korjaava tarkastus rectifying inspection
komplementtinen loglog-muunnos complementary loglog transformation	korkeamman kertaluokan asymptotiikka higher order asymptotics
komplementtinen virhefunktio complementary error function	korkotaso interest rate
komplementtitapahtuma complement event	korrelaatio correlation
koneoppiminen machine learning	korrelaatiokerroin correlation coefficient
konenäkö computer vision	korrelaatiomatriisi correlation matrix
koneäly machine intelligence	korrelaatiopuhdistettu decorrelated
konfiguraatio configuration	korrelaatiosta puhdistettu
konfiguraatiomalli configuration	
konkaavisuus concavity	

decorrelated	kumulanttiemäfunktio cumulant generating function
korreloimaton uncorrelated	kumulatiivinen frekvenssi cumulative frequency
korreloiva correlated	kumulatiivinen ilmaantuvuus cumulative incidence, incidence proportion
korrespondenssianalyysi correspondence analysis, dual scaling	kumulatiivinen ilmaantuvuusfunktio cumulative incidence function
korttelietäisyys city block distance, Manhattan distance	kumulatiivinen summa cumulative sum
kosinijakauma cosine distribution	kumulatiivinen virhe cumulative error
kotitalous household	kuolemanuhka instantaneous death-rate
kovariaatti covariate	kuolemaprosessi death process
kovarianssi covariance	kuolevuus mortality rate, death rate
kovarianssianalyysi analysis of covariance	kuolinuhka force of mortality, hazard of death
kovarianssimatriisi covariance matrix	kuperuus convexity
kovarianssistationaarinen covariance stationary, weakly stationary	kuplakuvio bubble plot
kovaydinmalli hard-core model	kurtoosi kurtosis
koveruus concavity	kustannus-hyötyanalyysi cost-benefit analysis
kreikkalais-latinalainen neliö Graeco-Latin square	kustannuskertymä cumulative cost
kriging-menetelmä kriging method	kustannus-tehokkuus cost-efficiency
kriittinen alue critical region	kustannus-vaikuttavuus cost-effectiveness
kriittinen arvo critical value	kutistusestimaattori shrinkage estimator
Kroneckerin tulo Kronecker product	kuuden sigman periaate six sigma principle
kruuna heads	kuukausikeskiarvo monthly average
k-tunnusluku k -statistic	kuumaimputointi hot deck imputation
kulkuaika passage time	kuumapaikkaus hot deck imputation
Kullbackin–Leiblerin poikkeama Kullback–Leibler divergence	kuutiohila cubic lattice
kulmakerroin slope	
kulmamuunnos angular transformation	
kumppanimuoto companion form	
kumulantit generoiva funktio cumulant generating function	
kumulantti cumulant	

kuutiosplini	cubic spline	kvasiuskottavuus	quasi likelihood
kuvaaja	graph ₃	kvintiili	quintile
kuva-analyysi	image analysis	kybernetiikka	cybernetics
kuvaileva analytiikka	descriptive analytics	kyllästetty malli	saturated model
kuvaileva tilastotiede	descriptive statistics	 kylmäimputointi	cold deck imputation
kuvaileva tunnusluku	descriptive statistic	kylmäpaikkaus	cold deck imputation
kuvaileva tutkimus	descriptive study	kymmenys	decile
kuvankäsittely	image processing	kynnysarvo	cut-off point
kuvio	graph ₁ , plot ₁	kynnysarvo	threshold value
kvadraattinen	quadratic	kynnysautoregressio	threshold autoregression
kvadraattinen erottelufunktio	quadratic discriminant function	kynnysmalli	threshold model
kvadraattinen estimaattori	quadratic estimator	kysely	inquiry
kvadraattinen vaihtelu	quadratic variation	kyselykierros	survey round
kvadranttiotanta	quadrant sampling	kyselylomake	questionnaire
kvadranttiriippuvuus	quadrant dependence	kyselytutkimus	questionnaire survey
kvalitatiivinen	qualitative	kytkentä	coupling
kvantiili	quantile, fractile	kytkentäanalyysi	linkage analysis
kvantiiliregressio	quantile regression	kytkentäepätasapaino	linkage disequilibrium
kvantitatiivinen	quantitative	kytkentämatriisi	incidence matrix ₂
kvanttitodennäköisyys	quantum probability	kytketyt lohkot	linked blocks
kvartiili	quartile	kytketyt otokset	linked samples
kvartiilipoikkeama	quartile deviation	käsittely	treatment
kvartiiliväli	interquartile range	käsittelyhaara	treatment arm
kvartimax-kierto	quartimax rotation	käsittelykontrasti	treatment contrast
kvartimax-rotaatio	quartimax rotation	käsittelyryhmä	treatment group
kvasikoe	quasi experiment	käsittelyvirhe	processing error
kvasiriippumattomuus	quasi independence	käyntihaastattelu	face-to-face interview
kvasisymmetria	quasi symmetry	käypä	admissible
		käypä estimaattori	admissible estimator
		käypä päätösfunktio	admissible decision function

käypä testi admissible test
käyräanalainen pinta-ala area
 under curve
käyrän sovitus curve fitting
käyräviivainen curvilinear
käyttökielen prosessointi natu-
 ral language processing
käännekohta turning point
käänteinen Bernoulli-poiminta
 inverse Bernoulli sampling
käänteinen betajakauma in-
 verted beta distribution
käänteinen Dirichlet-jakauma
 inverted Dirichlet distribution
käänteinen gammajakauma
 inverse gamma distribution
käänteinen Gaussin jakauma
 inverse Gaussian distribution
käänteinen khi-neliöjakauma
 inverse chi-square distribution
käänteinen normaali-jakauma
 inverse Gaussian distribution
käänteislukumuunnos reciprocal
 transformation
käänteismatriisi inverse matrix
käänteismuunnos inversion
käänteismuunnosmenetelmä
 inverse transformation method
käänteisongelma inverse prob-
 lem
käänteisotanta inverse sampling
käänteistodennäköisyys inverse
 probability
**käänteistodennäköisyyspaino-
 tus** inverse probability weighting
kääntyvä invertible
kääntyvä prosessi invertible pro-
 cess
käärretty jakauma wrapped dis-
 tribution

L

L^1 -etäisyys L^1 metric
 L^1 -normi L^1 norm
 L^2 -etäisyys L^2 distance
 L^2 -normi L^2 norm
laadullinen qualitative
laadullinen vertaileva analyysi
 qualitative comparative analysis
laadunvalvonta quality control
laadunvalvontakortti quality
 control chart
laadunvarmistus quality assur-
 ance
**laajassa mielessä stationaari-
 nen** wide sense stationary
laakea priori vague prior
laatikkokuvio box plot
laatuelinvuosi quality adjusted
 life years
Lagrangen kerrointesti La-
 grange multiplier test
lajien ordinaatio ordination of
 species
lajien runsaus species abundance
Laplace-jakauma Laplace distri-
 bution
Laplace-muunnos Laplace trans-
 form
lapsiluku parity₁
laskennallinen tilastotiede com-
 putational statistics
laskennan alipeitto underenu-
 meration, undercount
laskennan ylipeitto overenumer-
 ation, overcount
laskenta enumeration
laskentaintensiivinen
menetelmä computationally
 intensive method
laskentataulukko spreadsheet
**laskentojen välinen elossaolo-
 suhde** census survival rate

laskostuminen	aliasing ₂	likimääräinen	approximate
laskurimitta	counting measure	likimääräinen Bayes-laskenta	approximate Bayesian computa-
laskuri-prosessi	counting process		tion
Laspeyresin indeksi	Laspeyres' index	likimääräinen uusioluotta-	
LASSO	least absolute shrinkage and selection operator	musväli	approximate bootstrap confidence interval
latinalainen hyperkuutio-otanta	Latin hypercube sampling	likviditeetti	liquidity
latinalainen kuutio	Latin cube	lineaarinen	linear
latinalainen neliö	Latin square	lineaarinen ennustin	linear predictor
latinalainen suorakulmio	Latin rectangle	lineaarinen erotteluanalyysi	linear discriminant analysis
lattiavaikutus	floor effect	lineaarinen estimaattori	linear estimator
laukut	bagging, bootstrap aggregating	lineaarinen malli	linear model
laukutuvirhe	out-of-bag error	lineaarinen ohjelmointi	linear optimization, linear programming
laumasuoja	herd immunity	lineaarinen optimointi	linear optimization
lause	theorem	lineaarinen prosessi	linear process
lehtikuvio	stem-and-leaf diagram	lineaarinen rajoite	linear constraint
leikattu keskiarvo	trimmed mean	lineaarinen regressio	linear regression
leikkaus	intersection	lineaarinen riippuvuus	linear dependence
lepojakso	idle period	lineaarinen suodin	linear filter
Lexis-kuvio	Lexis diagram	lineaarinen todennäköisyysmalli	linear probability model
liikakaltaistus	overmatching	lineaarinen trendi	linear trend
liikaparametroitu	overparameterized	linearisoitu	linearized
liikasovitus	overfitting	linja-leikkausotanta	line-intersect sampling
liikavakiointi	overadjustment	linjaotanta	line sampling
liikkuja-pysyjämalli	mover-stayer model	linkkifunktio	link function
liitospisteregressio	joinpoint regression	linkkuveitsiestimaattori	jack-knife estimator
liittojakauma	conjugate distribution	linkkuveitsimenetelmä	jack-knife method
liittomatriisi	adjugate matrix	lisäinformaatio	auxiliary information
liittopriorijakauma	conjugate prior distribution		
Likertin asteikko	Likert scale		
likiarvo	approximation ₂		
likiarvoinen	approximate		
likiarvoistus	approximation ₁		

lisämuuttuja	auxiliary variable	lohko	block
lisäriski	excess risk	lohkodiagonaalimatriisi	block diagonal matrix
lisäselittäjäkuvio	added variable plot	lohkojen välinen	interblock
lisäuhkamalli	excess hazards model	lohkokaavio	block diagram
lisääntymisjakauma	offspring distribution	lohkolävistäjämatriisi	block diagonal matrix
liukuneliökuvio	sliding square plot	lohkonsisäinen	intrablock
liukuva keskiarvo	moving average	lohkosatunnaistaminen	block randomization
liukuvan keskiarvon malli	moving average model	lokeromalli	compartment model
liukuva tunnusluku	scan statistic	Lordin paradoksi	Lord's paradox
liukuva vaihteluväli	moving range	Lorenzin käyrä	Lorenz curve
logaritmimuunnos	logarithmic transformation	luetteleminen	enumeration
logaritminen asteikko	logarithmic scale	lukumäärä	count, frequency ₁ , number
logaritmisen sarjan jakauma	logarithmic series distribution	lukumääräaikaasarja	count time series
logistinen jakauma	logistic distribution	lukumääräaineisto	count data, frequency data
logistinen käyrä	logistic curve	lukumääräjakauma	frequency distribution
logistinen regressio	logistic regression	lukumäärätaulukko	frequency table
logit-malli	logit model	lume	placebo
logit-muunnos	logit transformation	lumevaikutus	placebo effect
log-lineaarinen malli	loglinear model	lumihiutalekuvio	snowflake diagram
loglog-muunnos	loglog-transformation	lumipallo-otanta	snowball sampling
log-normaalijakauma	lognormal distribution	lunastusarvo	exercise value
logrank-testi	logrank test	lunastushinta	exercise price
log-uskottavuusfunktio	log-likelihood function	lunastuspäivä	exercise date
log-vastasuhte	log-odds	luokitin	classifier
log-vetokerroin	log-odds	luokittelija	classifier
		luokittelu	classification
		luokitteluasteikko	nominal scale
		luokitteluasteikon muuttuja	
			categorical variable
		luokitteluattribuutti	classification attribute
		luokittelu- ja regressiopuu	classification and regression tree

sification and regression tree	natural direct effect
luokittelumääre classification attribute	luonnollisen kielen prosessointi natural language processing
luokittelun yhdenmukaisuus consistency of classification	luonnon koe natural experiment
luokittelusääntö classification rule	luotaus survey
luokittelutaulukko classification table	luotausinstrumentti survey instrument
luokittelutunnusluku classification statistic	luotausmetodologia survey methodology
luokitteluvirhe classification error	luotausotanta survey sampling
luokitus classification	luotautapa survey mode
luokitushila grouping lattice	luotettavuus reliability ₁
luokka-asteikko nominal scale	luotettavuusfunktio reliability function
luokkafrekvenssi class frequency	luotettavuusteoria reliability theory
luokkajakauma categorical distribution	luottamusalue confidence region
luokkakeskus class midpoint	luottamusjoukko confidence set
luokkamuuttuja categorical variable	luottamustaso confidence level
luokkatunnus class symbol	luottamusvyöhyke confidence band, confidence envelope
luokkaväli bin, class interval	luottamusväli confidence interval, compatibility interval
luokkien sisäinen	luotto credit
samankaltaisuus intra-class similarity	luottoalue credible region
luokkien välinen	luottoteoria credibility theory
samankaltaisuus inter-class similarity	luottoväli credible interval
luonnollinen epäsuora kokonaisvaikutus total natural indirect effect	luovuttajapaikkaus donor imputation
luonnollinen epäsuora vaikutus natural indirect effect	lyhimmän koodin pituus minimum description length
luonnollinen suora kokonaisvaikutus total natural direct effect	lyhin luottamusväli most accurate confidence interval
luonnollinen liittopriori natural conjugate prior	lyhyeksi myynti short sale
luonnollinen parametri natural parameter, canonical parameter	lyhythäntäinen jakauma short-tailed distribution
luonnollinen suora vaikutus	lyhytmuistinen prosessi short-memory process
	läheisyysmatriisi proximity matrix
	lähes vajaa-asteinen matriisi ill-conditioned matrix
	lähinaapurimenetelmät nearest-

neighbor methods
lähinaapuriryhmittely nearest-neighbor clustering
lähtöarvot baseline characteristics
lähtötaso baseline
lämmittelyjakso warm-up period
lämpökartta heat map
läpäisy penetrance
lävistäjämatriisi diagonal matrix
lääketieteellinen tilastotiede medical statistics
löytymisharha detection bias

