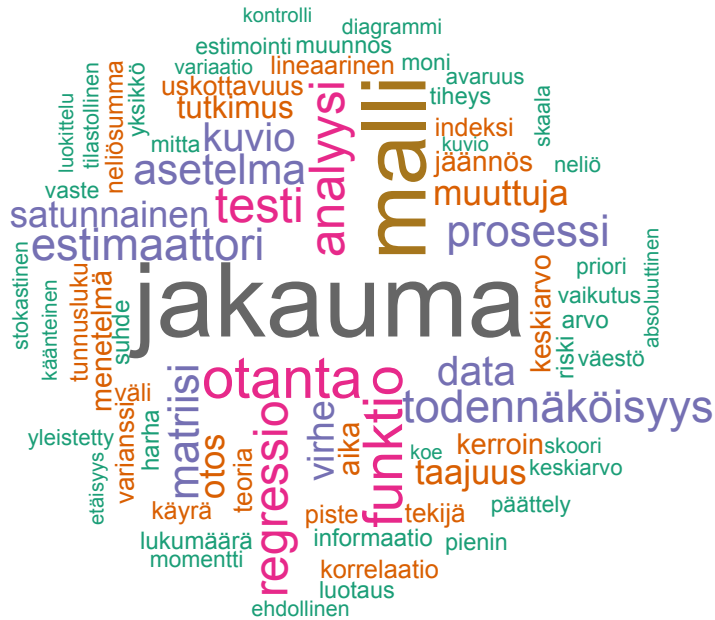


Tilastotieteen sanasto

Juha Alho, Elja Arjas, Juha Karvanen, Lasse Leskelä,
Esa Läärä ja Pekka Pere



Tilastotieteen sanasto

Juha Alho, Elja Arjas, Juha Karvanen, Lasse Leskelä,
Esa Läärä ja Pekka Pere

Suomen Tilastoseuran julkaisuja 8

3. laitos
Suomen Tilastoseura ry
Helsinki 2025

© Suomen Tilastoseura ry

3. laitos

ISSN 2954-2286 (verkkojulkaisu)

ISBN 978-952-9795-15-4 (pdf)

Kansien sanapilvet: Juha Alho 2020

Helsinki 2025

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	123
Lähteet	191

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. Vastanotamme 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

Alkusanat 3. laitokseen

Käyttäjät ovat löytäneet *Tilastotieteen sanaston*. Sanaston verkkosivu kirjaa suuria asioijamääriä, ja sanastosta haetaan suomennoksia aktiivisesti.

Verkkosivun Ehdota termiä -linkkiin on jätetty kahdessa vuodessa 4/2023–5/2025 noin seitsemänkymmentä ehdotusta uusiksi termeiksi. Kiitämme kaikista. On tähdellistä, että sanaston päivitykset eivät rajoitu allekirjoittaneiden erikoisaloille. Olemme pohtineet kaikille ehdotetuille termeille suomennoksia, ja useimmat ovat sanaston uudessa 3. laitoksessa. Jotkin suomennostoiveet askarruttivat ja työllistivät meitä suuresti.

Esimerkki on *non-inferiority trial*, jonka suomensimme **samanveroisuusko-**keeksi. Termin suomentamisen vaikeutta kuvanee, että raakalaina *inferiorinen* on vakiintunut taloustieteessä. Emme ole sisällyttäneet ehdotusta sanastoon, jos emme ole keksineet sopivaa suomennosta tai mielestämme termi ei ole vakiintunut tai tilastotieteellinen. Jotkut ovat ehdottaneet itse suomennoksia. Olemme ottaneet niitä sanastoon puntarointimme jälkeen sellaisinaan tai muokkaaminamme. Toivomme, että suomennosehdotusten laatijat ymmärtävät päätöksiämme.

Kiitämme Hannu Ojaa ja Klaus Nordhausenia monien suomentamista kaipaavien parametrittomien menetelmien termien osoittamisesta. Olemme myös itse keränneet suomennettavia termejä lisää. Tulos on, että englant-suomi-sanasto koostuu jo noin 3 700 termistä.

Olemme viilanneet muutamaa aiempaa suomennosta ja korjanneet painovirheitä. Esimerkiksi *signed rank test* suomentui edellisessä laitoksessa merkkijärjestystestiksi. Nyt ehdotamme termin ykkössuomennokseksi merkisijatestiä. Olemme noudattaneet periaatetta, että sanaston termit ovat ensisijaisesti substantiiveja. Päätimme kuitenkin, että *sample* on niin tärkeä verbi, että lisäsimme sanastoon rivit *sample*₃ poimia ja *sample*₄ näytteistää.

Olemme kiitollisia ja kannustettuja kaikesta saamastamme palautteesta. Jatkossa pyrimme päivittämään hakukonetta aiempaa tiheämmin ja ajoittain viemään päivitykset myös sanaston pdf-versioon. Hakukone ja päivitykset eivät tule siten vastedes olemaan aina täysin yhtäpitäviä.

Eino Leinon, runon ja suven päivänä 6.7.2025,
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 haittaavat 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* (Gausin 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 uudissan sanan englannin osalta 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 tiedonhakupäätelmästä 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 kirjoitusasun 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 suomenkielisessä 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* käännämme 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*”. Eri-tyisesti 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 mielekästä 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 määrättyä 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 standardinormaali jakauman 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 (Westzynthius 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 määrätty 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 [†]yhtisoutemme $x = 56.311 - 0.0851(v - 1800)$ 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 luentomonisteessaan 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 esitetyn 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 ”luvusto” (vrt. suomennos [†]*luwunlasku* alaluvussa 3.1). Tämä sana samoin kuin [†]*arwustikuteadus* eli ”luvustotiede” 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>hålbimus</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 neighbors 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 viroonnettu 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 viroonnettu 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)

ABC₂ (approximate bootstrap confidence interval)

absolute deviation poikkeaman itseisarvo, itseispoikkeama

absolute difference erotuksen itseisarvo, itseiserotus

absolute error virheen itseisarvo, itseisvirhe

absolute frequency lukumäärä, absoluuttinen frekvenssi

absolute moment itseismomentti

absolute risk absoluuttinen riski

absolute scale absoluuttinen asteikko

absolute value itseisarvo

absolutely continuous absoluutisesti jatkuva

absolutely convergent itseisesti suppeneva

absolutely summable itseisesti summautuva

absorbing barrier absorboiva reuna, nielevä reuna

absorbing region absorboiva alue, nielevä alue

absorbing state absorboiva tila, nielevä tila

absorption distribution imeytymisjakauma, absorptiojakauma

A/B-test A/B-testi

accelerated failure time model (AFT-model) nopeutetun vikaantumisajan malli, AFT-malli

accelerated life testing nopeutettu elinaikatestaus

acceptable quality limit hyväksyttävä laatutaso

acceptance boundary (acceptance line) hyväksymisraja

acceptance control chart hyväksymisvalvontakaavio

acceptance error (false acceptance) hyväksymisvirhe

acceptance line (acceptance boundary) hyväksymisraja

acceptance quality level (acceptable quality level) hyväksyttävä laatutaso

acceptance region hyväksymisalue

acceptance sampling hyväksymisotanta

accrual rate kertymisnopeus

accuracy tarkkuus₁

ACE₁ (alternating conditional expectations regression)

ACE₂ (average causal effect)

ACF (autocorrelation function)

acquiescence bias myöntöväisyys-harha

activation function aktivointifunktio

active learning aktiivinen oppiminen

actuarial estimator aktuaariestimattori

actuarial statistics vakuutustilastotiede

AdaBoost (adaptive boosting)

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, osittaisregressiokuvaio	affine equivariant affiinisti ekvivalentti
addition rule yhteenlaskusääntö	affine invariant affiinisti invariantti
additive hazards model additiivisten uhkien malli, summautuvien uhkien malli	affine transformation affiini muunnos
additive model additiivinen malli, summautuva malli	AFT (accelerated failure time)
additive process additiivinen prosessi, summautuva prosessi	age-adjusted ikävakioitu
additive seasonal model additiivinen kausivaihtelumalli, summautuva kausivaihtelumalli	age-dependent process ikäriippuva prosessi
adequate model kelpoinen malli	age heaping ikäsuosio, ikäkasautuminen
adherence to treatment (compliance with treatment) hoitositoutuneisuus, hoitomyöntyvyys	agent-based model agenttipohjainen malli
adjacency matrix vierusmatriisi	age-period-cohort model ikä-periodi-kohorttimalli
adjacent categories model vierusluokkamalli	age pyramid ikäpyramidi
adjugate matrix liittomatriisi	age-specific ikäryhmittäinen, i'ittäinen
adjusted estimator vakioitu estimaattori, rukattu estimaattori	age-specific death rate (age-specific mortality rate) ikäryhmittäinen kuolevuus, i'ittäinen kuolevuus
adjusted profile likelihood muokattu profiiliuskottavuus, rukattu profiiliuskottavuus	age-specific mortality rate (age-specific death rate) ikäryhmittäinen kuolevuus, i'ittäinen kuolevuus
adjusted R^2 muokattu R^2 , rukattu R^2	agglomerative hierarchical clustering kokoava hierarkkinen ryhmittely
adjusted sample weight muokattu otospaino, rukattu otospaino	aggregate data aggregoitu aineisto, summattu aineisto
adjustment vakiointi, muokkaus, rukkaus	aggregate imputation ryhmäpaikkaus, aggregaatti-imputointi
admissible käypä, hyväksyttävä	aggregate-level study ryhmätason tutkimus
admissible decision function käypä päätösfunktio	aggregation aggregointi
admissible estimator käypä esti-	AHP (analytic hierarchy process)
	AI (artificial intelligence)
	AIC (Akaike information crite-

algorithm algoritmi aliasing₁ täydellinen kollineaari- suus aliasing₂ laskostuminen allele frequency alleelifrekvenssi allocation₁ kiintiöinti allocation₂ kohdentaminen allocation concealment ryhmä- jaon salaaminen allometry allometria almost everywhere (a.e.) mel- kein kaikkialla, m.k. almost surely (a.s.) melkein var- masti, m.v. alternating conditional expec- tations regression (ACE₁) vuo- rottelevien ehdollistusten regres- sio alternative hypothesis vastahypoteesi, vaihtoehtoinen hypoteesi American option amerikkalainen optio amplitude amplitudi, värähdys- laajuus analysis of covariance (ANCOVA) kovarianssianalyysi analysis of deviance devianssia- nalyysi analysis of dispersion hajonta- analyysi analysis of variance (ANOVA) varianssianalyysi analytic hierarchy process (AHP) analyyttinen hierarkia- prosessi analytics analytiikka	ancillary information apuinformaatio ancillary statistic aputunnuslu- ku ANCOVA (analysis of cova- riance) angular transformation kulma- muunnos anisotropic anisotrooppinen annealed distribution hehkutet- tu jakauma annual data vuosiaineisto annuity₁ elinkorko, annuiteetti annuity₂ tasaerä, annuiteetti anomaly anomalia ANOVA (analysis of variance) antecedent of a rule säännön edeltäjäosa antedependence riippuvuus edel- tävästä anticipative control ennakoiva säätö antithetic vastakkaismerkkinen, antiteettinen antithetic variate vastamuuttuja antitonic regression antitoninen regressio aperiodic jaksoton applied statistics sovellettu tilas- totiede approximate likiarvoinen, liki- määräinen approximate Bayesian com- putation (ABC₁) likimääräinen Bayes-laskenta approximate bootstrap confi- dence interval (ABC₂) likimää- räinen uusioluottamusväli approximation₁ likiarvoistus, approksimointi approximation₂ likiarvo, approk- simaatio arbitrage arbitraasi
---	---

ARCH model (autoregressive conditional heteroskedastic model)	ascertainment ₁ jäljitys
ARCH-malli	ascertainment ₂ varmistus
arcsine distribution arkussinijakauma	ascertainment bias jäljitysharha
arcsine transformation arkussinimuunnos	assay määrittäminen, määrittäminen
ARE (asymptotic relative efficiency)	asset arvopaperi
area under curve (AUC) käyränalainen pinta-ala, käyräanalusa	association yhteys, assosiaatio
areal data alueaineisto	association analysis yhteysanalyysi, assosiaatioanalyysi
area sampling alueotanta	astrostatistics tähtitilastotiede
area under curve (AUC) käyränalainen pinta-ala, käyräanalusa	asymptotic asymptootinen
ARFIMA model (autoregressive fractionally integrated moving average model) ARFIMA-malli	asymptotic distribution (limiting distribution) rajajakautuma
ARIMA model (autoregressive integrated moving average model) ARIMA-malli	asymptotic efficiency asymptootinen tehokkuus
arithmetic average (arithmetic mean) aritmeettinen keskiarvo, keskiarvo	asymptotic relative efficiency (ARE) asymptootinen suhteellinen tehokkuus
arithmetic distribution aritmeettinen jakauma	asymptotics asymptotiikka
arithmetic mean (arithmetic average) aritmeettinen keskiarvo, keskiarvo	ATE (average treatment effect)
Arjas plot Arjaksen kuvio, Arjas-kuvio	atom atomi
ARMA model (autoregressive moving average malli) ARMA-malli	ATT (average treatment effect on the treated)
AR model (autoregressive model) AR-malli	attack rate sairastuneiden osuus
array (table) taulukko	attainable portfolio saavutettava salkku
arrival process saapumisprosessi	attenuation vaimeneminen
artificial intelligence (AI) tekoäly, keinoäly	attractor attraktori
a.s. (almost sure) m.v.	attributable fraction syyksilukuosuus
as-randomized analysis satunnaistuksen mukainen analyysi	attribute ominaisuus, piirre
	attrition harventuminen, kato
	ATU (average treatment effect on the untreated)
	AUC (area under curve)
	augmented Dickey-Fuller test täydennetty Dickeyn–Fullerin testi
	autocorrelation autokorrelaatio
	autocorrelation function (ACF) autokorrelaatiofunktio
	autocovariance autokovarianssi

autocovariance function auto-
kovarianssifunktio

**autocovariance generating
function** autokovarianssiemä-
funktio, autokovarianssit generoi-
va funktio

**automatic interaction detec-
tion** (AID) automaattinen inte-
raktiohaku, automaattinen yhdys-
vaikutushaku, AID

autoregression autoregressio

**autoregressive conditional het-
eroskedastic model** (ARCH-
model) autoregressiivinen ehdol-
lisen heteroskedastisuuden malli,
ARCH-malli

**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 kausaalivaikutus

average run length keskimää-
räinen merkkijakson pituus

average treatment effect (ATE)
keskimääräinen käsittelyvaikutus

**average treatment effect on
the treated** (ATT) keskimää-
räinen 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)
laukut

bag-of-words model sanapussi-malli	battery of tests testipatteri
bagplot (starburst plot) tähtisade-kuvio	Bayes factor Bayes-kerroin, Bayes-tekijä
balance equation tasapainoyhtälö	Bayes' formula Bayesin kaava
balanced incomplete block design (BIBD) tasapainoinen vajaa-lohkoasetelma	Bayesian bayesiläinen, bayesläinen
balanced sampling tasapainoinen otanta	Bayesian inference Bayes-päätely
BAN (best asymptotically normal estimator)	Bayesian information criterion (BIC) Bayes-informaatiokriteeri, BIC
bandwidth₁ kaistanleveys, taa-juusalue	Bayesian nonparametrics parametrittomat Bayes-menetelmät
bandwidth₂ (window width) ikkunanleveys	Bayesian updating Bayes-päivitys
bar chart₁ (bar plot) pylväskuvio, pylväskaavio ¹	Bayes network Bayes-verkko
bar chart₂ (bar plot) palkkikuvio, palkkikaavio ²	Bayes risk Bayes-riski
baseline lähtötaso, perustaso	BCa (bias-corrected and accelerated interval)
baseline characteristics lähtöarvot	belief function uskomusfunktio
baseline hazard perusuhka	belief network uskomusverkko
base period (reference period) perusjakso, viitejakso	Bernoulli distribution Bernoulli-jakauma
base rate bias (base rate fallacy) perustasoharha	Bernoulli experiment (Bernoulli trial) Bernoulli-koe
base rate fallacy (base rate bias) perustasoharha, perustasovirhe-päätelmä	Bernoulli trial (Bernoulli experiment) Bernoulli-koe
basic reproduction number (R_0) perustartuttavuusluku, perusuusiutumisluku, R_0	best asymptotically normal estimator (BAN) paras asymptootisesti normaalin estimaattori
basic sample weight perusotospaino	best linear unbiased estimator (BLUE) paras lineaarinen harhaton estimaattori
basis kanta	best linear unbiased prediction (BLUP) paras lineaarinen harhaton ennustin
basket trial korikoe	best practice life expectancy parhaan olosuhteen elinajanodote
batch variation erävaihtelu	beta-binomial distribution beta-binomijakauma
bathtub curve ammekäyrä	beta distribution betajakauma

¹Pylväskuviossa pylväät ovat pystysuorassa; palkkikuviossa vaakasuorassa.

²Pylväskuviossa pylväät ovat pystysuorassa; palkkikuviossa vaakasuorassa.

between-groups estimator	ryhmien välinen estimaattori	binomial test	binomitesti
between-groups sum of squares	ryhmien välinen neliösumma	binomial trial	(binomial experiment) binomikoe
between-groups variance	ryhmien välinen varianssi	bioassay	biomääritys
bias (systematic error)	harha, systemaattinen virhe	bioequivalence test	bioekvivalenssitesti
bias-corrected and accelerated interval (BCa)	nopeutettu harhakorjattu luottamusväli	bioinformatics	bioinformatiikka
biased estimator	harhainen estimaattori	biometric identification	(biometrics) biometrinen tunnistaminen, biometriikka
biased sampling	harhainen otanta	biometrics₁	(biometry) biometria
biased test	harhainen testi	biometrics₂	(biometric identification) biometrinen tunnistaminen, biometriikka
bias-variance tradeoff	harha-variانسsikompromissi	biometry	(biometrics ₁) biometria
BIBD (balanced incomplete block design)		biostatistics	biotilastotiede, biostatistiikka
BIC (Bayesian information criterion)		bipartite	kaksijakoinen
bidegree distribution	kaksiastejakauma	biplot	bikuvio, yksilö-muuttujakuvi
big data	suuraineisto, suurdata, massadata	birth cohort	syntymäkohortti
bilinear model	bilineaarinen malli	birth-death process	syntymä-kuolemaprosessi
bimodal distribution	kaksihuipunen jakauma	birth process	syntymäprosessi
bin	luokkaväli	birth rate	syntyvyys
binary variable (dichotomous variable)	kaksiarvoinen muuttuja, binäärimuuttuja, binaarimuuttuja, dikotominen muuttuja	biserial correlation	biseriaalinen korrelaatio
binned scatter plot	siilosirontakuvi	bispectrum	bispektri
binomial coefficient	binomikerroin	bit	bitti
binomial distribution	binomijakauma	bivariate	kaksimuuttujainen, kahden muuttujan, kaksiulotteinen
binomial experiment	(binomial trial) binomikoe	black box	musta laatikko
		Black–Scholes model	Blackin–Scholesin malli
		blinded experiment	sokkokoe
		blinding	sokkoutus
		blind source separation	(blind signal separation) lähteiden sokkoerotelu, signaalien sokkoerotelu
		blind signal separation	(blind source separation) signaalien sokkoerotelu, lähteiden sokkoerotelu

lu
block lohko
block diagonal matrix lohkolä-
vistäjämatriisi, lohkodiagonaali-
matriisi
block diagram lohkoakaavio
block randomization lohkosä-
tunnaistaminen
BLUE (best linear unbiased esti-
mator)
BLUP (best linear unbiased pre-
dictor)
BMI (body mass index)
body mass index (BMI) painoin-
deksi
body of a rule säännön vartalo
bond velkakirja
Bonferroni inequality Bonferro-
nin epäyhtälö
boosting tehostus
bootstrap aggregating (bagging)
uusioaggregointi, laukutus
bootstrap method uusiomenetel-
mä, uusio-otantamenetelmä
bootstrap confidence interval
uusioluottamusväli
bootstrap estimator uusioesti-
maattori
bootstrapping (bootstrap sam-
pling) uusio-otanta
bootstrap p -value uusio- p -arvo
bootstrap sampling
(bootstrapping) uusio-otanta
bootstrap testing uusiotestaus
borrowing of strength voiman
lainaaminen
bounded in probability stokas-
tisesti rajoitettu
bounded variation function ra-
joitetusti heilahteleva funktio
box-and-whisker plot (box plot)
laatikkokuvio
box plot laatikkokuvio

branching process haarautumis-
prosessi
breakdown point murtumispiste
Brownian bridge Brownin silta
Brownian motion Brownin liike
bubble chart (bubble plot) kupla-
kuvio
bubble plot (bubble chart) kupla-
kuvio
burn-in period (warm-up period)
sisäänaajojakso₁
business cycle suhdanne
busy period kiirejakso

C

càdlàg (continue à droite, limit à
gauche) cadlag
calculus of probability
(probability calculus) todennä-
köisyyslaskenta
calibrated weight kalibroitu pai-
no
calibration kalibointi
caliper matching välikaltaistus
call-back toistokutsu, toistokäyn-
ti, karhunta
call option osto-oikeus, osto-optio
candidate distribution
(proposal distribution) ehdokas-
jakauma, ehdotusjakauma
canonical analysis kanoninen
analyysi
canonical correlation kanoninen
korrelaatio
canonical link function kanoni-
nen 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

CASI (computer-assisted self-interview)

catastrophe theory katastrofiteoria

CATE₁ (complier average treatment effect)

CATE₂ (conditional average treatment effect)

categorical distribution luokkajakauma

categorical variable luokkamuuttuja, luokitteluasteikon muuttuja

caterpillar plot katepillarikuvio

CATI (computer-assisted telephone interview)

cauchit transformation cauchitmuunnos

Cauchy distribution Cauchy-jakauma

causal chain syyketju, kausaaliketju

causal diagram (causal graph) kausaalikaavio, kausaaliverkko

causal discovery kausaalirakenteen löytäminen

causal graph (causal diagram) kausaalikaavio, kausaaliverkko

causal inference syy-seurauspäättely, kausaalipäättely

causality syy-seuraussuhde, kausaalisuus

causal loop diagram kausaalisilmukkakaavio

causal model kausaalimalli, syy-seurausmalli

cause variable syymuuttuja

cause-specific hazard (cause-specific hazard rate) syykohtainen uhka

CAWI (computer-assisted web-interview)

CDE (controlled direct effect)

CDF (cumulative distribution function)

ceiling effect kattovaikutus

cell frequency solulukumäärä, solufrekvenssi

censored observation sensuroitu havainto

censoring sensurointi

census₁ (complete enumeration) kokonaistutkimus, kokonaislaskenta

census₂ väestölaskenta

census survival rate laskentojen välinen elossaolosuhde

centering keskistäminen	chi-squared statistic khi-neliösuure, khi-neliötunnusluku
centile sentiili	chi-squared test khi-neliötesti
central limit theorem keskeinen raja-arvolause	chi-statistic khi-suure, khi-tunnusluku
central moment keskusmomentti	choice-based sampling valinta-perusteinen otanta
centroid method sentroidimenetelmä	CIF (cumulative incidence function)
chain ketju	circular data kehäaineisto, kehädata
chain binomial model ketjubinomimalli	circular distribution kehäjakautuma
chain graph ketjukaavio	city block distance (Manhattan distance) korttelietäisyys, manhattanetäisyys
chain index ketjuindeksi	class frequency luokkafrekvenssi
chance variation (random variation) satunnaisvaihtelu	classification luokittelu, luokitus
changeover design (cross-over design) vaihtovuoroasetelma, riskikkäisasetelma	classification and regression tree (CART) luokittelu- ja regressiopuu
change point muutospiste	classification attribute luokitteluattribuutti, luokittelumääre
change score muutosarvo	classification error luokitteluvirhe
characteristic equation ominaisarvoyhtälö, karakteristinen yhtälö	classification rule luokittelusääntö
characteristic function ³ karakteristinen funktio	classification table luokittelutaulukko
characteristic value (eigenvalue) ominaisarvo	classification statistic luokittelutunnusluku
Chebyshev's inequality ⁴ Tshebyshevin epäyhtälö	classifier luokitin, luokittelija
chemometrics kemometria	class interval luokkaväli
Chernoff's faces Chernoffin kasvokuviot	class midpoint luokkakeskus
chi distribution khi-jakauma	clinical trial kliininen koe
Chinese restaurant process kiinalainen ravintolaprosessi	cliometrics kleiometria
chi-plot khi-kuvio	cloglog-transformation (complementary loglog transformation)
chi-squared distance khi-neliöetäisyys	
chi-squared distribution khi-neliöjakauma	

³Todennäköisyyslaskennassa karakteristinen funktio on satunnaismuuttujan X jakauman määrittävä kompleksiarvoinen funktio $\phi(t) = E[e^{itX}]$, mutta muualla matematiikassa termillä tarkoitetaan jonkin joukon A indikaattorifunktiota.

⁴Nimestä on vaihtelevia kirjoitusasuja niin englannin- kuin suomenkielissäkin teksteissä.

mation) cloglog-muunnos	nation yhteisselityssaste, yhteisde- terminaatiokerroin
closed multiple testing suljettu monitetaus	coefficient of variation variaatiokerroin
closed population suljettu väes- tö	cofactor kofaktori
closure sulkeuma	cognitive bias kognitiivinen har- ha
cluster ryväs ⁵	coherence johdonmukaisuus, ko- herenssi
cluster analysis ryhmittelyana- lyysi, ryvästämisanalyysi	cohort kohortti
cluster-dependent distribution ryväskohtainen jakauma	cohort study kohorttitutkimus
clustered data ryvästynyt aineis- to	cointegration yhteisintegroituvuus
clustering ryvästäminen	cold deck imputation kylmä- paikkaus, kylmäimputointi
cluster process ryväsprosessi	collapsibility koontuvuus, yhdis- tettävyys
cluster randomization ryvässa- tunnaistus	collider törmäytin
cluster sampling ryvästötanta	collider bias törmäytinharha
coalescence sulautuminen, koale- senssi	collinearity kollineaarisuus
coarse data karkea data	COMFAC model (common fac- tor model ₂)
coarsening karkeistus	commonality kommunaliteetti
COD (coefficient of divergence)	common factor yhteisfaktori
CoDA (compositional data analy- sis)	common factor model₁ yhteisfaktori malli
coefficient kerroin	common factor model₂ (COMFAC model) yhteisten te- kijöiden malli
coefficient of association assosi- aatiokerroin, yhteyskerroin	communicating class yhteyshuokka
coefficient of concentration kes- kittymiskerroin	community detection yhteisötunnistaminen
coefficient of concordance yhdenmukaisuuskerroin, konkor- danssikerroin	companion form kumppanimuo- to
coefficient of determination (R ²) selityssaste, determinaatioker- roin	comparative experiment (controlled experiment) vertaile- va koe, vertailukoe
coefficient of disarray (Kendall's tau) Kendallin tau	comparative parameter vertailuparametri
coefficient of divergence (COD) poikkeamiskerroin	comparison group (control
coefficient of multiple determi-	

⁵Taivutetaan "rypään".

group) vertailuryhmä	nent-plus-residual plot (partial residual plot) osittaisjäännöskuvio
compartment model lokeromalli	composite hypothesis yhdistetty hypoteesi
compatibility yhteensopivuus	compositional data analysis (CoDA) rakenneosajäännöskuvio
compatibility interval (confidence interval) yhteensopivuusväli, luottamusväli	compound interest korkoa korkolle
competing event kilpaileva tapahtuma	compound Poisson distribution yhdistetty Poisson-jakauma
competing risk kilpaileva riski	compound Poisson process yhdistetty Poisson-prosessi
complementary error function (erfc) komplementtinen virhefunktio	computationally intensive method laskentaintensiivinen menetelmä
complementary loglog transformation (cloglog-transformation) komplementtinen loglog-muunnos, cloglog-muunnos	computational statistics laskennallinen tilastotiede
complementary probability vastatodennäköisyys, vastatapahtuman todennäköisyys	computer-assisted personal interview (CAPI) tietokoneavustettu henkilökohtainen haastattelu
complement event vastatapahtuma, komplementti-tapahtuma	computer-assisted self-interview (CASI) tietokoneavusteinen itsehaastattelu
complete case analysis kokonaisten havaintojen analyysi	computer-assisted telephone interview (CATI) tietokoneavusteinen puhelinhaastattelu
complete enumeration (census ₁) kokonaistutkimus, kokonaislaskenta	computer-assisted web-interview (CAWI) tietokoneavustettu verkkohaastattelu
completely randomized design täyssatunnaistettu asetelmä	computer vision konenäkö
complete separation täydellinen erottelu	concavity konkaavisuus, kupeisuus
complete statistic erotteleva tunnusluku	concentration curve keskittyneisyyskäyrä
compliance with treatment (adherence to treatment) hoito-myöntyvyys, hoitositoutuneisuus	concentration matrix (precision matrix) keskittyneisyysmatriisi
complier average causal effect (CACE ₁) keskimääräinen kausaali-vaikutus sitoutuneilla	concentration parameter (precision parameter) keskittyneisyysparametri
complier average treatment effect (CATE ₁) keskimääräinen käsittelyvaikutus sitoutuneilla	concordance yhdenmukaisuus, konkordanssi
component and compo-	concordance index

yhteensopivuusindeksi	confidence limit luottamusraja
concordant pairs konkordantit	confidence region luottamusalue
parit, samanlaiset parit	confidence set (confidence re- gion) luottamusjoukko, luotta- musalue
concurrent study tosiaikainen tutkimus	configuration model konfiguraa- tiomalli
conditional association ehdolli- nen yhteys	confirmatory data analysis vahvistava data-analyysi
conditional autoregression (CAR) ehdollinen autoregressio	confounder (confounding vari- able) sekoittaja, sekoittava muut- tuja
conditional average causal ef- fect (CACE₂) ehdollinen keski- määräinen kausaalivaikutus	confounding factor (confounder) sekoittaja, sekoittava tekijä
conditional average treatment effect (CATE₂) ehdollinen keski- määräinen käsittelyvaikutus	confounding variable (confounder) sekoittaja, sekoittava muuttuja
conditional density function ehdollinen tiheysfunktio	confusion matrix ⁶ luokitteluvir- hetaulukko
conditional dependence ehdolli- nen riippuvuus	conjugate distribution liittoja- kauma
conditional distribution ehdolli- nen jakauma	conjugate prior distribution liittopriorijakauma
conditional expectation ehdolli- nen odotusarvo	consequent of a rule säännön seuraajaosa
conditional independence eh- dollinen riippumattomuus	conservative test konservatiivi- nen testi
conditional likelihood ehdolli- nen uskottavuus	consistency₁ tarkentuvuus
conditional probability ehdolli- nen todennäköisyys	consistency₂ yhdenmukaisuus
conditional variance ehdollinen varianssi	consistency of classification luokittelun yhdenmukaisuus
conditioning plot (coplot) ehdol- listettu sirontakuvio	consistent estimator tarkentuva estimaattori
confidence band (confidence en- velope) luottamusvyöhyke	content analysis sisällönanalyysi
confidence envelope (confidence band) luottamusvyöhyke	contingency coefficient kontin- genssikerroin
confidence interval (compatibility interval) luotta- musväli, yhteensopivuusväli	contingency table ristitaulukko, kontingenssitaulukko
confidence level luottamustaso	continuation ratio model jatko- suhdemalli
	continue à droite, limite à

⁶Tätä termiä ja suomennosta “sekaannusmatriisi” käytetään erityisesti koneop-
pimisen kirjallisuudessa.

gauche ⁷ (càdlàg) oikealle jatkuva vasemmalta raja-arvot, cadlag	convolution konvoluutio
continuity jatkuvuus	co-occurrence yhteisesiintyvyys
continuity correction jatkuvuuskorjaus	Cook's D (Cook's distance) Cookin etäisyys
continuous distribution jatkuva jakauma	Cook's distance (Cook's D) Cookin etäisyys
continuous-time process jatkuva-aikainen prosessi	coplot (conditioning plot) ehdollistettu sirontakuvio
contour line tasa-arvokäyrä	copula kopula
contrast kontrasti	correlated korreloiva
control verrokki	correlation korrelaatio
control arm vertailulinja, verrokkininja	correlation coefficient korrelaatiokerroin
control chart valvontakortti	correlation matrix korrelaatiomatriisi
control group (comparison group) vertailuryhmä, verrokki-ryhmä	correlogram korrelogrammi
controlled direct effect (CDE) kontrolloitu suora vaikutus	correspondence analysis (dual scaling) korrespondenssianalyysi
controlled trial (comparative trial) vertaileva koe, vertailukoe	cosine distribution (cardioid distribution) kosinijakauma
control theory säätöteoria	cost-benefit analysis kustannus-hyötyanalyysi
control variable ₁ säätömuuttuja	cost-effectiveness kustannus-vai-kuttavuus
control variable ₂ vakioiva muuttuja	cost-efficiency kustannus-tehokkuus
convenience sample mukavuusotos	count (frequency ₁) lukumäärä, frekvenssi
convergence suppeneminen, konvergenssi	countable numeroituva
convergence in distribution jakaumasuppeneminen, jakaumakonvergenssi	count data (frequency data) lukumääräaineisto
convergence in mean square keskineliösuppeneminen	counterfactual kontrafaktuaalinen
convergence in probability stokastinen suppeneminen, stokastinen konvergenssi	counter-matching vastakaltaistus
convex hull konvekksi peite	counting measure laskurimitta
convexity konveksisuus, kuperaus	counting process laskuri-prosessi
	count time series lukumääräaika-sarja
	coupling kytkentä

⁷Tämän ranskankielisen termin vastine englannissa on *right continuous with left limits*. Silti ranskankielistä termiä ja sen lyhennettä *càdlàg* käytetään paljon englanninkielisessä kirjallisuudessa.

coupling of processes	prosessien kytkentä	cross spectrum	ristispektri
covariance	kovarianssi	cross table	(contingency table) ristitaulukko, kontingenssitaulukko
covariance analysis	(ANCOVA) kovarianssianalyysi	cross tabulation	ristiintaulukointi
covariance matrix	kovarianssimatriisi	cross-validation	ristiinvalidointi, ristinvahvistus
covariance stationary	(weakly stationary) kovarianssistationaarinen	crude estimator	karkea estimaattori
covariate	kovariaatti	cubic lattice	kuutiohila
coverage probability	peittodennäköisyys	cubic spline	kuutiosplini
Cox model	(proportional hazards model) Coxin malli, verrannollisten uhkien malli	cumulant	kumulantti
Cramér–Rao lower bound	Cramérin–Raon alaraja	cumulant generating function	kumulanttiemäfunktio, kumulantit generoiva funktio
credibility theory	luottoteoria	cumulative cost	kustannuskertymä
credible interval ⁸	(posterior interval) luottoväli, posterioriväli	cumulative distribution function	(CDF) kertymäfunktio
credible region	luottoalue	cumulative error	virhekertymä, kumulatiivinen virhe
credit	luotto	cumulative frequency	frekvenssikertymä, kumulatiivinen frekvenssi
critical region	(rejection region) hylkäysalue, kriittinen alue	cumulative hazard	(integrated hazard) uhkakertymä
critical value	hylkäysraja, kriittinen arvo	cumulative incidence	ilmaantumisosuus, kumulatiivinen ilmaantuvuus
Cronbach’s alpha	Cronbachin alfa	cumulative incidence function	(CIF) kumulatiivinen ilmaantuvuusfunktio
cross-correlation	ristikorrrelaatio	cumulative intensity	(integrated intensity) intensiteettikertymä
cross-covariance	ristikovarianssi	cumulative model	kertymämalli, kumulatiivinen malli
cross-fitting	ristiinsovitus	cumulative odds ratio	kertymäristisuhde
cross-over design	(changeover design) vaihtovuoroasetelma, ristikkäisasetelma	cumulative risk	riskikertymä
cross-product ratio	(odds ratio) ristisuhde, vetosuhde, ristitulosuhde		
cross-sectional study	poikittais-tutkimus		
cross-spectral density	ristispekt-		

⁸Kirjoitusasua *credibility interval* käytetään myös.

cumulative sum kertymäsumma, kumulatiivinen summa
cumulative sum chart (CUSUM chart) summakaavio, CUSUM-kaavio
curse of dimensionality dimensioikiros
curtailed inspection keskeytetty tarkastus
curtate kokonaisosa
curved exponential family kaareva eksponenttiperhe
curve fitting käyrän sovitus
curvilinear käyräviivainen
CUSUM chart (cumulative sum chart) summakaavio, CUSUM-kaavio
cut-off point kynnyksarvo, 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
data⁹ 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 aineiston syvyys, havainnon sisäisyys
data dredging (fishing expedition) datan ruoppaus, kalastelu
data driven datavetoinen, aineistovetoinen
data envelopment analysis (DEA) verhoikäyräanalyysi, DEA
data frame datakehikko
data generating process havainnot tuottava prosessi
data linkage datan linkittäminen
data matrix havaintomatriisi, datamatriisi
data mining datanlouhinta, tiedonlouhinta
data peeking datan kurkistelu, aineiston kurkistelu
data pipeline dataputki
data poisoning datan myrkyttäminen
data preprocessing aineiston esikäsittely, datan esikäsittely
data reduction datan tiivistäminen
data science datatiede
data scientist datatieteilijä
data screening aineiston esitarkastus
data set datajoukko, tiedosto
data wrangling aineiston jalostaminen, datan jalostaminen
DEA (data envelopment analysis)
death process kuolemaprosessi

⁹Sana on monikollinen. Huolessissa englannissa kirjoitetaan “*the data are*” eikä “*the data is*”. Yksikkömuotoa *datum* ei juuri koskaan käytetä.

