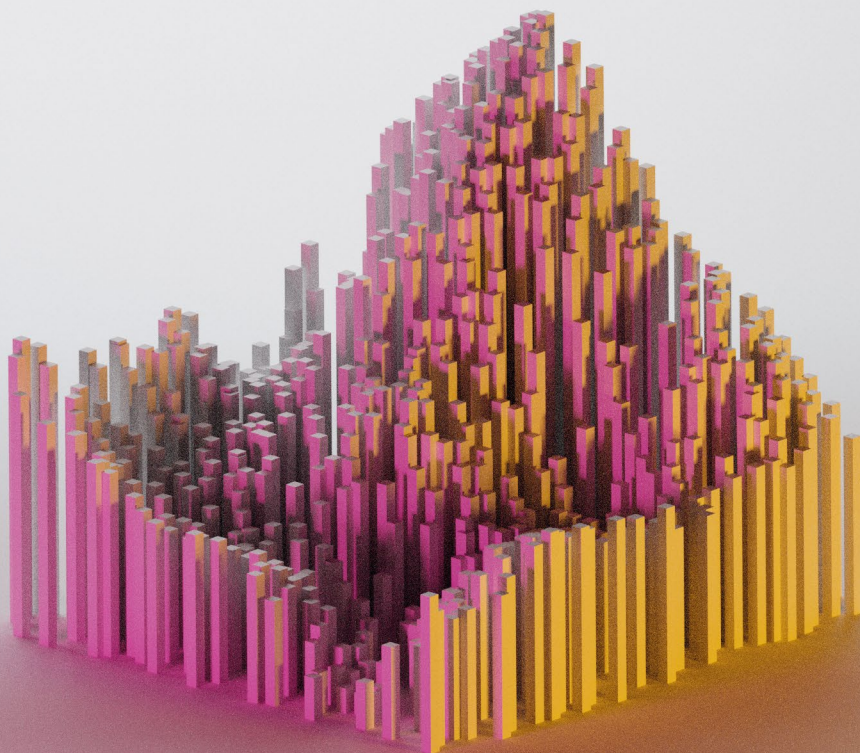


MARIANNE SILÉN

Sosiologian ja tilastotieteen rajoilla

Faktorianalyysi rajakohteena



MARIANNE SILÉN

Sosiologian ja tilastotieteen rajoilla.
Faktorianalyysi rajakohteena

Akateeminen väitöskirja,
joka Lapin yliopiston yhteiskuntatieteiden tiedekunnan
suostumuksella esitetään julkisesti tarkastettavaksi
Lapin yliopiston luentosalissa 3 (LS3)
helmikuun 26. päivänä 2021 klo 12.



LAPIN YLIOPISTO
UNIVERSITY OF LAPLAND

Rovaniemi 2021

Lapin yliopisto
Yhteiskuntatieteiden tiedekunta

Väitöskirjan ohjaajat:

Professori Suvi Ronkainen, Lapin yliopisto
YTT Jukka Mäkelä, Lapin yliopisto
Professori emeritus Seppo Pöntinen, Turun yliopisto

Väitöskirjan esitarkastajat:

Vanhempi yliopistonlehtori Pertti Jokivuori, Jyväskylän yliopisto
Professori emeritus Harri Melin, Tampereen yliopisto

Vastaväittäjä:

Vanhempi yliopistonlehtori Pertti Jokivuori, Jyväskylän yliopisto

Taitto: Taittotalo PrintOne

Kannet: Viivi Hyvärinen

Acta electronica Universitatis Lapponiensis 301

ISBN 978-952-337-248-1

ISSN 1796-6310

elektronisen väitöskirjan pysyvä osoite:
<http://urn.fi/URN:ISBN:978-952-337-248-1>

Tiivistelmä

Tutkimus on metodologinen, tieteiden välisyyttä tarkasteleva tutkimus. Tarkastelun kohteena ovat tilastotiede ja sosiologia sekä niiden kohtaaminen kvantitatiivisessa sosiologisessa tutkimuksessa. Tutkimuksessa tarkastellaan ja pohditaan tilastotieteen mahdollisuuksia ja rajoja sosiologisessa tutkimuskäytössä. Tilastotieteen ja sosiologian kohtaamista ja siihen liittyviä jännitteitä havainnollistetaan tutkimusprosessin eri vaiheiden kautta.

Tutkimuksen perusaineistona ja empiirisen tarkastelun kohteena ovat Suomessa vuosina 1960–2018 tarkastettua 532 sosiologian väitöskirjaa. Aineiston kuvauksessa esitellään aluksi tilastollisten menetelmien vaihteleva suosio sosiologien väitöstutkimuksissa, jonka jälkeen aineiston kuvaus kohdentuu 123:een monimuuttujamenetelmiä sisältävään väitöskirjaan.

Suosituin monimuuttujamenetelmä sosiologian väitöskirjoissa on faktorianalyysi, jota tutkimuksessa analysoidaan rajakohdemallia hyödyntäen. Rajakohteena on täten faktorianalyysin valinta ja käyttötapa sosiologisessa tutkimuksessa. Mallin kauttakulkupisteinä ovat tilastomatematiikkaan ja muuttujien ominaisuuksiin perustuva kauttakulkupiste, faktorianalyysin asetelmaan perustuva kauttakulkupiste sekä sosiologiseen tutkimuskysymykseen perustuva kauttakulkupiste. Rajakohdetta käyttävinä osapuolina ovat tilastotiedettä edustavat tahot, jotka voivat olla tilastotieteilijöitä tai esimerkiksi menetelmäoppaita sekä toisena tahona tutkimusta tekevät sosiologit. Rajakohdemallissa kutsutaan kääntämiseksi vaihetta, jossa osapuolet valitsevat vakuuttuvavimmaksi kokevansa kauttakulkupisteen. Kääntäminen voidaan nähdä kamppailuna tieteellisestä auktoriteetista.

Sosiologit ovat perustelleet faktorianalyysin valintaa väitöskirjoissaan monin eri tavoin. Jotkut perustelut ovat liittyneet tilastomatematiikkaan näkökulmiin, esimerkiksi muuttujien välisen riippuvuuden analysointitarpeeseen. Osa perusteluista nivoutuu asetelmallisiin tavoitteisiin, kuten muuttujien taustalla olevan rakenteen selvittämiseen. Faktorianalyysia on perusteltu myös sosiologiseen tutkimuskysymykseen kytkeytyen. Tällöin perusteluissa on esitetty muun muassa tarve löytää aineistosta empiirisiä vastineita teoreettisille käsitteille.

Tutkimuksessa on analysoitu myös sitä, millä tavalla sosiologit ovat faktorianalyysia käyttäneet, toisin sanoen, minkälaisia analyysin suorittamiseen liittyviä teknisiä valintoja he ovat tuoneet esiin. Tulososiossa esitellään eri aikoina esitettyjä muuttujavalintoja sekä faktorointiin ja rotatointiin liittyviä valintoja.

Tarkastelu paljastaa faktorianalyysin taipuneen tilastomatematiikkaan lisäksi myös sosiologista ajattelua palvelevaksi tutkimusmenetelmäksi. Tulos ei ole yllättävä,

koska faktorianalyysi on alun perinkin kehitetty erilaiseen ihmistieteelliseen, ihmisten toimintaan kohdistuvan tutkimuksen - kuten psykologian - tutkimustarpeisiin. Faktorianalyysin perustajaksi mainittu Louis Thurstonen mukaan tilastotieteilijät eivät aluksi edes hyväksyneet faktorianalyysia tilastotieteelliseksi menetelmäksi.

Lopuksi tutkimuksessa pohditaan mahdollisuutta laajentaa rajakohdetarkastelu kaikkiin monimuuttujamenetelmiin. Monimuuttujamenetelmien ja yleisesti tilastollisten menetelmien käyttö sosiologisissa tutkimuksissa on herättänyt keskustelua sosiologien keskuudessa.

Tutkimuksen tekeminen on parhaimmillaan yhteistyötä, jossa onnistuneen lopputuloksen saavuttamiseksi on hyvä löytää yhteinen ymmärrys. Toivon tämän tutkimuksen avaavan sekä tilastotieteilijöille että sosiologeille erilaisia näkökulmia tilastollisten menetelmien käyttöön ja siten edesauttavan yhteisen ymmärryksen saavuttamisessa.

Avainsanat: faktorianalyysi, monimuuttujamenetelmät, sosiologia, tilastotiede, rajakohdemalli, tieteentutkimus, tieteiden välisyys, tutkimusprosessi

Abstract

This interdisciplinary study focuses on the limits and possibilities of using statistical methods in sociological research. The combining of statistics and sociology and the related tension caused by their differing baselines are demonstrated through the various phases of the sociological quantitative research process.

The basic data of the empirical study consist of 532 Finnish sociological dissertations written between 1960 and 2018, according to which the use of statistical methods has varied over time. The study examines further 123 dissertations that involved multivariate methods, of which factor analysis is the most popular one in sociological dissertations.

The theoretical frame of reference of the study is *boundary object theory*, where the boundary object consists of sociologists' arguments concerning their choice and execution of factor analysis. There are three passage points. The first one emphasizes the principles of mathematical statistics, the second one the structure of the factor model, and the third one the sociological point of view and the aim of sociological research. Translation, that is, the passage point a sociologist has chosen, can be seen as competition for scientific authority.

Sociologists' arguments for the use of factor analysis are heterogenous, including statistical, structural, and sociological points of view. The method has also been used in a variety of ways. The study shows that factor analysis is not only a statistical multivariate method, but also a useful tool for social scientists who use quantitative data to find answers to conceptual research questions.

To conclude the study, the possibility of expanding the use of boundary object theory to all multivariate methods is discussed.

Key words: factor analysis, multivariate methods, sociology, statistics, boundary object theory, science studies, interdisciplinarity, research process

Sisällys

| | |
|--|----|
| Tiivistelmä | 3 |
| Abstract | 5 |
| Kiitokset | 11 |
| 1. Johdanto | 13 |
| 2. Tutkimuksen taustaksi: tieteiden eriytymisen historiaa | 20 |
| 2.1. Tieteiden yleinen kehittyminen..... | 20 |
| 2.2. Tilastotieteen moninaisuus..... | 23 |
| 2.3. Sosiologian panos tilastotieteen kehittymiseen..... | 26 |
| 3. Tutkimuksen kohteena tilastolliset menetelmät, erityisesti faktorianalyysi suomalaisessa sosiologiassa | 29 |
| 3.1. Aineistona sosiologian väitöskirjat 1960–2018..... | 29 |
| 3.2. Tutkimuskysymykset ja analysointitavat..... | 31 |
| 4. Sosiologian ja tilastotieteen kohtaamisen haasteet kvantitatiivisessa tutkimuksessa | 37 |
| 4.1. Tutkimusprosessi nähdään eri tavalla..... | 37 |
| 4.2. Aineistoon kohdistuvat haasteet..... | 40 |
| 4.3. Analyysivaiheen valinnat..... | 47 |
| 4.4. Tulosten tilastotieteellinen ja sosiologinen tulkinta..... | 58 |
| 5. Tilastollisten menetelmien vaihteleva suosio sosiologiassa vuosina 1960–2018 | 62 |
| 5.1. Määrällisestä laadulliseen..... | 62 |
| 5.2. Sosiologian väitöskirjoissa käytetyt monimuuttujamenetelmät..... | 70 |
| 6. Rajakohdemalli ja tieteiden välinen yhteistyö, esimerkkinä faktorianalyysi | 76 |
| 6.1. Rajakohdemalli..... | 76 |
| 6.2. Faktorianalyysin tilastomatematiikka..... | 78 |
| 6.3. Asetelmaan perustuva kauttakulkupiste..... | 82 |
| 6.4. Sosiologiseen tutkimuskysymykseen perustuva kauttakulkupiste..... | 84 |
| 7. Sosiologian väitöskirjoissa esitetyt faktorianalyysiperustelut | 88 |
| 7.1. Perustelujen määrällisyys ja laadullisuus..... | 88 |
| 7.2. Tilastotieteeseen, asetelmaan ja tutkimuskysymykseen nojaavat perustelut..... | 93 |
| 7.3. Faktorianalyysi määrällisenä ja laadullisena menetelmänä..... | 97 |

| | |
|--|-----|
| 8. Faktorianalyysin tekniset valinnat väitöskirjoissa | 103 |
| 8.1. Muuttujien valinta – tilastotieteen puhtasoppisuus ja sosiologian tulkinta..... | 104 |
| 8.2. Faktorointi, rotatointi ja jatkoanalyysi..... | 109 |
| 8.3. Faktorianalyysin hyöty ja haasteet..... | 125 |
| 9. Monimuuttujamenetelmien monet kasvot | 129 |
| 9.1. Rajakohdemallin laajentaminen monimuuttujamenetelmiin..... | 129 |
| 9.2. Erilaiset näkemykset kiistan aiheena..... | 133 |
| 9.3. Löytyykö yhteinen ymmärrys?..... | 135 |
| 9.4. Tulevaisuuden näkymät..... | 137 |
| Lähteet | 139 |
| Liitteet | 153 |

Kuvaluettelo

| | | |
|----------------|--|-----|
| <i>Kuva 1</i> | <i>Tieteiden eriytyminen</i> | 24 |
| <i>Kuva 2</i> | <i>Tilastotieteen osa-alueiden kehittyminen</i> | 25 |
| <i>Kuva 3</i> | <i>Sosiologin ja tilastotieteilijän näkemykset tutkimusprosessiin. Lähde: Silén (2008)</i> | 38 |
| <i>Kuva 4</i> | <i>Sosiologin ja tilastotieteilijän tutkimuskysymykset eroavat toisistaan</i> | 39 |
| <i>Kuva 5</i> | <i>Perusjoukko-otos-asetelma</i> | 43 |
| <i>Kuva 6</i> | <i>Muuttujien välisen riippuvuuden lajit</i> | 48 |
| <i>Kuva 7</i> | <i>Tilastotieteen ja sosiologian erilaiset ilmaisut riippuvuustarkasteluissa</i> | 48 |
| <i>Kuva 8</i> | <i>Selittämiseen perustuvia monimuuttujamenetelmiä</i> | 53 |
| <i>Kuva 9</i> | <i>Muuttujien erilaiset roolit mallituksessa. Lähde Oja (1990)</i> | 54 |
| <i>Kuva 10</i> | <i>Tiivistämiseen perustuvat monimuuttujamenetelmät</i> | 55 |
| <i>Kuva 11</i> | <i>Faktorianalyysin ja ryhmittelyanalyysin ero Hairin ym. (1998, 98) esittämänä</i> | 55 |
| <i>Kuva 12</i> | <i>Ryhmävertailun sisältävät monimuuttujamenetelmät</i> | 56 |
| <i>Kuva 13</i> | <i>Monia asetelmia sisältäviä monimuuttujamenetelmiä</i> | 57 |
| <i>Kuva 14</i> | <i>Lineaaristen rakenneyhtälömallien koostuminen faktori-, regressio- ja varianssianalyyseistä</i> | 57 |
| <i>Kuva 15</i> | <i>Sosiologin ja tilastotieteilijän näkemys tulkintaan</i> | 59 |
| <i>Kuva 16</i> | <i>Sosiologiassa vuosina 1960–2007 tarkastetut väitöskirjamäärät eroteltuna tilastollisia menetelmiä ja monimuuttujamenetelmiä sisältäviin</i> | 70 |
| <i>Kuva 17</i> | <i>Neljän suosituimman monimuuttujamenetelmän esiintyminen sosiologian väitöskirjoissa 1960-2018, yläriivi esittää lukumäärät ja alariivi osuudet</i> | 73 |
| <i>Kuva 18</i> | <i>Callonin, Latourin ja Lawin kehittämä toimijaverkkoteorian malli. Lähde Star ja Griesemer (1989)</i> | 76 |
| <i>Kuva 19</i> | <i>Starin ja Griesemerin muokkaama malli. Lähde Star ja Griesemer (1989)</i> | 77 |
| <i>Kuva 20</i> | <i>Faktorianalyysi rajakohteena</i> | 78 |
| <i>Kuva 21</i> | <i>(Pearsonin) korrelaatiokertoimen tulkinta</i> | 79 |
| <i>Kuva 22</i> | <i>Faktorimalli graafisesti ja yhtälömuodossa</i> | 80 |
| <i>Kuva 23</i> | <i>Faktorimallin asetelma</i> | 82 |
| <i>Kuva 24</i> | <i>Monimuuttujamenetelmien jaottelu latentin muuttujan olemassaolon mukaan. Lähde: Tacq (1997)</i> | 84 |
| <i>Kuva 25</i> | <i>Vasemmalla operationalisointi eli mittauskohteen spesifointi teoreettisesta konstruktiosta mittareihin. Oikealla faktorianalyysi graafisesti kuvattuna. (Lähde: Nummenmaa ym. 1997, 213; 265)</i> | 86 |
| <i>Kuva 27</i> | <i>Faktorianalyyseiperustelujen tiivistettyjen ilmausten esiintymismäärät</i> | 90 |
| <i>Kuva 28</i> | <i>ATLAS-ti-analyysin tulos faktorianalyyseiperusteluille</i> | 91 |
| <i>Kuva 29</i> | <i>Faktorianalyyseiperustelut rajakohdetarkasteluna</i> | 93 |
| <i>Kuva 30</i> | <i>Faktorianalyysin perustelujen kauttakulkupisteet</i> | 96 |
| <i>Kuva 31</i> | <i>Faktorianalyysin määrälliset ja laadulliset tavoitteet</i> | 98 |
| <i>Kuva 32</i> | <i>Faktorianalyyseille eri aikoina esitetyt tavoitteet subteutettuna faktorianalyysia sisältävien väitöskirjojen määrään</i> | 101 |
| <i>Kuva 33</i> | <i>Faktorianalyysin vaiheet</i> | 103 |
| <i>Kuva 34</i> | <i>Faktorianalyyseissä käytettyjen aineiston koot vuosina 1960–2007</i> | 108 |

| | |
|--|-----|
| <i>Kuva 35 Vasemmalla yhden faktorin malli ja oikealla yhden pääkomponentin malli graafisina esityksinä ja matemaattisina yhtälöinä. (Lähteet: Leskinen 1987; Nummenmaa ym. 1997, 230; 242.)</i> | 114 |
| <i>Kuva 36 Faktorimatriisin varianssien tyypit (Lähde: Hair ym. 1998, 102)</i> | 115 |
| <i>Kuva 37 Faktorianalyysi (vasemmalla) ja pääkomponenttianalyysi (oikealla) tiivistävinä menetelminä</i> | 115 |
| <i>Kuva 38 Rotatoinnilla pyritään yksinkertaiseen latausrakenteeseen</i> | 116 |
| <i>Kuva 39 Suora ja vino rotaatio koordinaatistossa ja faktorimallissa</i> | 118 |
| <i>Kuva 40 Muuttujien korrelaatio vektorilaskentana</i> | 119 |
| <i>Kuva 41 Faktorianalyysin jatkotoimenpiteet eri aikoina rajakohdemallin kauttakulkupisteiden mukaan</i> | 124 |
| <i>Kuva 42 Monimuuttujamenetelmien näyttäytyminen rajakohdemallin kauttakulkupisteinä</i> | 130 |
| <i>Kuva 43 Lineaarisen ja logistisen regressioanalyysin asetelmallinen yhtäläisyys ja matemaattinen eroavaisuus</i> | 132 |

Taulukkoluetelo

| | |
|--|-----|
| <i>Taulukko 1 Tilastollisten menetelmien esiintyminen sosiologian väitöskirjoissa</i> | 65 |
| <i>Taulukko 2 Tilastollisia menetelmiä sisältävien väitöskirjojen esiintyminen yliopistoittain</i> | 69 |
| <i>Taulukko 3 Sosiologian väitöskirjoissa käytetyt monimuuttujamenetelmät ja niiden käytön ajankohdat</i> | 72 |
| <i>Taulukko 4 Väitöskirjoissa esitetyt menetelmät ongelmanasettelunsa mukaan</i> | 74 |
| <i>Taulukko 5 Neljän y-muuttujan kovarianssimatriisin alakolmioesitys</i> | 79 |
| <i>Taulukko 6 Faktorimallin tilastomatemaattiset oletukset</i> | 81 |
| <i>Taulukko 7 y-muuttujien teoreettinen kovarianssimatriisi kolmella tavalla esitettynä</i> | 81 |
| <i>Taulukko 8 Sosiologiien faktorianalyyssiperusteluja ja niiden tiivistetyt ilmaukset</i> | 88 |
| <i>Taulukko 9 Faktorianalyyssin tavoitteet eri aikoina</i> | 101 |
| <i>Taulukko 10 Faktorianalyyssin suorittamisessa käytetyt tilastolliset ohjelmat eri aikoina</i> | 111 |
| <i>Taulukko 11 Faktorointimenetelmät eri aikoina</i> | 111 |
| <i>Taulukko 12 Sosiologiien käyttämät rotatointimenetelmät</i> | 118 |
| <i>Taulukko 13 Suorien ja vinojen rotaatioiden esiintyminen sosiologian väitöskirjoissa</i> | 119 |
| <i>Taulukko 14 Faktoripistemäärä- ja summamuuttujien esiintyminen faktorianalyyssin jatkotoimenpiteenä</i> | 123 |
| <i>Taulukko 15 Faktoripistemäärämuuttujan ja summamuuttujan ominaisuudet Jokivuoren ja Hietalan (2007) mukaan</i> | 124 |
| <i>Taulukko 16 Faktorointi- ja rotaatiomenetelmien sekä faktorianalyyssin jatkotoimenpiteiden esiintymismäärät eri aikoina</i> | 125 |
| <i>Taulukko 17 Faktorianalyyssin rooli väitöskirjoissa eri aikoina</i> | 126 |

Kiitokset

Tutkimuksen tekeminen on ollut minulle oppimisprosessi. Olen oppinut, mitä tutkimuksen tekeminen on tilastotieteen ulkopuolella, minkälaista se on ollut eri aikoina ja mitä keskusteluja ja pohdintoja on liittynyt tilastollisten menetelmien käyttöön sosiologiassa tai sosiaalitieteissä. Väitöskirja määrittellään akateemiseksi opinnäytetyöksi. Oma väitöstutkimukseni on aivan konkreettisesti opinnäytetyö, koska se rakentuu asioista, joita olen tässä prosessissa oppinut. Olen ollut tilastotieteen parissa vuodesta 1985 lähtien. Ensimmäiset yhdeksän vuotta opiskelijana ja sen jälkeen opettajana. Tätä tutkimusta olen tehnyt vuosien ajan. Tutkimus on edennyt kirjoittamisen ja lukemisen kannalta pätkittäisesti, mutta ajatustyönä tutkimuksen tekeminen on ollut katkeamaton prosessi. Ensimmäiset kipinät aiheeseen lähtivät kytemään jo vuonna 1994, kun minua pyydettiin tilastotieteen opettajaksi Lapin yliopistoon. Muistan ihmetelleeni, että mitä ihmettä yhteiskuntatieteiden ja kasvatustieteiden opiskelijat tekevät tilastotieteellä. Nyt vuonna 2020 siteeraan Pikku G:n sanoja: ”Kaikki kasvaa, lapset varttuu, virheist oppii ja kokemust karttuu”. Kiitos opiskelijoille, tutkijoille, työkavereille ja kaikille tahoille, jotka ovat vuosien varrella pyytäneet minulta monenlaisia neuvoja ja ohjausta, koska he ovat vastavuoroisesti opettaneet minua katsomaan tutkimusprosessia Konsta Pylkkäsen sanoin ”vähän laajemmasta näkökulmasta”.

Suvi Ronkainen organisoii 2000-luvun alkupuolella Lapin yliopistoon useita vierailuvia luennoitsijoita, joiden monipuolisten, erityisesti tutkimusmenetelmiin, liittyvien luentojen kuuntelu sekä keskustelut Suvin kanssa herättivät kiinnostukseni pohtimaan, olisiko minusta tekemään omaa tutkimusta sen ohella, että neuvon muita tutkijoita heidän analyysissään. Pääsy Suvin johtamaan tutkimusprojektiin 2000-luvun alkupuolella toimi lähtölaukauksena omalle tutkimukselleni. Haluan kiittää Suvia siitä, että hän tarjosi minulle mahdollisuuden aloittaa tutkimukseni ja jaksoi sitkeästi kannustaa ja ohjata sen eteenpäin viemisessä sekä maalisuoralle saattamisessa, vaikka monta kertaa ihmettelin, miksi aloin tutkimaan tällaista aihetta eikä tutkimusmatkakaan osoittautunut maratoniksi vaan suorastaan ultramatkaksi. Väitöstutkimukseni toisena ohjaajana on toiminut Jukka Mäkelä. Kiitän Jukkaa sekä ohjauksesta että siitä, että hän on omalla esimerkillään antanut askelmerkkejä, jotka ovat auttaneet minua sekä tilastotieteen opettajana että sosiologista tutkimusta tekemänä tutkijana. Aloitin opettajaurani Jukan sijaisena tilastotieteen lehtorina hänen kirjoittaessaan omaa sosiologian alaan kuuluvaa väitöskirjaansa. Vuodet 1998–2009 Jukan johtamalla menetelmätieteiden laitoksella tarjosivat aitiopaikan monitieteisyyteen. Kolmantena ohjaajana on toiminut Seppo Pöntinen. Kiitän

Seppoa asiantuntevista sosiologiaa ja sosiologian tutkimusmenetelmiä koskevista kommenteista ja neuvoista sekä kannustavasta ohjauksesta.

Kiitän esitarkastajia Pertti Jokivuorta ja Harri Meliniä rohkaisevista lausunnoista. Väitöstutkimukseni sijoittuminen tieteiden väliin on vaatinut heiltä monitieteistä näkökulmaa. Kiitän myös ensimmäisen vaiheen esitarkastajia Semi Purhosta ja Reijo Sundia huomioista, jotka auttoivat käsikirjoituksen saattamisessa lopulliseen muotoonsa.

Syksyllä 1994 Tarja Orjasniemi käveli minua yliopiston käytävällä vastaan ja esitti kysymyksen faktorianalyysista. En osannut vastata hänen kysymykseensä, mutta päätin ottaa asiasta selvää. Siitä käynnistyi vuosia kestävä tutkimusyhteistyömme alko-aineistoa analysoiden. Tutkimuksellinen yhteistyö oli korvaamattoman arvokasta ja opettavaista, mutta sitäkin arvokkaampaa on ollut ystävyys. Kiitos Tarja kaikesta tuesta, kannustuksesta ja hauskuuksista, mitä olet vuosien varrella tarjonnut.

Marja Alastalo auttoi minua suuresti erityisesti tutkimukseni alkutaipaleella antamalla muun muassa tiedot sosiologian väitöskirjoista ja vihjaamalla rajakohde-teorian hyödyntämisestä. Pekka Vasari on aina jaksanut tarjota tukea ja apua tilastotieteeseen liittyvissä kysymyksissä sekä tarjoutunut hoitamaan puolestani opetukseen liittyviä tehtäviä, jotta saisin aikaa väitöskirjani kirjoittamiseen. Yhteistyö ja keskustelut Ritva Linnakankaan kanssa ovat olleet todella opettavaisia ja auttaneet minua ratkaisevalla tavalla etenemään kirjoitusprosessissani.

Pääsääntöisesti olen tehnyt tutkimustani opetuksen ohessa. Kaksi jaksoa, jolloin olen voinut irtautua opetuksesta, ovat mahdollistuneet Suomen akatemian rahoituksella tutkimukseni suunnitteluvaiheessa 2000-luvun alkupuolella sekä rehtorin apurahalla keväällä 2018.

Tutkimuksen tekeminen on yhteistyötä, jossa jokainen oppii. Tästä esimerkkinä mainitsen yhteistyön Kirsi Pollarin kanssa. Samalla, kun toimin hänen väitöstutkimuksensa menetelmäohjaajana, auttoi hän kommenteillaan omaa tutkimustani eteenpäin. Lapin yliopiston sosiologit ovat vuosien varrella kannustaneet ja kommentoineet käsikirjoituksen vaiheita. Haluan kiittää heidän lisäksi koko yhteiskuntatieteiden tiedekunnan henkilökuntaa sekä myös menetelmätieteiden laitoksen aikaisia kollegoja keskusteluista, ideoista ja opeista. Hanna Peltomaan ja Katriina Uljas-Raution kanssa jakamamme huumori sekä Sari Mantilan ja Pilvikki Lantelan kanssa jaetut Ladyline-hehkutukset ovat mukavasti piristäneet mieltä väitöstuskailujen keskellä.

Miki ja Juha-Mikko eivät varmaankaan edes muista aikaa, jolloin en olisi puhunut tekeillä olevasta väitöskirjasta. Poijjaat - ja sukulaiset - tässä tämä nyt on.

Rovaniemellä 8.12.2020

Marianne Silén

1. Johdanto

Tieteet ovat kehityksensä aikana eriytyneet ja niiden osaamisalueet ovat syventyneet, mutta samalla kaventuneet. Eriytyminen on synnyttänyt tarpeen tieteiden väliselle yhteistyölle. Tieteentutkimukseen luettavassa tutkimuksessani olen tarkastellut sosiologian ja tilastotieteen välistä yhteistyötä. Tutkimukseni on aineistonsa kautta konkreettinen kuvaus tieteiden välisyydestä sekä empiirisen tutkimuksen tekemisen edellyttämästä tieteenalojen kohtaamisesta. Tutkimuksessani tarkastelen tilastollisten menetelmien käyttöä sosiologian väitöskirjoissa. Erityisesti katson faktorianalyysia tieteellisen tiedon tuottajana sosiologisessa tutkimuksessa. Kiinnostukseni tähän aiheeseen juontaa juurensa matemaattiseen tilastotieteeseen painottuvaan koulutustaustaani ja yhteiskuntatieteisiin painottuvaan työhistoriaani. Olen toiminut tilastotieteen opettajana Lapin yliopistossa vuodesta 1994 lähtien opettaen pääasiassa yhteiskuntatieteiden ja kasvatustieteiden opiskelijoita. Tutkimuksen sekä siihen liittyvän opetuksen kautta olen itse päässyt ja joutunut kohtaamaan sekä tilastollisen laskentatavan ajattelulogiikan, että sosiologisen tilastolliseen aineistoon perustuvan päättelylogiikan erilaisuuden. Tämä näkyy myös väitöskirjassani. Kiinnostukseni hahmottaa sitä, miten sosiologinen kvantitatiivinen tutkimusajattelu ja tilastotieteen tilastollisen päättelyn logiikka kohtaavat, on syntynyt opetuskokemuksen kautta. Tutkimukseni taustalla on myös Suomen Akatemian rahoittama tutkimusprojekti *Kvantitatiivisuus, tulkinnallisuus ja faktojen politiikka* (2003–2006). Sen yksi tavoite oli selvittää tilastollisiin menetelmiin liitettäviä merkityksiä osana kvantitatiivisen tutkimuksen paradigmaa ja tarkastella, liittyykö menetelmiin samoja merkityksiä perinteisen tilastotieteen näkökulmasta katsottuna.

Tilastollisen tietojenkäsittelyn ajattelutapa ja yhteiskuntatieteellisen kvantitatiivisen tutkimuksen edellyttämä tilastollinen analyysi kohtaavat ajattelussani. Tutkimukseni yleiseksi aiheeksi asettuu tilastotieteellisen päättelylogiikan ja yhteiskuntatieteellisen, sosiologisen tutkimuksen kohtaaminen. Kysyn, millaisella tavalla tilastotiede ja sosiologia kohtaavat kvantitatiivisessa tutkimuksessa. Pyrin konkreettisten esimerkkien kautta avaamaan tutkimusprosessin eri vaiheisiin liittyviä erilaisia menetelmällisiä ajattelutapoja. Keskityn erityisesti analysoimaan faktorianalyysia sosiologisena tutkimusmenetelmänä, mutta pyrin esittämään myös kuvailevaa tietoa siitä, kuinka paljon ja mitä tilastollisia menetelmiä sosiologit ovat käyttäneet eri aikoina.

Tutkimukseni pohja-aineistona ovat suomalaiset sosiologian väitöskirjat vuosilta 1960–2018. Olen rajannut tarkempaan tarkasteluun väitöskirjat, joissa käytetään määrällisiä aineistoja ja tilastollisia analyysimenetelmiä sekä vielä tarkempaan tar-

kasteluun väitöskirjat, joissa käytetään faktorianalyysia. Olen muodostanut väitöskirjoihin pohjautuvia tekstiaineistoja ja numeerisia aineistoja.

Tutkimuksen kohdistuessa lähes kuuden vuosikymmenen ajalle on sekä sosiologiassa, että tilastollisten menetelmien käyttömahdollisuuksissa tapahtunut monenlaisia muutoksia. 1960-luvulla – sosiologian ja kvantitatiivisen tutkimuksen kulta-ajalla – ajateltiin, että yhteiskunnan kehitystä voidaan ohjata tieteen keinoin ja uskottiin, että sosiologian avulla saadaan selville tosiasioita yhteiskunnasta ja sen säännönmukaisuudesta. (Eskola 1997; Alastalo 2005, 73; Puuronen 2005, 133; Stolte-Heiskanen 1992, 295; Toivonen 1999, 119). Yhteiskunnassa vallitsevat ristiriidat toivottiin tutkimuksen avulla saatavan näkyviksi ja sen myötä hallittaviksi, ja sosiologian uskottiin tarjoavan tapoja ymmärtää ja käsitteellistää yhteiskuntaa (Puuronen 2005, 52). Sosiologit olivat kiinnostuneita mielipiteiden, asenteiden ja arvostusten selvittämisestä, toteaa Pertti Rautio (1985, 159–160) ja jatkaa, että kiinnostus saattoi osaksi johtua tietokoneiden suomista uusista mahdollisuuksista ja tutkimusmenetelmistä, mutta sellaiselle tiedolle riitti myös kysyntää. Tilastollisia menetelmiä hallitseva sosiologi sai arvostusta tutkijayhteisössä, toteaa Rautio seuraavalla kuvauksellaan:

Mitä monimutkaisempia tilastollisia analyysimenetelmiä yhteiskuntatieteilijät käyttivät, sitä helpommin he erottautuivat harmaasta tutkijoiden massasta. Samalla yhteiskuntatieteilijät voivat lopultakin katsoa tekevänsä jotakin samantapaista kuin tutkijoiden keskuudessa perinteisesti arvostetut matemaatikot ja ns. kovien luonnontieteiden edustajat. (Rautio 1985, 160.)

1960-luvun lopulla sosiologia ajautui kriisiin. Tilastollisia menetelmiä hyödyntävää sosiologiaa kritisoitiin voimakkaasti. Osa sosiologeista vaati uudenlaisen yhteiskunta-analyysin tekemistä. Sosiologisessa ajattelussa alkoi vallita vahva marxilainen painotus ja sosiologiaa pidettiin muun muassa kapitalismianalyysina. (Alestalo ja Rätty 1994, 219–224.) Antti Eskola (1994, 34) on tuonut esiin, miten 1960-luvun lopulla tapahtuneen marxismin esiinnousun myötä hänenkin metodikonstruktioitaan ”iskettiin suoraan palleaan selittämällä surveytutkimus ihmisiä esineellistäväksi ja teoreettisesti mielivaltaiseksi”. 1970-luvun alussa julkaistiin useita menetelmäkirjoja ja opetusmonisteita, joissa kritisoitiin tilastollisia menetelmiä (Alastalo 2005, 91–94.) Marja Alastalo (2005, 291) toteaa myös, että numeeriset aineistot ja tilastolliset menetelmät olivat 1970-luvun kirjoissa edelleen korostuneesti esillä, mutta niissä tuotiin esille myös huoli kyseisten menetelmien hallitsevuudesta ja niitä käyttävien tutkimusten laadusta.

Kriittinen keskustelu sosiologiasta siirtyi myös yliopistoihin, joissa vaadittiin korkeakoululaitoksen uudistamista sekä sosiologiaa hallinneen positivistisen paradigman vallan murtamista. (Stolte-Heiskanen 1992, 290–292.) Näkemykset sosiologian painotuksista kuitenkin vaihtelivat yliopistoittain ja sosiologia menetti siihen

asti vallinneen sisäisen yhtenäisyytensä. Tämän myllerryksen kourissa yliopistojen sosiologian laitokset olivat koko 1970-luvun ajan. (Alestalo ja Rätty 1994, 219–224.) Sosiologian hajaantuminen jatkui 1980-luvulla sosiologian tutkimusalueiden ja teoreettisten pohdintojen eriytyessä. Suomalaisen sosiologian paradigmoja käsittelevässä artikkelissaan Vesa Puuronen (2005, 55) kuvaa sosiologian tilaa sanoin: ”Sosiologia oli 1980-luvun alussa entistä pahemmin sisäisesti hajalla”.

1980-luvulla suomalaisessa sosiologiassa tapahtui niin kutsuttu kulttuurinen käänne. Tutkimus suuntautui yhteiskunnan rakenteiden tutkimisesta kulttuuriin ja vuorovaikutukseen (Alastalo 2014, 283). Suomalaisen sosiologian teoriat monipuolistuivat ja empiiristen - useimmiten kvantitatiivisen tutkimusten - rinnalla tehtiin voittopuolisesti tai pelkästään teoreettisesti suuntautuneita marxilaisia tutkimuksia ja kvalitatiivisia elämäntapatutkimuksia. Alastalo (2005, 19) kuvaa 1980-luvulla sosiologiassa vallinnutta ilmapiiriä seuraavasti: ”Kertomus lomaketutkimuksen kelvottomuudesta oli vahva, kun aloitin sosiologian opinnot 1980-luvun puolivälissä”. Samassa yhteydessä Alastalo viittaa myös brittisosiologi Catherine Marshin samankaltaisiin kokemuksiin 1970-luvun Britanniassa. Puuronen (2005) kuvaa 1980-luvun suomalaista sosiologiaa sanoin:

Sosiologien suuntautuminen pienten värikkäiden alakulttuurien ja merkitysten tutkimukseen oli vastaus tieteen sisäisen kehityksen nostamiin ongelmiin. Kvantitatiivisten, suurilla aineistoilla operoivien, positivistisiksi kutsuttujen ja pinnallisina pidettyjen tutkimusten sijasta haluttiin tehdä syvälle luotavia pieniin ryhmiin kohdistuvia tutkimuksia. Vanhanaikaiselta kuulostavan kausaalisen selittämisen sijasta haluttiin ymmärtää ihmisten toiminnalleen antamia merkityksiä ja päästä perille toimintaa ohjaavista merkitysjärjestelmistä. Yhteiskuntatutkimuksen tuloksiin perustuva yhteiskunnan kehityksen ennustaminen ja ohjaaminen todettiin jotakuinkin toivottomiksi yrityksiksi, eritoten 1980-luvun loppupuolella, jolloin tieteelliseen suunnitteluun ja ohjaukseen perustuva sosialistinen järjestelmä romahti ja raunioista paljastui huomattavasti suurempia ongelmia kuin oli aikaisemmin ollut tiedossa. 1980-luvun lopulla sosiologian yhteiskunnallinen rooli perustui sen panokseen yliopisto-opiskelijoiden kouluttajana ja yhteiskunnallisen todellisuuden kuvaajana. (Puuronen 2005.)

Puuronen (2005) toteaa, miten 1990-luvullakaan sosiologit eivät olleet kiinnostuneita tilastollisesta analyysistä vaan diskursseista, narratiiveista, retoriikasta ja muista kielellisistä konstruktioista. Alastalon (2014) mukaan monien menetelmien käyttö eli metodologinen pluralismi alkoi vahvistua 1990-luvulla Britanniassa. Suomessa 1990-luvun sosiologiaa leimasi kädenvääntö kvantitatiivisten ja kvalitatiivisten menetelmien paremmuudesta (Lehto 1998, 209; Heiskanen 1998, 201–202; Pöntinen 2004, 82; Alastalo 2005, 17–19). Timo Toivonen (1999, 175) harmittelee

empiiristä sosiaalitutkimusta käsittelevässä kirjassaan, miten tilastolliset menetelmät joutuivat sosiologien keskuudessa epäsuosioon juuri kun tietokoneet olisivat mahdollistaneet niiden täysipainoisen suorittamisen.

2000-luvun sosiologian kiinnostuksen kohteet, tutkimusaiheet ja tutkimuksen tavat ovat hyvin moninaiset. Sosiologian moninaisuus on Jani Erolan ja Pekka Räsänen (2014a, 241) mukaan nähty sekä ongelmana että vahvuutena. Seppo Pöntinen (2004, 91) kirjoittaa sosiologisen tutkimuksen monipuolistumisesta seuraavasti:

Viime aikoina erilaisten aineistojen ja menetelmien yhteiskäyttö on yleistynyt. On huomattu, että tällä tavoin tutkimuskohteesta saadaan monipuolisempi kuva. Toisaalta tämä edellyttää menetelmien parempaa hallintaa. Tutkija ei voi enää rajoittua yhden suosikkimenetelmänsä käyttöön. Usein hänen on myös haettava yhteistyökumppaneita, niin että muodostuu monipuolisesti erilaisia menetelmiä hallitseva projektiryhmä. (Pöntinen 2004, 91.)

Sosiologiassa tapahtuneiden muutosten rinnalla ovat myös tilastollisten menetelmien käytön mahdollisuudet muuttuneet rajusti (ks. esim. Tarkkonen 2007, 12–13). Tutkijat tekivät tilastollista analyysia suurten keskustietokoneiden ehdoilla 1960-luvulta noin 1980-luvun puoliväliin asti. Henkilökohtaiset tietokoneet ja tilastolliset ohjelmat tulivat tutkijoiden käyttöön 1980-luvun loppupuolelta alkaen. Mikrotietokoneet olivat yleistyneet jo aikaisemmin - vuotta 1984 kutsutaan mikrotietokoneiden kotikäytön läpimurtovuodeksi – mutta niiden kotikäyttö oli vielä enimmäkseen viihdekäyttöä Commodore 64:n ollessa suosituin kotimikro (Suominen 2000.) Tilastolliset ohjelmat, kuten BMDP, SAS ja SPSS oli luotu jo 1960-luvun puolivälissä, mutta 1980-luvun puoliväliin asti ja jonkin aikaa vielä sen jälkeenkin ne olivat käytettävissä vain keskustietokoneiden kautta. Henkilökohtaiset tietokoneet muuttuivat helppokäyttöisemmiksi 1980-luvun puolivälissä, kun graafinen Microsoftin Windows-käyttöjärjestelmä tuli merkkipohjaisen DOS-käyttöjärjestelmän tilalle. (Suominen 2000.) Myös tilastollisten ohjelmien valikoima laajeni suurimpien tilasto-ohjelmien ollessa SPSS, SAS, SYSTAT ja BMDP. Suurista tilasto-ohjelmien valmistajista SPSS oli ensimmäinen, joka kehitti ohjelmansa Microsoft Windows-ympäristöön vuonna 1992.

Tietoverkot olivat alkaneet kehittyä vuonna 1984 FUNET:in perustamisen myötä. Suomi liittyi vuodenvaihteessa 1988–1989 FUNET:in ja pohjoismaiden yhteisen NORDUnet-tutkimusverkon kautta Internetiin. (Suominen 2000.) Verkko-yhteyksien kehittyminen helpotti monia tutkimuksen käytäntöjä, tiedonhaku helpottui ja hakumahdollisuudet laajenivat, sähköiset kontaktikanavat mahdollistivat tutkijoiden kansallisen ja kansainvälisen verkostoitumisen. Tietoverkot ovat mahdollistaneet tiedon ja ohjeiden lähes rajattoman saatavuuden. Tutkijoilla on 2000-luvulla käytössään tehokkaat henkilökohtaiset tietokoneet, monia tilastollisia ohjelmia ja hyvä valikoima menetelmäoppaita sekä kirjallisessa ja sähköisessä

muodossa. Juuri siksi kysymys siitä, miten sosiologiset kvantitatiivisten aineistojen tilastolliset analyysitavat ja tilastotieteen matematiikkaan pohjaavat analyysitavat kohtaavat, on kiinnostava.

Tutkimukseni nivoutuu aineistonsa osalta Alastalon vuonna 2005 valmistuneeseen väitöstutkimukseen, jossa hän tarkasteli lomaketutkimuksen käyttöä ja suhdanteita suomalaisessa sosiologiassa vuosina 1947–2000. Oma tutkimukseni kohdentuu osittain samaan aineistoon kuin Alastalon (2005) tutkimus, mutta eroaa tarkastelukohteen osalta. Alastalon tarkastelun keskiössä olivat sosiologien väitöstutkimuksissaan käyttämät aineistot, kun taas omassa tutkimuksessani tarkastelun keskiössä ovat sosiologien väitöstutkimuksissaan käyttämät tilastolliset analyysimenetelmät.

Tutkimuksessani on yhteneväisyyksiä myös Jukka Mäkelän (1991; 1996) tutkimusten kanssa. Lisensiaattitutkimuksessaan Mäkelä (1991) tutki tilastollisen ajattelun siirtymistä osaksi empiiristä sosiaalitutkimusta. Väitöstutkimuksessaan Mäkelä (1996) jatkoi saman aihepiirin parissa päätyen menetelmällisten ajattelutapojen kolmijakoon. Hän nimittää kolmea ajattelutapaa tapaustutkimukseksi, tilastolliseksi tutkimukseksi ja kvantitatiiviseksi tutkimukseksi. Mäkelän näkemyksen mukaan tilastollinen tutkimus oli alkuvaiheessaan lähinnä yleisana ajattelutavalla, jossa tapaustutkimuksen ”epätieteellisyys” halutaan korvata tarkalla mittaamisella ja populaatiota koskevilla hypoteeseilla, joita otosaineistojen perusteella hyväksyttiin tai hylättiin. Tutkija ei saanut etsiä aineistostaan piirteitä, joita hän ei ollut ennakoinut ennen aineiston keräämistä. Tämä siksi, että tutkija ei voisi manipuloida tuloksia ja tutkimus olisi täten objektiivista ja arvovapaata. Kvantitatiivisen tutkimuksen Mäkelä määrittelee lähtökohtaisesti sellaiseksi, jossa tehdään mittauksia, mutta niiden taustalla ei ole mitään, mikä edellyttäisi loogisesti kvantifointia. Kvantitatiivisessa tutkimuksessa on kyse tutkimusprosessista, jossa teoreettisella aineksella on keskeinen merkitys. (Mäkelä 1996.)

Myös Suvi Ronkaisen väitöstutkimuksesta (1999), joka käsittelee kyselytutkimuksen tilastollisen analyysin tapoja, löytyy paljon samankaltaista pohdintaa, jota olen omassa tutkimuksessani käynyt läpi. Tutkimukseni aiheesta löytyy paljon keskustelua ja kirjallisuutta. Niiden näkökulmat vaihtelevat tieteenfilosofisesta, metodologisesta, matemaattisesta, sosiologisesta ja sosiaalitieteellisestä historialliseen ja varsin käytännölliseen otteeseen. Omassa tutkimuksessani nostan esille muutamia näistä näkökulmista. Toivon, että voin myös omalla tutkimuksellani tuoda uuden näkökulman tähän keskusteluun.

Tampereen ja Jyväskylän yliopistojen järjestämällä metodifestivaaleilla vuonna 2011 Seppo Pöntinen esitelmöi otsikolla: ”Mitä mielekäs kvantitatiivisen sosiologian väitöskirjan tekeminen vaatii”. Pöntinen (2011) kiteytti kvantitatiivisen sosiologisen väitöskirjan tärkeimmäksi kulmakiveksi teoreettisen viitekehyksen, aineiston sisältämien muuttujien ja käytettävän menetelmän kolmiyhteisen yhteensopivuuden. Näiden kolmen elementin yhteensopivuuden haasteet kulminoituvat

mittaamiseen. Tutkimuskohteen mittaaminen sosiologisessa ja tilastotieteellisessä mielessä onnistuneesti ei ole helppoa eikä siihen läheskään aina pystytäkään. (Pöntinen 2011; ks. myös Byrne 2001; 2004.)

Jennifer Platt (1976; 1996) on tutkinut sosiaalitutkimuksen käytäntöjä Britanniassa sekä sosiologisten tutkimusmenetelmien historiaa Amerikassa. Hän tuo esille tutkimuksen rahoituksen, aikataulujen, tutkimukseen sidoksissa olevien organisaatioiden ja projektien, tutkijoiden uran ja myös yksityiselämään liittyvien asioiden moninaisuuden. Vaikka omassa tutkimuksessani keskityn paljon kapeampaan alueeseen tutkimusprosessia - tutkimuksessa käytettyihin tilastollisiin menetelmiin - on niihin suhtautumisessa ja niiden käytössä havaittavissa monenlaisia näkemyksiä ja tapoja. Tutkimukseni aineiston ytimen, väitöskirjojen, voi ajatella syntyneen Plattin kuvaamassa maastossa. Tutkimusten erot ovat varmasti vaikuttaneet tilastollisten aineistojen ja analyysiohjelmien käyttöön.

Tutkimukseni on luettavissa tieteen tutkimuksen alaan kuuluvaksi. Tutkimukseni kohteena ovat kaksi tiedettä, sosiologia ja tilastotiede. Sen vuoksi lähden luvussa 2 taustoittamaan tutkimustani tieteiden yleisellä tarkastelulla. Tiede- ja tutkimusmaailman laajeneminen sekä monipuolistuminen on samalla myös eriyttänyt tieteenaloja toisistaan. Elämmekin aikaa, jossa tieteiden välisen yhteistyön merkitystä korostetaan. Esittelen luvussa 2 monitieteisyyden ja tieteidenvälisyyden määritelmiä tuoden esiin myös tieteiden välisyyden ja tieteiden väliin putoamisen. Pitkä kokemukseni tilastotieteilijänä on herättänyt kiinnostuksen pohtia omaa tieteenalaani, kuten Mika Kiikeri ja Petri Ylikoski (2004, 9–10) sekä Heikki Mikkeli ja Jussi Pakkasvirta (2007, 132) toteavat monen tieteen tutkijan osalta käyneen. Luvussa 2.2 esittelen tilastotieteen moninaisuutta ja historiaa sekä luvussa 2.3 tuon esiin sosiaalitieteiden ja sosiologian panoksen tilastollisten menetelmien kehittämisessä.

Luvussa 3 kerron tutkimukseni toteuttamisesta. Esittelen tutkimukseni perusaineiston, joka koostuu vuosina 1960–2018 Suomessa tarkastetuista sosiologian väitöskirjoista. Ensimmäisessä tutkimuskysymyksessä tarkastelen, kuinka paljon tilastollisia menetelmiä ja mitä monimuuttujamenetelmiä sosiologian väitöskirjoissa on käytetty vuosina 1960–2018. Toinen tutkimuskysymys kohdistuu sosiologian väitöskirjoissa esitettyihin faktorianalyysin valintaa koskeviin perusteluihin ja kolmas tutkimuskysymys faktorianalyysin suorittamiseen liittyviin valintoihin. Faktorianalyysi on eräs vanhimmista tilastollisista menetelmistä ja sillä on ollut erityinen asema suomalaisessa sosiologiassa. Nämä seikat vaikuttivat siihen, että kiinnostuin tutkimaan nimenomaan faktorianalyysin valintaperusteita ja käyttämisen tapoja sosiologian väitöskirjoissa. Analysoin väitöskirjoista muodostamiani erilaisia aineistoja sekä määrällisten että laadullisten menetelmien avulla.

Luvussa 4 avaan tutkimusprosessin kautta tilannetta, jossa sosiologian ja tilastotieteen näkemykset kohtaavat. Tuon esille, että tilastotieteen näkökulmasta tutkimusprosessi on erilainen kuin tutkimustaan tekevän sosiologin näkökulmasta. Yhteistyötä tehdessään – sosiologin tehdessä kvantitatiivista tutkimusta, jossa

tilastotieteilijä on menetelmäasiantuntijana tai analyysin suorittajana – ei edes ole tarkoituksenmukaista se, että he näkisivät tutkimuksen vaiheet samalla tavalla. Oleellista on kuitenkin se, että he ovat tietoisia erilaisista näkemyksistään. Ja ehkä erojen tunnistamisen kautta myös löytäisivät yhteisen ymmärryksen niiltä osin, kuin se tutkimuksen onnistumisen kannalta on tarpeellista.

Luvussa 5 vastaan ensimmäiseen tutkimuskysymykseen kuvailemalla, kuinka paljon sosiologit ovat käyttäneet väitöskirjoissaan tilastollisia menetelmiä vuosien 1960–2010 aikana. Teen monimuuttujamenetelmien osalta tarkempaa tarkastelua esittelemällä, mitä ja minkälaisia monimuuttujamenetelmiä väitöskirjoissa esiintyy ja kuinka paljon niitä on eri aikoina käytetty.

Luku 6 keskittyy Susan Leigh Starin ja James R. Griesemerin (1989) kehittämään rajakohdemalliin, joka soveltuu tieteidenvälisyyden tutkimiseen. Yleisen rajakohdemallin esittelyn jälkeen kerron, millä tavalla se on sovellettavissa faktorianalyysin valintaperusteiden ja käyttötapojen tarkasteluun. Tutkimuksessani rajakohdetta käyttävinä osapuolina ovat sosiologit ja tilastotieteilijät. Kauttakulkupisteitä on kolme. Ensimmäinen kauttakulkupiste kuvastaa tilastomatemattista näkökulmaa faktorianalyysiin. Toinen kauttakulkupiste edustaa sosiologista näkökulmaa, jolla tarkoitan erityisesti sosiologisiin tutkimuskysymyksiin nojautumista faktorianalyysin valinta- ja käyttötilanteissa. Kolmas, asetelmaan perustuva kauttakulkupiste asetuu tilastomatemattisen ja sosiologisen näkökulman väliin. Tutkimukseni kannalta kiinnostavin vaihe rajakohdemallissa liittyi kääntämissiin, joiden tarkoituksena on saada osapuolet vakuuttumaan tietyn kauttakulkupisteen paremmuudesta ja valitsemaan sen. Kyse on siis tieteellisestä auktoriteetista: kenen tai minkä tieteenalan näkemys painaa vaakakupissa eniten.

Luvussa 7 esittelen sisällönanalyysin ja rajakohdemallin avulla havaitsemiani tilastomatematiikkaan ja muuttujien ominaisuuksiin, faktorianalyysin asetelmaan ja sosiologiseen tutkimuskysymykseen viittaavia perusteluja ja niiden suosiota eri aikoina. Tuon esiin väitöskirjoissa esitetyt erilaiset faktorianalyysille asetetut tavoitteet.

Luvussa 8 tarkastelen, minkälaisia valintoja sosiologit ovat tehneet muuttujien, faktorointimenetelmien ja rotatointimenetelmien suhteen. Katson myös, millä tavalla sosiologit ovat jatkokäsitelleet faktorianalyysin tuottamia faktoreita.

Luvussa 9 teen yhteenvetoa ja pohdintaa faktorianalyysin ja yleisesti monimuuttujamenetelmien hyödyllisyydestä ja haasteellisuudesta sosiologian väitöskirjoissa.

2. Tutkimuksen taustaksi: tieteiden eriytymisen historiaa

2.1. Tieteiden yleinen kehittyminen

Tieteiden ajateltiin 1800-luvulle asti olevan joko filosofiaa tai fysiikkaa (Mikkeli ja Pakkasvirta 2007, 38; Gronow, Noro ja Töttö 1997, 15). Tosin muutamat tieteenalat, kuten matematiikka, tähtitiede, eläin- ja kasvitiede, lääketiede, kielioppi ja historiantutkimus olivat jo eriytyneet omiksi tieteikseen (Haaparanta ja Niiniluoto 1986, 90). Theodore M. Porter (1986) kuvaa tuon aikaisten tieteenharrastajien olleen yleistieteilijöitä, joiden näkemykset kattoivat monia tieteenaloja heidän ollessa tietoisia niin oman tieteenalansa kuin sen ulkopuolellakin tapahtuvasta kehityksestä.

Tieteet alkoivat eriytyä, laitostua ja ammattimaistua 1800-luvun lopulla. Mikkeli ja Pakkasvirta (2007, 27–30) esittelevät kirjassaan eri aikoina vallinneita länsimaisia tieteenkäsityksiä tieteen luokitteluja. 1800-luvun lopulla vallinnutta tieteen eriytymisen aikaa kutsutaan heidän mukaansa diffusionistiseksi luokitteluksi, vaikkakin eriytyminen on jatkunut vilkkaana myös sen jälkeen (ks. myös Gronow ym. 1997, 15.) Matti Sintonen (2005, 257) toteaa tieteen olevan jatkuvan muutoksen kourissa ”niiden syntyessä, kasvaessa ja kuollessa ja tarvittaessa jakautuessa kuin amebat”. Ilkka Niiniluoto (2003) on tarkastellut tieteen syntytapoja. Uudet tieteet ovat hänen mukaansa syntyneet eriytymällä emotieteestä, yhdentymällä toisen tieteen kanssa, teoreettisen integraation myötä, uusien tutkimusaiheiden esiinnousun tai taitojen tieteellistymisen myötä (Niiniluoto 2003, 133–137; ks. myös Mikkeli ja Pakkasvirta 2007, 31–34). Timo Airaksinen (1987, 68) on kuvannut tieteen pirstoutumista ja uusien tieteen syntymisen nopeutta hieman kriittiseen sävyyn toteamalla, että ”Suomessa on taipumus luoda tieteenaloja hallinnollisin toimenpitein: perustetaan professuuri ja tiede on syntynyt”.