M

maahanmuutto immigration
maastamuutto emigration
Mahalanobisin etäisyys Mahalanobis distance
malli model
malliavusteinen model-assisted
mallidiagnostiikka model diagnostics
mallikeskiarvoistus model averaging
mallimatriisi model matrix
mallinnus modeling
mallinrakentaminen model building
mallin sopivuus model fit
malliperusteinen päättely model-based inference
mallitus modeling
mallivirhe model error
manhattanetäisyys city block distance, Manhattan distance
Mannin–Whitneyn testi Mann–Whitney test
Mantelin–Haenszelin testi Mantel–Haenszel test

marginaalihomogeenisuus marginal homogeneity
marginaalijakauma marginal distribution
marginaalimalli marginal model
marginaalirakennemalli marginal structural model
marginaaliuskottavuus marginal likelihood
marimekkokuvio Marimekko plot, mosaic plot
markov¹ Markov
markovinen Markovian
Markovin ketju Markov chain
Markovin Monte Carlo -menetelmä Markov chain Monte Carlo method
Markovin satunnaiskenttä Markov random field
Markovin vaihtomalli Markov switching model
Markov-ominaisuus Markov property
Markov-prosessi Markov process
martingaali martingale
martingaalijäännös martingale residual
martingaalimitta martingale measure
martingaaliresiduaali martingale residual
massadata big data
matokuvio worm plot
matriisialgebra matrix algebra
matriisihajotelma matrix decomposition
matriisilaskenta matrix algebra
matriisinormaalijakauma matrix normal distribution
matriisin tekijöihinjako matrix factorization

¹Ks. alaluku 2.5.

matriisiotanta	matrix sampling	algorithm
matriisivasteregressio	matrix variate regression	metsikkökuvio forest plot
mediaani	median	mielipidekysely opinion survey, opinion poll
mediaanielinaika	median survival time	minimaalinen tyhjentävä tunnusluku minimal sufficient statistic
mediaaniharhattomuus	median unbiasedness	minimaalisesti kytketty asetelma minimally connected design
mediaaniregressiokäyrä	median regression curve	minimax-kriteeri minimax criterion
mediaanitesti	median test	mitallinen measurable
mediaatio	mediation	mitta measure
mediaatioanalyysi	mediation analysis	mitta-asteikko measurement scale
melkein kaikkialla	almost everywhere	mittakaava scale ₂
melkein varma suppeneminen	almost sure convergence	mittateoria measure theory
melkein varmasti	almost surely	mittaussarja serial measurements
menettelyharha	procedural bias	mittausvirhe measurement error
menetys	regret	m.k. a.e.
menneistäminen	backcasting	momentit generoiva funktio moment generating function
merkintä-takaisinpyyntiotanta	capture-recapture sampling	momentti moment
merkitsevyys	significance	momenttiemäfunktio moment generating function
merkitsevyystaso	level of significance, significance level	momenttiestimaattori moment estimator
merkitsevyystesti	test of significance, significance test	momenttimatriisi moment matrix
merkkijakso	run ₁	momenttimenetelmä method of moments
merkkijaksotesti	runs test	momenttisuhde moment ratio
merkkijärjestystesti	signed rank test	monen selittäjän regressio multiple regression
merkkipisteprosessi	marked point process	monen vasteen regressio multivariate regression
merkkitesti	sign test	monen vasteen varianssianalyysi multivariate analysis of variance
M-estimaattori	M-estimator	moniasteinen estimointi multi-stage estimation
meta-analyysi	meta-analysis	moniasteinen otanta multistage
metametri	metameter	
metaregressio	meta regression	
metodologia	methodology	
Metropolisin–Hastingsin algoritmi	Metropolis–Hastings	

sampling

moniasteotanta multistage sampling

monihuippuinen jakauma multimodal distribution

moni-imputointi multiple imputation

monijaksomalli multiepisode model

monikehikkoluotaus multiple-frame survey

monikeskuskoe multicenter trial

monikeskustutkimus multicenter study

monikätkäinen rosvo multiarmed bandit

monilineaarinen prosessi multilinear process

moniluokkainen polytomous

monimuotoisuusindeksi index of diversity

monimuuttuja-analyysi multivariate analysis

monimuuttujainen multivariate

monimuuttujainen laskuriprosessi multivariate counting process

moninormaali-jakauma multinormal distribution

moniositus multiple stratification

moniosumamalli multihit model

monipaikkaus multiple imputation

monipoistumamalli multiple decrement model

monipäätösmenetelmät multiple criteria decision making

monisilotusmenetelmä multiple smoothing method

monisto manifold

monisto-oppiminen manifold learning

monitasoitusmenetelmä multiple smoothing method

monitasomalli multilevel model, hierarchical model

monitavoiteoptimointi multiple criteria optimization

monitavoitteinen päätösten-teko multiple criteria decision making

monitehtäväoppiminen multitask learning

monitekijäkoe multifactorial design

monitestaus multiple testing

monitilainen elinajantaulukko multistate life table

monitilamalli multistate model

moniulotteinen multidimensional

moniulotteinen aikasarja multiple time series

moniulotteinen jakauma multivariate distribution

moniulotteinen skaalaus multidimensional scaling

monivaiheinen otanta multiphase sampling

monivaiheotanta multiphase sampling

monivaihepäätöskaavio multivalued decision diagram

monivaiheiselontaa multiphase screening

monivastausanalyysi multiple response analysis

moniverkko multigraph

monivertailu multiple comparison

moniyhtälömalli multiequational model

monotoninen monotone

monotoninen uskottavuusosamäärä monotone likelihood ratio

Monte Carlo -menetelmä

Monte Carlo method	muokattu otospaino adjusted sample weight
moodi mode	muokattu profiiliuskottavuus modified profile likelihood
morfometria morphometrics	muokattu R^2 adjusted R^2
mosaiikkikuvio mosaic plot	muokkaus adjustment
muistamattomuusominaisuus memorylessness property	muotoanalyysi morphometrics
muistiharha recall bias	muotoparametri shape parameter
muistittomuus lack of memory property	murtopolynomi fractional polynomial
mukaanottokriteeri inclusion criterion	murtumispiste breakdown point
mukautettu prosessi adapted process	musta laatikko black box
mukautuva asetelma adaptive design	muunnos transformation
mukautuva päättely adaptive inference	muunnosarvo metameter
mukautuva regressio adaptive regression	muutosarvo change score
mukautuva tehostus adaptive boosting	muutospiste change point
mukavuusotos convenience sample	muutostahti rate of change
multikollinearisuus collinearity, multicollinearity	muutosvauhti rate of change
multilineaarinen prosessi multilinear process	muuttuja variable, variate
multinomijakauma multinomial distribution	muuttujakato item non-response
multinomikerroin multinomial coefficient	muuttujalavennus data augmentation
multinomilogistinen regressio multinomial logistic regression	muuttujavalinta variable selection
multinomilogit-malli multinomial logit model	m.v. a.s., almost surely
multinormaalijakauma multinormal distribution	myyntioptio put option
multiplikatiivinen kausivaihtelumalli multiplicative seasonal model	myöntyväisyysharha acquiescence bias
multiplikatiivinen malli multiplicative model	myötävirtaneuroverkko feed-forward neural network
multiverkko multigraph	määrittelyharha specification bias
	määrittely assay
	määrittyskoe assay
	määrällinen quantitative

N

naapurustouputus neighbor embedding
naiivi ennuste naive forecast
napakoordinaatti polar coordinate

nate	
napasuure	pivot, pivotal quantity
naulakkolause	portemanteau theorem
naulakkotesti	portemanteau test
negatiivinen binomijakauma	negative binomial distribution
negatiivinen ennustearvo	negative predictive value
negatiivinen harha	downward bias
negatiivinen hypergeometrisen jakauma	negative hypergeometric distribution
nelierotus	tetrad difference
nelikenttä	fourfold table, two-by-two table
nelikenttäkuvio	fourfold plot
neliöerottelufunktio	quadratic discriminant function
neliöestimaattori	quadratic estimator
neliöhila	square lattice
neliöjuurimuunnos	square root transformation
neliöjuurisääntö	square root rule
neliöllinen	quadratic
neliöllisesti summautuva	square summable
neliömatrissi	square matrix
neliömuoto	quadratic form
neliötappiofunktio	quadratic loss function
neliövaihtelu	quadratic variation
neljännesvuosiaineisto	quarterly data
nettikysely	internet survey
nettomuutto	net migration
nettouusiutumisluku	net reproduction rate
neurolaskenta	neural computing
neuroverkko	neural network
Newtonin–Raphsonin menetelmä	Newton–Raphson method
Neymanin kiintiöinti	Neyman allocation
Neymanin–Pearsonin apulause	Neyman–Pearson lemma
nielevä alue	absorbing region
nielevä reuna	absorbing barrier
nielevä tila	absorbing state
nimellinen koko	nominal size
nimellinen luottamustaso	nominal confidence level
nimellinen merkitsevyystaso	nominal significance level
nimellinen peittotodennäköisyys	nominal coverage probability
nimittäjä	denominator
nimeämissääntö	labeling rule
nimiöimätön data	unlabeled data
nimiöity data	labeled data
nolla-avaruus	null space
nolladevianssi	null deviance
nollahypoteesi	null hypothesis
nollainflatoitu	zero-inflated
nollajakauma	null distribution
nollajoukko	null set
nollakatkaistu jakauma	zero-truncated distribution
nollamalli	null model
nollamatriisi	zero matrix
nollapaisunta	zero-inflation
nollapaisunut jakauma	zero-inflated distribution
nollasummapeli	zero-sum game
nollatypistetty jakauma	zero-truncated distribution
nomogrammi	nomogram
nopea Fourier-muunnos	fast Fourier transform
nopea harhakorjattu luottamusväli	bias-corrected and accelerated

erated interval	
nopeutettu elinaikatestaus	accelerated life testing
nopeutetun vikaantumisajan malli	accelerated failure time model
normaaliapproksimaatio	normal approximation
normaaliarvo	normal score
normaalijakauma	normal distribution, Gaussian distribution
normaalijakaumakuvi	normal probability plot
normaalisuus	normality
normaaliyhtälöt	normal equations
normalisoiva muunnos	normalizing transformation
normitettu muuttuja	standardized variable
normitettu normaalijakauma	standard normal distribution
numeerinen optimointi	numerical optimization
numeraari	numeraire
numeroituva	countable
numeroituvasti ääretön	denumerable
numerosuosio	digit preference
nykyistäminen	nowcasting
näennäisesti riippumattomat regressiot	seemingly unrelated regressions
näennäiskorrelaatio	spurious correlation
näennäissatunnaisluku	pseudo-random number
näennäistarkkuus	spurious precision
näyte	sample ₂
näytön alaraja	evidence lower bound
näytön paino	weight of evidence
näytön yläraja	evidence upper bound
O	
oblimax-rotatio	oblimax rotation
oblimin-rotatio	oblimin rotation
odotettu lukumäärä	expected frequency, expected number
odotetun hyödyn teoria	expected utility theory
odotusaika	waiting time
odotusarvo	expectation, expected value
odotusarvoon palautuva	mean reversion
odotustappio	expected loss
odotustappioarvo	expected shortfall
offset-termi	offset term
ohjaamaton luokittelu	unsupervised classification
ohjaamaton oppiminen	unsupervised learning
ohjaava analytiikka	prescriptive analytics
ohjattu luokittelu	supervised classification
ohjattu oppiminen	supervised learning
ohuthäntäinen jakauma	light-tailed distribution
ohutlevysplini	thin plate spline
oikealle jatkuva vasemmalta raja-arvot	continue à droite, limite à gauche, càdlàg
oikealle vinous	positive skewness
oikeustilastotiede	jurimetrics
oktiili	octile
olennaiset piirteet	stylized facts
oletus	assumption

omatodennäköisyys	personal probability	osamäärä	ratio, quotient
omatoimiluotaus	self-administered survey	osaotos	subsample
omien lasten menetelmä	own children method	osaparametrinen	semi-parametric
ominaisarvo	eigenvalue	osaruutuasetelma	split-plot design
ominaisarvohajotelma	spectral decomposition	osataulukko	partial table
ominaisarvoyhtälö	characteristic equation	osayksikköasetelma	split-unit design
ominaisavaruus	eigenspace	osioanalyysi	item analysis
ominaisuus	attribute, characteristic	osiovasteteoria	item-response theory
ominaisvektori	eigenvector	osite	stratum ²
operaatiotutkimus	operations research	ositettu malli	stratified model
opetusaineisto	training data	ositettu matriisi	partitioned matrix
optimaalinen kiintiöinti	optimal allocation	ositettu otanta	stratified sampling
optimaalisuus	optimality	ositettu pylväskuvio	multiple bar chart
optimiasetelma	optimal design	osittainen pienin neliösumma	partial least squares
optimikiintiöinti	optimal allocation	osittainen toisto	fractional replication
optimiprojektio regressio	projection pursuit regression	osittaisassosiaatio	partial association
optimiskaalaus	optimal scaling	osittaisautokorrelaatio	partial autocorrelation
optimointi	optimization	osittaisjäännöskuvio	partial residual plot
optimointimenetelmä	optimization method	osittaiskorrelaatio	partial correlation
ordinaaliasteikko	ordinal scale	osittaisregressiokerroin	partial regression coefficient
ordinaaliregressio	ordinal regression	osittaisuskottavuus	partial likelihood
ortogonaalinen	orthogonal, perpendicular	osittaisyhteys	partial association
ortogonaalipolynomi	orthogonal polynomial	ositaminen	decomposition ₁
ortogonaaliprojektio	orthogonal projection	osituva malli	decomposable model
ortonormaali	orthonormal	ositus₁	partition
osajoukko	subset		