¹⁰Oppikirjojen mukaan *desiilit* ovat *fraktiileja*, jotka jakavat aineiston kymmeneen yhtä suureen osajoukkoon. Joissakin yhteyksissä, esim. tulojakaumia kuvattaessa, *desiileillä* viitataan näihin osajoukkoihin eli kymmenyksiin. Vastaavaa rinnakkaistulkinta

death rate (mortality rate) kuolevuus	density histogram tiheyshistogrammi
decile desiili, kymmenys ¹⁰	density ratio model tiheyssuhdemalli
decision function päätössääntö, päätösfunktio	density sampling tiheysotanta, tiheyspoiminta
decision space päätösavaruus	denumerable numeroituvasti ääretön
decision theory päätösteoria	dependence riippuvuus
decision tree päätöspuu	dependent variable ¹¹ (response variable) vastemuuttuja, selitettävä muuttuja
decomposable model osittuva malli, dekomponoituva malli	derivative ₁ derivaatta
decomposition ₁ osittaminen, dekomponointi	derivative ₂ johdannainen
decomposition ₂ hajotelma	descriptive analytics kuvaileva analytiikka
deconvolution dekonvoluutio	descriptive statistic kuvaileva tunnusluku
decorrelated korrelaatiopuhdistettu, korrelaatiosta puhdistettu	descriptive statistics kuvaileva tilastotiede
deep learning syväoppiminen	descriptive study kuvaileva tutkimus
de facto population de facto -väestö, hetkellinen väestö	design asetelma
deff (design effect)	design-based inference asetelmaperusteinen päättely
deflation deflaatio	design effect (deff) asetelmakerroin
degenerate distribution degeneroitunut jakauma	design matrix asetelmamatriisi
degree distribution astejakauma	design of experiments koesuunnittelu
degree of belief uskomusaste	design optimality asetelmaoptimaalisuus
degrees of freedom vapausasteet, vapausasteluku	design space asetelma-avaruus
de jure population de jure -väestö, vakinainen väestö	detailed balance condition tilapariakohtainen tasapainoehto
delayed entry viivästynyt aloitus	detection bias löytymisharha
delta method deltamenetelmä	determinant determinantti
demography väestötiede, demografia	deterministic deterministinen
dendrogram ryhmittelypuu, dendrogrammi	depth function syvyysfunktio
denoised kohinasta puhdistettu	detrending trendin poisto
denoising kohinanpoisto	deviance devianssi
denominator nimittäjä	
density estimation tiheysestimointi	
density function tiheysfunktio	

taa näkee käytettävän myös kvartiileille ja kvintiileille.

¹¹Parempia termejä ovat *outcome variable* ja *response variable*.

deviance information criterion (DIC) devianssi-informaatiokriteeri	eriävä virheluokittelu
deviance residual devianssijäännös	differential privacy differentiaalinen tietosuoja
deviation poikkeama	diffuse prior diffuusi prior
DFBETA (difference in betas)	diffusion index diffuusioindeksi
DFFIT (difference in fits)	diffusion process diffuusioprosessi
diagnostics diagnostiikka	digamma function digamma-funktio
diagnostic statistic diagnostinen tunnusluku, diagnostinen suure, vikasuure	digit preference numerosuosio
diagnostic test diagnostinen testi	dimension ulottuvuus, dimensio
diagonal matrix lävistäjämatriisi, diagonaalimatriisi	dimensionality reduction (dimension reduction) dimension pienennys, dimension pudotus
diagram kuvio, diagrammi, kaavio	dimension reduction (dimensionality reduction) dimension pienennys, dimension pudotus
DIC (deviance information criterion)	dimension reduction method dimensionpudotusmenetelmä
dichotomous variable (binary variable) kaksiarvoinen muuttuja, dikotominen muuttuja, binäärimuuttuja	Dirac measure Diracin mitta
dichotomy kahtiajako, dikotomia	Dirac's delta function Diracin deltafunktio
DID (difference in differences)	direct effect suora vaikutus
difference erotus	directed acyclic graph (DAG) suunnattu syklitön verkko, suunnattu silmukaton verkko, DAG
difference equation differenssiyhtälö	direct likelihood suora uskottavuus
difference in betas (DFBETA) kerroinero	direct standardization suora vakiointi
difference in differences (DID) erotusten erotus	directional data suunta-aineisto
difference in fits (DFFIT) soviteero	Dirichlet distribution Dirichlet-jakauma
difference operator erotusoperaattori	Dirichlet process Dirichlet-prosessi
differencing erotusten laskeminen, differenssointi, differointi	disability adjusted life years (DALY) toimintakykyiset elinvuodet, DALY
differentiability derivoituvuus, differentioituvuus	disclosure control tietosuoja
differential equation differentiaaliyhtälö	discontinuous epäjatkuva
differential misclassification	discordance diskordanssi
	discordant pairs diskordantit

parit, erilaiset parit	divergence ₁ hajaantuminen, divergenssi
discounting diskonttaus	divergence ₂ poikkeama ¹²
discrete diskreetti	do-calculus do-laskenta
discrete distribution diskreetti jakauma	domain osajoukko ¹³
discrete-time diskreettiaikainen	domain adaptation kohdesopeutus
discrete variable diskreetti muuttuja	domain of attraction attraktioalue
discriminant diskriminantti	donor imputation luovuttaja-paikkaus
discriminant analysis erottelu-analyysi	do-operator do-operaattori
discriminant function erottelufunktio	D-optimality <i>D</i> -optimaalisuus
discriminatory power erottelukyky	dose meter annosmuunnos, annosmetrometri
disjoint events (mutually exclusive events) erilliset tapahtumat, toisensa poissulkevat tapahtumat	dose-response curve annos-vastekäyrä
dispersion hajonta	dot chart pistekuvio
dispersion matrix hajontamatriisi	dot plot pistekuvio, pistediagrammi
dissimilarity coefficient erilaisuuskerroin	dot product (inner product) pistetulo, sisätulo
dissimilarity matrix erilaisuusmatriisi	double-blind experiment kaksoissokkokoe
distance measure etäisyysmitta	double exponential distribution (Laplace distribution) kaksoiseksponenttijakauma, Laplace-jakauma
distance sampling etäisyysotanta	double sampling (two-phase sampling) kaksivaiheotanta
distributed-lag model eksogeenisten viipeiden malli	doubly robust estimation kaksoisvakaa estimointi
distribution jakauma	doubly stochastic model kahdesti satunnainen malli
distribution-free jakaumasta riippumaton	downward bias negatiivinen harha, alasharha
distribution function (cumulative distribution function) kertymäfunktio	draftsman's plot kuviomatriisi
disturbance ₁ (perturbation ₁) häiriö	drift ₁ vieto, suunta
disturbance ₂ (error term) virhetermi	drift ₂ ajautuminen
dithering (jittering) ravistaminen	dropout kato
	d-separation <i>d</i> -separaatio

¹²Esim. *Kullback-Leibler divergence* on suomeksi Kullbackin–Leiblerin poikkeama.

¹³Otantateoriassa *domain* viittaa perusjoukon kiinnostavaan osajoukkoon.

duality duaalisuus
dual scaling (correspondence analysis) korrespondenssianalyysi
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 measure modification (moderation) vaikutusparametrin muovautuminen, vaikutusmitan muovautuminen
effect modifier vaikutuksen muovaaja
effect of treatment on the treated (ETT) käsittelyvaikutus altistuneilla
effect size vaikutuskoko
effect variable (outcome variable) vastemuuttuja, tulostuuttuja
efficacy teho
efficiency tehokkuus
efficient estimator tehokas estimaattori
efficient market hypothesis tehokkaiden markkinoiden hypoteesi
efficient score (score₁) skoori
eigendecomposition (spectral decomposition) spektrihajotelma, ominaisarvohajotelma
eigenspace ominaisavaruus
eigenvalue ominaisarvo
eigenvector ominaisvektori
elasticity jousto, elastisuus
elastic net joustava verkko
ELBO (evidence lower bound)
element alkio
elemental estimator elementtimestimaattori
elementary contrast alkeiskont-rasti

elementary event	alkeistapaus, alkeistapahtuma
elementary unit	alkeisyksikkö
elicitation of prior	priorin koostaminen, esijakauman koostaminen
elimination	poisto, eliminointi
elliptical distribution	elliptinen jakauma
EM-algorithm	(expectation-maximization algorithm) EM-algoritmi
embedded process	upotettu prosessi
emigration	maastamuutto
empirical	empiirinen, kokemusperäinen
empirical Bayes estimation	empiirinen Bayes-estimointi
empirical coverage	empiirinen peittotodennäköisyys
empirical distribution function	empiirinen kertymäfunktio
empirical likelihood	empiirinen uskottavuus
empirical measure	empiirinen mitta
empirical size of test	testin empiirinen koko
empty set	tyhjä joukko
encompassing model	kattava malli
endogeneity	endogeenisuus, sisämääräytyneisyys
endogenous variable	(endogenous variate) endogeeninen muuttuja, sisämääräytynyt muuttuja
endpoint	(outcome event) päätapahtuma, vastetapahtuma
engineering statistics	insinööri-tilastotiede
ensemble	(statistical ensemble) kooste, tilastollinen kooste
ensemble average	koostekeskiaarvo
ensemble learning	koosteoppiminen
ensemble prediction	kooste-ennuste
entrance time	(first passage time) saavutushetki, ensikulun hetki
entropy	entropia, haje
enumeration	laskenta, luetteleminen
enumerator	väestönlaskija
envelope function	verhofunktio
environmetrics	ympäristötilastotiede
<i>E</i>-optimality	<i>E</i> -optimaalisuus
epidemic model	epidemiamalli
epidemiology	epidemiologia
equilibrium	tasapaino
equilibrium distribution	tasapainojakauma
equi-tailed	samahäntäinen
equivalence class	ekvivalenssiluokka
equivalence testing	ekvivalenssitestaus
equivariant estimator	ekvivariantti estimaattori
erf	(error function)
erfc	(complementary error function)
ergodic	ergodinen, keskiarvosuopeneva
ergodicity	ergodisuus
error	virhe, erhe
error components model	virhekomponenttimalli
error correction model	(ECM) virheenkorjausmalli
error function	(erf) virhefunktio
error margin	(margin of error)

virhemarginaali	evaluation metric (operating characteristic) toimivuusmitta
error matrix (classification table) luokittelutaulukko	event tapahtuma
error mean square (residual mean square) virhekeskineliö, jäännöskeskineliö	event history analysis (life history analysis) tapahtumahistoria-analyysi
error propagation (propagation of error) virheen kasautuminen, virheen leviäminen	event space tapahtuma-avaruus
error rate virheosuus	event study tapahtumatutkimus
errors in variables model (measurement error model ¹⁴) virheet selittäjissä -malli, mittausvirhemalli	evidence näyttö
error sum of squares (residual sum of squares) jäännösneliösumma	evidence lower bound (ELBO) näytön alaraja, variaatioalaraja
error term virhetermi	evidence upper bound (EUBO) näytön yläraja, variaatioyläraja
error variance virhevarianssi	exact test eksakti testi
estimable estimoituva, arvioitua	exceedance curve ylitysarvokäyrä, ylityslukemakäyrä
estimand estimoitava suure, estimandi	exceedance test ylitysarvotesti, ylityslukematesti
estimate estimaatti, arvio	excess hazards model lisäuhkamalli
estimating equation estimointiyhtälö, arviointiyhtälö	excess risk lisäriski
estimating function estimointifunktio	exchangeability vaihdettavuus, vaihtokelpoisuus
estimation estimointi, arviointi	exchangeable variables keskenään vaihtokelpoiset muuttujat
estimation error estimointivirhe, arviointivirhe	exclusion criterion poissulkukriteeri
estimator estimaattori, arviofunktio	exercise date lunastuspäivä
etiologic fraction etiologinen osuus	exercise price lunastushinta
ETT (effect of treatment on the treated)	exercise value lunastusarvo
EUBO (evidence upper bound)	exogeneity eksogeenisuus, ulkomääräytyneisyys
Euclidean distance euklidinen etäisyys	exogenous variable eksogeenin muuttuja, ulkomääräytynyt muuttuja
European option eurooppalainen optio	expectation (expected value) odotusarvo
	expectation-maximization algorithm (EM-algorithm) EM-algoritmi

¹⁴Esim. monografiassa Fuller (1987) tämä käsite on laajempi.

expected frequency (expected number) odotettu lukumäärä
expected loss odotustappio
expected number (expected frequency) odotettu lukumäärä
expected shortfall (average Value at Risk) odotustappioarvo
expected utility theory odotetun hyödyn teoria
expected value (expectation) odotusarvo
experiment (trial) koe
experimental design koeasetelma
experimental error koevirhe
experimental plot koeala, koe-ruutu
experimental study kokeellinen tutkimus
experimental unit koeyksikkö
expert judgment asiantuntija-arvio
expert overconfidence asiantuntijaliikavarmuus
expert system asiantuntijajärjestelmä
explanatory power selityskyky
explanatory trial selittävä koe
explanatory variable (regressor) selittävä muuttuja, selittäjä
exploration vs. exploitation trade-off etsintä-hyödynnys-kompromissi
exploratory data analysis (EDA) etsivä data-analyysi
exploratory survey esitutkimus, alustava kartoitus, esiluotaus
explosive process räjähtävä prosessi
exponential curve eksponenttikäyrä
exponential dispersion family eksponenttinen hajontaperhe

exponential dispersion model eksponenttinen hajontamalli
exponential distribution eksponenttijakauma
exponential family eksponentti-perhe
exponential growth eksponenttinen kasvu, eksponentiaalinen kasvu
exponential smoothing eksponenttinen tasoitus, eksponenttinen silotus
exponential trend eksponenttinen trendi, eksponentiaalinen trendi
exposure altistus
exposure time altistusaika
external validity ulkoinen validiteetti, ulkoinen pätevyys
extinction probability sukupuuton todennäköisyys
extra-Poisson variation ylipoissonvaihtelu
extrapolation ekstrapolointi
extremal intensity ääriarvointensiteetti, äärilukemaintensiteetti
extremal quotient vaihtelusuhte
extreme gradient boosting (XGBoost) äärigradienttitehostus
extreme value ääriarvo, äärilukema
extreme value distribution ääriarvojakauma, äärilukemajakautuma

F

face-to-face interview käyntihaastattelu
factor₁ tekijä, selittävä luokkamuuttuja
factor₂ faktori

factor analysis	faktorianalyysi	false acceptance	(acceptance error) hyväksymisvirhe
factorial	kertoma	false discovery rate	(FDR) väärin löydösten osuus
factorial cumulant	kertomakumulantti	false negative	väärä negatiivinen
factorial design	tekijäkoeasetelma	false positive	väärä positiivinen
factorial experiment	tekijäkoe	false positive rate	väärin positiivisten osuus
factorial moment	kertomamomentti	false rejection	(rejection error) hylkäysvirhe
factorial moment generating function	kertomamomenttien emäfunktio, kertomamomentit generoiva funktio	familial correlation	sukukorrelaatio
factoring	faktorointi	family-wise error rate	(FWER) perhekohtainen virhetodennäköisyys
factorization	tekijöihinjako	fast Fourier transform	(FFT) nopea Fourier-muunnos
factorization criterion	tekijöihinjakokriteeri, faktorointikriteeri	father wavelet	isäaallocke, isäväre
factorization theorem	tekijöihinjakolause, faktorointilause	fatigue model	väsymismalli
factor loading	faktorilataus	fat-tailed distribution ¹⁶	paksuhäntäinen jakauma
factor matrix	faktorimatriisi	F-distribution	(Fisher's distribution) <i>F</i> -jakauma, Fisherin jakautuma
factor pattern	faktorirakenne	FDR	(false discovery rate)
factor reversal criterion	tekijänvaihtokriteeri	feasibility study	(pilot study) toteutettavuustutkimus
factor rotation	faktorirotaatio, faktorikierto	feature	piirre
factor score	faktoripistemäärä	feature selection	piirrevalinta
failure	vikaantuminen, vioittuminen	feature space	piirreavaruus
failure function	vikakertymä	feed-forward neural network	myötävirtaneuroverkko
failure probability	epäonnistumistodennäköisyys	fertility	hedelmällisyys
failure rate ₁ ¹⁵	vikataajuus	fertility rate	hedelmällisyysluku
failure rate ₂	(hazard rate) uhkatilays	FFT	(fast Fourier transform)
failure time	vikaantumisaika	fiducial inference	fidusiaalinen päättely
fair game	reilu peli		

¹⁵Ks. alaluku 2.8.

¹⁶Englannin 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*.

field experiment	kenttäkoe	nen indeksi	
field work	kenttätö	fixed-effects model	kiinteävaikutusmalli
filter	suodin, suodatin	fixed sample	kiinteä otos
filtering	suodatus	fixed variable	kiinteä muuttuja
filtration	(history) filtraatio, historia	floor effect	lattiavaikutus
fine ordering	hieno järjestys	flow chart	vuokaavio
finite mixture distribution	äärellinen jakaumasekoite	F measure (<i>F</i> score) <i>F</i> -mitta	
finite population correction	(fpc) äärellisyyskerroin, äärellisen perusjoukon korjauskerroin	folded contingency table	taitettu kontingenssitaulukko, taitettu ristitaulukko
first digit distribution	ensimmäisen numeron jakauma	folded distribution	taitettu jakauma
first-order inclusion probability	yksikön sisällymistodennäköisyys	folded median	taitettu mediaani
first order model	ensimmäisen kertaluvun malli	folding	taittaminen
first passage time	(entrance time) saavutushetki, ensikulun hetki	follow-up	seuranta, seuruu
first return time	ensipaluun hetki	force of mortality	(hazard of death) kuolintahke
first stage sampling unit	(primary sampling unit) ensias-teen otantayksikkö	forecast	ennuste ₁
Fisher consistency	Fisher-tar- kentuvuus	forecasting	ennustaminen
Fisher information	Fisher-infor- maatio	forensic statistics	rikostekninen tilastotiede
Fisher information matrix	Fis- her-informaatiomatriisi	forest plot	metsikkökuvio
Fisher's distribution	(<i>F</i> -distri- bution) Fisherin jakauma	forward contract	termiinisopi- mus
Fisher's scoring method	Fishe- rin estimointialgoritmi	forward equation	eteenyhtälö
fitted model	sovitettu malli	forward selection	etenevä valin- ta
fitted value	sovite, sovitettu arvo	fourfold plot	nelikenttäkuvio
fitting	sovittaminen, sovitus	fourfold table	(two-by-two table) nelikenttä, 2 × 2 -taulukko
five-number summary	viiden luvun tiivistys	Fourier analysis	Fourier-analyysi
fixed base index	kiinteäkantai-	Fourier transform	Fourier- muunnos
		fpc	(finite population correction)
		fractal	fraktaal
		fractile	(quantile) fraktiili, kvantii- li
		fractional Bayes factor	fraktio- naalinen Bayes-tekijä
		fractional Brownian motion	fraktionaalinen Brownin liike
		fractional factorial experiment	

osittainen tekijäkoe	koe
fractional polynomial murtopolynomi	full information method täyden informaation menetelmä
fractional replication osittainen toisto	fully conditional distribution ¹⁷ täysehdollistettu jakauma
fraction defective virhekappaleiden osuus	functional central limit theorem funktionaalinen keskeinen raja-arvolause
frailty model haurausmalli	functional data analysis funktionaalinen data-analyysi
frame kehikko, kehys	functional determinant (Jacobian determinant) funktionaalideterminantti, Jacobin determinantti
frame population (sampled population) kehikkoperusjoukko	functional relationship funktionaalinen yhteys
frequency ₁ (count) lukumäärä	fund rahasto
frequency ₂ (rate) ₁ taajuus	funnel plot suppilokuvio
frequency ₃ (occurrence) esiintyvyys	futility hyödyttömyys
frequency curve frekvenssikäyrä	futures contract futuurisopimus
frequency data (count data) lukumääräaineisto	fuzzy logic sumea logiikka
frequency distribution lukumääräjakauma, frekvenssijakauma	fuzzy set theory sumea joukkooppi
frequency domain taajuusalue	FWER (family-wise error rate)
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 etuovivaiointi	
F score (<i>F</i> measure) <i>F</i> -mitta	
F statistic <i>F</i> -suure, <i>F</i> -tunnusluku	
F-test <i>F</i> -testi	
full factorial design täysitekijä-	

G

gain ₁ vahvistus ₁
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

¹⁷Myös muotoa *full conditional distribution* käytetään.

gamma distribution gammajakauma

gamma-frailty model gamma-haurausmalli

gamma function gammafunktio

gap välys

gap statistic aukkosuure, aukkotunnusluku

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 ristiininvalidointi, 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 latent variable model yleistetty lineaarinen piilomuuttujamalli

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

generalized linear model (GLM) yleistetty lineaarinen malli, GLM¹⁸

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

¹⁸Joissakin yhteyksissä tällä lyhenteellä viitataan termiin *general linear model*.

general linear model yleinen lineaarinen malli

generating function emäfunktio, generoiva funktio

generator of process prosessin generaattori, prosessin kehitin

genetic algorithm geneettinen algoritmi

genetic heterogeneity geneettinen heterogeenisuus

genomewide association study (GWAS) genomilaajuinen assosiaatiokartoitus, GWAS

geographical information system (GIS) paikkatietojärjestelmä

geometric distribution geometrisen 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 sovituksen hyvyys

goodness of fit test sovituksen hyvyystesti, yhteensopivuustesti

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

H_0 (null hypothesis)

H_1 (alternative hypothesis)

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

halo effect halovaikutus

Hamming code Hamming-koodi

Hamming distance Hamming-etäisyys

hanging rootogram roikkuva histogrammi

Hann window¹⁹ Hann-ikkuna

hard-core model kovaydinmalli

HARKing jälkihypotisointi

harmonic distribution harmoninen jakauma

harmonic mean harmoninen keskiarvo

harmonic regression harmoninen regressio

hash tiiviste

hat matrix hattumatriisi

hazard uhka, hasardi

hazard function (hazard rate) uhkafunktio, uhkatiheys

hazard of death (force of mortality) kuolinuhka

hazard rate (hazard function) uhkatiheys, uhkafunktio

hazard ratio (HR) uhkasuhde, HR

heads kruuna

headship rate method perheen-pääkohtainen menetelmä

heat map lämpökartta

Heaviside function Heavisiden funktio

heavy-tailed distribution

paksuhäntäinen jakauma

hedging strategy suojautumisstrategia

hedonic regression hedoninen regressio

Helmert contrast Helmert-kontrasti

herd immunity laumasuoja, laumaimmuneetti

heritability periytyvyys, heritabiliteetti

Hessian Hessen matriisi

heterogeneous heterogeeninen, sekakoosteinen

heteroskedastic erivarianssinen, heteroskedastinen

heteroskedasticity-consistent erivarianssisuuden suhteen tarkentuva, heteroskedastisuuden suhteen tarkentuva

heteroskedasticity-robust erivarianssisuuden suhteen vakaa, heteroskedastisuuden suhteen vakaa

hidden layer piilokerros

hidden Markov model (HMM) piilo-Markov-malli

hierarchical classification (nested classification) hierarkkinen luokittelu, sisäkkäinen luokittelu

hierarchical clustering hierarkkinen ryhmittely

hierarchical model (multilevel model) hierarkkinen malli, monitasomalli

high-dimensional data korkeaulotteinen aineisto, suuriulotteinen aineisto

high-dimensional statistics korkeaulotteinen tilastotiede, suuri-

¹⁹Termiä *Hanning window* käytetään myös, mutta se ei ole suositeltava.

ulotteinen tilastotiede	HR (hazard ratio)
higher-order asymptotics korkeamman kertaluvun asymptotiikka	HSD test HSD-testi
highest posterior density interval (HPD interval) suurimman posterioritiheyden väli, suurimman jälkitiheyden väli, HPD-väli	hurdle model (zero-altered model) rimamalli
highest posterior density region (HPD region) suurimman posterioritiheyden alue, suurimman jälkitiheyden alue, HPD-alue	hyperexponential distribution hyperekspONENTtijakauma
high-pass filter ylipäästösuodin, ylipäästösuodatin	hypergeometric distribution hypergeometrinen jakauma
high-throughput data isovuodatta	hyperparameter hyperparametri
high-throughput screening isovuoseulonta, sarjaseulonta	hypothesis hypoteesi, oletus
hill-climbing algorithm vuorikiipeilyalgoritmi	hypothesis testing hypoteesin testaus
hinge (quartile) kvartiili	
histogram histogrammi	I
history (filtration) historia	ICA (independent component analysis)
hitting probability osumatoden näköisyys	ICS (invariant coordinate selection)
hitting time osuma-aika, osumahetki	ideal index ideaali-indeksi, ihanneindeksi
HMM (hidden Markov model)	idempotent idempotentti
homogeneous homogeeninen, tasakoosteinen	identifiability identifioituvuus
homogeneity test homogeenisuustesti	identifier tunnistin
homoskedastic samavarianssinen, homoskedastinen	identity function identiteettifunktio
honest significant difference test HSD-testi	identity matrix yksikkömatriisi
horseshoe effect hevosenkenkäilmiö	idle period joutojakso, lepojakso
hot deck imputation kuuma-paikkaus, kuumaimputointi	IFR (infection fatality ratio)
household kotitalous	ignorable sivuutettava
HPD interval (highest posterior density interval) HPD-väli	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		yhteensopimattomuus
immortal person-time	henkilö- aika kuolemattomana		incomplete block design vajaalohkoasetelma
immortal time bias	kuolematto- muusajan harha		incomplete Latin square vajaa latinalainen neliö
implicit stratification	implisiitti- nen ositus		inconsistent estimator tarkentu- maton estimaattori
importance sampling	painotus- otanta, tärkeysotanta		independence riippumattomuus
improper distribution	epäaito jakauma		independent and identically di- stributed (i.i.d.) riippumattomat ja samoin jakautuneet, i.i.d.
improper imputation	epäkelpo imputointi		independent censoring riippu- maton sensurointi
improper prior	epäaito prior, epäaito esijakauma		independent component analy- sis (ICA) riippumattomien kom- ponenttien analyysi, ICA
impulse response function	im- pulssivastefunktio, sysäysvaste- funktio		independent increments riippu- mattomat lisäykset
imputation	paikkaus, imputointi		independent subspace analysis riippumattomien aliavaruuksien analyysi
imputation model	paikkausmal- li, imputointimalli		independent variable ²⁰ (explanatory variable) selittävä muuttuja, selittäjä
inadmissible estimator	epäkel- po estimaattori		index (index number) indeksi, in- deksiluku
inception cohort	lähtökohortti		index number (index) indeksilu- ku, indeksi
incidence	ilmaantuvuus		index of concentration keskitty- misindeksi
incidence matrix₁	indikaattori- matriisi, osoitinmatriisi		index of dispersion (dispersion index) hajontaindeksi
incidence matrix₂	kytkentämat- riisi		index of dissimilarity erilaisuus- indeksi
incidence proportion	(cumulative incidence) ilmaantu- misosuus, kumulatiivinen ilmaan- tuvuus		index of diversity monimuotoi- suusindeksi, diversiteetti-indeksi
incidence rate	ilmaantumistiheys		index plot indeksikuvi
incidental parameters	tilanne- parametrit		Indian buffet process intialai- nen noutopöytäprosessi
inclusion criterion	mukaanotto- kriteeri		indicator variable osoitin, indi- kaattorimuuttuja, ilmaisain
inclusion probability	sisältymis- todennäköisyys		
incompatability			

²⁰Suosittelava termi on *explanatory variable*.