Tieteiden eriytyessä ovat niiden osaamisalueet syventyneet mutta samalla kaventuneet, minkä seurauksena tieteet ovat kohdanneet kysymyksiä, joihin ei ole pystytty vastaamaan pysyttelemällä yhden tieteenalan sisällä (Kiikeri ja Ylikoski 2004, 80). Niinpä tieteen eriytyminen on synnyttänyt tarpeen tieteen väliselle yhteistyölle. Länsimaisessa tieteen luokittelussa tieteen yhteistyötä kuvaavaa aikaa kutsutaan interdisiplinaariseksi vaiheeksi. 1960-luvulla alkaneeseen interdisiplinaariseen vaiheeseen liittyvät termit monitieteisyys, tieteenvälisyys ja poikkitieteellisyys (Sintonen 2005, 263; Mikkeli ja Pakkasvirta 2007, 26). Ne ovat englanninkielisen käsitteen interdisciplinary suomennoksia. Samankaltaista näille käsitteille on Mikkelin ja Pakkasvirran (2007, 59–72) mukaan se, että tieteitä yhteen tuotaessa tieteen väliset rajat hämärtyvät mutta eivät katoa. He näkevät käsitteiden myös eroavan

toisistaan. Monitieteisyys (*multidisciplinarity*) kuvaa heidän mukaansa tutkimusta, jossa ilmiötä tarkastellaan eri tieteiden näkökulmista, mutta tieteiden välillä ei välttämättä synny todellista vuorovaikutusta vaan kukin tiede toimii omien paradigmojensa mukaisesti. Monitieteisessä tutkimuksessa tieteet siten pysyvät ja toimivat itsenäisinä tieteinä, vaikka yhteinen tavoite onkin ison tutkimuskysymyksen tai ilmiön selittäminen. Poikkitieteellisessä (*transdisciplinarity, crossdisciplinarity*) tutkimuksessa perinteinen tieteiden erottelu on tietoisesti unohdettu. Tutkimusongelma ja tutkimusprosessi ikään kuin kuljetetaan eri tieteiden läpi. Poikkitieteellisessä tutkimuksessa voidaan Mikkelin ja Pakkasvirran mukaan harrastaa ”metodologista anarkiaa”. Se voi synnyttää uudenlaisia ratkaisuja, paradigmoja ja jopa tieteenaloja. Poikkitieteellinen tutkimusote vaatii heidän näkemyksensä mukaan kokeneen monialaisesti kouluttautuneen tutkijan, joka kykenee keskustelemaan usean tieteenalan kanssa. (Mikkeli ja Pakkasvirta 2007, 59–72.)

Sintonen (2005, 257–259) määrittelee tieteidenvälisyyden olevan kattokäsite, jonka alla on erilaisia tieteidenvälisyyden asteita. Multidisiplinäärisessä eli monitieteisessä tutkimuksessa tieteiden väliset yhteydet ovat vähäisiä ja tieteet toimivat toistensa rinnalla säilyttäen oman identiteettinsä ja käyttäen niille tyypillisiä käsitteitä ja menetelmiä. Pluridisiplinäärisessä tutkimuksessa toisiaan lähellä olevat tieteet, esimerkiksi matematiikka, fysiikka ja tähtitiede, harjoittavat yhteistyötä. Cross-disiplinäärinen tutkimus on tieteidenväliset rajat ylittävää tutkimusta, jossa yksi tieteenala vaikuttaa yksisuuntaisesti toiseen. Interdisiplinäärisessä tutkimuksessa tieteenalat ovat vuorovaikutuksessa toistensa vaihtaen ajatuksia käsitteitä ja metodeja sekä kalibroiden terminologiaansa ja metodologiaansa. Viidentenä on transdisiplinäärinen tutkimus, joka rakentuu tieteenalojen yhteiseen viitekehykseen ja yhteisiin aksioomeihin. Sintonen huomauttaa, että kyseinen luokittelu ei ole ainoa mahdollinen eikä myöskään ongelmaton. Olennaista siinä hänen mukaansa on tieteidenvälisyyden asteiden perustuminen käsitteellisen ja käytännöllisen yhteistyön kiinteeseen. (Sintonen 2005, 257–259.)

Joe Moran (2002, 14–15) tuo esiin käsitteen *interdisciplinarity* kaksijakoisuuden viittaamalla Geoffrey Benningtonin toteamukseen inter-etuliitteen moniselitteisyydestä. Yhtäältä inter tarkoittaa liittämistä tai yhteyden muodostumista ja toisaalta se tarkoittaa erottamista ja erillään pitämistä. Sama monimerkityksellisyys heijastuu Moranin mukaan *interdisciplinarity* -käsitteeseen: se voi tarkoittaa yhteyden luomista tieteiden välille, mutta se voi luoda myös jonkinlaisen tieteettömän tyhjiöalueen tieteiden väliin. Mikkeli ja Pakkasvirta (2007, 60) viittaavat Moranin esitykseen todeten, että tieteidenvälisyys-sanassa ilmenevä välissä olemisen ajatus voi toimia sillanrakentajana tieteiden välissä, mutta se voi tarkoittaa myös syvän kuilun olemassaoloa eli jonkinlaista väliinpuotoamista.

Tieteiden välissä olevaa kuilua tai tieteiden väliin puotoamista voisi kutsua myös yhteiseksi ymmärtämättömyydeksi. Eräänä esimerkkinä tieteiden väliin puotoamisesta ja yhteisestä ymmärtämättömyydestä ovat tilanteet, joissa tieteiden edustajat eivät

hallitse jotakin asiaa, mutta luulevat toisen tieteen edustajan hallitsevan sen. Toinen esimerkki liittyy ilmausten tai käsitteiden erilaisen merkitykseen eri osapuolille siten, että osapuolet eivät tiedosta samoja ilmauksia käyttäessään puhuvansakin eri asioista (ks. esim. Ronkainen 1998, 247–248). Ronkainen (2001, 23–28) tuo esille, että käsitteiden lainaaminen toisesta tieteenalasta saattaa tuoda mukanaan kokonaisen teorian. Tutkijan on tiedostettava ja myös tuotava esille, kuinka hän käyttää käsitettä ja miten käsite eroaa eri tieteenaloissa. (Ronkainen 2001, 23 – 28.) Myös Mäkelä (1991, 17) hahmottaa käsitteiden moninaisuutta kirjoittaessaan, miten erilaisissa konteksteissa samannimisille käsitteille on annettu erilaisia merkityksiä. Käsitteet ovat pitkän kehityskulun tuloksia, jolloin alkuperäiset merkitykset ovat hämärtyneet ja muuttuneet (Mäkelä 1991, 18).

Jari Aro (1999, 9–10) tuo esille kielen ja käsitteiden keskeisen roolin sosiologiassa. Kielellinen ilmaisu on sosiologian keskeinen työväline, sen vuoksi käsitteet on pyrittävä määrittelemään kurinalaisesti ja ilmaisuissa pyrittävä sellaiseen täsmällisyyteen, että lukijalla jää mahdollisimman vähän tilaa omille tulkinnoilleen, Aro toteaa. Raymond Boudon (1993, 12–14) kertoo esimerkin *metodologia* -ilmauksen monimerkityksellisyydestä. Boudon tuo esille, miten Paul Lazarsfeldiä häiritsi se, että monet samastivat metodologian teknologiaan. Lazarsfeld näki metodologian olevan teknologiaa laajempi, joustavampi, eläväisempi ja älyllistä pohdintaa enemmän vaativa. Teknologiaa hän kuvasi työkalupakiksi, joka sisältää tutkimukseen liittyvinä työkaluina erilaiset säännöt ja keinot tutkimuksen toteuttamiseksi. Metodologiaa hän kuvasi ajatustyöksi ja kehittelyksi, jonka pohjalta työkalut ovat syntyneet. Lazarsfeld tiivistä näkemyksensä seuraavasti: ”The sociologist studies man in society: The methodologist studies the sociologist at work”.¹

Omassa tutkimuksessani huomioin tieteiden eriytymisen vaikutukset monella tapaa. Tieteiden yleinen kehittyminen ja eriytyminen taustoittavat tutkimustani. Sosiologiaa ja tilastotiedettä pidetään tänä päivänä kahtena toisistaan erillään olevana ja ajoittain täysin erilaisena tieteenä. Historiatarkastelu kuitenkin paljastaa niillä olevan monia kytköksiä menneisyydessä (Porter 1986, 152–156); Lazarsfeld 1993a, 263; Töttö 1997a; Alastalo 2005, 34). Tieteiden väliin putoamista ja yhteisen ymmärryksen löytymisen haasteita konkretisoin luvussa 4 havainnollistaessani tilastotieteen ja sosiologian kohtaamista tutkimusprosessissa. Tutkimukseni myös sijoittuu tieteiden väliin siinä mielessä, että tutkimukseni sijoittuu sosiologiseen menetelmien tutkimukseen, mutta tutkijana minulla on tilastotieteilijän tausta. Eriytyminen, pirstoutuminen ja osaamisalueen kaventuminen näkyvät myös nykyisten tieteiden sisällä. Sosiologian pirstoutumisen ovat tuoneet esiin muun muassa Johanna Hokka (2016) sekä Erola ja Räsänen (2014a, 241). Myös tilastotieteen voidaan nähdä pirstoutuneen tai edelleen pirstoutuvan, kuten käy ilmi seuraavassa luvussa.

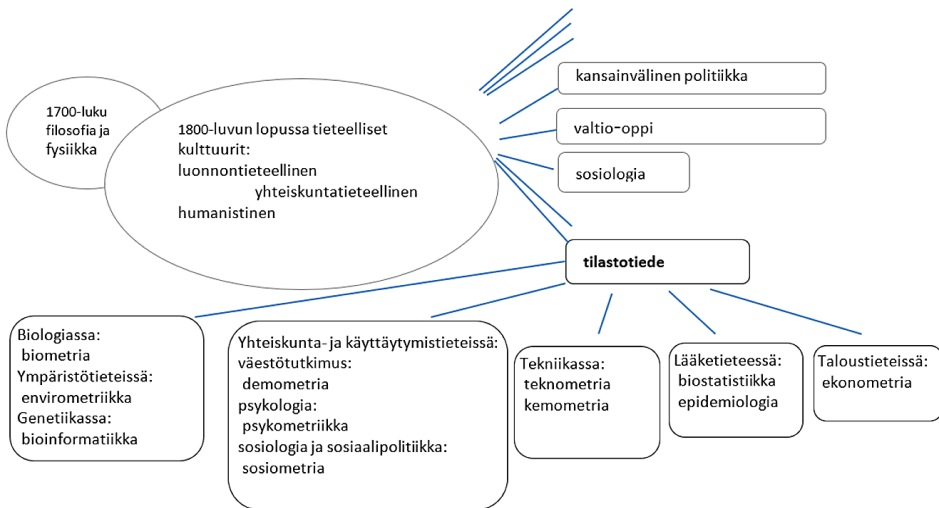
1 Lazarsfeld, Paul (1959). Problems in Methodology. Teoksessa Merton, Robert K., Broom, Leonard and Cottrell, Leonard S. Jr. (ed.) Problems and Prospects. Basic Books. New York. 39–78.

2.2. Tilastotieteen moninaisuus

Tiedettä tutkimuskohteenaan pitävää oppialaa kutsutaan tieteen tutkimukseksi. Monet tieteen tutkijat, eräänä tunnetuimmista Thomas Kuhn, ovat taustaltaan luonnontieteilijöitä, jotka ovat alkaneet pohtia ja tutkia omaa toimintaansa ja tieteenalaansa (Kiikeri ja Ylikoski 2004, 9–10; Mikkeli ja Pakkasvirta 2007, 132; Bird 2001). Oman tutkimukseni kipinä on myös toiminut pitkäaikainen kokemukseni tilastotieteen parissa ja sen myötä syntyneet huomioit tilastotieteen moninaisuudesta. Opiskeluajana tilastotiede näyttäytyi minulle lähinnä matemaattisena oppiaineena. Tilastotieteilijäksi valmistumisen jälkeen olen opettanut tilastotiedettä yhteiskunta- ja kasvatustieteiden opiskelijoille sekä ollut menetelmäasiantuntijana erilaisissa kvantitatiivisissa tutkimusprojekteissa. Vuosien kokemukset näissä tehtävissä ovat osoittaneet, että tilastotieteen sisältö, käytötapa ja rooli voivat vaihdella suuresti eri yhteyksissä. Joissakin tilanteissa tilastotiede palvelee matematiikkana, toisissa tilanteissa matemaattisena tutkimusmenetelmänä ja joissakin jopa melko laadullisena tutkimusmenetelmänä.

Seppo Raiski (1976) on todennut tilastollisten tutkimusmenetelmien opiskelussa olevan kaksi puolta. Ensimmäinen on hänen mukaansa tilastollisen teorian hallinta, joka on tilastotieteen tehtävä. Toinen on harjaantuminen niiden käyttöön yhteiskuntailmiöiden erittelyssä, mikä puolestaan on yhteiskuntatieteen metodinen tehtävä. (Raiski 1976, 59.) Tilastotiedettä ei siten voida samaistaa puhtaaseen matematiikkaan (Bellhouse 2000). Tilastotieteen sijoittuminen tieteiden jaottelussa on ollut vaihtelevaa. Historiansa aikana tilastotiede on voitu lukea niin matemaattiseksi kuin yhteiskuntatieteelliseksi ja joskus myös maantieteeseen kuuluvaksi tieteeksi (Porter 1986, 24–25). Tilastotieteen nimen on katsottu liittyvän valtioihin tai valtiolliseen, tilaan tai numeeriseen tietoon yhteiskunnasta (Porter 1986, 11–12; 23–25; 38; 152–153; Alastalo 2005, 30).

Hans Helenius (1992, 3) toteaa kirjassaan, miten tilastotieteen yhteistyö muiden tieteiden kanssa on hiljalleen häivyttänyt tieteiden rajat ja yhteistyön aluetta on alettu nimittää omalla nimellä. Esimerkiksi lääketieteen ja tilastotieteen yhteistyön aluetta kutsutaan biostatistiikaksi, psykologian ja tilastotieteen yhteistyön aluetta psykometriaksi ja taloustieteen ja tilastotieteen yhdistävää alaa ekonometriaksi. Kuva 1 havainnollistaa tilastotieteen käytön kirjoja eri tieteissä.

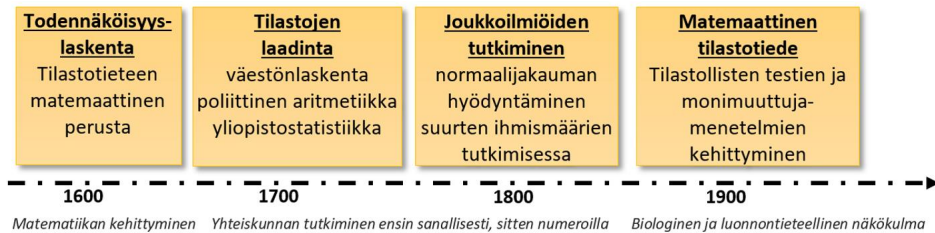


Kuva 1 Tieteiden eriytyminen

Kirjallisuudessa ei ole pystytty esittämään tiukkarajaista, yleisesti hyväksyttyä määritelmää tilastotieteelle, koska eri tahojen esittämissä määritelmässä tiivistyvät erilaiset näkökulmat, lähestymistavat ja painotukset (Helenius 1992, 1). David Bartholomew (1995a) esittää tilastotieteelle useita määritelmää päätyen siihen, että tilastotiedettä on melko mahdotonta määritellä kattavasti. Mäkelän (1991, 11) mukaan käytännössä ei ole olemassa mitään yksikäsitteistä tilastollista metodologiaa vaan useita erilaisia vaihtoehtoja, joiden taustalta on löydettävissä omat tieteenfilosofiset sitoumuksensa. Sosiologi ja tilastotieteilijä Clifford C. Clogg (1992) on halunnut erottaa aineiston analyysiin tarkoitetun tilastotieteen tilastomatematiikasta. Aineiston analyysiin soveltuvaa tilastotiedettä hän kutsuu vuonna 1992 julkaistussa artikkelissaan mieluummin tilastolliseksi metodologiaksi. David Hand (2009) on Cloggin kanssa samoilla linjoilla ilmaistessaan, että tilastotieteen voisi jakaa laajaan ja suppeaan tilastotieteeseen. Laaja tilastotiede liittyisi Handin mukaan kaikkeen aineistosta oppimiseen. Suppea tilastotiede taas viittaa tilastotieteilijöiden kehittämään metodologiseen ytimeen eli sellaiseen tilastotieteeseen, jota julkaistaan alan aikakauslehdissä ja josta tehdään esimerkiksi väitöskirjat. Handin näkemyksen mukaan moderni tilastotiede tulisi määritellä laajan tilastotieteen tapaan.

Eskola (1971) on todennut, että ”Ymmärtäksemme käyttämiämme menetelmiä meidän on tarkasteltava niitä historiallisesti ja rakenteellisesti. Menetelmä tulee tajuttavaksi, kun selvitetään sen rakenne; ja sen käyttö selittyy, kun tuo rakenne sijoitetaan laajempaan, tieteenfilosofiseen ja yhteiskunnalliseen rakenteeseen, joka on sitä ennen selvítettävä ja tajuttava.” Myös Alastalo (2005, 295) korostaa historian tuntemuksen tärkeyttä todetessaan, että historiaa tutkimalla voi tehdä näkyväksi ja ymmärrettäväksi tieteenalalla tapahtuneita muutoksia. Eskolan ja Alastalon linjaa

myötäillen teen lyhyen tieteiden, tilastollisen ajattelun ja tilastotieteen kehittymistä koskevan historiakatsauksen havainnollistaakseni tilastotieteen moninaisuutta. Tilastotieteen historiaan ja tilastollisen ajattelun syntyyn voi tarkemmin tutustua esimerkiksi D. B. Owenin, Paul D. Mintonin ja John W. Prattin (1976), Alain Desrosièresin (2002) tai Ian Hackingin (1990) teoksissa. Julian Champkin (2014) on esittänyt tilastotieteen kehittymisen aikajanana. Tässä yhteydessä esitän yksinkertistetun aikajanana tilastotieteen osa-alueiden kehittymisestä (kuva 2).



Kuva 2 Tilastotieteen osa-alueiden kehittyminen

Tilastotieteen matemaattinen perusta, todennäköisyyslaskenta, sai alkunsa 1600-luvulla. 1700-luvulla kehitettiin yliopistostatistiikan ja poliittisen aritmetiikan piirissä erilaisia tilastointeja ja taulukointeja, joiden avulla yhteiskuntia ja valtioita pyrittiin hallitsemaan. Eräs kytkös liittyy joukkoilmiöiden tutkimisen yhteydessä tehtyihin keksintöihin. Porter (1986) kuvaa kirjassaan, millä tavalla 1800-luvulla, joka oli merkittävä ajanjakso tilastollisen ajattelun kehittämisessä, oivallettiin muun muassa normaalijakauman hyödyntäminen joukkoilmiöiden tutkimisessä. Keskeisenä kehittäjänä tässä oli Adolphe Quételet, joka työskenteli observatoriossa tutkien taivaankappaleiden liikeratoja. Taivaankappaleilla ajateltiin tiettyinä aikoina olevan tietty sijainti. Mitattaessa liikeratoja ja sijainteja ajateltiin mittauksissa esiintyvän vaihtelun johtuvan puutteellisista mittausvälineistä. Mittausvirheiden arvioimiseksi kehiteltiin menetelmiä, joista hyödylliseksi osoittautui eksponentiaalinen funktio hyödyntävä virhelaki. Myöhemmin virhelaki opittiin tuntemaan Gaussin käyränä eli normaalijakaumafunktiona. Quételet oivalsi, että virhelaki ja erityisesti siihen liittyvä keskiarvo sopivat myös ihmisten fyysisten piirteiden jakaumien kuvailuihin ihmispopulaatioissa. Quételet kiinnostivat myös joukkoilmiöiden taustalla olevat syyt. Niitä tarkastellakseen hän taulukoi ilmiöiden esiintymismääriä erilaisten taustatekijöiden, esimerkiksi iän, sukupuolen, ammatin ja syntymäpaikan mukaan pohdittien huolellisesti havaitsemiensa lukujen merkityksiä ja vaikutuksia. Havaitsemiaan säännönmukaisuuksia Quételet kutsui tilastollisiksi laeiksi. Hän tulkitsi, että tilastolliset lait toteutuvat ryhmässä, vaikka ne eivät päde yksilötasolla. Porter kirjoittaa Quételetin päätyneen siihen, että tilastollinen säännöllisyys ja suurten lukujen laki olivat avaimia numeeriseen sosiaalitieteeseen. (Porter 1986.)

Toinen tilastotieteen ja sosiologian kytkös liittyy tilastotieteen nimeen. Sosiologian perustajana pidetty Auguste Comte kehitti 1800-luvulla positivistisen tieteen nimittäen sitä sosiaalifysiikaksi. Comte ajatteli tieteiden kehittyvän teologisesta metafyyisiseen ja siitä edelleen positiiviseen vaiheeseen. Matematiikkaan ja todennäköisyyksiin ihastunut Quételet omi Comten ideat kehitellessään joukkoilmiöitä koskevaa tiedettään. Quételet otti käyttöönsä myös Comten omalle tieteelleen keksimän nimen *physique sociale*. Comte, joka Porterin mukaan suorastaan karttoi matematiikan, numeroiden ja todennäköisyyksien käyttöä, oli tästä todella vihainen. Kun nimi vielä virheellisesti käännettiin tilastotieteeksi, oli Comten keksittävä omalle tieteelleen uusi nimi. Hän päätti nimetä tieteensä sosiologiaksi. (Porter 1986, 152–156; Lazarsfeld 1993b, 263; Töttö 1997b.)

Tilastotiede muuttui tieteenä matemaattiseksi ja erityisesti biologiaan ja luonnontieteisiin soveltuvaksi 1900-luvulla. (Porter 1986; Alastalo 2005; Kuusela 2010; Champkin 2014.) Porterin (1986, 24) mukaan on vaikea arvioida, milloin ihmiset tarkalleen alkoivat nähdä tilastotieteen matemaattisena tieteenä. Mäkelä (1996) esittää matemaattisen tilastotieteen kehittymisen tapahtuneen kahtena ajanjaksona. Ensimmäinen jakso (1890–1925) liittyi voimakkaasti biologiaan. Silloin kehitettiin korrelaatioajattelu, koesuunnittelu, estimointiteoria sekä merkitsevyytestit. Toisen vaiheen (1926–1934) aikana kehitettiin hypoteesien testaukseen ja luottamusvälien laskemiseen liittyvät teoriat. Matemaattisen tilastotieteen kehittyessä tilastotieteen omat tutkimusongelmat alkoivat muotoutua. Matemaattisen tilastotieteen kehittymisen myötä luonnontieteellisyys kiinnittyi vahvasti tilastotieteeseen ja tilastollisten menetelmien käyttöön. Tieteenalojen ja tutkijoiden, jotka käyttivät tilastotieteen menetelmiä, oli siten myös omaksuttava matemaattisesti painottunut, luonnontieteellinen ajattelutapa omissa tutkimuksissaan. Mäkelä (1996.)

2.3. Sosiologian panos tilastotieteen kehittymiseen

Sosiologian synty liittyy 1800-luvulla käytyyn pohdintaan tieteiden ykseydestä. Tuolloin pohdittiin erityisesti ihmistä, kulttuuria ja yhteiskuntaa tutkivien tieteiden omaleimaisuutta luonnontieteisiin verrattuna (Haaparanta ja Niiniluoto 1986, 97; Gronow ym. 1997, 22–26). C. P. Snow (1998, 98–99; 174–177) kuvaa 1800-luvun loppua kahden toisiaan ymmärtämättömän – luonnontieteellisen ja humanistisen – kulttuurin ajaksi. Snown mukaan kulttuurien väliin tarvittiin kolmas kulttuuri, joka kykenisi kommunikoimaan sekä luonnontieteellisen että humanistisen ajattelutavan kanssa. Tätä tarvetta täyttämään syntyivät yhteiskuntatieteet, myös sosiologia. Wolf Lepenies (1988, 1–15) näkee, että historiastaan johtuen sosiologia on joutunut olemassaolonsa aikana tasapainottelemaan luonnontieteellisen ja humanistisen ajattelutavan välillä. Sosiologia on hänen mukaansa korostanut tieteellisyyttään luonnontieteistä otetuilla vaikutteilla ja samalla sosiologia on kilpaillut kirjallisuus-

ja historiatieteiden kanssa todellisuuden kuvaamisessa. (ks. myös Gronow ym. 1997, 22–26; Allardt 1988, 12–14; Mikkeli ja Pakkasvirta 2007, 40–42.)

Sosiologian tasapainottelu erilaisten ajattelutapojen välillä on näkynyt muun muassa kiistana kvantitatiivisen ja kvalitatiivisen tutkimusotteen paremmuudesta. Tähän keskusteluun voi tutustua esimerkiksi George Lundbergin (1943), Alan Brymanin (1988), Christopher G. A. Bryantin (1985), John W. Cresswellin (1994), Jukka Mäkelän (1996), Pertti Tötön (1997b; 2000b; 2004; 2012), Pertti Alasuutarin (1989; 1999; 2000), Anna-Maija Lehdon (1998), Kyösti Raunion (1999), Suvi Ronkaisen (1999), Marja Alastalon (2005, 17–19), David Byrnen (2011, 19–39) ja Arto Selkälän (2013, 104–120) tekstien välityksellä.

Tutkimusmenetelmien kehittyminen on ollut keskeisessä roolissa tieteellisen tiedon hankkimisessa, toteavat Leila Haaparanta ja Niiniluoto (1986, 1). Vastavuoroisesti ovat monet tutkimusmenetelmät kehittyneet pyrittäessä vastaamaan eri tieteenalojen tutkimuskysymysten tarpeisiin (ks. esim. Harman 1976). Porterin (1986) tilastollisen ajattelun historiakuvauksesta käy myös ilmi, miten tilastollisen tilastolliset menetelmät kehittyivät 1800-luvulla paljon sen ansiosta, että eri tieteenalan edustajat – tähtitieteilijät, biologit, fyysikot ja sosiaalitieteilijät – tekivät yhteistyötä lainaten toisiltaan menetelmiä ja tekniikoita.

Cloggin (1992) mukaan sosiologian metodologinen kehitys on aina ollut läheisesti sidoksissa tilastotieteen kehittymiseen ja toisaalta monet nykyisin käytössä olevista tilastollisista menetelmistä ovat saaneet alkunsa muiden tieteiden tutkimustarpeista². Clogg korostaa, että tilastotieteellisellä metodologialla on ollut valtava vaikutus sodanjälkeiseen sosiaalitieteelliseen tutkimukseen, mutta myös sosiaalitutkimuksen metodologialla on ollut vaikutuksensa nykyaikaiseen tilastotieteeseen. Sosiaalitieteet, erityisesti sosiologia ovat hänen mukaansa kohdanneet niin haastavia metodologisia kysymyksiä, että niihin ei olisi voitu vastata keskittymällä vain kokeellisen aineiston varianssianalyyysiin tai matemaattisesti tyylikkäisiin ja siisteihin ongelmiin, mitä usein näkee teknisissä julkaisuissa. Esimerkiksi log-lineariset mallit, tapahtumahistoriamallit ja latentin muuttujan mallit ovat Cloggin mukaan syntyneet ja kehittyneet sosiaalitutkimuksen tarpeista ja niitä kehittäneillä tilastotieteilijöillä on ollut läheiset sidokset sosiaalitieteisiin. Alun perin sosiaali-

2 Cloggin mukaan Wilhelm Lexis loi perustan tapahtumahistorian väestötieteellisille malleille, joita käytetään taloustieteellisissä ja sosiologisissa tutkimuksissa. Toisena tutkijana Clogg mainitsee Gustav Fechnerin ja muut varhaiset psykofyysikot, jotka hänen mukaansa todistettavasti loivat perustan modernille kokeelliselle menetelmälle. Tärkeiksi vaikuttajiksi Clogg mainitsee myös Francis Galtonin ja Karl Pearsonin, joiden tutkimukset vaikuttivat suuresti sosiaalisen ja fyysisen periytyvyyden tutkimiseen. Myös regressiomallit ovat Cloggin mukaan silmiinpistävästi samantyyppisiä kuin esimerkiksi sosiaalisen liikkuvuuden tutkimuksissa käytetyt menetelmät. Clogg mainitsee myös aikasarja-analyysejä ja sosiaalitaloudellisia indikaattoreita kehittäneen Francis Ysidro Edgeworthin, joka loi perustan sosiaalitieteellisiin aineistoihin käytettävälle regressiolle ja korrelaatiolle ja teki pohjatyötä myös nykyisille kausaalimalleille tai rakenneyhtälömalleille. Viimeisenä Clogg mainitsee G. Udney Yulen, jolla oli hänen näkemysensä mukaan selvästi sosiaalitieteelliset aineistot mielessään kehittäessään assosiaatioteoriaansa.

tutkimuksen haasteisiin kehitetyistä metodologisista ratkaisuista on myöhemmin tullut merkittävää matemaattista ja sovellettua tilastotiedettä. Clogg tuo esiin, että sosiaalitutkimusta tekevät ovat myös kehittäneet monimuuttujamenetelmiä, jotka soveltuvat ei-numeerisille muuttujille. (Clogg 1992.)

Boudon (1993, 1–29) tuo esille sosiologian tutkimusmenetelmien keskeisen kehittäjän Lazarsfeldin huomattavan panoksen muun muassa surveytutkimuksen metodologiaan sekä luokitellun aineiston ja paneeliaineistojen analysointiin. (Boudon 1993, 1–29.) Unohtaa ei voi myöskään akateemisen sosiologian perustajaa Émile Durkheimia, jota pidetään tilastollisten menetelmien käytön pioneerina sosiologiassa ja jonka tutkimusten keskiössä oli teorian ja metodologian vuorovaikutus (Selvin 1958, 619; Töttö 1997b, 194; Alastalo 2005, 38–39; Aro ja Jokivuori 2015, 12; 282–283). Myös Adrian E. Raftery (2000; 2005), joka on tarkastellut tilastollisten menetelmien käyttöä ja kehittelyä amerikkalaisessa sosiologiassa vuosina 1950 – 2000, toteaa sosiologien olleen eturintamassa tilastollisten menetelmien kehittämässä. Raftery mainitsee muutamien menetelmäkehittelyjen saaneen alkunsa sosiologisista ongelmista, ja joitakin tilastollisia menetelmiä on ensimmäisen kerran esitetty nimenomaan sosiologisissa julkaisuissa. Sosiologialla ja tilastotieteellä on siten useampia historiallisia kohtaamispisteitä.

3. Tutkimuksen kohteena tilastolliset menetelmät, erityisesti faktorianalyysi suomalaisessa sosiologiassa

3.1. Aineistona sosiologian väitöskirjat 1960–2018

Tutkimukseni perusaineistona ovat Suomessa vuosina 1960–2018 tarkastetut sosiologian väitöskirjat (liite 1). Tiedot suomalaisista sosiologian väitöskirjoista on kerätty neljässä erässä. Vuosina 1947–1992 tarkastettujen väitöskirjojen tiedot on kerännyt Teuvo Rätty (1988; 1989). Alastalo on täydentänyt tiedot vuoteen 2000 asti (Alastalo 2005, 42–43). Sain bibliografian Alastalolta ja täydensin sitä edelleen vuoteen 2007 asti. Tietojen täydentämisessä noudatin Alastalon linjaa hyväksymällä bibliografiaan vain ne väitöskirjat, jotka on tarkastettu sosiologian ainelaitoksilla ja näin ollen jätin ulkopuolelle esimerkiksi koulutus-, liikunta- ja taloussosiologiassa tarkastetut väitöskirjat³ (ks. Alastalo 2005, 43). Toisen kerran täydensin listaa vuosien 2008–2018 osalta. Tällöin tiedot Suomessa tarkastetuista sosiologian väitöskirjoista löytyivät kohtalaisen helposti, koska ne olivat sähköisessä muodossa. Löysin tiedot hakukoneilla sekä yliopistojen ja kirjastojen sivuilta. Haastavuus uusimpien väitöskirjatietojen keräämisessä syntyi sen arvioimisesta, kuuluuko väitöskirja todella sosiologian alaan vai luetaanko se ennemmin sosiaalitieteisiin, sosiaalityöhön, sosiaalipolitiikkaan tai esimerkiksi liikunnan yhteiskuntatieteisiin. Tarkistin väitöskirjojen kuulumista sosiologian alaan eri tavoin. Käytin hyväkseni muun muassa yliopistojen väitöstiedotteita, joissa usein mainittiin, mihin oppiaineeseen tai alaan väitös sijoittuu. Vuosien 2008–2018 väitöskirjojen osalta löytyy väitöskirjoja, jotka on listattu kuuluvaksi usealle alalle. Olen ottanut ne aineistooni mukaan, jos sosiologia on kuvattu ensimmäiseksi alaksi.

Tilastollisten menetelmien, erityisesti monimuuttujamenetelmien, suosiota ja käyttöä tarkasteltaessa on väistämättä huomioitava tietokoneiden ja tilastollisten ohjelmien vaikutus (Hair ym. 1998, 5). Väitöskirjojen tarkastelun alkaminen vuodesta 1960 perustuu ensisijaisesti siihen, että tietokoneiden hyödyntäminen tilastollisten analyysin suorittamisessa alkoi vähitellen olla mahdollista. Vuosina 1955–1960 oli

3 Vertasin yliopistojen kotisivuilta keräämiäni tietoja Sosiologia-lehden julkaisemiin väitöskirjatietoihin. Lehdissä (4/2002, 1/2004, 1/2005, 1/2006, 1/2007) on listattu Suomessa ilmestyneet sosiologian alaan kuuluvat väitöskirjat vuosilta 2002–2006. Tämän jälkeen lähetin yliopistokohtaiset väitöskirjalistat tarkistettaviksi kunkin yliopiston sosiologian opintoja hoitavalle toimistosihteerille. Sain varmistukset Joensuun yliopistosta, Kuopion yliopistosta, Turun yliopistosta ja Åbo Akademista. Käytin tarkastuksessa apuna myös internetissä olevia yliopistojen tiedotteita väitöksistä, usein niissä mainittiin, mihin oppiaineeseen tai alaan väitös sijoittuu.

rakennettu ESKO-matematiikkakone. Suomen Kaapelitehdas perusti keväällä 1960 laskentakeskuksen, jonka ensimmäiseksi tietokoneeksi tuli Elliott 803 A. Se korvattiin pian tehokkaammalla 803 B-mallilla. Turun laskentakeskuksessa käynnistettiin joulukuussa 1960 Wenner-Grenin säätiön lahjoituksena saatu Wegematic 1000 -tietokone. Kaapelitehtaalle tuli vuoden 1961 lopussa toiseksi tietokoneeksi Siemens 2002. (Suominen 2000.) Vuosina 1965–1966 Suomessa oli käytössä 84 tietokonetta, joista suuri osa sijaitsi Helsingin seudulla ja joita enimmäkseen käytettiin kaupallishallinnolliseen tietojenkäsittelyyn. Sitra rahoitti korkeakoulujen yhteiskäyttöön ”suurtietokone” Univacin 1108 vuosina 1970–1971. Keväällä 1971 Suomessa oli 210 suurta tietokonetta ja noin 700–800 minitietokonetta. Ensimmäiset mikroprosessorit, tyypiltään Intel 8008, tuotiin Suomeen vuonna 1973. (Suominen 2000.)

Sosiologian väitöskirjoissa ensimmäinen maininta tietokoneen käytöstä löytyy vuonna 1964 Helsingin yliopistossa tarkastetusta Yrjö Uiton väitöskirjasta. Vuonna 1965 ja sen jälkeen valmistuneissa väitöskirjoissa analyysit on harvoja poikkeuksia lukuun ottamatta tehty tietokoneita hyödyntäen.

Toinen perustelu tarkastelun alkamiselle vuodesta 1960 liittyy monimuuttujamenetelmien käyttöön. Sosiologit käyttivät monimuuttujamenetelmiä jonkin verran 1950-luvulla (Mäkelä 1996, 209; Alastalo 2005, 73), mutta 1960-luvulla niiden suosio kasvoi huomattavasti. Suosion kasvuun vaikuttivat sekä tietokoneiden suomat mahdollisuudet, että lisääntynyt tietoisuus monimuuttujamenetelmistä (Tarkkonen 2007; Alastalo 2005; 2014).

Aikarajaukseen ovat vaikuttaneet myös Mäkelän (1991; 1996) tutkimukset. Mäkelä on tutkinut tilastollisten menetelmien käyttöä sosiologiassa toisesta maailmansodasta alkaen vuoteen 1957 asti. Lisäksi 1950-luvun lopussa ja 1960-luvulla julkaistiin tilastollisia menetelmiä esitteleviä oppikirjoja, jotka tulivat keskeisiksi sosiologisessa tutkimuksessa. Erik Allardtin kirjoittama Nuorison harrastukset ja yhteisön rakenne ilmestyi vuonna 1958, minkä myötä survey-analyysi ja ristiintaulukointi vakiintuivat sosiologien käyttöön (Eskola 1973, 300–301). Tilastollisia menetelmiä opetettiin myös Sakari Sariolan (1956) kirjassa Sosiaalitutkimuksen menetelmät sekä Toivo Vahervuon ja Yrjö Ahmavaaran (1958) Johdatus faktorianalyysiin -teoksessa. Erik Allardtin ja Yrjö Littusen vuonna 1958 julkaistu Sosiologia vakiintui standardioppikirjaksi kuten myös Eskolan vuonna 1962 julkaistu Sosiologian tutkimusmenetelmät I ja vuonna 1967 julkaistu Sosiologian tutkimusmenetelmät II. (Alastalo 2005, 58–84.) Myöhemmin Eskola (1994, 13) arvioi, että hän lienee vaikuttanut suomalaisen yhteiskuntatieteeseen ennen kaikkea tutkimusmenetelmäkirjojensa kautta viitaten siihen, että kirjat ovat olleet kymmenien vuosien ajan tenttikirjoina ja niihin on viitattu väitöskirjoissa usein metodisena lähteenä.

Toivo Salonen (2004, 61) toteaa tutkimustilanteiden olevan eri asia kuin siitä jälkikäteen kertovat kirjalliset esitykset. Myös Ronkainen (1999, 100–101) ja Mäkelä (1991; 1996) ovat tuoneet esille vertailun käytännössä tehdyn tutkimuksen ja oppikirjaparadigman eroavuudesta, jolla he tarkoittavat esimerkiksi kvantitatiivisen

tutkimuksen käytäntöjen eroavan oppikirjojen tarjoamista menetelmäkuvauksista. Salonen (2004, 61) mainitsee esimerkkinä nimenomaan akateemiset opinnäytetyöt, joissa pyritään loogiseen täydellisyyteen ja sen vuoksi lopullinen esitystapa usein etäännytty kauas siitä, miten tutkimuksessa on edetty. Näin ollen kirjalliset esitykset eivät kerro paljoakaan varsinaisista tutkimuksista, erehdyksistä ja vaivalloisesta tiestä asioitten hahmottamiseen (Salonen 2004, 62). Ronkainen ym. (2013, 123) tuovat esiin tutkimusraporteissa esiintyvän katveen todeten, että raporteissa analyysitapa on nimetty, mutta ei ole kuvattu sitä, mitä analyysiprosessissa on tehty.

Oman tutkimusaiheeni osalta tiedostan, että väitöskirja-aineistosta en saa selville sitä, mitä kaikkea väitöstutkija on analyysissaan oikeasti tehnyt. Alastaloa (2014) mukaillen totean, että vaikka väitöskirjoissa menetelmät kuvataan yleensä tarkemmin kuin artikkeleissa, jää kuvauksiin kuitenkin suuriakin katvealueita. Voin lukea väitöskirjoista sen, millä tavalla tutkimus kerrotaan kokonaisuudessaan toteutetun, mutta kuten Kiikeri ja Ylikoski (2004, 24) tuovat esille tieteellisten artikkeleiden osalta, antavat kuvaukset harhaanjohtavan suoraviivaisen kuvan tutkimusprosessin kulusta. Tutkimuksen harhapolut, umpikujat ja keksimisprosessin monet elementit jäävät kertomatta. He myös toteavat, että pelkkä julkaistuihin tuloksiin keskittyminen saattaa antaa vääristyneen kuvan tieteellisen tutkimuksen onnistumisesta, koska vain positiiviset tulokset ja onnistuneet asetelmat kerrotaan ja epäonnistumiset jätetään kertomatta. (emt.)

Alastalo (2005, 42–44) ja Mäkelä (1996, 16–18) ovat pohtineet väitöskirjojen käyttöä tutkimusaineistona. Väitöskirjat eivät aineistona mahdollista tarkan kuvan piirtämistä siitä, miksi ja millä tavalla sosiologit ovat faktorianalyysia käyttäneet enkä sitä tutkimuksessani tavoittelekaan. Tutkimuksellani haluan tuoda esiin tilastollisten menetelmien käyttöön ja valintoihin liittyviä monia näkökulmia ja siihen väitöskirja-aineisto tarjoaa aineksia.

3.2. Tutkimuskysymykset ja analysointitavat

Tieteentutkimuksessa käytetään lukuisia erilaisia tutkimusmenetelmiä ja -aineistoja, mutta tyypillisesti tiedettä tarkastelevalla tutkijalla on omasta tieteenalastaan peräisin oleva menetelmällinen ja teoreettinen tausta, jonka myötä hän valitsee tutkimusnäkökulmansa ja erikoistuu tiettyihin menetelmiin ja aineistoihin. (Kiikeri ja Ylikoski 2004, 10–20). Tutkimuksessani tarkastelen sosiologian väitöskirjoissa käytettyjä tilastollisia menetelmiä. Oman tutkimukseni aineisto, sosiologian väitöskirjat, edustaa lähtökohtaisesti kvalitatiivista tekstiaineistoa, mutta olen muodostanut väitöskirjojen pohjalta myös kvantitatiivisia aineistoja. Siten analysoin tutkimusaineistoja sekä määrällisin että laadullisin menetelmin. Aineiston käytön ja sen keskeisyyden näkökulmasta tutkimukseni on empiiristä tieteentutkimusta.

Tieteentutkimuksessa pohditaan tieteen rajoja ja mahdollisuuksia tavoitteena tieteen moninaisuuden ymmärtäminen ja siinä tapahtuneiden muutosten tutkiminen. Käsitteet tieteestä, sen tavoitteista ja menetelmistä ovat historian aikana vaihdelleet (Haaparanta ja Niiniluoto 1986, 8; Kiikeri ja Ylikoski 2004, 15). Inhimillisenä toimintana tiede voidaan katsoa tutkimuksen avulla toteutettavaksi tiedonhankinnaksi ja tieteitä voidaan hahmottaa tutkimuskohteidensa kautta (Haaparanta ja Niiniluoto 1986, 11; 90; Salonen 2004, 77). Joskus tieteellä viitataan yhteiskunnallisiin instituutioihin, jolla on oma yhteiskunnallinen tehtävänsä ja omat tavoitteensa (Haaparanta ja Niiniluoto 1986, 9). Tieteeksi voidaan nimittää myös tietoa hankkivia ja ylläpitäviä sosiaalisia muodostumia (Kiikeri ja Ylikoski 2004, 16). Matti Sintosen (2005, 252) ja Kristiina Rolinin (2006, 18–21) mukaan tiede voidaan nähdä sosiaalisena prosessina. Rolin kirjoittaa tieteellisen tiedon olevan subjektiivista ja sidoksissa kunkin ajan vallitsevaan ajatteluun (ks. myös Haaparanta ja Niiniluoto 1986, 15). Tieteen subjektiivisuutta voidaan vähentää julkisuusperiaatteella, jonka mukaan tutkijan on hyväksyttävä väitteensä ja tutkimustuloksensa tiedeyhteisössä. Rolin (2006, 18–21) toteaa, että tällöin tieteellisen tiedon todellisia subjekteja ovatkin tutkijayhteisöt ja tiede voidaan näin ollen nähdä myös sosiaalisena käytäntönä, jossa tiedolliset arvot perustuvat tutkimustulosten ja tieteentekijän hyveisiin, mutta myös tiedeyhteisön toimintaa ohjaaviin arvoihin.

Tieteentutkimuksen osa-alueita tai tarkastelukulmia nimetään tieteentieteiksi (Kiikeri ja Ylikoski 2004, 9; Haaparanta ja Niiniluoto 1986, 89). Tieteensosiologiassa tutkitaan esimerkiksi tiedeinstituution yhteiskunnallisia tehtäviä sekä tutkijoiden ammattikunnassa vallitsevia arvo- ja normijärjestelmiä (Haaparanta ja Niiniluoto 1986, 89). Tieteentutkimuksen osa-aluejako ei ole yksiselitteinen eikä yksimielinen, koska perinteiset tieteenalarajat eivät aina ole sopivia valittuun tarkastelukulmaan ja on hyvin mahdollista, että tieteen tarkastelu vaatii useiden näkökulmien yhdistämistä (Kiikeri ja Ylikoski 2004, 11).

Sijoitan oman tutkimukseni kuuluvaksi tieteensosiologian tutkimuksen kenttään, vaikka siinä on tunnistettavissa myös muita tieteentutkimuksen osa-alueita. Tieteensosiologian tunnetuimpia tutkijoita on Bruno Latour (1987; 1990; 1993). Latour on kehittänyt Michel Callonin ja John Law'n kanssa toimijaverkkoteorian, jonka pohjalta Star ja Griesemer (1989) puolestaan kehittivät rajakohdemallin, jota käytän omassa tutkimuksessani.

Empiirisen analyysin avulla haen vastauksia seuraaviin tutkimuskysymyksiin:

1. Kuinka paljon tilastollisia menetelmiä ja mitä monimuuttujamenetelmiä sosiologian väitöskirjoissa on käytetty vuosien 1960–2018 aikana?
2. Minkälaisia perusteluja sosiologit ovat väitöskirjoissaan esittäneet faktorianalyysin käytölle?
3. Minkälaisia faktorianalyysin suorittamiseen liittyviä valintoja sosiologit ovat tuoneet esille väitöskirjoissa?

Ensimmäiseen tutkimuskysymykseen vastaaminen toimii samalla väitöskirja-aineistoni kuvailuna. Selvitän, kuinka paljon tilastollisia menetelmiä sosiologian väitöskirjoissa on käytetty vuosina 1960–2018. Tätä varten olen käynyt läpi kaikki lähes kuuden vuosikymmenen aikana tarkastetut sosiologian väitöskirjat, joita on 532 kappaletta. Katson, kuinka paljon sosiologian väitöskirjoja on eri aikoina tarkistettu ja kuinka monessa on käytetty tilastollisia menetelmiä. Teen lyhyen yliopistokohtaisen katsauksen eri aikoina tarkastettujen tilastollisia menetelmiä sisältävien väitöskirjojen määristä. Yleisen tilastollisia menetelmiä koskevan katsauksen jälkeen keskityn katsomaan, kuinka paljon monimuuttujamenetelmiä on eri aikoina käytetty ja mitä menetelmät ovat olleet.

Suuren tietomäärän hallitsemiseksi olen tallentanut väitöskirjojen tiedot useaksi SPSS-aineistoksi. Laajin aineisto sisältää tiedot kaikista sosiologian väitöskirjoista koostuen väittelijän nimestä ja sukupuolesta, väitöskirjan tarkastusvuodesta ja yliopistosta sekä tiedot siitä, onko sosiologi käyttänyt tilastollisia menetelmiä ja esiintyykö väitöskirjassa monimuuttujamenetelmiä. Monimuuttujamenetelmiä sisältävistä väitöskirjoista olen tallentanut SPSS-aineistoon vapaamuotoisena tekstinä kussakin väitöskirjassa käytetyt tilastolliset menetelmät. Sen lisäksi tallensin numeerisina muuttujina (1=kyllä/0=ei) tiedot AID-analyysin, erotteluanalyysin, faktorianalyysin, korrespondenssianalyysin LISREL-analyysin, log-lineaarisen analyysin, logistisen regressioanalyysin, polkuanalyysin, regressioanalyysin, ryhmittelyanalyysin, transformaatioanalyysin ja varianssianalyysin esiintymisestä.

Toisessa tutkimuskysymyksessä rajaudun tarkastelemaan faktorianalyysia. Faktorianalyysi on ollut suosittu menetelmä suomalaisten sosiologiain keskuudessa. Sosiologit ovat itsekin pohtineet faktorianalyysin suosiota tieteenalallaan. Alastalo (2005, 105) toteaa kirjassaan seuraavasti: ”Vaikuttaa siltä, että faktorianalyysia on mahdotonta mainita suomalaisessa sosiologiassa ilman, että viitataan sen kohtuutomaan käyttöön ja sitä seuranneeseen kritiikkiin.” Alastalo (2005, 73) tuo esiin, miten faktorianalyysin suosioon suomalaisessa 1960-luvun sosiologiassa vaikutti ennen kaikkea Vahervuon ja Ahmavaaran (1958) kirja *Johdatus faktorianalyysiin*. Eskola (1971) näki faktorianalyysin suosion liittyvän historiallisiin kehityslinjoihin. Eskola linjaa, että yhdysvaltalainen positivismi behaviorismina tuli 1950-luvulla suomalaisen sosiaalitieteeseen, joka muutoin otti vaikutteita eurooppalaisesta tieteestä. Hänen mukaansa faktorianalyysissa oli mahdollista yhdistää yhdysvaltalainen ja eurooppalainen näkemys, koska se sisälsi behaviorismille tunnusomaista numeerisen aineiston tilastollisiin menetelmiin pohjautuvaa tarkastelua ja toisaalta myös eurooppalaiselle tutkimusperinteelle tunnusomaista intuitioon pohjautuvaa ajattelua. Eskolan mukaan Marxin, Weberin, Durkheimin ja muiden sosiologian klassikkojen luoma tulkinnallinen perinne löysi behaviorismin faktorianalyysissa (emt.).

Faktorianalyysin suosio 1960-luvulla johtunee osaksi faktorianalyysia suorittavien tietokoneohjelmien olemassaolosta (Harman 1976, 7; Leskinen 1987). Tie-

tokoneohjelmia oli olemassa myös varianssianalyysille ja regressioanalyysille, mutta erityisesti faktorianalyysin kehittäminen oli tuolloin voimakasta (Alastalo 2014).

Toinen syy siihen, miksi keskityn tutkimuksessani nimenomaan faktorianalyysiin, liittyy tutkimukseni tavoitteeseen avata sosiologisen kvantitatiivisen tutkimuksen ja tilastotieteen analyysitapojen samankaltaisuuksia ja eroja. Faktorianalyysi tarjoaa tähän hyvän mahdollisuuden.

Vastatakseni faktorianalyysia koskeviin tutkimuskysymyksiin olen muodostanut lähilukua hyväksikäyttäen väitöskirjoissa esitetyistä faktorianalyysia koskevista teksteistä useita tekstitiedostoja, jotka liittyvät faktorianalyysiperusteluihin sekä sosiologien tekemiin valintoihin analyysin suorittamisessa. Analysoin sosiologien esittämiä faktorianalyysiperusteluja sekä sisällönanalyysin että ATLAS.ti -ohjelman tarjoamia analyysimahdollisuuksia hyödyntäen.

Olen muodostanut faktorianalyysin perusteluista ja analyysiin liittyvistä valinnoista myös SPSS-aineiston, jonka muodostaminen mahdollistui sen jälkeen, kun olin ensin vaihe vaiheelta tutustunut tekstiaineistoihin. Faktorianalyysin perusteluja koskeviksi tiedoiksi olen tallentanut aineistoon kyllä/ei-muuttujiksi tiedot siitä, onko sosiologi viitannut perusteluissaan riippuvuuteen, rakenteeseen, tiivistämiseen, muuttujien vähentämiseen, analyysin yksinkertaistamiseen, faktorirakenteeseen, faktorimalliin, mittaamiseen, operationalisointiin, summamuuttujaan, jatkoanalyysiin, ulottuvuuksiin, ulottuvuuksien riippuvuuteen, latentiin muuttujaan, käsitteeseen tai teoriaan. Nämä ilmaukset valikoituivat sekä sisällönanalyysin ja faktorimallin pohjalta. Kuvailen perustelujen esiintymismääriä kvantitatiivisin menetelmin ja analysoin niitä rajakohdemallia hyödyntäen.

Esittelen rajakohdemallin luvussa 6.1. Mainitsen kuitenkin tässä vaiheessa lyhyesti, että rajakohdemalli sisältää itse rajakohteen sekä erilaisia kauttakulkupisteitä, joista rajakohdetta käyttävät osapuolet valitsevat sopivimman. Tutkimukseni rajakohdemallissa on kolme kauttakulkupistettä. Ensimmäinen kauttakulkupiste kuvastaa tilastomatemaattista ja muuttujien ominaisuuksiin perustuvaa näkökulmaa. Toinen pohjautuu faktorianalyysin asetelmaan ja kolmas sosiologiseen tutkimuskysymykseen. Tutkimukseni kannalta kiinnostavaa on tarkastella rajakohdemalliin liittyviä kääntämissä, joiden tarkoituksena on saada osapuolet vakuuttumaan tietyn kauttakulkupisteen paremmuudesta ja valitsemaan sen. Kyse on siis tieteellisestä auktoriteetista: kenen ja minkä tieteenalan näkemys painaa vaakakupissa eniten.

Faktorianalyysiin on sen syntymisestä asti liittynyt sisällöllisyyden ja matemaattisuuden kilvoittelu. Faktorianalyysin idea syntyi 1800-luvulla biologien keskuudessa heidän hahmotellessaan ajatusta latentin muuttujan olemassaolosta. Varsinainen faktorimalli kehittyi psykologian ja älykkyystutkimusten yhteydessä. Ensimmäisen yhden faktorin mallin esitti vuonna 1904 Charles Spearman, jonka älykkyystutkimukset johtivat myöhemmin niin sanottuun kahden faktorin teoriaan. Sen mukaan älykkyys jakaantuu yleisälykkyyteen ja erityiskykyihin. Yleisälykkyyttä kuvastaa yleisfaktori ja erityiskykyjä kuvastavat spesififaktorit.

(Bartholomew 1995b; Leskinen 1987; Harman 1976, 3–5.) Älykkyystutkimusten kehittyessä tutkijat totesivat Spearmanin kahden faktorin teorian riittämättömäksi älykkyiden rakenteen kattavaan kuvaamiseen. Tutkijat tiesivät, mitä he sisällöllisesti halusivat tavoitella, mutta heillä ei ollut välineitä sen toteuttamiseen. He kokeilivat kahden faktorin teorian korvaamista monen ryhmäfaktorin teorialla, mutta menetelmällisesti he joutuivat pysyttelemään yhden faktorin analyysissa. (Thurstone 1935.) Faktorianalyysin matematiikka kehittyi ratkaisevasti 1930-luvun alussa, kun matemaatikot ohjeistivat Louis Thurstonea hyödyntämään matriisilaskentaa. Se tarjosikin Thurstonelle keinon laajentaa yhden faktorin malli monen faktorin malliksi. (Thurstone 1935, ix.) Esittelen faktorianalyysiin liittyvää matriisilaskentaa hieman luvussa 6.2.

Thurstone (1947, v-xiv) kertoo vuonna 1947 julkaistun kirjansa esipuheessa, miten matemaattiset tilastotieteilijät eivät aluksi hyväksyneet faktorianalyysia tilastotieteelliseksi menetelmäksi. Tosin tilastotiede itsessäänkin haki vielä tuolloin paikkaansa. Faktorianalyysi esitettiin Thurstonen mukaan tilastotieteellisenä menetelmänä ensimmäisen kerran Holzingerin ja Harmanin kirjassa, joka on painettu vuonna 1941. Faktorianalyysin tilastotieteellistyminen oli Thurstonen mielestä hyvä tavoite vaikkakin hän ennakoி sen tulevan aiheuttamaan tutkimuksissa haasteita tai jopa ristiriitaisuuksia muiden tutkimuksellisten tavoitteiden kanssa. Eräs Thurstonen mainitsema haaste liittyy lineaarisuuteen. Lineaarisuuskeskusteluun voi tutustua esimerkiksi Byrnen (2002; 2004) ja Selkälän (2013, 117–119) tekstien kautta (ks. myös Toiviainen 1998, 314). Tilastotieteellisyyden ja tutkimuksellisuuden kamppailusta Thurstone (1947, v-xiv) kertoo esimerkin matemaattisen tilastotieteen edustajaan Harold Hotellingiin liittyen. Thurstonen mukaan Hotellingia harmitti se, että faktorianalyysin opiskelijat eivät hyödyntäneet matemaattisen tilastotieteen uusimpia edistysaskelaita eivätkä muotoilleet faktorianalyysissa käyttämiään käsitteitä uusimman tilastotieteellisen teorian mukaisesti. Thurstone kertoo vastanneensa Hotellingille, että faktorianalyysin tilastomatemattinen teoria oli toki tehnyt tärkeitä edistysaskelia edellisten kahden vuosikymmenen aikana, mutta kehitystä ei ollut tapahtunut tutkimuksellisten näkökulmien kannalta riittävästi.