²Monikkomuoto on *strata*.

ositus₂ stratification
osoitin indicator variable
osoitinmatriisi incidence matrix₁
osoittaja numerator
osto-optio call option
osuma-aika hitting time
osumatodennäköisyys hitting probability
osuus proportion
otannasta johtumaton virhe non-sampling error
otanta sampling
otanta-asetelma sample design
otantajakauma sampling distribution
otantakehikko sampling frame
otantaosuus sampling fraction
otanta palauttaen sampling with replacement
otanta palauttamatta sampling without replacement
otantasuunnitelma sampling plan, sample plan
otantatarkastus sampling inspection
otanta vaihtelevin todennäköisyyksin unequal probability sampling
otantavirhe sampling error
otantayksikkö sampling unit
otos sample₁
otosasetelma sample design
otosavaruus sample space
otosjakauma sample distribution
otuskertymäfunktio sample distribution function
otoskeskiarvo sample mean
otoskoko sample size
otoskovarianssimatriisi sample covariance matrix
otosluotaus sample survey
otosmomentti sample moment
otospaino sample weight

otospiste sample point
otospolku sample path
otossuunnitelma sample plan, sampling plan
otossuure sample statistic
otostunnusluku sample statistic
otostutkimus sample survey
otosvarianssi sample variance
otosyksikkö sample unit
oudokki outlier

P

paikallinen asymptoottinen tehokkuus local asymptotic efficiency
paikallinen regressio local regression
paikallisuuskottavuus local likelihood
paikkadata spatial data
paikkatietojärjestelmä geographical information system
paikkatilastotiede spatial statistics
paikkaus imputation
paikkausmalli imputation model
paino weight
painoindeksi body mass index
painotettu keskiarvo weighted mean, weighted average
painotettu pienimmän neliösumman menetelmä weighted least squares method
painotettu uskottavuus weighted likelihood
paksuhäntäinen jakauma heavy-tailed distribution
palautusotanta sampling with replacement
palautuva recurrent₁
palkkikuvio bar chart
pallokoordinaatti spherical coordinate

dinate	timaattori best linear unbiased estimator
pallopintajakauma spherical distribution	paremmuuskoe superiority trial
pallopintamuuttuja spherical variable	Pareto-jakauma Pareto distribution
pallopintatesti sphericity test	parhaan olosuhteen elinajanodote best practice life expectancy
paloittainen regressio piecewise regression	pariotokset paired samples
paloittain lineaarinen piecewise linear	pariteetti parity ₂
paloittain vakio piecewise constant	pariteettisuhde parity-progression ratio
paluu aika recurrence time ₁	parittainen riippumattomuus pairwise independence
paluujakso return period	parittainen <i>t</i>-testi paired <i>t</i> -test
paluukvantiili return level	parivertailu paired comparison
paneeliaalto panel wave	<i>p</i>-arvo <i>p</i> -value
paneeliaineisto panel data	<i>p</i>-arvon kalastelu <i>p</i> -hacking
paneelidata panel data	Pearson-jäännös Pearson residual
paneelitutkimus panel study	Pearsonin khi-neliösuure Pearson's chi-squared statistic
panos-tuotosmalli input-output model	Pearsonin khi-neliötesti Pearson's chi-squared test
panos-tuotostmatriisi input-output matrix	pehmo laskenta soft computing
parametri parameter	peittoharha noncoverage bias
parametriavaruus parameter space	peittotodennäköisyys coverage probability
parametrinen jakauma parametric distribution	peittovaikutus masking effect
parametrinen malli parametric model	peliteoria game theory
parametrinen päättely parametric inference	pelkistetty muoto reduced form
parametrinon nonparametric	pelkistymätön ketju irreducible chain
parametrivetoinen malli parameter-driven model	pelurin vararikko gambler's ruin
parametrointi parametrization	pelurin virhepäätelmä gambler's fallacy
paras asympotoottisesti normaali best asymptotically normal estimator	penetranssi penetrance
paras lineaarinen harhaton ennustin best linear unbiased prediction	perhekohtainen virhetodennäköisyys family-wise error rate
paras lineaarinen harhaton es-	perheenpääkohtainen menetelmä headship rate method
	periodogrammi periodogram

periytyvyys	heritability		estimation
perkolaatioteoria	percolation theory	pienimmän itseiskutistamisen ja valinnan operaattori	least absolute shrinkage and selection operator
permutaatio	permutation	pienimmän itseispoikkeaman menetelmä	least absolute deviation method
permutaatiotesti	permutation test	pienimmän neliösumman estimaattori	least squares estimator
perspektiivikuvio	perspective plot	pienimmän neliösumman menetelmä	method of least squares
perturbaatiomenetelmä	perturbation method	pienimmän varianssin harhaton estimaattori	minimum variance unbiased estimator
perusjakso	base period	pienin khi-neliö	minimum chi-square
perusjoukko	population ₁	pienten lukujen laki	law of small numbers
perusotos	master sample	Pietarin paradoksi	St. Petersburg paradox
perusotospaino	basic sample weight	piikki ja laatta	spike and slab
perustartuttavuusluku	basic reproduction number	piikki ja laatta -jakauma	spike and slab distribution
perustaso	baseline	piilokki	inlier
perustasoharha	base rate bias	piilomarkovmalli	hidden Markov model
perusuhka	baseline hazard	piiloluokka-analyysi	latent class analysis
perusuusiutumisluku	basic reproduction number	piilomuuttuja	latent variable
peräkkäinen todennäköisyys-suhdetesti	sequential probability ratio test	piilopiirreanalyysi	latent trait analysis
peräkkäisallokointi	sequential allocation	piilorakenne	latent structure
peräkkäisanalyysi	sequential analysis	piilosekoittaja	unobserved confounder
peräkkäisasetelma	sequential design	piirakkakuvio	pie diagram, pie chart
peräkkäiskoe	sequential experiment	piirre	attribute, feature, trait
peräkkäismalli	sequential model	piirreavaruus	feature space
peräkkäisotanta	sequential sampling	piirrevalinta	feature selection
peräkkäispoiminta	sequential sampling	piktogrammi	pictogram
peräkkäistesti	sequential test	pilotti	pilot study
pesitetty tapaus-verrokkitutkimus	nested case-control study		
Phillipsin käyrä	Phillips curve		
pienalue-estimointi	small area		

pistebiseriaallinen korrelaatio	poikkeaman itseisarvo
point biserial correlation	absolute deviation
piste-estimaatti	poimija
point estimate	sampler
piste-estimaattori	poimintaväli
point estimator	sampling interval
piste-estimointi	poismuutto
point estimation	out-migration
pistehypoteesi	Poisson-jakauma
simple hypothesis	Poisson distribution
pisteittäinen luottamusvyöhyke	Poisson-otanta
pointwise confidence band	Poisson sampling
pistekohina	Poisson-prosessi
shot-noise	Poisson process
pistekuvio₁	poissulkukriteeri
dot plot	exclusion criterion
pistekuvio₂	poisto
point pattern	elimination
pistemäärä	poistomenetelmä
score ₂	removal method
pisteotanta	poistovalinta
point sampling	backward elimination
pisteprosessi	polkuanalyysi
point process	path analysis
pistetiheys	polkukerroin
point density	path coefficient
pistetodennäköisyys	polkuriippuvainen
point probability	path dependent
pistetulo	Pólyan uurnamalli
dot product, inner product	Pólya urn model
pisteytys	polynominen trendi
scoring	polynomial trend
pistoke-estimaattori	polynomiregressio
plug-in estimator	polynomial regression
pitkittäisaineisto	polyspektri
longitudinal data	polyspectrum
pitkäaikaismuutos	populaatio
trend	population
pitkähäntäinen jakauma	populaatiogenetiikka
long-tailed distribution	population genetics
pitkämuistinen prosessi	populaatiokeskiarvoistettu malli
long-memory process	population-averaged model
pitkän kantaman riippuvuus	populaation kasvumalli
long-range dependence	population growth model
Pitman-tehokkuus	populaatiopyramidi
Pitman efficiency	population pyramid
pito-otanta	porrasasetelma
sampling without replacement	staircase design
pituusharhainen otanta	porraskiila-asetelma
length-biased sampling	stepped wedge design
PNS-estimaattori	porrasteinen aloitus
least squares estimator	staggered entry
poikittaistutkimus	positiividefiniitti
cross-sectional study	positive definite

positiivinen ennustearvo	positive predictive value	projektion haku	projection pursuit
positiivinen harha	upward bias	Prokrustes-analyysi	Procrustes analysis
positiivinen vinous	positive skewness	propensiteetti	propensity
positiivisesti definiitti	positive definite	propensiteettikaltaistus	propensity matching
positiivisesti semidefiniitti	positive semidefinite, nonnegative definite	propensiteettiluku	propensity score
posteriorijakauma	posterior distribution	propensiteettipainotus	propensity matching
posterioritodennäköisyys	posterior probability	prosenttiili	percentile
posterioriväli	posterior interval	prosenttijakauma	percentage distribution
postikysely	postal survey	prosenttipiste	percentile
potenssilaki	power law	prosenttiyksikkö	percentage point
potenssimuunnos	power transformation	prosessien kerrostaminen	superposition of processes
potentiaali	potential	prosessien kytkentä	coupling of processes
potentiaalinen vaste	potential outcome	prosessien summaus	superposition of processes
prediktiivinen päättely	predictive inference	prosessin generaattori	generator of process
priorijakauma	prior distribution	prosessin kehitin	generator of process
priorin koostaminen	elicitation of prior	prospektiivinen tutkimus	prospective study
prioritodennäköisyys	prior probability	prospektiteoria	prospect theory
probit-muunnos	probit transformation	protokolla	protocol
profiilianalyysi	profile analysis	protokollan mukainen analyysi	per-protocol analysis
profiilikuvio	profile plot	protokollarike	protocol violation
profiili-log-uskottavuus	profile log-likelihood	pseudosatunnaisluku	pseudo-random number
profiilitukifunktio	profile support function, profile log-likelihood	pseudouskottavuus	pseudo-likelihood
profiiliuskottavuus	profile likelihood	psykometriikka	psychometrics
profilointi	profiling	puhdas luonnollinen epäsuora vaikutus	pure natural indirect effect
projektio	projection	puhdas luonnollinen suora	
projektiokuvio	projection plot		

vaikutus	pure natural direct effect		principal components regression
puhdistumisjakso	wash-out period	pääkoordinaatti	principal coordinate
puheentunnistus	speech recognition	pääositus	principal stratification
puhelinnumeron satunnaisvalinta	random digit dialing	päätetapahtuma	endpoint
puoleen tehoava annos	median effective dose	päättelin	perceptron
puoleen tehoava taso	median effective level	päättele	inference
puolet tappava annos	median lethal dose	päätösavaruus	decision space
puolet tappava taso	median lethal level	päätösfunktio	decision function
puolimartingaali	semi-martingale	päätöspuu	decision tree
puolinormaalijakauma	half-normal distribution	päätössääntö	decision function
puolitusmenetelmä	split-half method	päätösteoria	decision theory
puukaavio	tree diagram	päävaikutus	main effect
puppukorrelaatio	nonsense correlation		
puumalli	tree model	R	
puuttuva data	missing data	R²	coefficient of determination
puuttuvuuden sekoitemalli	pattern mixture model	raaka-arvo	raw score
puuttuvuusrakenne	missing data pattern	raakamomentti	raw moment
pylväskuvio	bar chart	radiaalifunktio	radial function
pysyvä elinkorko	perpetuity	Radonin–Nikodymin derivaatta	Radon–Nikodym derivative
pysäytyshetki	stopping time, optional time	rahasto	fund
pysäytyssääntö	stopping rule	rajahinta	strike price
pyöristys	rounding	rajajakauma	asymptotic distribution, limiting distribution
pätevyys	validity	rajapiste	cut-off point
pääakselihajotelma	principal axes decomposition	rajatta jakautuva	infinitely divisible
pääkomponenttianalyysi	principal components analysis	rajatuloestimaattori	product limit estimator
pääkomponenttiregressio	prin-	rajoitettu satunnaistus	restricted randomization
		rajoitettu suurimman uskottavuuden menetelmä	restricted maximum likelihood
		rajoitetun informaation estimaattori	limited information estimator
		rajoitetusti heilahteleva funktio	bounded variation function
		rakenneaikasarjamalli	struc-