indirect effect	epäsuora vaikutus	tiivinen sensurointi
indirect least squares	epäsuora pienin neliösumma	informative missingness infor- matiivinen puuttuvuus
indirect sampling	epäsuora otan- ta	informative prior informatiivi- nen prior, informatiivinen esija- kauma
indirect standardization	epä- suora vakiointi	informed consent tietoinen suos- tumus
induction	induktio	initial distribution alkujakauma
inductive behaviour	induktiivi- nen käyttäytyminen	inlier piilokki
inductive inference	induktiivi- nen päättely	in-migration sisäänmuutto
inefficient estimator	vajaatehoinen estimaattori	inner fence sisäaita
inefficient markets	tehottomat markkinat	inner product (dot product) sisä- tulo, pistetulo
inequality coefficient	eriarvoi- suuskerroin	innovation gain innovaatiohyöty
infection fatality ratio (IFR)	infektiokuolleisuus	innovation term satunnaistermi, innovaatiotermi
inference	päättely	input layer syötekerros
infinitely decomposable	ääret- tömästi osittuva	input-output matrix (IO ma- trix) panos-tuotosmatriisi
infinitely divisible	rajatta jakau- tuva, äärettömästi jakautuva	input-output model panos-tuo- tosmalli
inflation	inflaatio	inquiry tutkimus, selvitys, kysely
influence	vaikutus ₂	instance tapaus ₂
influence function	vaikutusfunktio	instance space tapausavaruus
influence statistic	vaikutustunnusluku	instantaneous death-rate (force of mortality) kuolemanuhka
influential observation	vaikuttava havainto	instantaneous state hetkellinen tila
information	informaatio	instrumental variable (IV) välinemuuttuja, instrumentti- muuttuja
information criterion	informaa- tiokriteeri	insurance vakuutus
information function	informaa- tiofunktio	integrable integroituva
information matrix	informaatio- matriisi	integral integraali
information theory	informaatio- teoria	integrated hazard (cumulative hazard) uhkakertymä
informative censoring	informa-	integrated intensity (cumulative intensity) intensiteettikertymä
		integrated moving average model (IMA model) integroitu liukuvan keskiarvon malli, IMA- malli

integrated spectrum spektri-
tymäfunktio, integroitu spektri

intelligence quotient (IQ) älyk-
kysosamäärä

intensity (rate_1) intensiteetti, taa-
juus

intensity function intensiteetti-
funktio

intention-to-treat analysis hoi-
toaikeen mukainen analyysi, ITT-
analyysi

interaction yhdysvaikutus, inter-
aktio

interarrival time (interevent ti-
me) saapumisväliaika

interblock lohkojen välinen

intercept vakiotermi

inter-class similarity luokkien
välinen samankaltaisuus

interest rate korkotaso

interevent time (interarrival ti-
me) tapahtumaväliaika

interim analysis välianalyysi

internal validity sisäinen validi-
teetti

internal variance sisäinen va-
rianssi

internet survey verkkokysely,
nettikysely

interpenetrating samples (net-
work of samples) kytketyt otokset,
otosverkosto

interpolating spline interpoloiva
splini

interpolation interpolaatio

interquartile range (IQR) kvar-
tiiliväli

intersection leikkaus

interval censoring
välisensurointi

interval estimation
väliestimointi

interval scale välimatka-asteikko,
intervalliasteikko

intervention analysis interven-
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 coordinate selection
(ICS) invariantti koordinaatiston
valinta

invariant coordinate system
invariantti koordinaatisto

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 normaali-jakauma,
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, käänteismuunnos
inverted beta distribution käänteinen betajakauma
inverted Dirichlet distribution käänteinen Dirichlet-jakauma
invertible kääntyvä
invertible process kääntyvä prosessi
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 squares (IWLS) iteratiivisesti painotettu pienin neliösumma
Itô integral Itô-integraali
ITT analysis (intention-to-treat analysis) hoitoaikkeen mukainen analyysi, ITT-analyysi
IV (instrumental variable)

IWLS (iteratively weighted least squares)

J

jackknife estimator linkkuveitsiestimaattori, jackknife-estimaattori
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 diagonalization yhteisdiagonalisointi
joint distribution yhteisjakauma
jointly sufficient statistics yhteisesti tyhjentävät 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
Karhunen–Loève transform Karhusen–Loèven muunnos
k-class estimator k -luokan estimaattori
Kendall’s tau (coefficient of dissimilarity) Kendallin tau
kernel density estimator ydintiheysestimaattori
kernel estimator ydinestimaattori
kernel function ydinfunktio
kernel mean function ydinodotusarvofunktio
kernel method ydinmenetelmä
kernel regression estimator ydinregressioestimaattori
kernel smoothing ydintasoitus,

ydinsilotus
k-means clustering k :n keskiarvon ryvästys
k-nearest-neighbors method (kNN) k :n lähinaapurin menetelmä
kNN (k -nearest-neighbors method)
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 kurtoosi²¹
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
ladle estimator kauhaestimaatto-

²¹Kun *kurtosis* viittaa nimenomaisesti κ -parametriin (alaluku 2.6).

ri	nalainen hyperkuutio-otanta
ladle plot kauhakuvio	Latin rectangle latinalainen suorakulmio
lag viive ²²	Latin square latinalainen neliö
lag operator (backshift operator) viiveoperaattori	lattice hila
Lagrange multiplier test (score test) Lagrangen kerrointesti	lattice design hila-asetelma
lag window viiveikkuna	lattice distribution hilajakauma
landmark analysis virstanpylväsanalyysi	lattice sampling hilaotanta
Laplace distribution (double exponential distribution) Laplace-jakauma	lattice square hilaneliö
Laplace transform Laplace-muunnos	law of large numbers suurten lukujen laki
large deviation theory suurten poikkeamien teoria	law of small numbers pienten lukujen laki
large sample theory suurten otosten teoria	law of iterated expectations iteroitujen odotusarvojen laki
Laspeyres' index Laspeyresin indeksi	law of the iterated logarithm (LIL) iteroidun logaritmin laki
LASSO (least absolute shrinkage and selection operator)	law of truly large numbers todella suurten lukujen laki
last observation carried forward (LOCF) viimeisen havainnon toisto, LOCF-menetelmä	linear discriminant analysis (LDA) lineaarinen erotteluanalyysi
LATE (local average treatment effect)	LD (linkage disequilibrium)
latent piilo-, piilevä	LD50 (median lethal dose)
latent class analysis piiloluokkanalyysi	LDA (linear discriminant analysis)
latent projection piilotusprojektiio	leading indicator ennakoiva indikaattori, ennakoiva osoitin
latent structure piilorakenne	leading question johdatteleva kysymys
latent trait analysis piilopiirreanalyysi	leaps-and-bounds algorithm harpaten valinta
latent transition analysis piilo-siirtymäanalyysi	least absolute deviation method (LAD) pienimmän itseispoikkeaman menetelmä
latent variable piilomuuttuja	least absolute shrinkage and selection operator (LASSO) LASSO, pienimmän itseiskutistamisen ja valinnan operaattori
Latin cube latinalainen kuutio	least favorable distribution epäsuotuisin jakauma
Latin hypercube sampling lati-	least significant difference test

²²Taivutetaan "viipeen" tai "viiveen".

(LSD test) vähiten merkitsevän erotuksen testi, LSD-testi	likelihood interval uskottavuusväli
least squares estimator (LSE) pienimmän neliösumman estimaattori, PNS-estimaattori	likelihood principle uskottavuusperiaate
least squares method (method of least squares) pienimmän neliösumman menetelmä, PNS-menetelmä	likelihood ratio uskottavuusosamäärä
least variance difference method pienimmän varianssin erotusmenetelmä	likelihood ratio statistic uskottavuusosamääräsuure, uskottavuusosamäärätunnusluku
leave-one-out cross-validation (LOOCV) yksi-pois-ristiinvalidointi	likelihood ratio test (LRT) uskottavuusosamäärätesti
left censoring sensurointi vasemmalta	likelihood region uskottavuusalue
left truncation katkaisu vasemmalta, typistys vasemmalta	likely (probable) todennäköinen
length-biased sampling pituus-harhainen otanta	Likert scale Likertin asteikko
level map tasokartta	LIL (law of the iterated logarithm)
level of factor tekijän taso	l.i.m. (limit in the mean)
level of significance (significance level) merkitsevyystaso	limit in probability (plim) stokastinen raja-arvo
leverage vipuvoima	limit in the mean (l.i.m.) keskineliöraja-arvo
Lexis diagram Lexis-kuvio	limited dependent variable rajoitettu vaste
life course analysis elinkaariaanalyyssi	limited information estimator rajoitetun informaation estimaattori
life cycle utility elinkaarihyöty	limiting distribution rajajakautuma
life expectancy elinajanodote	linear lineaarinen
life insurance henkivakuutus	linear constraint lineaarinen rajoite
life table elinajantaulukko	linear dependence lineaarinen riippuvuus
lifetime elinaika	linear estimator lineaarinen estimaattori
lifetime data analysis (survival analysis) elinaika-analyysi	linear filter lineaarinen suodin
lifetime distribution elinaikajakautuma, elinajan jakauma	linearized linearisoitu
light-tailed distribution ohut-häntäinen jakauma	linear model lineaarinen malli
likelihood uskottavuus	linear optimization (linear programming) lineaarinen optimointi, lineaarinen ohjelmointi
likelihood function uskottavuusfunktio	linear predictor lineaarinen en-

nustin

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, line graph, line diagram) 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

list lista₁

listwise deletion listapoisto

loading matrix latausmatriisi

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 voimakas testi

locally estimated scatterplot smoothing (LOESS) paikallisesti estimoitu sirontakuviolositus, LOESS

locally weighted regression paikallisesti painotettu regressio

locally weighted scatterplot smoothing (LOWESS) paikallisesti painotettu sirontakuviolositus, 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 logaritimuunnos

logistic curve logistinen käyrä

logistic distribution logistinen jakauma

logistic regression logistinen regressio

logit model logit-malli

logit transformation logit-muunnos

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-vastasuhte, log-veto-kerroin, logitti
logrank test logrank-testi
logworth log-arvo²³
lollipop plot nuppineulakuvio
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 scatterplot smoothing)
low-pass filter alipäästösuodin, alipäästösuodatin
LRT (likelihood ratio test)
LSD test (least significant difference test) LSD-testi

LSE (least squares estimator)

L-statistic L-tunnusluku

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 ilminuuttuja

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

marginal causal effect reuna-kausaalivaikutus

marginal density function reu-

²³Tämä on p -arvon muunnos $-\log_{10}(p)$.

nätiheysfunktio

marginal dependence reunariippuvuus

marginal distribution reunajakauma, marginaalijakauma

marginal effect reunavaikutus

marginal frequency reunalukumäärä, reunafrekvenssi

marginal homogeneity reunahomogeenisuus, marginaalihomogeenisuus

marginal independence reunariippumattomuus

marginalize reunallistaa

marginal likelihood reunauskottavuus, marginaaliskottavuus

marginal model reunamalli, marginaalimalli

marginal probability reunatodennäköisyys

marginal structural model (MSM) reunarakennemalli, marginaalirakennemalli

marginal table reunataulukko

marginal total reunasumma

margin of error (error margin) virhemarginaali

Marimekko plot (mosaic plot) marimekkokuvio, mosaiikkikuvio

marked point process merkkipisteprosessi

Markov chain Markovin ketju, Markov-ketju

Markov chain Monte Carlo method (MCMC method) Markovin ketju Monte Carlo -menetelmä, MCMC-menetelmä

Markovian markovinen, markov²⁴

Markov process Markov-prosessi, Markovin prosessi

Markov property Markov-ominaisuus

Markov random field Markovin satunnaiskenttä

Markov switching model Markovin vaihtomalli

mark-recapture sampling (capture-recapture sampling) merkintä-takaisinpyyntiotanta, toistopyyntiotanta

MARS (multivariate adaptive regression splines)

martingale martingaali

martingale measure martingaalimitta

martingale residual martingaali-jäännös, martingaaliresiduaali

masking effect peittovaikutus

master sample perusotos

matched case-control study kaltaistettu tapaus-verrokki-tutkimus

matched pair kaltaistettu pari

matched samples (paired samples) kaltaistetut otokset

matched set kaltaistusjoukko

matching kaltaistus

mathematical statistics matemaattinen tilastotiede

mat operator matrisointioperaattori

matrix algebra matriisilaskenta, matriisialgebra

matricize matrisoida

matricization matrisointi

matrix decomposition matriisihajotelma

matrix factorization matriisin tekijöihinjako

matrix normal distribution matriisinormaalijakauma

²⁴Ks. alaluku 2.5.

matrix sampling	matriisiotanta	elinajan odote	
matrix-variate	matriisiarvoinen, matriisimuuttujainen	mean reversion	odotusarvoon palautuva
matrix-variate regression	mat- riisivasteregressio	mean square continuity	keski- neliöjatkuvuus
maturity date	eräpäivä	mean squared error (MSE)	kes- kineliövirhe
maximum a posteriori estima- tor (MAP estimator)	suurimman posterioritiheyden estimaatto- ri, suurimman jälkitiheyden esti- maattori, MAP-estimaattori	mean time to failure (MTTF)	keskimääräinen vikaantumisaika
maximum entropy method	suurimman entropian menetelmä	mean time between failures (MTBF)	keskimääräinen vikaantu- misväli
maximum likelihood estimator (MLE)	suurimman uskottavuus- den estimaattori, SU-estimaattori, ML-estimaattori	mean vector	keskiarvovektori
matched case-control study	kaltaistettu tapaus-verrokki- tutkimus-estimaattori	measurable	mitallinen
maximum likelihood method	suurimman uskottavuuden mene- telmä, SU-menetelmä, ML-mene- telmä	measure	mitta
maximum tolerated dose (MTD)	suurin siedetty annos	measurement error	mittausvir- he
MCAR (missing completely at random)		measurement error model	mit- tausvirhemalli
MCMC (Markov Chain Monte Carlo)		measurement scale	mitta-asteik- ko
MDL (minimum description length)		measure of association	yhteysmitta, assosiaatiomitta
MDS (multidimensional scaling)		measure of central tendency (measure of location)	keskiluku, sijaintimitta, sijaintiluku
mean₁	keskiarvo	measure of dispersion (measure of variation)	hajontamitta, hajon- taluku, vaihtelumitta
mean₂ (expected value)	odotusar- vo	measure of location (measure of central tendency)	sijaintimitta, sijaintiluku, keskiluku
mean absolute deviation (MAD)	keskipoikkeama	measure of variation (measure of dispersion)	vaihtelumitta, ha- jontamitta, hajontaluku
mean absolute error (MAE)	keski-itseisvirhe	measure theory	mittateoria
mean integrated squared error (MISE)	integroitu keskineliövirhe	median	mediaani
mean residual life	jäljellä olevan	median effective dose (ED50)	puoleen tehoava annos, ED50
		median effective level	puoleen tehoava taso
		median lethal dose (LD50)	puo- let tappava annos, LD50

- median lethal level** puolet tappava taso
- median regression curve** mediaaniregressiokäyrä
- median survival time** mediaanielinaika
- median test** mediaanitesti
- median unbiasedness** mediaaniharhattomuus
- mediation** vaikutuksen välittyminen, välitys, mediaatio
- mediation analysis** välitysanalyysi, mediaatioanalyysi
- mediator** välittävä tekijä
- medical statistics** lääketieteellinen tilastotiede
- memorylessness property** muistamattomuusominaisuus
- M*-estimator** *M*-estimaattori
- meta-analysis** meta-analyysi
- metameter** muunnosarvo, metametri
- meta regression** metaregressio
- method of least squares** (least squares method) pienimmän neliösumman menetelmä
- method of maximum likelihood** (maximum likelihood method) suurimman uskottavuuden menetelmä
- method of moments** momenttimenetelmä
- methodology** metodologia
- Metropolis–Hastings algorithm** Metropolisin–Hastingsin algoritmi
- MGF** (moment generating function)
- mid-hinge** ²⁵ kvartiilikeskisarvo
- mid-mean** keskipuolikkaan keskiarvo
- mid-*p*-value** keski-*p*-arvo
- 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
- minimax criterion** minimax-kriteeri
- minimax estimation** minimax-estimointi
- minimum chi-square** pienin khineliö
- minimum covariance determinant estimator** pienimmän kovarianssideterminantin estimaatti
- minimum description length principle** (MDL) lyhimmän kuvauspituuden periaate
- minimum distance index** minimietäisyysindeksi
- minimum norm quadratic unbiased estimator** (MINQUE) kvadraattinen, harhaton miniminormiestimaattori, MINQUE-estimaattori
- minimum variance unbiased estimator** (MVUE) pienimmän varianssin harhaton estimaattori
- MINQUE** (minimum norm quadratic unbiased estimator)
- mirrored density plot** (mirror density chart) peilattu tiheyskuvio
- misclassification** väärinluokittelu, virheluokittelu
- MISE** (mean integrated squared error)

²⁵1. ja 3. kvartiilin keskiarvo

missing at random (MAR) satunnaisesti puuttuva, MAR
missing by design suunnitellusti puuttuva
missing completely at random (MCAR) täysin satunnaisesti puuttuva, MCAR
missing data puuttuva data
missing data pattern puuttuvuusrakenne
missing information principle puuttuvan informaation periaate
missing not at random (MNAR) ei-satunnaisesti puuttuva
misspecification väärinspesifiointi
misspecified model väärin spesifioitu malli
mixed distribution sekajakauma
mixed-effects model (mixed model) sekamalli
mixed model (mixed-effects model) sekamalli
mixed spectrum sekaspektri, yhdistetty spektri
mixed strategy sekastrategia
mixing sekoittava, sekoittavuus
mixing distribution sekoittava jakauma
mixture distribution jakauma-seos, jakaumien sekoitus
mixture experiment sekoitekoe
mixture model sekoitemalli
MLE (maximum likelihood estimator)
MNAR (missing not at random)
mode moodi, tyyppiarvo
model malli
model-assisted malliavusteinen
model averaging mallikeskiarvoistus
model-based inference malliperusteinen päättely
model building mallinrakentaminen
model diagnostics mallidiagnostiikka
model error mallivirhe
model fit mallin sopivuus
modeling mallinnus, mallitus
model matrix mallimatriisi
moderation (effect measure modification) vaikutusparametrin muovautuminen, vaikutusmitan muovautuminen
moderator vaikutuksen muovaaja
modified profile likelihood muokattu profiiliuskottavuus
moment momentti
moment estimator momenttietimaattori
moment generating function (MGF) momenttiemäfunktio, momentit generoiva funktio
moment matrix momenttimatriisi
moment ratio momenttisuhte
monotone monotoninen
monotone likelihood ratio monotoninen uskottavuusosamäärä
Monte Carlo method Monte Carlo -menetelmä
monthly average kuukausikeskiarvo
moralization of graph verkon moralisointi
morphometrics muotoanalyysi, morfometria
mortality rate (death rate) kuolevuus
mosaic plot (Marimekko plot) mosaiikkikuvio, marimekkokuvio
most accurate confidence interval lyhyin luottamusväli
most powerful test voimakkain

testi

most stringent test tiukin testi

mother wavelet äitiaalloke, äiti-väre

mover-stayer model liikkuja-py-syjamalli

moving average liukuva keskiarvo

moving average model (MA model) liukuvan keskiarvon malli, MA-malli

moving range liukuva vaihteluväli

MSE (mean squared error)

MSM (marginal structural model)

M-statistic M-tunnusluku

MTBF (mean time between failures)

MTD (maximum tolerated dose)

MTTF (mean time to failure)

multiarmed bandit monikäätinen rosvo

multicategory logit model (multinomial logit model) multinomilogit-malli

multicenter study monikeskustutkimus

multicenter trial monikeskuskoe

multicollinearity (collinearity) kollineaarisuus, multikollineaarisuus

multidimensional moniulotteinen

multidimensional scaling (MDS) moniulotteinen skaalaus

multiepisode model monijakso-malli

multiequational model moniyhtälömalli

multifactorial design moniteki-jäkoe

multigraph moniverkko, multi-verkko

multihit model moniosumamalli

multilevel model monitasomalli, hierarkkinen malli

multilinear process multilineaarinen prosessi, monilineaarinen prosessi

multimodal distribution monihuippuinen jakauma

multinomial coefficient multinomikerroin

multinomial distribution multinomijakauma

multinomial logistic regression multinomilogistinen regressio

multinomial logit model (multicategory logit model) multinomilogit-malli

multinormal distribution (multivariate normal distribution) multinormaalijakauma, moninormaalijakauma

multipartite monijakoinen

multiphase sampling monivaiheinen otanta, monivaiheotanta

multiphase screening monivaiheiselontaa

multiple bar chart ositettu pylväskuvio, ositettu palkkikuvio

multiple comparison moniver-tailu

multiple correlation coefficient yhteiskorrelaatiokerroin

multiple criteria decision making monitavoitteinen päätöksen-teko

multiple criteria optimization monitavoiteoptimointi

multiple decrement model monipoistumamalli

multiple factor analysis (factor analysis) faktorianalyysi

multiple-frame survey monikehikkoluotaus

multiple imputation	monipaikkaus, moni-imputointi	(multiple regression)	monen selittäjän regressio
multiple range test	usean vaihteluvälin testi	multivariate	monimuuttujainen, monen muuttujan, moniulotteinen
multiple recapture design	toistopyyntien asetus, sarjatoistopyynti	multivariate adaptive regression spline	(MARS) monimuuttujainen mukautuva regressiosplini
multiple regression	monen selittäjän regressio	multivariate analysis	monimuuttuja-analyysi
multiple response analysis	monivastausanalyysi	multivariate analysis of variance	(MANOVA) monen vasteen varianssianalyysi, MANOVA
multiple smoothing method	monisilotusmenetelmä, monitasotusmenetelmä	multivariate counting process	monimuuttujainen laskuriprosessi
multiple stratification	moniositus	multivariate distribution	(joint distribution) moniulotteinen jakauma
multiple testing	monitestaus	multivariate Gaussian distribution	(multinormal distribution) moniulotteinen Gaussin jakauma, multinormaalijakauma, moninormaalijakauma
multiple time series	(vector time series) moniulotteinen aikasarja, vektoriaikasarja	multivariate imputation by chained equations	(MICE) ketjuimputointi
multiplicative model	multiplikaatiivinen malli, kerrannainen malli	multivariate median	moniulotteinen mediaani
multiplicative seasonal model	multiplikatiivinen kausivaihtelumalli, kerrannainen kausivaihtelumalli	multivariate moment	moniulotteinen momentti
multiplicity controlled	monivertailukontrolloitu, monivertailurukattu	multivariate normal distribution	(multivariate Gaussian distribution) multinormaalijakauma, moninormaalijakauma
multistage estimation	moniasasteinen estimointi	multivariate rank	sijapiste
multistage sampling	moniasteinen otanta, moniasteotanta	multivariate rank test	moniulotteinen sijatesti
multistate life table	monitilainen elinajantaulukko	multivariate regression ²⁶	monen vasteen regressio
multistate model	monitilamalli	multivariate sign test	moniulotteinen merkkitest
multi-task learning	monitehtävöppiminen		
multivalued decision diagram	monivaihepäätöskaavio		
multivariable regression			

²⁶Ei ole synonyymi termeille *multiple regression* ja *multivariable regression*, vaikka monet soveltajat siinä merkityksessä tätä termiä käyttävät.

mutual fund sijoitusrahasto
mutual information keskinäisin-
 formaatio
mutually exclusive events
 (disjoint events) erilliset tapah-
 tumat, toisensa poissulkevat ta-
 pahtumat
MVUE (minimum variance unbias-
 ed estimator)

N

Nadaraya–Watson estimator
 Nadarayan–Watsonin estimaattori
naive forecast naiivi ennuste
natural conjugate prior luon-
 nollinen 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 proses-
 sointi, käyttökielen prosessointi
natural parameter (canonical
 parameter) luonnollinen paramet-
 ri, kanoninen parametri
NBU (new better than used)
NDE (natural direct effect)
nearest-neighbors clustering
 lähinaapuriryhmittely
nearest-neighbors method lähi-
 naapurimenetelmä
near unit root yksikköjuurta lä-
 hestyvä
negative binomial distribution
 negatiivinen binomijakauma
**negative exponential distribu-
 tion** (exponential distribution)
 eksponenttijakauma
**negative hypergeometric dis-
 tribution** negatiivinen hypergeo-

metrinen jakauma
negative predictive value nega-
 tiivinen ennustearvo
negative skewness vasemmalle
 vinous
neighbor embedding naapurus-
 toupotus
nested case-control study pe-
 sitetty tapaus-verrokkitutkimus,
 upotettu tapaus-verrokkitutkimus
nested design sisäkkäinen asetel-
 ma
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) otos-
 verkosto, 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 Ney-
 manin–Pearsonin apulause, Ney-
 manin–Pearsonin lemma
NIC (network information crite-
 rion)
NID (normally and independently
 distributed)
NIE (natural indirect effect)
NLP (natural language proces-

sing)	
node (vertex) solmu	nonnested classification epäsi-
noise kohina, häly	säkkäinen luokitus
nominal confidence level nimel-	non-normal ei-normaalinen
linen luottamustaso	non-orthogonal epäortogonaali-
nominal coverage probability	nen
nimellinen peittotodennäköisyys	nonparametric parametriton,
nominal scale luokka-asteikko,	epäparametrinen
luokitteluasteikko	non-probability sampling epäai-
nominal significance level ni-	to otanta
mellinen merkitsevyystaso	non-random ei-satunnainen
nominal size nimellinen koko	non-randomized satunnaistama-
nomogram nomogrammi	ton
non-central epäkeskinen	nonresponse vastauskato
non-centrality parameter epä-	non-sampling error otannasta
keskisyysparametri	johtumaton virhe
noncoverage bias peittoharha	nonsense correlation hölynpöly-
non-differential misclassifica-	korrelaatio, puppu-korrelaatio
tion yhtäläinen virheluokittelu	nonsignificant ei-merkitsevä
non-experimental study epäko-	nonsingular epäsingulaarinen
keellinen tutkimus	non-terminal event
non-Gaussian component anal-	välitapahtuma
ysis epägaussisten komponenttien	normal approximation norma-
analyysi	liapproksimaatio
nonignorable sivuuttamaton	normal distribution (Gaussian
non-inferiority samanveroisuus	distribution) normaalijakauma,
non-inferiority trial samanveroi-	Gaussin jakauma
suuskoe	normal equations normaaliyhtä-
noninformative prior epäinfor-	löt
matiivinen prior, epäinformatiivi-	normality normaalisuus
nen esijakauma	normalizing transformation
non-life insurance	normalisoiva muunnos
vahinkovakuutus	normally and independently
nonlinear model epälineaarinen	distributed (NID) riippumatto-
malli	mat ja normaalijakautuneet
nonlinear optimization epäline-	normal probability plot nor-
aarinen optimointi	maalijakaumakuvio
nonlinear regression epälineaari-	normal score normaaliarvo
nen regressio	notched boxplot lovettu laatik-
nonnegative definite (positive	kokuvio
semidefinite) positiivisesti semide-	nowcasting nykyistäminen
finiitti, ei-negatiivisesti definiitti,	NR (net reproduction rate)
epänegatiivisesti definiitti	nuisance parameter kiusapara-
	metri, häittaparametri

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 numeerinen optimointi
nuptiality avioliittoisuus

O

objective function kohdefunktio
oblimax rotation oblimax-rotatio, oblimax-kierto
oblimin rotation oblimin-rotatio, oblimin-kierto
oblique vino
oblique projection vinoprojektio
oblique reflection vinoheijastus
observable havaittava
observation havainto
observational error havaintovirhe
observational study havainnoiva tutkimus
observational unit havaintoyksikkö, 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 taapahtuma-alttiussuhde
octile oktiili
odds vastasuhde, vetokerroin
odds ratio (OR) ristisuhde, vetosuhde, OR
official statistics virallinen tilasto
offset term offset-termi
offspring distribution jälkeläisjakauma, lisääntymisjakauma
ogive (sample distribution function) otoskertymäfunktio
Oja median Ojan mediaani, Oja-mediaani
oldest old vanhoista vanhimmat
OLS (ordinary least squares)
omitted variable bias puuttuvan muuttujan harha
omnibus test kattava testi, yhteistesti
one-armed bandit yksikätkäinen rosvo
one-compartment model yksilokeromalli
one-hit model yksiosumamalli
one-sample t-test yhden otoksen t-testi
one-sided alternative hypothesis yksisuuntainen vastahypoteesi, yksisuuntainen vaihtoehtoinen hypoteesi
one-sided test (one-tailed test) yksisuuntainen testi, yksitahoinen testi
one-tailed test (one-sided test) yksisuuntainen testi, yksitahoinen testi
one-way analysis of variance yksisuuntainen varianssianalyysi
one-way classification yksisuuntainen luokittelu
open data avoin data

open-ended class	avoin päätyluokka		nen aineisto, sija-aineisto
open population	avoin väestö	ordinal regression	järjestysregressio, sijaregressio, ordinaaliregressio
operating characteristic	(evaluation metric) toimivuusmitta	ordinal scale	järjestysasteikko, järjestysasteikko, sija-asteikko, ordinaaliasteikko
operations research	operaatio-tutkimus	ordinal scale variable	järjestysasteikollinen muuttuja, sija-asteikollinen muuttuja
opinion poll	(opinion survey) mielipidekysely, mielipidetiedustelu	ordinal variable	järjestysmuuttuja, sijamuuttuja
opinion survey	(opinion poll) mielipidekysely, mielipidetiedustelu	ordinary least squares estimator	(OLS) pienimmän neliösumman estimaattori
optimal allocation	optimaalinen kiintiöinti, optimikiintiöinti	ordination of species	lajien ordinaatio
optimal design	optimiasetelma	orthogonal	(perpendicular) ortogonaalinen, kohtisuora
optimal transport	optimikuljetus	orthogonal polynomial	ortogonaalipolynomi
optimality	optimaalisuus	orthogonal projection	ortogonaaliprojektio, kohtisuora projektiio
optimal scaling	optimiskaalaus	orthonormal	ortonormaallinen
optimization	optimointi	oscillation	heilahtelu, värähtely
optimization method	optimointimenetelmä	oscillatory process	värähtelevä prosessi
optional time	(stopping time) valintahetki, pysäytyshetki	outcome	(response) vaste, tulos
OR	(odds ratio)	outcome variable	(response variable) vaste, vastemuuttuja, selitetävä muuttuja
order determination	asteluvun määrittäminen	outer fence	ulkoaita
ordered categorical variable	(ordinal categorical variable) järjestetty luokkamuuttuja	outer product	ulkotulo
ordered sample	järjestetty otos	outlier	oudokki, vierashavainto
order of interaction	yhdyksvaikutuksen kertaluku	out-migration	poismuutto
order of magnitude	suuruusluokka	out-of-bag error	laukutusvirhe
order statistic	järjestystunnusluku, sijatunnusluku	out-of-sample prediction	ulkoinen ennustaminen
ordinal categorical variable	(ordered categorical variable) järjestetty luokkamuuttuja	out-of-sample testing	ulkoinen testaus
ordinal data	järjestysasteikollinen	output layer	vastekerros, tulostekerros

overadjustment liikavakiointi
overcount (overenumeration) laskennan ylipieitto
overcoverage ylipieitto
overenumeration (overcount) laskennan ylipieitto
overdispersion ylihajonta
overfitting liikasovitus, ylisovitus
overidentification yli-identifiointi
overmatching liikakaltaistus, ylikaltaistus
overparameterized liikaparamet-roitu, yliparametroituu
overplotting päällekkäistyminen
oversampling₁ ylipoiminta
oversampling₂ ylinäytteistys
own children method omien lasten menetelmä

P

paired comparison parivertailu
paired samples pariotokset
paired *t*-test parittainen *t*-testi
pairwise independence parittai-nen riippumattomuus
 pairwise *t*-test (paired *t*-test) pa-rittainen *t*-testi
panel data paneeliaineisto, panee-lidata
panel study paneelitutkimus
panel wave paneeliaalto
parallel coordinate plot rinnak-kaisten koordinaattien kuvio
parallel groups design rinnak-kaisten ryhmien asetelma
parameter parametri
parameter-driven model para-metrivetoinen malli
parameter space parametriava-

ruus
parametric distribution para-metrinen jakauma
parametric inference parametri-nen päättely
parametric model parametrinen malli
parametrization parametointi
Pareto distribution Pareto-ja-kauma
parity₁ lapsiluku
parity₂ pariteetti
parity-progression ratio pari-teettisuhde
parsimony vähäparametrisuus, säästeliäisyys
partial association osittaisyh-teys, osittaisassosiaatio
partial autocorrelation osittai-sautokorrelaatio
partial correlation osittaiskorre-laatio
partial least squares (PLS) osit-tainen pienin neliösumma
partial likelihood osittaisuskot-tavuus
partially balanced array osit-tain tasapainotettu asetelma
partial pie chart osittainen pii-rakkakuvio
partial proportional odds osit-tain verrannolliset vastasuhteet
partial rank correlation osit-taisjärjestyskorrelaatio
partial regression coefficient osittaisregressiokerroin
partial regression plot (added variable plot) osittaisregressioku-vio, lisäselittäjäkuvio
partial replacement osittainen palautus

²⁷Tietojenkäsittelytieteessä käytetään tietokantojen yhteydessä suomennosta “osit-

partial residual plot osittais-jäännöskuvio	pension fund eläkerahasto
partial table osataulukko ²⁷	pentad criterion viisikkokriteeri
participation rate osallistumis-osuus, osallistumisaste	percentage distribution prosenttijakauma
partition ositus	percentage point prosenttiyksikkö
partitioned matrix ositettu matriisi	percentile prosentiili, prosenttipiste
passage time kulkuaika	perceptron päättelin
path analysis polkuanalyysi	percolation theory perkolaatio-teoria
path coefficient polkukerroin	periodic process jaksollinen prosessi
path dependency polkuriippuvuus	periodic survey (repeated survey) toistoluotaus
pattern mixture model puuttuvuuden sekoitemalli	period of seasonality kausijakso
pattern recognition hahmontunnistus	periodogram jaksokuvio, periodogrammi
pay-off matrix tuottomatriisi	permissible estimator hyväksyttävä estimaattori
PCA (principal component analysis)	permutation permutaatio
PDF (probability density function)	permutation test permutaatiotesti
peakedness (kurtosis ₂) huipukkuus	perpendicular (orthogonal) kohtisuora, ortogonaalinen
Pearson residual Pearson-jäännös	perpetuity pysyvä elinkorko
Pearson's chi-squared statistic Pearsonin khi-neliösuure, Pearsonin khi-neliötunnusluku	per-protocol analysis protokollan mukainen analyysi
Pearson's chi-squared test Pearsonin khi-neliötesti	personal probability omato-dennäköisyys, henkilökohtainen todennäköisyys
penalized estimation sakotettu estimointi	person-time henkilöaika
penalized least squares sakotettu pienin neliösumma	person-year henkilövuosi
penalized likelihood sakotettu uskottavuus	perspective plot perspektiivikuvio
penalty function sakkofunktio	perturbation₁ (disturbance ₁) häiriö
penalty term sakkotermi	perturbation₂ sysäys
penetrance penetranssi, ilmene-misyleisyys, läpäisy	perturbation method häiriömenetelmä, perturbaatiomenetelmä
	PF (preventive fraction)

tainen taulu”.