Kolmannessa tutkimuskysymyksessä jatkan faktorianalyysiin liittyvää tarkastelua katsomalla, millä tavalla sosiologit ovat väitöskirjoissaan eri aikoina käyttäneet faktorianalyysia ja minkälaisia faktorianalyysin suorittamiseen liittyviä valintoja he ovat tehneet. Tarkastelen, minkälaisia muuttujavalintoja, faktorointivalintoja ja rotatointivalintoja eri aikoina on tehty ja miten faktorianalyysin tuloksena saatujen faktoreiden analyysia on jatkettu. Nämä tiedot on tallennettu SPSS-aineistoon ja analysoin niitä kvantitatiivisesti.

Teen sekä toisen että kolmannen tutkimuskysymyksen kohdalla myös ajallista vertailua katsoakseni, minkälaisia muutoksia perusteluissa ja faktorianalyysin käytötavoissa on tapahtunut vuosien 1960–2018 aikana. Muutoksen tarkastelussa on väistämättä huomioitava tietokoneiden vaikutus.

Tieteentutkimuksessa käytettävät aineistot ovat moninaisia erilaisten tieteenalataustojen ja tutkimuskohteen monitahoisuuden vuoksi (Kiikeri ja Ylikoski 2004, 10). Väitöskirjoista muodostamieni erilaisten aineistojen lisäksi tarkastelen suomalaisia ja kansainvälisiä eri aikoina julkaistuja menetelmäoppikirjoja sekä metodologisia – erityisesti suomalaisessa sosiologiassa – käytyjä keskusteluja. En kutsu niitä tutkimukseni aineistoksi, koska en tee niiden osalta systemaattista tarkastelua vaan käytän niitä ennemminkin eri aikoina vallitsevien näkökulmien peilaamiseen. Tutkimuksessani on piirteitä kirjallisuustieteistä, bibliometriikasta, meta-analyysistä, etnografiasta ja tilastollisesta analyysistä. Kirjallisuustieteitä sivuan siinä, että tutkimukseni kohteena on tieteellinen kirjallisuus, jota tarkastelen teoreettista ja historiallista tietoa hyväksi käyttäen (ks. esim. Lakomäki, Latvala ja Laurén 2011). Bibliometriikka on tieteellisen viestinnän painettujen julkaisujen määrällistä tutkimusta, jonka tavoitteena on analysoida ja mallintaa tieteen kehitystä (Forsman 2016; Miettunen 2004, 17). Tästä lähtökohdasta tutkimukseni on kytköksissä bibliometriikkaan. Erotun bibliometrisestä tutkimuksesta kuitenkin siinä, että en tutki julkaisujen saamaa suosiota. Analyysini muistuttaa joiltain osin myös meta-analyysia, joka tarkoittaa aiempien tutkimusten pohjalta tehtyä synteesiä (Glass, McGaw ja Smith 1987, 21; Cook ym. 1992). Yhteneväisiä piirteitä meta-analyysin kanssa on se, että tutkin toisten tekemiä tutkimuksia, teen meta-analyysille tyypillisen kirjallisuuskatsauksen ja käytän tilastollisia menetelmiä (Wolf 1986, 5). Tutkimustani ei kuitenkaan voi kutsua metatutkimukseksi, koska en tee synteesiä aiemmista tutkimustuloksista enkä tutkimusten sisällöistä vaan teen omassa tutkimuksessani synteesiä käytetyistä tilastollisista tutkimusmenetelmistä. Haluan yhteiskuntatieteelliseen opetus- ja tutkimusympäristöön siirtyneenä tilastotieteen edustajana kuvata ja selittää sosiologian väitöskirjoissa kuvattua toimintaa analysoimalla erityisesti faktorianalyysiin liittyviä argumentteja. Tutkimukseni lukeutuu tieteen sosiologiaan, mutta analyysitavoissani ja tulkinnoissani näkyy myös tilastotieteellinen ajattelutapani.

4. Sosiologian ja tilastotieteen kohtaamisen haasteet kvantitatiivisessa tutkimuksessa

4.1. Tutkimusprosessi nähdään eri tavalla

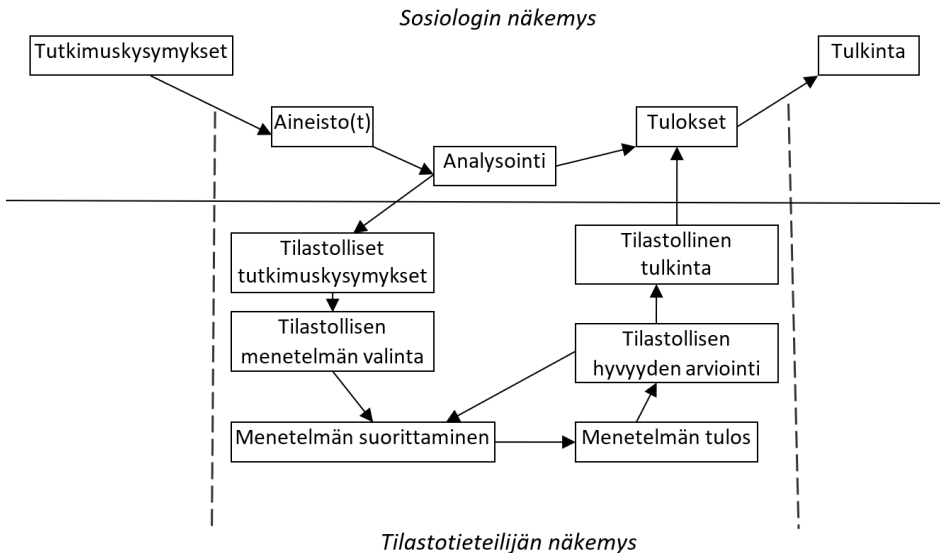
Nicholas Babchuk, Bruce Keith ja George Peters (1999) kuten myös Laura Hunter ja Erin Leahey (2008) ovat tutkineet sosiologian yhteistyötä muiden tieteenalojen kanssa. He toteavat yhteistyön olevan yhä yleisempää. Hunterin ja Leaheyn mukaan yhteistyö on nykyään jo normatiivista ja sitä tapahtuu erityisesti kvantitatiivisissa tutkimuksissa.

Ivo Molenaar (1992) tuo esiin yhteistyön haasteet toteamalla tilastotieteen käytön niin sanotuissa pehmeissä tieteissä olevan mielenkiintoista, koska se on niin vaikeaa. Hänen mukaan haasteet aiheutuvat tutkimuksen metodologisista ratkaisuksista, jotka liittyvät esimerkiksi empiirisen tutkimuksen suunnitteluun, mittaamiseen, mallintamiseen ja analyysiin.

Seuraavaksi esittelen sosiologisen kvantitatiivisen tutkimusprosessin (kuva 3) kuvatakseni sosiologian ja tilastotieteen yhteistyötä ja siinä esiintyviä haasteita. Nimityksellä kvantitatiivinen tutkimusprosessi tarkoitan sosiologian ja tilastotieteen kohtaamista, joka voi olla konkreettista, henkilöiden välillä tapahtuvaa tutkimuksellista yhteistyötä tai se voi olla kuvaannollista yhteistyötä, jolloin sosiologi tukeutuu menetelmäoppikirjojen ja menetelmäopetuksen tarjoamiin ohjeisiin ja opastuksiin tilastollisen analyysin suorittamiseksi. Alastalo (2005, 14–17) tuo esiin numeroaineistoja käyttävästä sosiologisesta tutkimuksesta käytettäviä monenlaisia nimityksiä. Yleisesti ottaen kvantitatiivisuudella voidaan viitata paitsi koko tutkimukseen, myös pelkkään aineistoon tai menetelmään. Esimerkiksi mixed methods -tutkimuksissa osa aineistoista on kvantitatiivisia, joita analysoidaan kvantitatiivisin – tilastollisin – menetelmin, mutta koko tutkimusta ei voida määritellä kvantitatiiviseksi. Mäkelä (1996) erottelee kvantitatiivisen tutkimuksen tilastollisesta tutkimuksesta sen perusteella, miten tutkija käyttää tilastollisia menetelmiä (ks. myös Alastalo 2005, 15–16). Jos tutkija painottaa analyysissään teknisiä ja tilastomatemattisia kriteerejä, kutsuu Mäkelä tätä tilastolliseksi tutkimukseksi. Kvantitatiivisella tutkimuksella hän tarkoittaa tutkimusta, jossa tutkija tilastollisia menetelmiä käyttäessään on kiinnostunut empirian ja teorian suhteesta.

Kuvan 3 tavoitteena ei ole määritellä tutkimusprosessia tai sen vaiheita, koska tutkimusprosessin kattava kuvaaminen ja määrittely on mahdotonta johtuen tieteen sisällä ja tieteiden välillä vallitsevista erilaisista käsityksistä tutkimuksen ja sen tavoitteiden asettamisen logiikan suhteen (Kiikeri ja Ylikoski 2004, 66; Blaxter, Hughes ja Tight 1996, 8–10). Joop Hox, Edith de Leeuw ja Don Dillman (2008,

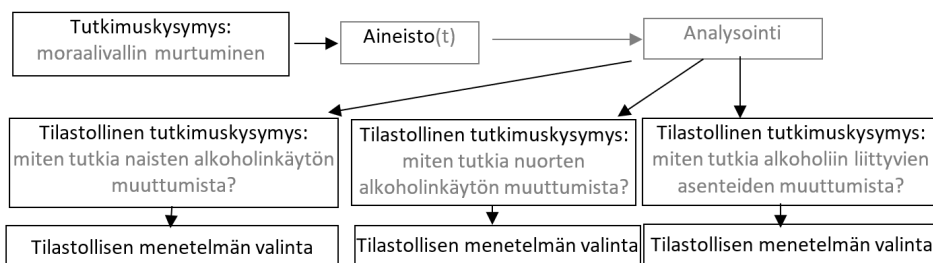
1–2) huomauttavat, että tutkimuksen suorittamisen tavat ovat jatkuvasti monipuolistuneet niiden ollessa tällä hetkellä monipuolisempia kuin koskaan aikaisemmin. Määrittelyn sijaan pyrin kvantitatiivista tutkimusprosessia kuvaavan esityksen avulla havainnollistamaan sosiologian ja tilastotieteen yhteistyötä ja siinä kohdattavia haasteita.



Kuva 3 Sosiologin ja tilastotieteilijän näkemykset tutkimusprosessiin. Lähde: Silén (2008)

Sosiologin näkökulmasta tutkimusprosessiin sisältyvät muun muassa tutkimuskysymysten muotoilu, sopivan aineiston tai aineistojen hankinta, analysointi ja sen tuottamien tulosten tulkinta (ks. esim. Bryman 1988, 20; Bryman ja Cramer 1990, 2; Alkula, Pöntinen ja Ylöstalo 1994, 24–25; Räsänen ym. 2005, 10). Tilastotieteilijä puolestaan hahmottaa prosessista aineiston analyysivaiheen erilaisine menetelmävalintoineen, menetelmäkokeiluineen, niiden tuottamine tuloksineen ja tulkintoineen (ks. esim. Hair ym. 1998, 94; 101; 160; 177; 340; 350; 480; 492).

Tutkimusprosessin hahmottamisen näkökulmasta on kaksi vaihetta, joiden kohdalla kvantitatiivista analyysia tekevän sosiologin ja tilastotieteen pohjalta kvantitatiivista aineistoa käsittelevän henkilön näkemykset eroavat toisistaan. Ensimmäinen niistä liittyy tutkimuskysymyksiin. Sosiologin tutkimuskysymykset ovat eri kysymyksiä, kuin tilastotieteilijän tutkimuskysymykset (ks. esim. Hand 1994). Kuvassa 4 avataan Tarja Orjasniemen (2005) väitöstutkimukseen liittyvän esimerkin kautta yhteiskuntatieteilijän ja tilastotieteilijän erilaisia tutkimuskysymyksiä.



Kuva 4 Sosiologin ja tilastotieteilijän tutkimuskysymykset eroavat toisistaan

Sosiologiset tutkimuskysymykset syntyvät kiinnostuksesta tarkastella yhteiskunnallisia ilmiöitä. Sosiologin näkökulmasta tutkimuksen aiheella on Pöntisen (2004, 72; 2011) sanoin oltava edes jonkinlaista yhteiskunnallista merkitystä ja sen on edes jossain määrin lisättävä ymmärrystä yhteiskunnasta. Aiheella on oltava myös teoreettista relevanssia, joka voi liittyä itse tutkimuskysymykseen tai tutkimuksen käsitteisiin. Siten sosiologin tutkimuskysymykset kohdistuvat tutkimuksen kohteena olevaan ilmiöön. (Pöntinen 2004, 72.) Tilastotieteilijän taas on hahmoteltava tutkimuskysymykset olemassa olevan numeerisen tutkimusaineiston perusteella. Tilastolliset tutkimuskysymykset perustuvat paljolti aineistossa olevien muuttujien tilastollisiin ja numeerisiin ominaisuuksiin, esimerkiksi muuttujien mittaustasoihin – ja toki jossain määrin myös muuttujien sisältöön ja oletuksiin niiden välisistä suhteista (ks. esim. Hair ym. 1998, 94; 160; 255; 446; 480). Tilastotieteessä tätä vaihetta ei edes nimetä tutkimuskysymysten luomiseksi vaan aineistoon tutustumiseksi, jonka tavoitteena on löytää perusteluja erilaisten tilastollisten menetelmien valintaan. Kuten Richard McElreath (2016, 3) ja Tapio Nummenmaa (1980) toteavat, kykenevät tilastolliset menetelmät analysoimaan ilmiötä vain tietyltä kannalta, mistä seuraa sosiologisen ja tilastollisen tutkimuskysymyksen eroaminen toisistaan myös siinä, että yhtä sosiologin tutkimuskysymystä varten on luotava useita tilastollisia tutkimuskysymyksiä ja käytettävä sen myötä monia tilastollisia analyysimenetelmiä.

Sosiologi ja tilastotieteilijä eivät välttämättä tiedosta tutkimuskysymystensä erilaisuutta ja niiden liittymistä eri asioihin. Vuonna 2008 julkaistussa artikkelissani (Silén 2008) analysoin omia kokemuksiani tilanteesta, jossa tilastotieteilijänä yritin etsiä sopivaa tilastollista menetelmää yhteiskuntatieteilijän esittämään tutkimuskysymykseen. Alkoholitutkimusta tekevä yhteiskuntatieteilijä halusi tietää, minkälaisella tilastollisella menetelmällä hän voisi analysoida moraalivallin murtumista (ks. Orjasniemi 2005). Tilastollisten menetelmien löytyminen mahdollistui vasta, kun ymmärsin yhteiskuntatieteilijän tutkimuskysymyksen olevan käsitteellinen. Yhteiskuntatieteilijän tutkimuskysymys oli pilkkottava ja sen myötä oli vasta mahdollista löytää kytkös aineistoon ja mahdollisiin menetelmiin. Moraalivallia ilmensivät tutkijan mukaan alkoholin käytön yleisyys, naisten alkoholin käyttö, nuorten alkoholi-

lin käyttö sekä alkoholipoliittiset mielipiteet ja asenteet. Moraalivallin murtumista tutkittaessa olisi siten analysoitava pitkittäisaineistosta, millä tavalla naisten alkoholin käyttö on muuttunut, millä tavalla nuorten alkoholin käyttö on muuttunut ja miten asenteet ovat muuttuneet. Näistä muodostui tilastollisia tutkimuskysymyksiä, jolle kullekin oli löydettävä sopivat tilastolliset menetelmät. Sosiologiseen tutkimuskysymykseen vastaaminen tilastollista analyysia hyödyntäen vaatii siis sekä tutkimuskysymyksen pilkkomista osa-alueisiin, että kuhunkin osa-alueeseen – tilastolliseen tutkimuskysymykseen – sopivan tilastollisen menetelmän valitsemista, suorittamista ja tulkintaa.

Osapuolten erilaisista näkökulmista asetettujen tutkimuskysymysten ymmärtäminen vaatii kommunikointia osapuolten välille. Kun erilaisuus tunnistetaan voivat osapuolet nivoa tutkimuskysymyksensä toisiaan täydentäviksi, jolloin ne ikään kuin kartoittavat samaa asiaa erilaisin keinoin. Siltikin tilastotieteilijän puhtaalla tilastotieteellisellä ajattelulla ja sosiologin kysymisen tavalla sekä analysoitavaksi taivutella aineistolla on kohtaamisalueiden etsintä edessä.

4.2. Aineistoon kohdistuvat haasteet

Aineistonkeruu ja aineiston laatu asettavat sekä tilastotieteilijälle että sosiologille haasteita, mutta haasteet ovat heille erilaisia. Tilastotieteellisestä näkökulmasta kiinnostavia ovat aineistossa olevien muuttujien numeeriset arvot ja se, minkälaisiin menetelmiin kyseiset arvot taipuvat. Sosiologisesta näkökulmasta taas aineistossa kiinnostavat sen viittaukset jonkin ilmiön olemassaoloon ja ilmiön piirteisiin (Kii-keri ja Ylikoski 2004, 52). Siten se, mitä analyysin tekemiseksi aineistolta kysytään sekä se, millaisella tavalla aineiston rakennetta hahmotetaan, eroaa monin tavoin.

Sosiologit keräävät tutkimusaineistojaan useilla eri tavoilla (ks. Alastalo 2005, 55–116). Aineisto(t) saatetaan varta vasten hankkia tutkimustarkoitukseen, mutta sosiologit käyttävät myös aineistoja, jotka ovat syntyneet spontaanisti yhteiskunnassa tapahtuvassa toiminnassa (Pöntinen 2004, 80–82). Hyvän aineiston kerääminen vaatii ammattitaitoa, vaivannäköä ja rahaa (Ketokivi 2009, 9; Paananen, Juntto ja Sauli 1998, 183). NykYTEknologia on joiltakin osin helpottanut tutkimusaineistojen keräämistä (Erola ja Räsänen 2014a, 246), mutta aineistonkeruussa on silti haasteensa, joihin voi tutustua esimerkiksi Bethlehemin (2010) artikkelissa.

Aineiston hankkiminen joko itse keräämällä tai muulla tavoin on vaihe, jonka tärkeyttä tutkimustaan aloittavat eivät aina oivalla. Ehkä jopa ajatellaan, että tutkimus alkaa vasta, kun aineisto on käsillä. Aineiston hankkiminen on kuitenkin tärkeä ja peruuttamaton vaihe tutkimuksen suorittamisesta. Siinä tehtyjä virheitä ei voi jälkikäteen korjata, vaikka yrityksiä toki on (ks. esim. Biemer ja Sharon 2008, 317–341; Stapleton 2008, 342–369; Rässler, Rubin ja Schenker 2008, 370–386; Hox 2008,

387–402). Alastalo (2005, 74) toteaa, että sosiologit ovat jo varhain tunnistaneeet aineiston keruun ja mittaamisen tärkeyden. Alastalo viittaa esimerkkinä Vesa Laakkoseen⁴ joka jo vuonna 1961 totesi, että tutkimuksen tuloksia ei saa luotettavammiksi teknisillä ja matemaattisilla keinoilla, jos aineiston keruu on tehty huolimattomasti.

Sosiologiassa tyypillinen kvantitatiivisen tutkimuksen aineisto on pitkään kerätty kysymyslomakkeiden kautta. Aineiston keruuseen liittyvien kysymysten ja kysymyslomakkeen suunnittelu ovat aikaa vieviä ja vaikeita harkintoja (ks. esim. Selkälä 2013, 104–140; Gehlbach 2015; Alwin ja Beattie 2016). Sosiologin näkökulmasta on olennaista saada kerättyä tutkittavaa ilmiötä koskevaa tietoa. Näin sosiologi joutuu pohtimaan sosiologisten ilmiöiden mitattavuutta ylipäätään (ks. esim. Ervasti ja Saari 2011, 193). Onnistuneen kvantitatiivisen aineiston hankkiminen edellyttää, että keruuvälineenä toimiva kysymyslomake läpäisee sosiologin, tilastotieteilijän ja vastaajien kriteerit. Kysymysten on toimittava sekä sanallisesti että numeerisesti. Kysymysten ja vastausvaihtoehtojen sanoittamisessa joudutaan huolella arvioimaan niiden ymmärrettävyys, vastaajastävällisyys, konkreettisuus ja yksinkertaisuus ja osuvuus (ks. esim. Ronkainen 1998, 242–244). Sähköiset kyselyt ovat tuoneet omat haasteensa aineistonkeruuseen. Näihin voi tutusta esimerkiksi Mick P. Couperin (2017) artikkelissa.

Sosiologit tunnistavat, että mittaamisella tavoitetaan ehkä vain osa tutkimuskohdetta (Alkula ym. 1994, 20). Mittaaminen on askarruttanut sosiologeja kauan (ks. esim. Raiski 1974a). Seppo Paanasen, Anneli Junton ja Hannele Saulin (1998) Fakta juttu -teoksen alkusanoissa todetaankin, että: ”Kvantitatiivinen mittaaminen koetaan usein ongelmattomaksi maailman tosiasioiden mittaamiseksi. Ihmisten toiminta ei kuitenkaan ole palautettavissa mitattaviksi käyttäytymisen osaksi, vaan inhimillisessä toiminnassa ymmärtäminen on tärkeintä.” Mäkelä (1991) tuo esiin tilastollisen mittaamisen ja yhteiskuntatieteellisten, tai sosiologisten, käsitteiden eron. Hän kirjoittaa, että luonnontieteissä, esimerkiksi biologiassa, mittarit ja käsitteet voidaan samaistaa, mutta yhteiskuntatieteissä ei. Kun tilastotiedettä sovelletaan yhteiskuntatieteisiin, niin keskeistä on se, miten käsitteet ja empiria saadaan pelamaan toistensa kanssa. (Mäkelä 1991.) Myös tilastotieteen edustaja Lauri Tarkkonen (1997) ottaa kantaa mittaamisen ongelmaan seuraavasti:

Olen varsin vakuuttunut siitä, että useimpia ilmiöitä tutkittaessa olisi syytä ymmärtää niiden moniulotteisuus. Osa tilastotieteen huonosta maineesta johtuu siitä, että tilastolliset mallit yksinkertaistavat ilmiöitä liiaksi. Erityisharrastukseni on mittaamisen ongelma, joka on keskeinen kaikessa empiirisessä tutkimuksessa. Useimmat tilastotieteilijät sysäävät sen maton alle, koska sen kunnollinen käsitteleminen edellyttää myös tutkittavan ilmiön käsittei-

⁴ Laakkonen, Vesa (1961) Tilastollisista menetelmistä sosiaalitutkimuksessa. Teoksessa Heikki Waris ja 15 tohtoria. Sosiaalipoliittisen yhdistyksen julkaisuja 7. WSOY. Porvoo, Helsinki. 215–230.

den tuntemista ja ymmärtämistä. Tilastotieteilijät karttavat tätä ongelmaa, koska eivät halua perehtyä ongelman sisällöllisiin piirteisiin ja soveltajat käsittelevät sitä kovin naivisti, koska heillä ei ole aseita sen kunnolliseen käsitelyyn. (Tarkkonen 1997.)

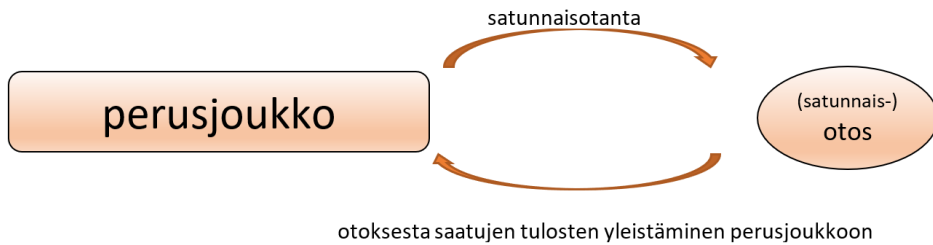
Menetelmäoppikirjat ohittavat usein mittaamiseen ja erityisesti käsitteisiin liittyvän mittaamisen haasteet. Niissä mittaaminen esitellään mittaustasojen ja mittauksen harhattomuuden näkökulmasta. Mitta-asteikkojen suhteen sosiologien tiedonkeruussa tapahtuva mittaaminen on useimmiten luokittelu- tai järjestysasteikollista. Sosiologit voivat keinotekoisesti pyrkiä korkeampiin mittaustasoihin lisäämällä kysymyksiin enemmän systemaattisessa järjestyksessä olevia vastausvaihtoehtoja, jolloin mittaustaso lähenee välimatka- tai suhdeasteikkoa. Kysymyslomakkeen vastaajan kannalta tällaisesta lomakkeesta tulisi kuitenkin liian työläs vastata. Jos taas tehdään vastaajan kannalta miellyttävät kysymykset, ovat ne usein tilastotieteellisesti vaatimattomia tai mahdottomia analysoida. Tapani Alkula, Seppo Pöntinen ja Pekka Ylöstalo (1994, 12) toteavatkin, että sosiologisten ilmiöiden mittaamisessa joudutaan tekemään kompromisseja mittauksen tarkkuus- ja systemaattisuusvaatimusten ja arkikielen monimerkityksisyyden välillä.

Kun kysymyslomakkeella saadut vastaukset tallennetaan numeeriseksi aineistoksi, voi tutkija itse päättää, mitkä numerot hän valitsee edustamaan kutakin kysymyslomakkeessa olevaa vastausvaihtoehtoa. Erityisesti tämä valinnanvapaus koskee luokitteluasteikolla kysytyjen tietojen tallentamista. Järjestysasteikon mittaamisessa vastausvaihtoehtoja edustavat numerot voidaan muutoin valita vapaasti, kunhan ne noudattavat samaa yksikäsitteistä järjestystä kuin vastausvaihtoehtojen sisältö. Tästä seuraa kvantitatiiviseen sosiologisen tutkimuksen suuri sudenkuopan vaara. On vaarana, että pian muuntamisen jälkeen unohdetaan numeroihin liittyvä keinotekoisuus ja niihin aletaan suhtautua, kuten ne olisivat tarkkoja mittaustuloksia, mitä ne eivät kuitenkaan ole. Tutkija myös helposti unohtaa kaikki mittaamiseen liittyneet epävarmuudet ja alkaa suhtautua aineistoonsa faktatietona (Alastalo 2005, 13; Kiikeri ja Ylikoski 2004, 30; Sund 2000; Porter 1995; Blumer 1956). Liisa Rantalaiho (1998, 20) vertaa osuvasti tutkijan suhtautumista aineistoonsa H.C. Andersenin kuuluisaan satuun Keisari ilman vaatteita.

Sosiologit joutuvat kohtaamaan mittaamisen haasteet myös silloin, kun eivät kerää tutkimusaineistoaan itse vaan saavat käyttöönsä valmiin, mutta eri tarkoitukseen kerätyn aineiston tai esimerkiksi käyttäessään rekisteriaineistoa (ks. esim. Sund 2008, 213; Sund, Nyländer ja Palonen 2004).

Tilastotieteilijän tutkimusprosessin alkaa usein siinä vaiheessa, kun aineisto on jo kerätty tai muutoin olemassa. Tämä lähtökohta näkyy monissa tilastotiedettä esittelevissä oppikirjoissa, joissa yleensä mainitaan kohtalaisen lyhyesti ja yksinkertaisesti muutamia aineiston hankkimisen periaatteita liittyen lähinnä mittaamiseen ja havaintoaineiston rakenteeseen (esim. Vasama ja Vartia 1980, 34–70; Helenius

1992, 14–15). Aineistonkeruun suunnittelu on alue, joka sosiologiassa ja yleisesti yhteiskuntatieteissä on tarkkaan oppikirjoissa käsitelty (ks. esim. Belson⁵ 1981; Oppenheim 1983; Hague 1993; Fowler⁶ 1995; Saris ja Gallhofer 2014). Suomenkielisessä kirjallisuudessa aiheetta on sivuttu muun muassa Sariolan (1956), Eskolan (1962; 1967), Erkki Jyringin (1974) sekä Tapani Alkulan, Seppo Pöntisen ja Pekka Ylöstalon (1994) teoksissa. Verkkomateriaalia aiheesta löytyy myös esimerkiksi Menetelmäopetuksen tietovaranto-sivustolta (MOTV 2018). Suomessa kysymyslomakkeisiin liittyvää tutkimusta ovat tehneet muun muassa Anja Ahola (1993; 1998; 2000a; 2000b), Alastalo (2005), Ronkainen ym. (2008) sekä Selkälä (2013, 103–128).



Kuva 5 Perusjoukko-otos-asetelma

Tilastotieteen näkökulmasta aineistonkeruu saa toisenlaisen roolin. Siihen liittyy keskeisesti perusjoukko-otos-asetelma (kuva 5). Silloin, kun perusjoukkoa ei voida tutkia kokonaisuudessaan, on siitä poimittava pienempi osajoukko (Bethlehem 2010, 162; Tarkkonen ja Vehkalahti 2005, 172–173). Tapoja, joilla pienempi osajoukko poimitaan, ovat kirjoissaan ohjeistaneet tilastotieteellisestä näkökulmasta muun muassa Erkki Pahkinen ja Risto Lehtonen (1989) sekä Pahkinen (2012). Yhteiskuntatieteellisen tutkimuksen näkökulmasta otantaa on käsitelty muun muassa David A. De Vausin (ks. esim. 2002) teoksissa.

Osajoukon toivotaan edustavan perusjoukkoa riittävän hyvin, parhaassa tapauksessa osajoukko olisi ikään kuin perusjoukko pienoiskoossa (Nummenmaa 2004, 21). Edustavuustavoitteen saavuttamiseksi osajoukon poimintaan käytetään otantamenetelmiä, joita ovat yksinkertainen satunnaisotanta, systemaattinen otanta, ositettu otanta ja ryväotanta. (Manninen 1974, 103–112.) Otoksen riittävyys tutkimusaineistoksi perustuu yleistettävyyssajatteluun, jonka kehittymisestä kirjoittavat muun muassa D. R. Bellhouse (2000), J. N. K. Rao (2006) sekä Rao ja W.A. Fuller (2017). Otoksesta saatujen tulosten yleistettävyyttä tilastotieteellisin keinoin voidaan arvioida tilastollisten testien avulla.

5 Belson, William A. (1981) *The design and understanding of survey questions*. Gower. Aldershot.

6 Fowler, Floyd J. Jr (1995) *Improving survey questions. Design and evaluation*. Applied social research methods series 38. SAGE Publ. Thousand Oaks.

Otannan merkitys kvantitatiivisessa sosiologisessa tutkimuksessa on kuitenkin tilastotieteen kuvausta problemaattisempi. Sosiologisessa tutkimuksessa perusjoukko-otos-asetelman toteuttaminen ei ole helppoa eikä se aina onnistu. Mäkelä (1991, 44–45; 125) on pohtinut asetelman käyttökelpoisuutta sosiaalitieteellisissä tutkimuksissa. Mäkelän mukaan vain harvat sosiaalitieteilijöiden keräämät aineistot voidaan tulkita otokseksi. Useimmat aineistot ovat hänen mukaansa näytteitä, jotka ovat ongelmallisia, koska niiden keräämisessä on voinut tapahtua esimerkiksi vastaajien valikoitumista. Valikoitumisen seurauksena tietynlaiset vastaajat ovat aineistossa yliedustettuina tai aliedustettuina. Monet aineistot muistuttavat Mäkelän mukaan myös Chicagon koulukunnan tapaustutkimusten aineistoja, joissa suhteellisen väljästi rajoitetusta joukosta tai alueesta kerätään mahdollisimman monipuolisesti tietoa. Mäkelän mukaan myös perusjoukon ja otoksen käsitteiden kautta maailma ikään kuin vakioidaan, jolloin muutosprosessien olemassaolo ainakin periaatteessa kielletään.

Tilastotieteen näkökulmasta perusjoukko määrittyy tutkimuksen aiheeseen liittyviksi kaikkien mahdollisten tilastoyksiköiden tai satunnaismuuttujan arvojen joukoksi (ks. esim. Manninen 1974, 104; Nummenmaa ym. 1997, 35). Sosiologiassa ja yleisemminkin yhteiskuntatieteissä perusjoukon määrittely on ongelmallisempaa (ks. esim. Rantalaiho 1972). Mäkelä (1991, 44–45; 125) huomauttaa, että tilastotieteilijät mieltävät perusjoukon muuttumattomaksi, kun taas sosiaalitieteelliset tutkimukset kohdistuvat rajoitettuihin, harkinnanvaraisesti valittuihin ja ajassa muuttuviin perusjoukkoihin (ks. esim. Stoop 2005, 27; Laaksonen 2013). Sosiaalitieteilijöiden tekemät yleistyksiset ovat Mäkelän mukaan aikaan ja paikkaan sidottuja. Orjasniemi (2005, 247–256) tuo tutkimuksessaan esille pohdintaansa tutkimuksensa perusjoukosta. Orjasniemi kirjoittaa, että vaikka hän lähtökohtaisesti määritteli tutkimuksensa suunnitteluvaiheessa perusjoukoksi yli 14-vuotiaat ranualaiset ja posiolaiset ja keräsi aineiston tämän mukaisesti, joutui hän pohtimaan vielä tutkimuksensa loppuvaiheessa sitä, mikä tutkimuksen todellinen perusjoukko on, toisin sanoen minkälaisilla alueilla ja minkälaisissa ihmisryhmissä hänen tutkimuksensa tulokset ovat päteviä.

Satunnaisotoksen poimiminen ja otantamenetelmien hyödyntäminen eivät ole sosiologian tutkimuksissa yksinkertaisia eivätkä aina edes mahdollisia toteuttaa. Otoksen poimiminen on mahdollista esimerkiksi rekistereitä hyödyntävissä tutkimuksissa silloin, kun rekisterit sisältävät perusjoukon määrittelyyn tarvittavat tiedot. Väestötietojärjestelmästä on mahdollista poimia vastaajat postiosoitteineen, mutta sähköisten kyselyjen osalta poiminta on haasteellista, koska väestötietojärjestelmät eivät automaattisesti sisällä vastaajien sähköpostiosoitteita. Bethlehem (2010, 184–185) pohtii, olisiko tällaisissa tilanteissa mahdollista poimia vastaajat joko postiosoitteen tai puhelinnumeron avulla ja lähettää heille kirjeitse tai puhelimitse pyyntö kirjautua tietyssä verkko-osoitteessa olevaan sähköiseen kyselyyn.

Aineiston edustavuutta on arvioitava monin tavoin (ks. esim. Bethlehem 2010; Cornesse ja Bosnjak 2018). Edustavuuden arvioimiseksi on olemassa tilastotieteelli-

siä kriteerejä, jotka perustuvat esimerkiksi aineiston kokoon (ks. esim. Nummenmaa 2004, 25–26) tai vastausprosenttiin, mutta ne eivät riitä kriteereiksi sosiologisen tutkimusaineiston edustavuuden arviointiin. Aineiston edustavuutta heikentävät monet kattavuuteen, otantaan, vastaamiseen ja mittaamiseen liittyvät asiat, joita esimerkiksi Hox ym. (2008, 1–17) ja Sharon L. Lohr (2008, 97–112) ovat tuoneet esille. Tötön (2012, 56) näkemyksen mukaan ”edustavuutta ylipäänsä” ei ole olemassa, sitä on olemassa vain suhteessa joihinkin ominaisuuksien tai muuttujiin. Mäkelä (1991, 125–127) on pohtinut edustavuuden määrittämistä sosiaalitieteellisissä tutkimuksissa. Hän pitää riittämättömänä sitä, että edustavuuden arviointi pelkistyisi vain taustamuuttujien vertaamiseen populaation vastaaviin jakaumiin. Hänen näkemyksensä mukaan populaatioon liittyvän edustavuuden sijaista pitäisi tarkastella tutkittavaan ilmiöön liittyvää edustavuutta.

Alhainen vastausprosentti ei automaattisesti tarkoita aineiston huonoa laatua (ks. esim. Cook, Heath ja Thompson 2000). Esimerkkinä tästä mainittakoon vuonna 2013 tapahtunut kolmen eri tavalla kerätyn aineiston tarkastelu (Silén ja Ronkainen 2013). Ensimmäinen aineisto kerättiin avoimena nettikyselynä, toinen tiettyjä valintoja noudattavin katuhaastatteluin ja kolmas otantaan perustuvalla postikyselyllä. Nettikyselyaineisto oli kooltaan suurin sisältäen 2073 vastaajaa. Katuhaastatteluaineistossa oli 1635 vastaajaa. Postikysely lähetettiin 2000 henkilölle, mutta vastausprosentiksi muodostui vain 26 %, minkä seurauksena postikyselyaineisto jäi pienimmäksi aineistoksi sisältäen vain 525 vastaajaa. Vastaajamääriä verrattaessa syntyy oletus, jonka mukaan eniten vastaajia sisältävä aineisto, tässä tapauksessa nettikyselyaineisto, olisi edustavin. Postikyselyaineisto puolestaan vaikuttaisi lähes käyttökelttomalta pienimmällä vastaajamäärällään ja alhaisella vastausprosentillaan. Aineistoja tarkemmin tarkasteltaessa, esimerkiksi taustamuuttujia vertailtaessa suomalaisen väestöjakaumaan, osoittautui, että postikyselyaineisto edusti parhaiten suomalaista väestörakennetta. Avoin nettikyselyaineisto puolestaan poikkesi eniten suomalaisesta väestörakenteesta, siinä olivat yliedustettuina nuoret korkeasti koulutetut naiset. Netissä avoimena oleva kysely aiheutti vastaajien valikoitumista johtuen osaksi aineistonkeruumenetelmästä, mutta paljolti myös kyselyn aiheesta. Kyselyssä kysyttiin vastaajien näkemyksiä seksuaaliseen väkivaltaan liittyen sekä heidän luottamustaan suomalaisen oikeuslaitokseen seksuaaliseen väkivaltaan liittyvien tuomioiden antamisessa. Eri tavoin kerätyt aineistot houkuttelivat erilaisia vastaajia ja erosivat toisistaan taustamuuttujajakaumien osalta. Aineistoja analysoitaessa kuitenkin yllätti, miten suuri vaikutus erilaisuudella oli tuloksiin. Aineistot tuottivat keskenään aivan erilaisia – melkein päinvastaisia – tuloksia. Pohdittavaksi jäi, mikä niistä olisi uskottavin.

Sosiologian väitöskirjoissa aineiston edustavuuteen on ottanut kantaa Tomi Kanainen (2007, 63) todeten seuraavasti: ”Yleisesti ottaen sosiaalitutkimuksia koskee valikoitumisen suhteen se ongelma, että sellaiset, joilla on elämässään vaikeuksia, ovat tyypillisesti aliedustettuina otoksissa”. Mainintoja aineiston edustavuudesta

löytyy myös muun muassa Johanna Lammi-Taskulan (2007), Sanna Sanaksenahon (2006), Jenni Blomgrenin (2005) Sakari Karvosen (1997), Tapani Köpän (1979), Seppo Toivaisen (1970) ja Esko Kalimon (1969) väitöskirjoista.

Pöntinen (2011) neuvoo suhtautumaan maltillisesti sosiologisen kvantitatiivisen aineiston edustavuusvaatimuksiin. Hänen mukaansa aineiston edustavuus on hyvä asia, mutta ei kuitenkaan välttämättömyys eikä aina mahdollistakaan. Edustavalla aineistolla sosiologi pystyy varmemmalla äänellä puhumaan asioiden yleisyydestä, niiden välisistä yhteyksistä ja jopa vaikutussuhteista. Ei-edustavalla aineistolla niistä voi myös puhua, mutta paljon suuremmalla epävarmuudella. (Pöntinen 2011.)

Aineiston edustavuutta voidaan yrittää parantaa painotuksella, jossa aliedustetuna olevaa vastaajaryhmää voimistetaan painottamalla sitä ykköistä suuremmalla kertoimella ja yliedustetun ryhmän vaikutusta puolestaan vaimennetaan ykköistä pienemmällä painokertoimella. Tilastomatemaattisesta näkökulmasta painojen käyttö on perusteltua, koska niiden avulla saadaan korjattua otoksen ryhmäkoot ja mahdollisesti myös painottamattoman aineiston harhaiset estimaatit (ks. esim. Biemer ym. 2008, 317–341). Biemerin ym. mukaan painokertoimet voidaan tulkita edustuskertoimiksi, jotka kertovat, kuinka montaa perusjoukon jäsentä vastaa yksi painotettu otoksen jäsen. Esimerkiksi painokerroin 100 tarkoittaa, että vastaaja edustaa itsensä lisäksi 99 muuta henkilöä perusjoukossa. Painotuksella siis monistetaan yhden vastaajan tai aliedustetun ryhmän tiedot painokertoimen ilmoittaman määrän verran.

Painottamisessa tapahtuu oletus, jonka mukaan otoksessa mukana olevat vastaajat ovat mitatuilta ominaisuuksiltaan samankaltaisia kuin otoksesta syystä tai toisesta pois jääneet henkilöt. Sosiologisen tutkimuksen näkökulmasta aineiston painottaminen, siis esimerkiksi tietyn ikäisen tai tiettyä sukupuolta olevan vastaajan monistaminen, herättää kysymyksen monistamisen realistisuudesta. Tilastokeskus (2018) toteaa sivuillaan lyhyesti painotuksen jättävän huomiotta sen, että painotuksen lähteenä oleva vastaaja(ryhmä) ei välttämättä todellisuudessa vastaa painotuksen kohteena tuotettua ryhmää. Pia Mäkelä ja Janne Härkönen (2017; ks. myös Mäkelä 1999) tuovat saman asian esiin suomalaisten alkoholin kulutusta koskevassa artikkelissaan, jossa he toteavat, että ”Painotetuissa taulukoissa kunkin osaryhmän osuus vastaa ryhmän todellista osuutta kohdeväestössä. Painojen käyttö ei kuitenkaan korjaa sitä, jos esimerkiksi tutkimuksiin osallistuvat nuoret miehet käyttävät alkoholia eri tavalla kuin ne nuoret miehet, jotka eivät osallistuneet tutkimuksiin.”

Hoxin ym. (2008, 8) mukaan erääksi virhetyypiksi aineistonkeruussa katsotaan niin sanottu ylipeitto. Ylipeitto tarkoittaa, että sama vastaaja tulee syystä tai toisesta useaan kertaan otokseen. Hieman askarruttaa se, miksi ylipeitto katsotaan virheeksi ja painottaminen korjaavaksi toimenpiteeksi, kun molemmissa on kuitenkin kyse siitä, että yhden vastaajan tiedot ovat aineistossa useaan kertaan. Oletettavasti sopi-

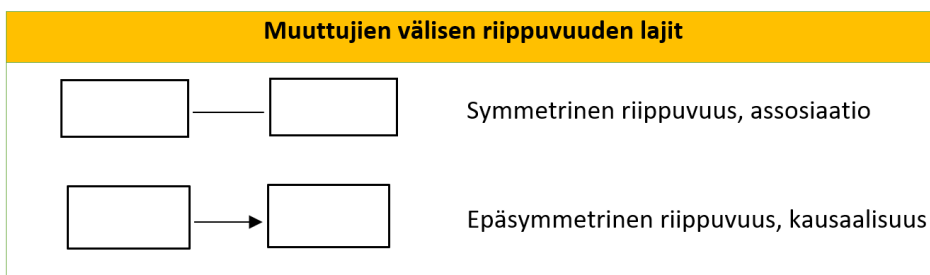
valla tavalla ja huolellisesti harkitulla painottamisella lopulta päästään lähimmäksi edustavaksi katsottua aineistoa. Tosin Bethlehem (2010) toteaa, että ei ole takuita siitä, että painottamisen avulla aineisto saataisiin edustavaksi.

4.3. Analyysivaiheen valinnat

Kuva 4 esitti, miten sosiologinen tutkimuskysymys on pilkottava useaan tilastolliseen tutkimuskysymykseen ja jokaiseen tilastolliseen tutkimuskysymykseen on löydettävä sopiva tilastollinen analyysimenetelmä. Sopivien tilastollisten analyysimenetelmien valintaa sosiologiseen tutkimuskysymykseen voisi kuvata kahden monimutkaisen ja laaja-alaisen asian yhteensovittamiseksi. Sosiologiset tutkimuskysymykset ovat laaja-alaisia ja tilastollisten menetelmien valintakriteerit moninaisia. Tutkimusmenetelmän valintaa ei kannata perustaa liian yksipuoliseen harkintaan. Loraine Blaxterin, Christina Hughesin ja Malcolm Tightin (1996, 73) mukaan tutkimusmenetelmän valinta on järkevää arvioida sen mukaan, mihin tutkijan taidot riittävät. Lisäksi on arvioitava, minkälaista tietoa tutkija haluaa saada, sopiiko tutkimuskysymykseen erityisesti tietty menetelmä, vaikuttavatko menetelmät saatiin vastauksiin ja miten tutkija itse vaikuttaa tutkimukseen (emt., 73). Reijo Sund (2000) suosittelee valitsemaan ja käyttämään niin yksinkertaisia menetelmiä kuin mahdollista, mutta ei yhtään yksinkertaisempia, kuitenkin sellaisia, jotka antavat aineistosta vastauksia haluttuihin kysymyksiin. Sund korostaa metodisen osaamisen lisäksi tarvittavan välttämättä myös substanssitetoutta, jotta pystytään arvioimaan ongelmanasettelun ja tulosten tarkoituksenmukaisuutta.

Kuvailevat menetelmät

Kuvailevat tilastolliset menetelmät ovat keskeisiä aineistoon tutustuttaessa. Yksinkertaisimmillaan tutkimusta tekevä sosiologi haluaa kuvailevia menetelmiä käyttäen saada selville, miten aineiston tiedot jakautuvat tai mikä on aineistossa tyypillistä (Alkula ym. 1994, 282). Useinkaan näin yksinkertainen aineiston kuvailu ei kuitenkaan ole sosiologisessa tutkimuksessa riittävää vaan sosiologi on kiinnostunut tutkimaan asioiden välisiä yhteyksiä. Tilastotieteessä puhutaan tällöin muuttujien välisten riippuvuuksien tarkastelusta. Riippuvuustarkasteluissa sosiologin on tärkeä erottaa toisistaan muuttujien välinen riippuvuus ja muuttujien väliset vaikutussuhteet (Alkula ym. 1994, 282). Tilastotieteessä muuttujien välinen riippuvuus voi olla epäsymmetristä tai symmetristä (kuva 6). Epäsymmetriset menetelmät käsittelevät riippuvuutta aina ikään kuin vaikutuksina, kausaalimaisena rakenteena. Tutkimusta tekevä sosiologin ei tarvitse kuitenkaan tulkinnassaan sitoutua kausaalisuuteen vaan sosiologisen tutkimuksen kannalta kyse on ryhmien välisistä eroista (Alkula ym. 1994, 282–283).



Kuva 6 Muuttujien välisen riippuvuuden lajit

Ristiintaulukoinnilla tehtävän riippuvuustarkastelun yhteydessä tilastotieteilijä nimeää muuttujat riippumattomiksi ja riippuviksi muuttujiksi, tai selittäjäksi ja selitettäväksi tai jopa syy- ja seurausmuuttujaksi (kuva 7). Tilastotieteilijälle sanavallinat eivät tarkoita sisällöllistä kannanottoa vaan ne ilmentävät muuttujan rakenteellista roolia. Tekniset termit eivät palvele sosiologista tarkastelua vaan voivat jopa aiheuttaa sekaannusta, mikä käy ilmi muun muassa Allardtin (1969) toteamuksessa siitä, että yhteiskuntatieteissä puhutaan usein selittämisestä, vaikka kyse on tilastollisesta riippuvuussuhteesta. Myös Jacques Tacq (1997, 31) korostaa menetelmien teknisen ilmaisujen välttämistä empiiristen tutkimuskysymysten yhteydessä. Yhteiskuntatieteelliseen tutkimukseen sopii hahmotustapa, jossa ristiintaulukoinnin avulla vertaillaan erilaisten vastaajaryhmien vastauksia, toisin sanoen eroavatko eri vastaajaryhmien vastaukset toisistaan. Ristiintaulukoinnilla voidaan myös selvittää, mitkä asiat ovat yhteydessä tarkastelun kohteena olevaan asiaan.



Kuva 7 Tilastotieteen ja sosiologian erilaiset ilmaisut riippuvuustarkasteluissa

Ristiintaulukointi on ollut keskeinen menetelmä sosiologisissa tutkimuksissa (Mäkelä 1996, 209). 1950- ja 1960-luvuilla ristiintaulukoinnin suosion syyt olivat käytännön sanelemia. Taulukointi oli ennen tietokoneiden käyttöönottoa

lähes ainoa analysointimahdollisuus (Alkula ym. 1994, 175; myös Alastalo 2005, 252–253). Vaikka nykyään menetelmävalikoima ja analysointimahdollisuudet ovat laajat, on ristiintaulukointi edelleenkin käyttökelpoinen ja tarpeellinen menetelmä aineiston analyysissä. Ristiintaulukointi sopii kaikenlaisen tiedon tarkasteluun. Taulukko esittää havainnollisesti tarkasteltavien asioiden välisen yhteyden (ks. esim. Alkula ym. 1994, 175). Ristiintaulukoinnilla voidaan toteuttaa myös Lazarsfeldin kehittämää elaboraatiota, joka on koettu hyödylliseksi sosiaalitieteellisessä tutkimuksessa ja johon myös monet kehittyneemmät menetelmät pohjautuvat (Eskola 1994, 27–28). Yksinkertaisuudestaan ja konkreettisuudestaan huolimatta sisältää ristiintaulukointi myös sudenkuoppia. Jos ristiintaulukoinnilla esitetään lukumäärätietoa, on sosiologin pidettävä mielessä tietojen vertailukelpoisuus. Jos taas taulukon avulla vertaillaan eri vastaajaryhmien vastauksia prosenttien avulla, on sosiologin taulukkoa rakentaessaan ja tuloksia tulkitessaan ymmärrettävä vertailun suunta.

Kuvailevat menetelmät ovat tarpeellisia myös tutkimusten tulosten esittämisessä (Helenius 1992, 16; 25–26.). Tutkimuksen tulokset voidaan esittää tekstin, taulukon tai kuvion avulla. Numeerisen tiedon visuaalisen esittämisen keinoja käsittelevää alaa kutsutaan tilastografiikaksi. Tilastografiikalla on Vesa Kuuselan (2000) mukaan muualla maailmassa pitkät perinteet ja siitä on kirjoitettu useita oppikirjoja, mutta Suomessa tilastografiikkaa ei juurikaan tunneta tai opeteta. Kuusela ihmettelee, miksi kuvioilla esittämisen haasteellisuutta ei oteta tosissaan. Hän näkee virheellisesti laaditun kuvan voivan valehdella enemmän kuin tuhat sanaa.

Jos kirjoitetaan sellaista, mikä ei pidä paikkaansa, sitä kutsutaan valehteluksi. Jos tehdään taulukko, jonka luvut eivät pidä paikkaansa, on kyseessä huijaus. Sen sijaan, jos tehdään tilastokuvio, joka antaa väärän käsityksen asioista, sitä ei pidetä valehteluna eikä huijaamisena. Miksi? Miksi sellaisia kuvioita vielä julkaistaankin? Yksi syy löytyy varmasti asenteista: tilastokuvion asetetaan vähäisemmät vaatimukset kuin muille esitystavoille. Toinen syy lienee taitojen puute. Suomessa tilastografiikan koulutusta ei taida olla käytännöllisesti ollenkaan. (Kuusela 2000.)

Kuusela vertaa hyvää tilastokuvion mainoksen vastakohtaksi kuvaamalla mainoksen sanomaa yksinkertaiseksi ja rakennetta monimutkaiseksi, kun taas hyvän tilastokuvion sanoma on monimutkainen, mutta rakenne yksinkertainen. Tällä Kuusela tarkoittaa, että numeerisia tietoja visuaalisesti esitettäessä kuvaan olisi mahdutettava niin paljon informaatiota kuin mahdollista niin helposti luettavassa muodossa kuin mahdollista ja vielä siten, ettei kuvasta vahingossakaan saa väärää käsitystä. (Kuusela 2000.)

Tilastolliset testit

Tilastolliset testit syntyivät 1920–1930-luvulla, kun haluttiin kehittää menetelmiä, jotka antaisivat luotettavia tuloksia pienilläkin havaintomäärillä. Siihen asti kehitetyt menetelmät nojautuivat suuriin havaintomääriin ja normaalijakaumaoletukseen. (Mäkelä 1991, 41.) Tilastollisen testauksen sisältävää tilastotieteen aluetta kutsutaan tilastolliseksi päättelyksi. Edellä esitetty kuva 5 perusjoukko-otos-asetelmasta havainnollistaa tilastollisen testauksen yleisen periaatteen. Tapani Valkonen (1981, 105) kuvaa periaatteen seuraavasti: havaittaessa perusjoukosta poimitussa otoksessa jokin tulos, esimerkiksi riippuvuus tai ryhmien välinen ero, halutaan tietää, päteekö sama tulos myös perusjoukossa.

Ronkainen (1999, 130–131) varoittaa tutkijaa kiinnittämästä tilastollisen testin tulokseen sellaisia merkityksiä, jotka eivät siihen kuulu. Hän tuo myös esille testeihin liittyvät piilo-oletukset, jotka sopivat hyvin alkuperäisiin kohdealueisiinsa biologiaan ja tuotevalvontaan, mutta ovat yhteiskunnallisille ilmiöille sopimattomia. Testit ovat myös yleisesti yhteiskuntatieteellisen kvantitatiivisen tutkimuksen alueella kiistanalainen aihe (ks. Mäkelä 1991; Engman 2013; Gorard 2016).

Testin suorittaminen alkaa valintojen tekemisellä. Testausasetelma siihen liittyvine oletuksineen sekä testattavien muuttujien ominaisuudet mitta-asteikkoineen ja jakaumineen vaikuttavat testaustilanteeseen sopivan mallin valintaan. Mallin valinnasta seuraa hypoteesien määrittely, merkitsevyytason valinta, testisuureen laskeminen ja päätöksen tekeminen. Ennen tilastollisia ohjelmia edellä mainitut vaiheet olivat konkreettisia valintoja. Tilastollisten ohjelmien myötä osa niistä on jäänyt pimentoon ja tämän päivän tutkijalle testin suorittamisen vaiheet ovat oikeastaan supistuneet kahteen: sopivan testin valintaan ja p-arvon tulkitsemiseen. Valkonen (1981, 105–106) tuo esille, että testin mekaaninen suorittaminen hypoteeseineen, testin suorittamisineen ja hypoteesin hylkäämisineen tai hyväksymisineen ei ole riittävää sosiologisessa tutkimuskäytössä. Esimerkkinä hän mainitsee testiin liittyvän hypoteesin erilaisuuden tai riittämättömyyden tutkimuskysymykseen liittyvään hypoteesiin verrattuna. Kun tutkimuksen tavoitteena on esimerkiksi kausaalinen tarkastelu, ei korrelaatiota koskevan testin hypoteesi ole riittävä. (Valkonen 1981, 105–106.)