tural time series model	regressiomalli regression model
rakennemuutos structural change	regressiomallinnus regression modeling
rakenneosa-aineiston analyysi compositional data analysis	regressiomallitus regression modeling
rakenneyhtälömalli structural equation model	regressioneliösumma regression sum of squares
rakenteellinen kausaalimalli structural causal model	regressio origon kautta regression through the origin
rakenteellinen nolla structural zero	regressiopinta regression surface
rakenteellinen vektoriautoregressio structural vector autoregression	regressiosplini regression spline
rankit-muunnos rankit transformation	regressiosuora regression line
Raon–Blackwellin tehostus Rao–Blackwellization	Rényin entropia Rényi entropy
raportointivirhe reporting error	Rényin poikkeama Rényi divergence
rationaaliset odotukset rational expectations	reilu peli fair game
ratkeava lohkoasetelma resolvable block design	reittiotanta route sampling
realisaatio realization	rekursiivinen recursive
redusoitu muoto reduced form	relaatiotietokanta relational database
regressio regression	reliabiliteetti reliability ₂
regressioanalyysi regression analysis	reliabiliteettikerroin reliability coefficient
regressiodiagnostiikka regression diagnostics	replikaatio replication
regressioepäjatkuvuusasetelma regression discontinuity design	replikaatti replicate
regressioestimaattori regression estimator	residuaali residual
regressiofunktio regression function	resistentti estimaattori resistant estimator
regressiokalibrointi regression calibration	retrospektiivinen tutkimus retrospective study
regressiokerroin regression coefficient	reunafrekvenssi marginal frequency
regressio keskiarvoa kohti regression to the mean	reunahomogeenisuus marginal homogeneity
regressiökäyrä regression curve	reunajakauma marginal distribution
	reunakausaalivaikutus marginal causal effect
	reunalukumäärä marginal frequency
	reunamalli marginal model
	reunarakennemalli marginal structural model

reunariippumattomuus	marginal independence	riskifunktio	risk function
reunariippuvuus	marginal dependence	riskihakuinen henkilö	risk prone person
reunasumma	marginal total	riskijoukko	risk set
reunataulukko	marginal table	riskijoukkopaiminta	risk set sampling
reunatiheysfunktio	marginal density function	riskikertymä	cumulative risk
reunatodennäköisyys	marginal probability	riskineutraali mitta	risk neutral measure
reunauuskottavuus	marginal likelihood	riskisuhde	relative risk, risk ratio
reunavaikutus	marginal effect	riskitekijä	risk factor
reunayhteys	marginal association	riskiteoria	risk theory
riippumaton muuttuja ³	explanatory variable	riskitön arvopaperi	riskless asset
riippumaton sensurointi	independent censoring	riskiä kaihtava henkilö	risk averse person
riippumattomat ja normaali-jakautuneet	normally and independently distributed	ristiintaulukointi	cross tabulation
riippumattomat ja samoin jakautuneet	independent and identically distributed	ristiinvahvistus	cross-validation
riippumattomat ja samoin jakautuneet	independent and identically distributed	ristiinvalidointi	cross-validation
riippumattomat lisäykset	independent increments	ristikkäisasetelma	cross-over design
riippumattomien komponenttien analyysi	independent component analysis	ristikorrrelaatio	cross-correlation
riippumattomuus	independence	ristikovarianssi	cross-covariance
riippuva muuttuja ⁴	outcome variable, response variable	ristispektri	cross spectrum
riippuvuus	dependence	ristispektritiheys	cross-spectral density
riippuvuus edeltävästä	antependence	ristisuhde	odds ratio
rimamalli	hurdle model	ristitaulukko	cross table
rinnakkaisten koordinaattien kuvio	parallel coordinate plot	ristitulosuhde	cross-product ratio, odds ratio
rinnakkaisten ryhmien asetelma	parallel groups design	robusti estimaattori	robust estimator
riski	risk	robustisuus	robustness
		ROC-käyrä	ROC curve
		roikkuva histogrammi	hanging rootogram
		rososakko	roughness penalty
		rotaatio	rotation
		rotaatioasetelma	rotatable design

³Suosittelavampi on selittävä muuttuja.

⁴Suosittelavampia ovat vastemuuttuja ja selitettävä muuttuja.

rukattu estimaattori adjusted estimator
rukattu otospaino adjusted sample weight
rukattu profiiliuskottavuus adjusted profile likelihood
rukkaus adjustment
ruudukko grid
ruusukaavio rose diagram
ruutuotanta grid sampling
ryhmien sisäinen neliösumma within-groups sum of squares
ryhmien välinen neliösumma between-groups sum of squares
ryhmitely aineisto grouped data
ryhmittelyanalyysi cluster analysis
ryhmittelypuu dendrogram
ryhmäfaktori group factor
ryhmäjaollinen asetelma group divisible design
ryhmäkaltaistus group matching, frequency matcing
ryhmäpaikkaus aggregate imputation
ryhmäperäkkäinen koe group-sequential experiment
ryhmäperäkkäisasetelma group-sequential design
ryhmäseulonta-asetelma group screening design
ryhmätason tutkimus aggregate-level study
ryväs cluster
ryväotanta cluster sampling
ryväprosessi cluster process
ryvässatunnaistus cluster randomization
ryvästys clustering
ryöppyprosessi cascade process
räjähtävä prosessi explosive process

S

saapumishetki entrance time
saapumisprosessi arrival process
saapumisväliaika interarrival time
saatavilla olevien havaintojen analyysi available case analysis
saavutettava salkku attainable portfolio
saavutushetki first passage time
sairastuneiden osuus attack rate
sairaus-kuolemamalli illness-death model
sakkofunktio penalty function
sakkotermi penalty term
sakotettu estimointi penalized estimation
sakotettu pienin neliösumma penalized least squares
sakotettu uskottavuus penalized likelihood
samanaikainen estimointi simultaneous estimation
samanaikainen päättely simultaneous inference
samanaikaiset luottamusvälit simultaneous confidence intervals
samanlaiset parit concordant pairs
samanlaisuuskerroin similarity coefficient
samanlaisuusmatriisi similarity matrix
samanlaisuusmitta similarity measure
samavarianssinen homoskedastic
sanapilvi word cloud
sanapussimalli bag-of-words model
sandwich-estimaattori sandwich estimator

sarjakorrelaatio	serial correlation	satunnaismuuttuja	random variable
sarjalaimennuskoe	serial dilution assay	satunnaisotanta	random sampling
sarjaseulonta	high-throughput screening	satunnaisotos	random sample
sarjatoistopyynti	multiple recapture design	satunnaispaikkaus	stochastic imputation
sarjavaihtelu	serial variation	satunnaispoiminta	random selection
s-arvo	s-value	satunnaisprosessi	random process, stochastic process
sateenvarjo-ongelma	umbrella problem	satunnaissuure	random variable
satulapisteapproksimaatio	saddle point approximation	satunnaistamaton	non-randomized
satunnainen	random	satunnaistapahtuma	random event
satunnaisesti puuttuva	missing at random	satunnaistermi	innovation term
satunnaisfunktio	random function	satunnaistessellaatio	random tessellation
satunnaisheilunta	stochastic volatility	satunnaistettu järjestys-	rank-randomization test
satunnaisilmiö	random phenomenon	lukutesti	
satunnaisjako	random allocation	satunnaistettu kliininen koe	randomized clinical trial
satunnaisjono	random sequence	satunnaistettu koe	randomized experiment, randomized trial
satunnaisjoukko	random set	satunnaistettu päätössääntö	randomized decision rule
satunnaiskenttä	random field	satunnaistettu testi	randomized test
satunnaiskerroin	random coefficient	satunnaistettu vastaus	randomized response
satunnaiskomponentti	random component	satunnaistettujen lohkojen	asetelma randomized blocks design
satunnaiskulku	random walk	asetelma	
satunnaisluku	random number	satunnaistetut lohkot	randomized blocks
satunnaislukugeneraattori	random number generator	satunnaistuksen mukainen	analyysi as-randomized analysis
satunnaislukukehitin	random number generator	analyysi	
satunnaismatriisi	random matrix	satunnaistus	randomization
satunnaismetsä	random forest	satunnaistustesti	randomization test
satunnaismitta	random measure		
satunnaismosaiikki	random tessellation		

satunnaisuus	randomness		
satunnaisvaihtelu	random variation	semimartingaali	semi-martingale
satunnaisvaikutus	random effect	semiparametrinen	semi-parametric
satunnaisvaikutusmalli	random-effects model	sensitiivisyys	sensitivity
satunnaisvektori	random vector	sensurointi	censoring
satunnaisverkko	random graph	sensurointi oikealta	right-censoring
satunnaisvirhe	random error	sensurointi vasemmalta	left censoring
saturoitu malli	saturated model	sensuroitu havainto	censored observation
segregaatio	segregation	sentiili	centile
segregaatioanalyysi	segregation analysis	sentroidimenetelmä	centroid method
sekajakauma	mixed distribution	separoituva avaruus	separable space
sekakoosteinen	heterogeneous	seulaestimaattori	sieve estimator
sekamalli	mixed model, mixed-effects model	seulonta	screening
sekanttimenetelmä	secant method	seulonta-asetelma	screening design
sekaspektri	mixed spectrum	seulontatesti	screening test
sekastrategia	mixed strategy	seuranta	follow-up
sekoitekoe	mixture experiment	seuruu	follow-up
sekoitemalli	mixture model	Shannon-informaatio	Shannon information
sekoittaja	confounder	sidos	tie
sekoittava jakauma	mixing distribution	sigma-algebra	sigma-algebra
sekoittava tekijä	confounder, confounding factor	sigmakenttä	sigma-field
sekoittava muuttuja	confounder, confounding variable	sigmoidikäyrä	sigmoid curve
sektorikuvio	pie diagram, pie chart	signaali-kohinasuhde	signal to noise ratio
sekvenssimalli	sequence model	signaalinkäsittely	signal processing
selitettävä muuttuja	outcome variable	sihteeriongelma	secretary problem
selittäjä	explanatory variable	siilosirontakuvio	binned scatter plot
selittävä koe	explanatory trial	siirrettävyys	transportability
selittävä muuttuja	explanatory variable, regressor	siirto	translation
selitysaste	coefficient of determination	siirtofunktio	transfer function
selvitys	inquiry	siirto-oppiminen	transfer learning
semimarkovprosessi	semi-Markov process		

ing	simulointi simulation
siirtymäintensiteetti transition intensity	simuloitu hehkus simulated annealing
siirtymämatriisi transition matrix	simuloitu verho simulated envelope
siirtymätaajuus transition rate	simultaaniluottamusväli simultaneous confidence intervals
siirtymätodennäköisyys transition probability	simultaaninen autoregressio simultaneous autoregression
sijaintiluku measure of location	simultaaninyhtälöharha simultaneous equations bias
sijaintimitta measure of location	simultaaninyhtälömalli simultaneous equations model
sijaintiparametri location parameter	singulaariarvohajotelma singular value decomposition
sijaintiperhe location family	singulaarinen jakauma singular distribution
sijainti-skaalaperhe location-scale family	singulaarinen matriisi singular matrix
sijaintitunnusluku measure of location	sironta scatter
sijaismuuttuja surrogate variable	sirontakuvio scatter plot
sijaisvaste surrogate outcome	sirontakuviomatriisi scatterplot matrix
sijakisa ranking	sisäaita inner fence
sijaluku rank ₁	sisäinen kasvuvauhti intrinsic growth rate
sijalukukorrelaatio rank correlation	sisäinen validiteetti internal validity
sijalukuregressio rank regression	sisäinen varianssi internal variance
sijatunnusluku rank order statistic	sisäkkäinen asetelma nested design
sijavertailu ranking	sisäkkäinen luokittelu hierarchical classification
sijoitusrahasto mutual fund	sisäkkäinen otanta nested sampling
silmukka loop	sisäkkäiset hypoteesit nested hypotheses
silotin smoother	sisäkkäiset mallit nested models
silottava splini smoothing spline	sisäkorrelaatio intraclass correlation
silotus smoothing	sisällönanalyysi content analysis
siluettikuvio silhouette plot	sisältymistodennäköisyys inclu-
similaarinen testi similar test	
similaariset alueet similar regions	
similaarit matriisit similar matrices	
simpleksialgoritmi simplex algorithm	
Simpsonin paradoksi Simpson's paradox	
simulaattori simulator	

sion probability	spatiaalinen data spatial data
sisämääräytyneisyys endogeneity	spatiaalinen mediaani spatial median
sisämääräytynyt muuttuja endogenous variable	spatiaalinen paripotentiaali spatial pair-potential
sisätulo inner product, dot product	spatiaalinen pisteprosessi spatial point process
sisävarianssi intraclass variance	spatiaalinen tilastotiede spatial statistics
sisäänajojakso₁ burn-in period	spektri spectrum
sisäänajojakso₂ run-in period	spektrianalyysi spectral analysis
sisäänmuutto in-migration	spektrihajotelma spectral decomposition
sivuutettava ignorable	spektrikertymäfunktio spectral distribution function
sivuuttamaton nonignorable	spektrisäde spectral radius
skaala scale ₁	spektritiheys spectral density
skaalakerroin scale factor	spesifisyys specificity
skaalaparametri scale parameter	splini spline
skaalattu devianssi scaled deviance	stabiili jakauma stable distribution
skaalaus scaling	stabiili tila stable state
skaalautuvuus scalability	stabiili väestö stable population
skalaari scalar	standardinormaalijakauma standard normal distribution
skeptinen priori sceptical prior	standardointi standardization ₂
skoori score ₁	standardoitu jännös standardized residual
skoorifunktio score function	standardoitu kerroin standardized coefficient
skoorijännös score residual	standardoitu muuttuja standardized variable
skoorisuure score statistic	stationaarinen jakauma stationary distribution
skooritesti score test	stationaarinen prosessi stationary process
skoorivektori score vector	stationaarinen väestö stationary population
S-käyrä sigmoid curve	stereologia stereology
sokkimalli shock model	stokastiikka stochastics
sokkokoe blinded experiment	stokastinen approksimaatio stochastic approximation
sokkoutus blinding	stokastinen integraali stochastic
solmu node, vertex	
solmupiste knot	
solufrekvenssi cell frequency	
solulukumäärä cell frequency	
solvenssi solvency	
sovite fitted value	
sovite-ero difference in fits	
sovitettu malli fitted model	
sovitteen hyvyys goodness of fit	
sovitus fitting	
spagettikuvio spaghetti plot	