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

population₁	perusjoukko, populaatio	syys	
population₂	väestö, populaatio	post hoc comparison	jälkikäteisvertailu, post hoc -vertailu
population-averaged model	populaatiokeskiarvoistettu malli	post-interventional distribution	intervention jälkeinen jakautuma
population genetics	populaatio-genetiikka	post-stratification	jälkiositus
population growth model	populaation kasvumalli, väestönkasvumalli	potential	potentiaali
population pyramid	populaatiopyramidi, väestöpyramidi	potential outcome	potentiaalin vaste
population structure	väestörakenne	power₁	voima
portemanteau test	portemanteau-testi, naulakkotesti	power₂	potenssi
portemanteau theorem	portemanteau-lause, naulakkolause	power function	voimafunktio
portfolio optimization	arvopaperisalkun optimointi	power law	potenssilaki
positive definite	positiividefiniitti, positiivisesti definiitti	power spectrum	spektritiheys ²⁸
positive predictive value	positiivinen ennustearvo	power transformation	potenssimuunnos
positive semidefinite	(non-negative definite) positiivisesti semidefiniitti	P-P plot	P-P-kuvio, todennäköisyyskuvio
positive skewness	oikealle vinous	pps	(probability proportional to size)
postal survey	postikysely	pragmatic trial	(effectiveness trial) pragmaattinen koe, vaikuttavuuskoe
post-clustering	jälkiryöstys	precision₁	toistotarkkuus, täsmällisyys
posterior distribution	posteriorijakauma, jälkijakauma	precision₂	keskittyneisyys, täsmällisyys
posterior interval	posterioriväli, jälkiväli	precision₃	(positive predictive value) positiivinen ennustearvo ²⁹
posterior predictive distribution	posterioriennustejakauma, jälkiennustejakauma	precision matrix	(concentration matrix) keskittyneisyysmatriisi
posterior probability	posterioritodennäköisyys, jälkitodennäköi-	precision parameter	(concentration parameter) keskittyneisyysparametri
		predictable process	ennakoitava prosessi
		predicted value	ennustettu arvo, ennuste ₃
		prediction	ennuste ₂ , ennustami-

²⁸Aikasarja-analyysisessä kirjallisuudessa termejä *power spectrum* ja *spectral density* voidaan käyttää synonyymeinä tai hieman eri merkityksissä.

²⁹Koneoppimisen alalla sanaa *precision* käytetään tässä merkityksessä.

nen	principal axis factoring pääakselifaktorointi
prediction error ennustevirhe	principal component analysis (PCA) pääkomponenttianalyysi
prediction interval ennusteväli	principal component regression pääkomponenttiregressio
prediction residual ennustejäännös	principal coordinate pääkoordinaatti
prediction residual sum of squares (PRESS) ennustejäännösten neliösumma	principal stratification pääositus
predictive analytics ennustava analytiikka	prior distribution priorijakauma, esijakauma
predictive distribution ennustejakauma	prior elicitation (elicitation of prior) priorin koostaminen, esijakauman koostaminen
predictive inference ennustava päättely, prediktiivinen päättely	prior predictive distribution esiennustejakauma
predictive mean matching ennustekeskia arvokaltaistus	prior probability (a priori probability) prioritodennäköisyys, esitodennäköisyys
predictive value ennustearvo	probabilistic inference todennäköisyyspäättely
predictor ennustin, ennustava muuttuja	probability todennäköisyys
preferential attachment process valikoivan kasvun prosessi	probability calculus todennäköisyyslaskenta
preregistration ennakkorekisteröinti	probability density function (PDF) tiheysfunktio, todennäköisyystiheys
prescriptive analytics ohjaava analytiikka	probability distribution todennäköisyysjakauma
pre-specified significance level ennalta asetettu merkitsevyystaso	probability generating function todennäköisyyssemäfunktio, todennäköisyydet generoiva funktio
PRESS (prediction residual sum of squares)	probability integral transformation kertymäfunktio muunnos
pre-treatment covariate käsitellyä edeltävä kovariaatti	probability interval todennäköisyysväli
prevalence vallitsevuus	probability judgement todennäköisyysarvio
prevention trial ehkäisykoe	probability kernel todennäköisyysydin
preventive fraction (PF) ehkäisyosuus	probability limits todennäköisyyden rajoitukset
prewhitening esivalkaisu	
price index hintaindeksi	
primary outcome päävaste, ensisijainen vaste	
primary sampling unit (PSU) ensiasteen otantayksikkö	
principal axes decomposition pääakselihajotelma	

syysrajat	profile likelihood profiiliuskottavuus
probability mass function (point probability function) pistetodennäköisyysfunktio	profile log-likelihood profiili-loguskottavuus
probability measure todennäköisyysmitta	profile plot profiilikuvio
probability model (stochastic model) todennäköisyysmalli	profiling profilointi
probability plot todennäköisyyskuvio	prognostic factor ennustetekijä
probability proportional to size (pps) kokoon verrannollinen todennäköisyys	prognostic model ennustemalli
probability sampling (random sampling) todennäköisyysotanta, satunnaisotanta	projection projektio
probability space todennäköisyysavaruus	projection plot projektiokuvio
probability theory todennäköisyysteoria	projection pursuit projektion haku
probability weighting todennäköisyyspainotus	projection pursuit regression optimiprojektio regressio
probable (likely) todennäköinen	propagation of error (error propagation) virheen kasautuminen, virheen leviäminen
probable error todennäköinen virhe	propensity propensiteetti, alttius
probit model probit-malli	propensity matching propensiteettikaltaistus, alttiuskaltaistus
probit transformation probit-muunnos	propensity score propensiteettiluku, alttiysluku
procedural bias menettelyharha	propensity weighting propensiteettipainotus, alttiuspainotus
processing error käsittelyvirhe	proper imputation kelvollinen imputointi
Procrustes analysis Prokrustes-analyysi	proper prior aito prior
product integral tulointegraali	proportion osuus
product limit estimator (Kaplan–Meier estimator) rajatuloestimaattori, Kaplanin–Meierin estimaattori	proportional allocation suhteellinen kiintiöinti
product measure tulomitta	proportional hazards model (Cox model) verrannollisten uhkien malli, Coxin malli
product moment tulomomentti	proportional odds model verrannollisten vastasuhteiden malli
product moment correlation tulomomenttikorrelaatio	proportional sampling suhteellinen otanta
product space tuloavaruus	proposal distribution (candidate distribution) ehdotusjakauma, ehdokasjakauma
profile analysis profiilianalyysi	prospect theory prospektiteoria, näkymäteoria

prospective study prospektiivinen tutkimus
protected data suojattu tieto
protocol protokolla
protocol violation protokollariike
proximity matrix läheisyysmatriisi
proxy variable sijaismuuttuja
proxy variance sijaisvarianssi
pruning karsinta
pseudo-likelihood pseudouskotavuus
pseudo-random number pseudosatunnaisluku, näennäissatunnaisluku
PSU (primary sampling unit)
psychometrics psykometria, psykometriikka
publication bias julkaisuharha
pure natural direct effect (PNDE) puhdas luonnollinen suora vaikutus
pure natural indirect effect (PNIE) puhdas luonnollinen epäsuora vaikutus
purposeful selection harkintavalinta
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 kvadranttiotanta

quadratic kvadraattinen, neliöllinen
quadratic discriminant function kvadraattinen erottelufunktio, neliöerottelufunktio
quadratic estimator kvadraattinen estimaattori, neliöestimaattori
quadratic form neliömuoto
quadratic loss function neliötappiofunktio
quadratic variation kvadraattinen vaihtelu, neliövaihtelu
qualitative laadullinen, kvalitatiivinen
qualitative comparative analysis (QCA) laadullinen vertaileva analyysi
quality adjusted life years (QALY) laatuelinvuosi
quality assurance laadunvarmistus
quality control laadunvalvonta
quality control chart laadunvalvontakortti
quantal dose-response kaksiarvoinen annos-vaste
quantile (fractile) kvantiili, fraktiili
quantile function kvantiilifunktio
quantile-quantile plot (Q-Q plot) kvantiili-kvantiilikuvio
quantile regression kvantiili-regressio
quantitative määrällinen, kvantitatiivinen
quantum probability kvanttitoennäköisyys
quarterly data neljännesvuosiaineisto
quartile kvartiili
quartile deviation kvartiilipoik-

keama
quartimax rotation kvartimax-
 rotaatio, kvartimax-kierto
quasi-experiment kvasikoe
quasi-independence kvasiriippu-
 mattomuus
quasilikelihood kvasiuskottavuus
quasisymmetry 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
 R^2 (coefficient of determination)
 R^2
radial function radiaalifunktio
radial plot sädekuvio
Radon–Nikodym derivative
 Radonin–Nikodymin derivaatta
raking (iterative proportional fit-
 ting) haravointi
Rand index Rand-indeksi
random satunnainen
random allocation satunnaisjako
random coefficient satunnaisker-
 roin
random component satunnais-
 komponentti
random digit dialing puhelin-
 numeron satunnaisvalinta
random effect satunnaisvaikutus
random effects model satun-

naisvaikutusmalli
random error satunnaisvirhe
random event satunnaistapahtu-
 ma
random field satunnaiskenttä
random forest satunnaismetsä
random function satunnaisfunk-
 tio
random graph satunnaisverkko
random-effects model satun-
 naisvaikutusmalli
randomization satunnaistus
randomization test satunnais-
 tustesti
randomized blocks design sa-
 tunnaistettujen lohkojen asetelma
randomized clinical trial (RCT)
 satunnaistettu kliininen koe
randomized decision rule sa-
 tunnaistettu päätössääntö
randomized experiment
 (randomized trial) satunnaistet-
 tu koe
randomized response satunnais-
 tettu vastaus
randomized test satunnaistettu
 testi
randomized trial (randomized
 experiment) satunnaistettu koe
random matrix satunnaismatrii-
 si
random measure satunnaismitta
randomness satunnaisuus
random number satunnaisluku
random number generator sa-
 tunnaislukugeneraattori, satun-
 naislukukehitin
random phenomenon satunnais-
 ilmiö
random process (stochastic pro-
 cess) satunnaisprosessi, stokasti-
 nen prosessi
random sample satunnaisotos

random sampling	satunnaisotanta	Rao–Blackwellization	Raon–Blackwellin tehostus
random selection	satunnaisvalinta	rate ₁	(intensity) taaajuus, tiheys
random sequence	satunnaisjono	rate ₂	(proportion) osuus, aste ₂
random set	satunnaisjoukko	rate ₃	(ratio) suhde
random tessellation	satunnais-tessellaatio, satunnaismosaiikki	rate of change	muutosvauhti, muutostahti
random variable	(random variate) satunnaismuuttuja, satunnaisuure	ratio	suhde, suhdeluku, osamäärä
random variation	(chance variation) satunnaisvaihtelu	ratio estimator	suhde-estimaattori
random vector	satunnaisvektori	rational expectations	rationaaliset odotukset
random walk	satunnaiskulku	ratio scale	suhdeasteikko
random walk with drift	suunnallinen satunnaiskulku	raw moment	raakamomentti
range ₁	vaihteluväli	raw score	raaka-arvo
range ₂	vaihteluvälin pituus	RCT	(randomized clinical trial)
range chart	vaihteluvälikortti	RDS	(respondent-driven sampling)
rank ₁	sija, sijaluku, järjestysluku	realization	toteuma, realisaatio, toteutuma
rank ₂	aste ₁	recall	(sensitivity ₁) herkkyyys, sensitiivisyys ₁ ³⁰
rank correlation	sijakorrelaatio, sijalukukorrelaatio, järjestyskorrelaatio	recall bias	muistiharha
rank correlation matrix	sijakorrelaatiomatriisi	receiver operating characteristic	(ROC) erottelukykysoitin
ranking ₁	sijavertailu, sijakisa	receiver operating characteristic curve	(ROC curve) erottelukykykäyrä, ROC-käyrä
ranking ₂	sijoitus	reciprocal transformation	käänteislukumuunnos
rankit transformation	rankitmuunnos	record ₁	tietue, kirjaus, merkintä
rank order statistic	sijatunnusluku	record ₂	ennätys
rank-randomization test	satunnaistettu sijatesti, satunnaistettu järjestyslukutesti	record linkage	tietuekytkentä
rank regression	sijalukuregressio	rectangular distribution	(uniform distribution) tasajakautuma
rank sum test	sijasummatesti	rectangular lattice	suorakaidehila
rank transformation	sijamuunnos, järjestysmuunnos	rectifying inspection	korjaava tarkastus, korjaileva laadunvalvon-

³⁰Koneoppimisen alalla termiä *recall* käytetään samassa merkityksessä kuin *sensitivity*₁ kuvattaessa kaksiarvoisen luokittelun erottelukykyä. Tiedonhankinnassa *recall* ja sen synonyymi *recall ratio* ovat merkitykseltään analogisia suomennoksenaan “saanti”.

ta	
recurrence risk	uusiutumisriski
recurrence time₁	paluu aika
recurrence time₂	uusiutumisaika
recurrent₁	palautuva
recurrent₂	uusiutuva
recursive	toistava, rekursiivinen
reduced form	pelkistetty muoto, redusoitu muoto
reduced inspection	supistettu tarkastus
reduced maximum likelihood (restricted maximum likelihood)	rajoitettu suurin uskottavuus
reference interval	viiteväli
reference limit	viiteraja
reference period	viitejakso
reference population	viitepopulaatio, viiteväestö
reference prior	viitepriori
reference value	viitearvo
reflecting barrier	heijastava reuna
reflection principle	heijastusperiaate
refusal rate	kieltäytymisosuus
regenerative	uudistuva
regime switching model	tilanvaihtomalli
regressand (response variable)	vaste, vastemuuttuja, selitettävä muuttuja
regression	regressio
regression analysis	regressioanalyysi
regression attenuation (regression dilution)	regression vaimeneminen
regression calibration	regressiokalibrointi
regression coefficient	regressiokerroin
regression curve	regressiokäyrä
regression diagnostics	regressiodiagnostiikka
regression dilution (regression attenuation)	regression laimeneminen
regression discontinuity design (RDD)	regressioepäjatkuvuusasetelma
regression estimator	regressioestimaattori
regression function	regressiofunktio
regression line	regressiosuora
regression model	regressiomalli
regression modeling	regressiomallinnus, regressiomallitus
regression specification error test (RESET)	RESET-testi
regression spline	regressiosplini
regression sum of squares	regressioneliösumma
regression surface	regressiopinta
regression through the origin	regressio origon kautta
regression to the mean	regressio keskiarvoa kohti
regressor (explanatory variable)	selittävä muuttuja, selittäjä
regret	menetys
regularization	sääntely
regularly varying	säännöllisesti vaihteleva
reinforcement learning	vahvistusoppiminen
reinsurance	jälleenvakuutus
rejectable quality level	hylättävä laatutaso
rejection error	hylkäysvirhe
rejection region (critical region)	hylkäysalue, kriittinen alue
rejection sampling	hylkäysotanta
rejective sampling	hylkäävä

otanta	funktio
relational database relaatiotietokanta	renewal process uusiutumisprosessi
relative efficiency suhteellinen tehokkuus	renewal theory uusiutumisteoria
relative frequency (proportion) osuus, suhteellinen frekvenssi	reparametrization uusparametointi, uudelleenparametointi
relative likelihood suhteellinen uskottavuus	repeatability toistuvuus
relative risk (risk ratio) riskisuhde, suhteellinen riski, RR	repeated measurements design (repeated measures design) toistomittausasetelma
relative risk reduction (RRR) suhteellisen riskin vähenemä	repeated measures design (repeated measurements design) toistomittausasetelma
relative survival (survival ratio) suhteellinen elossaolo, suhteellinen selviytyminen	repeated replications moninkertaiset toistot
relative variance (relvariance) suhteellinen varianssi	repeated experiment (repeated trial) toistokoe ₁
reliability ₁ luotettavuus	repeated survey (periodic survey) toistoluotaus
reliability ₂ reliabiliteetti	repetition toisto, uusinta, kertaus
reliability coefficient reliabiliteettikerroin	repeated trial (repeated experiment) toistokoe ₁
reliability function luotettavuusfunktio	replicability toistettavuus
reliability theory luotettavuusteoria	replicate toisinne, replikaatti
relvariance (relative variance) suhteellinen varianssi	replication toistaminen, replikaatio
remaining life expectancy jäljellä oleva elinajanodote	replication crisis toistettavuuskriisi
REML (restricted maximum likelihood)	replication experiment toistokoe ₂
Rényi divergence Rényin poikkeama	replication study tutkimuksen toisto
Rényi entropy Rényin entropia	reporting error raportointivirhe
remote sensing kaukokartoitus	representative sample edustava otos
removal method poistomenetelmä	representativeness edustavuus
renewal density uusiutumistiheys, uusiutumistaajuus	reproducibility uusittavuus
renewal distribution uusiutumisajakauma	reproducing kernel Hilbert space (RKHS) toisintavayttaminen Hilbertin avaruus
renewal function uusiutumis-	reproduction number (reproductive number) uusiutumisluku

reproductive number	response error
(reproduction number) uusiutu-	vastausvirhe
misluku	response group vastaajaryhmä
resample uudelleen poimia	response indicator vastausilmai-
resampling uudelleen poiminta,	sin, vastausindikaattori
toisto-otanta	response mechanism
rescaled sample weight skaalat-	vastausmekanismi
tu otospaino	response metameter vasteen
RESET (regression specification	muunnos
error test)	response propensity weight
residual jäännös, residuaali	vastausalttiuspaino
residual confounding jäännösse-	response rate vastausaste
koittuneisuus	response surface vastepinta
residual deviance jäännösde-	response surface methdodolo-
vianssi	gy (RSM) vastepintamethodiikka
residual diagnostics jäännös-	response time distribution
diagnostiikka	vasteaikajakauma
residual effect jäännösvaikutus	response variable (outcome vari-
residual lifetime jäljellä oleva	able) vaste, vastemuuttuja, selitet-
elinaika	tävä muuttuja
residual maximum likelihood	restricted maximum likelihood
(restricted maximum likelihood)	(REML) rajoitettu suurimman
rajoitettu suurimman uskottavuus-	uskottavuuden menetelmä, REML
den menetelmä, REML	restricted mean survival time
residual mean square jäännös-	(RMST) rajoitettu elinaikaodote
keskineliö	restricted randomization rajoi-
residual plot jäännöskuvio	tettu satunnaistus
residual standard deviation	retrospective study retrospektii-
jäännöshajonta	vinen tutkimus
residual sum of squares jään-	return level paluukvantiili
nösneliösumma	return period paluujakso
residual variance jäännösvarians-	return time paluuaika
si	reversal design käänteisasetelma
residual waiting time jäljellä	reversible jump Markov chain
oleva odotusaika	Monte Carlo (RJMCMC)
resistant estimator resistentti	vaihtelevaulotteinen MCMC,
estimaattori	RJMCMC
resolvable block design ratkea-	reversible process aikakääntyvä
va lohkoasetelma	prosessi
respondent-driven sampling	reweighting uudelleenpainotus,
(RDS) vastaajavetoinen otanta	uuspainotus
response ₁ (outcome) vaste, tulos	ridge regression harjanneregres-
response ₂ (answer) vastaus	sio, harjaregressio
	ridit analysis ridit-analyysi

right censoring	sensurointi oikealta
right truncation	katkaisu oikealta, typistys oikealta
risk	riski
risk averse	riskiä kaihtava, riskiä karttava
risk aversion	riskin kaihtaminen, riskin karttaminen
risk difference	riskiero
risk factor	riskitekijä
risk function	riskifunktio
riskless asset	riskitön arvopaperi
risk neutral measure	riskineutraali mitta
risk prone	riskihakuinen
risk proneness	riskihakuisuus
risk ratio	(relative risk) riskisuhde, suhteellinen riski, RR
risk set	riskijoukko
risk set sampling	riskijoukkopointiminta
risk theory	riskiteoria
RJMCMC	(reversible jump Markov chain Monte Carlo)
RKHS	(reproducing kernel Hilbert space)
RMST	(restricted mean survival time)
robust estimator	vakaa estimaattori, robusti estimaattori
robustness	vakaus, robustisuus
robust statistic	vakaa tunnusluku, robusti tunnusluku
ROC	(receiver operating characteristic) erottelukykysoitin
ROC curve	(receiver operating characteristic curve) erottelukykäykäyrä
rooted graph	juurellinen verkko
root mean squared error	(RMSE) jäännösvirrehajonta, RMSE
rose diagram	ruusukaavio
rotatable design	kiertoasetelma, rotaatioasetelma
rotation	kierto, rotaatio
rotation sampling	kierto-otanta
roughness penalty	rososakko
Round Robin design	kiertovuoroasetelma, round robin -asetelma
rounding	pyöristys
route sampling	reittiotanta
RR	(relative risk)
RSM	(response surface methodology)
r-statistic	r-tunnusluku
rug plot	hapsukuvio
ruin probability	vararikkotodennäköisyys
ruin problem	vararikko-ongelma
rule of three	kolmen sääntö
run₁	merkkijakso
run₂	ajo
run₃	koete ³¹
run-in period	sisäänajojakso ₂
runs test	merkkijaksotesti
RV-coefficient	RV-kerroin

S

saddle point approximation	satulapistepennäksimaatio
sample₁	otos
sample₂	näyte
sample₃	pomia
sample₄	näytteistä
sample covariance matrix	otoskovarianssimatriisi
sample design	(sampling design) otosasetelma, otanta-asetelma
sample distribution	otosjakau-

³¹Yksittäinen koe tai toisto. Koesuunnittelussa asetelmamatriisin rivi.

ma	sampling plan otantasuunnitelma
sample distribution function (ogive) otoskertymäfunktio	sampling unit otantayksikkö
sampled population (frame population) kehikkoperusjoukko	sampling without replacement otanta palauttamatta, pito-otanta
sample mean otoskeskiarvo	sampling with replacement otanta palauttaen, palautusotanta
sample moment otosmomentti	sandwich estimator kerrosestimäätori, sandwich-estimaattori
sample path otospolku, realisatio	SAR (simultaneous autoregression)
sample plan (sampling plan) otossuunnitelma, otantasuunnitelma	SARMA (seasonal autoregressive moving average process)
sample point otospiste	saturated model kyllästetty malli, saturoitu malli
sampler poimija	SAVE (sliced average variance estimation)
sample selection model (selection model) valikoitumismalli	scalability skaalautuvuus
sample size otoskoko	scalar skalaari
sample space otosavaruus	scale₁ skaala
sample standard error otoskeskivirhe	scale₂ mittakaava
sample statistic otostunnusluku, otossuure	scaled deviance skaalattu devianssi
sample survey otosluotaus, otostutkimus	scale factor skaalakerroin, skaalatekijä
sample unit otosyksikkö	scale parameter skaalaparametri
sample variance otosvarianssi	scaling skaalaus
sample weight otospaino	scan statistic liukuva tunnusluku
sampling₁ otanta, poiminta	scatter (dispersion) sironta, hajonta
sampling₂ näytteistys	scatter diagram (scatter plot) sirontakuvio, hajontakuvio
sampling bias otantaharha	scatter matrix hajontamatriisi
sampling design (sample design) otanta-asetelma, otosasetelma	scatter plot (scatter diagram) sirontakuvio, hajontakuvio
sampling distribution otantajakauma	scatterplot matrix sirontakuvio-matriisi, hajontakuviomatriisi
sampling error otantavirhe	scenario skenaarior
sampling fraction otantaosuus	sceptical prior skeptinen prior, skeptinen esijakauma
sampling frame otantakehikko	SCM (structural causal model)
sampling inspection otantatarkastus	score₁ (efficient score) skoori
sampling interval poimintaväli	score₂ pistemäärä
sampling method otantamenetelmä, poimintamenetelmä	

score function	skoorifunktio	ma	
score residual	skoorijäännös	secular trend	trendi
score statistic	skoorisuure, skooritunnusluku	seemingly unrelated regressions	(SUR) näennäisesti riippumattomat regressiot
score test	(Lagrange multiplier test) skooritesti, Lagrangen kerointesti	segmented regression	(piecewise regression) paloittainen regressio
score vector	skoorivektori	segregation₁	eriytyminen, segregatio
scoring	pisteytys	segregation₂	lohkeaminen, segregatio
screening	seulonta	segregation analysis	segregaatioanalyysi, eriytymisanalyysi
screening design	seulonta-asetelma	SEIR	(susceptible-exposed-infectious-recovered)
screening test	seulontatesti	SEIR model	SEIR-malli
SD	(standard deviation)	selection bias	valikoitumisharha, valintaharha
SE	(standard error)	selection model	(sample selection model) valikoitumismalli
series of experiments	koesarja	selection prior	valikoitumispriori, valintapriori
seasonal adjustment	kausipuhdistus	self-attention	itsehuomio
seasonal autoregressive moving average process	(SARMA) kausi-ARMA-malli	self-administered survey	omatoimiluotaus
seasonal root	kausijuuri	self-exciting process	itsekehkeytyvä prosessi
seasonal variation	kausivaihtelu	self-financing	itsensä rahoittava
secant method	sekanttimenetelmä	self-fulfilling prophecy	itsensä toteuttava ennuste
secondary outcome	toissijainen vaste	self-negating prophecy	itsensä estävä ennuste
secondary sampling unit (SSU)	toisen asteen otantayksikkö	self-organizing map	(SOM) itseorganisoituva kartta, SOM
second order efficiency	toisen kertaluvun tehokkuus	self-selected sample	itsevalikoitunut otos
second-order inclusion probability	kahden yksikön sisällymistodennäköisyys	self-similar process	itsesimilaarinen prosessi
second order stationary	(covariance stationary) kovariansistationaarinen	self-weighting sample	itsepainottuva otos
second-order source separation	lähteiden toisen kertaluvun erottelu	semi-invariant	(cumulant) kumulantti
secretary problem	sihteeriongel-	semi-Markov process	semimar-

kovprosessi	serial variation sarjavaihtelu
semi-martingale semimartingaa- li, puolimartingaali	S-estimator S-estimaattori
semi-parametric semiparametri- nen, osaparametrinen	set joukko
sensitivity ₁ herkkyys, sensitiivisyys ₁ , oikeiden positii- visten osuus	Shannon information Shannon- informaatio
sensitivity ₂ arkaluonteisuus	shape matrix muotomatriisi
sensitivity analysis herkkyysana- lyysi	shape parameter muotoparamet- ri
sensitivity curve herkkyyskäyrä	share of a stock osake
sentiment analysis tunneanaly- ysi	shock sokki
separable space separoituva ava- ruus	shock model sokkimalli
separating hyperplane erottava hypertaso	shortest probability interval lyhin todennäköisyysväli
sequence model sekvenssimalli	short-memory process lyhyt- muistinen prosessi
sequential allocation peräkkäis- allokointi	short sale lyhyeksi myynti
sequential analysis peräkkäis- analyysi	short-tailed distribution lyhyt- häntäinen jakauma
sequential model peräkkäismalli	shot-noise pistekohina
sequential probability ratio test (SPRT) peräkkäinen toden- näköisyysuhdetesti	shrinkage estimator kutistuses- timaattori
sequential design peräkkäisase- telma	sieve estimator seulaestimaattori
sequential experiment peräk- käiskoe	sigma-algebra (sigma-field) sigma-algebra, sigma kenttä
sequential sampling peräkkäis- otanta, peräkkäispoiminta	sigma-field (sigma-algebra) sigma- kenttä, sigma-algebra
sequential test peräkkäistesti	sigmoid curve sigmoidikäyrä, S- käyrä
sequential updating vaiheittainen päivittäminen	signal processing signaalinkäsit- tely
serial correlation (autocorrelation) autokorrelaatio, sarjakorrelaatio	signal to noise ratio signaali-ko- hinasuhde
serial dilution assay sarjalai- mennuskoe	signed rank test merkkisijatesti, merkkijärjestystesti
serial measurements mittaussar- ja	signed root-transformation etumerkkinen neliöjuurimuunnos
	significance merkitsevyys
	significance level (level of signi- ficance) merkitsevyystaso
	significance test (test of signi- ficance) merkitsevyystesti
	sign test merkkitesti
	silhouette plot siluettikuvio

similarity coefficient	samanlaisuuskerroin	simultaneous inference	saman- aikainen päättely, simultaaninen päättely
similarity matrix	samanlaisuus- matriisi	single-arm study	yhden linjan tutkimus
similarity measure	samanlai- suusmitta	single-blind experiment	yksöissokkokoe
similar matrices	similaarit mat- riisit	single imputation	kertapaikkaus, kertaimputointi
similar regions	similaariset alu- eet	single-index model	yhden indek- sin malli
similar test	similaarinen testi	single-linkage clustering	yksikytkösryhmittely
simple hypothesis	yksinkertainen hypoteesi, piste- hypoteesi	single-stage sampling	yksiasteotanta
simple linear regression	yhden selittäjän regressio	singular distribution	singulaari- nen jakauma
simple random sampling (SRS)	yksinkertainen satunnaisotanta	singular matrix	singulaarinen matriisi, vajaa-asteinen matriisi
simplex algorithm	simpleksi- algoritmi	singular value decomposition	(SVD) singulaariarvohajotelma
Simpson's paradox	Simpsonin paradoksi	SIR	(susceptible-infectious- recovered)
simulated annealing	simuloitu hehkutus	SIR model	SIR-malli
simulated envelope	simuloitu verho	six sigma principle	kuuden sig- man periaate
simulation	simulaatio, simulointi	size-biased	kokovalikoitunut
simulator	simulaattori	size of test	testin koko
simultaneous autoregression		skew distribution	vino jakauma
(SAR)	simultaaninen autoregres- sio	skewness	vinous
simultaneous confidence inter- vals	samanaikaiset luottamusvä- lit, simultaaniset luottamusvälit	skew-normal distribution	vinonormaali-jakauma
simultaneous diagonalization	samanaikainen diagonalisointi	skew-symmetric matrix	vinosymmetrinen matriisi
simultaneous equations bias	simultaanivyhtälöharha	sliced average variance estima- tion (SAVE)	viipaloitu keskiva- rianssiestimointi
simultaneous equations model	simultaanivyhtälömalli	sliced inverse regression	viipaloitu käänteinen regressio
simultaneous estimation	sa- manaikainen estimointi, simultaa- ninen estimointi	slice sampler	viipalepaimija
		sliding square plot	liukuneliöku- vio
		slope	kulmakerroin

slope ratio assay jyrkkyyssuhde-määrittäminen

slowly varying function hitaasti vaihteleva funktio

small area estimation pienalue-estimointi

smooth transition autoregressive model (STAR) tasaisen siirtymän autoregressiivinen malli, STAR-malli

smoother silotin, tasoitin

smoothing silotus, tasoitus

smoothing spline silottava splini

SMR (standardized mortality ratio)

snowball sampling lumipallo-otanta

snowflake diagram lumihiutalekuvio

social data science yhteiskuntatietiede

social statistics yhteiskuntatilastotiede

soft computing pehmolaskenta

softmax function softmax-funktio

sojourn time viipymisaika

solvency solvenssi

SOM (self-organizing map)

space avaruus

space-time cluster tila-aikaryvä

spaghetti plot spagettikuvio

sparse matrix harva matriisi

spatial data spatiaalinen data, aluedata, paikkadata

spatial experiment spatiaalinen koe

spatial median spatiaalinen mediaani

spatial pair-potential spatiaalinen paripotentiaali

spatial point pattern tilapistekuvio

spatial point process spatiaalinen pisteprosessi, tilapisteprosessi

spatial rank spatiaalinen sijaluku

spatial sign spatiaalinen merkki, suuntavektori

spatial statistics spatiaalinen tilastotiede, aluetilastotiede, paikkatilastotiede

speech recognition puheentunnistus

species abundance lajien runsaus

specification bias määrittelyharha

specification error määrittelyvirhe

specificity (true negative rate) tarkkuus₂, spesifisyys, oikeiden negatiivisten osuus

spectral analysis spektrianalyysi

spectral decomposition (eigendecomposition) spektriha-jotelma, ominaisarvohajotelma³²

spectral density spektritiheys³³

spectral distribution function spektrikertymäfunktio

spectral gap spektrivälily

spectral radius spektrisäde

spectral representation spektrisyys

spectrum spektri

speech recognition puheentunnistus

spherical coordinate pallokoordinaatti

spherical distribution₁ pallosymmetrinen jakauma

³²Symmetriselle matriisille pääakselihajotelma.