Edellä esitellyt testin suorittamisen vaiheet kuuluvat Mäkelän (1991, 40–55) mukaan hypoteesien testaukseksi kutsuttuun teoriaan. Toisenlaista testaamisen tapaa kutsutaan merkitsevyytestaukseksi. Näiden kahden tavan ero ei ole niinkään matematiikassa vaan niiden suhtautumisessa tutkittaviin ongelmiin, toteaa Mäkelä ja jatkaa, että useimmat oppikirjat esittelevät testauksen hypoteesien testauksen teoriana, josta on siten muodostunut oppikirjaparadigma. Mäkelä tuo esille testien historiaa, jossa aluksi syntyi merkitsevyytestaus Pearsonin, Gossetin ja Fisherin kehittämänä. Merkitsevyytestauksen tarkoituksena oli toimia tieteellisen päättelyn apuvälineenä. Jerzy Neyman ja Egon Pearson jatkoivat merkitsevyytestauksen kehittelyä tarkoituksenaan systematisoida sitä. Kehittelyn tuloksena syntyi hypoteesien testauksen

teoria. Mäkelä näkee heidän systematisointinsa menneen jopa mekaanisuuteen asti. Hän kirjoittaa, että Fisherille tilastotiede toimi päättelyn apuvälineenä, joka auttaisi käytännön tutkimuksessa esiin tulevien ongelmien ratkomisessa, kun taas Neymanin ja Pearsonin hypoteesien testauksen teoria muodostui enemmänkin mekaaniseksi päätöksenteoksi. Mäkelä (1991, 40–55) esittelee hypoteesien testauksen teorian vaiheet ja tuo kussakin vaiheessa esille niiden ristiriitaisuudet sosiaalitutkimuksen tavoitteiden suhteen. Esimerkiksi mallin valinnasta melko mekaanisesti seuraavat hypoteesit voivat yhteiskunnallisia ilmiöitä tutkittaessa olla jopa nurinkuriset. Hän toteaa myös, että hypoteesien kaksisuuntaisuus ei juurikaan vie tutkimusta eteenpäin. Merkitsevyytason valinta on vaihe, johon Mäkelän mukaan liittyy ehkä kaikkein eniten ristiriitoja, harhakuvia ja tutkimuksen piiloutumista tietyn fraseologian taakse. Mäkelä kirjoittaa hauskan kertomuksen siitä, miten merkitsevyytaset 0.05, 0.01 ja 0.001 ovat alkuaan syntyneet. Tasojen käytön taustalla ei ole mitään tieteellisyteen liittyviä syitä vaan ne saivat alkunsa siitä, kun Biometrikan päätoimittajana ollut Pearson valitsi tilaa säästääkseen julkaistavaksi Fisherin vuosikautia laskemista taulukoista vain tasoihin 0.05, 0.01 ja 0.001 liittyvät taulukot. Näihin tasoihin liitettiin vähitellen tilastollista merkitsevyyttä ilmaiseva terminologia ja niitä merkittiin yhdellä, kahdella tai kolmella tähdellä. Terminologia ja tähtien määrä eivät sinänsä ole olleet vaarallisia, mutta haitallista sen sijaan on ollut se, että tilastollinen merkitsevyyden samaistettu tieteellisesti merkittävien tai merkityksellisten tulosten kanssa (ks. myös Wasserstein, Schirm ja Lazar 2019). Hypoteesien testauksen mukaiseen teoriaan kuuluu vielä päätös nollahypoteesin hyväksymisestä tai hylkäämisestä. Mäkelä (1991, 166–167) näkee tämän mekaanisena vaiheena ja tutkijan asiantuntemuksen hukkaan heittämisena. Haitallisuus aiheutuu siitä, että tutkimuksen tulosten yleistämistä ei pidä rajoittaa pelkästään testien käyttöön liittyvään tilastolliseen yleistämiseen. Hänen mukaansa tilastollinen analyysi on ainoastaan työkalu varsinaisia ongelmia ratkaistaessa. Tilastollinen testi ei ole missään tapauksessa tutkimuksen loppupiste vaan usein sen alkukohta. (Mäkelä 1991, 166–167.)

Valkosen (1981, 105) mukaan sosiologiassa on käyty jo 1950-luvun loppupuolelta keskustelua siitä, tulisiko survey-analyysin yhteydessä käyttää tilastollisia testejä. Valkonen tuo esille, että vaikka testaamisesta oli keskusteltu 1980-luvun alkuun mennessä jo pitkään, ei siitä ollut päästy yksimielisyyteen. Orjasniemi (2005, 252) tuo väitöskirjassaan esiin pohdintansa testien käytöstä otantaan perustuvan tutkimusaineistonsa analyysissa.

Mäkelä (1995, 60–61) näkee testien suosion syiksi sen, että tieteellisyyden yleiset kriteerit toistettavuus, yleistettävyyden ja objektiivisuus on helppo samaistaa testauksen terminologian kanssa. Tutkimuksen objektiivisuus on ajateltu saavutettavan siten, että testejä suorittaessaan tutkija valitsee menettelytapansa ennakkoon eikä näin ollen voi jälkikäteen manipuloida tuloksia haluamaansa suuntaan. Mäkelä jatkaa, että tutkijaa on voitu jopa syyttää objektiivisuuden puutteesta, jos hän ei ole käyttänyt sopivia testejä tulostensa tukena. Tilastollisten testien käyttö on voinut

johtaa tiedeyhteisöissä myös äärimmäisen suppeaan objektiivisuuden käsitteeseen. (Mäkelä 1995, 60–61.)

Tilastollisen merkitsevyyden sekoittumisesta tieteelliseen merkittävyyteen on kirjoitettu paljon. Esimerkiksi Lehto (1998, 219) Raiskiin (1974b) viitaten kirjoittaa, että tilastollinen testaaminen ei tuo mitään teoreettisesti merkittävää lisää tutkimukseen, tilastollisella testauksella ei muuteta empiiristä tietoa teoreettiseksi. Myös Mikko Ketokivi (2009, 15) tuo kirjassaan esille, että tilastollinen merkitsevyys ei ole sama kuin tieteellinen löydös eikä tilastollinen tulos ei ole sama kuin tärkeä tulos. Tästä huolimatta usko tilastollisiin testeihin on lujassa ja testit saavat joskus tutkimuksessa suuremman roolin kuin mihin niistä todellisuudessa on. Esimerkkinä tästä ovat tilastollisesti merkitsevien tulosten ylikunnioittaminen ja ”tähtien” toiminen julkaisukynnyksenä tieteellisissä julkaisusarjoissa (Mäkelä 1991, 43). Töttö (2000a) ja Alasuutari (2000) ovat käyneet sananvaihtoa Sosiologia-lehdessä tilastollisista testeistä sekä perusjoukon ja otoksen määrittelyistä ja kokovaatimuksista.

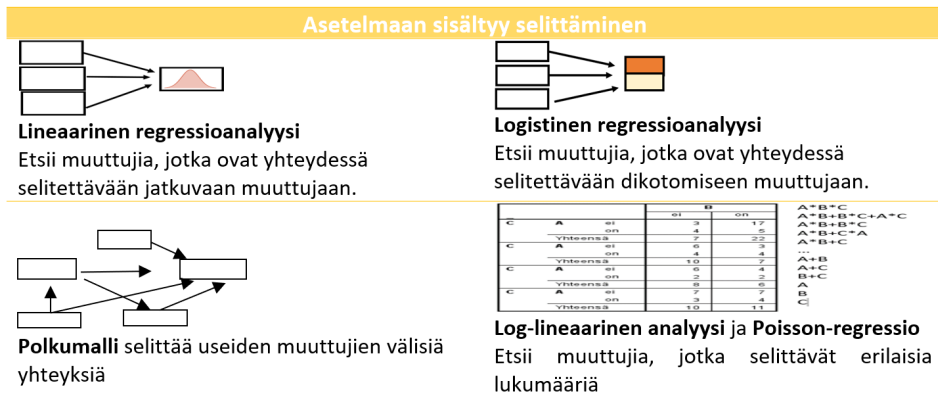
Tilastollisilla testeillä on paikkansa, kunhan niiden rajat tunnustetaan. Testit auttavat esimerkiksi aineistoon tutustumisessa ja antavat analyysin alkuvaiheessa alustavaa tietoa muuttujien välisistä yhteyksistä (ks. Eskola 1967, 110). Valkosen (1981, 107–108) kanta tilastollisten testien käyttöön (kausaalianalyyseissa) on se, että testejä tulisi käyttää sen verran, että tutkija saavuttaa kohtalaisen varmuuden tulostensa paikkansapitävyydestä, mutta testit eivät ole ainoa eivätkä tärkein keino sen saavuttamisessa. Stephen ja Jonathan Gorard (2015; ks. myös Gorard 2018; Wasserstein ym. 2019) esittelevät perinteiselle tilastolliselle testaukselle vaihtoehtoja. Tilastollisen merkitsevyyden eli p-arvon tarkastelun lisäksi tutkija voi arvioida luottamusvälejä tai efektikokoja, mutta niidenkin käytössä he näkevät heikkouksia. He ehdottavat tilalle niin sanottua NNTD-arvoa (‘number needed to disturb’), jonka avulla arvioidaan sitä, kuinka erilainen aineiston tulisi olla, jotta vertailun kohteena havaittu ero häviäisi. Laskenta siis perustuu sen arvioimiseen, kuinka monta vaihtoehtoista havaintoa tarvitaan, jotta tuloksessa havaittu ero häviää, kun otetaan huomioon myös puuttuvien havaintojen määrä. (Gorard ja Gorard 2015; Gorard 2016; 2018.)

Monimuuttujamenetelmät

Näkemykset siitä, mitä monimuuttujamenetelmät ovat ja mitä menetelmiä niihin luetaan kuuluvaksi, vaihtelevat. Tässä tutkimuksessa luen monimuuttujamenetelmiksi Hairia ym. (1998, 6) mukaillen kaikki sellaiset tilastolliset menetelmät, joiden avulla voidaan analysoida vähintään kolmea muuttujaa samanaikaisesti.

Monimuuttujamenetelmiä voidaan jaotella monin tavoin (ks. esim. Nummenmaa ym. 1997; Hair ym. 1998; Toivonen 1999). Seuraavassa olen jakanut menetelmät ongelmanasettelunsa mukaan selittämiseen perustuviin menetelmiin, aineistoa tiivistäviin menetelmiin, ryhmävertailun sisältäviin menetelmiin sekä monia asetelmia sisältäviin menetelmiin.

Selittämiseen pohjautuviin menetelmiin lukeutuvat erilaiset regressioanalyysit sekä log-lineaarinen analyysi (kuva 8). Linearisessa ja logistisessa regressioanalyysissä selitetään aineiston muuttujaa muilla aineiston muuttujilla. Lineaarisen regressioanalyysin käyttöä rajoittaa usein se, että sosiologisen tutkimusaineiston muuttujat eivät täytä menetelmän edellyttämiä tilastomatemattisia vaatimuksia. Vaihtoehtona on käyttää logistista regressioanalyysia, joka on asetelmaltaan samanlainen, mutta muuttujavaatimuksiltaan sosiologisille aineistoille sopivampi. Log-lineaarisisessa analyysissä sekä Poisson-regressioanalyysissä selitetään erilaisia lukumääriä.



Kuva 8 Selittämiseen perustuvia monimuuttujamenetelmiä

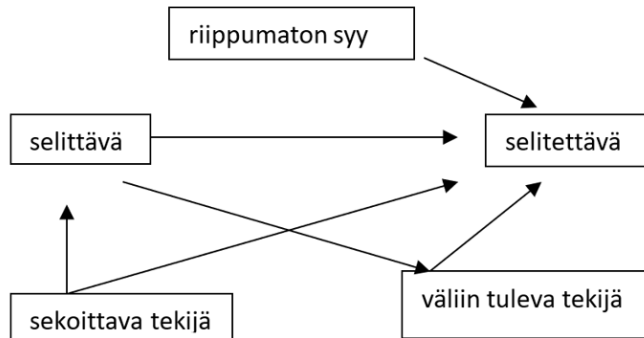
Selittämiseen pohjautuminen on asetelmaltaan hyödyllinen, koska tutkimuksissa luontaista etsiä selittäjiä erilaisille asioille, kuten Antti Eskola kuvaa vuonna 1977:

Sosiologian tutkimusasetelmaan kuuluu tavallisesti jokin riippuva variaabeli eli ilmiö, jota tutkimuksessa halutaan selittää, sekä joukko riippumattomia variaabeleita, joista selitystä etsitään. (Eskola 1977, 15.)

Eskola jatkaa toteamalla, että sosiologisella tutkimuksella voidaan kuvata sellaisia monimutkaisia vaikutussuhteita, joihin ei ehkä muutoin päästäisi käsiksi.

Sosiologia laajentaa arkikokemukseen perustuvaa tietoa, koska se saattaa löytää tärkeitä muuttujia, joihin arkielämän pohdiskelussa ei ole ymmärretty kiinnittää tarpeeksi huomiota, ja monimutkaisia useiden muuttujien samanaikaisia vaikutussuhteita voidaan sosiologian tutkimusmenetelmin kuvata paremmin kuin tavallisessa elämässä käytetyn kielen avulla. (Eskola 1977, 15.)

Polkuanalyysillä pystytään mallintamaan muuttujien välisiä yhteyksiä monipuolisesti. Polkumalliin ja yleisemminkin analyysin pohjana olevaan malliin muuttujat voidaan asettaa erilaisiin rooleihin (kuva 9). Lähtökohtana on tarkastella kahden muuttujan, joita kutsutaan selittäjäksi ja selitettäväksi, välistä yhteyttä. Sekoittavaksi tekijäksi kutsutaan asiaa, joka edeltää selittävää ja selitettävää tekijää joko ajallisesti tai loogisesti. Riippumaton syytekijä on yhteydessä vain selitettävään. Väliin tuleva tekijä on nimensä mukaisesti selittävän ja selitettävän tekijän välissä.



Kuva 9 Muuttujien erilaiset roolit mallituksessa. Lähde: Oja (1990)

Tilastotieteilijän näkökulmasta muuttujat voidaan asettaa mallissa eri rooleihin, kunhan niiden tilastomatemattiset ominaisuudet sen sallivat. Sosiologin näkökulmasta mallin rakentaminen on ongelmallisempaa (ks. esim. Rantalaiho 1972). Allardt (1969) pohtii sosiologista selittämistä mainiten esimerkkinä sukupuolen. Sukupuoli muuttujana on eri asia kuin sukupuoli sosiologisena käsitteenä. Muuttujana se on yksinkertainen, mutta sosiologisena käsitteenä sukupuoleen liittyy monia kytköksiä, esimerkiksi roolit ja odotukset. (Allardt 1969). Myös sosiaaliluokka on tilastollisena muuttujana yksinkertainen mutta sosiologian näkökulmasta sosiaaliluokka ei ole vain yksi ominaisuus vaan useita asioita sisällään pitävä ominaisuuskimppu. Sosiologin on siten pohdittava, mitä mallin elementit edustavat ja mitä kaikkea ne pitävät sisällään.

Selittämisen kytkeminen kausaalisuuteen on sosiologisessa tutkimuksessa ongelmallista. Kausaalisuuskeskusteluun voi tutustua muun muassa Pertti Tötön (2004, 112–136) ja Arto Selkälän (2013, 111–120) teksteissä (ks. myös Winship ja Sobel 2004, 481–504; Mäkelä 1996, 118–119; Sobel 1995, 1–38).

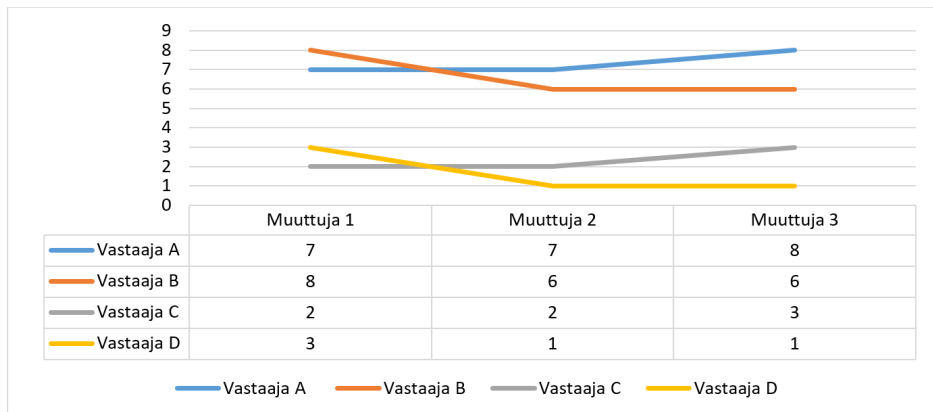
Asetelmaltaan tiivistäviin menetelmiin lukeutuvat faktorianalyysi, ryhmittelyanalyysi ja korrespondenssianalyysi. Kuvassa 10 esitetään faktorianalyysin niputtavan muuttujia ja ryhmittelyanalyysin vastaajia tai havaintoja. Eronteko on yksinkertaistettu. Todellisuudessa molemmat menetelmät tarjoavat monenlaisia vaihtoehtoja tiedon tiivistämiseen. Yksinkertainen eronteko auttaa tutkijaa ylei-

sellä tasolla hahmottamaan menetelmien käyttötarkoituksen. Usein kuitenkin on tarpeen kokeilla menetelmää tutkimusaineistoon, jotta voidaan arvioida, tuottaako menetelmä tutkimuksen kannalta hyödyllistä tietoa.



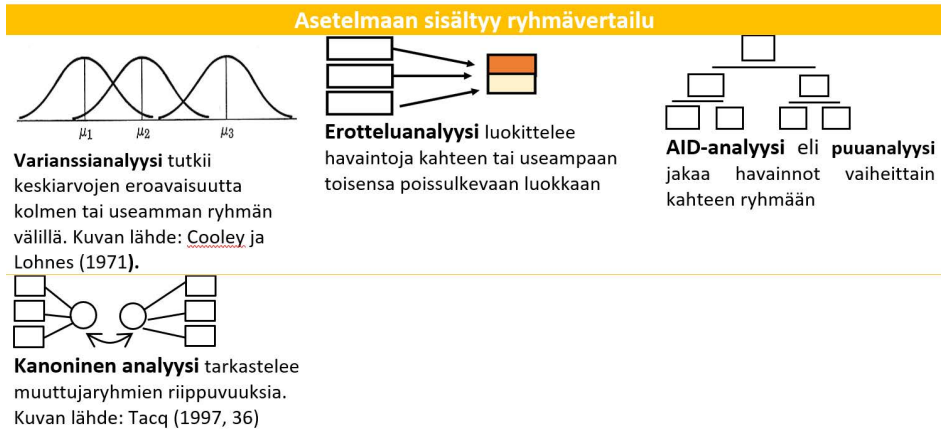
Kuva 10 Tiivistämiseen perustuvat monimuuttujamenetelmät

Faktorianalyysi ja ryhmittelyanalyysi perustuvat erilaiseen matematiikkaan ja erilaisiin tunnuslukuihin. Ryhmittelyanalyysissa tarkastellaan havaintojen välisiä etäisyyksiä, kun taas faktorianalyysi perustuu samankaltaisiin havaintoprofiileihin ja muuttujien korreloitumiseen. Faktorianalyysissa on mahdollisuus valita niin sanottu Q-faktorointimenetelmä, jolloin analyysissa ryhmitelläänkin muuttujien sijaan havaintoja. Kuvassa 11 näkyy Hairin ym. (1998, 98) esimerkki havaintoja niputtavan Q-faktorianalyysin ja ryhmittelyanalyysin tuottamista erilaisista ratkaisuista. Ryhmittelyanalyysi niputtaa lähellä toisiaan olevat havainnot (vastaajat) A ja B yhdeksi ryhmäksi sekä vastaajat C ja D toiseksi ryhmäksi. Faktorianalyysi puolestaan etsii samankaltaista profilia noudattavia havaintoja sijoittaen näin ollen vastaajat A ja C yhteen faktoriin ja vastaajat B ja D toiseen faktoriin.



Kuva 11 Faktorianalyysin ja ryhmittelyanalyysin ero Hairin ym. (1998, 98) esittämänä

Kuva 12 esittää asetelmaltaan ryhmävertailun sisältäviä monimuuttujamenetelmiä. Tällaisiksi luen varianssianalyysin, erotteluanalyysin, AID-analyysin ja kانونin analyysin.

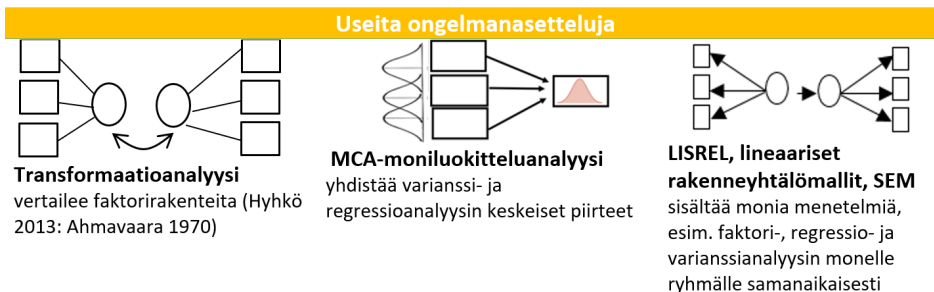


Kuva 12 Ryhmävertailun sisältävät monimuuttujamenetelmät

Varianssianalyysillä tutkitaan keskiarvojen eroavaisuutta kolmen tai useamman ryhmän välillä (esim. Nummenmaa ym. 1997, 78–80; Tacq 1997, 183–230). Erotteluanalyysillä luokitellaan havaintoja kahteen tai useampaan toisensa poissulkevaan luokkaan (Tacq 1997, 231–265; Hair ym. 1998, 239–325; Jokivuori ja Hietala 2007, 119–132). AID-analyysi⁷ eli puuanalyysi on graafinen vaiheittain tapahtuva analyysi, jossa havainnot jaetaan kussakin vaiheessa kahteen ryhmään. Analyysin edetessä syntyy puun mallinen kuvio, jossa latva on analyysin aloituskohta ja analyysin eteneminen näkyy puun oksien haarautumisena. Kanonisessa analyysissä tarkastellaan kahden muuttujaryhmän välisiä riippuvuuksia (Tacq 1997, 322–343).

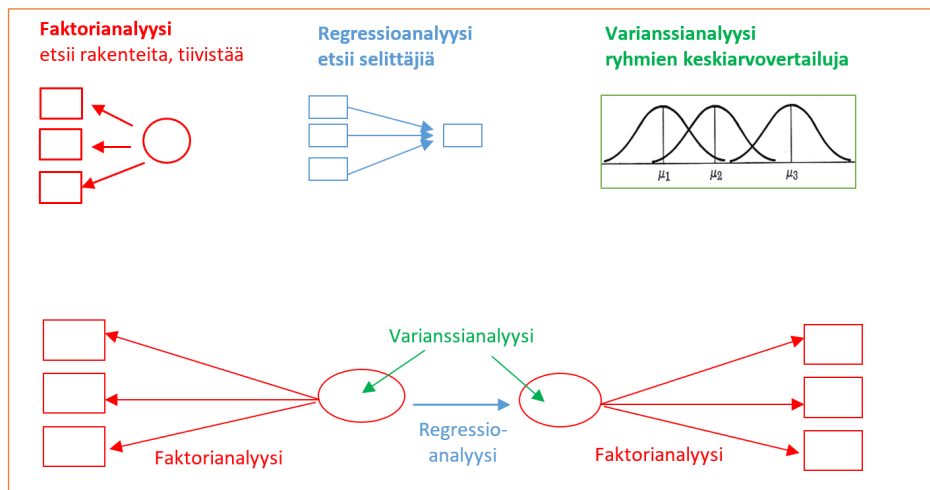
Monia ongelmanasetteluja sisältäviin menetelmiin kuvassa 13 luen Yrjö Ahmavaaran 1950-luvulla kehittämän transformaatioanalyysin, jota sosiologit ovat käyttäneet lähinnä 1960–1970-luvulla. Myös varianssi- ja regressioanalyysin keskeisiä piirteitä sisältävä MCA-moniluokitteluanalyysi, johon voi tutustua Pertti Jokivuoren ja Risto Hietalan (2007, 155–179) teoksessa, lukeutuu tähän ryhmään.

⁷ AID-analyysi kehittyi 1980-luvulla CHAID-analyysiksi (Chi-square Automatic Interaction Detector) ja se liitetään päätöspuuteoriaan



Kuva 13 Monia asetelmia sisältäviä monimuuttujamenetelmiä

Lineaariset rakenneyhtälömallit, jota kutsutaan myös LISREL-analyysiksi, tarjoaa useiden tilastollisten menetelmien sisällyttämistä yhteen malliin (kuva 14). Lineaariin rakenneyhtälömalleihin voidaan sisällyttää faktori-, regressio- ja varianssianalyysia (ks. esim. Hämmäläinen ja Leskinen 1985; Leskinen 2005; Töttö 2004, 213–226; Nummenmaa ym. 1997, 303; Ullman ja Bentler 2004, 431–458; Rafferty 2000; 2005; Bollen ja Long 1993).



Kuva 14 Lineaaristen rakenneyhtälömallien koostuminen faktori-, regressio- ja varianssianalyysistä

Orjasniemi (1997) on käyttänyt kuvan 14 kaltaista mallia sosiaalityön väitöskäytöksessään, jossa hän tarkasteli ranualaisten ja posiolaisten asennoitumisen muutosta alkoholia kohtaan vuodesta 1984 vuoteen 1989. Hän on sisällyttänyt

malliin konfirmatorisia faktorimalleja, joilla hän varmisti väittämät, jotka kuvastavat alkoholiin asennoitumista. Mallissa olevat regressiokertoimet kuvasivat, millä tavalla asennoituminen on muuttunut viiden vuoden aikana. Orjasniemi vertaili vielä asennoitumisen muutosta neljän ryhmän - ranualaisten naisten, posiolaisten naisten, ranualaisten miesten ja posiolaisten miesten - keskuudessa. Varianssianalyysin avulla hän pystyi selvittämään, millä näistä ryhmistä asennoitumisessa on tapahtunut eniten muutosta. (Orjasniemi 1997.)

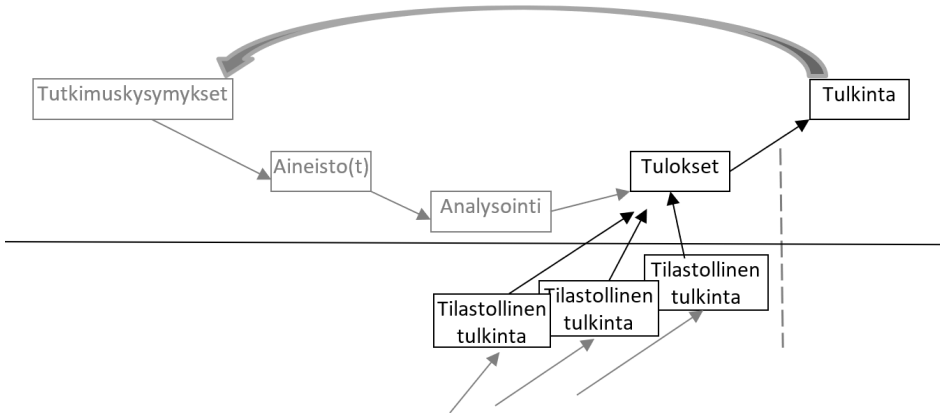
Lineaariset rakenneyhtälömallit mahdollistavat monipuolisen analysoimisen ja monenlaisen tiedon kokoamisen yhteen malliin sekä vielä ryhmävertailujen kautta saatavan tiedon, jota ei erillisillä menetelmillä saataisi. Rakenneyhtälömallit ovat kuitenkin vaativia rakentaa, koska tutkijalla on oltava vankka sisällöllinen ja tilastotieteellinen osaaminen. Tästä varoittavat myös Esko Leskinen ja Jorma Kuusinen (1991):

Nykyisin käytettävissä olevat tilastolliset eksploratiiviset ja konfirmatoriset faktorianalyysimenetelmät ja niiden perusteella luodut tietokoneohjelmat antavat nykyään tutkijoille hyvin tehokkaita keinoja aineistojensa analysointiin. Tutkijoiden on myös syytä käyttää niitä hyväkseen, mutta käytön on oltava älykästä, lujaluontoista ja rehellistä. Tämä tarkoittaa muun muassa sitä, että tutkijoiden on itse hallittava ja ymmärrettävä käyttämiensä menetelmien perusteet. (Leskinen ja Kuusinen 1991.)

Rakenneyhtälömallit sopivat analyysin loppuvaiheeseen, koska silloin tutkijalle on jo muodostunut hyvä käsitys aineistonsa sisällöstä. Tällöin rakenneyhtälömallit voivat toimia yhteen kokoavana analyysinä ja ehkä vielä paljastaa sellaista tietoa, mitä ei erillisillä analyysimenetelmillä saatu.

4.4. Tulosten tilastotieteellinen ja sosiologinen tulkinta

Tutkimusprosessin aikana saadaan monenlaisia ja monen tasoisia tuloksia: kukin analyysimenetelmä tuottaa oman tuloksensa, sosiologin tutkimuskysymyksiin saadaan omat tuloksensa ja ehkä vielä koko tutkimuksellekin nimetään tuloksensa. Näin ollen myös tulosten tulkinta on asia, joka tarkoittaa tilastotieteilijälle ja sosiologille eri asiaa (kuva 15).



Kuva 15 Sosiologin ja tilastotieteilijän näkemys tulkintaan

Kuvassa esitetyt tilastolliset tulkinnat viittaavat kunkin tilastollisen analyysimenetelmän tuloksen tulkintaan, joka tapahtuu kunkin analyysimenetelmän termistöllä ja tilastotieteen ehdoilla. Tilastotieteilijän näkökulmasta tutkimusprosessi myös loppuu tähän tulkintaan. Sosiologille tulkinta merkitsee vastaamista tutkimuskysymyksiin (Pöntinen 2004, 72). Sosiologin on jatkettava tilastollisesta tulkinnasta sellaiseen tulkintaan, jonka avulla hän voi vastata tutkimuskysymyksiinsä. Se tarkoittaa tulosten kontekstointia ja liittämistä tutkimuksen yleiseen näkökulmaan.

Monet sosiologit ovat tunnistaneeet tilastollisen tulkinnan ja sisällöllisen tulkinnan eron, kuten Eskolan (1977, 19–20) kuvauksesta ilmenee: ”Määrällisten numerolaskelmien ohella sosiologi pyrkii tavallisesti myös tulkitsemaan tuloksiaan teoreettisesti sekä ymmärtämään niitä tutkittujen henkilöiden ja ryhmien omasta kokemusmaailmasta käsin”. Myös Allardt (1969) on todennut, että tilastollisella analyysillä ja tilastollisilla menetelmillä voidaan matemaattisesti selittää analyysin kohde, mutta enempään ne eivät kykene. Allardtin mukaan matemaattinen selittäminen ei ole sosiologisesti mielenkiintoista. Sosiologisesti mielenkiintoiseksi menetelmän tulos tulee vasta silloin, kun se liitetään sosiologiseen teoriaan ja yleiseen käsitteistöön. (Allardt 1969, 12.) Myös Suvi Ronkainen, Leila Pehkonen, Sari Lindblom-Yläne ja Eija Paavilainen (2013, 123–125) toteavat, että tutkija, hänen oivalluksensa sekä ymmärryksen kasvunsa ovat keskeisiä. He jakavat analyytisen prosessin kolmeen vaiheeseen. Ensimmäinen vaihe ”analyysi pikku a:lla” on aineistoon tutustumista ja sen rajojen kokeilua. Toinen vaihe on käsitteellistämistä tai tulkintaa, jossa tutkija siirtää analyysin tuottamat havainnot teoreettiseen tai tutkimuskeskustelujen luomaan yhteyteen. Kolmannessa vaiheessa, luennassa, irrottaudutaan aineistokeskeisestä analyysistä ja haetaan tulkintaideoita. Tulkinta syntyy heidän mukaansa aineiston ja teorian keskusteluttamisen tuottamista oivalluksista. (Ronkainen ym. 2013, 123–125.)

On myös tutkijoita, joille tilastollisen tulkinnan ja sisällöllisen tulkinnan ero ei ole selvä, kuten Töttö (2012, 71; ks. myös Byrne 2002) toteaa oppikirjassaan:

Määrällisessä tutkimuksessa argumentoidaan luvuilla: tunnusluvuilla, riippuvuusluvuilla, regressiokertoimilla, selitysprosenteilla, tilastollisilla merkitsevyyksillä ja mallin hyvyttä kuvaavilla indekseillä. Silti on täydellinen väärinkäsitys luulla, että luvut olisivat määrällisen tutkimuksen tuloksia. Luvut ovat pelkkiä merkkejä, jotka pitää lukea. Tuloksia ovat ne sisällölliset johtopäätökset, joita luvuista tehdään, eli se mitä luvut tutkittavasta ilmiöstä kertovat. (Töttö 2012, 71.)

Myös Paananen ym. (1998, 7) tuovat esiin kvantitatiivisen tiedon tulkinnan haasteellisuuden seuraavassa:

Työssämme olemme huomanneet, että kvantitatiivisen tiedon tulkinta ei ole helppoa; edes tutkijat eivät aina osaa tunnistaa sen sävyjä, mahdollisuuksia ja rajoituksia. Aivan liian usein tilastotietoja esitetään faktajuttuina eli tietona, jonka uskottavuutta ei tarvitse perustella. (Paananen ym. 1998, 7.)

Joissakin menetelmäkirjoissa tuodaan esille tilastollisen tulkinnan ja sisällöllisen tulkinnan ero. Esimerkkinä mainittakoon Tapio Nummenmaan, Raimo Konttisen, Jorma Kuusisen ja Esko Leskisen (1997, 15) teos, jossa kuvataan tilastollisia menetelmiä käytävällä tutkimuksella olevan seitsemän vaihetta. Nämä vaiheet ovat 1) tutkimuskysymys, 2) aineisto, 3) tilastollisen menetelmän valinta, 4) tilastollisen ohjelman valinta ja toteutus, 5) tilastollinen tulkinta, 6) sisällöllinen tulkinta ja 7) arviointi. Myös Pertti Jokivuori ja Risto Hietala (2007, 3) viittaavat menetelmäoppikirjassaan sisällöllisen tulkinnan tärkeyteen ihmistieteellisessä tutkimuksessa kirjoittaessaan, että: ”Pelkät analyysin avulla tuotetut numerot ovat kuitenkin vain symboleja, joista ei vielä voi tehdä syvällisiä tulkintoja ilmiöstä”.

Robert Rosenthal (1986, 19) esittää mielenkiintoisen kysymyksen: mitä tarkoitetaan tutkimuksen tuloksella, mikä mielletään tutkimuksen tulokseksi. Rosenthal ei pidä tutkimustuloksina tutkijan tekemiä johtopäätöksiä, koska ne liittyvät hänen mukaansa liian ympärilyövästi ja epäselvästi varsinaisiin tuloksiin. Mielenkiintoa herättää myös Rosenthalin epäily tutkimustulosten tulkinnan oikeellisuudesta. Hänen mielestään muutos, joka joskus tapahtuu tulososion ja pohdintaosion välillä, on itsessään tutkimisen ja perusteellisen mietinnän arvoista.

Pöntinen (2004, 78) kirjoittaa, että sosiologian tutkimusmenetelmien kannalta on merkitystä sillä, että tutkimuksen kohteena oleva yhteiskunta toimii merkitysvälitteisesti. Sosiologia toimii kahden kielen alueella. On olemassa arkikieli, jonka varassa yhteiskunta toimii ja teoriakieli, jonka avulla sosiologia yhteiskuntaa tulkitsee (Pöntinen 2004, 78). Tällöin – Rosenthalia (1986, 19) mukailen – olisi kiinnos-

tavaa tietää, minkälaisia muutoksia tapahtuu siirryttäessä tilastollisesta tulkinnasta sisällölliseen tulkintaan. Sisällöllinen tulkinta on sosiologien itsensä varassa, siihen ei tilastotieteilijällä ole juurikaan kykyä vaikuttaa. Sisällöllisen tulkinnan myötä tutkimuksen tulokset alkavat elämään ja tulevat kiinnostaviksi, mutta onko vaarana, että tulokset alkavat elämään liikaakin omaa elämäänsä. Ketokivi (2009, 6) tuo tilastolliseen analyysiin liittyvän tulkinnan haasteen esille kirjoittaessaan seuraavaa:

Ongelma näkyy myös monissa tilastotieteen oppikirjoissa, joissa käsitellään tilastollista analyysia, mutta lähinnä tilastollisen laskennan näkökulmasta, irrallaan tutkimuskysymyksistä, tulosten tulkinnasta ja erityisesti teoreettisesta selittämisestä. Tilastollisen tulosten tulkinnan moniselitteisyydestä ja tilastollisen päättelyn tietynlaisesta metodologisesta keskeneräisyydestä vaietaan.” (Ketokivi 2009, 6.)

Onkin mielenkiintoista pohtia, kuuluisiko tulosten tulkinta ja teoreettinen selittäminen opettaa tilastotieteen oppikirjoissa vai sosiologian oppikirjoissa. Vai tarvitaanko tieteiden väliin sijoittuva oppikirja? Voidaan myös pohtia sitä, onko mahdollista kirjoittaa oppikirjaa tulkinnasta, kuten Toivonen (1999, 134) tuo esille seuraavassa:

Onko diskurssianalyysi periaatteessakaan sen tyyppinen menetelmä, että joku voisi kirjoittaa siitä keittokirjan? Tutkijan oma mielikuvitus, harkinta ja kokemus ovat tämän tyyppisessä analyysissa aina avainasemassa. Tavallaan sama pätee tilastolliseen analyysiin, vaikka jonkin tilastollisen ohjelman opiskelijaa voidaankin opastaa tyyliin: ”Ensin painat tätä nappulaa ja sen jälkeen tuota”. Vasta analyysimenetelmän käytössä osoitettu luovuus ratkaisee analyysin laadun. (Toivonen 1999, 134.)

Ronkainen ym. (2013, 123) toteavat myös menetelmäkirjojen rajoittuneisuuden tulkinnan apuna. Menetelmäkirjallisuus tarjoaa heidän mukaansa yleisiä ohjeita tutkimuksen tekemiseen, mutta ei siitä, mitä analyysiprosessissa todella tapahtuu. Esimerkkinä menetelmäkirjasta, joissa menetelmiä ei esitelläkään vain tiettyinä menetelmänä – kuten tilastotieteessä – vaan ennemminkin osana tutkimusprosessia, mainittakoon Jokivuoren ja Hietalan (2007) monimuuttujamenetelmien käyttöä ja tulkintaa koskeva teos.

5. Tilastollisten menetelmien vaihteleva suosio sosiologiassa vuosina 1960–2018

5.1. Määrällisestä laadulliseen

Suomalainen sosiologia muodostuu monenlaisen sosiologian ja monenlaisten sosiologien yhteissummasta (Stolte-Heiskanen 1992, 287). Syntyessään 1800-luvun lopussa ja 1900-luvun alussa suomalainen sosiologia kytkeytyi etnologiaan, sosiaaliantropologiaan, sosiaalihistoriaan ja konkreettiseen sosiaalitutkimukseen sekä sosiaalipolitiikkaan. Myös nykyisellä sosiologialla on yhtymäkohtia moniin tieteisiin, esimerkiksi muihin yhteiskuntatieteisiin, käyttäytymistieteisiin ja historiaan (ks. Allardt 1988, 12–14; Kaukonen 1984).

Suomalaisen sosiologian suurin muutos tapahtui toisen maailmansodan jälkeen. Muutoksen myötä sosiologia vaihtui etnologisesta ja sosiaaliantropologisesta painotuksesta sosiaalisia ilmiöitä *mittaamaan* pyrkiväksi, tilastomatemattisia menetelmiä käyttäväksi tieteeksi painottaen enemmän luonnontieteellistä kuin humanistista näkökulmaa (Eskola 1977, 12). Pekka Haatanen (1992, 149) kuvaa muutosta sanoin ”Kyselyineen ja monimuuttujamenetelmineen ei uusi sosiologia juuri kiinnittänyt huomiota historialliseen aineistoon tai historian antamiin selitysmalleihin”. Muutokseen vaikuttivat amerikkalaisesta sosiologiasta ja psykologiasta tulleet vaikutteet (Allardt, Alapuro ja Alestalo 1992, 16). Uudistuneen sosiologian tutkijat olivat 1950- ja 1960-luvulla kiinnostuneita suomalaisen, erityisesti teollisen yhteiskunnan tutkimisesta. Kuten Quételet aikoinaan, halusivat suomalaiset sosiologit saada selville yhteiskunnan toimintaa ohjaavat lakeja ja säännönmukaisuuksia. Tutkimuksen ihanteeksi tulivat systemaattisuus ja pyrkimys hypoteeseihin. Luonnontieteellisen lähestymistavan tarkkuus, tutkittavien kohteiden mittaaminen ja matemaattiset mallit tarjosivat tähän hyvät lähtökohdat. (Allardt 1973, 2–4; Ahtiainen ja Tervonen 1994, 50.)

Allardt⁸ vuonna 1952 valmistunut väitöskirja nähdään suomalaisen sosiologian uudistumisen tärkeäksi maamerkiksi, toteaa Eskola (1992, 267–273). Hän kuvaa, että väitöskirjassaan Allardt kytkee metodisesti vanhan ja uuden sosiologian tutkimusperinteet käyttäessään sekä historiallisia että sosiaalitieteellisiä tutkimusperinteitä, kenttätutkimusta ja myös amerikkalaisperäistä asennetutkimusta. Eskola näkee Allardt⁸in väitöstutkimuksen osoittaneen, että sosiologian ei tarvitse rajoittua sosiaalihistoriaan, tilastotieteeseen tai sosiaalipolitiikkaan vaan sosiologialla on oma

8 Miljöbetingade differenser i skilsmässofrekvensen: olika normsystem och andra sociala faktorerers inverkan på skilsmässofrekvenserna i Finland 1891–1950.

tutkimuskohteensa, oma teoriansa ja omat metodinsa ollen itsenäinen tieteenala muiden tieteiden joukossa. 1970-luvulta alkaen amerikkalaisen sosiologian vaikutusvalta alkoi väistyä antaen tilaa eurooppalaiselle sosiologialle.

Puuronen (2005; 2015) on tarkastellut suomalaisessa sosiologiassa vallinneita muutoksia 1960-luvulta vuoteen 2004 asti paradigma-ajattelun kautta. Puuronen mainitsee tarkoittavansa paradigmalla lähinnä sosiologisia koulukuntia. Paradigma on alun perin Thomas Kuhnin (1994) vuonna 1962 esittämä käsite, jolla Kuhn halusi kuvata tieteen vakiintuneita toiminta. Eri paradigmoissa on omat näkemyksensä sen suhteen, mitä tulee tarkkailla ja tutkia, minkälaisia kysymyksiä tulee kysyä tutkimustilanteessa tai tutkia aiheeseen liittyen, kuinka nämä kysymykset esitetään ja kuinka tieteellisiä tutkimustuloksia tulkitaan (ks. myös Haaparanta ja Niiniluoto 1986, 91). Paradigmat koostuvat erilaisista ontologisista, epistemologisista ja metodologisista näkemyksistä. Ontologinen näkemys viittaa olemassaoloon, mitä nähdään olevan olemassa, mitä voidaan tutkia. Epistemologinen näkemys liittyy tietämiseen, mitä voidaan tietää. Metodologinen näkemys puolestaan viittaa tietämisen tapaan, millä tavalla voidaan tietää tai millä tavalla voidaan hankkia tietoa. Kiikeri ja Ylikoski (2004, 57–58) tulkitsevat paradigman tarkoittavan oikeana pidettyä, yleisesti hyväksyttyä ja auktoriteetin asemassa olevaa teoriaa tai viitekehystä. He arvioivat paradigmojen keskeiseksi ansioksi tutkijoiden vapautumisen hankalista lähtökohtia koskevista pohdinnoista, koska paradigman sisällä on lupa toimia ikään kuin tällaiset ongelmat olisi jo ratkaistu. Tiettyä paradigmaa noudattavat tiedeyhteisön jäsenet omaavat samankaltaisen näkemyksen tieteen ja oman erityisalansa tiedollisista tavoitteista, kuten myös siitä, millaisiin kysymyksiin tieteenala pyrkii vastaamaan ja mitkä ovat mielekkäitä tutkimuskysymyksiä (Kiikeri ja Ylikoski 2004, 59–60).

Puuronen (2005; 2015) toteaa paradigmojen vaihtuneen suomalaisessa sosiologiassa noin kymmenen vuoden välein. Uuden paradigman tullessa valtaan eivät aikaisemmat paradigmat eivät ole Puuronen esityksen mukaan täysin unohtettuja vaan ne ovat edelleen tarjolla uuden ohella. 1960-luvulla suomalaisessa sosiologiassa oli vallitsevana ”positivismi”-paradigma, jonka Puuronen esittää lainausmerkeissä viitaten sillä positivismin moniselitteisiin merkityksiin. Igor Konin (1979, 6) mukaan sosiologian piirissä vaikuttaneen positivismin kattava kuvaaminen on vaikeaa, koska positivismi, kuten myös sosiologia, ovat käsitteellisiltä sisällöiltään historiallisen kehityksen kuluessa vaihdelleet huomattavasti. Usein positivismilla viitataan kvantitatiivisuuteen tai sitä pidetään tilastollisia menetelmiä hyödyntävän tutkimuksen tieteenfilosofisena taustana ja periaatteena. 1950- ja 1960-luvun positivismin yhteydessä mainitaan usein myös Eino Kailan kehittämä looginen empirismi (Allardt ym. 1992, 16; ks. myös Niiniluoto ja Koskinen 2002). Positivismin samastamista kvantitatiivisuuteen ovat kritisoineet esimerkiksi Töttö (1997b; 2000b; 2004; 2012) sekä Kristiina Rolin, Marja-Liisa Kakkuri-Knuuttila ja Elina Henttonen (2006, 11).

Samaan aikaan ”positivismin” kanssa on suomalaisessa sosiologiassa ollut Puuronen (2005; 2015) mukaan vallalla ”jälkipositivismi”. Puuronen esittää ”positivismin”

metodisiksi ihanteiksi muun muassa kokeellisuuden, hypoteesien todentamisen ja pääasiassa kvantitatiiviset menetelmät. ”Jälkipositivismissa” nämä ihanteet puolestaan ovat Puurosen listaamina muunneltu kokeellinen, manipuloiva kriittinen monimetodisuus, hypoteesien falsifointi ja lupa käyttää myös laadullisia menetelmiä.

Kriittinen realismi nousi vallitsevaksi paradigmaksi 1970-luvun puolivälissä. Sen metodisena ihanteena ovat dialogisuus ja kvalitatiiviset aineistot. 1990-luvulla alkoi konstruktionismin aika. Konstruktionismin analyysitapoja ovat Puurosen mukaan diskurssianalyysi ja keskusteluanalyysi. Puuronen nimeää 2000-luvun alun paradigmatoksi uuden kriittisen realismin, jonka metodiseksi ihanteeksi hän mainitsee monenlaiset näkemykset periaatteella ”kaikki käy mutta ei kaikkeen”. (Puuronen 2005.)

Allardt (1973, 1–4) on todennut sosiologian paradigmojen vaihtuvan tiheästi. Syyksi sosiologian nopeatahtiseen paradigmatvaihteluun Allardt on arvellut sosiologian tutkimuskohteiden luonteen. Sosiologian on ollut pakko nopeasti reagoida tutkimuskohteissa tapahtuneisiin muutoksiin. Allardt myös toteaa sosiologian paradigmatkriisien tiheyden antaneen paljon aiheita ironisille kommenteille. Esimerkkinä tällaisesta hän esittää sosiologi Johan Asplundin maininnan, jonka mukaan sosiologia muistuttaa välistä turhuuden markkinoita: pukeudutaan kernaasti uusiin vaatteisiin.

Taulukossa 1 näkyy sosiologiassa tarkastettujen väitöskirjojen viisivuosittaiset määrät. Keskimmaisessä sarakkeessa näkyy tilastollisia menetelmiä sisältävien väitöskirjojen määrät ja viimeisessä sarakkeessa niiden osuudet kaikista väitöskirjoista. 532 tutkimuksen kohteena olevan sosiologian väitöskirjan joukossa on 180 sellaista väitöskirjaa, jossa on käytetty tilastollisia menetelmiä edes jonkin verran. Tilastollisia menetelmiä sisältävien väitöskirjojen osuus on siten 34 % kaikista vuosina 1960 – 2018 tarkastetuista sosiologian väitöskirjoista.

Taulukko 1 Tilastollisten menetelmien esiintyminen sosiologian väitöskirjoissa

| | Väitöskirjojen määrä | Tilastollisia menetelmiä sisältävien väitöskirjojen määrä | Tilastollisia menetelmiä sisältävien väitöskirjojen osuus |
|-------------|-------------------------|---|--|
| 1960 – 1964 | 9 | 8 | 89 % |
| 1965 – 1969 | 13 | 13 | 100 % |
| 1970 – 1974 | 16 | 12 | 75 % |
| 1975 – 1979 | 13 | 8 | 62 % |
| 1980 – 1984 | 11 | 7 | 64 % |
| 1985 – 1989 | 18 | 9 | 50 % |
| 1990 – 1994 | 51 | 17 | 33 % |
| 1995 – 1999 | 61 | 14 | 23 % |
| 2000 – 2004 | 90 | 21 | 23 % |
| 2005 – 2009 | 80 | 28 | 35 % |
| 2010 – 2014 | 112 | 27 | 24 % |
| 2015 – 2018 | 58 | 16 | 28 % |
| Yhteensä | 532 | 180 | 34 % |

Suosituinta tilastollisten menetelmien käyttö on ollut 1960-luvulla, jolloin lähes jokainen väitöskirja sisälsi tilastollisia menetelmiä. 1960-lukua on aivan perustellusti luonnehdittu kvantitatiivisen tutkimuksen kultakaudeksi (Alastalo 2005, 73; Puuronen 2005; Stolte-Heiskanen 1992, 295; Toivonen 1999, 119). Sosiologit olivat tuolloin suosittuja asiantuntijoita ja tärkeitä yhteiskunnallisia vaikuttajia (Puuronen 2015, 132). 1960-luvulla uskottiin, että tieteen avulla voidaan ohjata yhteiskunnan kehitystä ja sosiologian avulla voidaan saada selville tosiasioita yhteiskunnasta ja sen säännönmukaisuudesta (Eskola 1997; ks. myös Puuronen 2015, 133). Yhteiskunnassa vallitsevat ristiriidat toivottiin tutkimuksen avulla saatavan näkyviksi ja sen myötä hallittaviksi, ja sosiologian uskottiin tarjoavan tapoja ymmärtää ja käsitteellistää yhteiskuntaa (Puuronen 2005, 52). Sosiologit olivat Raution (1985, 159–160) mukaan erittäin kiinnostuneita mielipiteiden, asenteiden ja arvostusten selvittämisestä, mikä saattoi osaksi johtua tietokoneiden suomista uusista mahdollisuuksista ja tutkimusmenetelmistä, mutta sellaiselle tiedolle riitti myös kysyntää. Tilastollisia menetelmiä hallitseva sosiologi sai myös arvostusta tutkijayhteisössä, toteaa Rautio seuraavalla kuvauksellaan:

Mitä monimutkaisempia tilastollisia analyysimenetelmiä yhteiskuntatieteilijät käyttivät, sitä helpommin he erottautuivat harmaasta tutkijoiden massasta. Samalla yhteiskuntatieteilijät voivat lopultakin katsoa tekevänsä jotakin samantapaista kuin tutkijoiden keskuudessa perinteisesti arvostetut matemaatikot ja ns. kovien luonnontieteiden edustajat. (Rautio 1985, 160.)

Toiviainen (1970, 14) kirjoittaa vuonna 1970 tarkastetussa väitöskirjassaan, että ”sosiologeilla on taipumusta potea alemmuuskompleksia kahteen suuntaan: luonnontieteille teorioistaan ja tilastomatematiikalle menetelmistään”. Toivaiasen kommentti heijastaa 1960-luvulla vallinnutta ajattelutapaa sosiologiassa, jolloin tutkimuksessa otettiin mallia luonnontieteistä ja ihannoitiin kokeellista tutkimusta. Sosiologia nähtiin 1960-luvulla vakuuttavana tieteenalana. Koska tilastollisten menetelmien käyttö lisäsi sekä sosiologian vakuuttavuutta, että sosiologiain arvostusta, on helppo ymmärtää sosiologiain innostus tilastollisten menetelmien opetteluun ja käyttöön.

Tilastollisia menetelmiä sosiologit pystyivät opiskelemaan harvalukuisista menetelmäkirjoista, jotka olivatkin ahkerassa käytössä. Kirjojen lisäksi sosiologit pääsivät opettelemaan tilastollisia menetelmiä käytännönläheisiä ohjeita sisältävistä monisteista, joita esimerkiksi julkaisi Alkoholipoliittinen tutkimuslaitos julkaisusarjassaan (Alastalo 2005, 79). Niitä käytettiin väitöstutkimustenkin lähteinä, Alastalo toteaa. Tilastollisiin menetelmiin oli mahdollista tutustua myös erilaisilla kursseilla sekä saada oppia kollegoilta, tilastotieteilijöiltä ja projekteissa työskentelemällä (Alastalo 2005, 79).

Tilastollisten menetelmien käyttö alkoi vähentyä 1970-luvulla, mutta edelleen niitä kuitenkin käytettiin noin joka toisessa väitöskirjassa aina 1980-luvun loppuun saakka. 1990-luvulta lähtien tilastollisia menetelmiä esiintyy 23-35 %:ssa sosiologiain väitöskirjoista.

Väitöskirjojen määrä lisääntyi selvästi 1990-luvulla. Määrän lisääntyminen ei ole vain sosiologiassa tapahtunut ilmiö vaan väitöskirjojen määrät alkoivat 1990–2000-luvulla kasvaa myös muissa oppiaineissa (Alastalo 2005). Samalla kun väitöskirjojen määrät kasvoivat 1990-luvun alussa, alkoivat laadulliset menetelmät olla määrällisiä menetelmiä suositumpia. Puuronen (2005) on todennut, että 1990-luvulla sosiologit eivät enää olleet kiinnostuneita tilastollisesta analyysistä vaan diskursseista, narratiiveista, retoriikasta ja muista kielellisistä konstruktioista.

Kädenvääntö kvantitatiivisten ja kvalitatiivisten menetelmien paremmuudesta oli näkyvää 1990-luvulla (Lehto 1998, 209; Heiskanen 1998, 201–202; Pöntinen 2004, 82; Alastalo 2005, 17–19). Kädenvääntö näkyy esimerkiksi Tötön ja Alasuutarin kirjoissa. Alasuutari julkaisi vuonna 1989 yhteiskuntatutkimusta ja sen menetelmiä käsittelevän kirjansa, jonka väitteisiin Töttö (1997b) otti kantaa vuonna 1997 ilmestyneellä kirjallaan *Pirallinen positivismi. Kysymyksiä laadulliselle tutkimukselle*. Vuonna 1999 ilmestyneen *Laadullinen tutkimus* -kirjan 3. uudistetussa painoksessa Alasuutari (1999) kommentoi Tötön kysymyksiä, johon Töttö (2000b, 14) puolestaan vastaa seuraavassa menetelmiä käsittelevässä kirjassaan *Pirallisen positivismin paluu*. Kirjansa alaotsikossa Töttö ei enää esitä kysymyksiä pelkästään laadulliselle tutkimukselle vaan tekee ”laadullisen ja määrällisen tarkastelua” (Töttö 2000b). Seuraavassa menetelmiä koskevan kirjansa Töttö (2004) aloittaa toteamalla: ”Puhe laadullisesta ja määrällisestä tutkimuksesta on suuri erehdys”. Hän jatkaa, että kysei-

set adjektiivit voitaisiin unohtaa ja ajatella kahtiajakoa aivan toisin käsittein (Töttö, 2004, 9). Vuonna 2012 ilmestyneessä kirjassaan Töttö pohtii jälleen laadullista tutkimusta. Erilaisten esimerkkien kautta hän päätyy kuvaamaan, että kvalitatiivisessa tutkimuksessa käytetään yhtä lailla määrää kuvaavia ilmaisuja kuin kvantitatiivisesakin, mutta ilmaisut ovat vain luonteeltaan epätarkempia.

Sosiologian väitöskirjoissa on otettu kantaa kvantitatiivisuus-kvalitatiivisuus-jakoon jonkin verran 1990-luvulta alkaen. Väitöskirjoissa pätee Alastalon (2005) huomio liittyen yleiseen kvantitatiivisuus-kvalitatiivisuus-keskusteluun. Alastalon mukaan kvantitatiivisten menetelmien käyttäjät peilaavat ja perustelevat valintaansa suhteessa kvalitatiivisiin, mutta toisinpäin ei tehdä eli kvalitatiivisten menetelmien käyttäjät eivät ota kantaa siihen, miksi he eivät ole käyttäneet kvantitatiivisia menetelmiä.

Töttö (2004) on kritisoinut kvalitatiivisen otteen sitomista syvällisyyteen ja kvantitatiivisen sitomista pinnallisuuteen. Muutamassa väitöskirjassa on kvalitatiivisen ja kvantitatiivisen otteen etuja puntaroitu juurikin tällä syvällisyys-pinnallisuus-vertauksella.

Kvalitatiivisen aineiston ottaminen mukaan täydentämään määrällistä aineistoa oli onnistunut ratkaisu. Sen avulla pystyin paneutumaan syvällisemmin pieniin yksityiskohtiin paremmin kuin pelkästään määrällisen aineiston käytöllä. Yksinään kvalitatiivinen aineisto ei kuitenkaan olisi saanut sitä painoarvoa, jonka kvantitatiivisen tutkimuksen etuna on. (Valli 1998, 187.)

Pinnallisesti ajatellen kvalitatiivisilla tutkimuksilla ja esitutkimuksilla hiotaan kvantitatiivisen menetelmän särmiä, tehdään se ”siedettävän salonkikel-poiseksi”... Siten on hedelmällistä aloittaa tilastollisista vertailuista ja edetä niiden jälkeen syventäviin tutkimuksiin kvalitatiivisilla menetelmillä. (Heiskanen 2002, 243.)