integral	
stokastinen järjestys	stochastic ordering
stokastinen malli	stochastic model
stokastinen matriisi	stochastic matrix
stokastinen optimointi	stochastic optimization
stokastinen paikkaus	stochastic imputation
stokastinen prosessi	stochastic process
stokastinen raja-arvo	limit in probability
stokastinen riippuvuus	stochastic dependence
stokastinen suppeneminen	stochastic convergence
stokastinen volatiliiteetti	stochastic volatility
stokastisesti rajoitettu	bounded in probability
Studentin jakauma	Student's distribution, <i>t</i> -distribution
stylometria	stylometry
subjektiivinen todennäköisyys	subjective probability
submartingaali	submartingale
SU-estimaattori	maximum likelihood estimator, ML estimator
suhdanne	business cycle
suhde	ratio
suhdeasteikko	ratio scale
suhde-estimaattori	ratio estimator
suhdeluku	ratio
suhteellinen elossaolo	relative survival
suhteellinen frekvenssi	relative frequency
suhteellinen kiintiöinti	proportional allocation
suhteellinen otanta	proportional sampling
suhteellinen riski	relative risk, risk ratio
suhteellinen tehokkuus	relative efficiency
suhteellinen uskottavuus	relative likelihood
suhteellinen varianssi	relative variance
sukukorrelaatio	familial correlation
sukupuuton todennäköisyys	extinction probability
sulautuminen	coalescence
suljettu väestö	closed population
sumeaa joukko-oppi	fuzzy set theory
sumeaa logiikka	fuzzy logic
summakaavio	cumulative sum chart
summakontrasti	sum contrast
summautuva malli	additive model
summautuvien uhkien malli	additive hazards model
suodatin	filter
suodatus	filtering
suodin	filter
suojattu tieto	protected data
suojausstrategia	hedging strategy
suorakaidehila	rectangular lattice
suora uskottavuus	direct likelihood
suora vaikutus	direct effect
suora vakiointi	direct standardization
supereksogeenisuus	super exogeneity
supermartingaali	supermartingale

supernormaalisuus	supernormality	suuriulotteinen aineisto	high-dimensional data
superoppija	super learner	suurten lukujen laki	law of large numbers
superpopulaatiomalli	superpopulation model	suurten otosten teoria	large sample theory
supistettu tarkastus	reduced inspection	suurten poikkeamien teoria	large deviation theory
suppeneminen	convergence	suuruusluokka	order of magnitude
suppilokuvio	funnel plot	sykli	cycle
suunnallinen satunnaiskulku	random walk with drift	syklinen malli	cyclic model
suunnattu syklitön verkko	directed acyclic graph	symmetrinen jakauma	symmetric distribution
suunnitellusti puuttuva	missing by design	symmetrinen otanta	symmetric sampling
suunniteltu vertailu	planned comparison	synteettinen aineisto	synthetic data
suunta	drift ₁	synteettinen estimaattori	synthetic estimator
suunta-aineisto	directional data	syntymäkohortti	birth cohort
suuntaamaton verkko	undirected graph	syntymä-kuolemaprosessi	birth-death process
suuraineisto	big data	syntymäprosessi	birth process
suurdata	big data	syntyvyys	birth rate
suurimman entropian menetelmä	maximum entropy method	systemaattinen allokonti	systematic allocation
suurimman posterioritiheyden alue	highest posterior density region	systemaattinen jako	systematic allocation
suurimman posterioritiheyden estimaattori	maximum a posteriori estimator	systemaattinen otanta	systematic sampling
suurimman posterioritiheyden väli	highest posterior density interval	systemaattinen virhe	systematic error, bias
suurimman uskottavuuden estimaattori	maximum likelihood estimator	sysäys	perturbation ₂
suurimman uskottavuuden menetelmä	method of maximum likelihood	sysäysvastefunktio	impulse response function
suurin siedetty annos	maximum tolerated dose	syväoppiminen	deep learning
		syyketju	causal chain
		syykohtainen uhka	cause-specific hazard
		syyksilukuosuus	attributable fraction
		syymuuttuja	cause variable

syy-seurausmalli causal model
syy-seurauspäätely causal inference
syy-seuraussuhde causality
sädekuvio radial plot
säännöllisesti vaihteleva regularly varying
säännön edeltäjäosa antecedent of a rule
säännön seuraajaosa consequent of a rule
säännön vartalo body of a rule
sääntely regularization
säästeliäisyys parsimony
säätömuuttuja control variable₁
säätöteoria control theory

T

taajuus frequency₂, rate₁
taajuusalue₁ frequency domain
taajuusalue₂ bandwidth₁
taajuuspinta frequency surface
taajuusvastefunktio frequency response function
taakselaskenta backcalculation
taakseprojisointi backprojection
taaksesovitus backfitting
taakseyhtälö backward equation
taikaneliöasetelma magic square design
taitekohta turning point
taitettu jakauma folded distribution
taitettu kontingenssitaulukko folded contingency table
taitettu mediaani folded median
taitettu normaalijakauma half-normal distribution
taitettu ristitaulukko folded contingency table
taittaminen folding
takaovikorjaus backdoor adjust-

ment
takaovipolku backdoor path
takaovivakiointi backdoor adjustment
taksonomia taxonomy
taluskuvio scree plot
talustesti scree test
tapahtuma event
tapahtuma-aika time-to-event
tapahtuma-aika-analyysi time-to-event analysis
tapahtuma-alttiussuhde occurrence-exposure rate
tapahtuma-avaruus event space
tapahtumahistoria-analyysi event history analysis
tapahtumaväli aika interevent time
tapaus₁ case
tapaus₂ instance
tapaus-alustatutkimus case-base study
tapausavaruus instance space
tapaus-kohorttitutkimus case-cohort study
tapauskuolleisuus case fatality ratio
tapaus-pohjatutkimus case-base study
tapaus-verrokitutkimus case-control study, case-referent study
tappio loss
tappioarvo Value at Risk
tappiofunktio loss function
tappiokolmio loss triangle
tappiomatriisi loss matrix
tappion karttaminen loss aversion
tarkentuvuus consistency
tarkkuus₁ accuracy
tarkkuus₂ specificity
tarttuvuus transmission rate₁
tartuntatodennäköisyys trans-

mission probability	teoreettinen tilastotiede theoretical statistics
tasa-arvokäyrä contour line	teho efficacy
tasa-arvosuora line of equal distribution	tehokas estimaattori efficient estimator
tasaerä annuity ₂	tehokkaiden markkinoiden hypoteesi efficient market hypothesis
tasahavainto tie	tehokkuus efficiency
tasainen integroituvuus uniform integrability	tehollinen otoskoko effective sample size
tasainen suppeneminen uniform convergence	tehollinen vapausasteluku effective degrees of freedom
tasaisen jakautumisen suora line of equal distribution	tehostus boosting
tasaisen siirtymän autoregressiivinen malli smooth transition autoregressive model	tekijä factor ₁
tasaisesti integroituva uniformly integrable	tekijäkoe factorial experiment
tasajakauma uniform distribution	tekijäkoeasetelma factorial design
tasakoosteinen homogeneous	tekijän taso level of factor
tasakorrelaatio uniform correlation	tekijänvaihtokriteeri factor reversal criterion
tasapaino equilibrium	tekijöihinjako factorization
tasapainoinen vajaalohkoasetelma balanced incomplete block design	tekijöihinjakokriteeri factorization criterion
tasapainojakauma equilibrium distribution	tekoäly artificial intelligence
tasapainotilajakauma steady-state distribution	tekstinlouhinta text mining
tasapainoyhtälö balance equation	tensori tensor
tasaväliotanta systematic sampling	teoreema theorem
tasoitin smoother	termiinisopimus forward contract
tasoitus smoothing	testi test
tasoituskäyrä graduation curve	testiaineisto test data
tasokartta level map	testin empiirinen koko empirical size of test
taudinkantaja carrier of disease	testin koko size of test
taulukko table, array	testipatteri battery of tests
tavoitekoe target trial	testisarja test series
tavoitekokeen jäljittely target trial emulation	testisuure test statistic
	tetrakorinen korrelaatio tetrachoric correlation
	tiedonlouhinta data mining
	tiedonsiirtonopeus transmission rate ₂

tiedosto	data set	tem	
tietoinen suostumus	informed consent	tilastollinen avaruus	statistical space
tietokanta	database	tilastollinen hypoteesi	statistical hypothesis
tietokoneavusteinen	computer assisted	tilastollinen järkeily	statistical reasoning
tietosuoja	disclosure control	tilastollinen koneoppiminen	statistical machine learning
tietue	record ₁	tilastollinen kooste	statistical ensemble
tietuekytkentä	record linkage	tilastollinen malli	statistical model
tietämystekniikka	knowledge technology, knowledge engineering	tilastollinen merkitsevyys	statistical significance
tiheys	density	tilastollinen oppiminen	statistical learning
tiheysestimointi	density estimation	tilastollinen perustelu	statistical reasoning
tiheysfunktio	density function	tilastollinen päättely	statistical inference
tiheyshistogrammi	density histogram	tilastollinen testi	statistical test
tiheysotanta	density sampling	tilastollisesti erottuva	statistically discernible
tiheyspoiminta	density sampling	tilastollisesti merkitsevä	statistically significant
tiheys suhdemalli	density ratio model	tilastolukutaito	statistical literacy
tikasindeksi	ladder index	tilastotiede	statistics ₂ , statistical science
tikasmuuttuja	ladder variable	tilastotieteellinen	statistical ₂
tikkausmenetelmä	staircase method, up-and-down method	tilastotieteilijä	statistician
tikunkatkaisuprosessi	stick breaking process	tilastotoimi	statistical service
tila-aikaryvä	space-time cluster	tilastoviranomainen	statistical authority
tila-avaruus	state space	tilastoyksikkö	observational unit
tilajoukko	state space	tiukennettu tarkastus	tightened inspection
tilanneparametrit	incidental parameters	tiukin testi	most stringent test
tilapistekuvio	spatial point pattern	tiukka järjestys	tight ordering
tilapisteprosessi	spatial point process	<i>t</i>-jakauma	<i>t</i> -distribution
tilasto	statistics ₁	<i>t</i>-jäännös	studentized residual
tilastoahdistus	statistics anxiety	<i>t</i>-kerroinero	studentized differ-
tilastografiikka	statistical graphics		
tilastollinen	statistical ₁		
tilastollinen asiantuntijajärjestelmä	statistical expert sys-		

ence in betas	kernel
t-sovite-ero studentized difference in fits	toimenpideanalyysi intervention analysis
tobit-malli tobit"-model	toimintakykyiset elinvuodet disability adjusted life years
todella suurten lukujen laki law of truly large numbers	toimivuusmitta operating characteristic
todennäköinen probable	toisen asteen otantayksikkö secondary sampling unit
todennäköinen virhe probable error	toisen kertaluvun tehokkuus second order efficiency
todennäköisyydet generoiva funktio probability generating function	toisensa poissulkevat tapahtumat mutually exclusive events
todennäköisyys probability	toisinne replicate
todennäköisyysarvio probability judgement	toisintavaytiminen Hilbertin avaruus reproducing kernel Hilbert space
todennäköisyysavaruus probability space	toistaminen replication
todennäköisyysmäfunktio probability generating function	toistava recursive
todennäköisyysjakauma probability distribution	toistettavuus replicability, repeatability
todennäköisyyskuvio probability plot	toisto repetition
todennäköisyyslaskenta probability calculus	toistokutsu call-back
todennäköisyysmalli probability model	toistoluotaus periodic survey, repeated survey
todennäköisyysmitta probability measure	toistomittausasetelma repeated measurements design, repeated measures design
todennäköisyysotanta probability sampling	toisto-otanta resampling
todennäköisyyspainotus probability weighting	toistopyyntien asetelma multiple recapture design
todennäköisyyspäättely probabilistic inference	toistopyyntiotanta capture-recapture sampling
todennäköisyysrajat probability limits	toistotarkkuus precision ₁
todennäköisyysteoria probability theory	toleranssi tolerance
todennäköisyystiheys probability density function, density function	toleranssiraja tolerance limit
todennäköisyysydin probability	toleranssiväli tolerance interval
	tontiini tontine
	tosiaikainen tutkimus concurrent study
	toteuma realization
	toteutettavuustutkimus feasibility study

toteutuma	realization	cient statistic	
totuudenmukaisuus	veracity	tyhjä joukko	empty set
translaatio	translation	typistetty havainto	truncated observation
transpoosi	transpose	typistys oikealta	right-truncation
trendi	trend	typistys vasemmalta	left-truncation
trendin poisto	detrending	tyylitelty ominaispiirteet	stylized facts
trimmattu keskiarvo	trimmed mean	tyylitilastotiede	stylometry
troijalainen neliö	Trojan square	tyyppi	mode
Tukeyn vaihteluvälitest	Tukey's range test	työmuuttuja	working variate
Tshebyshevin epäyhtälö	Chebyshev's inequality	työvoimatutkimus	labor-force survey
t-suure	<i>t</i> -statistic	työttömyysaste	unemployment rate
t-testi	<i>t</i> -test	työttömyysosuus	unemployment rate
t-tunnusluku	<i>t</i> -statistic	tähtisadekuvio	starburst plot
tukifunktio	support function	tähtitilastotiede	astrostatistics
tukivektorikone	support vector machine	tärkeysotanta	importance sampling
tuloavaruus	product space	täsmälleen identifioituva	just identified
tulointegraali	product integral	täydellinen erottelu	complete separation
tulomitta	product measure	täyden informaation menetelmä	full information method
tulomomentti	product moment	täydennetty Dickeyn–Fullerin testi	augmented Dickey–Fuller test
tulomomenttikorrelaatio	product moment correlation	täysehdollistettu jakauma	fully conditional distribution
tunneanalyysi	sentiment analysis	täysin satunnaisesti puuttuva	missing completely at random
tunnistin	identifier	täysiteikijäkoe	full factorial design
tunnusluku	statistic	täyssatunnaistettu asetelma	completely randomized design
tuottomatriisi	pay-off matrix	törmäytin	collider
tutkimuksen toisto	replication study	törmäytinharha	collider bias
tutkimus	study		
tutkimuspopulaatio	study population		
tutkimusväestö	study population		
tutkimusyksikkö	study unit		
t-vaihteluvälisuure	studentized range statistic		
tyhjentyvyys	sufficiency		
tyhjentävä dimension pienenys	sufficient dimension reduction		
tyhjentävä tunnusluku	suffi-		