³³Aikasarja-analyttisessä kirjallisuudessa termejä *spectral density* ja *power spectrum* voidaan käyttää synonyymeinä tai hieman eri merkityksissä.

spherical distribution ₂	pallopin-tajakauma	stacking	pinoaminen
spherical variable ₁	pallosymmet-rinen muuttuja	stacking model	pinomalli
spherical variable ₂	pallopinta-muuttuja	staggered entry	porrasteinen sisääntulo
sphericity test	pallopointatesti	staircase design	porrasasetelma
spike and slab	piikki ja laatta	staircase method	(up-and-down method) ylös-alasmenetelmä, tik-kausmenetelmä
spike and slab distribution	piikki ja laatta -jakauma	standard deviation	keskihajonta
spin bias	spinnausharha	standard error	keskivirhe
spline	splini	standardization ₁	vakiointi
split-half method	puolitusmene-telmä	standardization ₂	standardointi, normitus
split-plot design	osaruutuasetel-ma	standardized coefficient	stan-dardoitu kerroin
split-unit design	osayksikköase-telma	standardized effect size	stan-dardoitu vaikutuskoko ³⁴
spread (dispersion)	hajonta	standardized mortality ratio	(SMR) vakioitu kuolevuussuhde, SMR
spreadsheet	laskentataulukko	standardized residual	standar-doitu jäännös
spurious correlation	näennäis-korrelaatio	standardized variable	standar-doitu muuttuja, normitettu muut-tuja
spurious precision	näennäistäs-mällisyys	standard normal distribution	standardinormaalijakauma, normi-tettu normaalijakauma
square lattice	neliöhila	standard population	vakioväestö
square matrix	neliömatriisi	standard rate	vakioilmaantuvuus
square root rule	neliöjuurisään-tö	STAR	(smooth transition auto-regressive model)
square root transformation	ne-liöjuurimuunnos	starburst plot	(bagplot) tähtisa-dekuvio
square summable	neliöllisesti summautuva	star plot	tähtikuvio
SRS (simple random sampling)		state space	tilajoukko, tila-avaruus
SSU (secondary sampling unit)		stationary distribution	statio-naarinen jakauma
stable distribution	stabiili ja-kauma	stationary population	stationaa-
stable population	stabiili väestö		
stable state	stabiili tila		
stable-unit-treatment-value assumption	(SUTVA), stabiili yksilövaste, vakaa yksilövaste		

³⁴Ks. alaluku 2.2.

rinen väestö	statistical test tilastollinen testi
stationary process stationaari- nen prosessi	statistician tilastotieteilijä
stationary subspace analysis stationaarinen aliavaruusanalyysi	statistics₁ tilasto
statistic tunnusluku, havaintosuure	statistics₂ (statistical science) tilastotiede
statistical₁ tilastollinen	statistics anxiety tilastoahdistus
statistical₂ tilastotieteellinen	steady-state distribution tasapainotilajakauma
statistical authority tilastoviranomainen	stem-and-leaf diagram varsilehtikuvio, lehtikuvio
statistical ensemble (ensemble) tilastollinen kooste, kooste	step-down procedure alasaskeltava menettely
statistical expert system tilastollinen asiantuntijajärjestelmä	stepped wedge design porraskiila-asetelma
statistical graphics tilastografiikka	stepwise regression askeltava regressio
statistical hypothesis tilastollinen hypoteesi	stereology stereologia
statistical inference tilastollinen päättely	stick breaking process tikunkatkaisuprosessi
statistical learning tilastollinen oppiminen	stochastic stokastinen, satunnainen
statistical literacy tilastolukutaito	stochastic approximation stokastinen approksimaatio
statistically discernible tilastollisesti erottuva	stochastic block model stokastinen lohkomalli
statistically significant tilastollisesti merkitsevä	stochastic convergence stokastinen suppeneminen
statistical machine learning tilastollinen koneoppiminen	stochastic dependence stokastinen riippuvuus
statistical model tilastollinen malli	stochastic imputation satunnaispaikkaus, stokastinen imputointi
statistical reasoning tilastollinen järkeily, tilastollinen perustelu	stochastic integral stokastinen integraali
statistical significance tilastollinen merkitsevyys	stochastic matrix stokastinen matriisi
statistical science (statistics ₂) tilastotiede	stochastic model (probabilistic model) todennäköisyysmalli, stokastinen malli
statistical service tilastotoimi	stochastic optimization stokastinen optimointi
statistical space tilastollinen avaruus	stochastic ordering stokastinen järjestys

stochastic process (random process) satunnaisprosessi, stokastinen prosessi	rakenneaikasarjamalli
stochastics stokastiikka	structural vector autoregression rakenteellinen vektoriautoregressio
stochastic volatility stokastinen volatiliteetti, satunnaisheilunta	structural zero rakenteellinen nolla
stopping rule pysäytyssääntö	studentized difference in betas t -kerroinero
stopping time (optional time) pysäytyshetki, valintahetki	studentized difference in fits t -sovite-ero
stopword yleissana, stop-sana	studentized range statistic t -vaihteluvälisuure, t -vaihteluvälitunnusluku
St. Petersburg paradox Pietarin paradoksi	studentized residual t -jäännös
stratification ositus	Student's distribution (t -distribution) Studentin jakauma, t -jakauma
stratified model ositettu malli	study tutkimus
stratified sampling ositettu otanta	study population tutkimuspopulaatio, tutkimusväestö
stratum ³⁵ osite	study unit tutkimusyksikkö
strict exogeneity aito eksogeenisuus	stylized facts tyylitelty ominaispiirteet, olennaiset piirteet
strictly stationary aidosti stationaarinen, vahvasti stationaarinen	stylometry stylometria, tyylitilastotiede
strike price rajahinta	subdensity function alitiheysfunktio
stripes plot viivakoodikuvio, juovakuvio	subdistribution alijakauma
strong consistency vahva tarkentuvuus	subdistribution function alikertymäfunktio
strong convergence (almost sure convergence) vahva suppeneminen, melkein varma suppeneminen	sub-Gaussian distribution aligaussinen jakauma
strong exogeneity vahva eksogeenisuus	subgroup analysis alaryhmäanalyysi
strong law of large numbers vahva suurten lukujen laki	subjective probability subjektiivinen todennäköisyys
structural equation model rakenneyhtälömalli	subject-specific model yksilökohtainen malli
structural change rakennemuutos	submartingale submartingaali, alimartingaali
structural causal model (SCM) rakenteellinen kausaalimalli	subexponential aliekspONENTTI-
structural time series model	

³⁵Monikko on *strata*.

nen	piminen
subsample osaotos, aliotos	support kantaja
subset osajoukko	support function (log-likelihood function) tukifunktio, log-uskottavuusfunktio
subspace aliavaruus	support vector machine (SVM) tukivektorikone
substantive model ilmiömalli	suppressor variable alistaja, alistava muuttuja
success probability onnistumistodennäköisyys	SUR (seemingly unrelated regressions)
sufficiency tyhjentävyys	surprisal value (s -value) yllätysarvo, s -arvo
sufficient dimension reduction tyhjentävä dimension pienennys, tyhjentävä ulottuvuuksien vähentäminen	surrogate outcome sijaisvaste
sufficient statistic tyhjentävä tunnushuku	surrogate variable sijaismuuttuja
sum contrast summakontrasti	survey ³⁶ luotaus, kartoitus, kyselytutkimus
summary statistic tiivistävä tunnushuku	survey instrument luotausinstrumentti
sunflower plot auringonkukkakuvio	survey methodology luotausmetodologia
superexogeneity supereksogeneisuus	survey mode luotautapa
superefficient ylitehokas	survey round kyselykierros
superiority trial paremmuuskoe	survey sampling luotausotanta
super learner superoppija	survey statistics luotaustilastotiede
superleverage supervipuvoima	survival analysis (lifetime data analysis) elinaika-analyysi, elosaoloanalyysi, kestoanalyysi
supermartingale supermartingaali, ylimartingaali	survival cohort selviytyjäkohortti
supernormality supernormaalisuus	survival curve välttökäyrä, elosaolokäyrä, kestokäyrä
superpopulation model superpopulaatiomalli	survival function välttöfunktio, elossaolofunktio, kestofunktio
superposition of processes prosessien summaus, prosessien kerrostaminen	susceptible-exposed-infectious-recovered ³⁷ (SEIR) altis-tartunnan saanut-tartuttava-toipunut, SEIR
supersaturated design ylikyllästetty asetelma	
supervised ohjattu	
supervised classification ohjattu luokittelu	
supervised learning ohjattu op-	

³⁶Ks. alaluku 2.11.

³⁷Adjektiivien *infectious* ja *recovered* paikalla käytetään myös adjektiiveja *infected* ja *removed*.

susceptible-infectious-recovered ³⁷ (SIR) altis-tartuttava-toipunut, SIR	täeksponentti
SUTVA (stable-unit-treatment-value assumption)	tail of a distribution jakauman häntä
s-value (surprisal value) <i>s</i> -arvo, yllätysarvo	tails klaava
SVD (singular value decomposition)	tapering kartiosilotus
switch-back design vaihtoasetelma	TAR (threshold autoregression) kynnysautoregressio
switching regression model vaihtoregressiomalli	targeted maximum likelihood estimation kohdistettu suurimman uskottavuuden menetelmä
symmetric distribution symmetrinen jakauma	target population kohdeperusjoukko, kohdepopulaatio, kohdeväestö
symmetric sampling symmetrisen otanta	target trial tavoitekoe
symmetrized scatter matrix symmetroituu hajontamatriisi	target trial emulation tavoitekokeen jäljittely
synthetic data synteettinen aineisto	taxonomy taksonomia
synthetic estimator synteettinen estimaattori	<i>t</i>-distributed stochastic neighborhood embedding (<i>t</i> -SNE) <i>t</i> -ympäristöupotus, <i>t</i> -jakaumaan perustuva stokastinen naapurustouputus, <i>t</i> -naapurustouputus
systematic allocation systemaattinen jako, systemaattinen allokointi	<i>t</i>-distribution (Student's distribution) <i>t</i> -jakauma, Studentin jakauma
systematic error (bias) systemaattinen virhe, harha	technometrics teknometria
systematic review järjestelmällinen katsaus, systemaattinen katsaus	temporal aggregation aika-aggregointi
systematic sampling systemaattinen otanta, tasaväliotanta	temporal homogeneity (time-homogeneity) aikahomogeenisuus
	tensor tensori
	tensorial data (tensor data) tensoriaineisto
	tensor data (tensorial data) tensoriaineisto
	tensor-variate tensoriarvoinen, tensorimuuttujainen
	terminal event (endpoint) pääte-tapahtuma
	test testi
	test data (test set) testiaineisto

T

table taulukko³⁸

tag cloud (word cloud) sanapilvi

tail exponent (tail index) häntäeksponentti

tail index (tail exponent) häntä-

³⁸Tietojenkäsittelytieteessä on tietokantojen yhteydessä termille vakiintunut suomenos "taulu".

test of significance (significance test) merkitsevyystesti	analyysi
test series testisarja	time-series econometrics aika-sarjaekonometria
test set (test data) testiaineisto	time-to-event tapahtuma-aika
test statistic testisuure, testitunnusluku	time-to-event analysis tapahtuma-aika-analyysi
tetrachoric correlation tetrakorinen korrelaatio	time-varying aikamuuttuva
tetrad difference nelierotus	TNDE (total natural direct effect)
text mining tekstinlouhinta	TNIE (total natural indirect effect)
theorem lause, teoreema	tobit-model tobit-malli
thinning harvennus	Toeplitz matrix Toeplitz-matriisi
thin plate spline ohutlevysplini	tolerance toleranssi
thin-tailed distribution ohuthäntäinen jakauma	tolerance interval toleranssiväli
three series theorem kolmen sarjan lause	tontine tontiini
three-stage least squares kolmivaiheinen pienin neliösumma	topic modeling aihemallitus
three-stage sampling kolmias-teotanta	total kokonaismäärä
threshold autoregression (TAR) kynnysautoregressio, TAR	total correlation kokonaiskorrelaatio
threshold model kynnysmalli	total natural direct effect (TNDE) luonnollinen suora kokonaisvaikutus
threshold value kynnysarvo	total natural indirect effect (TNIE) luonnollinen epäsuora kokonaisvaikutus
tie sidos, tasahavainto	total probability kokonaistodennäköisyys
tight ordering tiukka järjestys	total sum of squares kokonaisneliösumma
tightened inspection tiukennettu tarkastus	total variation distance (variational distance) kokonaisvaihteluetäisyys
time-dependent aikariippuva	trace jälki
time-dependent covariate aikariippuva selittäjä	training data (training set) opetusaineisto
time-dependent distribution aikariippuva jakauma	training set (training data) opetusaineisto
time domain aika-alue	trait piirre
time-frequency distribution aika-taajuusjakauma	trajectory kehityspolku, trajektorii
time-homogeneity (temporal homogeneity) aikahomogeenisuus	transfer function siirtofunktio
time-independent aikariippumaton, ajasta riippumaton	transfer learning siirto-oppiminen
time series aikasarja	
time-series analysis aikasarja-	

transformation	muunnos	trim and fill method	leikkuu- täyttömenetelmä
transient state	väistyvä tila	trimmed mean	leikattu keskiar- vo, trimmattu keskiarvo
transition intensity	(transition rate) siirtymäintensiteetti, siirty- mätaajuus	trimodal	kolmihiippuinen
transition matrix	siirtymämatrii- si	tripartite	kolmijakoinen
transition probability	siirtymä- todennäköisyys	triples test	kolmikkotesti
transition rate	(transition inten- sity) siirtymätaajuus, siirtymäin- tensiteetti	trivariate	kolmimuuttujainen, kol- men muuttujan, kolmiulotteinen
translation	translaatio, siirto	Trojan square	troijalainen neliö
transmission probability	tart- tuntatodennäköisyys	true negative rate	(specificity) tarkkuus ₂ , spesifisyys, oikeiden negatiivisten osuus
transmission rate ₁	tarttuvuus	true positive rate	(sensitivity ₁) herkkyys, sensitiivisyys ₁ , oikeiden positiivisten osuus
transmission rate ₂	tiedonsiirto- nopeus	truncated distribution	katkais- tu jakauma, typistetty jakauma
transportability	siirrettävyys	truncated observation	katkais- tu havainto, typistetty havainto
transpose	transpoosi	truncation	kokonaisosa
trapdoor variable	ansaluukku- muuttuja	t-SNE	(<i>t</i> -distributed stochastic neighborhood embedding)
traveling salesman problem	kauppamatkustajan ongelma	t-statistic	<i>t</i> -suure, <i>t</i> -tunnusluku
treatment	käsittely, hoito	t-test	<i>t</i> -testi
treatment arm	hoitolinja, käsit- telyhaara	tuple	lista ₂
treatment contrast	käsittely- kontrasti	turning point	taitekohta, käänne- kohta
treatment difference	hoitoero	twin study	kaksostutkimus
treatment group	hoitoryhmä, käsittelyryhmä	two-armed bandit	kaksikätkäinen rosvo
tree diagram	puukaavio, puudia- grammi	two-by-two table	(fourfold table) 2×2 -taulukko, nelikenttä
tree model	puumalli	two-phase sampling	(double sampling) kaksivaiheotanta
trend	trendi, kehityssuunta, pitkä- aikaismuutos	two-sided alternative hypothe- sis	kaksisuuntainen vastahypotee- si, kaksisuuntainen vaihtoehtoinen hypoteesi
trial	(experiment) koe	two-sample t test	kahden otok- sen <i>t</i> -testi
triangular distribution	kolmio- jakauma	two-sided test	(two-tailed test) kaksisuuntainen testi, kaksitahoi-
triangular matrix	kolmiomatriisi		
triangular table	kolmiotaulukko		
triangular test	kolmiotesti		

nen testi
two-stage least squares kaksi-
vaiheinen pienin neliösumma
two-stage sampling kaksiaste-
otanta
two-tailed test (two-sided test)
kaksisuuntainen testi, kaksitahoi-
nen testi
two-way analysis of variance
kaksisuuntainen varianssianalyysi
two-way classification kaksi-
suuntainen luokittelu
two-way contingency table kak-
sisuuntainen ristitaulukko, kaksi-
ulotteinen ristitaulukko
type I error (rejection error) hyl-
käysvirhe
type II error (acceptance error)
hyväksymisvirhe

U

ultrametric tree ultrametrisen
puu
umbrella problem sateenvarjo-
ongelma
umbrella trial sateenvarjokoe
UMP (uniformly most powerful)
UMPU (uniformly most powerful
unbiased)
unadjusted estimator
vakioimaton estimaattori
unbalanced data epätasapainoi-
nen aineisto
unbalanced design epätasapai-
noinen asetelma
unbiased harhaton
uncertainty epävarmuus
uncertainty interval epävar-
muusväli
uncertainty quantification epä-
varmuuden määrittäminen
unconditional distribution eh-

dollistamaton jakauma
uncorrelated korreloimaton
unemployment rate työttömyys-
osuus, työttömyysaste
underadjustment vajaavakiointi
undercount (underenumeration)
laskennan alipeitto
undercoverage alipeitto
underdispersion alihajonta
underenumeration (undercount)
laskennan alipeitto
underfitting vajaasovitus
underidentification ali-identifioi-
tuvuus
undersampling₁ alipoiminta
undersampling₂ alinäytteistys
undirected graph suuntaamaton
verkko
unequal probability sampling
otanta vaihtelevin todennäköi-
syyksin
unidentifiability identifioimatto-
muus
uniform convergence tasainen
suppeneminen
uniform correlation tasakorre-
laatio
uniform distribution tasajakau-
ma
uniformly integrable tasaisesti
integroituva
uniformly minimum risk kaut-
taaltaan pienin riski
**uniformly most accurate inter-
val** kauttaaltaan tarkin väli
uniformly most powerful
(UMP) kauttaaltaan voimakkain
**uniformly most powerful un-
biased** (UMPU) kauttaaltaan
voimakkain harhaton
uniform sampling fraction
vakio-otantaosuus
unimodal yksihiippuinen

union yhdiste, unioni
union-intersection test yhdiste-leikkaustesti
unipartite yksijakoinen
uniqueness yksikäsitteisyys
unit yksikkö
unit circle yksikköympyrä
unit interval yksikköväli
unit non-response yksikkövastauskato
unit root yksikköjuuri
unit-root test yksikköjuuritest
univariate yksimuuttujainen, yhden muuttujan, yksiulotteinen
universe (population) perusjoukko, populaatio
unlabeled data nimiöimätön data
unobserved confounder piilosekoittaja
unordered set järjestämätön joukko
unreduced design pelkistymätön asetelma
unsupervised ohjaamaton
unsupervised classification ohjaamaton luokittelu
unsupervised learning ohjaamaton oppiminen
up-and-down method ylös-alasmenetelmä, 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 muotoinen jakauma
U-statistic *U*-tunnusluku
utility hyöty, utiliteetti

utility function hyötyfunktio
utility theory hyötyteoria, utiliteettiteoria

V

vague convergence löyhä suppeneminen
vague prior heikko prior, laakea prior
validation validointi, vahvistus
validation data (validation set) validointiaineisto, vahvistusaineisto
validation set (validation data) validointiaineisto, vahvistusaineisto
validity validiteetti, pätevyys
Value at Risk (VaR) tappioarvo, VaR
value index arvoindeksi
value of information informaatioarvo
VaR (Value at Risk)
VAR (vector autoregression)
variability (variation) vaihtelu
variable (variate) muuttuja
variable selection muuttujavalinta
variance varianssi
variance analysis (analysis of variance) varianssianalyysi
variance component varianssikomponentti
variance-covariance matrix (covariance matrix) kovarianssimatriisi
variance inflation factor (VIF) varianssin inflaatiokerroin, VIF
variance stabilizing transformation varianssin vakioiva muunnos
variate (variable) muuttuja

variation (variability) vaihtelu
variational distance (total variation distance) variaatioetäisyys
variational inference variaatiopäättely
varimax rotation varimax-rotatio, varimax-kierto
variogram variogrammi
vec operator vektorointioperaattori
vector autoregression (VAR) vektoriautoregressio
vector time series vektoriaikasarja
vector-variate vektoriarvoinen, vektorimuuttujainen
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
volunteer sample vapaaehtoisnäyte, vapaaehtoisotos
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 eksogeneisuus
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 painoarvo
white noise valkoinen kohina
wide sense stationary (covariance stationary) kovarianssistationaarinen, laajassa mielessä stationaarinen
Wiener process Wiener-prosessi
Wilcoxon test Wilcoxonin testi
wild bootstrap villi uusio-otanta
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

X

XGBoost (extreme gradient boosting)

Y

Yates' correction Yatesin kor-
jaus

Yule–Walker equations Yulen–
Walkerin yhtälöt

Z

zero-altered model³⁹ (hurdle
model) rimamalli

zero-inflated distribution nolla-
paisunut jakauma

zero-inflation nollapaisunta

zero matrix nollamatriisi

zero-sum game nollasummapeli

zero-truncated distribution
nollakatkaistu jakauma, nollaty-
pistetty jakauma

zeta distribution zetajakauma

z-score *z*-arvo

z-test *z*-testi

z-transformation *z*-muunnos

³⁹*Zero-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
- affiini muunnos** affine transformation
- affinisti ekvivariantti** affine equivariant
- affinisti invariantti** affine invariant
- agenttipohjainen malli** agent-based model
- aggregaatti-imputointi** aggregate imputation
- aggregointi** aggregation
- aggregoitu aineisto** aggregate data
- 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
- aineiston syvyys** data depth
- aineistovetoinen** data driven
- aito eksogeenisuus** strict exogeneity
- aito priori** proper prior

annuiteetti	annuity		
anomalia	anomaly		
ansaluukkumuuttuja	trapdoor variable		
antiteettinen	antithetic		
antitoninen regressio	antitonic regression		
approksimaatio	approximation ₂		
approksimointi	approximation ₁		
apuinformaatio	ancillary information		
apumuuttuja	auxiliary variable		
aputunnusluku	ancillary statistic		
arbitraasi	arbitrage		
aritmeettinen jakauma	arithmetic distribution		
aritmeettinen keskiarvo	arithmetic mean		
arkaluonteisuus	sensitivity ₂		
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	rank ₂		
aste₂	proportion, rate ₂		
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 infer-

ence
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 distribution
binomikerroin binomial coefficient
binomikoe binomial experiment, binomial trial
binomitesti binomial test
binaarimuuttuja binary variable, dichotomous variable
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 Bonfer-

roni 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 preprocessing

datan jalostaminen data wrangling

datan kurkistelu data peeking

datan linkittäminen data link-

age

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

datatieteilijä data scientist

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	diskreetti jakauma	discrete distribution
diagnostinen tunnusluku	diagnostic statistic	diskreetti muuttuja	discrete variable
diagonaalimatriisi	diagonal matrix	diskriminantti	discriminant
diagrammi	diagram	divergenssi	divergence ₁
differenssiyhtälö	difference equation	diversiteetti-indeksi	index of diversity
differensointi	differencing	do-laskenta	do-calculus
differentiaalinen tietosuoja	differential privacy	do-operaattori	do-operator
differentiaaliyhtälö	differential equation	D-optimaalisuus	<i>D</i> -optimality
differentioituvuus	differentiability	d-separaatio	<i>d</i> -separation
differointi	differencing	duaalilause	dual theorem
diffuusiaindeksi	diffusion index	duaalisuus	duality
diffuusioprosessi	diffusion process	dynaaminen kiintiöinti	dynamic treatment allocation
diffuusi prior	diffuse prior	dynaaminen ohjelmointi	dynamic programming, dynamic optimization
digammafunktio	digamma function	dynaaminen optimointi	dynamic optimization
dikotomia	dichotomy	dynaaminen populaatio	dynamic population, open population
dikotominen muuttuja	dichotomous variable, binary variable		
dimensiokirous	curse of dimensionality	E	
dimension pienennys	dimension reduction	editointisääntö	edit rule
Diracin deltafunktio	Dirac's delta function	edustava otos	representative sample
Diracin mitta	Dirac measure	edustavuus	representativeness
Dirichlet-jakauma	Dirichlet distribution	ehdokasjakauma	candidate distribution, proposal distribution
Dirichlet-prosessi	Dirichlet process	ehdollinen autoregressio	conditional autoregression
diskonttaus	discounting	ehdollinen jakauma	conditional distribution
diskordanssi	discordance	ehdollinen keskimääräinen	
diskordantit parit	discordant pairs	kausaalivaikutus	conditional average causal effect
diskreetti	discrete	ehdollinen keskimääräinen	
diskreettiaikainen	discrete-time	käsittelyvaikutus	conditional average treatment effect

ehdollinen odotusarvo	conditional expectation	nous variable
ehdollinen riippumattomuus	conditional independence	eksogeenisten viipeiden malli
ehdollinen riippuvuus	conditional dependence	distributed-lag model
ehdollinen tiheysfunktio	conditional density function	eksogeenisuus
ehdollinen todennäköisyys	conditional probability	exogeneity
ehdollinen uskottavuus	conditional likelihood	eksponentiaalinen kasvu
ehdollinen varianssi	conditional variance	exponential growth
ehdollinen yhteys	conditional association	eksponentiaalinen trendi
ehdollistamaton jakauma	unconditional distribution	exponential trend
ehdollistettu sirontakuvio	conditioning plot, coplot	eksponenttijakauma
ehdotusjakauma	proposal distribution, candidate distribution	exponential distribution
ehkäisykoe	prevention trial	eksponenttikäyrä
ehkäisyosuus	preventive fraction	exponential curve
ei-merkittävä	nonsignificant	eksponenttinen hajontamalli
ei-negatiivisesti definiitti	non-negative definite	exponential dispersion model
ei-normaali	non-normal	eksponenttinen hajontaperhe
ei-satunnainen	non-random	exponential dispersion family
ei-satunnaisesti puuttuva	missing not at random	eksponenttinen kasvu
ekologinen harha	ecological fallacy, ecological bias	exponential growth
ekologinen korrelaatio	ecological correlation	eksponenttinen silotus
ekologinen tutkimus	ecological study	exponential smoothing
ekologinen virhepäätelmä	ecological fallacy	eksponenttinen tasoitus
ekonometria	econometrics	exponential smoothing
ekonometrikko	econometrician	eksponenttinen trendi
eksakti testi	exact test	exponential trend
eksogeeninen muuttuja	exoge-	eksponenttiperhe
		exponential family
		ekstrapolointi
		extrapolation
		ekvivalenssiluokka
		equivalence class
		ekvivalenssitestaus
		equivalence testing
		ekvivariantti estimaattori
		equivariant estimator
		elastisuus
		elasticity
		elementtiestimaattori
		elemental estimator
		eliminointi
		elimination
		elinaika
		lifetime
		elinaika-analyysi
		lifetime data analysis, survival analysis
		elinaikajakauma
		lifetime distribution
		elinajan jakauma
		lifetime distribution

bution	
elinajanodote	life expectancy
elinajantaulukko	life table
elinkaarianalyysi	life course analysis
elinkaarihyöty	life cycle utility
elinkorko	annuity ₁
elliptinen jakauma	elliptical distribution
elossaoloanalyysi	survival analysis
elossaolofunktio	survival function
elossaolokäyrä	survival curve
eläkerahasto	pension fund
EM-algoritmi	EM-algorithm, expectation-maximization algorithm
empiirinen	empirical
empiirinen Bayes-estimointi	empirical Bayes estimation
empiirinen kertymäfunktio	empirical distribution function
empiirinen mitta	empirical measure
empiirinen peittotodennäköisyys	empirical coverage
empiirinen uskottavuus	empirical likelihood
emäfunktio	generating function
endogeeninen muuttuja	endogenous variable
endogeenisuus	endogeneity
ennakkorekisteröinti	preregistration
ennakoituva prosessi	predictable process
ennakoiva indikaattori	leading indicator
ennakoiva säätö	anticipative control
ennalta asetettu merkitsevyystaso	pre-specified significance level
ennustaminen	forecasting, prediction
ennustava analytiikka	predictive analytics
ennustava muuttuja	predictor
ennustava päättely	predictive inference
ennuste₁	forecast
ennuste₂	prediction
ennuste₃	predicted value
ennustearvo	predictive value
ennustejakauma	predictive distribution
ennustejäännös	prediction residual
ennustejäännösten neliösumma	prediction residual sum of squares
ennustekeskiarvokaltaistus	predictive mean matching
ennustemalli	prognostic model
ennustetekijä	prognostic factor
ennustettu arvo	predicted value
ennustevirhe	prediction error
ennusteväli	prediction interval
ennustin	predictor
ennätys	record ₂
ensiasteen otantayksikkö	first stage sampling unit, primary sampling unit
ensimmäisen kertaluvun malli	first order model
ensinumeron jakauma	first digit distribution
ensipaluun hetki	first return time
ensikulun hetki	first passage time
ensisijainen vaste	primary outcome
entropia	entropy
E-optimaalisuus	<i>E</i> -optimality
epidemiamalli	epidemic model

epidemiologia	epidemiology		
epäaito esijakauma	improper prior	epätasapainoinen aineisto	unbalanced data
epäaito jakauma	improper distribution	epätasapainoinen asetelma	unbalanced design
epäaito otanta	non-probability sampling	epävarmuuden määrittäminen	uncertainty quantification
epäaito prior	improper prior	epävarmuus	uncertainty
epäinformatiivinen prior	non-informative prior	epävarmuusväli	uncertainty interval
epäjatkuva	discontinuous	ergodinen	ergodic
epäkelpo estimaattori	inadmissible estimator	ergodisuus	ergodicity
epäkelpo imputointi	improper imputation	erhe	error
epäkeskinen	non-central	eriarvoisuuskerroin	inequality coefficient
epäkeskisyyssparametri	non-centrality parameter	erilaiset parit	discordant pairs
epäkokeellinen tutkimus	non-experimental study	erilaisuus	discordance
epälineaarinen	nonlinear	erilaisuusindeksi	index of dissimilarity
epälineaarinen malli	nonlinear model	erilaisuuskerroin	dissimilarity coefficient
epälineaarinen optimointi	nonlinear optimization	erilaisuusmatriisi	dissimilarity matrix
epälineaarinen regressio	nonlinear regression	erilliset tapahtumat	mutually exclusive events, disjoint events
epäonnistumistodennäköisyys	failure probability	erivarianssinen	heteroskedastic
epäortogonaalinen	non-orthogonal	erivarianssisuuden suhteen tarkentuva	heteroskedasticity-consistent
epäparametrinen	nonparametric	erivarianssisuuden suhteen vakaa	heteroskedasticity-robust
epäsingulaarinen	nonsingular	eriytyminen	segregation ₁
epäsisäkkäinen luokitus	nonnested classification	eriytymisanalyysi	segregation analysis
epäsuora otanta	indirect sampling	eriävä virheluokittelu	differential misclassification
epäsuora pienin neliösumma	indirect least squares	erottava hypertaso	separating hyperplane
epäsuora vaikutus	indirect effect	erotteleva tunnusluku	complete statistic
epäsuora vakiointi	indirect standardization	erotteluanalyysi	discriminant analysis
epäsuotuisin jakauma	least favorable distribution	erottelufunktio	discriminant

