

# Luku 3 - Tilastotiede tieteenalana

## Tiivistelmä

Roope Rihtamo

# Luvun ydinviesti

Tilastotiede on useille korkeakouluopinnot aloittavalle varsin vieras tieteenala. Tämä johtuu siitä että tilastotiedettä ei juuri opeteta lukioissa tai ammattikouluissa. Ei ainakaan sellaisena alana, joksi se korkeakouluissa mielletään!

Tässä luvussa tarkastellaankin siis **mitä tilastotiede oikeastaan on?**

Erityisesti keskitymme erottelemaan tilastotieteen sen lähitieteistä, matematiikasta ja tietojenkäsittelytieteestä. Syvempi tarkastelu mahdollistaisi myös tilastotieteen ja tieteenfilosofian välisen suhteen tarkastelun, mutta jätetään se tällä kertaa väliin!

# Keskeiset termit

## **Populaatio**

Konkreettinen tai hypoteettinen tutkimuskohteiden joukko, joka koostuu kaikista tilastoyksiköistä.

## **Tilastoyksikkö ja tilastollinen muuttuja**

Populaation muodostavilta tilastoyksiköiltä tarkastellaan tilastollisia muuttujia, joita voidaan mitata tai havaita.

## **Havainto ja havaintoaineisto**

Havainto muodostuu tilastoyksikön tarkasteltavien tilastollisten muuttujien havaitusta arvoista ja havaintoaineisto, data, on näiden havaintojen joukko.

# Tilastotieteen karakterisointeja

Luotan tässä määritelmässä itseäni pätevämpiin.

Leo Törnqvist, Suomen ensimmäinen tilastotieteen professori:

*“Tilastotiede on tietotuotannon teknologiaa, jonka avulla voidaan suorittaa kvantitatiivisten tietojen joukkotuotantoa ja havaintoihin perustuvia tieteellisiä ja käytännöllisiä päätöksiä.”*

Ilkka Mellin (2004): *“Tilastotiede on yleinen menetelmätiede, jota sovelletaan, jos reaalimaailman ilmiöstä halutaan tehdä johtopäätöksiä ilmiötä kuvaavien kvantitatiivisten tai numeeristen tietojen perusteella sellaisissa tilanteissa, joissa tietoihin liittyy epävarmuutta tai satunnaisuutta.”*

Mark Twain: “Vale, emävale, tilasto”, teoksessaan *Chapters from My Autobiography* vuonna 1907, mutta koska valtaosa “modernin” tilastotieteen teoriakehityksestä on tapahtunut vasta Twainin teoksen julkaisun jälkeen, niin tällä lentävällä lausahduksella ei ole mitään tekemistä nykyisten tilastollisten menetelmien kanssa!

# Mitä tilastotiede ei ole

Toisin kuin usein luullaan, tilastotiede ei ole *ainoastaan* tilastojen, eli **numeeristen tietojen järjestelmällisten kokoelmien**, tuotantoa ja niiden harrastamista.

Tilastotiede sijoittuu tieteiden kentässä matematiikan, filosofian ja tietojenkäsittelytieteen rinnalle. Tästä huolimatta se ei kuitenkaan ole yksiselitteisesti minkään näiden osa-alue.

Tilastotiede ei ole matematiikkaa, sillä se lähestyy tieteellistä ongelmanratkaisua eri tavoin: tilastotiede on aina konteksti- ja aineistopohjaista ja perustuu induktiiviseen päättelyyn.

Tilastotiede ei ole tietojenkäsittelytieteen osa-alue, koska niillä on omat erilliset ja ehjät teoriapohjansa joista jälkimmäinen ei perustu ajatukselle satunnaisista reaalimaailman ilmiöistä.

# Mitä tilastotiede (ainakin) on

Tieteellistä tietoa hankitaan tieteellisellä menetelmällä (ks. luku 2), jonka avulla tutkitaan mielenkiinnon kohteena olevaa ilmiötä tai sen generoimaa kvantitatiivista mutta epävarmaa tietoa sisältävää aineistoa.

Tilastotieteessä kehitetään menetelmiä, jotka antavat eri alojen tutkijoille yhtenevät ja yleisesti hyväksytyt raamit, jotka mahdollistavat (keinot) päättelyn epävarmuuden vallitessa ja keinot tämän epävarmuuden mittaamiseen.

► *Tilastotiede kehittää matemaattisia malleja satunnaisilmiöitä kuvaavia kvantitatiivisia tietoja generoiville prosesseille!*

Tilastotieteen voidaan katsoa kuuluvan ns. menetelmätieteisiin, joissa kehitetään menetelmiä, mutta jolla on myös oma sovelluksista vapaa teorianmuodostuksensa.

# Tilastotieteen osa-alueet

Tilastotiede kehittyi pitkään ns. “ongelmasta menetelmään”, kun yhteiskunnan eri aloilla tarvittiin keinoja tilastojen analysointiin. Tämä johti siihen, että tilastotiede on jakautunut moniin osa-alueisiin. Seuraava karkea kahtiajako voidaan kuitenkin tehdä

## **Soveltava tilastotiede**

on nimensä mukaisesti teoreettisen tilastotieteen kehittämien menetelmien soveltamista jonkin tutkimusalan empiiriseen ongelmaan.

**Teoreettinen tilastotiede** kehittää (tilasto)matemaattisia malleja kuvaamaan satunnaisilmiöitä, jotka generoivat reaallimaailman ilmiöitä kuvaavia numeerisia tietoja, joihin liittyy epävarmuutta ja satunnaisuutta.

# Tilastotieteen kritiikkiä

Tilastotiedettä, kuten mitä tahansa alaa, on myös kritisoitu vuosien saatossa eri tavoin. *Usein kritiikki on aiheetonta ja perustuu väärinymmärryksiin.*

- ▶ **“Yleismaailmallinen kritiikki”** tiivistyy seuraavaan *“Jos on toinen jalka jäässä ja toinen kiehuva vedessä, niin tilastotieteilijän mielestä ihmisellä on tällöin keskimäärin sopivan lämmen”*.
- ▶ **Kritiikki matemaattisuutta kohtaan:** tilastotieteen matemaattinen esitystapa on paikoin vaikeasti lähestyttävää.
- ▶ **Kritiikki yksnikertaistuksia kohtaan:** tilastotiedettä on kritisoitu siitä että se ei kykene riittävällä tasolla huomioimaan reaali maailman kompleksisuutta.
- ▶ **Temppukokoelmakritiikki:** tilastotiedettä pidetään paikoin vain kokoelmana erilaisia menetelmiä tai “temppuja”.



# Tilastotieteen sovellusaloja ja “rajatieteitä”

Yleisenä menetelmätieteenä tilastotiedettä sovelletaan useilla eri tieteenaloilla, joilla on oma erillinen teoriapohjansa sekä empiiriset käytänteet ja täten substanssitietous on sovellettaessa tärkeää!

Jokaisella tieteenalalla, jonka tutkimusaineistot voidaan esittää numeerisessa tai kvantitatiivisessa muodossa voi soveltaa/voisi soveltaa/pitäisi soveltaa tilastollisia menetelmiä sekä tutkimusaineistoja kerättyä että niitä analysoitaessa.

Muutamia esimerkkejä aloista, joilla tilastotieteen soveltaminen on muodostunut omaksi tutkimuskohteekseen: psykologia (psykometriikka), bio- ja lääketiede (biometria), taloustiede (ekonometria) ja kemia (kemometria).