Tässä tutkimuksessa kvantitatiivisella otteella kuvataan ... määrällistä ulottuvuutta ja siinä mielessä pintaa, kvalitatiivinen ote taas mahdollistaa ... pinnan alle kurkistamisen. (Hilden-Paajanen 2005, 103.)

Kvantitatiivisten ja kvalitatiivisten menetelmien vastakkainasettelua on myös kritisoitu (ks. Eskola ja Suoranta 1995, 93–94). Toivonen (1999, 139) kommentoi 1990-luvun lopulla julkaistussa kirjassaan kvantitatiivisen ja kvalitatiivisen vastakkainasettelua toteamalla, että vaikka raja kvalitatiivisen ja kvantitatiivisen välillä on teknisessä mielessä hämärtynyt, on osa epäkohdista tutkijan päässä ja ne voi poistaa vain paremmalla pään käytöllä. Pöntinen toi vuonna 2004 esille kahtiajaon tuomat haasteet sosiologian tutkimusmenetelmäopetukseen:

Opiskelijoille pitäisi antaa monipuolinen kuva menetelmäkirjoista, mutta tuloksena on helposti pintaliipaisu, joka ei anna todellisia tutkimusvalmiuksia. Toinen ongelma on se, että käytännön syistä kvalitatiivisten ja kvantitatiivisten menetelmien opetus on pitkälle erillään toisistaan, mikä korostaa keinotekoisesti tätä kahtiajakoa. Toisaalta tähän jakoon on liittynyt vahvoja erimielisyyksiä, jotka juontuvat tieteenfilosofisista periaatteista ja heijastavat myös sosiologian sisäisiä koulukuntaeroja. (Pöntinen 2004.)

Määrällisen ja laadullisen yhdistämisestä on käytetty nimityksiä triangulaatio (ks. Lehto 1998, 224) ja mixed methods (ks. Ronkainen 2014; Teddlie ja Tashakkori 2009). Metodologinen pluralismi eli monien menetelmien käyttö alkoi yleistyä 1990-luvulla Britanniassa (Alastalo 2014). Suomalaisissa sosiologian väitöskirjoissa tutkimusotteiden yhdistäminen alkaa enenevässä määrin näkymään 2000-luvulla, vaikka viitteitä siitä näkyy jo 1990-luvun väitöskirjoissa. Esimerkiksi vuonna 1996 tarkastetussa väitöskirjassaan Jukka Niemelä (1996, 28) tuo esille, miten ”kvalitatiivinen aineisto antoi virikkeitä kyselylomakkeen kysymysten muotoiluun ja auttoi myös kvantitatiivisen analyysin tulosten tulkinnassa. Toisaalta kyselyn tulokset herättivät uusia kysymyksiä ja ohjasivat näin kvalitatiivisen aineiston keruuta.”

Kiinnostuksen kohteet, tutkimusaiheet ja tutkimuksen tavat ovat 2000-luvun sosiologiassa monipuolisia. 2000-luvulla valmistuneet sosiologian väitöskirjat ovat usein artikkelimuotoisia pitäen sisällään monien erilaisten aineistojen ja analyysimenetelmien hyödyntämistä. Pöntinen (2004, 91) kirjoittaa sosiologisen tutkimuksen monipuolistumisesta seuraavasti:

Viime aikoina erilaisten aineistojen ja menetelmien yhteiskäyttö on yleistynyt. On huomattu, että tällä tavoin tutkimuskohteesta saadaan monipuolisempi kuva. Toisaalta tämä edellyttää menetelmien parempaa hallintaa. Tutkija ei voi enää rajoittua yhden suosikkimenetelmänsä käyttöön. Usein hänen on myös haettava yhteistyökumppaneita, niin että muodostuu monipuolisesti erilaisia menetelmiä hallitseva projektiryhmä. (Pöntinen 2004, 91.)

Monen tyyppisten aineistojen, monien tutkimusmenetelmien ja tutkijoiden verkostoitumisen ja tutkijoiden muodostamien projektiryhmien seurauksena tutkimusenteko on monipuolistunut ja ammattitaitoistunut. Erolan ja Räsänen (2014a, 241) mukaan sosiologian monipuolistuminen – tai pirstaloituminen – on nähty sekä ongelmana että vahvuutena. He arvelevat, että menetelmien merkitys sekä erityisesti kvantitatiivisen menetelmäosaamisen vaatimustaso tulevat kasvamaan sosiologiassa. Tällä Erola ja Räsänen (2014b, 68) viittaavat uusiin teknologisiin tiedonkeräystapoihin ja big dataan.

Vincent Traag ja Thomas Franssen (2016) ovat tarkastelleet bibliometrisen visualisoinnin avulla sosiologien käyttämää kieltä ja sanavalintoja sosiologisissa

aikakauslehdissä. He huomasivat, että sosiologien käyttämä kieli tuo esille, että kvantitatiivinen-kvalitatiivinen-jako on edelleen vahvasti näkyvissä sosiologiassa.

Yliopistokohtainen tarkastelu näkyy taulukossa 2. Suomen vanhin sosiologiaa opettava yliopisto on Helsingin yliopisto. Sosiologian opetus alkoi Helsingin yliopistossa käytännöllisen filosofian ohessa vuonna 1890 kunnes vuonna 1927 se itsenäistyi omaksi oppiaineekseen. Turun yliopistoon sosiologia tuli oppiaineeksi vuonna 1926 ja Åbo Academiin vuonna 1942. Tampereen yliopistoon perustettiin sosiologian professorin virka vuonna 1944. Joensuun korkeakoulun (vuodesta 1984 Joensuun yliopisto) yhteiskuntapolitiikan apulaisprofessorin virka perustettiin vuonna 1974. Ensimmäinen sosiologian pääaineopiskelija valmistui 1982. Jyväskylän yliopistossa sosiologian opetus alkoi 1970-luvun loppupuolella. Lapin korkeakoulun yhteiskuntatieteiden tiedekunta perustettiin vuonna 1982. Kuopion yliopistossa sosiologian opetus alkoi vuonna 1990. Sosiologian syventävien opintojen opetus päättyi Oulussa vuonna 2013. Joensuun yliopisto ja Kuopion yliopisto yhdistyivät Itä-Suomen yliopistoksi vuonna 2010.

1960-luvulla sosiologian väitöskirjoja valmistui Helsingin, Tampereen ja Turun yliopistoista. Åbo Academiasta ensimmäinen sosiologian väitöskirja valmistui vuonna 1971, Jyväskylän yliopistosta vuonna 1986, Joensuun yliopistosta vuonna 1988, Oulun yliopistosta vuonna 1992, Lapin yliopistosta vuonna 1996, Kuopion yliopistosta vuonna 1997 ja Itä-Suomen yliopistosta vuonna 2010.

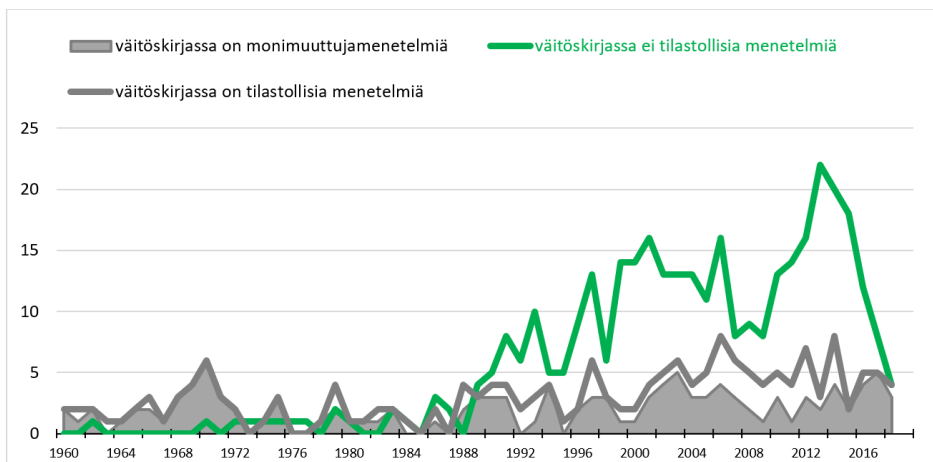
Taulukko 2 Tilastollisia menetelmiä sisältävien väitöskirjojen esiintyminen yliopistoittain

| | Tilastollisia menetelmiä sisältävien väitöskirjojen määrä yliopistoittain | | | | | | | | | | Tilastollisia menetelmiä sisältäviä väitöskirjoja yhteensä | Sosiologian väitöskirjoja yhteensä |
|---------------------------|---|-----|----|----|----|----|----|-----|----|----------------------|--|------------------------------------|
| | HY | TaY | TU | ÅA | JY | KY | LY | JoY | OY | Itä-Suomen yliopisto | | |
| 1960-1964 | 4 | 2 | 2 | | | | | | | | 8 | 9 |
| 1965-1969 | 8 | 4 | 1 | | | | | | | | 13 | 13 |
| 1970-1974 | 5 | 4 | 2 | 1 | | | | | | | 12 | 16 |
| 1975-1979 | 6 | 1 | 1 | 0 | | | | | | | 8 | 13 |
| 1980-1984 | 4 | 1 | 2 | 0 | | | | | | | 7 | 11 |
| 1985-1989 | 5 | 0 | 2 | 0 | 1 | | | 1 | | | 9 | 18 |
| 1990-1994 | 12 | 1 | 1 | 2 | 0 | | | 1 | 0 | | 17 | 51 |
| 1995-1999 | 8 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | | 14 | 61 |
| 2000-2004 | 13 | 0 | 3 | 0 | 4 | 0 | 1 | 0 | 0 | | 21 | 90 |
| 2005-2009 | 12 | 8 | 3 | 1 | 2 | 0 | 1 | 0 | 1 | | 28 | 80 |
| 2010-2014 | 10 | 4 | 3 | 0 | 3 | | 1 | | | 5 | 27 | 112 |
| 2015-2018 | 7 | 1 | 0 | 0 | 4 | | 0 | | 0 | 4 | 16 | 58 |
| Yhteensä | 94 | 28 | 22 | 4 | 14 | 1 | 4 | 2 | 2 | 9 | 180 | |
| Sosiologian väitöskirjoja | 218 | 83 | 49 | 22 | 66 | 5 | 23 | 17 | 11 | 38 | | 532 |
| Osuus yliopistoittain | 43 | 34 | 45 | 18 | 21 | 20 | 17 | 12 | 18 | 24 | | |

Tilastollisia menetelmiä sisältäviä väitöskirjoja on valmistunut määrällisesti eniten Helsingin yliopistosta. 218 sosiologian väitöskirjasta 94:ssä (43 %) on käytetty tilastollisia menetelmiä. Suhteellisesti tarkasteltuna on tilastollisten menetelmien käyttö ollut suosituinta Turun yliopistossa, jossa 45 % sosiologian väitöskirjoista sisältää tilastollisia menetelmiä.

5.2. Sosiologian väitöskirjoissa käytetyt monimuuttujamenetelmät

Kuvassa 16 esitetään vuosina 1960–2018 tarkastettujen sosiologian väitöskirjojen määrät eroteltuna tilastollisia menetelmiä sisältäviin ja niitä sisältämättömiin. Kuvassa kiinnittää huomiota 1990-luvulla alkanut väitöskirjojen määrän kasvu ja laadullisia menetelmiä sisältävien väitöskirjojen lisääntyminen. Tämän luvun kannalta keskeinen asia liittyy tilastollisia menetelmiä sisältävien väitöskirjojen määrän vertailuun siltä osin, ovatko väitöskirjoissa käytetyt tilastolliset menetelmät kuvailevia menetelmiä ja testejä ilman monimuuttujamenetelmiä. Kuvaan on merkitty vihreällä viivalla ne väitöskirjat, joissa ei ole käytetty mitään tilastollisia menetelmiä ja harmaa viiva esittää tilastollisia menetelmiä sisältävien väitöskirjojen määrät. Harmaan viivan sisällä oleva harmaa alue osoittaa monimuuttujamenetelmien osuuden tilastollisia menetelmiä sisältävistä väitöskirjoista. Harmaa alue pyrkii siis kuvaamaan, minkälainen osuus monimuuttujamenetelmillä on ollut kaikkien tilastollisten menetelmien käytössä. 180:n tilastollisia menetelmiä sisältävän väitöskirjan joukossa on 123 väitöskirjaa, joihin sisältyy monimuuttujamenetelmiä ja 57 sellaista, joissa on käytetty pelkästään kuvailevia menetelmiä ja tilastollisia testejä.



Kuva 16 Sosiologiassa vuosina 1960–2007 tarkastetut väitöskirjamäärät eroteltuna tilastollisia menetelmiä ja monimuuttujamenetelmiä sisältäviin

Sosiologien väitöskirjoissaan käyttämien monimuuttujamenetelmien kirjaaminen tutkimusaineistoon osoittautui haasteelliseksi. Sosiologit ovat maininneet menetelmät hyvin vaihtelevin käytännöin ja erilaisilla tarkkuuksilla. Joistakin väitöskirjoista en esimerkiksi löytänyt mainintaa menetelmän nimestä, mutta menetelmän yhteydessä käytetyistä sanavalinnoista olen päätellyt tutkijan käyttäneen tiettyä menetelmää. Esimerkiksi sanoista odds ratio, OR tai ristitulo suhde olen päätellyt tutkijan käyttäneen logistista regressioanalyysia tai log-lineaarista analyysia. Joissakin väitöskirjoissa esiintyi ristiriitaisuuksia menetelmän nimeämisen, menetelmäesittelyiden ja tulosten välillä. Tutkija on esimerkiksi kirjoittanut käyttäneensä varianssianalyysia, mutta hänen kuvaksensa menetelmästä ja hänen esittämänsä tulokset vaikuttavat logistiselta regressioanalyysilta (esim. Kainulainen 1998). Tällaisissa tilanteissa olen kirjannut menetelmän tutkijan käyttämää nimitystä kunnioittaen. Lisähaastetta menetelmien kirjaamisessa tutkimustani varten aiheutti myös se, että joidenkin menetelmien nimet ovat aikojen saatossa muuttuneet. Erityisesti logistisen regressioanalyysin ja log-lineaarisen analyysin kirjaamiset aiheuttivat päänvaivaa. Nämä menetelmät olivat 1990-luvulla vielä melko uusia ja vähän tunnettuja eivätkä niiden nimet olleet vielä vakiintuneet. Tuolloin väitöskirjoissa käytettiin vaihtelevasti nimityksiä log-lineaarinen analyysi, logit-analyysi ja logistinen regressioanalyysi menetelmästä, joka nykyisestä näkövinkkelistä katsottuna olisi logistinen regressioanalyysi. 2000-luvulla nimitystapa näyttää vakiintuneen logistiseksi regressioanalyysiksi.

Väitöskirjoissa mainittuja tilastollisia analyysimenetelmiä ovat aakkosjärjestyksessä mainittuna AID-analyysi, Cox-regressioanalyysi, demografiset menetelmät, eksploratiivinen faktorianalyysi, elinaika-analyysi, erotteluanalyysi, faktorianalyysi, hierarkkinen analyysi, kanoninen analyysi, klusterianalyysi, kohorttianalyysi, konfirmatorinen faktorianalyysi, korrespondenssianalyysi, lineaarinen regressioanalyysi, LISREL-analyysi, logistinen regressioanalyysi, logit-mallit, log-lineaarinen analyysi, MCA-analyysi, monisuuntainen varianssianalyysi, Poisson-regressioanalyysi, polkuanalyysi, pääkomponenttianalyysi, Q-metodologia, regressioanalyysi, ryhmittelyanalyysi, transformaatioanalyysi ja varianssianalyysi yksisuuntaisena ja kaksisuuntaisena. Olen ryhmitellyt menetelmät seuraaviin kokonaisuuksiin: **Faktorianalyysin** alle olen sijoittanut maininnat faktorianalyysista, eksploratiivisesta faktorianalyysista sekä pääkomponenttianalyysista. **Regressioanalyysin** alle olen lukenut kuuluvaksi maininnat yhden selittäjän ja monen selittäjän regressioanalyysista sekä lineaarisesta regressioanalyysista. **Logistisen regressioanalyysin, log-lineaarisen analyysin ja logit-mallien** ryhmittely oli haastavaa, koska niiden osalta erityisesti on tapahtunut muutoksia menetelmien nimeämiskäytännöissä. Päädyin jakamaan menetelmät logistiseen regressioanalyysiin ja log-lineaariseen analyysiin. **Varianssianalyysi** pitää sisällään yksisuuntaisen, kaksisuuntaisen ja monisuuntaisen varianssianalyysin. **Ryhmittelyanalyysi** on toiselta nimeltään klusterianalyysi, joten olen sijoittanut näihin liittyvät maininnat samaan ryhmään.

Taulukossa 3 näkyy sosiologian väitöskirjoissa eri aikoina käytetyt monimuuttujamenetelmät. Taulukossa esitetään erikseen sellaiset väitöskirjat, joissa tutkimusaineistona on ollut rekisteriaineisto. Taulukosta käy ilmi, että rekisteriaineistoon sovelletut monimuuttujamenetelmät ovat erilaisia kuin ei-rekisteripohjaisiin aineistoihin - joita kutsun taulukossa kvantitatiivisiksi aineistoiksi - sovelletut monimuuttujamenetelmät. 123:sta monimuuttujamenetelmää sisältävästä väitöskirjasta 100 on sellaisia, joissa analyysin kohteena on ollut kvantitatiivinen aineisto. 23 väitöskirjaa perustuu puolestaan rekisteriaineistojen analyysiin. Ensimmäinen rekisteriaineistoa hyödyntävä väitöskirja on tarkastettu vuonna 1989.

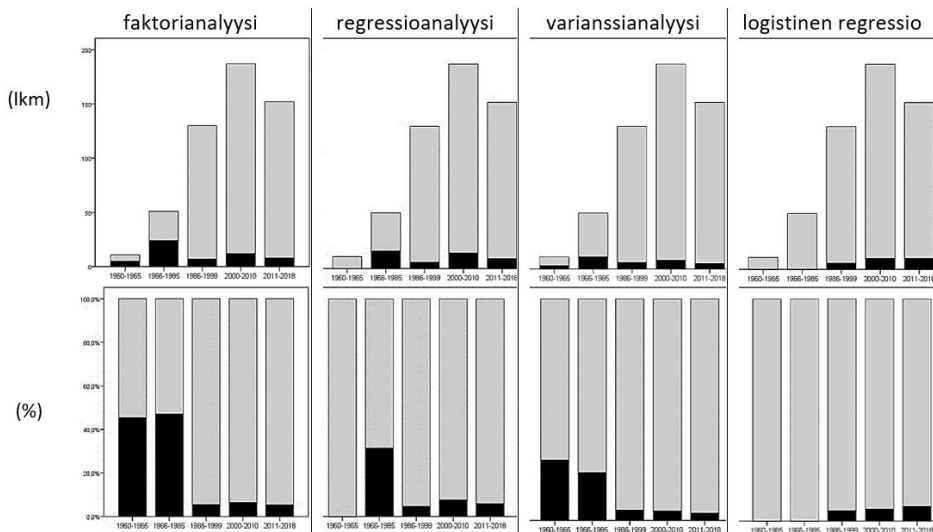
Taulukko 3 Sosiologian väitöskirjoissa käytetyt monimuuttujamenetelmät ja niiden käytön ajankohdat

| Kvantitatiivinen aineisto | 1960-1965 | 1966-1985 | 1986-1999 | 2000-2018 | Yhteensä |
|----------------------------------|------------------|------------------|------------------|------------------|-----------------|
| faktorianalyysi | 5 | 24 | 7 | 20 | 56 |
| lineaarinen regressioanalyysi | 0 | 16 | 5 | 14 | 35 |
| varianssianalyysi | 3 | 11 | 6 | 13 | 33 |
| logistinen regressioanalyysi | 0 | 0 | 6 | 16 | 22 |
| ryhmittelyanalyysi | 0 | 0 | 2 | 8 | 10 |
| log-lineaarinen analyysi | 0 | 1 | 6 | 0 | 7 |
| kanoninen analyysi | 0 | 5 | 0 | 0 | 5 |
| erotteluanalyysi | 0 | 4 | 0 | 1 | 5 |
| AID-analyysi | 0 | 4 | 0 | 0 | 4 |
| MCA-analyysi | 0 | 1 | 2 | 1 | 4 |
| LISREL | 0 | 0 | 1 | 1 | 2 |
| korrespondenssianalyysi | 0 | 0 | 1 | 1 | 2 |
| Yhteensä | 8 | 33 | 20 | 39 | 100 |
| Rekisteri-aineisto | 1960-1965 | 1966-1985 | 1986-1999 | 2000-2018 | Yhteensä |
| poisson-regressio | | | 1 | 11 | 12 |
| lineaarinen regressioanalyysi | | | 1 | 7 | 8 |
| cox-regressio | | | 0 | 8 | 8 |
| elinaika-analyysi | | | 3 | 1 | 4 |
| logistinen regressioanalyysi | | | 0 | 4 | 4 |
| log-lineaarinen analyysi | | | 1 | 1 | 2 |
| LISREL | | | 1 | 0 | 1 |
| Yhteensä | | | 6 | 17 | 23 |

Käytetyin monimuuttujamenetelmä sosiologian väitöskirjoissa on faktorianalyysi. Faktorianalyysia sisältävien väitöskirjojen määrä on 56 ja niiden osuus on 10,5 % kaikista sosiologian väitöskirjoista. Erityisen suosittua faktorianalyysi oli vuosina

1968 ja 1969. Tuolloin tarkastettiin seitsemän väitöskirjaa, joista jokaisessa on käytetty faktorianalyysia. 1970-luvulla faktorianalyysin käyttö väheni, mutta edelleen enemmistö sosiologian väitöskirjoista (60 %) sisälsi faktorianalyysia. 1980-luvulla faktorianalyysin suosio laski, tuolloin vain kahdessa väitöskirjassa on käytetty faktorianalyysia. Vuosien 2000–2009 aikana 12 väitöstutkijaa on käyttänyt faktorianalyysia tutkimuksessaan. Vuosina 2010–2018 faktorianalyysia esiintyy kahdeksassa väitöskirjassa.

Toiseksi käytetyin menetelmä on lineaarinen regressioanalyysi. Menetelmää on käytetty 43 väitöstutkimuksessa, joista kahdeksan liittyy rekisteriaineistojen analysointiin. Kolmanneksi käytetyin on varianssianalyysi, jota esiintyy 33 väitöskirjassa. Varianssianalyysin suosio sosiologien keskuudessa on ollut maltillista verrattuna kasvatustieteisiin, jossa se on nauttinut huomattavasti suurempaa suosiota. Lauri Nummenmaa (2004, 173) arvelee, että varianssianalyysi lienee yksi eniten käytetyistä tilastomenetelmistä kasvatustieteissä. Näiden kolmen menetelmän suosio perustuu osaltaan siihen, että ne olivat 1960-luvulla ainoat menetelmät, joita tietokoneilla pystyi suorittamaan, koska vain niille oli saatavissa tietokoneohjelmat (Alastalo 2014). Logistinen regressioanalyysi on uudempi menetelmä kuin edellä mainitut. Ensimmäinen logistista regressioanalyysia sisältävä väitöskirja on tarkastettu vuonna 1989.



Kuva 17 Neljän suosituimman monimuuttujamenetelmän esiintyminen sosiologian väitöskirjoissa 1960–2018, yläriivi esittää lukumäärät ja alariivi osuudet

Kuva 17 havainnollistaa faktorianalyysin, regressioanalyysin, varianssianalyysin ja logistisen regressioanalyysin suosion sosiologian väitöskirjoissa vuosina 1960–2018. Lukumäärätarkastelu paljastaa, että menetelmiä on käytetty määrälli-

sesti melko saman verran eri aikoina. Osuuksien tarkastelu tuo hyvin esiin faktorianalyysin 1960-luvulla vallinneen suosion. Myös regressioanalyysin suhteellinen suosio 1960-luvun jälkipuoliskolla tulee näkyviin. Menetelmien suhteellinen suosio 1980-luvulta alkaen on ollut jokseenkin vaatimatonta ja laskusuhdanteista, ainoastaan logistisen regressioanalyysin suosiossa on havaittavissa hienoista kasvua.

Monimuuttujamenetelmien esiintyminen on jaoteltu taulukossa 4 menetelmien sisältämään ongelmanasetteluun perustuen. Taulukossa on jälleen erotettu omaan sarakkeeseen rekisteripohjaisia aineistoja sisältävät väitöskirjat. On huomattava, että väitöskirjoissa on voitu käyttää useita eri menetelmiin pohjautuvia menetelmiä, minkä vuoksi taulukon 4 yhteensä-lukumäärien summa on suurempi kuin monimuuttujamenetelmiä sisältävien väitöskirjojen määrä.

Taulukko 4 Väitöskirjoissa esitetyt menetelmät ongelmanasettelunsa mukaan

| | Kvantitatiivinen aineisto | Rekisteriaineisto |
|-------------------------------------|---------------------------|-------------------|
| Tiivistäminen, ja ryhmittely | | |
| faktorianalyysi | 56 | 0 |
| ryhmittelyanalyysi | 10 | 0 |
| korrespondenssianalyysi | 2 | 0 |
| Yhteensä | 68 | 0 |
| Selittäminen | | |
| lineaarinen regressioanalyysi | 35 | 8 |
| logistinen regressioanalyysi | 22 | 4 |
| log-lineaarinen analyysi | 7 | 2 |
| Poisson-regressio | 0 | 12 |
| Cox-regressio | 0 | 8 |
| Yhteensä | 64 | 34 |
| Ryhmiä vertailu | | |
| variassianalyysi | 33 | 0 |
| erotteluanalyysi | 5 | 0 |
| AID-analyysi eli puuanalyysi | 4 | 0 |
| kanoninen analyysi | 5 | 0 |
| Yhteensä | 47 | 0 |
| Monia ongelmanasetteluja | | |
| Lineaariset rakenneyhtälömallit | 2 | 1 |
| MCA-moniluokitteluanalyysi | 4 | 0 |
| transformaatioanalyysi | 3 | 0 |
| Yhteensä | 9 | 1 |

Ongelmanasettelultaan tietoa tiivistäviä menetelmiä on käytetty 68 väitöskirjassa. Rekisteriaineistojen analysoinnissa ei tähän ryhmään kuuluvia menetelmiä ole käytetty lainkaan. Suosituin tiivistäviin menetelmiin lukeutuva monimuuttujamenetelmä on ollut faktorianalyysi. Ryhmittelyanalyysia on käytetty yllättävän vähän siihen nähden, että erilaisten vastaajaryhmien etsiminen voisi hyvin sopia sosiologiseksi tutkimustavoitteeksi. Vähäisen käytön syinä ovat luultavasti menetelmän nuoruus ja se, että menetelmä on huonosti tunnettu. Vähäisen käytön syinä

on luultavasti myös se, että väitöskirjoissa on valittu menetelmäksi faktorianalyysi tilanteissa, joissa tutkimuskysymyksen, perustelujen ja tulkinnan perusteella menetelmäksi olisi sopinut ryhmittelyanalyysi. Vaikka ryhmittelyanalyysi ja faktorianalyysi eroavat toisistaan tilastomatematisesti, tuottavat ne kvantitatiivisen aineiston analyysissa lopulta hyvin samankaltaisia tuloksia. Vähiten käytetty menetelmä on ollut korrespondenssianalyysi.

Jos katsotaan kvantitatiivisiin, ei-rekisteripohjaisiin, aineistoihin käytettyjä monimuuttujamenetelmiä, ovat toisella sijalla ovat asetelmaltaan selittämiseen pohjautuvat menetelmät. Jos taas katsotaan kaikkia monimuuttujamenetelmiä sisältäviä väitöskirjoja, ovat selittämiseen perustuvat menetelmät suosituimpia. Kvantitatiivisia aineistoja käyttäneissä väitöskirjoissa yleisimmät selittävät menetelmät ovat lineaarinen ja logistinen regressioanalyysi. Rekisteriaineistoja hyödyntävät väitöstutkimukset puolestaan ovat käyttäneet erityisesti Poisson-regressioanalyysia, Cox-regressioanalyysia mutta myös lineaarista regressioanalyysia.

Kolmannella sijalla taulukossa 4 ovat vastaajaryhmien analysointiin liittyvät menetelmät, joita on käytetty 47 väitöskirjassa. Monia ongelmanasetteluja sisältäviä menetelmiä on käytetty vähiten. Syynä vähäiseen käyttöön ei varmastikaan ole se, että menetelmät olisi koettu hyödyttömiksi vaan se, että tähän ryhmään kuuluvat menetelmät ovat kohtalaisen uusia ja vaativat erityisiä tilasto-ohjelmia.

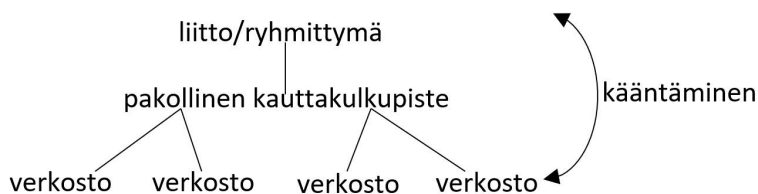
6. Rajakohdemalli ja tieteiden välinen yhteistyö, esimerkkinä faktorianalyysi

6.1. Rajakohdemalli

Star ja Griesemer (1989) ovat tutkineet tieteiden välistä yhteistyötä. Lähtökohtana heillä on ollut Callonin, Latourin ja Lawin kehittämä toimijaverkkoteorian malli. Ylikoski (2000) määrittelee toimijaverkkoteorian analyysityökaluksi tai käsitteistöksi, joka auttaa tutkijaa vapautumaan liian rajoittavista erotteluista esimerkiksi luonnollisen ja sosiaalisen välillä ja näkemään toimijoiden maailmaa ilman yhteiskuntatieteen perinteen muodostamaa taakkaa. Ylikoski pohjaa kuvauksensa toimijaverkkoteoriasta Latouriin⁹ ja esittää seuraavan kuvauksen:

Keskeisenä tutkimuskohteena on toimijuus verkostossa. Toimijat yrittävät värvätä muita toimijoita verkostoonsa vahvistaakseen sitä. Onnistuessaan toimija saavuttaa *pakollisen kauttakulkupisteen aseman*. Uusien toimijoiden synnyttämistä ja olemassa olevien muokkaamista kutsutaan *kääntämiseksi*. Rihmaston analyysi on siis periaatteessa aina valta-analyysia. (Ylikoski 2000, 300–301.)

Toimijaverkkoteorian mallissa pakollisia kauttakulkupisteitä on vain yksi ja kääntäminen tapahtuu kertaalleen (kuva 18).

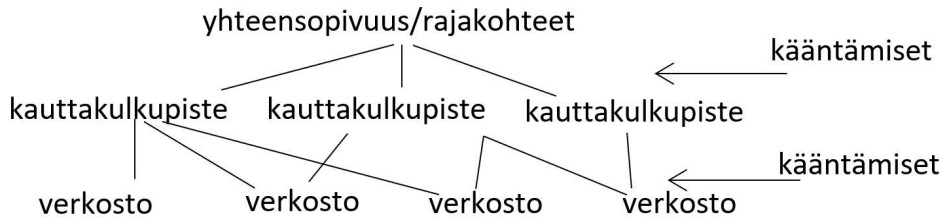


Kuva 18 Callonin, Latourin ja Lawin kehittämä toimijaverkkoteorian malli. Lähde: Star ja Griesemer (1989)

Star ja Griesemer (1989, 388–390) ovat muokanneet Callonin, Latourin ja Lawin mallia lisäämällä siihen useita kauttakulkupisteitä, joista verkostot pystyvät myös valitsemaan useita. Tapoja, millä tavalla toimijat voivat muodostaa pakollisia

⁹ Latour, Bruno (1987) *Science in Action*. Open University Press, Milton Keynes. 71–74, 118–19; Latour, Bruno (1988) *The Pasteurization of France*. Harvard University Press, Cambridge. 195–201.

kauttakulkupisteitä on olemassa Starin ja Griesemerin mukaan rajattomasti ja sen myötä kääntämisiäkin voi tapahtua lukematon määrä (kuva 19).



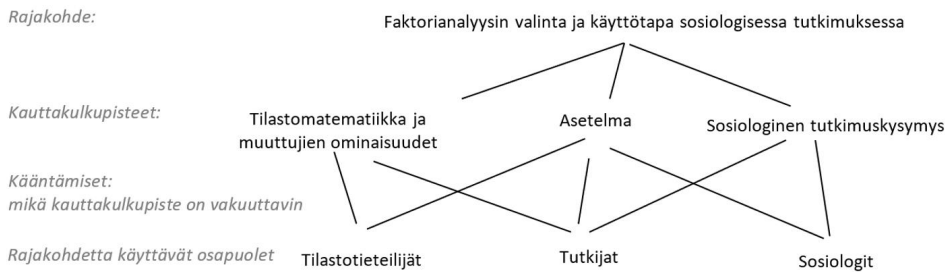
Kuva 19 Starin ja Griesemerin muokkaama malli. Lähde: Star ja Griesemer (1989)

Verkostojen sijaan Star ja Griesemer (1989, 412; Star 2010) puhuvat sosiaalisista maailmoista. Järjestelyjä, jotka mahdollistavat erilaisten sosiaalisten maailmojen yhteistyön itsenäisinä ja ilman varsinaista yksimielisyyttä, kutsutaan kääntämiseksi. Kääntämisen keinoja on monia ja niiden avulla luodaan niin sanottuja rajakohteita. Kääntäminen voi tapahtua esimerkiksi siten, että sosiaaliset maailmat yrittävät löytää näkemyksistään pienimmät yhteiset tekijät. Toinen kääntämisen tapa on etsiä sellaisia muuntautuvia tekijöitä, joita osapuolet pystyvät muokkaamaan tarkoituksiinsa sopiviksi. Kolmas kääntämisen tapa on abstrahoida tai yksinkertaistaa yhteinen kohde omaan tarkoitukseen sopivaksi ja jättää sopimattomat ominaisuudet huomiotta. (Star ja Griesemer 1989, 412; Star 2010.) Star ja Griesemer määrittävät rajakohteen seuraavasti:

Rajakohteet ovat kohteita, jotka ovat riittävän joustavia mukautumaan usean niitä tarvitsevan osapuolen paikallisiin tarpeisiin ja rajoituksiin, ja silti ne ovat riittävän robusteja säilyttääkseen yleisen identiteettinsä. Yleisessä käytössä ne ovat väljästi strukturoituja ja tulevat vahvarakenteisiksi yksilökäytössä. Ne voivat olla abstrakteja tai konkreettisia. Niillä on erilainen merkitys erilaisissa sosiaalisissa maailmoissa, mutta niiden rakenne on riittävän yleinen useampaan maailmaan tehdäkseen niistä tunnistettavia. Rajakohteiden luominen ja hallinta on avain koherenssin kehittämiseen ja säilyttämiseen toisensa leikkaavissa sosiaalisissa maailmoissa. (Star ja Griesemer 1989, 388.)

Rajakohteet tarjoavat erilaisen tietämyksen omaaville tai erilaisissa yhteisöissä työskenteleville yksilöille keinoja tietojensa esittämiseen, kyseenalaistamiseen ja muuntamiseen (Carlile 2002, 451–452). Rajakohteiden tarkoituksena on maksimoida sekä autonomia että kommunikaatio sosiaalisten maailmojen välillä. Rajakohteelta ja sen liittymäpinnoilta eri sosiaalisiin maailmoihin vaaditaan sisällön ja esitysmuodon huolellista harkintaa. (Star 2010.) Starin mukaan rajakohteeseen liitetään usein tulkinnallinen joustavuus.

Omassa tutkimuksessani asetan rajakohteeksi faktorianalyysin valinnan ja käytettävän sosiologisessa tutkimuksessa (kuva 20). Kauttakulkupisteitä, jotka voidaan ajatella myös näkökulmiksi, on kolme. Tilastomatematisesta näkökulmasta faktorianalyysin valintaperusteet ja käyttötavat liittyvät muuttujien ominaisuuksiin ja niiden matemaattisiin operaatioihin. Asetelman näkökulmasta katsottuna faktorianalyysin valinta ja käyttö nähdään muuttujien määrää vähentävänä, tietoa tiivistävänä ja muuttujia niputtavana menetelmänä. Sosiologisen tutkimuskysymyksen näkökulmasta faktorianalyysi nähdään keinona ottaa haltuun käsitteellinen tarkastelu.



Kuva 20 Faktorianalyysi rajakohteena

Seuraavaksi esittelen faktorianalyysia näistä kolmesta näkökulmasta. Luvussa 6.2 kauttakulkupisteenä on tilastomatematiikkaan ja muuttujien ominaisuuksiin perustuva näkökulma. Luvussa 6.3. esittelen faktorianalyysin asetelmaan perustuvan kauttakulkupisteen ja luvussa 6.4. sosiologiseen tutkimuskysymykseen perustuvan kauttakulkupisteen.

6.2. Faktorianalyysin tilastomatematiikka

Kovarianssi ja korrelaatio

Faktorianalyysin tilastomatematisena tavoitteena on löytää sellaiset faktorilataukset, että ne selittävät mahdollisimman hyvin muuttujien välisiä kovariansseja tai korrelaatioita. (Leskinen 1987; Nummenmaa ym. 1997, 242.)

Faktorianalyysin lähtökohtana on siten tarkastella muuttujien välisiä riippuvuuksia, jotka on esitetty joko kovarianssien¹⁰ tai korrelaatioiden avulla. Taulukossa 5 on kovarianssimatriisi, joka esittää neljän muuttujan väliset kovarianssit¹¹. Kovariansseja

10 y -muuttujien teoreettinen kovarianssi ilmaistaan kaavalla $\text{kovarianssi}(y_i, y_j) = \text{odotusarvo}(y_i y_j)$ eli $\text{cov}(y_i, y_j) = E y_i y_j$. Teoreettinen varianssi ilmaistaan kaavalla $\text{varianssi}(y_i) = \text{odotusarvo}(y_i^2)$ eli $\text{var}(y_i) = E y_i^2$.

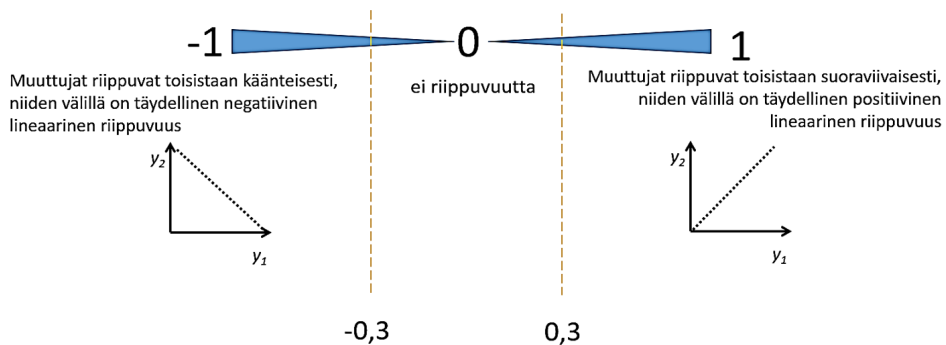
11 Muuttujan kovarianssi itsensä kanssa on nimeltään varianssi.

laskettaessa muuttujien järjestyksellä ei ole merkitystä, eli esimerkiksi kovarianssi_{y₂y₁} on sama kuin kovarianssi_{y₁y₂}. Sen seurauksena kovarianssimatriisi on symmetrinen diagonaalien suhteen ja kovarianssimatriisissa on riittävää esittää joko diagonaalien yläpuolella tai alapuolella olevat tiedot.

Taulukko 5 Neljän y-muuttujan kovarianssimatriisin alakolmioesitys

| | y ₁ | y ₂ | y ₃ | y ₄ |
|----------------|---|---|---|---|
| y ₁ | varianssi _{y₁y₁} | | | |
| y ₂ | kovarianssi _{y₂y₁} | varianssi _{y₂y₂} | | |
| y ₃ | kovarianssi _{y₃y₁} | kovarianssi _{y₃y₂} | varianssi _{y₃y₃} | |
| y ₄ | kovarianssi _{y₄y₁} | kovarianssi _{y₄y₂} | kovarianssi _{y₄y₃} | varianssi _{y₄y₄} |

Kovarianssi ei ole empiiristä tutkimusta tekevälle useinkaan käyttökelpoinen muuttujien riippuvuusmitta, koska muuttujien mittayksiköt vaikuttavat kovarianssin arvoon ja siten eri tyyppisten muuttujien välisiä kovariansseja ei voi vertailla keskenään (Taanila 2020). Kovarianssista muunnettu korrelaatiokerroin¹² sen sijaan on muuttujien mittayksiköistä riippumaton tunnusluku. Korrelaatiokertoimen¹³ arvot rajoittuvat -1:n ja 1:n välillä nollan ilmaistessa riippumattomuutta (kuva 21).



Kuva 21 (Pearsonin) korrelaatiokertoimen tulkinta

Tutkijoiden pohdittavaksi jää se, minkä suuruiset kertoimet pitäisi tulkita riippuvuutta osoittaviksi. Kuvaan 21 on merkitty raja-arvoiksi -0,3 ja 0,3, mutta tutkijan

$$12 \text{ korrelaatio}_{y_1y_2} = \frac{\text{kovarianssi}_{y_1y_2}}{\text{keskihajonta}_{y_1} \text{keskihajonta}_{y_2}}$$

13 Tunnetuimmat korrelaatiokertoimet ovat Pearsonin korrelaatiokerroin ja Spearmanin korrelaatiokerroin. Pearsonin korrelaatiokerroin ilmaisee, kuinka vahva lineaarinen riippuvuus muuttujien välillä on. Spearmanin korrelaatiokerroin sopii myös epälineaarisen riippuvuuden tarkasteluun.

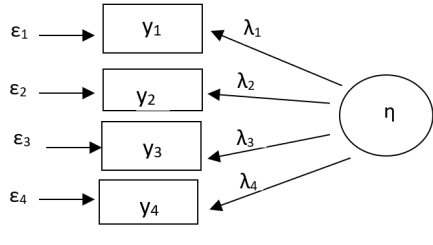
on korrelaatioita tulkitessaan huomioitava monia asioita, esimerkiksi aineiston koko ja muuttujien ominaisuudet.

Yhden faktorin malli matriisilaskentana

Avaan seuraavaksi faktorianalyysia matriisilaskentana yhden faktorin mallin avulla. Esitykseni nojautuu Leskisen (1987) teokseen. Monta faktoria sisältävän faktorimallin matriisilaskentaan voi tutustua esimerkiksi Harry H. Harmanin (1976) teoksessa.

Kuvassa 22 esitetään yhden faktorin faktorimalli. Faktoria merkitään kreikkalaisella kirjaimella η (eta). Faktorimallissa oletetaan, että faktori vaikuttaa muuttujiin, mitä graafisessa mallissa symboloivat nuolien suunnat faktorista muuttujiin. Faktorilataukset λ_i (lamda) ilmaisevat faktorin vaikutuksen voimakkuutta kuhunkin muuttujaan. Muuttujien arvot eivät ole kuitenkaan täysin selitettävissä faktorin avulla. Selittämättä jääneitä osuuksia kutsutaan jäännöstermeiksi ja niiden symbolina ovat ε_i :t (epsilon).

Kuvan 22 oikea puoli esittää faktorimallin matemaattisina yhtälöinä. Kunkin muuttujan y_1, y_2, y_3 , ja y_4 arvon katsotaan muodostuvan sen saamasta faktorilatauksesta ($\lambda_1, \lambda_2, \lambda_3$ tai λ_4) kerrottuna faktori η :lla ja tuloon lisätynä jäännöstermi ($\varepsilon_1, \varepsilon_2, \varepsilon_3$ tai ε_4).

| Yhden faktorin malli graafisesti | Yhden faktorin malli matemaattisesti |
|---|---|
|  | $y_1 = \lambda_1\eta + \varepsilon_1$ $y_2 = \lambda_2\eta + \varepsilon_2$ $y_3 = \lambda_3\eta + \varepsilon_3$ $y_4 = \lambda_4\eta + \varepsilon_4$ |

Kuva 22 Faktorimalli graafisesti ja yhtälömuodossa

Kuten luvun alussa jo tuotiin esiin, on faktorianalyysin tavoitteena löytää sellaiset faktorilataukset $\lambda_1, \lambda_2, \lambda_3$ ja λ_4 , että ne selittävät mahdollisimman hyvin muuttujien y_1, y_2, y_3 ja y_4 välisiä kovariansseja tai korrelaatioita (Leskinen 1987; Nummenmaa ym. 1997, 242). Laskutoimitusten mahdollistamiseksi joudutaan tekemään taulukossa 6 esiin tuodut tilastomatemattiset oletukset.

Taulukko 6 Faktorimallin tilastomatematitiset oletukset

| Oletukset | Tilastomatematitinen merkintä |
|---|---|
| Faktorin ja mittausvirheiden väliset kovarianssit ja korrelaatiot ovat nollia | $E\eta\varepsilon_i = 0$ |
| Mittausvirheet eivät korreloi keskenään | $E\varepsilon_i\varepsilon_j = 0, i \neq j$ |
| Merkintätavat | |
| Merkitään η -faktorin varianssia ω :lla | $\text{var}(\eta) = E\eta^2 = \omega$ |
| Merkitään ε_i -mittausvirheiden variansseja θ_i :llä | $\text{var}(\varepsilon_i) = E\varepsilon_i^2 = \theta_i$ |

Muuttujien teoreettisen kovarianssin ja varianssin laskukaavat esitettiin alaviitteessä 9. Kun niihin sijoitetaan faktorimallin yhtälö $y_i = \lambda_i\eta + \varepsilon_i$, saadaan tulokseksi, että muuttujien kovarianssi on ilmaistavissa faktorilatausten ($\lambda_i\lambda_j$) ja faktorin varianssin (ω) avulla¹⁴ ja muuttujan varianssi on ilmaistavissa faktorilatauksen (λ_i^2), faktorin varianssin (ω) ja jäännösvaihtelun (θ_i) avulla¹⁵.

Faktorilatauksille tai faktorin varianssille on asetettava tilastomatematitisia rajoituksia, koska ilman niitä syntyisi lukematon määrä faktorimalliesityksiä, jotka tuottaisivat saman kovarianssimatriisin. Eräs rajoitusvaihtoehto on määrittää faktorin varianssi ykkösen suuruiseksi. Tällainen rajoitus saa aikaan sen, että faktorimalli on faktorilatausten etumerkkiä lukuun ottamatta yksikäsitteinen. Faktorin varianssin kiinnittäminen ykköseksi merkitään tilastomatematitisesti $\text{var}(\eta) = \omega = 1$. Sen myötä taulukossa 7 esitettävä y -muuttujan teoreettinen varianssi oikenee muotoon $\text{var}(y_i) = \lambda_i^2 + \theta_i$. Teoreettinen kovarianssi oikenee muotoon $\text{cov}(y_i, y_j) = \lambda_i\lambda_j$. Näin päästiin ilmaisemaan muuttujien riippuvuudet faktorilatausten λ avulla.

Taulukko 7 y -muuttujien teoreettinen kovarianssimatriisi kolmella tavalla esitettynä

| | | | | |
|--|--|--|--|--------------------------|
| Teoreettinen kovarianssimatriisi | $\text{var}(y_{11})$ $\text{cov}(y_2, y_1)$ $\text{cov}(y_3, y_1)$ $\text{cov}(y_4, y_1)$ | $\text{var}(y_2)$ $\text{cov}(y_3, y_2)$ $\text{cov}(y_4, y_2)$ | $\text{var}(y_3)$ $\text{cov}(y_4, y_3)$ | $\text{var}(y_4)$ |
| Käytetään laskukaavoissa. | σ_{11} σ_{21} σ_{31} σ_{41} | σ_{22} σ_{32} σ_{42} | σ_{33} σ_{43} | σ_{44} |
| Teoreettinen kovarianssimatriisi faktorilatauksin ja jäännöstermein esitettynä | $\lambda_1^2 + \theta_1$ $\lambda_2\lambda_1$ $\lambda_3\lambda_1$ $\lambda_4\lambda_1$ | $\lambda_2^2 + \theta_2$ $\lambda_3\lambda_2$ $\lambda_4\lambda_2$ | $\lambda_3^2 + \theta_3$ $\lambda_4\lambda_3$ | $\lambda_4^2 + \theta_4$ |

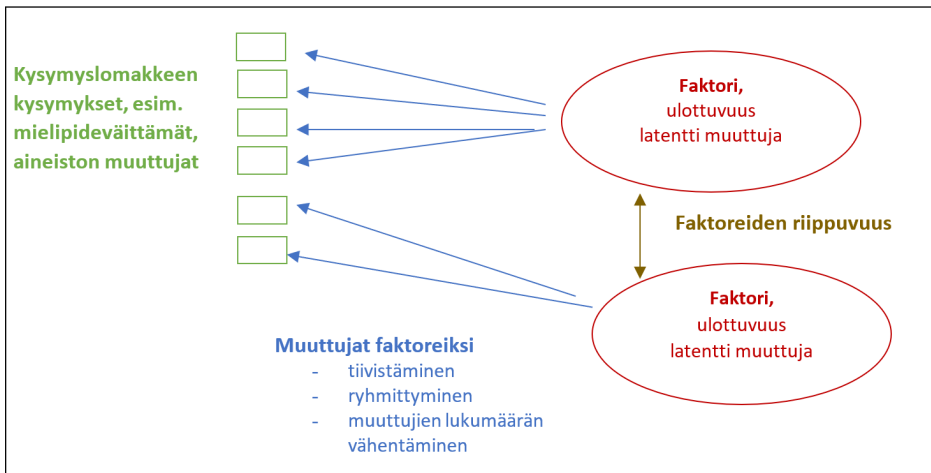
¹⁴ $\text{cov}(y_i, y_j) = \sigma_{ij} = E y_i y_j = E(\lambda_i\eta + \varepsilon_i)(\lambda_j\eta + \varepsilon_j) = \lambda_i\lambda_j E\eta^2 + \lambda_i E\eta\varepsilon_j + \lambda_j E\varepsilon_i\eta + E\varepsilon_i\varepsilon_j = \lambda_i\lambda_j\omega$

¹⁵ $\text{var}(y_i) = \sigma_{ii} = E y_i^2 = E(\lambda_i\eta + \varepsilon_i)^2 = \lambda_i^2 E\eta^2 + E\varepsilon_i^2 + 2\lambda_i E\eta\varepsilon_i = \lambda_i^2\omega + \theta_i$

Faktorianalyysissä muuttujien väliset riippuvuudet voidaan näin kuvata faktorilatausten avulla. Muuttujien varianssit kuvataan faktorilatausten neliöiden λ_i^2 ja jäännösvaihtelun θ_i avulla siten, että λ_i^2 ilmaisee faktorin ”selittämää” osaa y-muuttujan vaihtelusta ja θ_i ilmaisee jäännösvaihtelun, esimerkiksi mittausvirheen osuutta y_i :n vaihtelussa. Faktorilataus λ_i puolestaan kuvaa muuttujan y_i ja faktorin välistä kovarianssia. Jos faktorimalliesitysten perustana on korrelaatiomatriisi, ilmaisevat lataukset λ_i havaitun muuttujan y_i ja faktorin η välisen korrelaation. (Leskinen 1987.)

6.3. Asetelmaan perustuva kauttakulkupiste

Asetelmaan perustuva kauttakulkupiste sijaitsee rajakohdemallissa (kuva 20) tilastomatemattisen kauttakulkupisteen ja sosiologiseen tutkimuskysymykseen pohjaavan kauttakulkupisteen välissä. Sijainnilla ilmaistaan yhtäältä sitä, että asetelma-kauttakulkupiste mahdollistaa sekä tilastotieteellisesti että sosiologisesti painottuneita näkökulmia faktorianalyysiin ja toisaalta se tarjoaa myös mahdollisuuden yhteiseen näkökulmaan. Kuvassa 23 on kahden faktorin malli, jota seuraavaksi avaan asetelmallisena kauttakulkupisteenä.



Kuva 23 Faktorimallin asetelma

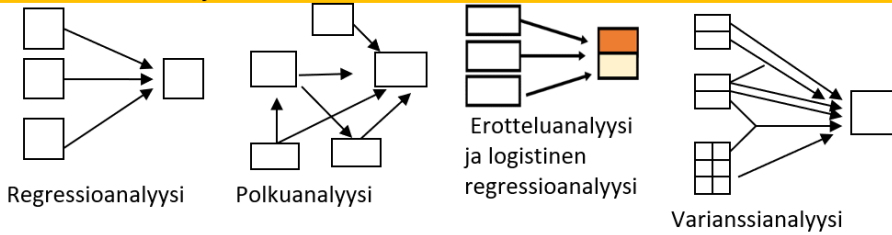
Tilastomatemattisesti faktorianalyysin tavoitteena on löytää sellainen faktorilatausmatriisi, joka selittää mahdollisimman hyvin muuttujien välistä kovarianssi- tai korrelaatiomatriisia. Faktorianalyysin kohteena olevat muuttujat edustavat tilastotieteilijälle matemaattisia mittaustuloksia, joiden osalta kiinnostus kohdistuu mittaustulosten keskimääräisyyden, vaihtelun ja harhattomuuden tarkasteluun.

Sosiologille muuttujat edustavat kysymyslomakkeen kysymyksiä ja niiden toivotaan heijastavan sisällöllisiä piirteitä tutkimuksen kohteena olevasta ilmiöstä. Sosiologisen tutkimuskysymyksen näkökulmasta faktorianalyysin tavoitteena on selvittää, kuinka monta kysymyskokonaisuutta syntyy ja mitkä kysymykset niputuvat yhteen kokonaisuuteen. Oleellista on arvioida kysymyskokonaisuuksien sisällöllinen järkevyyden ja pohtia, onko kysymyskokonaisuudelle löydettävissä yhteinen nimittäjä tai nimi.

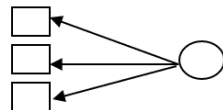
Asetelmallisessa kauttakulkupisteessä kiinnostuksen kohteena on selvittää muuttujien riippuvuuden taustalla oleva rakenne. Tämä tarjoaa sosiologille ja tilastotieteilijälle yhteisen näkökulman. Faktorianalyysi on Hairin ym. (1998, 87–138; 2014, 89–149) mukaan yleisnimi sellaiselle monimuuttujamenetelmien ryhmälle, jonka ensisijaisena tarkoituksena on määrittää aineiston taustalla oleva rakenne. Faktorianalyysin avulla tutkija voi ensin identifoida rakenteen eri ulottuvuudet ja sen jälkeen päätellä, kuinka vahvasti kukin muuttuja selittyy kullakin ulottuvuudella. Yhteisen ymmärryksen ja kiinnostuksen sosiologi ja tilastotieteilijä jakavat siis sen selvittämisessä, kuinka monta faktoria syntyy ja mitkä muuttujat kuuluvat mihinkin faktoriin. Asetelmaan perustuvassa kauttakulkupisteessä faktorianalyysi nähdään rakennetta etsivänä ja muuttujia niputtavana menetelmänä.

Kuten kuvassa 23 esitetään, voidaan faktoreita kutsua myös ulottuvuuksiksi tai latenteiksi muuttujiksi. Tacq (1997, 35–55) esittää monimuuttujamenetelmille useita erilaisia jakoperusteita. Eräs monimuuttujamenetelmien jaottelutapa perustuu latentin muuttujan olemassaoloon ja määrään (kuva 24). Menetelmiä, jotka eivät sisällä latenttia muuttujaa, ovat regressioanalyysi, polkuanalyysi, yksisuuntainen varianssianalyysi (ANOVA) sekä erotteluanalyysi. Yhden latentin muuttujan sisältäväksi menetelmäksi Tacq esittää faktorianalyysin silloin, kun mallissa on vain yksi faktori. Useita latenteja muuttujia sisältäväksi menetelmäksi Tacq lukee faktorianalyysin siinä tapauksessa, kun mallissa on monta faktoria. Muita useita latenteja muuttujia sisältäviä menetelmiä ovat Tacqin mukaan usean kertaluvun faktorimallit sekä lineaariset rakenneyhtälömallit. Neljäs ryhmä muodostuu menetelmistä, joiden asetelmaan sisältyy latenttien muuttujien ryhmä. Tähän sisältyvinä menetelminä Tacq mainitsee esimerkiksi kanonisen korrelaatioanalyysin. (Tacq 1997, 48–50.)