function	estimoituva estimable
erottelukyky discriminatory power	eteenyhtälö forward equation
erottelukykyosoitin receiver operating characteristic	etenevä valinta forward selection
erottelukykykäyrä receiver operating characteristic curve, ROC curve	etiologinen osuus etiologic fraction
erotuksen itseisarvo absolute difference	etsintä-hyödynnyskompromissi exploration vs. exploitation trade-off
erotus difference	etsivä data-analyysi exploratory data analysis, EDA
erotusoperaattori difference operator	etumerkkinen neliöjuurimuunnos signed root-transformation
erotusten erotus difference in differences	etuovivakiointi frontdoor adjustment
erotusten laskeminen differencing	etäisyysmitta distance measure
erä lot	etäisyysotanta distance sampling
eräkat item non-response	euklidinen etäisyys Euclidean distance
eräpäivä maturity date	eurooppalainen optio European option
erävaihtelu batch variation	
esiennustejakauma prior predictive distribution	F
esiintyvyys occurrence	faktori factor ₂
esijakauma prior distribution	faktorianalyysi factor analysis
esijakauman koostaminen elicitation of prior	faktorikierto factor rotation
esiluotaus exploratory survey	faktorilataus factor loading
esitodennäköisyys prior probability	faktorimatriisi factor matrix
esitutkimus exploratory survey, pilot study	faktoripistemäärä factor score
esivalkaisu prewhitening	faktorirakenne factor pattern
estimaatti estimate	faktorirotaatio factor rotation
estimaattori estimator	faktorointi factoring, factorization
estimandi estimand	faktorointilause factorization theorem
estimointi estimation	fenotyyppi phenotype
estimointifunktio estimating function	fidusiaalinen päättely fiducial inference
estimointivirhe estimation error	fi-kerroin phi-coefficient
estimointiyhtälö estimating equation	filtraatio filtration
estimoitava suure estimand	Fisherin estimointialgoritmi Fisher's scoring method
	Fisher-informaatio Fisher infor-

mation
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 geometric 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-formula
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
hajontaluku measure of dispersion, measure of variation
hajontamatriisi dispersion ma-

trix, scatter matrix
hajontamitta measure of dispersion, measure of variation
hajotelma decomposition₂
hakutaulu lookup table
halovaikutus halo effect
Hamming-etäisyys Hamming distance
Hamming-koodi Hamming code
Hann-ikkuna Hann-window
hapsukuvio rug plot
haravointi raking
harha bias, systematic error
harhainen estimaattori biased estimator
harhainen otanta biased sampling
harhainen testi biased test
harhaton unbiased
harha-varianssikompromissi bias-variance tradeoff
harjaregressio ridge regression
harjanneregressio ridge regression
harkintanäyte judgment sample, purposive sample
harkintavalinta purposeful selection
harmaan laatikon malli grey box model
harmoninen jakauma harmonic distribution
harmoninen keskiarvo harmonic mean
harmoninen regressio harmonic regression
harva matriisi sparse matrix
harvennus thinning
harventuminen attrition
hasardi hazard
hattumatriisi hat matrix
haurausmalli frailty model
havainnoiva tutkimus observa-

tional study	process
havainnon sisäisyys data depth	heilahtelu oscillation
havainnot tuottava prosessi	heilunta volatility
data generating process	Helmert-kontrasti Helmert contrast
havainto observation	henkilöaika person-time
havaintoaineisto data	henkilöaika kuolemattomana
havaintokuutio data cube	immortal person-time
havaintomatriisi data matrix	henkilökohtainen todennäköisyys personal probability
havaintosuure statistic	henkilövuosi person-year
havaintovirhe observational error	henkivakuutus life insurance
havaintoyksikkö observational unit	heritabiliteetti heritability
havaittava observable	herkkyys sensitivity ₁
havaittu informaatio observed information	herkkyysanalyysi sensitivity analysis
havaittu informaatiomatriisi observed information matrix	Hessen matriisi Hessian
havaittu lukumäärä observed frequency, observed number	heterogeeninen heterogeneous
havaittu merkitsevyystaso <i>p</i> -value, observed significance level	heteroskedastinen heteroskedastic
Heavisiden funktio Heaviside function	heteroskedastisuuden suhteen tarkentuva heteroskedasticity-consistent
hedelmällisyys fertility	heteroskedastisuuden suhteen vakaa heteroskedasticity-robust
hedelmällisyysluku fertility rate	hetkellinen tila instantaneous state
hedoninen regressio hedonic regression	hetkellinen väestö de facto population
hehkutettu jakauma annealed distribution	hevosenkenkäilmiö horseshoe effect
heijastava reuna reflecting barrier	hieno järjestys fine ordering
heijastusperiaate reflection principle	hierarkkinen luokittelu hierarchical classification
heikko eksogeenisuus weak exogeneity	hierarkkinen malli hierarchical model, multilevel model
heikko priori vague prior	hierarkkinen ryhmittely hierarchical clustering
heikko suppeneminen weak convergence	hila grid, lattice
heikoimman lenkin malli weakest-link model	hila-asetelma lattice design
heikosti stationaarinen weakly stationary	hilajakauma lattice distribution
heilahteleva prosessi oscillatory	hilaneliö lattice square
	hilaotanta lattice sampling

hintaindeksi	price index	hypergeometrisen jakauman	hypergeometric distribution
histogrammi	histogram	hyperparametri	hyperparameter
historia	history, filtration	hypoteesi	hypothesis
hitaasti vaihteleva funktio	slowly varying function	hypoteesin testaus	hypothesis testing
hoito	treatment	hyppytaajuus	jump rate
hoitoaikkeen mukainen analyysi	intention-to-treat analysis	hyppytodennäköisyys	jump probability
hoitoero	treatment difference	hyväksymisalue	acceptance region
hoitolinja	treatment arm	hyväksymisotanta	acceptance sampling
hoitomyöntävyys	compliance with treatment, adherence to treatment	hyväksymisraja	acceptance boundary, acceptance line
hoitoryhmä	treatment group	hyväksymisvalvontakaavio	acceptance control chart
hoitositoutuneisuus	compliance with treatment, adherence to treatment	hyväksymisvirhe	acceptance error, false acceptance
homogeeninen	homogeneous	hyväksyttävä	admissible
homogeenisuustesti	homogeneity test	hyväksyttävä estimaattori	permissible estimator
homoskedastinen	homoskedastic	hyväksyttävä laatutaso	acceptable quality limit
HPD-alue	HPD region, highest posterior density region	hyödyttömyys	futility
HPD-väli	HPD interval, highest posterior density interval	hyöty	gain ₂
HSD-testi	HSD-test, honest significant difference test	hyöty	utility
huipukkuus	peakedness	hyötyfunktio	utility function
huonosti asetettu ongelma	ill-posed problem	hyötyteoria	utility theory
hylkäysalue	rejection region, critical region	häiriö	perturbation ₁
hylkäysotanta	rejection sampling	häiriöaltis matriisi	ill-conditioned matrix
hylkäysraja	critical value	häiriömenetelmä	perturbation method
hylkäysvirhe	rejection error, false rejection	häly	noise
hylkäävä otanta	rejective sampling	hätäeksponentti	tail exponent
hylättävä laatutaso	rejectable quality limit	hölynpölykorrelaatio	nonsense correlation
hypereksponenttijakauma	hyperexponential distribution		
hypergeometrinen jakauma			

I

ideaali-indeksi	ideal index
idempotentti	idempotent
identifioimattomuus	unidentifiable

bility	indeksi index, index number
identifioituvuus identifiability	indeksikuvio index plot
identiteettifunktio identity function	indeksiluku index number, index
ihanneindeksi ideal index	indikaattorifunktio characteristic function ₂
i.i.d. independent and identically distributed	indikaattorimatriisi incidence matrix ₁
i'ittäinen age-specific	indikaattorimuuttuja indicator variable
i'ittäinen kuolevuus age-specific mortality rate, age-specific death rate	induktiivinen käyttäytyminen inductive behaviour
ikkuna window	induktiivinen päättely inductive inference
ikkunanleveys window width, bandwidth ₂	induktio induction
ikäkasautuminen age heaping	infektiokuolleisuus infection fatality ratio
ikä-periodi-kohorttimalli age-period-cohort model	inflaatio inflation
ikäpyramidi age pyramid	informaatio information
ikäriippuva prosessi age-dependent process	informaatioarvo value of information
ikäryhmittäinen age-specific	informaatiofunktio information function
ikäryhmittäinen kuolevuus age-specific death rate, age-specific mortality rate	informaatiokriteeri information criterion
ikäsuosio age heaping	informaatiomatriisi information matrix
ikävakioitu age-adjusted, age-standardized	informaatioteoria information theory
ilmaantumisosuus incidence proportion, cumulative incidence	informatiivinen esijakauma informative prior
ilmaantumistiheys incidence rate	informatiivinen priori informative prior
ilmaantuvuus incidence	informatiivinen puuttuvuus informative missingness
ilmais indicator variable	informatiivinen sensurointi informative censoring
ilmiasu phenotype	innovaatiohyöty innovation gain
ilmimuuttuja manifest variable	innovaatiotermi innovation term
ilmiömalli substantive model	insinööritilastotiede engineering statistics
implisiittinen ositus implicit stratification	instrumenttimuuttuja instrumental variable
impulssivastefunktio impulse response function	integraali integral
imputointi imputation	
imputointimalli imputation model	

integroitu keskineliövirhe	mean integrated squared error	isovuoseulonta	high-throughput screening
integroitu liukuvan keskiarvon malli	integrated moving average model	isäaaloke	father wavelet
integroitu spektri	integrated spectrum	isäväre	father wavelet
integroituva	integrable	iteraatio	iteration
intensiteetti	intensity, rate ₁	iteratiivinen verrannollinen sovitus	iterative proportional fitting
intensiteettifunktio	intensity function	iteratiivisesti painotettu pienin neliösumma	iteratively weighted least squares
intensiteettikertymä	cumulative intensity, integrated intensity	iteroidun logaritmin laki	law of the iterated logarithm
interaktio	interaction	iterointi	iteration
interpolaatio	interpolation	iteroitujen odotusarvojen laki	law of iterated expectations
interpoloiva splini	interpolating spline	iteroitu uusio-otanta	iterated bootstrap
intervallasteikko	interval scale	Itô-integraali	Itô integral
interventioanalyysi	intervention analysis	itsehuomio	self-attention
intervention jälkeinen jakauma	post-interventional distribution	itseisarvo	absolute value
interventiotutkimus	intervention study	itseiserotus	absolute difference
intialainen noutopöytäprosessi	Indian buffet process	itseisesti summautuva	absolutely summable
invarianssi	invariance	itseisesti suppeneva	absolutely convergent
invariantti	invariant	itseismomentti	absolute moment
invariantti jakauma	invariant distribution	itseispoikkeama	absolute deviation
invariantti koordinaatisto	invariant coordinate system	itseisvirhe	absolute error
invariantti koordinaatiston valinta	invariant coordinate selection	itsekehkeytyvä prosessi	self-exciting process
inversio	inversion	itsensä estävä ennuste	self-negating prophecy
inversio-ongelma	inverse problem	itsensä rahoittava	self-financing
isometria	isometry	itsensä toteuttava ennuste	self-fulfilling prophecy
isotoninen regressio	isotonic regression	itseorganisoituva kartta	self-organizing map
isotrooppinen	isotropic	itsepainottuva otos	self-weighting sample
isovuodata	high-throughput data	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 matrix

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

juovakuvio 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

jälkiryöstys post-clustering

jälkitodennäköisyys posterior probability

jälkivaikutus carry-over effect

jälkiväli posterior interval
jälleenvakuutus reinsurance
järjestelmällinen katsaus systematic review
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
jäännössekoittuneisuus residual confounding
jäännösvaikutus residual effect
jäännösvarianssi residual vari-

ance
jäännösvirhehajonta 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 sisällymisdennäköisyys second-order inclusion probability
kahdesti satunnainen malli doubly stochastic model
kahtiajako dichotomy
kaistanleveys bandwidth₁
kaksiarvoinen annos-vaste quantal dose-response
kaksiarvoinen muuttuja binary variable, dichotomous variable
kaksiasteinen pienin neliösumma two-stage least squares
kaksiastejakauma bidegree distribution
kaksiasteotanta two-stage sampling
kaksihuippuinen jakauma bimodal distribution
kaksijakoinen bipartite
kaksikätkäinen rosvo two-armed bandit
kaksisuuntainen luokittelu two-way classification
kaksisuuntainen ristitaulukko two-way contingency table
kaksisuuntainen testi two-sided test, two-tailed test
kaksisuuntainen vaihtoehtoinen hypoteesi two-sided alter-

native hypothesis	correlation
kaksisuuntainen varianssianalyysi two-way analysis of variance	kanoninen linkkifunktio canonical link function
kaksisuuntainen vastahypoteesi two-sided alternative hypothesis	kanoninen parametri canonical parameter, natural parameter
kaksitahoinen testi two-sided test, two-tailed test	kanta basis
kaksiulotteinen jakauma bivariate distribution	kantaja support
kaksiulotteinen normaali-jakauma bivariate normal distribution	kantajamuuttuja carrier variable
kaksiulotteinen ristitaulukko two-way contingency table	Kaplanin–Meierin estimaattori Kaplan–Meier estimator, product limit estimator
kaksivaiheotanta two-phase sampling, double sampling	kappakerroin kappa coefficient
kaksoiseksponenttijakauma double exponential distribution	karakteristinen funktio characteristic function
kaksoisrekisteriestimointi dual systems estimation	karakteristinen yhtälö characteristic equation
kaksoissokkokoe double-blind experiment	karhunta call-back
kaksoisvakaa estimointi doubly robust estimation	karkaistu jakauma quenched distribution
kaksostutkimus twin study	karkea data coarse data
kalibrointi calibration	karkea estimaattori crude estimator
kalibroitu paino calibrated weight	karkein tyhjentävä tunnusluku minimal sufficient statistic
Kalman-hyöty Kalman gain	karkeistus coarsening
Kalman-suodin Kalman filter	karsinta pruning
kaltaistettu pari matched pair	kartiosilotus tapering
kaltaistettu tapaus-verrokki-tutkimus matched case-control study	kartogrammi cartogram
kaltaistetut otokset matched samples	kartoitus survey
kaltaistus matching	kaskadiprosessi cascade process
kaltaistusjoukko matched set	kasvukäyrä growth curve
kanoninen analyysi canonical analysis	kasvunopeus growth rate
kanoninen korrelaatio canonical	katastrofiteoria catastrophe theory
	katerpillarikuvio caterpillar plot
	katkaistu jakauma truncated distribution
	katkaisu oikealta right-truncation
	katkaisupiste cut-off point
	kato dropout, attrition, loss
	kattava malli encompassing model

kattovaikutus	ceiling effect	imputation	
kaukokartoitus	remote sensing	kemometria	chemometrics
kauppamatkustajan ongelma	traveling salesman problem	Kendallin tau	Kendall's tau
kausaalikaavio	causal diagram, causal graph	kenttäkoe	field experiment
kausaaliketju	causal chain	kenttätyö	field work
kausaalimalli	causal model	kerrannainen kausivaihtelu- malli	multiplicative seasonal model
kausaalipäätely	causal inference	kerrannainen malli	multiplica- tive model
kausaalirakenteen löytäminen	causal discovery	kerroin	coefficient
kausaalisilmukkakaavio	causal loop diagram	kerroinero	difference in betas
kausaalisuus	causality	kerrosetestimaattori	sandwich estimator
kausaaliverkko	causal graph, causal diagram	kertaimputointi	single imputa- tion
kausi-ARMA-malli	seasonal autoregressive moving average process	kertapaikkaus	single imputation
kausijakso	period of seasonality	kertoma	factorial
kausijuuri	seasonal root	kertomakumulantti	factorial cumulant
kausipuhdistus	seasonal adjust- ment	kertomamomentit generoiva funktio	factorial moment gener- ating function
kausivaihtelu	seasonal variation	kertomamomentti	factorial mo- ment
kauttaaltaan pienin riski	uni- formly minimum risk	kertomamomenttien emäfunk- tio	factorial moment generating function
kauttaaltaan tarkin väli	uni- formly most accurate interval	kertymisnopeus	accrual rate
kauttaaltaan voimakkain	uni- formly most powerful	kertymäfunktio	distribution function, cumulative distribution function
kehikko	frame	kertymäfunktio muunnos	proba- bility integral transformation
kehikkoperusjoukko	frame pop- ulation, sampled population	kertymämalli	cumulative model
kehityspolku	trajectory	kertymäristisuhde	cumulative odds ratio
kehityssuunta	trend	kertymäsumma	cumulative sum
kehys	frame	keskeinen raja-arvolause	cen- tral limit theorem
kehäaineisto	circular data	keskenään vaihtokelpoiset muuttujat	exchangeable vari- ables
kehädata	circular data		
kehäjakauma	circular distribu- tion		
keinoäly	artificial intelligence		
kelpoinen malli	adequate model		
kelvollinen imputointi	proper		

keskeytetty tarkastus	curtailed inspection		deviation
keskiarvo	mean_1 , average	keskipuolikkaan keskiarvo	mid-mean
keskiarvosuppeneva	ergodic	keskisijaluku	mid-rank
keskiarvovektori	mean vector	keskistäminen	centering
keskihajonta	standard deviation	keskittymisindeksi	index of concentration
keski-itseisvirhe	mean absolute error	keskittymiskerroin	coefficient of concentration
keskiluku	measure of location	keskittyneisyys	precision ₂
keskimääräinen kausaali vaikutus	average causal effect	keskittyneisyyskäyrä	concentration curve
keskimääräinen kausaali vaikutus sitoutuneilla	complier average causal effect	keskittyneisyysmatriisi	precision matrix, concentration matrix
keskimääräinen käsittely vaikutus	average treatment effect	keskittyneisyysparametri	concentration parameter
keskimääräinen käsittely vaikutus altistumattomilla	average treatment effect on the untreated	keskivirhe	standard error
keskimääräinen käsittely vaikutus altistuneilla	average treatment effect on the treated	keskusmomentti	central moment
keskimääräinen käsittely vaikutus sitoutuneilla	complier average treatment effect	kesko	mid-range
keskimääräinen merkkijakson pituus	average run length	kesto	duration
keskimääräinen vikaantumisaika	mean time to failure	kestoanalyysi	survival analysis
keskimääräinen vikaantumisväli	mean time between failures	kestofunktio	survival function
keskineliösuppeneminen	convergence in mean square	kestokäyrä	survival curve
keskineliöjatkuvuus	mean square continuity	ketju	chain
keskineliöraja-arvo	limit in the mean	ketjubinomimalli	chain binomial model
keskineliövirhe	mean squared error	ketjuimputointi	multivariate imputation by chained equations (MICE)
keskinäisinformaatio	mutual information	ketjuindeksi	chain index
keski-p-arvo	mid- p value	ketjukaavio	chain graph
keskipoikkeama	mean absolute	khi-jakauma	chi distribution
		khi-kuvio	chi-plot
		khi-neliöetäisyys	chi-squared distance
		khi-neliöjakauma	chi-squared distribution
		khi-neliösuure	chi-squared statistic
		khi-neliötesti	chi-squared test
		khi-suure	chi-statistic
		kieltäytymisosoitus	refusal rate
		kierto	rotation

kierto-otanta	rotation sampling	kohdepopulaatio	target population
kiertoasetelma	rotatable design	kohdesopeutus	domain adaptation
kiertovuoroasetelma	Round Robin design	kohdeväestö	target population
kiinalainen ravintolaprosessi	Chinese restaurant process	kohdistettu suurimman uskottavuuden menetelmä	targeted maximum likelihood estimation
kiinteävaikutusmalli	fixed-effects model	koherenssi	coherence
kiinteäkantainen indeksi	fixed base index	kohina	noise
kiinteä otos	fixed sample	kohinanpoisto	denoising
kiintiöinti	allocation ₁	kohinasta puhdistettu	denoised
kiintiöotanta	quota sampling	kohortti	cohort
kiirejakso	busy period	kohorttitutkimus	cohort study
kilpaileva riski	competing risk	kohtisuora	perpendicular, orthogonal
kilpaileva tapahtuma	competing event	kohtisuora projektio	orthogonal projection
kiusaparametri	nuisance parameter	kokeellinen tutkimus	experimental study
klaava	tails	kokemusperäinen	empirical
kliininen koe	clinical trial	kokoava hierarkkinen ryhmitely	agglomerative hierarchical clustering
<i>k</i>-luokan estimaattori	<i>k</i> -class estimator	kokonaiskeskiarvo	grand mean
<i>k:n</i> keskiarvon ryvästys	<i>k</i> -means clustering	kokonaiskorrelaatio	total correlation
<i>k:n</i> lähinaapurin menetelmä	<i>k</i> -nearest-neighbors method	kokonaislaskenta	complete enumeration
koalesenssi	coalescence	kokonaismäärä	total
koe	experiment, trial	kokonaisneliösumma	total sum of squares
koeala	experimental plot	kokonaisosa	curtate, truncation
koeasetelma	experimental design	kokonaisten havaintojen analyysi	complete case analysis
koeruutu	experimental plot	kokonaistodennäköisyys	total probability
koesarja	series of experiments	kokonaistutkimus	census ₁ , complete enumeration
koesuunnittelu	design of experiments	kokonaisvaihteluetaisyys	total variation distance
koete	run ₃	kokoon verrannollinen todennäköisyys	probability propor-
koevirhe	experimental error		
koeyksikkö	experimental unit		
kofaktori	cofactor		
kognitiivinen harha	cognitive bias		
kohdefunktio	objective function		
kohdentaminen	allocation ₂		

tional to size	gency table
kokovalikoitunut size-biased	kontrafaktuaalinen counterfac- tual
kollinearisuus collinearity	kontrasti contrast
kolmen sarjan lause three series theorem	kontrolloitu suora vaikutus controlled direct effect
kolmen sääntö rule of three	konvekksi peite convex hull
kolmiasteotanta three-stage sam- pling	konveksisuus convexity
kolmihiippuinen trimodal	konvergenssi convergence
kolmijakoinen tripartite	konvoluutio convolution
kolmikkotesti triples test	koontuvuus collapsibility
kolmimuuttujainen trivariate	kooste ensemble
kolmiojakauma triangular distri- bution	kooste-ennuste ensemble predic- tion
kolmiomatriisi triangular matrix	koostekeskiarvo ensemble aver- age
kolmiotaulukko triangular table	koosteoppiminen ensemble learn- ing
kolmiotesti triangular test	koosteregressio regression by composition
kolmivaiheinen pienin neliö- summa three-stage least squares	kopula copula
kommunaliteetti commonality	korikoe basket trial
komplementtinen loglog-muun- nos complementary loglog trans- formation	korjaava tarkastus rectifying inspection
komplementtinen virhefunktio complementary error function	korkeamman kertaluokan asymptotiikka higher order asymptotics
komplementtitapahtuma com- plement event	korkeaulotteinen aineisto high- dimensional data
koneoppiminen machine learning	korkeaulotteinen tilastotiede high-dimensional statistics
konenäkö computer vision	korkotaso interest rate
koneäly machine intelligence	korrelaatio correlation
konfiguraatio configuration	korrelaatiokerroin correlation coefficient
konfiguraatiomalli configuration	korrelaatiomatriisi correlation matrix
konkaavisuus concavity	korrelaatiopuhdistettu decorre- lated
konkordanssi concordance	korrelaatiosta puhdistettu decorrelated
konkordanssikerroin coefficient of concordance	korrelogrammi correlogram
konkordantit parit concordant pairs	korreloimaton uncorrelated
konservatiivinen testi conserva- tive test	
kontingenssikerroin contingency coefficient	
kontingenssitaulukko contin-	

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

kuutiosplini	cubic spline	kvaskoe	quasi-experiment
kuvaaja	graph ₃	kvasi riippumattomuus	quasi-independence
kuva-analyysi	image analysis	kvasisymmetria	quasisymmetry
kuvaileva analytiikka	descriptive analytics	kvasi uskottavuus	quasilikelihood
kuvaileva tilastotiede	descriptive statistics	kvintiili	quintile
kuvaileva tunnusluku	descriptive statistic	kybernetiikka	cybernetics
kuvaileva tutkimus	descriptive study	kyllästetty malli	saturated model
kuvankäsittely	image processing	kylmäimputointi	cold deck imputation
kuvio	graph ₁ , plot ₁	kylmäpaikkaus	cold deck imputation
kvadraattinen	quadratic	kymmenys	decile
kvadraattinen erottelufunktio	quadratic discriminant function	kynnysarvo	cut-off point
kvadraattinen estimaattori	quadratic estimator	kynnysarvo	threshold value
kvadraattinen vaihtelu	quadratic variation	kynnysautoregressio	threshold autoregression
kvadranttiotanta	quadrant sampling	kynnysmalli	threshold model
kvadranttiriippuvuus	quadrant dependence	kysely	inquiry
kvalitatiivinen	qualitative	kyselykierros	survey round
kvantiili	quantile, fractile	kyselylomake	questionnaire
kvantiilifunktio	quantile function	kyselytutkimus	questionnaire survey
kvantiili-kvantiilikuvio	quantile-quantile plot	kytkentä	coupling
kvantiiliregressio	quantile regression	kytkentäanalyysi	linkage analysis
kvantitatiivinen	quantitative	kytkentäepätasapaino	linkage disequilibrium
kvanttiodennäköisyys	quantum probability	kytkentämatriisi	incidence matrix ₂
kvartiili	quartile	kytketyt lohkot	linked blocks
kvartiilikeskisarvo	mid-hinge	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
		käsittelyvirhe	processing error
		käsittelyä edeltävä kovariaatti	pre-treatment covariate
		käyntihaastattelu	face-to-face interview

käypä admissible
käypä estimaattori admissible estimator
käypä päätösfunktio admissible decision function
käypä testi admissible test
käyräanalainen pinta-ala area under curve
käyräanalusala area under curve
käyrän sovitus curve fitting
käyräviivainen curvilinear
käyttökielen prosessointi natural language processing
käännekohta turning point
käänteinen Bernoulli-poiminta inverse Bernoulli sampling
käänteinen betajakauma inverted 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 normaalijakauma inverse Gaussian distribution
käänteislukumuunnos reciprocal transformation
käänteismatriisi inverse matrix
käänteismuunnos inversion
käänteismuunnosmenetelmä inverse transformation method
käänteisongelma inverse problem
käänteisotanta inverse sampling
käänteistodennäköisyys inverse probability
käänteistodennäköisyyspainotus inverse probability weighting
kääntyvä invertible
kääntyvä prosessi invertible pro-

cess
kääritty jakauma wrapped distribution

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 assurance
laajassa mielessä stationaarinen wide sense stationary
laakea priori vague prior
laatikkokuvio box plot
laatuelinvuosi quality adjusted life years
Lagrange kerrointesti Lagrange multiplier test
lajien ordinaatio ordination of species
lajien runsaus species abundance
Laplace-jakauma Laplace distribution
Laplace-muunnos Laplace transform
lapsiluku parity₁
laskennallinen tilastotiede computational statistics
laskennan alipeitto underenumeration, undercount
laskennan ylipeitto overenumeration, overcount
laskenta enumeration
laskentaintensiivinen menetelmä computationally

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

knife estimator	loglog-muunnos loglog-transformation
linkkuveitsimenetelmä jack-knife method	log-normaalijakauma lognormal distribution
lista ₁ list	logrank-testi logrank test
lista ₂ tuple	log-uskottavuusfunktio log-likelihood function
listapoisto listwise deletion	log-vastasuhde log-odds
lisäinformaatio auxiliary information	log-vetokerroin log-odds
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	lohkolävistäjämatrissi 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	lovettu laatikkokuvio notched boxplot
logaritminen asteikko logarithmic scale	luetteleminen enumeration
logaritmisen sarjan jakauma logarithmic series distribution	lukumäärä count, frequency ₁ , number
logistinen jakauma logistic distribution	lukumääräaikaasarja count time series
logistinen käyrä logistic curve	lukumääräaineisto count data, frequency data
logistinen regressio logistic regression	lukumääräjakauma frequency distribution
logit-malli logit model	lukumäärätaulukko frequency table
logit-muunnos logit transformation	lume placebo
logitti log-odds	lumevaikutus placebo effect
log-lineaarinen malli loglinear model	lumihiutalekuvio snowflake diagram
	lumipallo-otanta snowball sampling
	lunastusarvo exercise value

lunastushinta	exercise price	luonnollinen epäsuora kokonaisvaikutus	total natural indirect effect
lunastuspäivä	exercise date	luonnollinen epäsuora vaikutus	natural indirect effect
luokitin	classifier	luonnollinen suora kokonaisvaikutus	total natural direct effect
luokittelija	classifier	luonnollinen liittopriori	natural conjugate prior
luokittelu	classification	luonnollinen parametri	natural parameter, canonical parameter
luokitteluasteikko	nominal scale	luonnollinen suora vaikutus	natural direct effect
luokitteluasteikon muuttuja	categorical variable	luonnollisen kielen prosessointi	natural language processing
luokitteluattribuutti	classification attribute	luonnon koe	natural experiment
luokittelu- ja regressiopuu	classification and regression tree	luotaus	survey
luokittelumääre	classification attribute	luotausinstrumentti	survey instrument
luokittelun yhdenmukaisuus	consistency of classification	luotausmetodologia	survey methodology
luokittelusääntö	classification rule	luotausotanta	survey sampling
luokittelutaulukko	classification table	luotautapa	survey mode
luokittelutunnusluku	classification statistic	luotaukilastotiede	survey statistics
luokitteluvirhe	classification error	luotettavuus	reliability ₁
luokitteluvirhetaulukko	confusion matrix	luotettavuusfunktio	reliability function
luokitus	classification	luotettavuusteoria	reliability theory
luokitushila	grouping lattice	luottamusalue	confidence region
luokka-asteikko	nominal scale	luottamusjoukko	confidence set
luokkafrekvenssi	class frequency	luottamusraja	confidence limit
luokkajakauma	categorical distribution	luottamustaso	confidence level
luokkakeskus	class midpoint	luottamusvyöhyke	confidence band, confidence envelope
luokkamuuttuja	categorical variable	luottamusväli	confidence interval, compatibility interval
luokkatunnus	class symbol	luotto	credit
luokkaväli	bin, class interval	luottoalue	credible region
luokkien sisäinen samankaltaisuus	intra-class similarity	luottoteoria	credibility theory
luokkien välinen samankaltaisuus	inter-class similarity	luottoväli	credible interval

luovuttajapaikkaus donor imputation
lyhimmän koodin pituus minimum description length
lyhin luottamusväli most accurate confidence interval
lyhyeksi myynti short sale
lyhyhäntäinen jakauma short-tailed distribution
lyhytmuistinen prosessi short-memory process
läheisyysmatriisi proximity matrix
lähes vajaa-asteinen matriisi ill-conditioned matrix
lähinaapurimenetelmät nearest-neighbors methods
lähinaapuriryhmittely nearest-neighbors clustering
lähtöarvot baseline characteristics
lähtökohortti inception cohort
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
marginaalihakkuisuus marginal homogeneity
marginaalijakauma marginal distribution
marginaalimalli marginal model
marginaalirakennemalli marginal structural model
marginaaliskottavuus marginal likelihood
marimekkokuva Marimekko plot, mosaic plot
markov¹ Markov
markovinen Markovian
Markovin ketju Markov chain
Markovin ketju 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

¹Ks. alaluku 2.5.