U

- uhka** hazard
- uhkafunktio** hazard function
- uhkakertymä** cumulative hazard, integrated hazard
- uhkasuhde** hazard ratio
- uhkatiheys** hazard rate, failure rate₁
- ulkoaita** outer fence
- ulkoinen pätevyys** external validity
- ulkoinen validiteetti** external validity
- ulkomääräytyneisyys** exogeneity
- ulkomääräytynyt muuttuja** exogenous variable
- ulkotulo** outer product
- ulotteisuus** dimensionality
- ulottuvuuksien vähentäminen** dimension reduction
- ulottuvuus** dimension
- ultrametrinen puu** ultrametric tree
- unioni** union
- U:n muotoinen jakauma** U-shaped distribution
- unohtavaisuus** lack of memory property
- upotettu prosessi** embedded process
- upotettu tapaus-verrokkitutkimus** nested case-control study
- usean vaihteluvälin testi** multiple range test
- uskomusaste** degree of belief
- uskomusfunktio** belief function
- uskomusverkko** belief network
- uskottavuus** likelihood
- uskottavuusfunktio** likelihood function
- uskottavuusosamäärä** likelihood ratio
- uskottavuusosamääräsuure** likelihood ratio statistic
- uskottavuusosamäärätesti** likelihood ratio test
- uskottavuusperiaate** likelihood principle
- utiliteetti** utility
- utiliteettiteoria** utility theory
- U-tunnusluku** *U*-statistic
- uudelleenpainotus** reweighting
- uudelleenparametointi** reparametrization
- uudistuva** regenerative
- urnamalli** urn model
- uusioaggregointi** bootstrap aggregating, bagging
- uusioestimaatti** bootstrap estimate
- uusioestimaattori** bootstrap estimator
- uusioluottamusväli** bootstrap confidence interval
- uusiomenetelmä** bootstrap method
- uusio-otanta** bootstrap sampling, bootstrapping
- uusio-otantamenetelmä** bootstrap method
- uusio-otos** bootstrap sample
- uusio-*p*-arvo** bootstrap *p*-value
- uusioreplikaatti** bootstrap replicate
- uusiotestaus** bootstrap testing
- uusioisto** bootstrap replication
- uusiuutumisaika** recurrence time₂
- uusiuutumisaikajakauma** renewal distribution
- uusiuutumisfunktio** renewal function
- uusiuutumisprosessi** renewal process
- uusiuutumisriski** recurrence risk
- uusiuutumistaajuus** renewal rate

uusiutumisteoria renewal theory
uusiutumistiheys renewal rate
uusiutuva recurrent₂
uuspainotus reweighting
uusparametrointi reparametrization

V

vaaleanpunainen kohina pink noise
vahinkovakuutus non-life insurance
vahva eksogeenisuus strong exogeneity
vahvasti stationaarinen prosessi strictly stationary process
vahva suppeneminen strong convergence
vahva suurten lukujen laki strong law of large numbers
vahva tarkentuvuus strong consistency
vahvistava data-analyysi confirmatory data analysis
vahvistus₁ gain₁
vahvistus₂ validation
vahvistusaineisto validation data, validation set
vahvistusoppiminen reinforcement learning
vaihdettavuus exchangeability
vaihe phase
vaihedigrammi phase diagram
vaiheittainen päivittäminen sequential updating
vaihekaavio phase diagram
vaihekulma phase
vaihemuutos phase transition
vaihespektri phase spectrum
vaihtelevaulotteinen MCMC reversible jump Markov chain Monte Carlo

vaihtelu variability, variation
vaihtelumitta measure of variation, measure of dispersion
vaihtelusuhte extremal quotient
vaihteluväli range₁
vaihteluvälin pituus range₂
vaihteluvälin keskipiste mid-range
vaihteluvälin pituus range
vaihtoasetelma switch-back design
vaihtoehtoinen hypoteesi alternative hypothesis
vaihtokelpoisuus exchangeability
vaihtoregressiomalli switching regression model
vaihtovuoroasetelma cross-over design, changeover design
vaikuttava havainto influential observation
vaikuttavuus effectiveness
vaikuttavuuskoe pragmatic trial
vaikutuksen muovaja effect modifier
vaikutuksen suuruus effect magnitude
vaikutus₁ effect
vaikutus₂ influence
vaikutusfunktio influence function
vaikutuskoko effect size
vaikutustunnusluku influence statistic
vaimeneminen attenuation
vajaa-asteinen matriisi singular matrix
vajaa latinalainen neliö incomplete Latin square
vajaalohkoasetelma incomplete block design
vajaasovitus underfitting
vajaatehoinen estimaattori inefficient estimator

vajaavakiointi	underadjustment	valvontakortti	control chart
vakaa estimaattori	robust estimator	vanhoista vanhimmat	oldest old
vakaa tunnusluku	robust statistic	vapaaehtoisharha	volunteer bias
vakaus	robustness	vapausasteet	degrees of freedom
vakinainen väestö	de jure population	vapausasteluku	degrees of freedom
vakioilmaantuvuus	standard rate	vararikko-ongelma	ruin problem
vakioimaton estimaattori	unadjusted estimator	vararikkotodennäköisyys	ruin probability
vakiointi	adjustment, standardization ₁	variaatioetäisyys	variational distance
vakioitu estimaattori	adjusted estimator	variaatiokerroin	coefficient of variation
vakioitu kuolevuussuhde	standardized mortality ratio	variaatiopäätely	variational inference
vakioiva muuttuja	adjusting variable, control variable ₂	varianssi	variance
vakio-otantaosuus	uniform sampling fraction	varianssianalyysi	analysis of variance
vakiotermi	intercept	varianssikomponentti	variance component
vakioväestö	standard population	varianssin inflaatiokerroin	variance inflation factor
vakuutus	insurance	varianssin vakioiva muunnos	variance stabilizing transformation
vakuutustilastotiede	actuarial statistics	varimax-rotaatio	varimax rotation
validiteetti	validity	variogrammi	variogrammi
validointi	validation	varmistus	ascertainment ₂
validointiaineisto	validation data	varsilehtikuvio	stem-and-leaf diagram
valikoitumisharha	selection bias	vasemmalle vinous	negative skewness
valikoitumismalli	selection model	vastaaajaryhmä	response group
valikoitumispriori	selection prior	vastaaajavetoinen otanta	respondent-driven sampling
valikoivan kasvun prosessi	preferential attachment process	vastahypoteesi	alternative hypothesis
valintaharha	selection bias	vastakaltaistus	counter-matching
valintahetki	optional time, stopping time	vastakkaismerkkinen	antithetic
valintaperusteinen otanta	choice-based sampling	vastamuuttuja	antithetic variate
valkoinen kohina	white noise	vastasuhde	odds
vallitsevuus	prevalence	vastatapahtuma	complement

event	tion of graph
vastaus response ₂ , answer	verkosto configuration
vastausalttiuspaino response propensity weight	verrannollisten uhkien malli proportional hazards model
vastausaste response rate	verrannollisten vastasuhteiden malli proportional odds model
vastausilmais response indicator	verrokki control
vastausindikaattori response indicator	verrokkiryhmä control group
vastauskato nonresponse	vertaileva koe comparative experiment
vastausmekanismi response mechanism	vertailuparametri comparative parameter
vastausvirhe response error	vertailuryhmä comparison group, control group
vastavirta-algoritmi backpropagation algorithm	vetokerroin odds
vaste outcome, response ₁	vetosuhde odds ratio
vasteaikajakauma response time distribution	vierashavainto outlier
vasteen muunnos response metameter	vierusluokkamalli adjacent categories model
vastemuuttuja outcome variable, response variable	vierusmatriisi adjacency matrix
vastepinta response surface	vietto drift ₁
vastepintamethodiikka response surface methodology	viiden luvun tiivistys five-number summary
vastetapahtuma endpoint	viimeisen havainnon toisto last observation carried forward
vektoriaikasarja vector time series	viipalepoimija slice sampler
vektoriautoregressio vector auto-regression	viipaloitu käänteinen regressio sliced inverse regression
velkakirja bond	viipymisaika occupation time, sojourn time
Venn-diagrammi Venn diagram	viisikkokriteeri pentad criterion
verhofunktio envelope function	viitearvo reference value
verhokäyräanalyysi data envelopment analysis	viitejakso reference period, base period
verkko graph ₃	viitepopulaatio reference population
verkkoinformaatiokriteeri network information criterion	viitepriori reference prior
verkkohaastattelu web-interview	viiteraja reference limit
verkkokysely internet survey	viiteväestö reference population
verkkomalli graphical model	viiteväli reference interval
verkko-otanta network sampling	viivakoodikuvio stripes plot
verkon moralisointi moraliza-	viivakuvio line plot
	viivaspektri line spectrum

viive ⁵ lag	virhekomponenttimalli error components model
viiveikkuna lag window	virheluokittelu misclassification
viiveoperaattori lag operator, backshift operator	virhemarginaali margin of error, error margin
viivästynyt aloitus delayed entry	virheosuus error rate
viivästysoperaattori backshift operator	virhetermi error term, disturbance ₂
vikaantuminen failure	virhevarianssi error variance
vikaantumisaika failure time	virstanpylväsanalyysi landmark analysis
vikakertymä failure function	visualisointi visualisation
vikasuure diagnostic statistic	viulukuvio violin plot
vikataajuus failure rate ₂	voima power ₁
vino oblique	voimafunktio power function
vinoheijastus oblique reflection	voimakkain testi most powerful test
vino jakauma skew distribution	voiman lainaaminen borrowing of strength
vinonormaalijakauma skew-normal distribution	volatiliteetti volatility
vinoprojektio oblique projection	volyymi-vaihtelu-vauhti volume, variety, velocity
vinosymmetrinen matriisi skew-symmetric matrix	vuokaavio flow chart
vinous skewness	vuorikiipeilyalgoritmi hill-climbing algorithm
vioittuminen failure	vuorottelevien ehdollistusten regressio alternating conditional expectations regression
vipuvoima leverage	vuosiaineisto annual data
virallinen tilasto official statistics	vyörykuvio scree plot
virhe error	vyörytesti scree test
virheen itseisarvo absolute error	väestö population ₂
virheen kasautuminen propagation of error, error propagation	väestölaskenta census ₂
virheenkorjausmalli error correction model	väestönkasvumalli population growth model
virheen leviäminen propagation of error, error propagation	väestönlaskija enumerator
virheet selittäjissä -malli errors in variables model	väestönmuutostiedot vital statistics
virhefunktio error function	väestönmuutostilasto vital statistics
virhekappaleiden osuus fraction defective	väestöpyramidi population pyra-
virhekertymä cumulative error	
virhekeskineliö error mean square	

⁵Taivutetaan “viipeen” tai “viiveen”.

mid
väestörakenne population structure
väestötiede demography
vähiten merkitsevän erotuksen testi least significant difference test
vähäparametrisuus parsimony
väistytvä tila transient state
välialalyysi interim analysis
väliestimointi interval estimation
välikaltaistus caliper matching
välimatka-asteikko interval scale
välinemuuttuja instrumental variable
välisensurointi interval censoring
välitapahtuma non-terminal event
välittäjä mediator
välitys mediation
välitysanalyysi mediation analysis
väre wavelet
värähdyslaajuus amplitude
värähtelevä prosessi oscillatory process
värähtely oscillation
väsymismalli fatigue model
väärien löydösten osuus false discovery rate
väärinluokittelu misclassification
väärinspesifiointi misspecification
väärin spesifioitu malli misspecified model
väärä negatiivinen false negative
väärä positiivinen false positive

W

Waldin suure Wald statistic
Waldin testi Wald test
Wiener-prosessi Wiener process

Wilcoxonin testi Wilcoxon test
Wilks-ilmio Wilks phenomenon
winsorkeskiarvo winsorized mean
Wishartin jakauma Wishart distribution