Ei latentia muuttujaa

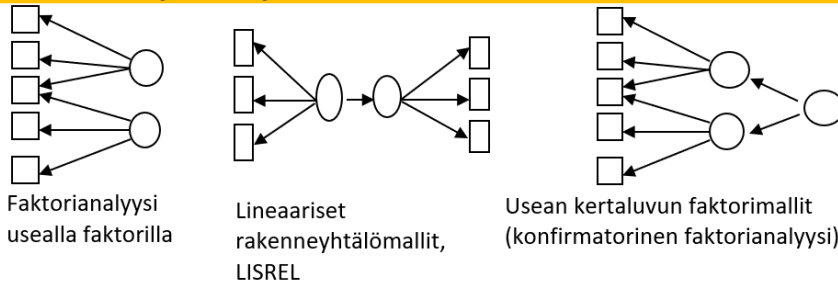


Yksi latentti muuttuja



Faktorianalyysi yhdellä faktorilla

Useita latentteja muuttujia



Kuva 24 Monimuuttujamenetelmien jaottelu latentin muuttujan olemassaolon mukaan. Lähde: Tacq (1997)

Latentin muuttujan olemassaolon ja määrään perustuva jaottelu palvelee hyvin kvantitatiivisia sosiologisia tutkimuksia, koska niissä on usein tarpeen etsiä empiirisiä vastineita tutkimuksessa käytettäville teoreettisille käsitteille.

6.4. Sosiologiseen tutkimuskysymykseen perustuva kauttakulkupiste

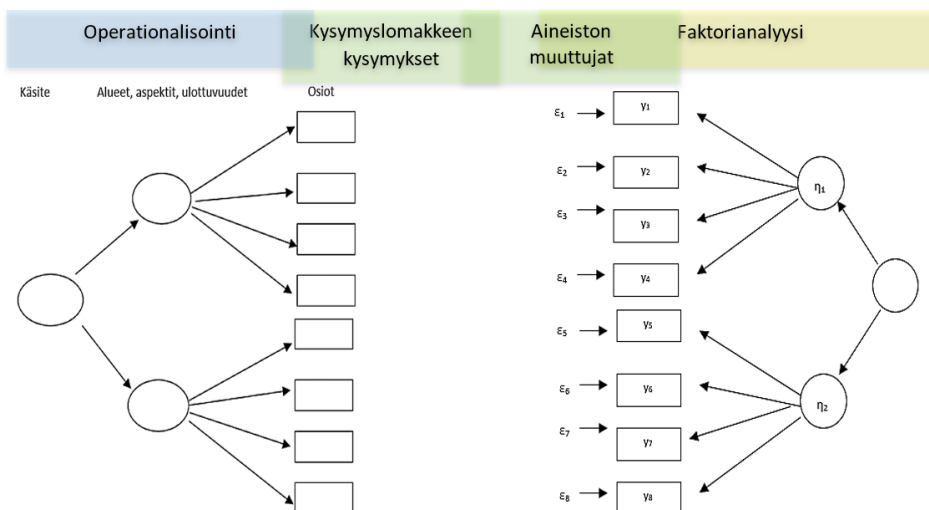
Sosiologiassa tutkimuskohteet määritellään usein käsitteellisten hahmotusten varassa (Alkula ym. 1994, 10; Pöntinen 2004, 71). Tutkimusaineiston keruuvaiheessa käsitteelliset hahmotukset on jollakin tavalla pyrittävä muuttamaan esimerkiksi kysymyslomakkeen kysymyksiksi tai tutkimusaineiston numeroilla kuvattaviksi muuttujiksi. Prosessointia, jonka avulla teoreettiset käsitteet pyritään yhdistämään kohteessa havaittaviin mitattaviin ominaisuuksiin, kutsutaan operationalisoinniksi (Jokivuori ja Hietala 2007, 182). Alkula ym. (1994, 80) korostavat operationalisoinnin tärkeyttä sosiologisessa tutkimuksessa. Tutkijan on mietittävä, mitä hänen

tutkimuksessaan käyttämä käsite merkitsee ja minkälaisista osista se koostuu. Operationalisointiprosessissa tutkija etenee vaihe vaiheelta käsitteestä osakäsitteisiin ja edelleen pienemmiksi ja konkreettisimmiksi käsitteiksi, kunnes pystyy muotoilemaan niistä konkreettisia kysymyksiä. (Alkula ym. 1994, 80.) Operationalisointia tehdessään tutkija tuo myös näkyväksi, miten hän on määritellyt tutkimuksensa käsitteet. Operationalisointi on näin ollen empiirisen sosiaalitutkimuksen ydinalueetta sen toimiessa sekä tutkijan että tutkimuksen lukijoiden kehikkona käsitteiden viidakossa. (Porter 1995, 125; Alkula ym. 1994, 75–76; Ketokivi 2009, 43–46.) Operationalisointikeskusteluun voi tutustua esimerkiksi Mäkelän (1991; 1996, 64–88) ja Alastalon (2005, 150–189) tekstien kautta.

Operationalisoinnin taustalla on muuttujakieli, jonka Lazarsfeld kehitti teoreettisten käsitteiden ja latenttien muuttujien analysoimiseksi (Selkälä 2013, 110). Muuttujakielen pyrkimyksenä oli kääntää teoreettiset käsitteet vaiheittain empiiriksi indikaattoreiksi (ks. Lazarsfeld 1993b, 236–254). Selkälä kuvaa Lazarsfeldin muuttujakielen vaiheita seuraavasti:

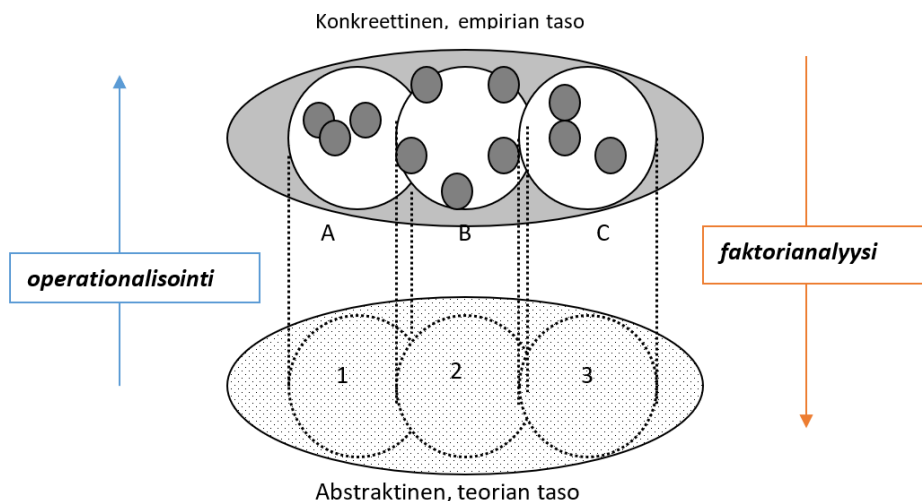
Ensimmäisessä vaiheessa käsitteestä muodostetaan jonkinlainen mielikuva teoreettisten tarkastelujen perusteella. Toisessa vaiheessa käsitettä täsmennetään rikkomalla se useaksi osaksi. Kolmannessa vaiheessa etsitään kutakin ulottuvuutta vastaavia indikaattoreita, jotka lopulta yhdistetään indekseiksi. Näin saadut indeksit ovat käsitteiden empiirisiä vastineita. (Selkälä 2013, 110.)

Lazarsfeldin muuttujakielen vaiheet on tunnistettavissa myös Nummenmaan ym. (1997, 213; 265) esityksessä. Katsottaessa kuvaa 25 vasemmalta oikealle lähdetään liikkeelle käsitteestä. Käsite jaetaan alueisiin, aspekteihin tai ulottuuksiin, mikä on muuttujakielen toinen vaihe. Muuttujakielen kolmas vaihe – ulottuvuuksia vastaavat indikaattorit – nimetään kuvassa 25 osioiksi. Indikaattorit tai osiot ovat kysymyslomakkeen kysymyksiä, jotka tallennetaan aineistoon tilastollisiksi muuttujiksi. Muuttujakielen neljäs vaihe – indikaattoreiden yhdistäminen indekseiksi – vastaa kuvassa 25 oikeaa puolta, jossa Nummenmaa ym. kuvaavat faktorianalyysia. Ideaalitulanteessa faktorianalyysi tuottaisi käsitteen ulottuvuuksia vastaavat faktorit η_1 ja η_2 , jotka ehkä vielä yhdistyisivät yhdeksi - toisen kertaluvun - faktoriksi. Näin operationalisoinnin ja faktorianalyysin voi nähdä toistensa peilikuvana tai käänteistoimenpiteenä.



Kuva 25 Vasemmalla operationalisointi eli mittauskohteen spesifiointi teoreettisesta konstruktiosta mittareihin. Oikealla faktorianalyysi graafisesti kuvattuna. (Lähde: Nummenmaa ym. 1997, 213; 265)

Faktorianalyysin toimivuus operationalisoinnin käännteistomenpiteenä on näkyvissä myös kuvassa 26, jonka pohjana on Alkulan ym. (1994, 80–81) esitys operationalisoinnista.



Kuva 26 Faktorianalyysi operationalisoinnin käännteistomenpiteenä (Lähde: Alkula ym. 1994)

Pöntinen (2004, 79) toteaa, että käsitteellisiä hahmotuksia käyttäessään sosiologin on ensin operationalisoitava ja pilkottava käsite lukuisiin konkreettisiin, mitattavissa oleviin osa-alueisiin. Analyysivaiheessa hänen on tiivistettävä osa-alueet faktorianalyysia hyödyntäen takaisin suurempiin kokonaisuuksiin ja faktoripistemäärien tai summamuuttujien avulla luotava niistä aineistoon empiirisiä muuttujia.

7. Sosiologian väitöskirjoissa esitetyt faktorianalyysiperustelut

7.1. Perustelujen määrällisyys ja laadullisuus

Miksi faktorianalyysia käytetään sosiologian väitöskirjoissa? Mitä faktorianalyysin käytöllä haetaan osana tutkimusta? Vastaan näihin kysymyksiin tarkastelemalla sosiologien esittämiä faktorianalyysiperusteluja. Taulukkoon 8 on poimittu sosiologian väitöskirjoissa eri aikoina esitettyjä faktorianalyysiperusteluja. Sisällönanalyysia hyödyntäen olen nostanut perusteluista tiivistettynä esiin sen, mihin faktorianalyysia on haluttu käyttää. Olen hyödyntänyt myös rajakohdeteoriasta ja tieteiden välisen yhteistyön tutkimisesta kirjoittaneen Starin (2010, 605) ohjeistusta. Hän on neuvonut kiinnittämään huomiota osapuolten käyttämiin kielikiviin, ilmauksiin ja sanoihin, ja niiden tiivistetyistä ilmauksista.

Taulukko 8 Sosiologien faktorianalyysiperusteluja ja niiden tiivistetyt ilmaukset

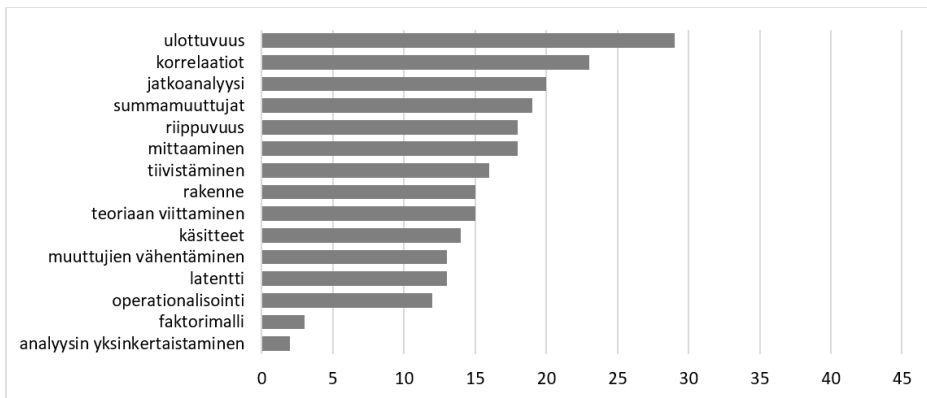
| Alkuperäinen ilmaus sosiologian väitöskirjassa | Tiivistetty ilmaus |
|--|---|
| Faktorianalyysilla on mahdollisuus testata kolmen oletetun ulottuvuuden riippuvuutta toisistaan. (Piepponen 1960.) | ulottuvuus riippuvuus |
| Vaikka mukaan otettujen muuttujien keskinäiset korrelaatiot ovat enimmäkseen hyvin pieniä, on (faktori)analyysin katsottu valottavan esillä olevien asenteiden rakennetta. (Asp 1965.) | korrelaatiot riippuvuus rakenne |
| ...alun perin käytetyt asteikot voidaan korvata pienemmällä joukolla uusia ulottuvuuksia... (Randell 1965.) | tiivistäminen ulottuvuus |
| .. As indicated in the beginning of the study, the structural dimension of pressure toward conformity and social differentiation could be operationalized with the aid of the results of a study carried out by O. Riihinen based on data collected on all Finnish communes. (Stolte-Heiskanen 1967.) | rakenne ulottuvuus operationalisointi |
| Tällaista esitutkimuksen kaltaista eri mittareiden validiuden ja niitä vastaavien käsitteiden välisten riippuvuuksien analyysia voitaneenkin pitää faktorianalyysin omimpana sovellutusalueena.... Faktorien avulla tiivistettiin korrelaatioiden antamaa informaatiota. (Kalimo 1969.) | mittaaminen käsite riippuvuus tiivistäminen korrelaatiot riippuvuus |
| Niinpä korrelaatiomatriisi käsiteltiin faktorianalyysilla.... Kun nämä viisi ...muuttujaa laskettiin korrelaatioiksi ja tehtiin faktorianalyysi, saatiin esille ...harrastusten yleinen rakenne.... ...tiivistää musiikkimaun vaihtelu muutamaa perusulottuvuuteen...(Toiviainen 1970.) | korrelaatiot rakenne ulottuvuus |
| Faktorianalyysin avulla esiin saatujen ongelmaluottuvuuksien pohjalta... (Heikkilä 1990.) | ulottuvuus jatkoanalyysi |
| Faktorianalyysissa muodostetaan useiden havaittujen muuttujien keskinäisten riippuvuuksien kautta yleisellä käsitteellisellä tasolla olevia ulottuvuuksia eli faktoreita, joista kussakin kiteytyy joidenkin havaittujen muuttujien yhteinen teoreettinen sisältö.... Näiden ulottuvuuksien operationalisoiminen faktoreiksi eli latenteiksi (ei-havaituiksi) muuttujiksi mahdollistaa abstraktien käsitteiden mittaamisen. (Nyyssölä 1994.) | riippuvuus käsite ulottuvuus teoria operationalisointi latentti käsite mittaaminen |

| | |
|--|---|
| Faktoriulottuvuuksien ja käsitteellisen jäsentämisen pohjalta on mahdollista muodostaa summamuuttujia kuvaamaan erilaisia... (Kainulainen 1998.) | käsite summamuuttuja |
| Jotta likert-asteikosta muodostamieni summamuuttujien sisältöjen vastaavuutta todellisuuteen voitaisiin empiirisesti osoittaa, tulisi tietää, mitä nämä sisällöt tyhjentävästi määriteltynä sisältävät... Käsitteen aluetta voidaan jäsentää faktorianalyyksillä (Valli 1998.) | summamuuttuja käsite |
| Muuttujajoukon tiivistämiseen käytettiin faktorianalyyksia...Summamuuttujia nimitetään jatkossa ilmapiirifaktoreiksi tai ilmapiiriulottuvuuksiksi... Ilmapiiriulottuvuudet perustuvat yksittäisiin kysymyksiin, joiden on ajateltu mittaavan takana olevaa teoreettista vastinetta. (Nakari 2003.) | tiivistäminen summamuuttuja ulottuvuus mittaaminen teoria |
| ...muodostin aineiston harrastusmuuttujista erityyppistä harrastusaktiivisuutta kuvaavat harrastusulottuvuudet faktorianalyyksin avulla... (Sanaksenaho 2006.) | ulottuvuus rakenne |
| Tämä muuttuja rakennettiin niin, että ensin tehtiin faktorianalyyksin avulla summamuuttuja kolmesta kysymyksestä... (Niemelä 1996.) | summamuuttuja |
| Empiirisen analyysin aikana esimerkiksi luottamuksen eri ulottuvuuksia on – tarvittaessa faktorianalyyksia hyödyntäen – yhdistetty keskiarvosummamuuttujiksi. (Kankainen 2007.) | (käsite) ulottuvuus summamuuttuja |
| Ennen käsitteen varsinaista faktorianalyyttistä operationalisointia... Yhteisöllisyys, kuten se tullaan pian faktorianalyyksin perusteella määrittelemään... Mutta toimivatko mittarit? Sen selvittämisessä hyödynnetään faktorianalyyksiä. (Koivumäki 2008.) | käsite operationalisointi mittaaminen |
| Tältä pohjalta kokosin yksilöiden verkostokontakteja kartoittavia muuttujia faktorianalyyksin avulla summamuuttujiksi, jotka läpäisivät reliabiliteettitestin mutta olivat myös tutkimusteorian ja muiden sosiaalisen pääoman tutkimuksissa saatujen tulosten kannalta mielekkäitä (sosiaalisen pääoman operationalisoinnista enemmän luvussa... (Salminen 2012.) | summamuuttuja teoria käsite |
| Pääkomponenttianalyyksin tulosten perusteella muodostettiin kaksi summamuuttujaa. (Turunen 2012.) | summamuuttuja |
| Faktorianalyysi helpottaa useiden Likert-asteikollisten muuttujien tuottaman tiedon tiivistämistä. Analyysimenetelmää käytettiin, koska haluttiin löytää muuttujajoukosta yhteisiä piirteitä tai ulottuvuuksia ja löytää keskenään eniten korreloivia muuttujia. Metodi auttoi löytämään niitä muuttujia, joilla on samankaltaista vaihtelua ja jotka ovat toisista muuttujista riippumattomia. (Ohisalo 2017.) | tiivistäminen ulottuvuus korrelaatio riippuvuus |
| Artikkelissa 1 faktorianalyyksillä ryhmitellään samankaltaisia muuttujia eli tutkitaan, millä kaikilla muuttujilla on keskenään samankaltaista vaihtelua (miten muuttujien korrelaatiot kimputtuvat). Aluksi selvitettiin, että aineiston muuttujajoukko soveltuu faktorianalyyksin käyttöön. Faktorianalyyksin perusideana on yhdistää muuttujia toisiinsa niin, että ne muodostavat sisällöllisesti tai teoreettisesti kiinnostavan oman ulottuvuutensa. (Laihiala, 2018.) | korrelaatio teoria ulottuvuus |

Tarkasteltaessa sosiologian väitöskirjoissa eri aikoina esitettyjä faktorianalyyksipe-
rustelua kiinnittyy ensimmäisenä huomio niiden moninaisuuteen. Sekä perustelujen
määrä että perustelujen pituus vaihtelevat väitöskirjoissa lyhyestä maininnasta lähes
koko väitöskirjan kattavaan esittelyyn. Perustelujen moninaisuus näkyy myös niiden
sisällöissä. Faktorianalyyksin valintaa perustellessaan sosiologit ovat nojautuneet tilas-
tomatematiikkaan mainiten muun muassa tarpeen tarkastella muuttujien välisiä riip-
puvuuksia, muuttujien vaihtelua, varianssia tai harhaa. Perusteluissa on tuotu esiin
myös analyysin yksinkertaistamistarve, mikä on liittynyt tarkasteltavien muuttujien
määrän vähentämiseen tai aineiston muunlaiseen tiivistämiseen myös mahdollista

jatkoanalyysia silmällä pitäen. Faktorianalyysilla on haluttu myös tyypitellä tai ryhmitellä vastaajia tai muuttujia tai selittää yksilöiden erilaisuutta. Faktorianalyysiperusteluissa on viitattu myös sosiologiseen selittämiseen, teoriaan, operationalisointiin, latenttiin muuttujaan tai tulkinnallisen rakenteen löytämiseen. Faktorianalyysin avulla on haluttu myös tarkastella teoreettisen käsitteen osatekijöitä tai löytää ulottuvuuksia. Joistakin perusteluista käy ilmi faktorianalyysin valinta sen takia, että se on ollut monimuuttujamenetelmä tai sillä hetkellä suosiossa oleva menetelmä.

Tiivistin perustelujen sisällöt 15 ilmaukseen. Kuva 27 esittää näiden tiivistettyjen ilmausten esiintymislukumäärät väitöskirjoissa. Yleisin väitöskirjoissa mainittu ilmaus on ulottuvuus. 56:sta faktorianalyysia sisältävästä väitöskirjasta 29:ssä puhuttiin faktorianalyysin yhteydessä ulottuvuuksista. Mutta kuten kuvasta näkyy, ovat myös muut ilmaukset olleet tasaisen suosittuja lukuun ottamatta faktorimallia ja analyysin yksinkertaistamista.

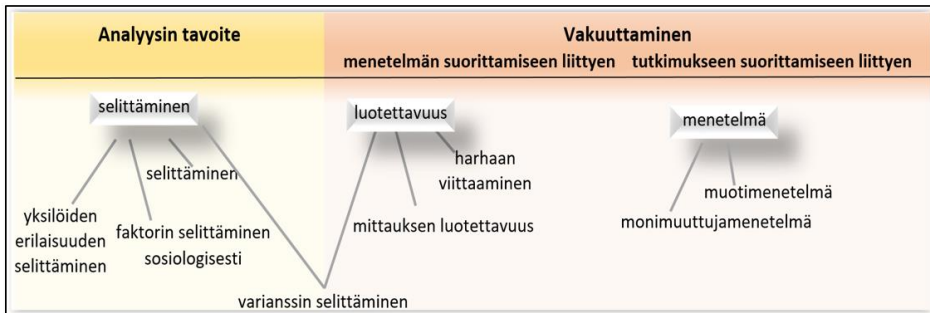


Kuva 27 Faktorianalyysiperustelujen tiivistettyjen ilmausten esiintymismäärät

Sanojen ja ilmaisujen tiivistämisessä sekä niiden esiintymismäärien laskemisessa olen kohdannut saman haasteen, jonka numeerisia aineistoja keräävät sosiologit yleensäkin kohtaavat: tarkka mittaaminen on mahdotonta (ks. esim. Stouffer 1953; Patty ja Penn 2015; Smith ja Atkinson 2016; Uprichard 2016; Franzosi 2016). Sosiologien esittämien faktorianalyysin valintaa koskevien perustelujen tarkka haltuun ottaminen laskemisen ja mittaamisen kautta on siten myös mahdotonta. Perusteluissa esiintyvien sanojen laskeminen ja mittaaminen ei ole voinut olla tarkkaa muun muassa sen takia, että sosiologit ovat käyttäneet useita ilmauksia yhdessä perustelussaan, kuten taulukosta 8 kävi ilmi. Joidenkin perustelujen kohdalla olen tehnyt myös omia tulkintojani. Olen esimerkiksi tulkinnut sosiologin viittaavan sosiologisen tutkimuskysymyksensä teoriaan, vaikka sosiologi ei olisikaan sitä täsmälleen ilmaissut. Perustelujen tarkkaa haltuunottoa vaikeuttavat myös erilaiset kirjoitustyylliset

seikat. Tarkastelemani väitöskirjat on kirjoitettu vuosina 1960–2018, jona aikana on tapahtunut monia kirjoitustyyliin liittyviä muutoksia. Myös kirjoittajat ovat erityyppisiä joidenkin ollessa ilmaisuisaan hyvin lyhytsanaisia ja toisten kirjoittaessa taas runsassanaisesti. Myös faktorianalyysin rooli väitöskirjoissa on vaikuttanut esitettyjen perustelujen laajuuteen. Käytännössä kyse on samasta ongelmasta, kuin mitä kyselytutkimusten avoimien kysymysten vastausten teemoittamisessa ja luokittelussa kohdataan.

Mittaamiseen liittyvistä haasteista huolimatta voi perustelujen esiintymislukumääriä tarkastelemalla saada alustavaa tietoa siitä, minkälaisia asioita sosiologit ovat kokeneet tarpeellisiksi – tai vakuuttaviksi – tuoda esiin faktorianalyysiperusteluissaan. Jatkoin perustelujen analysointia laadullisin menetelmin ATLAS.ti -ohjelmaa hyödyntäen. Analyysi tuotti seitsemän perustelujen ryhmää ja ne ovat tiivistettävissä kahdeksi kokonaisuudeksi. Niistä ensimmäinen kuvaa analyysin tavoitteeseen ja analyysillä vakuuttamiseen kytkeytyviä perusteluja. Siihen luen kuuluvaksi selittämiseen, luotettavuuteen ja menetelmään itsessään viittaavat perustelut (kuva 28). Toinen - tutkimukseni kannalta mielenkiintoisempi – kokonaisuus koostuu riippuvuuteen, rakenteeseen, niputtamiseen ja teoriaan viittaavista perusteluista (kuva 29), jota käsittelen luvussa 7.2.



Kuva 28 ATLAS.ti-analyysin tulos faktorianalyysiperusteluille

Kun faktorianalyysia on perusteltu *selittämisen* tarpeella, on tavoitteeksi esitetty joko selittäminen sellaisenaan tai yksilöiden erilaisuuden selittäminen, mutta myös faktorin itsensä selittäminen sosiologisesti. Koska faktorianalyysi ei tilastollisena menetelmänä sisällä selittäjä-selitettyä-asetelmaa (Tacq 1997, 35), viittaavat sosiologien esille tuomat selittämiset todennäköisesti tulkinnalliseen selittämiseen.

Faktorianalyysilla on haettu myös vakuuttavuutta liittyen joko aineiston ja muuttujien luotettavuustarkasteluihin tai yleisesti menetelmävalintaan liittyen. *Luotettavuuteen* kytkeytyvissä viittauksissa sosiologit ovat ottaneet esille harhan, mittauksen luotettavuuden tai varianssin selittämisen, jotka kaikki ovat hyvin tilastotieteellisiä

perusteluja. Varianssin selittäminen on yhteydessä myös selittämistä kuvaavaan kokonaisuuteen.

Yleinen menetelmävalintaan liittyvä vakuuttaminen, joka kuvassa 28 näkyy *menetelmään* liittyvinä viittauksia, liittyy tutkimuksesta kirjoitettaessa todennäköisesti sen vakuuttamiseen, että tutkimus on suoritettu asianmukaisesti valitsemalla tutkimukseen sopivia tutkimusmenetelmiä. Monimuuttujamenetelmien käyttäminen koetaan vakuuttavammaksi kuin pelkkien kuvailevien menetelmien käyttäminen. Muotimenetelmäksi nimetty perustelu taas viittaa faktorianalyysin suosioon: on ollut turvallista valita faktorianalyysi, koska monet muutkin ovat niin tehneet. Muotimenetelmän voi ymmärtää sekä positiivisessa että negatiivisessa tarkoituksessa. Positiivisessa mielessä se viittaa menetelmän käyttökelpoisuuteen ja hyödyllisyyteen sosiologisessa tutkimuksessa. Negatiivisessa mielessä sana herättää mielikuvan toisten tutkimisesta ja ehkä jopa osaamattomuudesta.

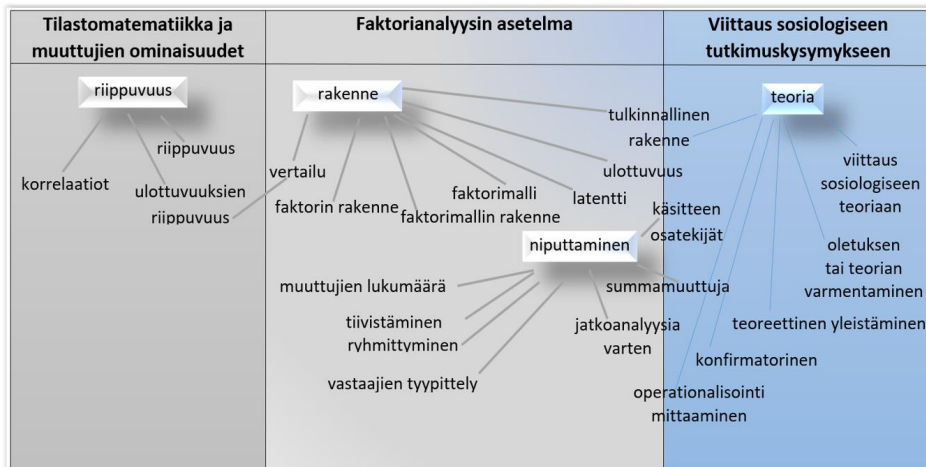
Menetelmän uutuus ja sen kokeilunhalu vaikuttavat joskus tutkijoiden menetelmävalintaan. Menetelmän kehittäminen on sinänsä usein merkittävä saavutus, ja menetelmän tultua keksityksi löytyy yleensä runsaastikin tutkijoita, jotka ovat sitä halukkaat kokeilemaan (Nummenmaa 1980, 4). Ei ole harvinaista sekään, että muo-dissa tai ”vallassa” olevat eli kaikkien muidenkin tutkijoiden käyttämät menetelmät ohjaavat tutkijaa valitsemaan tietyt menetelmät omaan tutkimukseensa. Ennen pitkää muotimenetelmien liiallinen – ja harkitsematon – käyttö herättää kritiikkiä. Esimerkiksi Esa Läärä (1991) pohtii hieman kriittiseen sävyyn syitä logistisen regressioanalyysin suureen suosioon lääketieteellisissä tutkimuksissa 1980-luvun puolivälin jälkeen. Yhtenä syynä hän mainitsee juuri sen, että menetelmä on helppoa ja turvallista valita omaan tutkimukseen, koska kaikki muutkin ovat sitä käyttäneet. Toisena syynä menetelmien yksipuoliseen valikoimaan tutkimuksissa Läärä mainitsee tietokoneohjelmien vaikutuksen:

Toisaalta nykykäytäntöä ylläpitävät vähintään yhtä paljon ne tietokoneohjelmat, joita regressiomallien sovittamiseen on käytettävissä. On monia soveltajia, jotka eivät ole käyneet mitään kurssia eivätkä lukeneet asianmukaisia oppikirjoja, mutta jotka sujuvasti pystyvät käyttämään BMDP:n tai SAS:in proseduuria logistisen regression sovittamiseksi aineistoonsa ilman syvällisempiä tietoja tai valmiuksia mallien käyttöön. He valitsevat tietenkin tämän mallin, koska muita vaihtoehtoja ei tarjota, ja tyytyvät ohjelman oletusarvoiseen toimintaan ja tulostukseen. Näillä ohjelmilla on aivan ilmeisesti liian alhainen käyttökynnys. (Läärä 1991.)

Sosiologian väitöskirjoissa faktorianalyysin käyttöä perustellaan useilla tavoilla. Sitä perustellaan sekä tilastomatematisilla että sosiologisilla painotuksilla. Kyse ei siis ole pelkästään faktorianalyysin mahdollistavan tilastollisen ohjelman alhaisesta käyttökynnyksestä vaan analyysin tavoitteesta.

7.2. Tilastotieteeseen, asetelmaan ja tutkimuskysymykseen nojaavat perustelut

Kun sosiologian väitöskirjoissa faktorianalyysia perustellaan riippuvuuteen, rakenteeseen, niputtamiseen tai teoriaan kytkeytyvillä tarkastelutarpeilla, kuvastavat ne tulkintani mukaan rajakohdemallissa esittämäni kolme kauttakulkupistettä (kuva 29). Olen sijoittanut riippuvuuteen liittyvät perustelut tilastomatematiikkaa ja muuttujien ominaisuuksia korostavaan kauttakulkupisteeseen. Rakenteen ja niputtamisen näen asetelmaa painottavana kauttakulkupisteenä. Teoriaan kytkeytyvät perustelut puolestaan sijoittuvat sosiologista tutkimuskysymystä painottavaan kauttakulkupisteeseen.



Kuva 29 Faktorianalyysiperustelut rajakohdetarkasteluna

Erilaiset perustelut ja perusteluryhmät ovat myös kytköksissä toisiinsa. Kun faktorianalyysilla on esimerkiksi haluttu tarkastella ulottuvuuksien riippuvuuksia, kytkeytyy tarkastelu samalla vertailun kautta rakenteeseen. Tulkinnallinen rakenne kytkeytyy sekä rakenteeseen että teoriaan, mikä kiteyttää tilastollisia menetelmiä käyttävien sosiologisten keskeisen tavoitteen analyysin suhteen: analyysin olisi oltava rakenteellisesti kunnossa ja teoreettisesti tulkittavissa.

Seuraavaksi avaan kuhunkin kauttakulkupisteeseen sijoittuvia faktorianalyysiperusteluja tarkemmin. Aloitan tilastomatematiikkaa ja muuttujien ominaisuuksia korostavasta kauttakulkupisteestä. Sen jälkeen esitän asetelmaa painottavan kauttakulkupisteen perusteluja ja viimeiseksi esittelen teoriaan liittyvän kauttakulkupisteen.

Tilastomatematiikkaa ja muuttujien ominaisuuksia korostavissa perusteissa ovat sosiologit halunneet tutkia faktorianalyysin avulla riippuvuuksia.

Riippuvuuteen on viitattu eri tavoin. Tyypillisimmin on tuotu esiin muuttujien korreloituminen tai tarve tehdä riippuvuustarkastelua muuttujien välillä. Väitöskirjoissa löytyy myös mainintoja ulottuvuuksien riippuvuuksista, riippuvuussuhteista, tai esimerkiksi ainesten riippuvuussuhteesta, kuten käy ilmi seuraavasta lainauksesta.

... tutkitaan monimuuttuja-analyysin avulla sitä, miten uskonnollisuuden eri ainekset ovat toisiinsa riippuvuussuhteessa... (Koskelainen, 1968.)

Riippuvuuteen kytkeytyvät ilmaukset sopivat selkeästi tilastomatematiikkaan nojautuviksi perusteluiksi. Väitöskirjoja, joissa riippuvuustarkastelu on tuotu esiin faktorianalyysin perusteluissa tavalla tai toisella, on 30 kappaletta.

Silloin, kun faktorianalyysin käytön perusteluissa korostetaan faktorianalyysin asetelmaa, on tavoitteena löytää muuttujajoukolle rakenne. Faktorianalyysin käyttöä perustellaan tavalla tai toisella rakenteen tunnistamisella ja tutkimisella, josta esimerkkinä on ote Uiton vuonna 1964 tarkastetusta väitöskirjasta.

Faktorianalyysin avulla on tarkasteltu tyytyväisyyden rakennetta tutkittavissa ammateissa. (Uitto 1964.)

Rakenteeseen viittaavia perusteluita on mainittu 41 väitöskirjassa. Niissä on tuotu esiin tarve tietää, kuinka monta faktoria analyysiin valitusta muuttujajoukosta syntyy ja minkälaisia faktorit ovat sisällöltään. Rakenteeseen viittaavat perustelut ilmentävät faktorianalyysin asetelmaa, mutta kuten kuvasta 29 näkyy, painottuvat tietyt ilmaukset lähemmäksi tilastomatematiikkaa näkökulmaa ja tietyt ilmaukset, kuten tulkinnallinen rakenne, latentit muuttujat ja perustelujen suosituin ilmaus ulottuvuus¹⁶ taas lähemmäksi sosiologista näkökulmaa. Ulottuvuus onkin tarkoittanut perusteluissa monenlaisia asioita. 1950- ja 1960-luvulla faktorianalyysin suorittamisessa hyödynnettiin koordinaattistoa (ks. esim. Vahervuo ja Ahmavaara 1958), jolloin ulottuvuus saattoi viitata koordinaattiakseleiden sijaintiin. Tyypillisimmin ulottuvuudella on viitattu faktoriin pitäen ulottuvuutta ja faktoria toistensa synonyymeinä. Näin on myös monessa oppikirjassa. Esimerkiksi Hair ym. (1998) käyttävät ulottuvuus-sanaa faktorin synonyymina kirjoittaessaan, että ”Faktorianalyysin avulla tutkija voi ensin identifioida eri ulottuvuudet ja sen jälkeen päätellä, kuinka vahvasti kukin muuttuja selittyy kullakin ulottuvuudella”. Ulottuvuuden voi ymmärtää paitsi tilastotieteellisenä ilmaisuna faktoreihin liittyen, myös sosiologisena ilmaisuna käsitteisiin liittyen. Esimerkiksi Alkula ym. (1994) toteavat, miten faktorianalyysin avulla voidaan etsiä käsitteen erilaisia ulottuvuuksia. Sosiologian väitöskirjoissa ulottuvuus-sanaa käytetään usein myös viittaamaan käsitteen erilaisiin osa-alueisiin.

¹⁶ Erityisen suosittu ilmaisu ulottuvuus on ollut 2000-luvulla, jolloin 20:stä faktorianalyysia sisältävästä väitöskirjasta 15:ssä on mainittu ulottuvuus.

Faktorianalyysin korostamiseen asetelmana nivoutuu myös tavoite niputtaa tai tiivistää. Faktorianalyysi tarjoaa tutkijalle tietoa siitä, millä tavalla tutkimusaineiston sisältämää tietoa voisi tiivistää tai millä tavalla muuttujia voisi niputtaa. 42:ssa sosiologian väitöskirjassa on mainittu jollakin tavalla niputtamisen, tiivistämisen tai ryhmittelyn tarve.

Monilukuisten muuttujien informaatio pyrittiin tiivistämään harvalukuisiin faktoreihin. (Tauriainen, 1970.)

Kuten rakenteeseen viittaavissa perusteluissa, myös niputtamiseen ja tiivistämiseen viittaavissa perusteluissa on havaittavissa sekä tilastotieteellisiä ilmauksia että sosiologiseen tutkimustarpeeseen liittyviä ilmauksia. Tilastotieteelliseksi ilmaukseksi luen muuttujien lukumäärään, tiivistämiseen ja ryhmittelyyn sekä vastaajien tyyppittelyyn viittaavat ilmaukset. Tällöin tutkijalle on voinut riittää faktorianalyysin tuottama tieto muuttujien taustalla vaikuttavien faktoreiden määrästä ja niiden sisällöstä. Usein sosiologit ovat kuitenkin halunneet hyödyntää faktorianalyysia alustavana analyysina jatkoanalyysia silmällä pitäen. Tällaisiksi sosiologista tutkimustarvetta kuvastaviksi ilmaukseksi näen summamuuttujiin, jatkoanalyysiin ja käsitteisiin viittaavat ilmaukset. Faktoreiden ollessa vain osoituksia muuttujien ryhmittymisestä on jatkoanalyysia varten faktorit jollakin tavalla rakennettava empiirisiksi muuttujiksi. Tämä tapahtuu joko faktoripistemäärämuuttujien tai summamuuttujien avulla. Esittelen luvussa 8.2, kuinka usein sosiologit ovat päätyneet näihin ratkaisuihin.

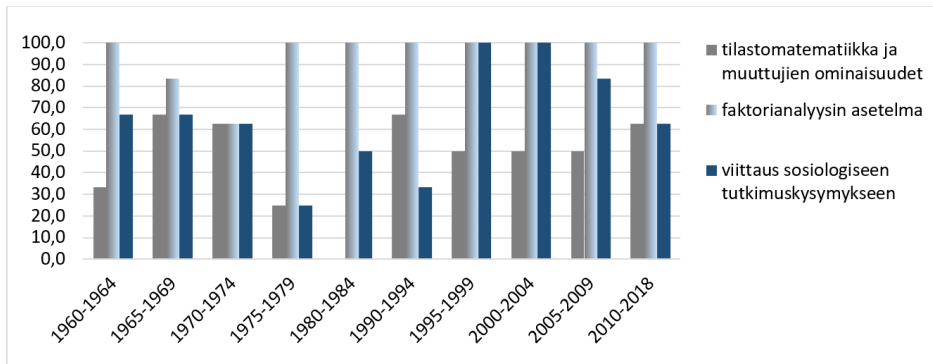
Sosiologit ovat valinneet faktorianalyysin myös teoreettisin perustein. Tällöin kyse on ryhmästä, jossa viitataan myös sosiologiseen tutkimuskysymykseen. Neljännen perustelujen kokonaisuuden katson tavalla tai toisella liittyvän sosiologisten tutkimusten teorioihin. Tällöin perustelut viittaavat operationalisointiin, oletuksen tai teorian varmentamiseen, teoreettiseen yleistämiseen tai tulkinnalliseen rakenteeseen.

Faktorianalyysin käyttö teoriaorientoituneessa tutkimuksessa liittyy paremmin empiiristen vastineiden verifikaatiostrategiaan, eli kyselylomaketta tehdessä on jo ollut käsitys niistä teoreettisista käsitteistä, joita yksittäisillä kysymyksillä ja väittämillä haetaan. (Jokivuori 2002.)

Väitöskirjoja, joissa teoria tulee faktorianalyysin perusteluissa tavalla tai toisella esille, on 41 kappaletta. Operationalisoinnin kautta faktorianalyysiin on kytköksissä myös mittaaminen. Faktorianalyysiperusteluissa mittaaminen on tuotu esiin 18 väitöskirjassa vuosien 1960–2018 aikana. Esimerkkinä mainittakoon Marjatta Marinin vuonna 1970 tarkastettu väitöskirja, jossa hän kertoo käyttäneensä faktorianalyysia mittareiden valitsemiseen.

Tämä saattaa tietenkin johtua 1) käytetyistä mittareista, joista esimerkiksi käyttäytymismittarit on jo valikoitu yhden faktorianalyysin perusteella (Marin 1970.)

Kuvassa 30 näkyy kolmeen rajakohdemallin kauttakulkupisteeseen kytkeytyvien perustelujen osuudet eri aikoina. Faktorianalyysin asetelmaan kytkeytyvät perustelut ovat olleet suosituimpia koko tarkastelujakson ajan. On kuitenkin myös huomattava, että asetelmaa kuvaavaan kauttakulkupisteeseen kuuluu enemmän perusteluja kuin kahteen muuhun kauttakulkupisteeseen.



Kuva 30 Faktorianalyysin perustelujen kauttakulkupisteet

Kuvan 30 tarkastelussa mielenkiinto kohdistuu tilastomatematiikkaa sekä muuttujien ominaisuuksia kuvaavan kauttakulkupisteen ja sosiologiseen tutkimuskysymykseen viittaavan kauttakulkupisteen vertailuun. Kuvan perusteella voisi varovaisesti – muistaen mittaamiseen liittyvät haasteet - tulkita, että sosiologit ovat valinneet sosiologiseen tutkimuskysymykseen viittaavan kauttakulkupisteen useammin kuin tilastomatematiikkaan ja muuttujien ominaisuuksiin viittaavan kauttakulkupisteen.

7.3. Faktorianalyysi määrällisenä ja laadullisena menetelmänä

Kirjallisuudessa on esitetty erilaisia faktorianalyysin käyttötapoja (ks. esim. Yong ja Pearce 2013, 79; Costello ja Osborne 2005, 1). Brett Williams, Andrys Onsmann ja Ted Brown (2010, 2) esittävät kahdeksan erilaista käyttötarkoitusta. Faktorianaalyysin tavoitteena voi heidän mukaansa olla muuttujien määrän vähentäminen, muuttujien välisten suhteiden ja muuttujien taustalla vaikuttavien rakenteiden tarkastelu, teoreettisen käsitteen ulottuvuuksien tarkastelu, mittaamisen liittyvän konstruktiivallisuuden arviointi (ks. Nummenmaa ym. 1997, 201–209; Jokivuori ja Hietala 2007, 100–105), analyysin ja tulokinnan yksinkertaistaminen, teoreettisten käsitteiden ja ehdolla olevien teorioiden käyttökelpoisuuden arviointi. Myös An Gie Yong and Sean Pearce (2013, 79) sekä Anna B. Costello ja Jason Osborne (2005, 1) toteavat faktorianaalyysillä olevan lukuisia erilaisia käyttötarkoituksia, joista osa on matemaattisia ja teknisiä ja osa teoreettisiin käsitteisiin painottuvia. Alkula ym. (1994, 278) kirjoittavat, että yhteiskuntatieteellisessä tutkimuksessa faktorianaalyysia voidaan mukaan käyttää ainakin kolmella tavalla: 1) suuren muuttujajoukon keskinäisten riippuvuuksien yleiskartoitukseen, 2) mittareiden tai summamuuttujien luomiseen ja kehittämiseen ja 3) varsinaisen faktorimallin rakentamiseen. He tarkoittavat toisen ja kolmannen käyttötavan olevan lähellä toisiaan. Faktorianaalyysin voi ajatella taipuvan tässä mielessä sekä määrälliseksi että laadulliseksi menetelmäksi.

Havainnollistan faktorianaalyysin määrällisyyttä ja laadullisuutta kuvan 31 avulla. Siinä faktorianaalyysin tavoitteet ja käyttötarkoitukset on jaettu viiteen kokonaisuuteen: a) muuttujien välisten riippuvuuksien tutkimiseen, b) rakenteen löytämiseen, c) ulottuvuuksien etsimiseen, d) käsitteiden empiiristen vastineiden löytämiseen ja e) käsitteiden empiiristen vastineiden analysointiin.

| | |
|---|---|
| Tavoitteena riippuvuuksien tutkiminen | |
| | <p>a) Tarkastelu kohdistuu muuttujiin ja niiden korreloitumiseen.</p> $y_1 = \lambda_1\eta + \varepsilon_1$ $y_2 = \lambda_2\eta + \varepsilon_2$ $y_3 = \lambda_3\eta + \varepsilon_3$ $y_4 = \lambda_4\eta + \varepsilon_4$ |
| Tavoitteena rakenteen löytäminen | |
| | <p>b) Muuttujien korreloitumiselle etsitään taustalla vaikuttava rakenne.</p> |
| Tavoitteena ulottuvuuksien etsiminen | |
| | <p>c) Tavoitteena arvioida faktoreiden eli ulottuvuuksien sisältöjä ja pyrkiä keksimään ulottuvuuksille sisältöjä kuvaavat nimet.</p> |
| Tavoitteena käsitteiden empiiristen vastineiden löytäminen (operationalisoinnin käännteistoimenpide) | |
| | <p>d) Arvioidaan faktoreiden, ulottuvuuksien, vastaavuutta teoreettisiin käsitteisiin</p> |
| Tavoitteena käsitteiden empiiristen vastineiden analysointi | |
| | <p>e) Käsitteiden empiiristen vastineiden muodostaminen summamuuttujien tai faktoripistemäärämuuttujien avulla ja niiden jatkoanalysointi</p> |

Kuva 31 Faktoriansalyysin määrälliset ja laadulliset tavoitteet

Kun faktorianalyysin tavoitteena on riippuvuuksien tutkiminen, on tarkastelun painopiste muuttujissa ja niiden korreloitumisessa (kuva 31 a). Sosiologian väitöskirjojen faktorianalyysiperusteluissa on tuotu esiin muuttujien keskinäisen riippuvuuden tai korreloitumisen tarkastelutarve. Harvaksen esitettyjä perusteluja ovat muuttujien varianssin selittäminen tai mittausten tai muuttujien sisältämän harhan tarkastelu. Faktorianalyysi nojautuu tällöin täysin tilastomatemattisiin kriteereihin ja on siten määrällinen menetelmä.

Jos faktorianalyysin tavoitteena on rakenteen löytäminen, siirtyy tarkastelun painopiste riippuvuuden taustalla vaikuttavien tekijöihin. Tällöin ollaan kiinnostuneita siitä, kuinka monta riippuvuuden taustalla vaikuttavaa tekijää löytyy ja miten ne vaikuttavat kuhunkin muuttujaan (kuva 31 b). Sosiologian väitöskirjojen perusteluissa on tuotu esiin tarve tiivistää, ryhmitellä tai tyyppitellä. Faktoreihin kytkeytyvät perustelut sisältävät mainintoja ulottuvuuksista, latenteista muuttujista ja summamuuttujista. Perusteluissa on mainittu myös faktoreiden tai ulottuvuuksien välisen riippuvuuden tarkastelutarve tai vertailu. Perusteluissa on mainittu myös tarve löytää rakenne, faktorin rakenne, faktorimallin rakenne tai tulkinnallinen rakenne. Tässäkin tapauksessa faktorianalyysi nojautuu paljolti tilastomatemattisiin kriteereihin ollen määrällinen menetelmä.

Kun faktorianalyysin tavoitteena on ulottuvuuksien etsiminen, muistuttaa tavoite hyvin paljon edellä mainittua rakenteen etsimistä, koska ulottuvuuksien etsimisessäkin halutaan saada selville muuttujien taustalla vaikuttavien faktoreiden lukumäärä sekä selvittää, millä tavalla faktorit vaikuttavat muuttujiin. Asia, jossa ulottuvuuksien etsiminen eroaa rakenteen etsimisestä, on tutkijan tekemä tulkinta. Ulottuvuuksien etsimisessä painopiste on siten faktoreiden eli ulottuvuuksien sisältöjen arvioinnissa ja ulottuvuuksien nimeämisessä (kuva 31 c). Ulottuvuuksien sisällön arviointi ja ulottuvuuksien nimeäminen vaativat tutkijalta sisällöllistä asiantuntemusta ja siten faktorianalyysi ei voi nojautua pelkästään tilastomatemattisiin kriteereihin. Tuloksen viimeistelyssä on käytettävä laadullista otetta.

Jos faktorianalyysin tavoitteena on käsitteiden empiiristen vastineiden löytäminen, jonka voi ajatella myös operationalisoinnin käännteistoimenpiteeksi, voi analyysi olla teknisesti samanlaista kuin edellä rakenteen ja ulottuvuuksien etsimisessä. Faktorianalyysi eroaa oleellisesti edellä mainituista tavoitteista siinä, että nyt keskiössä on sosiologisen asiantuntemuksen hyödyntäminen faktoreiden tulkinnassa. Sosiologin on arvioitava, kuinka hyvin analyysin tuottamat faktorit vastaavat hänen tutkimansa käsitteen osa-alueita (kuva 31 d). Faktoreiden arviointi ja tulosten tulkinta muistuttaa tämän vuoksi jo hyvinkin laadullista analyysia.






Sosiologian väitöskirjoissa faktorianalyysi on valittu analyysimenetelmäksi myös siksi, että sen avulla on haluttu tehdä teoreettista yleistämistä tai varmentaa sosiologista teoriaa. Kuvan 31 viimeinen vaihe (e) esittää faktorianalyysia, jonka tavoitteena on summamuuttujien tai faktoripistemäärämuuttujien muodostaminen jatkoanalysointia varten. Summamuuttujien tai faktoripistemäärämuuttujien ajatellaan olevan

käsitteiden empiirisiä vastineita. Tällöin faktorianalyysi toimii alustavana analyysina antaen tutkijalle viitteitä siitä, minkälaisilla summamuuttujilla tutkija mahdollisesti pääsee analyysissaan käsitteellisemmälle, teoreettisemmalle tasolle. Faktorianalyysi voi jälleen olla teknisesti samanlaista kuin kaikissa edellä kuvatuissa tilanteissa, mutta nyt sosiologin asiantuntemus on ratkaisevan tärkeässä asemassa. Sosiologin on käytettävä asiantuntemustaan kyetäkseen arvioimaan faktoreiden vastaavuutta käsitteiden tai niiden osa-alueita kanssa. Vain tutkittavaa kohdetta ymmärtävä sekä siihen liittyvää teoretisointia hahmottava sosiologi pystyy arvioimaan, vastaavatko faktorit ja hänen tutkimuksensa käsitteelliset hahmotukset toisiaan niin hyvin, että hänen on perusteltua muodostaa kunkin faktorin muuttujista summamuuttujia, joita voi käyttää jatkoanalyysissa käsitteiden empiirisinä vastineina. Faktorianalyysi vaatii teknisen ja tilastomatematiikan otteen lisäksi hyvin vahvan laadullisen otteen tulosten arvioinnissa ja tulkinnassa. Tämä on myös se kohta, jossa tilastotieteen ja matematiikan edustama faktorianalyysi sekä sosiologia eroavat toisistaan. Ehkä tämä on osasy sille, miksi Eskola harmitteli vuonna 1971 Sosiologia-lehdessä faktorianalyysiin liittyvää kapeakatseisuutta:

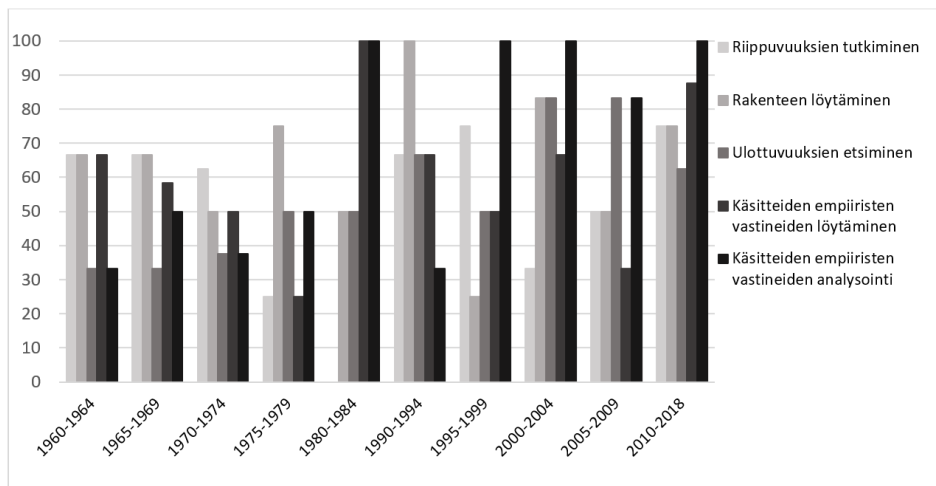
Valitettavasti faktorianalyysin myötä suomalainen sosiaalitiede yhä tiukemmin joutui amerikkalaiselle sosiaalitutkimukselle tunnusomaisiin pikku ympyröihin ja sen siteet laajakatseiseen eurooppalaiseen näkemykseen heikkenivät, vaikka faktorianalyysi ehkä pystyisi luomaan pohjan myös laajalle sosiaalitieteelliselle teorianmuodostukselle. (Eskola 1971.)

Sosiologian väitöskirjoissa eri aikoina esiin tuodut tavoitteet faktorianalyysin käytölle on koottu taulukkoon 9. Kuten edellä on tullut ilmi, esitetyt tavoitteet eivät ole toisensa pois sulkevia. Yhdessä väitöskirjassa on voitu esittää useita tavoitteita yhdelle faktorianalyysille tai väitöskirjassa on suoritettu useita faktorianalyysseja.

Taulukko 9 Faktorianalyysin tavoitteet eri aikoina

| | Väitöskirjojen määrä, joissa faktorianalyysin tavoitteena ollut | | | | | |
|-----------------|---|---|---|---|---|---|
| |  |  |  |  |  | Faktorianalyysia sisältävien väitöskirjojen määrä |
| | riippuvuuksien tutkiminen | rakenteen löytäminen | ulottuvuuksien löytäminen | käsitteiden empiiristen vastineiden löytäminen | käsitteiden empiiristen vastineiden analysointi | |
| 1960-1964 | 2 | 2 | 1 | 2 | 1 | 3 |
| 1965-1969 | 8 | 8 | 4 | 7 | 6 | 12 |
| 1970-1974 | 5 | 4 | 3 | 4 | 3 | 8 |
| 1975-1979 | 1 | 3 | 2 | 1 | 2 | 4 |
| 1980-1984 | 0 | 1 | 1 | 2 | 2 | 2 |
| 1990-1994 | 2 | 3 | 2 | 2 | 1 | 3 |
| 1995-1999 | 3 | 1 | 2 | 2 | 4 | 4 |
| 2000-2004 | 2 | 5 | 5 | 4 | 6 | 6 |
| 2005-2009 | 3 | 3 | 5 | 2 | 5 | 6 |
| 2010-2018 | 6 | 6 | 5 | 7 | 8 | 8 |
| Yhteensä | 32 | 36 | 30 | 33 | 38 | 56 |

Kuva 32 esittää samat tiedot kuin taulukko 9, mutta siinä väitöskirjoissa kulloinkin esitetyt faktorianalyysitavoitteet on suhteutettu eri aikoina tarkastettujen faktorianalyysia sisältävien väitöskirjojen määrään. Faktorianalyysille esitetyt tavoitteet ovat 1970-luvun loppuun asti painottuneet riippuvuuksien tutkimiseen tai rakenteen löytämiseen. 1980-luvulta alkaen aina tarkastelujakson loppuun asti eli vuoteen 2018 on faktorianalyysille esitetty lähes poikkeuksetta tavoitteeksi käsitteiden empiiristen vastineiden analysointi.