martingaali	martingale	merkitsevyys	significance
martingaalijäännös	martingale residual	merkitsevyystaso	level of significance, significance level
martingaalimitta	martingale measure	merkitsevyystesti	test of significance, significance test
martingaaliresiduaali	martingale residual	merkkijakso	run_1
massadata	big data	merkkijaksotesti	runs test
matokuvio	worm plot	merkkijärjestystesti	signed rank test
matriisialgebra	matrix algebra	merkkipisteprosessi	marked point process
matriisihajotelma	matrix decomposition	merkkisijatesti	signed rank test
matriisilaskenta	matrix algebra	merkkitesti	sign test
matriisinormaalijakauma	matrix normal distribution	M-estimaattori	M-estimator
matriisin tekijöihinjako	matrix factorization	meta-analyysi	meta-analysis
matriisiotanta	matrix sampling	metametri	metameter
matriisivasteregressio	matrix variate regression	metaregressio	meta regression
matrisoida	matricize	metodologia	methodology
matrisointi	matricization	Metropolisin–Hastingsin algoritmi	Metropolis–Hastings algorithm
mediaani	median	metsikkökuvio	forest plot
mediaanielinaika	median survival time	mielipidekysely	opinion survey, opinion poll
mediaaniharhattomuus	median unbiasedness	minimaalinen tyhjentävä tunnsluku	minimal sufficient statistic
mediaaniregressiokäyrä	median regression curve	minimaalisesti kytketty asetelma	minimally connected design
mediaanitesti	median test	minimax-kriteeri	minimax criterion
mediaatio	mediation	mitallinen	measurable
mediaatioanalyysi	mediation analysis	mitta	measure
melkein kaikkialla	almost everywhere	mitta-asteikko	measurement scale
melkein varma suppeneminen	almost sure convergence	mittakaava	scale_2
melkein varmasti	almost surely	mittateoria	measure theory
menettelyharha	procedural bias	mittaussarja	serial measurements
menetys	regret	mittausvirhe	measurement error
menneistäminen	backcasting	m.k.	a.e.
merkintä-takaisinpyyntiotanta	capture-recapture sampling	momentit generoiva funktio	moment generating function

momentti	moment	variate analysis	
momenttiemäfunktio	moment generating function	monimuuttujainen	multivariate
momenttiestimaattori	moment estimator	monimuuttujainen laskuriprosessi	multivariate counting process
momenttimatriisi	moment matrix	moninormaalijakauma	multinormal distribution
momenttimenetelmä	method of moments	moniositus	multiple stratification
momenttisuhde	moment ratio	moniosumamalli	multihit model
monen selittäjän regressio	multiple regression	monipaikkaus	multiple imputation
monen vasteen regressio	multivariate regression	monipoistumamalli	multiple decrement model
monen vasteen varianssianalyysi	multivariate analysis of variance	monipäätösmenetelmät	multiple criteria decision making
moniasteinen estimointi	multistage estimation	monisilotusmenetelmä	multiple smoothing method
moniasteinen otanta	multistage sampling	monisto	manifold
moniasteotanta	multistage sampling	monisto-oppiminen	manifold learning
monihuippuinen jakauma	multimodal distribution	monitasoitusmenetelmä	multiple smoothing method
moni-imputointi	multiple imputation	monitasomalli	multilevel model, hierarchical model
monijakoinen	multipartite	monitavoiteoptimointi	multiple criteria optimization
monijaksomalli	multiepisode model	monitavoitteinen päätösten-teko	multiple criteria decision making
monikehikkoluotaus	multiple-frame survey	monitehtäväoppiminen	multitask learning
monikeskuskoe	multicenter trial	monitekijäkoe	multifactorial design
monikeskustutkimus	multicenter study	monitestaus	multiple testing
monikätkäinen rosvo	multiarmed bandit	monitilainen elinajantaulukko	multistate life table
monilineaarinen prosessi	multilinear process	monitilamalli	multistate model
moniluokkainen	polytomous	moniulotteinen	multidimensional
monimuotoisuusindeksi	index of diversity	moniulotteinen aikasarja	multiple time series
monimuuttuja-analyysi	multi-	moniulotteinen jakauma	multivariate distribution

moniulotteinen mediaani	multi-variate median	design	
moniulotteinen merkkitesti	multivariate sign test	mukautuva päättely	adaptive inference
moniulotteinen sijatesti	multi-variate rank test	mukautuva regressio	adaptive regression
moniulotteinen skaalaus	multi-dimensional scaling	mukautuva tehostus	adaptive boosting
monivaiheinen otanta	multi-phase sampling	mukavuusotos	convenience sample
monivaiheotanta	multiphase sampling	multikollineaarisuus	collinearity, multicollinearity
monivaihepääätöskaavio	multi-valued decision diagram	multilineaarinen prosessi	multilinear process
monivaiheiselontaa	multiphase screening	multinomijakauma	multinomial distribution
monivastausanalyysi	multiple response analysis	multinomikerroin	multinomial coefficient
moniverkko	multigraph	multinomilogistinen regressio	multinomial logistic regression
monivertailu	multiple comparison	multinomilogit-malli	multinomial logit model
monivertailukontrolloitu	multiplicity controlled	multinormaalijakauma	multinormal distribution
moniyhtälömalli	multiequational model	multiplikatiivinen kausivaihtelumalli	multiplicative seasonal model
monotoninen	monotone	multiplikatiivinen malli	multiplicative model
monotoninen uskottavuusosamäärä	monotone likelihood ratio	multiverkko	multigraph
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	muunnos	transformation

muunnosarvo metameter
muutosarvo change score
muutospiste change point
muutostahti rate of change
muutosvauhti rate of change
muuttuja variable, variate
muuttujakato item non-response
muuttujalavennus data augmentation
muuttujavalinta variable selection
m.v. a.s., almost surely
myyntioptio put option
myöntyväisyysarha acquiescence bias
myötävirtaneuroverkko feed-forward neural network
määrittelyarha specification bias
määrittelyvirhe specification error
määrittely assay
määrittelyskoe assay
määrällinen quantitative

N

naapurustouputus neighbor embedding
naiivi ennuste naive forecast
napakoordinaatti polar coordinate
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 hypergeometri-
nen 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	mal approximation
nimellinen peittotoden-		normaaliarvo normal score
näköisyys nominal coverage prob-	ability	normaalijakauma normal distri-
nimittäjä denominator		bution, Gaussian distribution
nimeämissääntö labeling rule		normaalijakaumakuvio normal
nimiöimätön data unlabeled	data	probability plot
nimiöity data labeled data		normaalisuus normality
nolla-avaruus null space		normaaliyhtälöt normal equa-
nolladevianssi null deviance		tions
nollahypoteesi null hypothesis		normalisoiva muunnos normal-
nollainflatoitu zero-inflated		izing transformation
nollajakauma null distribution		normitettu muuttuja standard-
nollajoukko null set		ized variable
nollakatkaistu jakauma zero-	truncated distribution	normitettu normaalijakauma
nollamalli null model		standard normal distribution
nollamatriisi zero matrix		numeerinen optimointi numeri-
nollapaisunta zero-inflation		cal optimization
nollapaisunut jakauma zero-	inflated distribution	numeraari numeraire
nollasummapeli zero-sum game		numeroituva countable
nollatypistetty jakauma zero-	truncated distribution	numeroituvasti ääretön denu-
nomogrammi nomogram		merable
nopea Fourier-muunnos fast	Fourier transform	numerosuosio digit preference
nopea harhakorjattu luotta-		nuppineulakuvio lollipop plot
musväli bias-corrected and accel-	erated interval	nykyistäminen nowcasting
nopeutettu elinaikatestaus ac-	celerated life testing	näennäisesti riippumattomat
nopeutettu harhakorjattu lu-		regressiot seemingly unrelated
ottamusväli bias-corrected and	accelerated interval	regressions
nopeutetun ikääntymisen		näennäiskorrelaatio spurious
malli accelerated life model		correlation
nopeutetun vikaantumisajan		näennäissatunnaisluku pseudo-
malli accelerated failure time	model	random number
normaaliapproksimaatio nor-		näennäistarkkuus spurious preci-
		sion
		näyte sample ₂
		näytteistys sampling ₂
		näytteistää sample ₄
		näyttö evidence
		näytön alaraja evidence lower
		bound
		näytön painoarvo weight of evi-
		dence
		näytön yläraja evidence upper
		bound

O

oblimax-rotatio	oblimax rotation	oletus	assumption
oblimin-rotatio	oblimin rotation	omatodennäköisyys	personal probability
odotettu lukumäärä	expected frequency, expected number	omatoimiluotaus	self-administered survey
odotetun hyödyn teoria	expected utility theory	omien lasten menetelmä	own children method
odotusaika	waiting time	ominaisarvo	eigenvalue
odotusarvo	expectation, expected value	ominaisarvohajotelma	spectral decomposition
odotusarvoon palautuva	mean reversion	ominaisarvoyhtälö	characteristic equation
odotustappio	expected loss	ominaisavaruus	eigenspace
odotustappioarvo	expected shortfall	ominaisuus	attribute, characteristic
offset-termi	offset term	ominaisvektori	eigenvector
ohjaamaton luokittelu	unsupervised classification	onnistumistodennäköisyys	success probability
ohjaamaton oppiminen	unsupervised learning	operaatiotutkimus	operations research
ohjaava analytiikka	prescriptive analytics	opetusaineisto	training data
ohjattu luokittelu	supervised classification	optimaalinen kiintiöinti	optimal allocation
ohjattu oppiminen	supervised learning	optimaalisuus	optimality
ohuthäntäinen jakauma	light-tailed distribution	optimiasetelma	optimal design
ohutlevysplini	thin plate spline	optimikiintiöinti	optimal allocation
oikealle jatkuva vasemmalta raja-arvot	continue à droite, limite à gauche, càdlàg	optimiprojektio	projection
oikealle vinous	positive skewness	optimiskalaus	optimal scaling
oikeiden negatiivisten osuus	true negative rate	optimointi	optimization
oikeiden positiivisten osuus	true positive rate	optimointimenetelmä	optimization method
oikeustilastotiede	jurimetrics	ordinaaliasteikko	ordinal scale
oktiili	octile	ordinaaliregressio	ordinal regression
olennaiset piirteet	stylized facts	ortogonaalinen	orthogonal, perpendicular
		ortogonaalipolynomi	orthogonal polynomial
		ortogonaaliprojektio	orthogonal projection
		ortonormaallinen	orthonormal

osajoukko	subset		hood
osamäärä	ratio, quotient	osittaisyhteys	partial association
osaotos	subsample	osittaminen	decomposition ₁
osaparametrinen	semi-parametric	osittuva malli	decomposable model
osaruutuasetelma	split-plot design	ositus₁	partition
osataulukko	partial table	ositus₂	stratification
osayksikköasetelma	split-unit design	osoitin	indicator variable
osioanalyysi	item analysis	osoitinmatriisi	incidence matrix ₁
osiovasteteoria	item-response theory	osoittaja	numerator
osite	stratum ²	osto-optio	call option
ositettu malli	stratified model	osuma-aika	hitting time
ositettu matriisi	partitioned matrix	osumatodennäköisyys	hitting probability
ositettu otanta	stratified sampling	osuus	proportion
ositettu pylväskuvio	multiple bar chart	otannasta johtumaton virhe	non-sampling error
osittainen pienin neliösumma	partial least squares	otanta	sampling ₁
osittainen piirakkakuvi	partial pie chart	otanta-asetelma	sample design
osittainen toisto	fractional replication	otantaharha	sampling bias
osittain verrannolliset vastasuhteet	partial proportional odds	otantajakauma	sampling distribution
osittaisassosiaatio	partial association	otantakehikko	sampling frame
osittaisautokorrelaatio	partial autocorrelation	otantamenetelmä	sampling method
osittaisjäännöskuvio	partial residual plot	otantaosuus	sampling fraction
osittaiskorrelaatio	partial correlation	otanta palauttaen	sampling with replacement
osittaisregressiokerroin	partial regression coefficient	otanta palauttamatta	sampling without replacement
osittaisuskottavuus	partial likeli-	otantasuunnitelma	sampling plan, sample plan
		otantatarkastus	sampling inspection
		otanta vaihtelevin todennäköisyyksin	unequal probability sampling
		otantavirhe	sampling error
		otantayksikkö	sampling unit
		otos	sample ₁
		otosasetelma	sample design

²Monikkomuoto on *strata*.

otosavaruus sample space
otosjakauma sample distribution
otoskertymäfunktio sample distribution function
otoskeskiarvo sample mean
otoskeskivirhe sample standard error
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
palkkikaavio bar chart
palkkikuvio bar chart
pallokoordinaatti spherical coordinate
pallopintajakauma spherical distribution₂
pallopintamuuttuja spherical variable
pallopintatesti sphericity test
pallosymmetrisen jakauma spherical distribution₁
paloittainen regressio piecewise regression
paloittain lineaarinen piecewise linear
paloittain vakio piecewise constant
paluuaika recurrence time₁
paluujakso return period
paluukvantiili return level
paneeliaalto panel wave
paneeliaineisto panel data
paneelidata panel data
paneelitutkimus panel study
panos-tuotosmalli input-output model
panos-tuotosematriisi input-output matrix
parametri parameter
parametriavaruus parameter space
parametrinen jakauma para-

metric distribution	peittoharha noncoverage bias
parametrinen malli parametric model	peittotodennäköisyys coverage probability
parametrinen päättely parametric inference	peittovaikutus masking effect
parametriton nonparametric	peliteoria game theory
parametrivetoinen malli parameter-driven model	pelkistetty muoto reduced form
parametrointi parametrization	pelkistymätön ketju irreducible chain
paras asymptoottisesti normaalin estimaattori best asymptotically normal estimator	pelurin vararikko gambler's ruin
paras lineaarinen harhaton ennustin best linear unbiased prediction	pelurin virhepäättelmä gambler's fallacy
paras lineaarinen harhaton estimaattori best linear unbiased estimator	penetranssi penetrance
paremmuuskoe superiority trial	perhekohtainen virhetodennäköisyys family-wise error rate
Pareto-jakauma Pareto distribution	perheenpääkohtainen menetelmä headship rate method
parhaan olosuhteen elinajanodote best practice life expectancy	periodogrammi periodogram
pariotokset paired samples	periytyvyys heritability
pariteetti parity ₂	perkolaatioteoria percolation theory
pariteettisuhde parity-progression ratio	permutaatio permutation
parittainen riippumattomuus pairwise independence	permutaatiotesti permutation test
parittainen <i>t</i>-testi paired <i>t</i> -test	perspektiivikuvio perspective plot
parivertailu paired comparison	perturbaatiomenetelmä perturbation method
<i>p</i>-arvo <i>p</i> -value	perusjakso base period
<i>p</i>-arvon kalastelu <i>p</i> -hacking	perusjoukko population ₁
Pearson-jäännös Pearson residual	perusotos master sample
Pearsonin khi-neliösuure Pearson's chi-squared statistic	perusotospaino basic sample weight
Pearsonin khi-neliötesti Pearson's chi-squared test	perustartuttavuusluku basic reproduction number
pehmo-laskenta soft computing	perustaso baseline
peilattu tiheyskuvio mirrored density plot, mirror density chart	perustasoharha base rate bias
	perusuhka baseline hazard
	perusuusiutumisluku basic reproduction number
	peräkkäinen todennäköisyys-suhdetesti sequential probability ratio test

peräkkäisallokointi	sequential allocation	piilokerros	hidden layer
peräkkäisanalyysi	sequential analysis	piilevä	latent
peräkkäisasetelma	sequential design	piilokki	inlier
peräkkäiskoe	sequential experiment	piilomarkovmalli	hidden Markov model
peräkkäismalli	sequential model	piiloluokka-analyysi	latent class analysis
peräkkäisotanta	sequential sampling	piilomuuttuja	latent variable
peräkkäispoiminta	sequential sampling	piilopiirreanalyysi	latent trait analysis
peräkkäistesti	sequential test	piilorakenne	latent structure
pesitetty tapaus-verrokkitutkimus	nested case-control study	piilosekoittaja	unobserved confounder
Phillipsin käyrä	Phillips curve	piilosiiirtymäanalyysi	latent transition analysis
pienalue-estimointi	small area estimation	piilotusprojektiio	latent projection
pienimmän itseiskutistamisen ja valinnan operaattori	least absolute shrinkage and selection operator	piirakkakaavio	pie chart, pie diagram
pienimmän itseispoikkeaman menetelmä	least absolute deviation method	piirakkakuvio	pie chart, pie diagram
pienimmän neliösumman estimaattori	least squares estimator	piirre	attribute, feature, trait
pienimmän neliösumman menetelmä	method of least squares	piirreavaruus	feature space
pienimmän varianssin harhaton estimaattori	minimum variance unbiased estimator	piirrevalinta	feature selection
pienin khi-neliö	minimum chi-square	piktogrammi	pictogram
pienien lukujen laki	law of small numbers	pilotti	pilot study
Pietarin paradoksi	St. Petersburg paradox	pinoaminen	stacking
piikki ja laatta	spike and slab	pinomalli	stacking model
piikki ja laatta -jakauma	spike and slab distribution	pistebiseriaallinen korrelaatio	point biserial correlation
		piste-estimaatti	point estimate
		piste-estimaattori	point estimator
		piste-estimointi	point estimation
		pistehypoteesi	simple hypothesis
		pisteittäinen luottamusvyöhyke	pointwise confidence band
		pistekohdistettu aineisto	point referenced data
		pistekohina	shot-noise
		pistekuvio₁	dot plot
		pistekuvio₂	point pattern

pistekuvioaineisto	point pattern data	Poisson-otanta	Poisson sampling
pistemäärä	score ₂	Poisson-prosessi	Poisson process
pisteotanta	point sampling	poissulkevat tapahtumat	mutually exclusive events, disjoint events
pisteprosessi	point process	poissulkukriteeri	exclusion criterion
pistetiheys	point density	poisto	elimination
pistetodennäköisyys	point probability	poistomenetelmä	removal method
pistetulo	dot product, inner product	poistovalinta	backward elimination
pisteytys	scoring	polkuanalyysi	path analysis
pistoke-estimaattori	plug-in estimator	polkukerroin	path coefficient
pitkittäisaineisto	longitudinal data	polkuriippuvainen	path dependent
pitkäaikaismuutos	trend	Pólyan uurnamalli	Pólya urn model
pitkähäntäinen jakauma	long-tailed distribution	polykorinen korrelaatio	polychoric correlation
pitkämuistinen prosessi	long-memory process	polynominen trendi	polynomial trend
pitkän kantaman riippuvuus	long-range dependence	polynomiregressio	polynomial regression
Pitman-tehokkuus	Pitman efficiency	polyspektri	polyspectrum
pito-otanta	sampling without replacement	populaatio	population
pituusharhainen otanta	length-biased sampling	populaatiogenetiikka	population genetics
PNS-estimaattori	least squares estimator	populaatiokeskiarvoistettu malli	population-averaged model
poikittaistutkimus	cross-sectional study	populaation kasvumalli	population growth model
poikkeaman itseisarvo	absolute deviation	populaatiopyramidi	population pyramid
poimia	sample ₃	porrasasetelma	staircase design
poimija	sampler	porraskiila-asetelma	stepped wedge design
poiminta	sampling ₁	porrasteinen aloitus	staggered entry
poimintamenetelmä	sampling method	positiividefiniitti	positive definite
poimintaväli	sampling interval	positiivinen ennustearvo	positive predictive value
poismuutto	out-migration		
Poisson-jakauma	Poisson distribution		

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	protokollariike	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	psykometria	psychometrics
profiointi	profiling	psykometriikka	psychometrics
projektio	projection	puhdas luonnollinen epäsuora vaikutus	pure natural indirect effect
projektiokuvio	projection plot	puhdas luonnollinen suora vaikutus	pure natural direct ef-
projektion haku	projection pursuit		

fect	pal component analysis
puhdistumisjakso wash-out period	pääkomponenttiregressio principal component regression
puheentunnistus speech recognition	pääkoordinaatti principal coordinate
puhelinnumeron satunnaisvalinta random digit dialing	päällekkäistyminen overplotting
puoleen tehoava annos median effective dose	pääositus principal stratification
puoleen tehoava taso median effective level	päätetapahtuma endpoint
puolet tappava annos median lethal dose	päättelin perceptron
puolet tappava taso median lethal level	päätely inference
puolimartingaali semi-martingale	päätösavaruus decision space
puolinormaali jakauma half-normal distribution	päätösfunktio decision function
puolitusmenetelmä split-half method	päätöspuu decision tree
puukaavio tree diagram	päätössääntö decision function
puppukorrelaatio nonsense correlation	päätösteoria decision theory
puumalli tree model	päävaikutus main effect
puuttuva data missing data	päävaste primary outcome
puuttuvuuden sekoitemalli pattern mixture model	
puuttuvuus rakenne missing data pattern	
pylväskaavio bar chart	
pylväskuvio bar chart	
pysyvä elinkorko perpetuity	
pysäytyshetki stopping time, optional time	
pysäytyssääntö stopping rule	
pyöristys rounding	
pätevyys validity	
pääakselifaktorointi principal axis factoring	
pääakselihajotelma principal axes decomposition	
pääkomponenttianalyysi principal component analysis	

R

R² coefficient of determination
raaka-arvo raw score
raakamomentti raw moment
radiaalifunktio radial function
Radonin–Nikodymin derivaatta Radon–Nikodym derivative
rahasto fund
rajahinta strike price
rajajakauma asymptotic distribution, limiting distribution
rajapiste cut-off point
rajatta jakautuva infinitely divisible
rajatuloestimaattori product limit estimator
rajoitettu elinajanodote restricted mean survival time
rajoitettu satunnaistus restricted randomization
rajoitettu suurimman uskottavuuden menetelmä restricted maximum likelihood

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

reunakausaalivaikutus	marginal causal effect	riippuvuus edeltävästä	antependence
reunallistaa	marginalize	rimamalli	hurdle model
reunalukumäärä	marginal frequency	rinnakkaisten koordinaattien	
reunamalli	marginal model	kuvio	parallel coordinate plot
reunarakennemalli	marginal structural model	rinnakkaisten ryhmien	
reunariippumattomuus	marginal independence	asetelma	parallel groups design
reunariippuvuus	marginal dependence	riski	risk
reunasumma	marginal total	riskiero	risk difference
reunataulukko	marginal table	riskifunktio	risk function
reunatiheysfunktio	marginal density function	riskihakuinen	risk prone
reunatodennäköisyys	marginal probability	riskihakuisuus	risk proneness
reunauskottavuus	marginal likelihood	riskijoukko	risk set
reunavaikutus	marginal effect	riskijoukkopaiminta	risk set sampling
reunayhteys	marginal association	riskikertymä	cumulative risk
riippumaton muuttuja ³	explanatory variable	riskineutraali mitta	risk neutral measure
riippumaton sensurointi	independent censoring	riskin kaihtaminen	risk aversion
riippumattomat ja normaali-jakautuneet	normally and independently distributed	riskisuhde	relative risk, risk ratio
riippumattomat ja samoin jakautuneet	independent and identically distributed	riskitekijä	risk factor
riippumattomat lisäykset	independent increments	riskiteoria	risk theory
riippumattomien komponenttien analyysi	independent component analysis	riskitön arvopaperi	riskless asset
riippumattomuus	independence	riskiä kaihtava	risk averse
riippuva muuttuja ⁴	outcome variable, response variable	ristiinsovitus	cross-fitting
riippuvuus	dependence	ristiintaulukointi	cross tabulation
		ristiinvahvistus	cross-validation
		ristiinvalidointi	cross-validation
		ristikkäisasetelma	cross-over design
		ristikorraatio	cross-correlation
		ristikovarianssi	cross-covariance
		ristispektri	cross spectrum
		ristispektritiheys	cross-spectral density
		ristisuhde	odds ratio
		ristitaulukko	cross table
		ristitulosuhde	cross-product ratio, odds ratio

³Suosittelavampi on selittävä muuttuja.

⁴Suosittelavampia ovat vastemuuttuja ja selitettävä muuttuja.

robusti estimaattori	robust estimator	screening design
robustisuus	robustness	ryhmätason tutkimus
ROC-käyrä	ROC curve	aggregate-level study
roikkuva histogrammi	hanging rootogram	ryväs
rososakko	roughness penalty	cluster
rotaatio	rotation	ryväskohtainen jakauma
rotaatioasetelma	rotatable design	cluster-dependent distribution
rukattu estimaattori	adjusted estimator	ryväsootanta
rukattu otospaino	adjusted sample weight	cluster sampling
rukattu profiiliuskottavuus	adjusted profile likelihood	ryväsprosessi
rukkaus	adjustment	cluster process
ruudukko	grid	ryvässatunnaistus
ruusukaavio	rose diagram	cluster randomization
ruutuotanta	grid sampling	ryvästynyt aineisto
ryhmien sisäinen neliösumma	within-groups sum of squares	clustered data
ryhmien välinen neliösumma	between-groups sum of squares	ryvästäminen
ryhmitelty aineisto	grouped data	clustering
ryhmittelyanalyysi	cluster analysis	ryvästämisanalyysi
ryhmittelypuu	dendrogram	cluster analysis
ryhmäfaktori	group factor	ryöppyprosessi
ryhmäjaollinen asetelma	group divisible design	cascade process
ryhmäjaon salaaminen	allocation concealment	räjähtävä prosessi
ryhmäkaltaistus	group matching, frequency matching	explosive process
ryhmäpaikkaus	aggregate imputation	
ryhmäperäkkäinen koe	group-sequential experiment	
ryhmäperäkkäisasetelma	group-sequential design	
ryhmäseulonta-asetelma	group	

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
samahäntäinen	equi-tailed

samanaikainen estimointi	simultaneous estimation	satunnaisheilunta	stochastic volatility
samanaikainen päättely	simultaneous inference	satunnaisilmiö	random phenomenon
samanaikaiset luottamusvälit	simultaneous confidence intervals	satunnaisjako	random allocation
samanlaiset parit	concordant pairs	satunnaisjono	random sequence
samanlaisuuskerroin	similarity coefficient	satunnaisjoukko	random set
samanlaisuusmatriisi	similarity matrix	satunnaiskenttä	random field
samanlaisuusmitta	similarity measure	satunnaiskerroin	random coefficient
samanveroisuus	non-inferiority	satunnaiskomponentti	random component
samanveroisuuskoe	non-inferiority trial	satunnaiskulku	random walk
samavarianssinen	homoskedastic	satunnaisluku	random number
sanapilvi	word cloud	satunnaislukugeneraattori	random number generator
sanapussimalli	bag-of-words model	satunnaislukukehitin	random number generator
sandwich-estimaattori	sandwich estimator	satunnaismatriisi	random matrix
sarjakorrelaatio	serial correlation	satunnaismetsä	random forest
sarjalaimennuskoe	serial dilution assay	satunnaismitta	random measure
sarjaseulonta	high-throughput screening	satunnaismosaiikki	random tessellation
sarjatoistopyynti	multiple recapture design	satunnaismuuttuja	random variable
sarjavaihtelu	serial variation	satunnaisotanta	random sampling
s-arvo	s-value	satunnaisotos	random sample
sateenvarjokoe	umbrella trial	satunnaispaikkaus	stochastic imputation
sateenvarjo-ongelma	umbrella problem	satunnaispoiminta	random selection
satulapisteapproksimaatio	saddle point approximation	satunnaisprosessi	random process, stochastic process
satunnainen	random	satunnaissuure	random variable
satunnaisesti puuttuva	missing at random	satunnaistamaton	non-randomized
satunnaisfunktio	random function	satunnaistapahtuma	random event
		satunnaistermi	innovation term
		satunnaistessellaatio	random tessellation

satunnaistettu järjestys-	sekaspektri mixed spectrum
lukutesti rank-randomization	sekastrategia mixed strategy
test	sekoitekoe mixture experiment
satunnaistettu kliininen koe	sekoitemalli mixture model
randomized clinical trial	sekoittaja confounder
satunnaistettu koe randomized	sekoittava jakauma mixing dis-
experiment, randomized trial	tribution
satunnaistettu päätössääntö	sekoittava tekijä confounder,
randomized decision rule	confounding factor
satunnaistettu testi randomized	sekoittava muuttuja confounder,
test	confounding variable
satunnaistettu vastaus random-	sektorikuvio pie diagram, pie
ized response	chart
satunnaistettujen lohkojen	sekvenssimalli sequence model
asetelma randomized blocks	selitettävä muuttuja outcome
design	variable
satunnaistetut lohkot random-	selittäjä explanatory variable
ized blocks	selittävä koe explanatory trial
satunnaistuksen mukainen	selittävä muuttuja explanatory
analyysi as-randomized analy-	variable, regressor
sis	selitysaste coefficient of determi-
satunnaistus randomization	nation
satunnaistustesti randomization	selityskyky explanatory power
test	selvitys inquiry
satunnaisuus randomness	selviytyjäkohortti survival co-
satunnaisvaihtelu random varia-	hort
tion	selviytymissuhde survival ratio,
satunnaisvaikutus random effect	relative survival
satunnaisvaikutusmalli random-	semimarkovprosessi semi-Mar-
effects model	kov process
satunnaisvektori random vector	semimartingaali semi-martingale
satunnaisverkko random graph	semiparametrinen semi-
satunnaisvirhe random error	parametric
saturoitu malli saturated model	sensitiivisyys sensitivity
segregaatio segregation	sensurointi censoring
segregaatioanalyysi segregation	sensurointi oikealta right-cen-
analysis	soring
sekajakauma mixed distribution	sensurointi vasemmalta left
sekakoosteinen heterogeneous	censoring
sekamalli mixed model, mixed-	sensuroitu havainto censored
effects model	observation
sekanttimenetelmä secant meth-	sentiili centile
od	sentroidimenetelmä centroid

method	scale family
separoituva avaruus separable space	sijaintitunnusluku measure of location
seulaestimaattori sieve estimator	sijaismuuttuja proxy variable, surrogate variable
seulonta screening	sijaisvarianssi proxy variance
seulonta-asetelma screening design	sijaisvaste proxy outcome, surrogate outcome
seulontatesti screening test	sijakisa ranking ₁
seuranta follow-up	sijakorrelaatiomatriisi rank correlation matrix
seuruu follow-up	sijaluku rank ₁
Shannon-informaatio Shannon information	sijalukukorrelaatio rank correlation
sidos tie	sijalukuregressio rank regression
sigma-algebra sigma-algebra	sijapiste multivariate rank
sigmakenttä sigma-field	sijasummatesti rank sum test
sigmoidikäyrä sigmoid curve	sijatunnusluku rank order statistic
signaali-kohinasuhde signal to noise ratio	sijavertailu ranking ₁
signaalinkäsittely signal processing	sijoitus ranking ₂
sihteeriongelma secretary problem	sijoitusrahasto mutual fund
siilosirontakuvio binned scatter plot	silmukka loop
siirrettävyys transportability	silotin smoother
siirto translation	silottava splini smoothing spline
siirtofunktio transfer function	silotus smoothing
siirto-oppiminen transfer learning	siluettikuvio silhouette plot
siirtymäintensiteetti transition intensity	similaarinen testi similar test
siirtymämatriisi transition matrix	similaariset alueet similar regions
siirtymätaajuus transition rate	similaarit matriisit similar matrices
siirtymätodennäköisyys transition probability	simpleksialgoritmi simplex algorithm
sijaintiluku measure of location	Simpsonin paradoksi Simpson's paradox
sijaintimitta measure of location	simulaatioajo simulation run
sijaintiparametri location parameter	simulaatiotoisto simulation run
sijaintiperhe location family	simulaattori simulator
sijainti-skaalaperhe location-	simulointi simulation
	simuloitu hehkutus simulated annealing
	simuloitu verho simulated enve-

lope	uct
simultaaninen autoregressio	sisävarianssi intraclass variance
simultaneous autoregression	sisäänajojakso₁ burn-in period
simultaaniyhtälöharha simulta-	sisäänajojakso₂ run-in period
neous equations bias	sisäänmuutto in-migration
simultaaniyhtälömalli simulta-	sivuutettava ignorable
neous equations model	sivuuttamaton nonignorable
singulaariarvohajotelma singu-	skaala scale ₁
lar value decomposition	skaalakerroin scale factor
singulaarinen jakauma singular	skaalaparametri scale parameter
distribution	skaalattu devianssi scaled de-
singulaarinen matriisi singular	viance
matrix	skaalaus scaling
sironta scatter	skaalautuvuus scalability
sirontakuvio scatter plot	skalaari scalar
sirontakuviomatriisi scatterplot	skeptinen priori sceptical prior
matrix	skoori score ₁
sisäaita inner fence	skoorifunktio score function
sisäinen kasvuvauhti intrinsic	skoorijäännös score residual
growth rate	skoorisuure score statistic
sisäinen validiteetti internal va-	skooritesti score test
lidity	skoorivektori score vector
sisäinen varianssi internal vari-	S-käyrä sigmoid curve
ance	softmax-funktio softmax func-
sisäkkäinen asetelma nested	tion
design	sokkimalli shock model
sisäkkäinen luokittelu hierarchi-	sokkokoe blinded experiment
cal classification	sokkoutus blinding
sisäkkäinen otanta nested sam-	solmu node, vertex
pling	solmupiste knot
sisäkkäiset hypoteesit nested	solufrekvenssi cell frequency
hypotheses	solulukumäärä cell frequency
sisäkkäiset mallit nested models	solvenssi solvency
sisäkorrelaatio intraclass correla-	sovite fitted value
tion	sovite-ero difference in fits
sisällönanalyysi content analysis	sovitettu malli fitted model
sisältymistodennäköisyys inclu-	sovitteen hyvyys goodness of fit
sion probability	sovitteen hyvyystesti goodness
sisämääräytyneisyys endogene-	of fit test
ity	sovitus fitting
sisämääräytynyt muuttuja en-	spagettikuvio spaghetti plot
dogenous variable	spatiaalinen data spatial data
sisätulo inner product, dot prod-	spatiaalinen mediaani spatial