Y

Yatesin korjaus Yates' correction
ydinestimaattori kernel estimator
ydinfunktio kernel function
ydinmenetelmä kernel method
ydinodotusarvofunktio kernel mean function
ydinregressioestimaattori kernel regression estimator
ydinsilotus kernel smoothing
ydintasoitus kernel smoothing
ydintiheysestimaattori kernel density estimator
yhden indeksin malli single-index model
yhdenmukaisuus consistency₂
yhden muuttujan funktio univariate function
yhden muuttujan jakauma univariate distribution
yhden otoksen t-testi one-sample t-test
yhden selittäjän regressio simple linear regression
yhdiste union
yhdiste-leikkaustesti union-intersection test
yhdistetty estimaattori pooled estimator
yhdistetty hypoteesi composite hypothesis
yhdistetty Poisson-jakauma compound Poisson distribution
yhdistetty Poisson-prosessi

compound Poisson process	function
yhdistetty spektri mixed spectrum	yhteistodennäköisyys joint probability
yhdistettävyys collapsibility	yhteisötunnistaminen community detection
yhdysvaikutuksen kertaluku order of interaction	yhteys association
yhdysvaikutus interaction	yhteysanalyysi association analysis
yhteenlaskusääntö addition rule	yhteyskerroin coefficient of association
yhteensopivuus₁ concordance	yhteysluokka communicating class
yhteensopivuus₂ goodness of fit	yhteysmitta measure of association
yhteensopivuusindeksi concordance index	yhtäläinen virheluokittelu non-differential misclassification
yhteensopivuusväli compatibility interval, confidence interval	yksiasteotanta single-stage sampling
yhteisennustealue joint prediction region	yksihuippuinen unimodal
yhteisesiintyvyys co-occurrence	yksikkö unit
yhteisestimointi simultaneous estimation	yksikköjuuri unit root
yhteisesti tyhjentävät tunnusluvut jointly sufficient statistics	yksikköjuuritest unit-root test
yhteisfaktori common factor	yksikköjuurta lähestyvä near unit root
yhteisfaktori malli common factor model ₁	yksikkömatriisi identity matrix
yhteisintegroituvuus cointegration	yksikkövastauskato unit non-response
yhteisjakauma joint distribution	yksikköympyrä unit circle
yhteiskertymäfunktio joint cumulative distribution function	yksikytkösryhmittely single-linkage clustering
yhteiskorrelaatiokerroin multiple correlation coefficient	yksikäsitteisyys uniqueness
yhteiskuntatiete social data science	yksikätkäinen rosvo one-armed bandit
yhteiskuntatilastotiede social statistics	yksikön sisällysmistodennäköisyys first-order inclusion probability
yhteisluottamusalue joint confidence region	yksilokeromalli one-compartment model
yhteismomentti joint moment	yksilökohtainen malli subject-specific model
yhteisregressio joint regression	yksilö-muuttujakuvi biplot
yhteisten tekijöiden malli common factor model ₂	yksinkertainen hypoteesi simple hypothesis
yhteistesti omnibus test	
yhteistiheysfunktio joint density	

yksinkertainen satunnaisotanta	simple random sampling	generalized method of moments
yksiosumamalli	one-hit model	yleistetty pienin neliösumma
yksi-pois-ristiinvalidointi	leave-one-out cross-validation	generalized least squares
yksisuuntainen luokittelu	one-way classification	yleistetty stokastinen prosessi
yksisuuntainen testi	one-sided test, one-tailed test	generalized stochastic process
yksisuuntainen vaihtoehtoinen hypoteesi	one-sided alternative hypothesis	yleistetty summautuva malli
yksisuuntainen varianssianalyysi	one-way analysis of variance	generalized additive model
yksisuuntainen vastahypoteesi	one-sided alternative hypothesis	yleistetty ääriarvojakauma
yksitahoinen testi	one-sided test, one-tailed test	generalized extreme-value distribution
yksittäisristiinvalidointi	leave-one-out cross-validation	yleistettävyyys
yksiulotteinen jakauma	univariate distribution	generalizability
yksöissokkokoe	single-blind experiment	yliahajonta
yleinen lineaarinen malli	general linear model	overdispersion
yleisfaktori	general factor	yli-identifiointi
yleiskeskiarvo	grand mean	overidentification
yleistetty additiivinen malli	generalized additive model	ylikaltaistus
yleistetty autoregressiivinen ehdollisen heteroskedastisuuden malli	generalized autoregressive conditional heteroskedasticity model	overmatching
yleistetty estimointiyhtälö	generalized estimating equation	ylilyllästetty asetelma
yleistetty käänteismatriisi	generalized inverse	super-saturated design
yleistetty lineaarinen malli	generalized linear model	ylimartingaali
yleistetty momenttimenetelmä		supermartingale
		yliparametroitu
		overparameterized
		ylipeitto
		overcoverage
		ylipoissonvaihtelu
		extra-Poisson variation
		ylisovitus
		overfitting
		ylitehokas
		superefficient
		ylitysarvokäyrä
		exceedance curve
		ylitysarvotesti
		exceedance test
		yllätysarvo
		surprisal value
		yläkolmiomatriisi
		upper triangular matrix
		yläkvartiili
		upper quartile
		ylävalvontaraja
		upper control limit
		ylös-alasmenetelmä
		staircase method, up-and-down method
		ympäristötilastotiede
		environmental statistics
		YSO
		SRS
		Yulen–Walkerin yhtälöt
		Yule–Walker equations

Z

z-arvo *z*-score

zetajakauma zeta distribution

z-muunnos *z*-transformation

z-testi *z*-test

Ä

äitiaalloke mother wavelet

äitiväre mother wavelet

älykkyyssosamäärä intelligence
quotient

äänestysteoria voting theory

äärellinen jakaumasekoite fi-
nite mixture distribution

**äärellisen perusjoukon korjaus-
kerroin** finite population correc-

tion

äärellisyyskerroin finite popula-
tion correction

äärettömästi jakautuva in-
finitely divisible

äärettömästi osittuva infinitely
decomposable

ääriarvo extreme value

ääriarvointensiteetti extremal
intensity

ääriarvojakauma extreme value
distribution

äärilukema extreme value

äärilukemaintensiteetti ex-
tremal intensity

äärilukemajakauma extreme
value distribution

Lähteet

- Ahlman F. (1865) *Ruotsalais-suomalainen sanakirja*. Suomalaisen kirjallisuuden seuran toimituksia. 38 Osa. Frenckell. Helsinki. https://books.google.fi/books/about/Svenskt_Finskt_lexikon.html?id=E5pBAQAAMAAJ (haettu 27.10.2019).
- Airy G. B. (1835) Report on the Progress of Astronomy during the Present Century. Teoksessa *Report of the First and Second Meetings of the British Association for the Advancement of Science. 2. laitos*, 125–188. John Murray. Lontoo.
- Alameri R., Pöyhönen P. (1953) *Johdatusta tilastolliseen tutkimukseen*. Helsingin yliopisto. Print oy Helsinki. Helsinki.
- Aleksanterin yliopisto (1890–1905) Keisarillinen Suomen Aleksanderin-Yliopisto. Ohjelma lukuvuodelle 1890–1891 – Ohjelma lukuvuodelle 1905–1906 (painatteita eri lukuvuosille). Frenckell tai Weilin & Göös (lukuvuodesta riippuen). Helsinki.
- Anonyymi (1860) Lefnadsteckning (Johan Henrik Eklöfin elämäkerta ruotsiksi). *Suomi, Tidskrift i fosterländska ämnen*, 20, 334–336.
- Andmeanalyyti ja statistika oskussõnastik* (2019) <https://term.eki.ee/termbase/view/8917007/> (haettu 28.2.2020).
- Arppe A. E. (1888) *Suomen tiede-seura 1838–1888, sen järjestys ja toiminta*. J. Simeliuksen perill. osakeyhtiön kirjapaino. Helsinki.
- Bessel F. W. (1816) Untersuchungen über die Bahn des Olbersschen Kometen. *Abhandlungen der Königlischen Akademie der Wissenschaften in Berlin*, 119–160.
- Bonsdorff E. (1919) *Vakuutustieteen matemaattiset alkeet*. Otava. Helsinki.
- Bremer P. (1945) Keskiarvomenetelmistä todennäköisyyslaskennassa. *Matemaatikosten aineiden aikakauskirja*, 9, 135–141.
- Cox D. R., Hinkley D. V. (1974) *Theoretical Statistics*. Chapman & Hall/CRC. Lontoo.
- Cox D. R., Snell E. J. (1981) *Applied Statistics: Principles and Examples*. Chapman & Hall/CRC. Lontoo.
- Cumming G., Calin-Jageman R. (2017) *Introduction to the New Statistics: Estimation, Open Science, and beyond*. Routledge. New York.
- Desrosières A. (2010) *The Politics of Large Numbers: A History of Statistical Reasoning*. Kääntänyt C. Naish. Harvard University Press. Cambridge, Massachusetts.
- Draper N. R., Mäkeläinen T., Nordström K., Pukelsheim F. (1999) Gustav Elfving: An Appreciation. Teoksessa J. Alho (toim.) *Statistics, Registries, and Science. Experiences from Finland*. Statistics Finland. Helsinki.
- Eberl, A., Karl, B. (2023) Centre-Free Kurtosis Orderings for Asymmetric Distributions. *Statistical Papers*. <https://doi.org/10.1007/s00362-023-01403-6> (haettu 23.3.2023).
- Eklöf J. H. (1848) *Kolmiomittaviivain – Merkillisimmät Yhtisoudet ynnä Tasannes-Kolmiomitanto*. Suomalaisen kirjallisuuden seuran toimituksia, 12. Pränättä J. Simeliuksen perillisten tykönä. Imprimatur G. Rein. Helsinki.

- Eklöf J. H. (1850) Jäänlähtöajaat Kokemäenjoen virrassa vuosina 1901–1849, todenvaihe-laskulla määräsi Johan Henrik Eklöf. *Suomi, Tidskrift i fosterländska ämnen*, 9, 189–195. <https://digi.kansalliskirjasto.fi/aikakausi/bindin g/498473> (haettu 27.10.2019).
- Eklöf J. H. (1852) *200 vuotinen Almanakka, alkaen vuodesta 1801, loppuen vuodella 2000*. Näköispainos Jyväskylän yliopiston kirjaston kokoelmista 3. Jyväskylän yliopistopaino 1997.
- Eklöf J. H. (1856) *Pallo-kolmio-mitanto*. Suomalaisen kirjallisuuden seuran toimituksia. 12 Osa, Toinen wihko. Helsingissä, Suomalaisen Kirjallisuuden Seuran Kirjapainossa.
- Elfving F. (1938) Suomen Tiedeseuran historia 1838–1938. Suomentanut Aarni Penttilä. Teoksessa *Suomen Tiedeseura 1838–1938*. Suomen Tiedeseura. Keskuskirjapaino. Helsinki.
- Elfving E. G. (1946) Todennäköisyyslaskenta. Helsingin yliopiston monistustoitisto. Helsinki.
- Elfving E. G. (1956) *Todennäköisyyslaskenta*. Tiedekirjasto N:o 28. Otava. Helsinki.
- Elfving E. G. (1985) Finnish Mathematical Statistics in the Past. Teoksessa T. Pukkila ja S. Puntanen (toim.) *Proceedings of the First International Tampere Seminar on Linear Statistical Models and Their Applications*. Department of Mathematical Sciences/Statistics. University of Tampere. Vammala.
- Europaeus D. E. D. (1852–1853) *Ruotsalais-Suomalainen Sanakirja, 1–2 osa*. Suomalaisen Kirjallisuuden Seuran toimituksia, 16 osa, 1–2 jako. Helsinki.
- Fisher R. A. (1925) *Statistical Methods for Research Workers*. Oliver and Boyd. Lontoo.
- Fuller W. (1987) *Measurement Error Models*. Wiley. New York.
- Gelman A., Carlin J. (2017) Some Natural Solutions to the p -Value Communication Problem – and Why They Won’t Work. *Journal of the American Statistical Association*, 112, 899–901.
- Greenland S. (2017) The Need for Cognitive Science in Methodology. *American Journal of Epidemiology*, 186, 639–645.
- Greenland S. (2019) Valid P -Values behave Exactly as They Should: Some Misleading Criticisms of P -Values and Their Resolution with S -Values, *The American Statistician*, 73 (Suppl 1), 106–111.
- Gylling E. (1917) Tilasto. Teoksessa *Tietosanakirja, IX*. Tietosanakirja-osakeyhtiö. Otava. Helsinki.
- Gylling E., Forsman J. (1917) Tilastotiede. Teoksessa *Tietosanakirja, IX*. Tietosanakirja-osakeyhtiö. Otava. Helsinki.
- Hakulinen L. (1979) *Suomen kielen rakenne ja kehitys. 4. korjattu ja lisätty painos*. Otava. Helsinki.
- Harmaja L. (1939) *Tilastotieteen oppikirja*. Otava. Helsinki.
- Hedges L. V., Olkin I. (1985) *Statistical Methods for Meta-analysis*. Academic Press. San Diego.
- Heikel I. A. (1940) *Helsingin yliopisto 1640–1940*. Otava. Helsinki.
- Heikinheimo I. (toim.; 1955) *Suomen elämäkerrasto*. WSOY. Porvoo.