Kuva 32 Faktorianalyysille eri aikoina esitetyt tavoitteet suhteutettuna faktorianalyysia sisältävien väitöskirjojen määrään

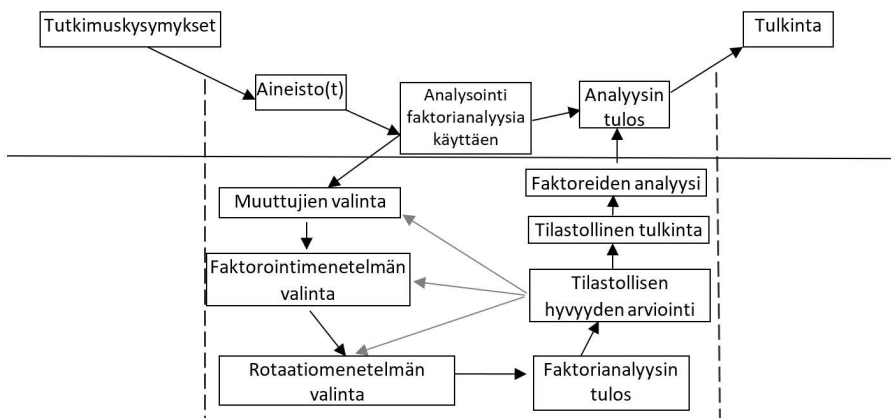
Yhteenvedona todettakoon, että asia, joka tekee faktorianalyysistä määrällisen tai laadullisen menetelmän, on tutkijan tekemä tulkinta. Eskola (1971) toteaa, että faktorianalyysissä tilastollisiin menetelmiin pohjautuvan tarkastelun voi yhdistää intuitioon pohjautuvaan ajatteluun. Tutkija itse on oman tutkimuksensa paras asiantuntija. Paras tulos syntyy, kun tutkija yhdistää asiantuntemuksensa menetelmän tuottamaan tietoon. Toivon, että edellä esittämäni rajakohdemalli eri tavalla painottuvine kauttakulkupisteineen rohkaisee faktorianalyysia käyttäviä sosiologeja huomioimaan tämän.

8. Faktorianalyysin tekniset valinnat väitöskirjoissa

Leskinen (1987) mainitsee faktorianalyysin ensimmäisen suoritusvaiheen olevan faktoreiden lukumäärän arvioinnin, jossa matemaattinen lähestymistapa on katsoa korrelaatiomatriisista johdettuja ominaisarvoja. Leskinen toteaa, että eräs käytännön nyrkkisääntö on valita yhtä monta faktoria kuin on arvoltaan yli ykkösen olevien ominaisarvoja, mikä on esimerkiksi SPSS-ohjelmassa oletuksena. Leskinen jatkaa, että mikäli muuttujat noudattavat moniulotteista normaalijakaumaa ja otoskoko on riittävän suuri, voidaan faktoreiden lukumäärää arvioida myös tilastollisin testein.

Toinen faktorianalyysin vaihe on Leskisen (1987) mukaan faktorointi, jolla tarkoitetaan latausmatriisin alustavan ratkaisun laskemista. 1960-luvun alkupuolella, kun tietokoneita ei juurikaan ollut käytettävissä, käytettiin sentroidimenetelmää, jossa oli mahdollista suorittaa faktorointi laskutoimituksilla. Sentroidimenetelmään voi tutustua esimerkiksi Vahervuon ja Ahmavaaran (1958, 56–77) teoksessa. Tilastollisten ohjelmien tarjoamia faktorointimenetelmiä ovat muun muassa pääakselimenetelmä, suurimman uskottavuuden menetelmä, pienimmän neliösumman menetelmä ja pääkomponenttimenetelmä. Pääkomponenttimenetelmä johtaa keskusteluun faktorianalyysin ja pääkomponenttianalyysin eroista, joita käsittelemme luvussa 8.2.

Leskinen (1987) toteaa, että rotaatiomenetelmiä, joilla faktoriratkaisuille pyritään löytämään tulkinnallinen ratkaisu, on olemassa lukuisia. Ne perustuvat kuitenkin erilaisiin teoreettisiin ja empiirisiin periaatteisiin. Rotatoinnilla pyritään useimmiten yksinkertaisen latausrakenteen periaatteeseen, jossa kukin muuttuja latautuisi mahdollisimman voimakkaasti yhteen faktoriin ja mahdollisimman vähän muihin faktoreihin. Kuten faktorointimenetelmiin, palaan myös rotaatiomenetelmiin luvussa 8.2.



Kuva 33 Faktorianalyysin vaiheet

Faktorianalyysin käytölle on esitetty sosiologian väitöskirjoissa useita perusteluja. Siksi on kiinnostavaa tarkastella, millaisia valintoja faktorianalyysia käytettäessä on tehty. An Gie Yong ja Sean Pearce (2013), Brett Williams, Andrys Onsmen ja Ted Brown (2010) kuten myös Costello ja Osborne (2005) ohjeistavat artikkeleissaan parhaiksi havaitsemiaan käytäntöjä faktorianalyysin soveltamiseksi tutkimusaineistojen analyysissa.

Kuvassa 33 on oma hahmotelmani faktorianalyysin suorittamisen vaiheista. Faktorianalyysia suorittaessaan tutkija valitsee muuttujat, faktorointimenetelmän ja rotaatiomenetelmän. Parasta mahdollista lopputulosta voidaan etsiä useilla kokeiluilla, esimerkiksi muuttujien poistamalla tai lisäämällä, erilaisilla faktorointi- ja rotaatiomenetelmillä tai eri faktorimäärillä.

Kolme ensimmäistä vaihetta kuvassa 33 ovat samat, mitä Leskinen (1987) toi edellä esiin. Faktorianalyysin tuloksella tarkoitan yhden analyysikokeilun tulosta, jonka tilastollista hyvyttä voidaan arvioida erilaisin tilastomatematisin keinoin. Leskinen (1987) puhuu tässä kohden riittävyystarkasteluista. Niiden antamien tulosten perusteella faktorianalyysi voidaan suorittaa lukuisia kertoja uudelleen tehden erilaisia muuttujavalintoja, faktorointivalintoja ja rotaatiomenetelmän valintoja, kunnes ollaan lopputulokseen tyytyväisiä ja voidaan tulkita saatu faktoriratkaisu sekä tehdä päätös mahdollisesta faktoreiden jatkoanalyysistä.

8.1. Muuttujien valinta – tilastotieteen puhdasoppisuus ja sosiologian tulkinta

Tilastotieteessä mittaaminen ja mittaamisen tuloksena syntyvät muuttujat nähdään eri tavalla kuin sosiologiassa. Tilastotieteessä mittaaminen yksinkertaistuu mitatasteikkoajatteluun, jossa muuttujat voidaan mitata neljällä eri tavalla. Alkeellisin mittaamisen tapa on sanallisten arvojen sijoittamista luokkiin, joille ei voida asettaa yksikäsitteistä järjestystä. Hieman kehittyneempi mittaaminen tapahtuu sanallisten arvojen sijoittamisella luokkiin, joille on löydettävissä yksikäsitteinen järjestys. Kaksi kehittyneintä mittaamisen tapaa, välimatka- ja suhdeasteikon mittaaminen, perustuvat numeeristen määrien mittaamiseen. Sosiologiassa mittaaminen on yleisesti haasteellisempaa ja monivaiheisempaa kuin tilastotieteessä, kuten edellä on useaan otteeseen tuotu esiin (ks. esim. Stouffer 1953; Patty ja Penn 2015; Smith ja Atkinson 2016; Uprichard 2016; Franzosi 2016). **Mitta-asteikkojen** näkökulmasta faktorianalyysiin valittavien muuttujien on oltava vähintään välimatka-asteikolla mitattuja, mitä vaatimusta sosiologisten aineistojen muuttujat harvoin täyttävät. Sosiologit ovat kuitenkin huomanneet, että mitatasteikkovaatimus ei estä menetelmien käyttämistä myös alemman mitatasteikon muuttujiin (ks. esim. Valkonen 1971, 39–40). Väitöskirjoissa on vaihtelevasti huomioitu tai tuotu esiin muuttujien mitatasteikot.

Faktorianalyysin muuttujavalintakriteerinä **korrelaatiokertoimet** on mainittu etenkin 1960-luvun väitöskirjoissa. Faktorianalyysin yhteydessä sosiologit ovat kuitenkin harvoin eritelleet käyttämäänsä korrelaatiokerrointa. Joitakin mainintoja siihen liittyen kuitenkin löytyy, kuten esimerkiksi Kauko Kämäräisen vuonna 1966 tarkastetussa väitöskirjassa.

Faktorianalyysin lähtökohtana käytettävät korrelaatiot laskettiin ns. Pearsonin tulomomenttikertoimina, joita lähes poikkeuksetta faktorianalyttisissä tutkimuksissa näkyvä käytetyin. (Kämäräinen 1966.)

Muuttujamuunnokset ovat myös aluetta, jossa tilastotieteen ja sosiologian ajattelu eroavat. Sosiologit tuovat väitöskirjoissaan jonkin verran esille myös muita muuttujien ominaisuuksille tilastotieteessä asetettuja vaatimuksia, joita heidän aineistossaan olevat muuttujat eivät kuitenkaan täytä. Tällaisia ovat esimerkiksi jakaumien normalisuus, symmetrisyys tai jatkuvuus. Muuttujille, jotka eivät noudata tilastotieteessä asetettuja ehtoja, voidaan tehdä erilaisia matemaattisia muunnoksia, joiden tavoitteena on saada muuttujat tilastomatemaattisesti sopivaan muotoon.

Olavi Riihinen kuvaa vuonna 1965 painetussa sosiaalipolitiikan väitöskirjassaan faktorianalyysin käyttöä kattavasti. Riihinen ottaa esiin muuttujien normalisoinnin. Hänen mukaansa useat tuon ajan tutkijat ovat normalisoineet faktorianalyysissä käytetyt muuttujat (Riihinen 1965, 114). Syiksi normalisointiin Riihinen arvelee nimenomaan tilastollisten menetelmien perustumista normaalijakaumaoletukseen. Sosiologian väitöskirjoissa mainitaan normalisointi muun muassa vuonna 1966 ja 1994 tarkastetuissa väitöskirjoissa:

Varsinkin ekologisissa tutkimuksissa yleisesti tarpeelliseksi katsottua muuttujien muuntamista samaa tilastojakaumaa noudattaviksi sanotaan normalisoimiseksi. Normalisointi suoritettiin T-asteikon perusteella. (Kämäräinen 1966.)

Lopulliseen faktorianalyysiin jäi näin neljä muuttujaa. Ensin havaitut muuttujat standardoitiin. Pääkomponenttianalyysiä käyttäen muodostettiin initiaaliratkaisu ja siitä valittiin kaksi pääkomponenttia rotatoitavaksi. (Nyyssölä 1994.)

Muuttujien jakauma voidaan saada muuttujamuunnosten avulla noudattamaan tilastotieteellisiä ihanteita, mutta niiden haittapuolena on muuttujan tulkinallisuuden menettäminen. Kahdesta sosiologian väitöskirjasta löytyy perusteluita, miksi esimerkiksi normalisointiin ei ole päädytty.

Faktorianalyysin muuttujille on tehty myös muita muunnoksia. Muutamassa väitöskirjassa kerrotaan, että muuttujille on tehty logaritmimuunnos (esim. Niemi

1978). Muuttujia on myös vakioitu osittaiskorrelaatiokertoimen avulla (esim. Rantalaiho 1968) ja joissakin tapauksissa myös dikotomisoitu (Leimu 1983; Heikkilä 1990). Dikotomisointia eli muuttujien luokittelua kaksiarvoisiksi on joissakin tapauksissa perusteltu sillä, että alkuperäisessä muodossaan muuttujat eivät täytä tilastotieteellisiä mitta-asteikkovaatimuksia, ja dikotomisoinnilla päästään toisenlaisen, tilastotieteellisessä mielessä luvallisen, menetelmäversion soveltamiseen.

Tavallisissa faktorianalyysimenetelmissä oletetaan, että havaittavat muuttujat on mitattu vähintään intervalliasteikolla. Koska tässä tarkasteltavat muuttujat eivät täytä kyseistä ehtoa, käytettiin erästä dikotomisten muuttujien käsittelyyn soveltuvaa faktorianalyysimallia. Tätä varten kaikki muuttujat dikotomisoitiin. (Leimu 1983)

Dikotomisointi on voinut toimia myös normalisoinnin vaihtoehtona, kun faktorianalyysiin on valittu eri asteikkoa olevia muuttujia.

Analyysiin sisällytetään kaikki seitsemän keskeistä elinolomuuttujaa sekä joukko sellaisia taustamuuttujia, joilla aiemman analyysin valossa voidaan olettaa olevan erottelevaa vaikutusta suhteessa elinolojen puutteisiin ja jotka ovat loogisesti toisistaan riippumattomia. Taustamuuttujat mittaavat asuinalueen taajama-astetta ja alueellista sijaintia, sukupuolta, ikää sekä perhetyyppejä (alaviite: kaikki koodattu 1/0-muuttujiksi...). (Heikkilä 1990.)

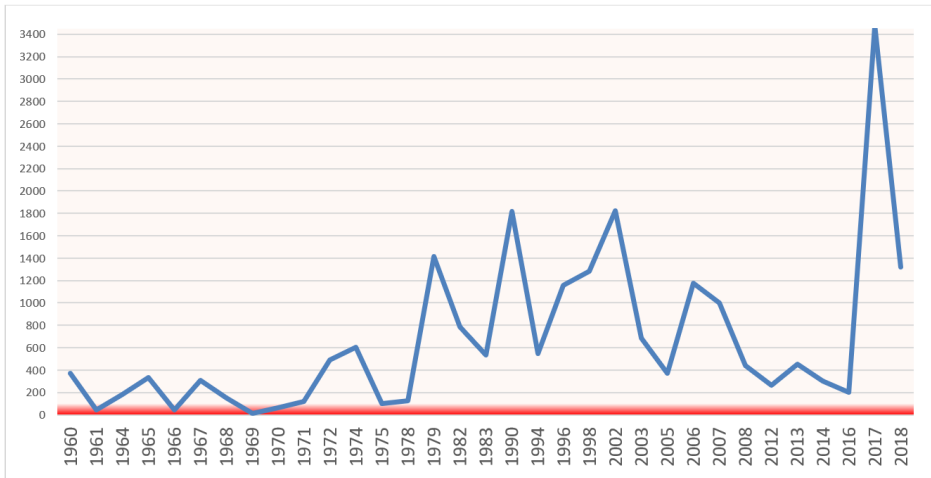
Erilaiset ja eri asteikkoa olevat muuttujat saadaan edellä mainittujen muunnosten avulla ”tasapäistettyä” eli muunnettua saman asteikkoisiksi ja ehkä vaikutuksiltaan samanarvoisiksi ja näin ollen tilastomatemaattisesti hyvin toimiviksi. Koska tilastomatemaattisissa muunnoksissa on vaarana tulkinnallisuuden menettäminen, on muuttujamuunnoksia käytettäessä syytä pohtia, kuinka paljon niitä voidaan tehdä sisällön ja tulkinnallisuuden kustannuksella. Erityisesti tieteissä, joissa tutkittavan kohteen mittaaminen on epätarkkaa, olisi mitattujen muuttujien analysoinnissa syytä olla maltillinen ja käyttää vain sellaisia menetelmiä, joiden tulkinnallisuuden pystyy takaamaan. Rantalaiho (1998) ja Alastalo (2005, 13) ovat tuoneet esiin vertauksen Keisarin uudet vaatteet -satuun viitatessaan kvantitatiivisen tutkimuksen sudenkuoppiin, erityisesti mittaamiseen ja aineiston keruuseen liittyen. Sama vertaus sopii myös muuttujamuunnosten käyttämiseen kvantitatiivisessa analyysissä. Muuttujamuunnokset ovat matemaattisesti vakuuttavia ja niiden käyttäminen on tilastomatemaattisesti perusteltua. Liiallinen matemaattisuus ja sisällöstä irrottautuminen voi kuitenkin aiheuttaa tilanteen, jossa tutkija ja tutkimusta lukeva yleisö eivät enää pysty ymmärtämään analyysin tuottamaa tulosta, mutta vakuuttuneina menetelmän matemaattisesta hienoudesta ovat ymmärtävinään sen.

Muuttujien arvot. Muutamissa 1960- ja 1970-luvulla suoritetuissa faktorianaalyyseissa on mukana hyvin erilaisia arvoja saavia muuttujia. Syyksi tähän arvelen faktorianaalyydin suorittamisen suurissa keskustietokoneissa. Koska yhden faktorianaalyydin suorittamiseen meni aikaa joskus jopa viikko (Töttö 2004, 196; Kutvonen 2004), on ymmärrettävää, että analyyseissa haluttiin pitää mukana niin monia muuttujia kuin mahdollista.

Aineiston koko. Siitä, kuinka paljon havaintoja, vastaajia, aineistossa on oltava faktorianaalyydin suorittamiseksi, on olemassa monia näkemyksiä (ks. esim. Hair ym. 1998). Kuva 34 esittää sosiologian väitöskirjojen faktorianaalyyseissa käytettyjen aineistojen koot. Niissä havaintojen määrät vaihtelevat 15:stä 16715:een. Pienimmät aineistot ovat olleet käytössä vuosina 1961–1970 tarkastetuissa väitöskirjoissa. Alle sadan havainnon alue on merkitty punaisella lähinnä osoittamaan erästä tilastomatematisessa mielessä asetettua minimirajaa aineiston koolle. Pääsääntöisesti sosiologit ovat pystyneet ”ylittämään” tämän rajan. Aineistojen koot kasvavat selvästi 1970-luvun lopulta alkaen. Ensimmäinen mieleen tuleva selitys aineistojen havaintomäärien kasvuun on tietokoneiden ja tietoverkkojen tuoma helpotus työmäärään. Aluksi tietokoneet helpottivat aineiston tallentamista ja analyydin suorittamista, mutta 2000-luvulla tietoverkot mahdollistivat myös tietojen sähköisen keräämisen ja niiden automaattisen tallentumisen aineistoiksi. Sähköiset kyselyt ovat mahdollistaneet postikyselyä suuremman vastaajamäärän hankkimisen. Postikyselyä tehtäessä tutkijan on jo aineistonkeruuta suunnitellessaan laskettava vastaajamäärän aiheuttamat kustannukset, jotka aiheutuvat otannan suorittamisesta ja postimaksuista. Kustannusten ohella myös muut seikat, esimerkiksi maantieteellinen sijainti, rajoittavat aineistonkeruun kattavuutta. Toisaalta kyselyjen yleistyessä on vastaajakadon määrä lisääntynyt. 1970-luvulla pyrittiin 80 %:n vastausprosenttiin, 1980-luvulla oltiin tyytyväisiä 70 %:iin. Vastausprosenttien alkaessa laskea pidettiin ”kipurajana” 50 %:a, mutta sitäkin on luovuttu 2000-luvulla (ks. esim. Taanila 2019; Ruskoaho ym. 2010; de Leeuw 2008). Vastausprosentista ja erilaisista tiedonkeruutavoista on olemassa kansainvälisiä tutkimuksia, mutta suomalaisia tutkimuksia on vähän. Tällä hetkellä tiedonkeruutapojen tutkimisessa on haasteena se, että käytännöt muuttuvat nykyisinä aikoina niin nopeasti, että tutkimusten on vaikeaa pysyä muutosten tahdissa.

Analyysiin valittujen muuttujien määrä. Analyysiin valittavien muuttujien määrän suhteen sekä menetelmäoppaat että sosiologit tasapainottelevat matemaattisten ja sisällöllisten kriteerien välillä. Matemaattiselta kannalta on suositeltavaa pitää muuttujien määrä maltillisena. Hair ym. (1998, 99) havainnollistavat muuttujien määrän vaikutusta esimerkiksi, jossa faktorianaalyyssiin valitaan 30 muuttujaa, jolloin niiden välisiä korrelaatioita syntyy 435 kappaletta. Näin suuresta korrelaatioiden määrästä seuraa, että 5 %:n merkitsevyydellä voi jopa 20 korrelaatiota olla tilastollisesti merkitseviä vain sattumalta. Sisällölliset seikat huomioiden taas on suositeltavaa, että muuttujia valitaan riittävästi niin, että ne kattavat faktorianaalyydin

kohteena oleva aihepiirin riittävässä määrin. Sosiologian väitöskirjoissa faktorianaalyyksien muuttujamäärä vaihtelee välillä 3–56.



Kuva 34 Faktorianaalyyksissä käytettyjen aineiston koot vuosina 1960–2017

Faktorianaalyyksien palvelee sosiologista tutkimusta parhaiten silloin, kun muuttujavalinnoissa huomioidaan myös sisällölliset kriteerit. Ne voivat perustua esimerkiksi omaan harkintaan, aikaisempaan tutkimustietoon ja/tai teoreettisiin lähtökohtiin. Tätä korostavat myös Hair ym. (1998, 96–97) kirjoittaessaan, että faktorianaalyysi ei pysty tekemään ihmeitä. Jos tutkija ei käytä lainkaan harkintaansa muuttujien valinnassa, on turha odottaa analyyksinkaan paljastavan mitään kovin järkeviä tuloksia. Tällöin faktorianaalyyksissä on vaarana ”roskaa sisään, roskaa ulos” -ilmiö. (Hair ym. 1998, 96–97.) Sosiologit tuovat väitöskirjoissaan usein esille muuttujavalintojen sisällölliset perusteet. Faktorianaalyysi on rajattu muuttujiin, jotka sisällöllisesti liittyvät tutkimuksen kohteena olevaan asiakokonaaisuuteen tai käsitteeseen.

Faktorianaalyyksiin valittujen muuttujien valinnassa on kiinnitetty nimenomaan huomiota paitsi siihen, että mukaan on otettu suoraan tyytyväisyyttä mittaavia muuttujia (3 kpl), myös siihen, että mukaan on saatu riittävä määrä sellaisia muuttujia, joilla on tutkimuksen aikaisemman vaiheen tulosten perusteella syytä olettaa olevan yhteyttä tyytyväisyyteen. (Uitto 1964.)

Edellä käsitellyjä elinkeinoihin liittyviä kysymyksiä on lopuksi tarkasteltu faktorianaalyyksien avulla. (Asp, 1965.)

Faktorianalyysiin sisällytettiin kaikki kuusi yksilön hyvinvointia mittaavaa johdettua deprivatiomuuttujaa sekä toisistaan erikseen köyhyyttä indikoiva muuttuja ja pienituloisuutta indikoiva muuttuja. (Heikkilä 1990.)

Ensimmäisessä vaiheessa analyysistä jätettiin pois muuttuja d, koska sen sisällöllinen tulkinta oli vaikeaa. Lisäksi kyseinen muuttuja latautui lähes yksinään omalle faktorilleen, joten sen mukaan ottaminen olisi tästäkin syystä ollut ongelmallista. (Kuussaari 2006.)

Aineiston koko ja muuttujien määrä eivät kumpikaan yksinään tarkastellen kerro riittävästi vaan ne on suhteutettava toisiinsa. Matemaattisena suosituksena havaintoja olisi oltava vähintään viisi kertaa - joidenkin näkemyksen mukaan jopa kymmenen tai kaksikymmentä kertaa - enemmän kuin faktorianalyysiin valittuja muuttujia. Sosiologian väitöskirjoissa tämä suhde vaihtelee välillä 0,31–1193,93. Faktorianalyysin innostuksen vuosina 1960–1970-luvuilla muuttujien ja havaintojen määrän suhde jäi usein alle viiden. Tuolloin joko faktorianalyysiin otettiin paljon muuttujia (jopa kaikki aineiston muuttujat) tai sitten aineisto oli pienikokoinen. 1970-luvun loppupuolelta alkaen havaintojen ja muuttujien määrän suhde on tyydyttänyt nämä tilastomatemattiset ehdot eli faktorianalyysissä on ollut vähintään viisi kertaa enemmän havaintoja kuin muuttujia.

8.2. Faktorointi, rotatointi ja jatkoanalyysi

Koska faktorianalyysin suorittamiseen liittyvien valintojen tarkastelussa on huomioitava tietokoneiden ja erityisesti tilastollisten ohjelmien vaikutus, olen jakanut vuodesta 1960 vuoteen 2018 ulottuvan tarkasteluni neljään ajanjaksoon tietokoneiden ja tilastollisten ohjelmien käyttömahdollisuuksien mukaan. Ensimmäinen ajanjakso sisältää vuodet 1960–1965, jolloin faktorianalyysi on suoritettu itse laskien tai keskustietokoneella laskentakeskuksissa. Toinen ajanjakso alkaa vuodesta 1966 ja jatkuu vuoteen 1985. Tällöin analyysit on suoritettu keskustietokoneilla. Tutkijoilla on ollut vuodesta 1986 alkaen mahdollisuus käyttää henkilökohtaisia tietokoneita ja tilastollisia ohjelmia. Ajanjaksolla 1986–1999 tietokoneiden kapasiteetti ja tietoverkkojen käyttömahdollisuudet olivat kuitenkin vielä kohtalaisen rajoitettuja. Viimeinen ajanjakso kattaa 2000-luvun, jolle on ominaista etenkin tietoverkkojen tehokas hyödyntäminen.

Keskustietokoneiden ja reikäkorttien aikaan eli suurin piirtein 1980-luvun puoliväliin asti tutkijoiden analyysimahdollisuudet olivat rajalliset. Jos tutkija olisi mielinnyt lisä- tai tarkistusanalyysija, ne olisi pitänyt tehdä käsityönä. Siihen, minkälaisena aineiston analyysi toteutui, vaikuttivat siis käytännölliset syyt yhtä lailla kuin puhdas tieteellinen harkinta, kirjoittaa Alastalo (2005, 252–253).

Myös Töttö (2004, 196) tuo esiin, miten erilaisiin kokeiluihin ei juurikaan ollut mahdollisuutta:

Reikäkortit piti kantaa tietokonekeskukseen ja odottaa viikko, jotta pääsi keskustelemaan huoneen kokoisen koneen kanssa. Kone vastaili näytöttömältä päätteeltä naputeltuihin komentoihin lipposmaisesti ”syntax error” eikä muuta. Hermostua ei auttanut, piti vain miettiä tarkkaan, mitä oli tekemässä. Tulokset sai aikaisintaan seuraavana päivänä. (Töttö 2004, 196.)

1960-luvun alussa Suomessa ei vielä ollut osaamista tietokoneiden käyttöön vaan laskentakeskusten työntekijät lähetettiin aluksi ulkomaille koulutautumaan. Kun osaamista alkoi karttua, pidettiin henkilökunnalle sisäisiä koulutustilaisuuksia ja tietokoneen käyttäjille laadittiin opetusmonisteita. Myös laskentakeskusten tietokoneiden kapasiteetti oli rajoitettu. Yhden analyysimenetelmän suorittamiseen saattoi kulu viikko, vaikka tietokoneajoja suoritettiin yölläkin, kirjoittaa 1960-luvulla Laskentakeskuksessa työskennellyt Heikki Kutvonen (2004). Näistäkin kuvauksista voi päätellä, että tutkimustaan tekevät sosiologit eivät pystyneet kokeilemaan analyysien suorittamista useita kertoja vaan heidän oli hyväksyttävä yhden analyysikokeilun tuottama tulos. Laskentakeskuksen työntekijä Monika Partonen (2004) muistelee 1960-luvun toimintaa Kaapelitehtaan tietokoneosastolla, jonne monet 1960-luvun sosiologit lähettivät aineistonsa analysoitavaksi: ”Kaapelitehtaalla tehtiin melkein kaikkea faktorianalyyseistä ja Survo-ohjelmointikielestä laivan vakavuuslaskentaan ja musiikkiohjelmiin”.

2000-luvulla analyysien suoritusmahdollisuudet ovat aivan toista luokkaa kuin edellisillä vuosikymmenillä. Alastalo (2014, 283) tuo esille lainauksen Sosiologia-lehden numerosta 3 vuodelta 1995, jossa sanotaan, että tietotekniikan kehittyminen on ravistellut yhteiskuntatieteilijöiden työtapoja viimeisten vuosikymmenien aikana ankarasti ja mahdollistanut tilastollisten analyysien tekemisen innovatiivisemmassa ja helppokäyttöisemmässä mikrotietokonemaailmassa. Tutkijoilla on 2000-luvulla käytössään tehokkaat henkilökohtaiset tietokoneet, monia tilastollisia ohjelmia ja hyvä valikoima menetelmäoppaita sekä kirjallisissa ja sähköisessä muodossa. Tietoverkot mahdollistavat tiedon ja ohjeiden melkein pärajattoman saatavuuden. Käänteisenä puolena tälle rajattoman tiedon saatavuudelle syntyy niin sanottu informaatioähy tai tietotulva, kuten Jussi Simpura ja Jussi Melkas (2013, 16) toteavat.

Taulukko 10 Faktorianalyysin suorittamisessa käytetyt tilastolliset ohjelmat eri aikoina

| | 1960-1965 | 1966-1985 | 1986-1999 | 2000-2018 | Yhteensä |
|-----------------|-----------|-----------|-----------|-----------------|----------|
| laskettu itse | 3 | 1 | 0 | 0 | 4 |
| keskustietokone | 2 | 12 | 0 | 0 | 14 |
| GLIM | 0 | 0 | 1 | 0 | 1 |
| SAS | 0 | 0 | 2 | 1 | 3 |
| SPSS | 0 | 0 | 2 | 12 ¹ | 14 |
| R | 0 | 0 | 0 | 1 | 1 |
| ei tietoa | 0 | 11 | 2 | 6 | 19 |
| Yhteensä | 5 | 24 | 7 | 20 | 56 |

1. SPSS:n ohella yhdessä väitöskirjassa käytetty Survoa ja yhdessä Stata-ohjelmaa

Taulukossa 10 näkyy faktorianalyysia sisältävien väitöskirjan osalta tieto siitä, mitä tilastollisia ohjelmia tutkijat ovat käyttäneet. Useimmiten analyysiin käytetty ohjelma on jätetty mainitsematta. Henkilökohtaisten tietokoneiden aikakaudella SPSS (Statistical Package for Social Sciences) on ollut suosituin tilastollinen ohjelma.

Faktorointi

Faktorointimenetelmä tarkoittaa tapaa, jolla korrelaatiomatriisista lähdetään hakemaan rakennetta, faktoreita. Sosiologien väitöskirjoissaan ilmoittamia faktorointimenetelmiä ovat sentroidi-, pääakseli-, pääkomponentti- sekä pienimmän neliösumman (pns) menetelmät. 16:n faktorianalyysia sisältävän väitöskirjan osalta ei faktorointimenetelmästä ole tietoa.

Taulukko 11 Faktorointimenetelmät eri aikoina

| | 1960-1965 | 1966-1985 | 1986-1999 | 2000-2018 | Yhteensä |
|---------------------------------|-----------|-----------|-----------|-----------|----------|
| sentroidi | 2 | 0 | 0 | 0 | 2 |
| pääakseli | 2 | 11 | 1 | 5 | 19 |
| pääkomponentti | 0 | 5 | 4 | 6 | 15 |
| pääakseli ja pääkomponentti | 0 | 1 | 0 | 1 | 2 |
| pienimmän neliösumman menetelmä | 0 | 0 | 0 | 2 | 2 |
| ei mainittu | 1 | 7 | 2 | 6 | 16 |
| Yhteensä | 5 | 24 | 7 | 20 | 56 |

Taulukosta 11 käy ilmi, että suosituin faktorointimenetelmä sosiologien väitöskirjoissa on ollut pääakselimenetelmä, jota on käytetty 19 väitöskirjassa sellaisenaan ja kahdessa pääkomponenttimenetelmän ohella. Erityisen suosittua pääakselimenetelmä on ollut vuosina 1966–1985, jolloin sitä mainitaan käytetyn

14 väitöskirjassa. 2000-luvulla tarkastetuissa väitöskirjoissa kuudessa kerrotaan käytetyn faktorianalyysin yhteydessä pääakselimenetelmää. Toiseksi suosituin – jos näin voi sanoa, koska 16 väitöskirjan osalta tieto puuttuu – on ollut pääkomponenttimenetelmä, joka on mainittu 16 väitöskirjassa. Pääkomponenttimenetelmän suosio jakautuu tasaisemmin verrattuna pääakselimenetelmään. Vuosina 1966–1985 kuudessa väitöskirjassa mainitaan pääkomponenttimenetelmä, vuosina 1986–1999 neljässä ja 2000-luvulla seitsemässä väitöskirjassa. Kahdessa väitöskirjassa, jotka on tarkastettu vuosina 1970 ja 2005 mainitaan käytetyn sekä pääakseli- että pääkomponenttimenetelmiä.

Leskinen ja Kuusinen (1991) ovat tutkineet faktorianalyysin käyttöä kasvatustieteellisissä väitöstutkimuksissa vuosina 1965–1989. Kasvatustieteissä yleisin faktorointimenetelmä oli tuolloin pääakselimenetelmä, jota oli käytetty 56 tutkimuksessa (88 %). Pääkomponenttimenetelmä oli valittu viidessä väitöstutkimuksessa (7 %) ja sentroidimenetelmä kolmessa (5 %). Sosiologiassa käytettyihin faktorointimenetelmiin verrattuna huomataan, että menetelmien suosio on järjestykseltään samanlainen, mutta sosiologiassa suosion erot eivät ole niin suuret kuin kasvatustieteissä.

Sosiologit eivät juurikaan ole perustelleet valitsemaansa faktorointimenetelmää. Arvelen faktorointimenetelmän olleen kytköksissä kulloinkin käytettävissä olevan tilastollisen ohjelman oletusarvoon. Liitetaulukossa 1 esitetään, mitä faktorointimenetelmiä eri tietokoneohjelmia käyttäneet sosiologit ovat valinneet. 1960-luvun alkupuolella, kun tietokoneita ei juurikaan ollut käytettävissä, on faktorointimenetelmänä käytetty sentroidimenetelmää. Sentroidimenetelmä mahdollisti faktoroinnin ja faktorianalyysin suorittamisen itse tehdyillä laskutoimituksilla. Vuosina 1966–1985 faktorianalyysit on suoritettu keskustietokoneilla. Keskustietokoneet sijaitsivat suurissa laskentakeskuksissa ja niitä käyttivät laskentakeskusten työntekijät. Tutkijalla ei tällöin juurikaan ollut sananvaltaa siihen, miten faktorianalyysi suoritettiin tai mikä faktorointimenetelmä valittiin. Keskustietokoneiden aikaan eniten käytetty faktorointimenetelmä on ollut pääakselimenetelmä ja seuraavaksi eniten pääkomponenttimenetelmä. Tosin kuudessa väitöskirjassa faktorointimenetelmiä ei ole mainittu. Henkilökohtaiset tietokoneet ja tilastolliset ohjelmat alkoivat yleistyä 1980-luvun puolivälissä, jonka jälkeen käytetyin tilastollinen ohjelma sosiologien väitöstutkimuksissa on ollut SPSS-ohjelma. SPSS-ohjelmassa faktorianalyysin oletusmenetelmänä on pääkomponenttimenetelmä. Liitetaulukosta 1 on luettavissa pieniä viitteitä siitä, että faktorianalyysi olisi suoritettu ohjelman oletusarvoa käyttäen eikä niinkään tietoisesti valiten.

Koska sosiologian väitöskirjoissa ei valittua faktorointimenetelmää useinkaan ole perusteltu, ei myöskään pääakseli- ja pääkomponenttimenetelmien eroa ole paljolti kommentoitu. Kahdesta väitöskirjasta erojen vertailua kuitenkin löytyy. Ensimmäinen on vuonna 1969 tarkastettu Kalimon väitöstutkimus, jossa Kalimo kirjoittaa seuraavasti:

Faktorointi suoritettiin pääakselimenetelmällä käyttämällä kommunaliteetien estimaatteina kunkin muuttujan itseisarvoltaan korkeinta korrelaatiota. Tätä kommunaliteetin estimointimenetelmää ei pidetä täysin tyydyttävänä muuttujien luvun ollessa pienempi kuin 10–20. Luotettavampaa vähätöistä estimointimenetelmää ei kuitenkaan ole kehitetty. Ykkösen käyttäminen kommunaliteetin estimaattina olisi johtanut faktorianalyysin sijasta pääkomponenttianalyysiin ja faktoreita lukuisampiin pääkomponentteihin. (Kalimo 1969, 173.)

Toinen sosiologian väitöskirja, jossa perustellaan, miksi on päädytty käyttämään pääkomponenttimenetelmää, on Juhani Tauriainen väitöstutkimus vuodelta 1970. Tauriainen (1970, 193–194) nojaa perustelunsa kokonaisvarianssin, spesifivarianssin ja virhevarianssin erotteluun (ks. kuva 36) ja kirjoittaa päätyneensä pääkomponenttianalyysiin, koska oli edullista lähteä kokonaisvarianssianalysoinnista, kun tutkimuksessa pyrittiin yhdistettyjen muuttujien tutkimiseen. Tauriainen viittaa perusteluissaan Cooleyn ja Lohnesin¹⁷ teokseen vuodelta 1962.

Tämän tutkimuksien analyysien lähtökohtana oli muuttujien kokonaisvarianssi. Laskentateknisessä mielessä se merkitsee alkuperäisten muuttujien korrelaatiomatriisin faktorointia niin, että ykkösiä käytetään matriisin päädiagonaalilla eikä kommunaliteettien estimaatteja. (Tauriainen 1970, 193.)

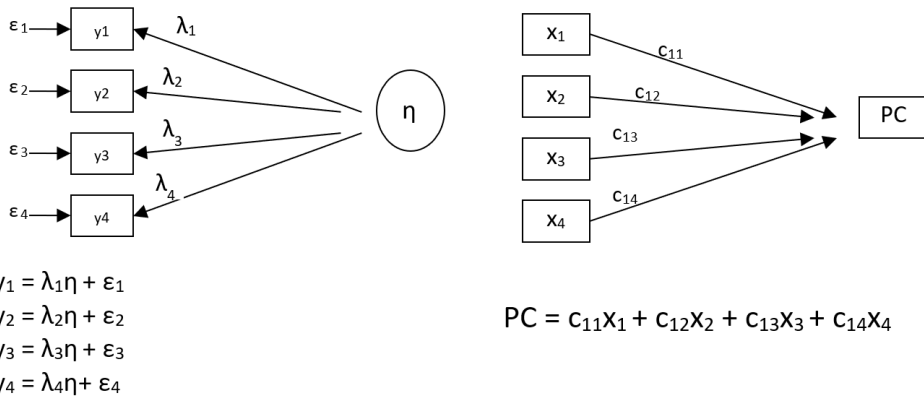
Pääakseli- ja pääkomponenttimenetelmän valinta johtaa keskusteluun faktorianalyysin ja pääkomponenttianalyysin erosta (ks. esim. Joliffe ja Morgan 1992; Brown 2009; Bandalos ja Boehm-Kaufman 2010). Pääakselimenetelmä luetaan yksimielisesti faktorianalyysiksi, mutta pääkomponenttimenetelmän osalta on vaihtelevia näkemyksiä sen suhteen, voidaanko analyysia kutsua faktorianalyysiksi¹⁸ vai täytyykö se erottaa faktorianalyysista kutsumalla sitä pääkomponenttianalyysiksi (ks. esim. Costello ja Osborne 2005, 1; Ranta, Rita ja Kouki 1997; ks. myös Reunanen ja Suikkanen 1998; Nummenmaa ym. 1997).

Mikä on siis faktorianalyysin ja pääkomponenttianalyysin ero? Tilastomatemattisesti tarkasteltuna eroja on kolme. Ensimmäinen eroavaisuus on siinä, että faktorianalyysin taustalla on tilastollinen malli, kun pääkomponenttianalyysissa ei sellaista ole. Tilastollisessa faktorimallissa faktori η vaikuttaa havaittuihin y_i -muuttujiin. Sen lisäksi y_i -muuttujiin vaikuttavat jäännöstermit ε_i . Pääkomponenttianalyysin pääkomponentit ovat vain havaittujen muuttujien lineaariyhdistelmiä ilman tilastollis-

17 Cooley, W.W. ja Lohnes, P.R. (1962) *Multivariate procedures for the behavioral sciences*. John Wiley & Sons. New York. 160–164.

18 Esimerkiksi Everitt, Brian ja Hay, Dale (1992). *Talking about statistics. A Psychologist's Guide to Data Analysis*. Edward Arnold. London, Melbourne, Auckland. 112–114.

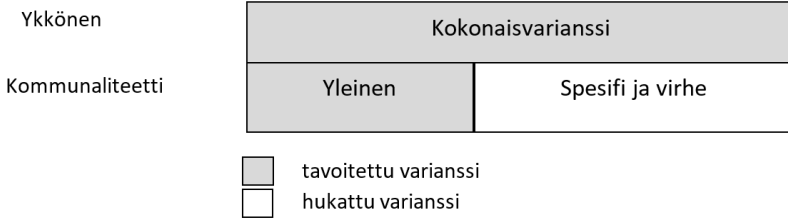
ta mallia (kuva 35). Tässä mielessä faktorianalyysin ja pääkomponenttianalyysin ero on selvä. Mutta kvantitatiivisessa sosiologisessa tutkimuksessa – haluttaessa löytää suuresta muuttujajoukosta keskenään yhteenkuuluvia muuttujia – herää kysymys, onko erolla merkitystä. Molemmat menetelmät antavat samanlaisen tiedon siitä, kuinka monta muuttujaryhmää – ovat ne faktoreita tai pääkomponentteja – syntyy ja mitkä muuttujat kuuluvat mihinkin ryhmään.



Kuva 35 Vasemmalla yhden faktorin malli ja oikealla yhden pääkomponentin malli graafisina esityksinä ja matemaattisina yhtälöinä. (Lähteet: Leskinen 1987; Nummenmaa ym. 1997, 230; 242.)

Toinen faktorianalyysin ja pääkomponenttianalyysin tilastomatemattinen ero liittyy vaihtelun selittämiseen. Hair ym. (1998, 100–103) esittävät kuvassa 36 faktorianalyysiin ja pääkomponenttianalyysiin liittyvät varianssit. Yleinen varianssi tarkoittaa muuttujan sellaista varianssia, joka on jaettu kaikkien analyysissä olevien muuttujien kanssa. Spesifi varianssi puolestaan liittyy vain muuttujaan itseensä. Virhevarianssi on mittausvirheestä, satunnaisvirheestä ym. johtuva varianssi. Hair ym. ohjeistavat valitsemaan pääkomponenttimenetelmän silloin, kun ensisijaisesti halutaan löytää mahdollisimman pieni määrä komponentteja eli muuttujaryhmiä kattamaan mahdollisimman suuren osuuden muuttujien varianssista. Pääkomponenttimenetelmä on heidän mukaansa sopiva myös silloin, kun oletetaan spesifin ja virhevarianssin edustavan suhteellisen pientä osuutta kokonaisvarianssista. Faktorianalyysiin kannattaa Hairin ym. mukaan päätyä silloin, kun ensisijaisena tavoitteena on identifioida latentit ulottuvuudet ja kun spesifin ja virhevarianssin määrästä on vain vähän tietoa ja ne halutaan eliminoida tarkastelusta. Faktorianalyysillä pyritään muuttujien välisen vaihtelun selittämiseen, kun taas pääkomponenttianalyysissä tavoitteena on selittää muuttujien kokonaisvaihtelua. (Hair ym. 1998, 100–103.)

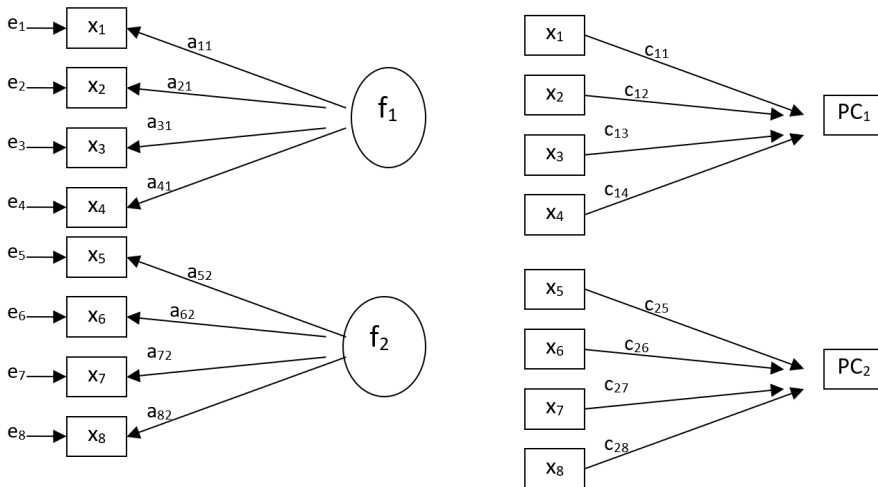
Diagonaaliarvo



Kuva 36 Faktorimatriisin variانسien tyypit (Lähde: Hair ym. 1998, 102)

Kolmas tilastomatemattinen ero liittyy faktoreiden tai pääkomponenttien riippuvuuteen. Faktorianalyyssissa faktorit voivat olla riippuvia tai riippumattomia keskenään. Pääkomponenttianalyyssissa ne ovat lähtökohtaisesti riippumattomia, kuten Nummenmaa ym. (1997, 229–230) kuvaavat. He kirjoittavat, että pääkomponenttimenetelmän tarkoituksena on tiivistää muuttujien vaihtelu muutamaan pääkomponenttiin siten, että ne selittävät mahdollisimman paljon muuttujien kokonaisvaihtelua. Muodostuvien pääkomponenttien järjestyksellä on väliä. Ensimmäinen pääkomponentti selittää eniten muuttujien kokonaisvariانسsia. Toinen pääkomponentti selittää jäljelle jäänyttä variانسia ja on korreloimaton ensimmäisen pääkomponentin kanssa. Lopputuloksena pääkomponentit ovat variانسiansa suhteen suuruusjärjestyksessä ja keskenään korreloimattomia. (Nummenmaa ym. 1997, 229–230.)

Kuvassa 37 esitetään vielä kertaalleen rinnakkain pääakselimenetelmää hyödyntävä faktorimalli ja pääkomponenttimenetelmää hyödyntävän analyysi, jota kutsutaan sekä pääkomponenttianalyyssiksi että faktorianalyyssiksi.



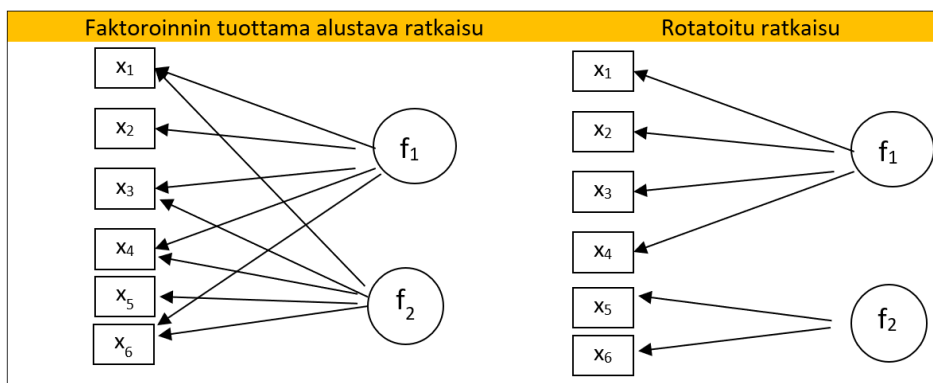
Kuva 37 Faktorianalyysi (vasemmalla) ja pääkomponenttianalyysi (oikealla) tiivistävinä menetelminä

Tutkimustaan tekevää sosiologia kiinnostaa, mitkä muuttujat ryhmittyvät yhteen ja kuinka monta muuttujaryhmää syntyy. Molemmat menetelmät antavat tämän tiedon aika yhteneväisin tuloksin (Bryman ja Cramer 2004, 28). Menetelmien tilastomatemattiset erot ovat selkeitä, mutta sosiologisen tutkimuksen kannalta oleellisemmaksi näen tutkijan eli sosiologin omaan asiantuntemukseen perustuvan arvioinnin siitä, ovatko analyysin tuottamat muuttujaryhmät eli faktorit tai pääkomponentit järkeviä ja tulkittavia, pystyykö hän nimeämään ne ja miten hän käsittelee niitä jatkoanalyysissa. Tässä mielessä näen faktorianalyysin ja pääkomponenttianalyysin eron merkitykselliseksi.

Sosiologian väitöskirjoissa ei eroihin juurikaan ole otettu kantaa ja niissä on vaihtelevia käytäntöjä sen suhteen, onko käytetty menetelmä nimetty faktorianalyysiksi, vaikka olisi käytetty pääkomponenttimenetelmää vai onko menetelmää kutsuttu pääkomponenttianalyysiksi.

Rotatointi

Faktoroinnin tuottama esitys muuttujien latautumisesta eri faktoreihin on tulkinnallisesti haastava, koska tässä vaiheessa muuttujat latautuvat useaan faktoriin. Faktoroinnin tuottamaa tulosta pidetäänkin vasta alustavana faktoriratkaisuna, johon lähdetään hakemaan tulkinnallista selkeyttä rotaation avulla. Rotatoinnilla pyritään yksinkertaiseen latausrakenteeseen, jossa kukin muuttuja latautuisi selkeästi vain yhteen faktoriin (kuva 38).

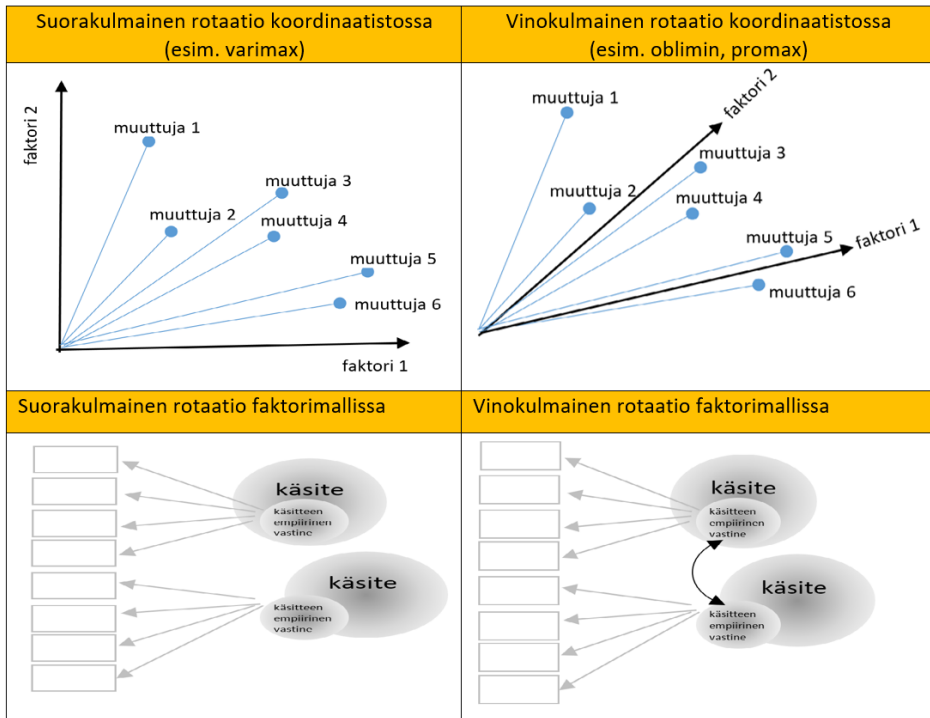


Kuva 38 Rotatoinnilla pyritään yksinkertaiseen latausrakenteeseen

Faktorianalyysia käyttävien opiskelijoiden ja tutkijoiden taholta kuulee usein kysymyksen, miten toimitaan tilanteissa, joissa yksinkertainen latausrakenne ei toteudu, vaan muuttujat latautuvat useaan faktoriin. Hair ym. (2005, 133) esittävät teoksessaan kolme vaihtoehtoa. Ensimmäinen on muuttujan poistaminen analyysistä. Toinen vaihtoehto on sijoittaa muuttuja siihen faktoriin, johon se latautuu

korkeammin. Kolmas vaihtoehto on pitää muuttuja mukana kaikissa faktoreissa, mihin se latautuu riittävän korkeasti. Kaikissa näissä vaihtoehtoissa on ratkaisevassa asemassa tutkijan asiantuntemus ja sen pohjalta tehty harkinta. Kiistanalaisin näistä vaihtoehtoista lienee kolmas. Kiistanalaisuudestaan huolimatta se on sosiologisessa tutkimuksessa perusteltu vaihtoehto, koska tutkimuksen kohteena olevat sosiologiset ilmiöt ja käsitteet ovat itsessään moniulotteisia ja monitulkintaisia. Siten on järkeenkäypää, että niitä vastaavat muuttujat sisältävät myös moniulotteista tietoa ja ovat siten kytköksissä myös moneen faktoriin.

Rotaatiomenetelmien osalta keskeisin valinta liittyy suorakulmaiseen ja vinokulmaiseen rotaatioon. Kuvassa 39 on kaksi esitystapaa suoran ja vinon rotaation eroavaisuudelle. Rotaation matemaattinen idea ilmenee ylemmästä esitystavasta, jossa faktorianalyysin muuttujat on piirretty koordinaatistoon origosta lähteviksi vektoreiksi. Vektoreiden loppupisteet määräytyvät faktorilatausten perusteella. Koordinaattiakselit kuvaavat faktoreita, joiden avulla muuttujien väliset yhteydet pyritään selittämään. Suorassa rotaatiossa faktorit ovat 90 asteen kulmassa toisiinsa nähden eli ne ovat toisistaan riippumattomia. Vinokulmaisessa rotaatiossa faktoriakselit eivät ole kohtisuorassa toisiaan vasten, mikä tarkoittaa, että faktoreiden välillä vallitsee jonkinlainen riippuvuus. Alemmassa esitystavassa suoran ja vinon rotaation ero esitetään sosiologiseen tutkimukseen istutetulla faktorimallilla, jossa faktoreiden tulkitaan edustavan käsitteellisiä ulottuvuuksia. Suorassa rotaatiossa oletetaan, että käsitteen ulottuvuudet eivät ole kytköksissä keskenään, kun taas vinossa rotaatiossa huomioidaan ulottuvuuksien välillä mahdollisesti olevan jonkinlainen riippuvuus.



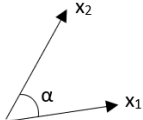
Kuva 39 Suora ja vino rotaatio koordinaatistossa ja faktorimallissa

Sosiologian väitöskirjoissa mainitut rotaatiomenetelmät esitetään taulukossa 12. Sosiologien esittämät kuvaukset rotaation suorittamisesta vaihtelevat paljon. Jotkut kuvaukset ovat hyvin yksityiskohtaisia, tilastomatemattisia, monia vaihtoehtoja esille tuovia tai pohtivia. Enimmäkseen kuvaukset ovat kuitenkin lyhyitä mainintoja ilman perusteluita. Rotatointivaiheen suorittaminen on myös vaihdellut aikojen kuluessa. 1960- ja 1970-luvun väitöstutkimuksissa tuodaan esille erillisenä päätöksenä muuttujien tai faktoreiden valinta rotaatiovaiheeseen. Tämä vaihe jää piiloon nyky-aikaisissa tilasto-ohjelmissa ja sen myötä myös väitöskirjojen kuvauksissa.

Taulukko 12 Sosiologien käyttämät rotatointimenetelmät

| | Väitöskirjojen lkm | Väitöskirjojen tarkastusvuodet |
|------------------------|--------------------|--------------------------------|
| Kosini | 4 | 1960–1965 |
| Kosini ja varimax | 4 | 1968–1974 |
| Varimax | 29 | 1966–2018 |
| Combined ja varimax | 2 | 1967–1971 |
| Oblique - vinorotaatio | 6 | 1965, 2002–2014 |
| Promax - vinorotaatio | 3 | 2007-2018 |

Vanhin väitöskirjoissa esiintyvä rotaatiomenetelmä on nimeltään kosiniratkaisu, jota on käytetty neljässä 1960-luvun alkupuolella tarkastetussa väitöskirjassa, jotka ovat Paavo Piepposen (1960), Antti Eskolan (1961), Yrjö Uiton (1964) ja Seppo Randellin (1965). Kosiniratkaisu perustuu tapaan nähdä faktorit vektoreina. Muuttujien välinen korrelaatio lasketaan muuttujia vastaavien vektoreiden pituuden ja vektoreiden välisen kulman avulla (kuva 40).

| Muuttujat vektoriesityksenä | Muuttujien välinen korrelaatio |
|---|---|
|  | $r_{xy} = \ x_1\ \cdot \ x_2\ \cdot \cos(\alpha)$ <p>muuttujien välinen korrelaatio = vektorin x1 pituus • vektorin x2 pituus • vektoreiden välisen kulman kosini</p> |

Kuva 40 Muuttujien korrelaatio vektorilaskentana

Kosiniratkaisu oli mahdollista suorittaa ilman tietokonetta tehtävillä laskutoimituksilla. Tietokoneiden yleistyttyä tuli suosituimmaksi rotaatiomenetelmäksi varimax-rotatio. Ensimmäisen kerran se esiintyy väitöskirjoissa vuonna 1966. Muutamissa 1960- ja 1970-luvun taitteessa tarkastetuissa väitöskirjoissa varimax-rotatiota on käytetty kosiniratkaisun (Rantalaiho 1968; Kalimo 1969; Talikka 1969; Olkinuora 1974) ja combined-rotatiion (Stolte-Heiskanen 1967; Honkala 1971) rinnalla. Varimax-rotatio on ollut käytetyin rotaatiomenetelmä näihin päiviin asti. Varimax on ollut Leskisen ja Kuusisen (1991) mukaan suosituin rotaatiomenetelmä myös kasvatustieteissä. Nummenmaa ym. (1997, 247) kuvaavat varimax-rotatiion käyttöä jopa pakonomaiseksi. Vinoa kosini-rotatiota käytettiin 1960-luvulla. Viinot rotaatiot ovat alkaneet yleistyä uudestaan 2000-luvulla (taulukko 13).