median	population
spatiaalinen merkki spatial sign	stereologia stereology
spatiaalinen paripotentiaali	stokastiikka stochastics
spatial pair-potential	stokastinen approksimaatio
spatiaalinen pisteprosessi spa-	stochastic approximation
tial point process	stokastinen integraali stochastic
spatiaalinen sijaluku spatial	integral
rank	stokastinen järjestys stochastic
spatiaalinen tilastotiede spatial	ordering
statistics	stokastinen malli stochastic
spektri spectrum	model
spektrianalyysi spectral analysis	stokastinen matriisi stochastic
spektrihajotelma spectral de-	matrix
composition	stokastinen optimointi stochas-
spektrikertymäfunktio spectral	tic optimization
distribution function	stokastinen paikkaus stochastic
spektrisäde spectral radius	imputation
spektritiheys spectral density	stokastinen prosessi stochastic
spektrivällys spectral gap	process
spesifisyys specificity	stokastinen raja-arvo limit in
splini spline	probability
stabiili jakauma stable distribu-	stokastinen riippuvuus stochas-
tion	tic dependence
stabiili tila stable state	stokastinen suppeneminen
stabiili väestö stable population	stochastic convergence
stabiili yksilövaste stable-unit-	stokastinen volatiliiteetti
treatment-value assumption	stochastic volatility
standardinormaalijakauma	stokastisesti rajoitettu bounded
standard normal distribution	in probability
standardointi standardization ₂	Studentin jakauma Student's
standardoitu jännös standard-	distribution, <i>t</i> -distribution
ized residual	stylometria stylometry
standardoitu kerroin standard-	subjektiivinen todennäköisyys
ized coefficient	subjective probability
standardoitu muuttuja stan-	submartingaali submartingale
dardized variable	SU-estimaattori maximum likeli-
standardoitu vaikutuskoko	hood estimator, ML estimator
standardized effect size	suhdanne business cycle
stationaarinen jakauma station-	suhde ratio
ary distribution	suhdeasteikko ratio scale
stationaarinen prosessi station-	suhde-estimaattori ratio estima-
ary process	tor
stationaarinen väestö stationary	suhdeluku ratio

suhteellinen elossaolo	relative survival	suojaumisstrategia	hedging strategy
suhteellinen frekvenssi	relative frequency	suorakaidehila	rectangular lattice
suhteellinen kiintiöinti	proportional allocation	suora uskottavuus	direct likelihood
suhteellinen otanta	proportional sampling	suora vaikutus	direct effect
suhteellinen riski	relative risk, risk ratio	suora vakiointi	direct standardization
suhteellinen selviytyminen	relative survival	supereksogeenisuus	super exogeneity
suhteellinen tehokkuus	relative efficiency	supermartingaali	supermartingale
suhteellinen uskottavuus	relative likelihood	supernormalisuus	supernormality
suhteellinen varianssi	relative variance	superoppija	super learner
suhteellisen riskin vähenemä	relative risk reduction	superpopulaatiomalli	superpopulation model
sukukorrelaatio	familial correlation	supistettu tarkastus	reduced inspection
sukupuuton todennäköisyys	extinction probability	suppeneminen	convergence
sulautuminen	coalescence	suppilokuvio	funnel plot
suljettu monitestausta	closed multiple testing	suunnallinen satunnaiskulku	random walk with drift
suljettu väestö	closed population	suunnattu syklistön verkko	directed acyclic graph
sulkeuma	closure	suunnitellusti puuttuva	missing by design
sumea joukko-oppi	fuzzy set theory	suunniteltu vertailu	planned comparison
sumea logiikka	fuzzy logic	suunta	drift ₁
summakaavio	cumulative sum chart	suunta-aineisto	directional data
summakontrasti	sum contrast	suuntaamaton verkko	undirected graph
summautuva malli	additive model	suuraineisto	big data
summautuvien uhkien malli	additive hazards model	suurdata	big data
suodatin	filter	suurimman entropian menetelmä	maximum entropy method
suodatus	filtering	suurimman jälkitiheyden alue	highest posterior density region
suodin	filter	suurimman jälkitiheyden estimaattori	maximum a posteriori
suojaattu tieto	protected data		

estimator	syntymä-kuolemaprosessi birth-death process
suurimman jälkitiheyden väli	syntymäprosessi birth process
highest posterior density interval	syntyvyys birth rate
suurimman posterioritiheyden alue highest posterior density region	systemaattinen allokointi systematic allocation
suurimman posterioritiheyden estimaattori maximum a posteriori estimator	systemaattinen jako systematic allocation
suurimman posterioritiheyden väli highest posterior density interval	systemaattinen katsaus systematic review
suurimman uskottavuuden estimaattori maximum likelihood estimator	systemaattinen otanta systematic sampling
suurimman uskottavuuden menetelmä method of maximum likelihood	systemaattinen virhe systematic error, bias
suurin siedetty annos maximum tolerated dose	sysäys perturbation ₂
suuriulotteinen aineisto high-dimensional data	sysäysvastefunktio impulse response function
suuriulotteinen tilastotiede high-dimensional statistics	syväoppiminen deep learning
suurten lukujen laki law of large numbers	syyketju causal chain
suurten otosten teoria large sample theory	syykohtainen uhka cause-specific hazard
suurten poikkeamien teoria large deviation theory	syyksilukuosuus attributable fraction
suuruusluokka order of magnitude	syymuuttuja cause variable
sykli cycle	syy-seurausmalli causal model
syklinen malli cyclic model	syy-seurauspäätely causal inference
symmetrinen jakauma symmetric distribution	syy-seuraussuhde causality
symmetrinen otanta symmetric sampling	syötekerros input layer
synteettinen aineisto synthetic data	sädekuvio radial plot
synteettinen estimaattori synthetic estimator	säännöllisesti vaihteleva regularly varying
syntymäkohortti birth cohort	säännön edeltäjäosa antecedent of a rule
	säännön seuraajaosa consequent of a rule
	säännön vartalo body of a rule
	sääntely regularization
	säästeliäisyys parsimony
	säätömuuttuja control variable ₁
	säätöteoria control theory

T

taaajuus	frequency ₂ , rate ₁	tapahtumaväliaika	interevent time
taaajuusalue₁	frequency domain	tapaus₁	case
taaajuusalue₂	bandwidth ₁	tapaus₂	instance
taaajuuspinta	frequency surface	tapaus-alustatutkimus	case-base study
taaajuusvastefunktio	frequency response function	tapausavaruus	instance space
taakselaskenta	backcalculation	tapaus-kohorttitutkimus	case-cohort study
taakseprojisointi	backprojection	tapauskuolleisuus	case fatality ratio
taaksesovitus	backfitting	tapaus-pohjatutkimus	case-base study
taakseyhtälö	backward equation	tapaus-verrokkitutkimus	case-control study, case-referent study
taikaneliöasetelma	magic square design	tappio	loss
taitekohta	turning point	tappioarvo	Value at Risk
taitettu jakauma	folded distribution	tappiofunktio	loss function
taitettu kontingenssitaulukko	folded contingency table	tappiokolmio	loss triangle
taitettu mediaani	folded median	tappiomatriisi	loss matrix
taitettu normaalijakauma	half-normal distribution	tappion karttaminen	loss aversion
taitettu ristitaulukko	folded contingency table	tarkentuvuus	consistency
taittaminen	folding	tarkkuus₁	accuracy
takaovikorjaus	backdoor adjustment	tarkkuus₂	specificity
takaovipolku	backdoor path	tarttuvuus	transmission rate ₁
takaovivakiointi	backdoor adjustment	tartuntatodennäköisyys	transmission probability
taksonomia	taxonomy	tasa-arvokäyrä	contour line
taluskuvio	scree plot	tasa-arvosuora	line of equal distribution
talustesti	scree test	tasaerä	annuity ₂
tapahtuma	event	tasahavainto	tie
tapahtuma-aika	time-to-event	tasainen integroituvuus	uniform integrability
tapahtuma-aika-analyysi	time-to-event analysis	tasainen suppeneminen	uniform convergence
tapahtuma-alttiussuhde	occurrence-exposure rate	tasaisen jakautumisen suora	line of equal distribution
tapahtuma-avaruus	event space	tasaisen siirtymän autoregressiivinen malli	smooth transition autoregressive model
tapahtumahistoria-analyysi	event history analysis	tasaisesti integroituva	uni-
tapahtumatutkimus	event study		

formly integrable	tekijä factor ₁
tasajakauma uniform distribution	tekijäkoe factorial experiment
tasakoosteinen homogeneous	tekijäkoeasetelma factorial design
tasakorrelaatio uniform correlation	tekijän taso level of factor
tasapaino equilibrium	tekijänvaihtokriteeri factor reversal criterion
tasapainoinen otanta balanced sampling	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	tensoriarvoinen tensor-valued, tensor-variate
tasoitin smoother	teoreema theorem
tasoitus smoothing	termiinisopimus forward contract
tasoituskäyrä graduation curve	testi test
tasokartta level map	testiaineisto test data
taudinkantaja carrier of disease	testin empiirinen koko empirical size of test
taulukko table, array	testin koko size of test
tavoitekoe target trial	testipatteri battery of tests
tavoitekokeen jäljittely target trial emulation	testisarja test series
teoreettinen tilastotiede theoretical statistics	testisuure test statistic
teho efficacy	tetrakorinen korrelaatio tetrachoric correlation
tehokas estimaattori efficient estimator	tiedonlouhinta data mining
tehokkaiden markkinoiden hypoteesi efficient market hypothesis	tiedonsiirtonopeus transmission rate ₂
tehokkuus efficiency	tiedosto data set
tehollinen otoskoko effective sample size	tietoinen suostumus informed consent
tehollinen vapausasteluku effective degrees of freedom	tietokanta database
tehostus boosting	tietokoneavusteinen computer assisted
	tietosuoja disclosure control
	tietue record ₁
	tietuekytkentä record linkage
	tietämystekniikka knowledge technology, knowledge engineering
	tiheys density

tiheysestimointi	density estimation	tilastollinen kooste	statistical ensemble
tiheysfunktio	density function	tilastollinen malli	statistical model
tiheyshistogrammi	density histogram	tilastollinen merkitsevyys	statistical significance
tiheysotanta	density sampling	tilastollinen oppiminen	statistical learning
tiheyspoiminta	density sampling	tilastollinen perustelu	statistical reasoning
tiheysuhdemalli	density ratio model	tilastollinen päättely	statistical inference
tiiviste	hash	tilastollinen testi	statistical test
tiivistävä tunnusluku	summary statistic	tilastollisesti erottuva	statistically discernible
tikasindeksi	ladder index	tilastollisesti merkitsevä	statistically significant
tikasmuuttuja	ladder variable	tilastolukutaito	statistical literacy
tikkausmenetelmä	staircase method, up-and-down method	tilastotiede	statistics ₂ , statistical science
tikunkatkaisuprosessi	stick breaking process	tilastotieteellinen	statistical ₂
tila-aikaryvä	space-time cluster	tilastotieteilijä	statistician
tila-avaruus	state space	tilastotoimi	statistical service
tilajoukko	state space	tilastoviranomainen	statistical authority
tilanneparametrit	incidental parameters	tilastoyksikkö	observational unit
tilapistekuvio	spatial point pattern	tiukennettu tarkastus	tightened inspection
tilapisteprosessi	spatial point process	tiukin testi	most stringent test
tilasto	statistics ₁	tiukka järjestys	tight ordering
tilastoahdistus	statistics anxiety	<i>t</i>-jakauma	<i>t</i> -distribution
tilastografiikka	statistical graphics	<i>t</i>-jäännös	studentized residual
tilastollinen	statistical ₁	<i>t</i>-kerroinero	studentized difference in betas
tilastollinen asiantuntijajärjestelmä	statistical expert system	<i>t</i>-sovite-ero	studentized difference in fits
tilastollinen avaruus	statistical space	tobit-malli	tobit"-model
tilastollinen hypoteesi	statistical hypothesis	todella suurten lukujen laki	law of truly large numbers
tilastollinen järkeily	statistical reasoning	todennäköinen	probable
tilastollinen koneoppiminen	statistical machine learning	todennäköinen virhe	probable error

todennäköisyydet generoiva funktio	probability generating function	toisen asteen otantayksikkö	secondary sampling unit
todennäköisyys	probability	toisen kertaluvun tehokkuus	second order efficiency
todennäköisyysväli	probability interval	toisensa poissulkevat tapahtumat	mutually exclusive events
todennäköisyysarvio	probability judgement	toisinne	replicate
todennäköisyysavaruus	probability space	toisintavayttaminen Hilbertin avaruus	reproducing kernel Hilbert space
todennäköisyysmäfunktio	probability generating function	toissijainen vaste	secondary outcome
todennäköisyysjakauma	probability distribution	toistaminen	replication
todennäköisyyskuvio	probability plot	toistava	recursive
todennäköisyyslaskenta	probability calculus	toistettavuus	replicability
todennäköisyysmalli	probability model	toistettavuuskriisi	replication crisis
todennäköisyysmitta	probability measure	toisto	repetition
todennäköisyysotanta	probability sampling	toistokoe₁	repeated experiment, repeated trial
todennäköisyyspainotus	probability weighting	toistokoe₂	replication experiment, replication trial
todennäköisyyspäättely	probabilistic inference	toistokutsu	call-back
todennäköisyysrajat	probability limits	toistoluotaus	periodic survey, repeated survey
todennäköisyysteoria	probability theory	toistomittausasetelma	repeated measurements design, repeated measures design
todennäköisyystiheys	probability density function, density function	toisto-otanta	resampling
todennäköisyysydin	probability kernel	toistopyyntien asetelma	multiple recapture design
toimenpideanalyysi	intervention analysis	toistopyyntiotanta	capture-recapture sampling
toimintakykyiset elinvuodet	disability adjusted life years	toistotarkkuus	precision ₁
toimivuusmitta	operating characteristic	toistuvuus	repeatability
		toleranssi	tolerance
		toleranssiraja	tolerance limit
		toleranssiväli	tolerance interval
		tontiini	tontine
		tosiaikainen tutkimus	concurrent study
		toteuma	realization
		toteutettavuustutkimus	feasibil-

ity study	tion
toteutuma realization	tyhjentävä tunnusluku sufficient 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-arvo mode
Tukeyn vaihteluvälitesti Tukey's range test	työmuuttuja working variate
Tshebyshevin epäyhtälö Chebyshev's inequality	työttömyysaste unemployment rate
t-suure <i>t</i> -statistic	työttömyysosuus unemployment rate
t-testi <i>t</i> -test	työvoimatutkimus labor-force survey
t-tunnusluku <i>t</i> -statistic	tähtikuvio star plot
tukifunktio support function	tähtisadekuvio starburst plot
tukivektorikone support vector machine	tähtitilastotiede astrostatistics
tuloavaruus product space	tärkeysotanta importance sampling
tulointegraali product integral	täsmälleen identifioituva just identified
tulomitta product measure	täsmällisyys precision
tulomomentti product moment	täydellinen erottelu complete separation
tulomomenttikorrelaatio product moment correlation	täyden informaation menetelmä full information method
tulostekerros output layer	täydennetty Dickey–Fullerin testi augmented Dickey–Fuller test
tunneanalyysi sentiment analysis	täysehdoollistettu jakauma fully conditional distribution
tunnistin identifier	täysin satunnaisesti puuttuva missing completely at random
tunnusluku statistic	täysitekijäkoe full factorial design
tuottomatriisi pay-off matrix	
tutkimuksen toisto replication study	
tutkimus study	
tutkimuspopulaatio study population	
tutkimusväestö study population	
tutkimusyksikkö study unit	
t-vaihteluvälisuure studentized range statistic	
tyhjentyvyys sufficiency	
tyhjentävä dimension pienennys sufficient dimension reduction	

täyssatunnaistettu asetus completely randomized design
törmäytin collider
törmäytinharha collider bias

U

uhka hazard
uhkafunktio hazard function
uhkakertymä cumulative hazard, integrated hazard
uhkasuhde hazard ratio
uhkatiheys hazard rate, failure rate₁
ulkoaita outer fence
ulkoinen ennustaminen out-of-sample prediction
ulkoinen pätevyys external validity
ulkoinen testaus out-of-sample testing
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 multi-

ple range test
uskomusaste degree of belief
uskomusfunktio belief function
uskomusverkko belief network
uskottavuus likelihood
uskottavuusalue likelihood region
uskottavuusfunktio likelihood function
uskottavuusosamäärä likelihood ratio
uskottavuusosamääräsuure likelihood ratio statistic
uskottavuusosamäärätesti likelihood ratio test
uskottavuusperiaate likelihood principle
uskottavuusväli likelihood interval
utiliteetti utility
utiliteettiteoria utility theory
U-tunnusluku *U*-statistic
uudelleenpainotus reweighting
uudelleenparametointi reparametrization
uudelleen poimia resample
uudelleenpoiminta resampling
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
uusiotointo bootstrap replication
uusittavuus reproducibility
uusitumis aika recurrence time₂
uusitumis aikajakauma renewal distribution
uusitumis funktio renewal function
uusitumisluku reproduction number, reproductive number
uusitumis prosessi renewal process
uusitumis riski recurrence risk
uusitumistaajuus renewal rate
uusitumisteoria renewal theory
uusitumistiheys renewal rate
uusituva recurrent₂
uuspainotus reweighting
uusparametointi 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
vaihediatrammi 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 range
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 muovaaja effect modifier, moderator
vaikutuksen suuruus effect mag-

nitude	
vaikutuksen välittyminen	medi- ation
vaikutus₁	effect
vaikutus₂	influence
vaikutusfunktio	influence func- tion
vaikutuskoko	effect size
vaikutusmitan muovautuminen	effect measure modification
vaikutusparametrin muovautu- minen	effect measure modifica- tion
vaikutustunnusluku	influence statistic
vaimeneminen	attenuation
vajaa-asteinen matriisi	singular matrix
vajaa latinalainen neliö	incom- plete Latin square
vajaalohkoasetelma	incomplete block design
vajaasovitus	underfitting
vajaatehoinen estimaattori	in- efficient estimator
vajaavakiointi	underadjustment
vakaa estimaattori	robust esti- mator
vakaa tunnusluku	robust statis- tic
vakaa yksilövaste	stable-unit- treatment-value assumption
vakaus	robustness
vakinainen väestö	de jure popu- lation
vakioilmaantuvuus	standard rate
vakioimaton estimaattori	unad- justed estimator
vakiointi	adjustment, standardiza- tion ₁
vakioitu estimaattori	adjusted estimator
vakioitu kuolevuussuhde	stan- dardized mortality ratio
vakioiva muuttuja	adjusting variable, control variable ₂
vakio-otantaosuus	uniform sam- pling fraction
vakiotermi	intercept
vakioväestö	standard population
vakuuutus	insurance
vakuumustilastotiede	actuarial statistics
validiteetti	validity
validointi	validation
validointiaineisto	validation data
valikoitumisharha	selection bias
valikoitumismalli	selection model, sample selection model
valikoitumispriori	selection prior
valikoivan kasvun prosessi	pref- erential attachment process
valintaharha	selection bias
valintahetki	optional time, stop- ping time
valintaperusteinen otanta	choice-based sampling
valkoinen kohina	white noise
vallitsevuus	prevalence
valvontakortti	control chart
vanhoista vanhimmat	oldest old
vapaaehtoisharha	volunteer bias
vapaaehtoisnäyte	volunteer sam- ple
vapaaehtoisotos	volunteer sam- ple
vapausasteet	degrees of freedom
vapausasteluku	degrees of free- dom
vararikko-ongelma	ruin problem
vararikkotodennäköisyys	ruin probability
variaatioetäisyys	variational dis- tance
variaatiokerroin	coefficient of

variation	
variaatiopäätely	variational inference
varianssi	variance
varianssianalyysi	analysis of variance
varianssikomponentti	variance component
varianssin inflaatiokerroin	variance inflation factor
varianssin vakioiva muunnos	variance stabilizing transformation
varimax-rotaatio	varimax rotation
variogrammi	variogram
varmistus	ascertainment ₂
varsilehtikuvio	stem-and-leaf diagram
vasemmalle vinous	negative skewness
vastaaajaryhmä	response group
vastaaajavetoinen otanta	respondent-driven sampling
vastahypoteesi	alternative hypothesis
vastakaltaistus	counter-matching
vastakkaismerkkinen	antithetic
vastamuuttuja	antithetic variate
vastasuhde	odds
vastatapahtuma	complement event
vastatapahtuman todennäköisyys	complementary probability
vastatodennäköisyys	complementary probability
vastaus	response ₂ , answer
vastausalttiuspaino	response propensity weight
vastausaste	response rate
vastausilmaisoin	response indicator
vastausindikaattori	response indicator
vastauskato	nonresponse
vastausmekanismi	response mechanism
vastausvirhe	response error
vastavirta-algoritmi	backpropagation algorithm
vaste	outcome, response ₁
vasteaikajakauma	response time distribution
vasteen muunnos	response metameter
vastekerros	output layer
vastemuuttuja	outcome variable, response variable
vastepinta	response surface
vastepintamethodiikka	response surface methodology
vastetapahtuma	endpoint
vektoriaikasarja	vector time series
vektoriautoregressio	vector autoregression
vektorimuuttujainen	vector-valued, vector-variate
velkakirja	bond
Venn-diagrammi	Venn diagram
verhofunktio	envelope function
verhokäyräanalyysi	data envelopment analysis
verkko	graph ₃
verkkoinformaatiokriteeri	network information criterion
verkkohaastattelu	web-interview
verkkokysely	internet survey
verkkomalli	graphical model
verkko-otanta	network sampling
verkon moralisointi	moralization of graph
verkosto	configuration
verrannollisten uhkien malli	proportional hazards model

verrannollisten vastasuhteiden malli	proportional odds model	viive ⁵	lag
verrokki	control	viiveikkuna	lag window
verrokkilinja	control arm	viiveoperaattori	lag operator, backshift operator
verrokkiryhmä	control group	viivästynyt aloitus	delayed entry
vertaileva koe	comparative experiment	viivästysoperaattori	backshift operator
vertailulinja	control arm	vikaantuminen	failure
vertailuparametri	comparative parameter	vikaantumisaika	failure time
vertailuryhmä	comparison group, control group	vikakertymä	failure function
vetokerroin	odds	vikasuure	diagnostic statistic
vetosuhde	odds ratio	vikataajuus	failure rate ₂
vierashavainto	outlier	villi uusio-otanta	wild bootstrap
vierusluokkamalli	adjacent categories model	vino	oblique
vierusmatriisi	adjacency matrix	vinoheijastus	oblique reflection
vietto	drift ₁	vino jakauma	skew distribution
viiden luvun tiivistys	five-number summary	vinonormaalijakauma	skew-normal distribution
viimeisen havainnon toisto	last observation carried forward	vinoprojektio	oblique projection
viipalepoimija	slice sampler	vinosymmetrinen matriisi	skew-symmetric matrix
viipaloitu käänteinen regressio	sliced inverse regression	vinous	skewness
viipymisaika	occupation time, sojourn time	vioittuminen	failure
viisikkokriteeri	pentad criterion	vipuvoima	leverage
viitearvo	reference value	virallinen tilasto	official statistics
viitejakso	reference period, base period	virhe	error
viitepopulaatio	reference population	virheen itseisarvo	absolute error
viitepriori	reference prior	virheen kasautuminen	propagation of error, error propagation
viiteraja	reference limit	virheenkorjausmalli	error correction model
viiteväestö	reference population	virheen leviäminen	propagation of error, error propagation
viiteväli	reference interval	virheet selittäjissä -malli	errors in variables model
viivakoodikuvio	stripes plot	virhefunktio	error function
viivakuvio	line plot	virhekappaleiden osuus	fraction defective
viivaspektri	line spectrum	virhekertymä	cumulative error
		virhekeskineliö	error mean

⁵Taivutetaan “viipeen” tai “viiveen”.

square	väestötiede demography
virhekomponenttimalli error components model	vähiten merkitsevän erotuksen testi least significant difference test
virheluokittelu misclassification	vähäparametrisuus parsimony
virhemarginaali margin of error, error margin	väistytvä tila transient state
virheisuus error rate	välialalyysi interim analysis
virhetermi error term, disturbance ₂	väliestimointi interval estimation
virhevarianssi error variance	välikaltaistus caliper matching
virstanpylväsanalyysi landmark analysis	välimatka-asteikko interval scale
visualisointi visualisation	välinemuuttuja instrumental variable
viulukuvio violin plot	välisensurointi interval censoring
voima power ₁	välitapahtuma non-terminal event
voimafunktio power function	välittäjä mediator
voimakkain testi most powerful test	välitys mediation
voiman lainaaminen borrowing of strength	välitysanalyysi mediation analysis
volatiliteetti volatility	vällys gap
volyyymi-vaihtelu-vauhti volume, variety, velocity	väre wavelet
vuokaavio flow chart	värähdyslaajuus amplitude
vuorikiipeilyalgoritmi hill-climbing algorithm	värähtelevä prosessi oscillatory process
vuorottelevien ehdollistusten regressio alternating conditional expectations regression	värähtely oscillation
vuosiaineisto annual data	väsymismalli fatigue model
väestö population ₂	väärien löydösten osuus false discovery rate
väestölaskenta census ₂	väärien positiivisten osuus false positive rate
väestönkasvumalli population growth model	väärinluokittelu misclassification
väestönlaskija enumerator	väärinspesifointi misspecification
väestönmuutostiedot vital statistics	väärin spesifioitu malli misspecified model
väestönmuutostilasto vital statistics	väärä negatiivinen false negative
väestöpyramidi population pyramid	väärä positiivinen false positive
väestörakenne population structure	

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
ydfunktio kernel function
ydinmenetelmä kernel method
ydinodotusarvofunktio kernel mean function
ydinregressioestimaattori kernel regression estimator
ydinsilotus kernel smoothing
y dintasoitus kernel smoothing
y dintiheystestimaattori kernel density estimator
yhden indeksin malli single-index model
yhden linjan tutkimus single-arm study
yhden mukaisuus consistency₂
yhden mukaisuus₁ consistency₂
yhden mukaisuus₂ concordance
yhden mukaisuuskerroin coefficient of concordance
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
yhdistetty spektri mixed spectrum
yhdistettävyyys collapsibility
yhdysvaikutuksen kertaluku order of interaction
yhdysvaikutus interaction
yhteenlaskusääntö addition rule
yhteensopimattomuus incompatibility
yhteensopivuus₁ concordance
yhteensopivuus₂ compatibility
yhteensopivuusindeksi concordance index
yhteensopivuusväli compatibility interval, confidence interval
yhteisdeterminaatiokerroin coefficient of multiple determination
yhteisennustealue joint prediction region
yhteisesiintyvyys co-occurrence
yhteisesti tyhjentävät tunnusluvut jointly sufficient statistics
yhteisfaktori common factor
yhteisfaktorigalli common factor model₁
yhteisintegroitu vuus cointegration
yhteisjakauma joint distribution
yhteiskertymäfunktio joint cumulative distribution function
yhteiskorrelaatiokerroin multiple correlation coefficient
yhteiskuntatiete social data science
yhteiskuntatilastotiede social statistics

yhteisluottamusalue	joint confidence region		bandit
yhteismomentti	joint moment	yksikön sisältymistoden-	
yhteisregressio	joint regression	näköisyys	first-order inclusion probability
yhteisselitysaste	coefficient of multiple determination	yksilokeromalli	one-compartment model
yhteisten tekijöiden malli	common factor model ₂	yksilökohtainen malli	subject-specific model
yhteistesti	omnibus test	yksilö-muuttujakuvi	biplot
yhteistiheysfunktio	joint density function	yksimuuttujainen	univariate
yhteistodennäköisyys	joint probability	yksinkertainen hypoteesi	simple hypothesis
yhteisötunnistaminen	community detection	yksinkertainen satunnaisotanta	simple random sampling
yhteys	association	yksiosumamalli	one-hit model
yhteysanalyysi	association analysis	yksi-pois-ristiinvalidointi	leave-one-out cross-validation
yhteyskerroin	coefficient of association	yksisuuntainen luokittelu	one-way classification
yhteysluokka	communicating class	yksisuuntainen testi	one-sided test, one-tailed test
yhteysmitta	measure of association	yksisuuntainen vaihtoehtoinen hypoteesi	one-sided alternative hypothesis
yhtäläinen virheluokittelu	non-differential misclassification	yksisuuntainen varianssianalyysi	one-way analysis of variance
yksiasteotanta	single-stage sampling	yksisuuntainen vastahypoteesi	one-sided alternative hypothesis
yksihiippuinen	unimodal	yksitahoinen testi	one-sided test, one-tailed test
yksijakoinen	unipartite	yksittäisristiinvalidointi	leave-one-out cross-validation
yksikkö	unit	yksiulotteinen jakauma	univariate distribution
yksikköjuuri	unit root	yksöissokkokoe	single-blind experiment
yksikköjuuritesti	unit-root test	yleinen lineaarinen malli	general linear model
yksikköjuurta lähestyvä	near unit root	yleisfaktori	general factor
yksikkömatriisi	identity matrix	yleiskeskisarvo	grand mean
yksikkövastauskato	unit non-response	yleistetty additiivinen malli	generalized additive model
yksikköympyrä	unit circle		
yksikytkösryhmittely	single-linkage clustering		
yksikäsitteisyys	uniqueness		
yksikääinen rosvo	one-armed		

yleistetty autoregressiivinen ehdollisen heteroskedastisuuden malli	generalized autoregressive conditional heteroskedasticity model	ylitysarvotesti	exceedance test
yleistetty estimointiyhtälö	generalized estimating equation	yllätysarvo	surprisal value
yleistetty käänteismatriisi	generalized inverse	yläkolmiomatriisi	upper triangular matrix
yleistetty lineaarinen malli	generalized linear model	yläkvartiili	upper quartile
yleistetty momenttimenetelmä	generalized method of moments	ylävalvontaraja	upper control limit
yleistetty pienin neliösumma	generalized least squares	ylös-alasmenetelmä	staircase method, up-and-down method
yleistetty stokastinen prosessi	generalized stochastic process	ympäristötilastotiede	environmental statistics
yleistetty summautuva malli	generalized additive model	Yulen–Walkerin yhtälöt	Yule–Walker equations
yleistetty ääriarvojakauma	generalized extreme-value distribution		
yleistettävyyys	generalizability	Z	
ylisahonta	overdispersion	z-arvo	z-score
yli-identifiointi	overidentification	zetajakauma	zeta distribution
ylikaltaistus	overmatching	z-muunnos	z-transformation
ylikäyllästetty asetelma	super-saturated design	z-testi	z-test
ylimartingaali	supermartingale	Ä	
ylinäytteistys	oversampling ₂	äitiaalloke	mother wavelet
yliparametroitu	overparameterized	äitiväre	mother wavelet
ylipeitto	overcoverage	älykkyysosamäärä	intelligence quotient
ylipoiminta	oversampling ₁	äänestysteoria	voting theory
ylipoissonvaihtelu	extra-Poisson variation	äärellinen jakaumasekoite	finite mixture distribution
ylipäästösuodatin	high-pass filter	äärellisen perusjoukon korjauskerroin	finite population correction
ylipäästösuodin	high-pass filter	äärellisyyskerroin	finite population correction
ylisovitus	overfitting	äärettömästi jakautuva	infinitely divisible
ylitehokas	superefficient	äärettömästi osittuva	infinitely decomposable
ylitysarvokäyrä	exceedance curve	ääriarvo	extreme value
		ääriarvointensiteetti	extremal intensity
		ääriarvojakauma	extreme value

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 oskussönastik* (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 Königlichen 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 Kokemäen virrassa vuosina 1801–1849, toden-
vaihe-laskulla määräsi Johan Henrik Eklöf. *Suomi, Tidskrift i fosterländska
ämnen*, 9, 189–195. <https://digi.kansalliskirjasto.fi/aikakausi/binding/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 toimi-
tuksia. 12 Osa, Toinen wihko. Helsingissä, Suomalaisen Kirjallisuuden Seuran
Kirjapainossa.
- Elfving F. (1938) Suomen Tiedeseuran historia 1838–1938. Suomentanut Aar-
ni Penttilä. Teoksessa *Suomen Tiedeseura 1838–1938*. Suomen Tiedeseura.
Keskuskirjapaino. Helsinki.
- Elfving E. G. (1946) Todennäköisyyslaskenta. Helsingin yliopiston monistustoi-
misto. Helsinki.
- Elfving E. G. (1956) *Todennäköisyyslaskenta*. Tiedekirjasto N:o 28. Otava. Hel-
sinki.
- 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 Commu-
nication 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 Mis-
leading 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-osakeyh-
tiö. Otava. Helsinki.
- Gylling E., Forsman J. (1917) Tilastotiede. Teoksessa *Tietosanakirja, IX*. Tieto-
sanakirja-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).
- Huimo K. (2005) *Perkeleen kieli: suomen kieli ja poliittisesti korrekti 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 määritykset 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 Törnqvist – 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 kannanmunahintojen 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.



© Suomen Tilastoseura ry

3. laitos

ISSN 2954-2286 (verkkojulkaisu)

ISBN 978-952-9795-15-4 (pdf)

Helsinki 2025