- Hiltunen E., Linko L., Hemminki S., Hägg M., Järvenpää E., Saarinen P., Simonen S., Kärhä P. (2011) *Laadukkaan mittaamisen perusteet*. Mittaustekniikan keskus MIKES Julkaisu J4/2011, Vantaa. http://www.vtt.fi/inf/pdf/MIKE_S/2011-J4.pdf (haettu 28.2.2020).
- Huomo K. (2005) *Perkeleen kieli: suomen kieli ja poliittisesti korrekki tiede 1800-luvulla*. Väitöskirja. Helsingin yliopisto. Bidrag till kännedom av Finlands natur och folk 166. Suomen Tiedeseura. Helsinki.
- Häkkinen K. (2020) *MOT Nykysuomen etymologinen sanakirja*. MOT Sanakirjat. 2021 Kielikone. www.sanakirja.fi/fin_etymology/finnish-finnish (kaupallinen verkkopalvelu; haettu 16.8.2021).
- Hällström G. G. (1823) Undersökning om vattens volum-förändring af värme, och bestämmelse af den varmegrad, hvarvid vattens täthet är störst. *Kongliga vetenskapsakademiens handlingar*, 193–228. Nordstedt. Tukholma. <https://www.biodiversitylibrary.org/item/107450> (haettu 17.3.2019). Korjauksia artikkelin kaavoihin: *Kongliga vetenskapsakademiens handlingar*, 1824, 510. <https://www.biodiversitylibrary.org/item/107454> (haettu 18.2.2019).
- Ignatius K. E. F. (toim., 1872) *Tilastollinen käsikirja Suomenmaalle*. G. W. Edlund'in kustannuksella. HELSINGISSÄ, J. C. Frencellin ja Pojan tykönä.
- Ingman S. (E. A.) (1850) Ovatko taudit itsenäisiä, eläväisiä olentoja? *Finska Läkare-Sällskapets Handlingar*, IV/II, 171–188.
- International Statistical Institute (2011) *Multilingual Glossary of Statistical Terms*. <https://isi-web.org/index.php/publications/glossary-of-statistical-terms> (haettu 28.2.2020).
- Ioannidis J. P. A. (2019) What Have We (Not) Learnt from Millions of Scientific Papers with P values? *The American Statistician*, 73(Suppl 1), 20–25.
- Itkonen T. (1990) Toimivia välineitä. Paavo Tikkanen suomen kielen kehittäjänä. *Kielikello*. <https://www.kielikello.fi/-/toimivia-valineita-paavo-tikkanen-suomen-kielen-kehittajana> (haettu 1.6.2020).
- Johanson V. F. (1935) *Tilastotieteen perusteet ja pääkohtia Suomen tilastosta*. I. Luentoja Helsingin yliopistossa. Helsinki. Julkaisematon. Moniste on osa Kansallisarkiston Universitas-kokoelmaa.
- Johanson V. F. (1940) *Suomen tilaston pääkohdat*. Kirjapaino ABC. Helsinki.
- Kaitila I. V. (1945) *Liiketaloustilasto*. Liiketaloudellisen tutkimuslaitoksen julkaisuja 5. Kauppatieteellinen yhdistys. Weilin & Göös. Helsinki.
- Kannisto V. (1947) Kuolemansyyt väestöllisinä tekijöinä Suomessa. Väitöskirja. Helsingin yliopisto. Filosofinen tiedekunta. Mercatorin kirjapaino. Helsinki.
- Kansallisarkiston Sakari Leväsen II arkisto. Konseptit lähetetyistä puheista, kirjeistä ja tutkielmista 1885–1893. Jyväskylän toimipiste.
- Kansanvalistus-Seura (1889) *Kansanvalistus-Seuran Kalenteri 1890*. Kansanvalistus-seuran kustantama. Helsingissä 1889.
- Kelson M. (2013) Significantly Misleading. *Significance Magazine*. <https://www.statslife.org.uk/the-statistics-dictionary/1000-the-statistics-dictionary-significantly-misleading> (haettu 28.2.2020).
- Kendall M. G., Buckland W. R. (1971) *A Dictionary of Statistical Terms*, 3. laitos. International Statistical Institute. Oliver & Boyd. Edinburgh.
- Kilpinen W. (1847) *Neljä ensimmäistä Kirjaa ynnä viidennen määrytykset Euklideen alkeista mittaustieteessä*. Präntätty J. Simeliuksen perillisten tykönä. Helsinki.

- Kirk R. E. (2007) Effect Magnitude: A Different Focus. *Journal of Statistical Planning and Inference*, 137, 1634–1646.
- Koistinen P. (2013) Tilastotieteen ja todennäköisyyslaskennan englantilais-suomalainen sanasto. <http://www.math.helsinki.fi/petrin/sanasto/tilastosanasto.html> (haettu 28.2.2020).
- Kotimaisten kielten keskus (2019) Rapolan 1800-luvun sanasto. Sanalipuston pohjalta laadittu tietokanta. Kielipankki. https://sanat.csc.fi/wiki/Rapolan_1800-sanasto (haettu 1.6.2020).
- Lehmann E. L., Romano J. P. (2005) *Testing Statistical Hypotheses, 3. laitos*. Springer. New York.
- Lindeberg J. W. (1927) *Todennäköisyyslasku ja sen käytäntö tilastotieteessä. Alkeellinen esitys*. Otava. Helsinki.
- Lokki O. (1956) Tilastollisen laaduntarkkailun perusteet. Luentomoniste. Helsinki.
- Lokki O. (1959) Tilastomatematiikan perusteet: STS:n täydennyskoulutuskurssilla joulukuussa 1959 pidettyjen luentojen luentoaineisto. Suomen teknillinen seura.
- Luther G. (1993) *Suomen tilastotoimen historia vuoteen 1970*. Tilastokeskus. WSOY. Helsinki.
- Läärä E., Luostarinen T., Hakulinen T., Lyytikäinen O., Sarna S., Virtala A.-M., Riihimäki H., Hakama M. (2008) Epidemiology dictionary: English–Finnish. Suomen Epidemiologian Seura ja Duodecim. <https://www.finepi.fi/epidemiologian-sanasto-englanti-suomi> (haettu 28.2.2020).
- Lönnrot E. (1874) *Suomalais-Ruotsalainen Sanakirja. Edellinen Osa. A–R*. Suomalaisen Kirjallisuuden Seuran toimituksia, 50, Osa I. Suomalaisen Kirjallisuuden Seuran kirjapaino. Helsinki.
- Mattila K. (2017) Merkkipaaluja ja episodeja observaattori Henrik Walbeckin elämästä. *Tieteessä tapahtuu*, 35(6), 45–50.
- Metsikkö T. (1951) Kirjasuomen opetus- ja koulusanastoa – Agricolasta vuoden 1866 kansakouluasetukseen. Väitöskirja. Helsingin yliopisto. Suomalaisen Kirjallisuuden Seuran toimituksia, 242. Helsinki.
- Mohn H. (1892) *Ilmatieteen pääpiirteet*. Suomalaisen kirjallisuuden seuran toimituksia, 79. Suomalaisen Kirjallisuuden Seuran kirjapaino. Helsinki.
- Niitamo O. (1991) Tilastopäivien avajaispuhe 1990. Teoksessa I. Mellin (toim.) *Suomen tilastoseura 70 vuotta*. Helsinki.
- Nordberg L. (1999) Leo Törnvist – The “Grandfather” of Finnish Statistics. Teoksessa J. Alho (toim.) *Statistics, Registries, and Science. Experiences from Finland*. Statistics Finland. Helsinki.
- Oja, H. (1981) On Location, Scale, Skewness and Kurtosis of Univariate Distributions. *Scandinavian Journal of Statistics*, 8, 154–168.
- Pahkinen E. (2012) *Kyselytutkimuksen otantamenetelmät ja aineistoanalyysi*. Jyväskylän University Library Publishing Unit. Jyväskylä.
- Paulauskas V. (1999) J. W. Lindeberg and The Central Limit Theorem. Teoksessa J. Alho (toim.) *Statistics, Registries, and Science. Experiences from Finland*. Statistics Finland. Helsinki.
- Pearson K. (1905) ”Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson.” A Rejoinder. *Biometrika*, 4, 169–212.

- Pere P. (2019) Olemattoman taiteilijan eläväinen piirustus tutkijan ruumiista. *Tieteessä tapahtuu*, 37(6), 44–48.
- Pere P., Nyblom J. (2020) Oikaisuja suomalaisen tieteen historiaan. *Tieteessä tapahtuu*, 38(1), 42–46.
- Pesola V. (1924) *Kenttäkoeopas*. Werner Söderström. Porvoo.
- Pihkala E. (2000) Taloustieteet. Teoksessa Tommila P. (toim.) *Suomen tieteen historia 2. Humanistiset ja yhteiskuntatieteet*. WSOY. Helsinki.
- Pihkala K. U. (1941) Suomen kananmunahintojen muodostumisesta maailmansodan jälkeen. Suomen maataloustieteellisen seuran julkaisuja, 46. Karisto. Hämeenlinna.
- Pöyhönen P. (1955) Ekonometrinen tutkimus tonttien hinnoista. Valtion teknillinen tutkimuslaitos. Helsinki.
- R Core Team (2019) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Wien. <https://www.R-project.org/> (haettu 28.2.2020).
- Rapola M. (1943) Lisiä kirjasuomen sanaston historiaan. *Virittäjä*, 47, 15–30.
- Rapola M. (1960) *Sanojemme ensiesiintymiä*. Tietolipas 22. Suomalaisen kirjallisuuden seura. Helsinki.
- Salminen V. (1906) D. E. D. Europæus. *Suomi – Kirjoituksia isänmaallisista aineista*, Neljäs Jakso, 4 Osa, 1–232.
- Schildt W. (1842) Förslag till nya ordbildningar i Finskan. *Helsingfors' Morgonblad* 28.7.1842.
- Stigler S. M. (1986) *The History of Statistics. The Measurement of Uncertainty before 1900*. Harvard University Press. Cambridge, Massachusetts.
- Student (1908) The Probable Error of a Mean. *Biometrika*, VI, 1–25.
- Sullström R. (2000) Tilastotieteilijästä tuli ensimmäinen ekonometri. *Helsingin Sanomat* 21.3.2000.
- Suomen historiallinen seura (1883) *Biografinen nimikirja*. G. W. Edlund. Helsinki.
- Suomen Tilastoseura (1954) *Pohjoismainen tilastosanasto – Nordisk statistisk nomenklatur*. Suomen Tilastoseura. Kööpenhamina.
- Suomen Tilastoseura (1975) *Pohjoismainen tilastosanasto – Nordisk statistisk nomenklatur, toinen uusittu laitos*. Suomen Tilastoseuran julkaisuja 3. Gummerus. Jyväskylä.
- Teräsvirta T. (1987) How We Got the Data. Teoksessa T. Pukkila ja S. Puntanen (toim.) *Proceedings of the Second international Tampere Conference in Statistics*. Department of Mathematical Sciences/Statistics. University of Tampere.
- Tiit E.-M., Tooding L.-M. (2019) *Statistikaleksikon*. Tartu Ülikooli Kirjastus. Tartto.
- Tikkanen P. (1846) Suomen Maatiede. Teoksessa Tikkanen P. (toim.) *Lukemisia Suomen Kansan hyödyksi*, 2, 99–141. Helsingissä, J. Simeliuksen perillisten tykönä.
- Tikkanen P. (1847) Suomen Maatiede. Teoksessa Tikkanen P. (toim.) *Lukemisia Suomen Kansan hyödyksi*, 3, 65–156. Helsingissä, J. Simeliuksen perillisten tykönä.
- Tikkanen P. (1848) *Suomen Suuriruhtinanmaan Nykynen Tilasto: Yritös Alkeis- ja Rahwaan-kouluin tarpeeksi*. Helsingissä, J. Simeliuksen perillisten kirjapajassa.

- Tikkanen P. (1859) *Väki-luvun ja Asukas-vaiheiden suhteita Suomessa*. Väitöskirja. Keisarillinen Aleksanterin yliopisto. Frenckell. Helsinki.
- Tilastokeskus (2021) Suomen tilastojen historia -opas <https://guides.stat.fi/suomentilastojenhistoria> (haettu 25.7.2021). Jakso 1.3 Tilastoviraston synnyn juuret kansallisuusaatteessa.
- Tooding L.-M. (1996) *Statistikasõnastik*. Tartu Ülikooli Kirjastus. Tartto.
- Topelius Z. (1876) *Maamme Kirja, 1. painos*. Suomentanut Joh. Bäckwall. Edlund. Helsinki.
- Topelius Z. (1878) *Maamme Kirja, 2. painos*. Suomentanut Joh. Bäckwall. Suomentannosta muokannut P. Cajander. Edlund. Helsinki.
- Topelius Z. (1905) *Maamme Kirja, 17. painos*. Kymmenennen, tekijän kuoleman jälkeen korjatun, painoksen mukaan suomentanut P. Cajander. Edlund. Helsinki.
- Tudeer A. E. (1946) *Suomen Tilastoseura 1920–1945*. Suomen Tilastoseura. Helsinki.
- Törnqvist L. (1946) Tilastollisten tunnuslukujen laskukaavoja. Helsingin yliopiston monistustoimisto. Helsinki.
- Vahervuo T. (1946) *Johdatus matemaattis-tilastollisiin menettelytapoihin*. WSOY. Helsinki.
- Vallinkoski J. (1966) Turun Akatemian väitöskirjat 1642–1828. Helsingin yliopiston kirjaston julkaisija 30. Helsinki.
- Wænerberg G. M. (1856) Keisarillisen Majesteetin Armollinen Gymnaasi- ja Koulu-järjestys Suomen Isoruhtinanmaalle. Annettu, Helsinginkaupungissa, 7 p. Huhtikuuta 1856. Helsingissä, J. Simeliuksen Perillisten tykönä 1856.
- Westzyntius C. G. (1844) Keisarillisen Majesteetin Armollinen Gymnaasi- ja Koulu-järjestys Suomen Isoruhtinanmaalle. Annettu Helsinginkaupungissa, 6 p. Marraskuuta 1843. Helsingissä, J. Simeliuksen Perillisten tykönä 1844.
- Wasserstein R. L., Schirm A. L., Lazar N. A. (2019) Moving to a World beyond “ $p < 0.05$ ”. *The American Statistician*, 73 (Suppl 1), 1–19.