Taulukko 13 Suorien ja vinojen rotaatioiden esiintyminen sosiologian väitöskirjoissa

| | 1960-1965 | | 1966-1985 | | 1986-1999 | | 2000-2018 | | Yhteensä | |
|-----------|-----------|-----|-----------|-----|-----------|-----|-----------|-----|----------|-----|
| | lkm | % | lkm | % | lkm | % | lkm | % | lkm | % |
| suora | 0 | 0 | 17 | 71 | 5 | 71 | 8 | 40 | 30 | 54 |
| vino | 4 | 80 | 0 | 0 | 0 | 0 | 8 | 40 | 12 | 21 |
| molempia | 1 | 20 | 5 | 21 | 0 | 0 | 1 | 5 | 7 | 13 |
| ei tietoa | 0 | 0 | 2 | 8 | 2 | 29 | 3 | 15 | 7 | 13 |
| Yhteensä | 5 | 100 | 24 | 100 | 7 | 100 | 20 | 100 | 56 | 100 |

Liitetaulukossa 2 voi jälleen tutustua suorien ja vinojen rotaatiomenetelmien valintaan tietokoneen käytön ja tilastollisten ohjelmien mukaan. Suora rotaatiomenetelmä oli suosittu keskustietokoneiden aikaan, mutta myös tilastollisten ohjelmien, erityisesti GLIM-, SAS- ja R-ohjelmien yhteydessä se on ollut suosittu. SPSS-ohjelmaa käyttäneet ovat valinneet vinon rotaatiomenetelmän viidessä väitöskirjassa ja yhdessä väitöskirjassa on käytetty sekä suoraa että vinoja menetelmiä.

Varimax-rotatation suosio perustuu paljolti suorakulmaisuudesta ja sen myötä faktoreiden riippumattomuudesta johtuvaan selkeyteen. Esimerkiksi Matti Heikkilä (1990) perustelee väitöskirjassaan varimaxin valintaa faktoreiden korreloimattomuudella, joka helpottaa faktoreiden tulkintaa. Tutkimuksen kohteena olevan ilmiön kannalta voi kuitenkin vino rotaatio olla todenmukaisempi, kuten esimerkiksi Marjatta Marin tuo vuonna 1970 tarkastetussa väitöskirjassaan ilmi:

Ilmeisesti jokin vinorotaatio olisi tällaisessa ongelmanasettelussa ollut järkevämpi. Toisaalta juuri suorakulmaisen rotatation avulla voidaan pyrkiä samaan selvälle toisistaan riippumattomia ulottuvuuksia, jolloin selvästi eri faktoreille sijoittuneiden muuttujien väliset korrelaatiot eivät ole kovin voimakkaita. (Marin 1970.)

Keskustelua suoran ja vinon rotatation sopivuudesta ja paremmuudesta on käyty pitkään, kuten käy ilmi Touko Markkasen vuonna 1964 kirjoitetusta tekstistä:

Paljon kiistaa on käyty siitä, kumpi, suora- vai vinokulmainen, rotatioratkaisu on parempi. (...) Suorakulmainen rotatio on teknillisesti helpompi ja vinokulmainen rotatio teknillisesti vaikeampi toteuttaa. Tämä vaikeusero oli erittäin selvä kaikkina niinä vuosina, jolloin laskentatyöt jouduttiin suorittamaan joko kynä-paperi-menetelmällä tai mekaanisilla pöytälaskukoneilla. (Markkanen 1964, 10.)

Leskinen ja Kuusinen (1991) kyseenalaistavat suorakulmaisen varimax-rotatation sopivuutta kasvatustieteellisten ja yhteiskuntatieteellisten ilmiöiden tarkastelussa:

Kasvatustieteessä ja muissa yhteiskuntatieteissä ilmiöiden todellinen rakenne on varmaankin perin harvoin luonteeltaan sellainen, että niiden kuvaamista ortogonaalisin faktorein voitaisiin sisällöllisin syin perustella. Siksi onkin huolestuttavaa, että 97 % väitöskirjatutkimuksen faktorianalyysejä käyttäneistä tutkijoista kuvasi tutkimuksen kohdetta varimax-rotatation avulla. (Leskinen ja Kuusinen 1991.)

Hair ym. (1998, 110–111) tuovat esille, että ei ole olemassa erityisiä sääntöjä sen suhteen, pitäisikö tutkijan valita jokin tietty suora tai vino rotaatiomenetelmä:

Useimmiten tutkija vain käyttää tietokoneohjelman tarjoamia rotatointimahdollisuuksia. Useimmissa ohjelmissa varimax on oletuksena, mutta saatavilla on monia muitakin. Ei kuitenkaan ole mitään pakottavaa analyttistä syytä suosia tiettyä rotaatiomenetelmää. Valinta suoran ja vinon rotaation välillä pitäisi tehdä tutkimusongelman perusteella. Jos tutkimuksen tavoitteena on vähentää muuttujien määrää, riippumatta siitä miten merkityksellisiä tuloksena olevat faktorit ovat, sopiva valinta olisi suora rotaatio. Myös, jos tutkija haluaa tiivistää suuren muuttujajoukon pienempää määrään korreloimattomia muuttujia jatkoanalysoitavaksi regressioanalyysissä tai muissa ennustemalleissa, suora rotaatio on paras. Mutta jos tavoitteena on saavuttaa muutamia teoreettisesti merkityksellisiä faktoreita tai käsitteitä, vino rotaatio on sopiva, koska todellisuudessa faktorit harvoin ovat korreloimattomia. (Hair ym. 1998, 110–111.)

Harmanin (1976, 8–9) mukaan tilastotieteellisen teorian eräs tavoite on tarjota tieteellinen laki tai matemaattinen malli selittämään aineiston taustalla olevaa rakennetta. Sellaiset lait antavat hänen mukaansa sallivuuden havaitun aineiston satunnaiselle vaihtelulle teoreettisesti odotetusta arvosta ja on ymmärrettävää, että mikä tahansa monesta melko erilaisesta mallista voi tarjota yhtä hyvän sopivuuden tai selityksen aineistolle tai sen osalle. Ei siis ole olemassa yhtä ainoaa oikeaa mallia tai ratkaisua, toteaa Harman (1976, 8–9).

Myös Leskinen (1987) toteaa, että: ”Yleinen faktorimalliesitys ei ole (latausmatriisin ja kovarianssimatriisin) parametrisoinnin suhteen yksikäsitteinen. Jos faktorimalliesitykselle löydetään joku ratkaisu parametrien Λ ja Ω suhteen, on tämä ratkaisu vain yksi ratkaisu useiden muiden ratkaisujen joukossa, jotka tuottavat saman kovarianssimatriisiesityksen.” Hän jatkaa, että sopiva tulkinnallinen ratkaisu haetaan rotaatiomenetelmällä. Nummenmaa ym. (1997, 245–248) toteavat, että tilastollisen yhteensopivuuden mielessä kaikki rotaatiomenetelmät ovat samantarvoisia, ne tuottavat samat otoskorrelaatiomatriisin sovitteet. Koska rotaatioiden avulla faktoriratkaisu pyritään muuntamaan sisällöllisesti tulkinnalliseen muotoon, perustuu niiden keskinäinen vertailu myös sisällöllisiin tulkintoihin. (Nummenmaa ym. 1997, 245–248.)

Rotaatiomenetelmät auttavat siis tutkijaa valitsemaan faktoriratkaisuista tulkinnallisesti käyttökelpoisimman. ”Tähän tilanteeseen liittyy myös eksploraatiivisten faktorimallien käytön pahin ongelma: koska ei ole olemassa tilastomatemattista objektiivista kriteeriä valita ’oikea’ latausratkaisu, jää valinta tutkijan subjektiivisen sisällöllisen tulkinnan varaan”, toteaa Leskinen (1987, 42–46). Matemaattisuuden ja tulkinnallisuuden ero tulee esille jo Vahervuon ja Ahmavaaran (1958, 25–26)

teoksessa heidän todetessaan, että voidaan löytää lukematon määrä faktorimatriiseja, jotka matemaattisesti katsoen ovat yhtä hyviä ja yhtä oikeutettuja, mutta joiden tulkinta muodostuu erilaiseksi. Onkin mielenkiintoista, jos tutkijat perustelevat rotaatiomenetelmävälintaansa matemaattisin perustein, kun rotaation tarkoituksena on auttaa tutkijaa valitsemaan paras sisällöllinen ja tulkinnallinen ratkaisu. Toki monet tutkijat kokeilevat useita faktorointi- ja rotatointimenetelmiä ja valitsevat niistä parhaaksi arvioimansa, kuten esimerkiksi Esko Kalimo (1969) ja Heikki Leimu (1983) tuovat sosiologian väitöskirjoissaan esille:

Rotaatoratkaisujen välillä ei ollut faktoritulkinnoissa eroja, mutta vinorotaatiotulokset valittiin julkaistaviksi, koska rotatointujen faktorimatriisien rakenteen yksinkertaisuus oli niissä suurempi. (Kalimo, 1969.)

Muuttujavalinnoissa käytettiin hyväksi useille työtyytyvyyteen liittyville muuttujille (28 kpl) tehtyä faktorianalyysia, johon sovellettiin useita eri faktorimääriin perustuvia rotatointeja ja jossa pyrittiin testaamaan Herzbergin työtyytyvyyden perusulottuvuuksien olemassaoloa. Muuttujat on pääsääntöisesti valittu siten, että (2 faktorin ratkaisun) kunkin tulkittavissa olevan faktorin edustajaksi on valittu mahdollisimman konkreettisesti työmotivaatiota ja työolosuhteita kuvaava muuttuja sen mukaan, kumpaan näistä kahdesta ulottuvuudesta faktori kuului. (Leimu 1983.)

Kuten Kalimokin tuo esiin, voi rotaatiomenetelmän valinta olla sosiologisessa tutkimuksessa enemmän periaatteellinen kuin käytännöllinen, koska eri rotaatiomenetelmillä saadaan lähes samanlaiset tulokset. Myös monimuuttujamenetelmiä esittelevässä oppikirjassa Hair ym. (1998, 103) toteavat faktorianalyysin olevan kohdallaisen luotettava menetelmä, koska se antaa suurin piirtein samanlaisia tuloksia riippumatta siitä, mitä faktorointi- ja rotaatiomenetelmiä valitsee.

Faktorianalyysin jatkotoimenpiteet

Faktorianalyysin tarjoama tieto faktoreiden lukumäärästä ja koostumuksesta voi olla tutkijalle riittävä tieto ja tulos sellaisenaan eikä tutkija koe tarvetta jatkaa faktoreiden analyysia muulla tavoin. Sosiologian väitöskirjoissa 56:sta faktorianalyysia sisältävästä väitöskirjasta 14:ssä faktoreita ei ole lähdetty jatkoanalysoimaan (taulukko 14). Nämä väitöstutkimukset on tehty pääasiassa 1960- ja osittain vielä 1970-luvulla. Erityisesti 1960-luvun alkupuolella, kun faktorianalyysit jouduttiin suorittamaan ilman tietokoneita, oli analyysin suorittaminen itsessään jo niin suuritöistä lukuisine laskutoimituksineen, että on oletettavasti koettu riittäväksi saada tieto siitä, kuinka monta faktoria ja minkä sisältöisiä faktoreita analyysi tuotti. Taulukossa 14 näkyy myös, kuinka monessa väitöskirjassa faktoreita on lähdetty jatkoanalysoimaan muodostamalla niistä joko faktoripistemäärämuuttujia tai

summamuuttujia¹⁹. Faktoripistemäärämuuttujat ovat olleet summamuuttujia suosittumia ainoastaan vuosina 1966–1985. Faktoripistemäärämuuttujien tuolloinen suosio selittyy ainakin osaltaan sillä, että analyysit suoritettiin keskustietokoneilla eivätkä tutkijat voineet juurikaan vaikuttaa analyysin vaiheisiin. Kun tutkijat alkoivat saada käyttöönsä henkilökohtaisia tietokoneita 1980-luvun puolivälissä, alkoi summamuuttujien muodostaminen olla suosituin tapa jatkaa faktoreiden analyysia. Liitetaulukossa 3 voi tutustua yksityiskohtaisemmin väitöskirjoissa eri aikoina esitettyihin faktorianalyysin jatkotoimenpiteisiin jaoteltuna tietokoneiden käytön ja tilastollisten ohjelmien mukaan.

Taulukko 14 Faktoripistemäärä- ja summamuuttujien esiintyminen faktorianalyysin jatkotoimenpiteenä

| | 1960-1965 | 1966-1985 | 1986-1999 | 2000-2018 | Yhteensä |
|--------------------|-----------|-----------|-----------|-----------|----------|
| faktoripistemäärät | 0 | 11 | 1 | 6 | 18 |
| summamuuttujat | 0 | 5 | 6 | 13 | 24 |
| ei jatkoa | 5 | 8 | 0 | 1 | 14 |
| Yhteensä | 5 | 24 | 7 | 20 | 56 |

Jokivuori ja Hietala (2007, 100–105) kirjoittavat pääkomponentti- tai faktoripistemäärien avulla muodostuneen summamuuttujan²⁰ ja tutkijan itse rakentaman summamuuttujan ominaisuuksista. Muutamia heidän esille nostamistaan ominaisuuksista on koottu taulukkoon 15. He toteavat, että faktoripistemäärämuuttujilla on tilastomatematiikan näkökulmasta hyviä ominaisuuksia ja ne antavat tutkijalle tietoa aineistosta. Tutkimuskäytössä niihin sisältyy kuitenkin heikkouksia. Faktoripistemäärien standardointi tekee niistä vaikeasti tulkittavia ja lisäksi faktoripistemäärämuuttujiin sisältyy virhetulkintojen vaara tilanteissa, joissa vain vahvimmin latautuneet muuttujat sisällytetään tulkintaan, mutta mukana ovat kuitenkin kaikki muuttujat eri painoillaan. Jos taas tutkija itse valitsee harkintaansa käyttäen voimakkaimmin latautuneet kärkimuuttujat ja rakentaa niistä muuttujien keskiarvoa kuvastavan summamuuttujan, pystyy hän tulkitsemaan summamuuttujan arvoja ja kertomaan tutkimastaan ilmiöstä. He korostavat käsitevaliditeetin tärkeyttä faktoripistemäärämuuttujan ja summamuuttujan valinnassa, koska teoreettisen käsitteen mittaamisen validiteetti on tutkimuksen sisäisen luotettavuuden peruskallio. He kehottavat tutkijoita puntaroimaan, onko analyysin tavoitteena saavuttaa tilastomatematisesti paras mahdollinen vai käsitteellisesti validi muuttujakombinaatio. (Jokivuori ja Hietala 2007, 100–105.)

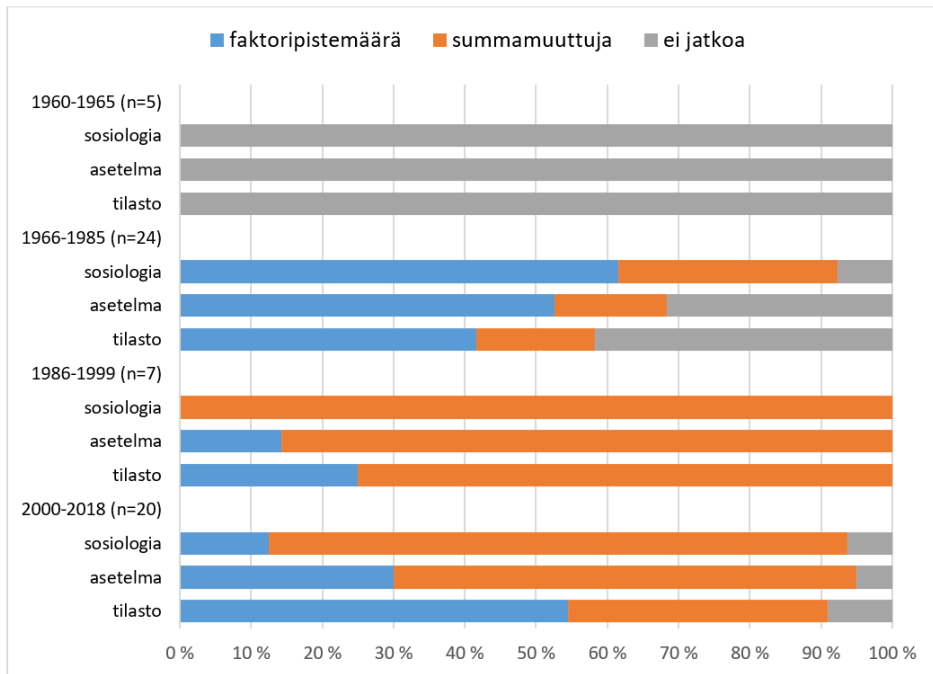
19 Muutamasosiologion tosin käyttänyt 1960-luvulla faktoreiden jatkoanalyysiin transformaatioanalyysia.

20 Tutkimuksessani kutsun pääkomponentti- tai faktoripistemäärien avulla muodostunutta summamuuttujaa lyhyesti faktoripistemäärämuuttujaksi. Samaa nimitystä käytän myös taulukossa 15.

Taulukko 15 Faktoripistemäärämuuttujan ja summamuuttujan ominaisuudet Jokivuoren ja Hietalan (2007) mukaan

| Faktoripistemäärämuuttuja | Summamuuttuja |
|---|--|
| <ul style="list-style-type: none"> • sisältää kaikki faktorianalyysin muuttujat erilaisilla painoilla • sisältö amebamainen | <ul style="list-style-type: none"> • tutkija valitsee voimakkaimmin latautuneet kärkimuuttuja harkintaansa käyttäen • tutkija tietää sisällön |
| <ul style="list-style-type: none"> • standardoitu muuttuja, jonka keskiarvo on 0 • negatiiviset pistemäärät ilmaisevat keskimääräistä pienempiä arvoja ja positiiviset pistemäärät keskimääräistä suurempia arvoja • yleistason arvioiminen mahdotonta • kertovat aineistosta | <ul style="list-style-type: none"> • erilaisia rakentamismahdollisuuksia • keskiarvoon pohjautuvassa summamuuttujassa sama asteikko alkuperäisten muuttujien kanssa • kuvaavat tutkittavan ilmiön yleisiä piirteitä |

Kuvassa 41 esitetään faktorianalyysin jatkotoimenpiteiden prosentuaaliset jakaumat sekä ajallisen että rajakohdemallin kauttakulkupisteiden mukaan.



Kuva 41 Faktorianalyysin jatkotoimenpiteet eri aikoina rajakohdemallin kauttakulkupisteiden mukaan

Keskustietokoneiden aikaan vuosina 1966–1985 on ollut tavanomaista jatkaa analyysia faktoripistemäärien avulla. Henkilökohtaisten tietokoneiden aikakau-

della 1986–1999 faktorianalyysia on käytetty sosiologian väitöskirjoissa vähän (vain seitsemässä), mutta niissä jatkoanalyysit on tehty pääosin summamuuttujia hyödyntämällä. 2000-luvulla on jatkoanalyysit tehty sekä faktoripistemääriä että summamuuttujia hyödyntäen. On mielenkiintoista huomata lievä yhteys kauttakulukupisteiden ja jatkoanalyysitavan välillä. Erityisesti 2000-luvun osalta näkyy se, että väitöskirjoissa, joissa faktorianalyysia on perusteltu sosiologiseen tutkimuskysymykseen viitaten, on käytetty summamuuttujia. Ja väitöskirjoissa, joissa faktorianalyysia on perusteltu vedoten tilastomatematiikkaan syihin, on jatkoanalyysi suoritettu enimmäkseen faktoripistemäärien avulla.

8.3. Faktorianalyysin hyöty ja haasteet

Faktorianalyysi on osoittautunut käyttötavoiltaan monipuoliseksi menetelmäksi. Taulukkoon 16 on koottu yhteen faktorianalyysin suorittamisessa tehdyt valinnat eri aikoina. Faktorointi on tapahtunut yleensä pääakseli- ja pääkomponenttimenetelmien välillä. Joskus valinta on tehty tietoisesti ja perustellen, mutta usein valintaa ei ole tietoisesti tehty vaan faktorointi on tapahtunut tilastollisen ohjelman oletusarvojen mukaisesti. Hyvin usein faktorointimenetelmä on jätetty myös kertomatta.

Taulukko 16 Faktorointi- ja rotaatiomenetelmien sekä faktorianalyysin jatkotoimenpiteiden esiintymismäärät eri aikoina

| | 1960- 1965 | 1966- 1985 | 1986- 1999 | 2000- 2018 | Yhteensä |
|--|---------------|---------------|---------------|---------------|----------|
| Faktorointi | | | | | |
| sentroidi | 2 | 0 | 0 | 0 | 2 |
| pääakseli | 2 | 11 | 1 | 5 | 19 |
| pääakseli ja pääkomponentti | 0 | 1 | 0 | 1 | 2 |
| pääkomponentti | 0 | 5 | 4 | 6 | 15 |
| pns | 0 | 0 | 0 | 2 | 2 |
| ei tietoa | 1 | 7 | 4 | 6 | 16 |
| Rotatointi | | | | | |
| suora | 0 | 17 | 5 | 7 | 29 |
| vino | 4 | 0 | 0 | 8 | 12 |
| molempia | 1 | 5 | 0 | 1 | 7 |
| ei tietoa | 0 | 2 | 2 | 4 | 8 |
| Faktorianalyysin jatko | | | | | |
| ei jatkoa | 3 | 6 | 0 | 0 | 9 |
| transformaatioanalyysi | 2 | 2 | 0 | 0 | 4 |
| faktoripistemäärät | 0 | 11 | 1 | 6 | 18 |
| summamuuttujat | 0 | 5 | 6 | 11 | 22 |
| pistemäärä- ja summamuuttuja | 0 | 0 | 0 | 2 | 2 |
| ei tietoa | 0 | 0 | 0 | 1 | 1 |
| Faktorianalyysiväitöskirjoja | 5 | 24 | 7 | 20 | 56 |
| Kaikki sosiologian väitöskirjat | 11 | 51 | 130 | 340 | 532 |

Rotatointimenetelmissä on ollut hallitsevana suoriin menetelmiin lukeutuva varimax-rotatatio. Vinot rotaatiomenetelmät ovat nousseet suosituiksi 2000-luvulla. Faktoreiden jatkoanalyysi on tapahtunut useimmiten summamuuttujien tai faktoripistemäärämuuttujien avulla. Faktoripistemäärämuuttujien käyttö on ollut yleisintä keskustietokoneiden aikaan vuosina 1966–1985. Summamuuttujat taas ovat olleet suosittuja 2000-luvun väitöskirjoissa.

Edellä olevista tuloksista ei voida tietää sitä, kuinka monia valintoja ja kokeiluja sosiologit ovat tehneet muuttujien, faktorointimenetelmien ja rotaatiomenetelmien suhteen. Tutkijat ovat voineet etsiä parasta mahdollista lopputulosta, faktorimallia, useilla kokeiluilla. Useimmat ovat luultavasti kokeilleet muuttujien poistamista ja lisäämistä, erilaisia faktorointi- ja rotaatiomenetelmiä ja myös eri faktorimääriä. Hair ym. (1998, 103) vertaavat faktoreiden lukumäärän valitsemista mikroskoopilla tarkasteluun, liian suuri tai liian pieni tarkkuus eli liian paljon tai liian vähän faktoreita hämärtää kokonaiskuvan. Myös he suosittelevat faktorianalyysin suorittamista useita kertoja parhaaseen ratkaisuun pääsemiseksi.

Tarkastelen taulukossa 17, kuinka suuressa roolissa faktorianalyysi on väitöskirjoissa eri aikoina ollut. Olen tulkinnut faktorianalyysin olevan pienessä roolissa 17 väitöskirjassa, keskimääräisessä roolissa 26:ssä ja suuressa roolissa 12 väitöskirjassa. Roolilla viitataan sekä väitöskirjassa esitettyjen faktorianalyysien lukumäärään, että myös faktorianalyysin näkyvyyteen esimerkiksi sivumäärissä.

Taulukko 17 Faktorianalyysin rooli väitöskirjoissa eri aikoina

| Faktorianalyysin rooli | 1960–1965 | 1966–1985 | 1986–1999 | 2000–2010 | 2011–2018 | Yhteensä |
|------------------------|-----------|-----------|-----------|-----------|-----------|----------|
| pieni | 0 | 6 | 3 | 6 | 2 | 17 |
| keskimääräinen | 3 | 8 | 4 | 6 | 5 | 26 |
| suuri | 2 | 9 | 0 | 0 | 1 | 12 |
| Yhteensä | 5 | 23 | 7 | 12 | 8 | 55 |

Faktorianalyysin rooli on ollut suurimmillaan 1960-luvulla. Muutamassa tuolloin tarkastetussa väitöskirjassa koko tutkimus on jopa rakentunut faktorianalyysin ympärille. Esimerkkinä tällaisesta mainittakoon Annikki Talikan vuonna 1969 tarkastettu väitöskirja. 1960-luvun jälkeen faktorianalyysin rooli ja näkyvyys on pienentynyt. Väitöskirjojen tekstien perusteella pienin rooli faktorianalyysilla on ollut Irja Kandolinin (1997) väitöskirjassa, jossa faktorianalyysi mainitaan yhdellä sanalla. Olen jakanut 2000-luvun tarkastelun taulukossa 19 poikkeuksellisesti kahteen ajanjaksoon, koska näyttää siltä, että faktorianalyysi on alkanut saada vuosina 2011–2018 tarkastetuissa väitöskirjoissa aikaisempaa suuremman roolin, vaikkakin faktorianalyysia sisältävien väitöskirjojen määrä on pieni.

Risto Sänkiaho (1974, 12) näkee faktorianalyysin enemmän aineistoa siistivänä menetelmänä kuin varsinaisena asioiden tarkkaa tutkimista ja yhteyksien selvittämistä koskevana analyysinä. Erityisesti 2000-luvun väitöskirjojen menetelmäkuvauksissa on havaittavissa faktorianalyysin hyödyntäminen ennemminkin aineistoa siistivänä menetelmänä tai alustavana analyysinä ennen varsinaisia analyyseja. Esimerkiksi Kankainen (2007, 67) kirjoittaa käyttäneensä faktorianalyysia summamuuttujaa edeltävänä menetelmänä –tarvittaessa. Faktorianalyysin sijoittaminen alustavaksi analyysiksi vahvistuu, kun Kankainen kertoo käyttäneensä väitöstutkimuksensa monimuuttujamenetelminä varianssianalyysia ja regressioanalyysia eikä siis lue faktorianalyysia kuuluvaksi tähän joukkoon. Myös Jokivuori (2002, 50) kirjoittaa käyttäneensä faktorianalyysia aineiston informaation tiivistämiseksi ja käyttämiseen monimuuttujamenetelmiksi hän lukee regressioanalyysin, monisuuntaisen varianssianalyysin ja erotteluanalyysin. Sanna Sanaksenahon (2006, 13) väitöskirjasta on myös luettavissa faktorianalyysin sijoittaminen alustavaksi menetelmäksi.

Faktorianalyysi on ollut suosittu tilastollinen menetelmä suomalaisten sosiologien keskuudessa, mutta se on herättänyt myös kritiikkiä. Hannu Takala (1973) otsikoi Sosiologia-lehdessä tekstinsä kriittiseen sävyyn: ”Miten faktoripistemäärillä arvioidaan sopeutumattoman nuoren huollontarve, kun isä juo ja äiti huoraa”. Sänkiaho (1974, 12) toteaa seuraavassa faktorianalyysin käytön olleen täysin kritiikitöntä:

Nuorille tutkijoille tuli faktorianalyysista jonkinlainen oikotie onneen, sillä jos ei tutkimuksesta päässyt muuten eteenpäin, niin voihan aina suorittaa faktorianalyysin. Ja kyllähän niitä tehtiinkin. Tämän vuoksi faktorianalyysin käyttö tuli täysin kritiikittömäksi ja sitä useimmissa tapauksissa käytettiin sellaistenkin tutkimusongelmien selvittelyssä, joihin se ei luonteeltaan soveltunut. (Sänkiaho 1974, 12.)

Faktorianalyysin lisäksi kritiikki kohdistui myös yleisesti tietokoneiden käyttöä kohtaan. Sänkiaho (1974) syytti tietokoneita siitä, että ne aikaansaivat 1950–1960-luvulla monimuuttujamenetelmien räjähdysmäisesti lisääntyneen käytön ja sen myötä sosiaalitieteellisen tutkimuksen vinoutumisen. Myös Rantalaiho (1972) esitti Sosiologia-lehdessä kritiikkiä tietokoneiden käyttämisestä kohtaan toteamalla, että ”automaation pitäisi tuki vapauttaa ihminen rutiineista myös tutkimuksen ja ajatustyön alueella, mutta nykyisin sitä käytetään etupäässä ajattelusta vapautumiseen”.

Faktorianalyysiin kohdistunut kritiikki on huomioitu muutamissa 1960-1970-luvun väitöskirjoissa. Esimerkiksi vuonna 1969 tarkastetussa väitöskirjassaan Kalimo kirjoittaa seuraavaa:

Toisaalta on korostettu, että faktorianalyysin suorittaminen ilman ohjaavaa teoriaa saattaa tuottaa harhaanjohtavia tuloksia. Tämä vaara on nykyään

huomattavan suuri, kun hieman kärjistään voidaan sanoa, että faktorianalyysin laskenta tietokoneella tulee halvemmaksi kuin huolellinen harkinta sen käytön sopivuudesta ko. tilanteessa. (Kalimo 1969, 173.)

Toivosen (1999, 334) mukaan faktorianalyysin käytön vähenemisen perusteena oli sen huomaaminen, että näennäisestä helppoudestaan huolimatta faktorianalyysi vaatiikin paljon enemmän teoreettista harkintaa kuin miltä aluksi näytti.

Faktorianalyysia kehittänyt Thurstone ennakoivat tulevan kritiikin jo 1900-luvun alkupuolella toteamalla, että todennäköisesti matemaattisten tilastotieteilijöiden ja tutkijoiden kohtaaminen tulisi olemaan haasteellista. Ensimmäisenä haasteena Thurstone mainitsi tasapainottelun tutkijoiden ja tilastotieteilijöiden näkökulmien välillä tutkijoiden ollessa kiinnostuneita teorioista ja tilastotieteilijöiden puolestaan otosjakaumista. Toisena haasteena Thurstone näki tutkimusnäkökulman rajallisuuden. Tällä hän tarkoitti, että tilastotieteilijöiden olisi hyväksyttävä tiettyjä epätarkkuuksia faktorianalyysin matematiikassa, koska tieteelliset formuloinnit eivät pysty kattamaan minkään ilmiön kokonaisvaihtelua. Thurstone mainitsi myös lineaarisuusoletuksen ongelmallisuuden: tilastomatematiikka perustuu lineaarisuuteen, kun taas tutkittavat ilmiöt harvoin ovat luonteeltaan lineaarisia tai lineaarisesti mitattavia. Hän näki tilastotieteilijöiden puutteena keskeisesti myös heidän suhtautumisensa faktoreiden rotaatioon, jonka tarkoituksena on jalostaa faktorianalyysin tulos sellaiseksi, että tutkija osaa valita hedelmällisimmän parametrijoukon kuvaamaan tutkittavan ilmiön vaihtelua. Thurstone kirjoitti, että tutkijoille parametrit edustavat tieteellisiä käsitteitä, ne eivät ole vain numeerisia kertoimia. Jos nämä tieteellisesti mielekkäät parametrit sattuvat korreloimaan keskenään koetoksessa, se ei ole peruste hylätä niitä. Tilastotieteilijän on ymmärrettävä tämä vinojen akselien ja niihin liittyvien tieteellisten käsitteiden ongelma. Tilastotieteilijät eivät saa määrätä mielivaltaista sellaista rajoitusta, että ihmisten kuvaamisessa käyttämiemme tieteellisten käsitteiden olisi oltava korreloimattomia otoksessa tai populaatiossa. (Thurstone 1947, v-xiv.) Thurstonen esille tuomat haasteet ovat yllättävän ajankohdaisia edelleen ja ne pätevät yleisesti kaikkien tilastollisten menetelmien, eivät vain faktorianalyysin, käyttöön.

Vaikka faktorianalyysia ja faktorianalyysin käyttäjiä on kritisoitu menetelmän liiallisesta, mekaanisesta ja harkitsemattomasta käytöstä ja epäilty, että analyysilla pystytään saamaan sellaisia tuloksia kuin halutaan, suurin osa sosiologian faktorianalyysia sisältävistä väitöstutkimuksista on sellaisia, että faktorianalyysi on ajanut asiansa ja tuottanut edes jonkinlaista tutkimusta palvelevaa tietoa. Väitöskirjat, joissa faktorianalyysin käyttö ansaitsee kritiikkiä, ovat poikkeustapauksia. Tosin niissäkään ei ole syytä osoittaa sormella menetelmää vaan tutkijaa. Tutkija on joko analyysivaiheessa tehnyt harkitsemattomia muuttujavalintoja tai on intoutunut tulkitsemaan tuloksia liian mielikuvituksellisesti.

9. Monimuuttujamenetelmien monet kasvot

9.1. Rajakohdemallin laajentaminen monimuuttujamenetelmiin

Olen tutkimuksessani sivunnut tieteiden moninaisuutta, sosiologian moninaisuutta ja tilastotieteen moninaisuutta. Sosiologit ovat valinneet faktorianalyysin monin eri perustein, he ovat käyttäneet sitä monin eri tavoin. Menetelmä on kehittynyt monen eri tieteen panoksena ja se voidaan nähdä monin matemaattisin ja tutkimusmenetelmällisin tavoin. Kaiken moninaisuuden haltuunotto yhdessä tutkimuksessa on mahdotonta enkä ole sellaista tavoitellutkaan.

Tutkimusta voisi jatkaa laajentamalla tässä tutkimuksessa esitettyä rajakohdemallia (kuva 43) koskemaan kaikkien monimuuttujamenetelmien valintaa ja käyttötapaa sosiologisessa tutkimuksessa. Kauttakulkupisteinä toimisivat kriteerit, joilla menetelmävalinta perustellaan ja joiden mukaan menetelmää käytetään. Yksi kauttakulkupiste tarjoaisi tilastomatematiikkaan ja muuttujiin matemaattisiin ominaisuuksiin pohjautuvia perusteluja. Toinen kauttakulkupiste perustuisi monimuuttujamenetelmien sisältämiin asetelmiin. Kolmas kauttakulkupiste puolestaan korostaisi sosiologisen tutkimuskysymyksen keskeisyyttä menetelmävalinnassa ja tavassa, miten menetelmää käytetään. Rajakohdemallin verkostoina toimisivat monimuuttujamenetelmiä käyttävät tilastotieteilijä(t), tutkija(t) ja sosiologi(t). Rajakohdemallin kääntämiset puolestaan olisivat kamppailuja tieteellisestä auktoriteetista eli siitä, mikä kauttakulkupiste tarjoaa vakuuttavimmat perustelut monimuuttujamenetelmän valintaan ja käyttöön.

Kauttakulkupisteet – monimuuttujamenetelmät tilastomatematiikkana, ongelmanasetteluina ja tutkimusmenetelminä

Kuvassa 42 havainnollistetaan faktorianalyysin lisäksi muiden monimuuttujamenetelmien eli regressioanalyysin, varianssianalyysin ja ryhmittelyanalyysin näyttäytyminen kolmessa kauttakulkupisteessä.

Tilastomatematiikka ja muuttujien ominaisuudet

$$y_1 = \lambda_1\eta + \varepsilon_1$$

$$y_2 = \lambda_2\eta + \varepsilon_2$$

$$y_3 = \lambda_3\eta + \varepsilon_3$$

$$y_4 = \lambda_4\eta + \varepsilon_4$$

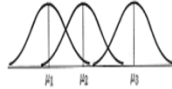
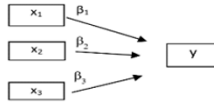
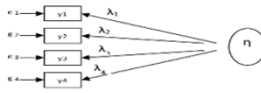
$$y = \beta_0 + \beta_1x_1 + \beta_3x_3$$

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$$

$$ryhmäkeskus_i = \frac{1}{n_i} \sum_{x \in R_i} x$$

Monimuuttujamenetelmän asetelma



Sosiologisen tutkimuskysymyksen elementtejä

Käsitteiden analysointi.

"Haluan tutkimuksessani löytää käsitteen eri ulottuvuuksia"

Usean tekijän yhteys tarkastelun kohteena olevaa tekijään.

"Haluan tutkimuksessani löytää käsitteeseen yhteydessä olevat tekijät"

Vastaajaryhmien keskiarvojen vertailu

"Onko ryhmillä eroa?"

Havaintojen ryhmittäminen

"Haluan löytää erilaisia vastaajaryhmiä"

Kuva 42 Monimuuttujamenetelmien näyttäytyminen rajakohdemallin kauttakulkupisteinä

Useimmissa monimuuttujamenetelmiä esittelevissä teoksissa menetelmien esittely ja niiden jaottelu pohjautuu muuttujiin ja niiden ominaisuuksiin (Tacq 1997, 1). Tilastomatematiikkaan perustuvassa kauttakulkupisteessä menetelmät näyttäytyvät muuttujina, matemaattisina yhtälöinä ja laskutoimituksina. Faktoriansalyysi esitetään yhtälöryhmänä, jossa havaitut y_i -muuttujat pyritään kuvaamaan faktorin η , faktorilatausten λ_i ja virhetermien ε_i avulla. Regressioanalyysiin sisältyy yhtälö, jossa x_i -muuttujien vaikutusta y -muuttujan vaihteluun tarkastellaan regressiokerrointen β_i avulla. Varianssianalysissa vertaillaan yhden muuttujan keskiarvoja eri havaintoryhmissä. Vertailu perustuu yhtälöihin, joissa lasketaan erilaisia poikkeamia kokonaiskeskiarvosta ja ryhmäkeskiarvosta sekä määritellään erilaisia vaihteluja kuvaavia neliösummia. Ryhmittelyanalyysi perustuu myös keskiarvolaskelmiin, joiden perusteella määrättyvät niin sanotut ryhmäkeskukset. Ryhmittelyanalyysissä vertaillaan havaintojen etäisyyttä ryhmäkeskuksiin pyrkimyksenä sijoittaa kukin havainto lähimpään ryhmäkeskukseen kuuluvaksi.

Asetelmaan perustuvassa kauttakulkupisteessä monimuuttujamenetelmät voidaan jakaa monin tavoin. Luvussa 4.3 menetelmät jaettiin tiivistäviin ja ryhmittäviin menetelmiin, selittämiseen perustuviin menetelmiin, ryhmävertailuja tekeviin menetelmiin sekä monia menetelmiä sisältäviin menetelmiin.

Sosiologisen tutkimuskysymyksen kauttakulkupisteessä monimuuttujamenetelmän on palveltava tutkimusta tukevana analyysimenetelmänä. Kvantitatiivisen aineiston analyysiin etsitään monimuuttujamenetelmiä, jotka asetelmaltaan vastaavat sosiologisen tutkimuskysymyksen käsitteellistä rakennetta ja sopivat yh-

teen tutkimuskysymysten valittujen elementtien kanssa. Faktorianalyysi voi sopia aineiston analyysimenetelmäksi, jos sosiologisessa tutkimuskysymyksessä viitataan teoreettisiin käsitteisiin tai käsitteiden ulottuvuuksiin. Regressioanalyysiin viittavia elementtejä puolestaan ovat maininnat tarkasteltavaan asiaan yhteydessä olevista tekijöistä. Jos sosiologisessa tutkimuskysymyksessä tuodaan esille tarve löytää erilaisia vastaajaryhmiä, saattaisi ryhmittelyanalyysi olla hyödyllinen.

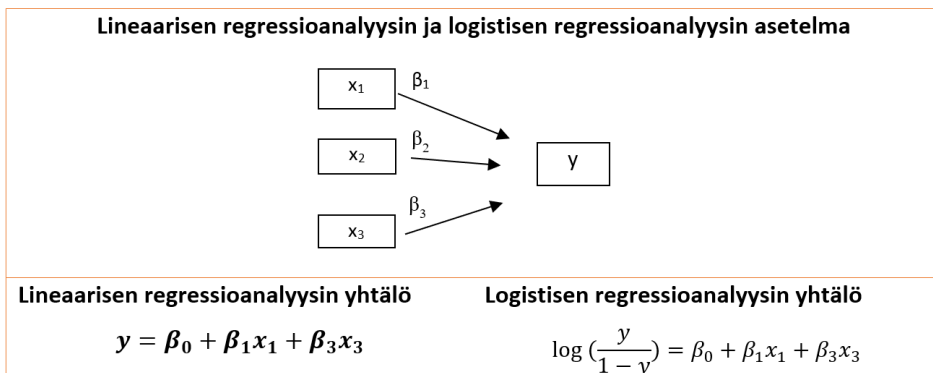
Kauttakulkupisteen valinta tapahtuu periaatteessa siten, että osapuolet puntaaroivat niiden tarjoamia näkemyksiä ja valitsevat niistä vakuuttavimman. Todellisuudessa monimuuttujamenetelmien valinnat ja käyttötavat sisältävät usean kauttakulkupisteen tarjoamia perusteluja. Tilastotieteilijä katsoo menetelmiä muuttujien ja niiden tilastomatemattisten ominaisuuksien näkökulmasta, mutta hän näkee menetelmät myös asetelmallisesta näkökulmasta. Sosiologille keskeisintä on asetelman vastaavuus tutkimusongelman käsitteellisen rakenteen kanssa, mutta hänen on jossakin määrin huomioitava myös menetelmään valittavien muuttujien tilastomatemattiset kriteerit (Tacq 1997; Alkula ym. 1994, 284–285; Byrne 2001; 2004; Ronkainen ym. 2013, 125–126).

Tilastomatemattisten kriteerien huomioiminen on sosiologisessa tutkimuksessa haasteellista. Ensimmäiseksi haasteeksi asettuu normaalijakaumavaatimus, jonka useat monimuuttujamenetelmät asettavat muuttujille (Tacq 1997, 31). Tähän vaatimukseen sosiologiset aineistot harvoin pystyvät vastaamaan, koska kvantitatiivisissa sosiologisissa tutkimusaineistoissa muuttujien jakaumat ovat usein vinoja.

Toinen haaste liittyy muuttujien mitta-asteikoihin. Sosiologit joutuvat kamppailemaan mittaamisen problematiikan kanssa tasapainotellen sen suhteen, kuinka paljon huomioida aineistossa olevien muuttujien matemaattisia ominaisuuksia ja toisaalta, kuinka paljon tulisi kiinnittää huomioita sisältöön, jota ne kuvastavat tutkittavasta ilmiöstä. Jos sosiologit noudattaisivat mitta-asteikkovaatimuksia tiukasti, olisi heidän käytettävissään melko suppea menetelmävalikoima. Analysoidakseen kvantitatiivisia aineistojaan mahdollisimman monipuolisesti sosiologit joutuvat rikkomaan menetelmien tilastotieteellisiä mitta-asteikkosääntöjä. Toisaalta ei voida puhua varsinaisesti sääntöjen rikkomisesta, koska mittaaminen ja mittaustulokset ovat sosiologiassa muutakin kuin numeroilla määriteltyä tietoa. Kuten Ronkainen ym. (2013, 122) toteavat, tutkijalle aineisto on tiedon tuottamisen väline eikä se koskaan edusta puhtaasti todellisuutta. Sosiologi saa numeroiden avulla suuntaa antavan käsityksen ilmiöstä. Sosiologi ei voi kuitenkaan rajautua mittaustulosten käsittelyssä vain tilastomatematiikan sääntöihin vaan hänen on koko ajan pohdittava, millä tavalla ilmiötä heijastavia numeroita voisi analysoida, jotta ilmiön saisi kuvattua parhaalla mahdollisella tavalla.

Vaikka sosiologille mitta-asteikkoajattelu ei ole niin keskeistä, on kuitenkin olemassa tilanteita, joissa muuttujien ominaisuuksien – mitta-asteikkojen ja jakaumien – tunnistaminen auttaa sosiologia valitsemaan parhaan mahdollisen menetelmän aineistonsa analyysiin. Esimerkkinä mainittakoon Köpän (1979, 37–40) sosiolo-

gian väitöskirja, jossa hän esittää tarkan kuvauksen käyttämästään AID-analyysistä eli puuanalyysistä. Kuvaus muistuttaa paljon ryhmittelyanalyysia. Köppä arvioi, että AID-analyysin vaihtoehtona olisi ollut moniulotteinen ristiintaulukointi ja regressioanalyysi. Hän kertoo kuitenkin päätyneensä AID-analyysiin, koska piti sitä taulukointia tehokkaampana ja taloudellisempänä ja alhaisille mittaustasojille sopivampana menetelmänä kuin lineaarinen regressioanalyysi. Toinen esimerkki liittyy tilanteeseen, jossa sosiologi haluaa etsiä aineistostaan selittäviä tekijöitä jollekin asialle. Tällöin asetelmaltaan sopivia menetelmiä ovat lineaarinen regressioanalyysi ja logistinen regressioanalyysi (kuva 43). Lineaarisen regressioanalyysin tilastomatematiikka vaatii muuttujilta vähintään välimatka-asteikollisuutta ja muuttujien normaalijakautuneisuutta, kun taas logistisen regressioanalyysin tilastomatematiikka perustuu todennäköisyyslaskentaan ja sen kaltaisiin matemaattisiin muunnoksiin, että muuttujat voivat olla mitta-asteikoltaan mitä tahansa. Sosiologille menetelmävalinnan matemaattiset kriteerit ja mitta-asteikkovaatimukset eivät ole ensisijaisia, mutta jonkinlainen ymmärrys menetelmiin liittyvästä matemaattisista ja niihin kytköksissä olevista muuttujien mitta-asteikko- ja jakaumavaatimuksista on kuitenkin hänelle hyödyksi (ks. esim. Saari ja Melin 2020, 50).



Kuva 43 Lineaarisen ja logistisen regressioanalyysin asetelmallinen yhtäläisyys ja matemaattinen eroavaisuus

Menetelmiin liittyvät lineaarisuus- ja kausaalisuusoletukset ovat myös sosiologisen tutkimuksen kannalta haasteellisia. Selkälä (2013, 117–119) tuo esille Byrnen (2002; 2004) näkemyksen, jonka mukaan regressioanalyysia ja faktorianaalyyseja ei sopisi käyttää, koska ne kausaalisuuteensa ja lineaarisuuteensa takia eivät vastaa tutkittavan todellisuuden rakennetta. Byrnen mukaan sopivia ei-lineaarista menetelmiä ovat ryhmittelyanalyysi, korrespondenssianalyysi ja neuroverkko-mallit. Selkälä (2013, 118) näkee, että menetelmät on kuitenkin nähtävä vain menetelminä, mutta toki niiden kohdalla on pohdittava kuinka suuressa määrin

ne muovaavat aineistoa sellaiseen suuntaan, että se ei enää vastaa todellisuuden teoreettista rakennetta.

Alkula ym. (1994, 17) toteavat, että erilaisten lähestymistapojen ja näkökulmien vaihdellessa ei mikään menetelmällinen ratkaisu ole ainoa oikea jonkin tutkimusongelman selvittämisessä. Käyttökelpoisia ratkaisuja voi olla heidän mukaansa olla useita ja valinta niiden välillä joskus vaikeakin. Pöntinen (2004, 90–91) toteaa, että olisi hyvä, jos tutkija valitsisi käyttämänsä menetelmät aina niiden käyttökelpoisuuden perusteella. Kuten Alkula ym. (1994, 285) toteavat, ovat kuvailevat menetelmät yksinkertaisia ja havainnollisia. Kuvailevien menetelmien tuloksia on kohtalaisen helppo ymmärtää. Monimuuttujamenetelmät puolestaan pystyvät tiivistämään informaatiota tehokkaammin, mutta monimuuttujamenetelmien haasteena on tiedon abstraktius. Kunkin monimuuttujamenetelmän tulkinta on erikseen opeteltava. (Alkula ym. 1994, 285.) Näen monimuuttujamenetelmien abstraktissa muodossa esitettyjen tulosten tulkinnassa ylitulkittamisen ja virheellisten päätelmien tekemisen riskin. Tässä mielessä monimuuttujamenetelmät voivat olla jopa haitallisia kuvaileviin menetelmiin verrattuna. Pöntinen (2004, 90–91) jatkaa, että väistämättä tutkijan menetelmävalintaan vaikuttavat myös kunkin ajan metodologiset ihanteet. Erola ja Räsänen (2014a, 246–247) näkevät menetelmien merkityksen ja menetelmäosaamisen vaatimusten kasvaneen 2000-luvun sosiologiassa. He toteavat, että aikaisemmin väitöstutkimuksen analyysimenetelmiksi riittivät kuvailevat menetelmät, mutta nykyisin ne eivät heidän mukaansa riitä edes maisteriopintojen läpäisemiseen. Erolan ja Räsänen näkemyksen vastineeksi esitän Ronkaisen ym. (2013, 123) näkemyksen, jonka mukaan tutkimuksessa tieto ei piile 'tuloksissa' sinänsä vaan avainasemassa on tutkija, joka nostaa tiettyjä asioita tutkimustuloksiksi eli tulkitsee ja käsitteellistää aihettaan oman ymmärryksensä varassa.

9.2. Erilaiset näkemykset kiistan aiheena

Pekka Räsänen, Anu-Hanna Anttila ja Harri Melin (2005, 9) toteavat, että erilaiset metodologiset näkemyserot ja kiistat ovat vuosikymmenien ajan olleet näkyvästi esillä sosiaalitieteellisessä keskustelussa ja ne koetaan heidän mukaansa välttämättömäksi osaksi sosiaalitieteellistä tutkimusta. Myös Lehto (1998, 207) ja Alastalo (2005, 45) toteavat sosiaalitieteilijöiden käyneen varsin paljon keskustelua menetelmäkysymyksistä usean vuosikymmenen ajan. Juha Varto (1995, 17) edustaa erilaista näkemystä keskustelujen määrästä. Varton mukaan ihmistieteissä ei ole voinut syntyä metodologista keskustelua, koska siellä on pitäydytty normitetussa todellisuudessa ja muualta annetuissa metodeissa. Varton näkemyksen mukaan ihmistieteissä on omaksuttu hyvin erilaisista filosofisista lähtökohdista syntyneitä menetelmiä ja käsitteitä välittämättä niihin liittyvistä maailma- ja ihmiskäsityksistä. Pöntinen (2004, 90–91) toteaa, että sosiologien käymien menetelmiin liittyvää kiistojen ja

erimielisyyksien syynä ovat voineet olla itse menetelmät, mutta usein menetelmät ovat toimineet vain taistelukenttänä muiden erimielisyyksien ratkomisessa (ks. myös Räsänen ym. 2005, 10).

Sosiaalitieteelliseen menetelmäkeskusteluun voi tutustua muun muassa Tötön (1997b, 11–29), Lehdon (1998, 207–232) ja Alastalon (2005, 17–24) tekstien välityksellä. Tässä yhteydessä nostan esiin Toiviaisen puheenvuoron sosiologipäivillä vuonna 1969. Toiviaisen (1969) puheenvuoro herätti keskustelua. Toiviainen kritisoi siihenastista sosiologiaa sanomalla, että sosiologiasta puuttui filosofia. Hänen mukaansa sosiologiassa oli vain koottu yhteiskunnallisia tosiasioita tilastolliseen todennäköisyyteen nojautuen. Tosiasiat jäävät inhimillisesti tyhjiksi, ellei niitä sijoiteta yhteiskunnalliseen todellisuuteen, Toiviainen totesi. Tilastotieteilijä Anders Ekholm (1969) vastasi Toiviaiselle suomalaisen sosiologian olevan metodisesti avutonta eikä siihen ollut hänen mielestään kuin yksi lääke: sosiologian opintoihin olisi lisättävä tilastotieteen opintoja. Ekholmin mielestä sosiologisessa keskustelussa hyvä keskustelija on sellainen, joka osaa lukea tilastollisia tutkimuksia.

1970-luvun alkupuolella Sosiologia-lehdessä käytiin keskustelua, joissa puitiin tilastollisten menetelmien merkitystä ja käyttökelpoisuutta sekä niihin liittyen yleisemminkin selittämiseen ja kausaliteettiin liittyviä pohdintoja. Kritiikin kohteena olivat surveytutkimus ja positivismi, joihin tilastollisten menetelmien käyttö liitettiin (ks. myös Puuronen 2005; Alastalo 2005, 18; 2014; Sulkunen 1972). Pekka Sulkusen (1972, 166) mukaan surveyn ongelmat eivät johtuneet välttämättä sen logiikasta vaan sen suhteesta teoriaan ja sosiologiaan tieteenä. Surveyille ja tilastollisille menetelmille oli Sulkusen mielestä annettu liian suuret saappaat sosiologisessa tutkimuksessa.

Rantalaihon (1972) Sosiologia-lehteen kirjoittama varsin mielenkiintoinen, monipuolinen ja hauska artikkeli on yksi tutuimmista kvantitatiivisen tietoajattelun kriittisistä analyyseistä. Rantalaiho näki vuonna 1972 silloisen positivistiseksi kutsutun metodologian perustuvan luonnontieteistä lainattuun mittaamisen ihanteeseen ja loogiseen päättelyrakenteeseen, tilastotieteelliseen todennäköisyysajatteluun otantateorioineen ja yhteiskuntatieteen aineelliseen työvälineistöön, lähinnä tietokoneisiin ohjelmiseen. Hän tuo kirjoituksessaan esille näihin liittyviä näkökohtia, joista olen poiminut muutaman esimerkin tähän. Rantalaiho (1972) mainitsee muun muassa joidenkin tilastotieteen peruskäsitteiden olevan ideologisesti harhauttavia. Kun ihmisiin tilastollisina yksikköinä sovelletaan hehkulamppujen käyttöaikoihin ynnä muita sarjavalmisteen kehitettyjä otantateorioita ja testejä, sisältää se vielä luonnontieteelliseen positivismiin mittaushanteeseen verrattuna astetta pitemmälle menevän esioletuksen. Ajattelu johtaa näkemään ihmisen tavaranäkökulmana. Artikkelissa todetaan myös yhteiskuntatieteellisen tutkimusajattelun rajoittuvan surkean usein ”aineiston hankkimiseen” tai ”aineiston käsittelyyn”. Rantalaiho nimittää vulgaari-positivismiksi sellaista metodologista ajattelua ja tutkimuskäytäntöä, jossa tutkimus samastetaan datamatriisin tuottamiseen ja käsittelyyn yksilöitä tapauksina käyttäen

(Rantalaiho 1972, 5–6). Hän syyttää tästä lähinnä tietokoneita. Vulgaaripositivisti-ssä tutkimuksessa aineisto on Rantalaihon kuvauksen mukaan möhkäle raaka-ainetta, jonka työstämisen tuloksena saadaan vastauksia, mutta ei aina tiedetä mihin kysymyksiin. Rantalaihon mielestä asioiden tutkiminen pitäisi tapahtua siten, että tutkittava asia on raaka-aineena ja sen työstämisen tuloksena saadaan kysymyksiä, joihin haetaan empiirinen vastaus mahdollisesti erilaisia lähteitä käyttäen. (Rantalaiho 1972.)

Mäkelä (1995, 59) toteaa, että sosiaalitutkimus on monessa mielessä jopa menetelmiensä vanki. Hän perustelee huomiolla, jonka mukaan menetelmien esittely on saanut tutkimuksissa suuremman roolin kuin itse tutkimustulosten esittely. Ronkainen (1998, 252) toteaa empiirisen sosiaalitutkimuksen olevan paitsi tutkimustapa myös valtaa käyttävää tiedon retoriikkaa (ks. myös Ronkainen 2002).

9.3. Löytyykö yhteinen ymmärrys?

Kiikerin ja Ylikosken (2004, 59–60; 66) mukaan paradigman sisällä olevat tiedeyhteisön jäsenet omaavat samankaltaisen koulutuksen ja tieteellisen taustan, he ovat oppineet samankaltaisista malliesimerkeistä syntyneet tutkimusta ohjaavat intuiot ja heillä jakavat samat muodolliset ja epämuodolliset kommunikaatioverkot. Näin ollen voi paradigmojen välinen kommunikaatio olla haastavaa eivätkä eri paradigmojen edustajat välttämättä edes ymmärrä toisiaan. (Kiikeri ja Ylikoski 2004, 59–60; 66.) Ronkainen (1999, 102–103) jakaa Kiikerin ja Ylikosken näkemyksen todetessaan, että tiedeyhteisön jäsenten samankaltainen tausta, vallitsevat samankaltaiset normit ja puhutavat voivat tehdä tiedeyhteisöstä varsin rajatun, jolloin yhteisön piirissä vallitsevat näkökulmat alkavat vaikuttaa itsestään selviltä. Myös Sintonen (2005) on esittänyt tieteiden eriytymiseen liittyvän huolensa kysymyksellä: ”Jos koulutuksella tähdätään vain yhden tieteenalan kielen, käsitteellisen ja aineellisen välineistön omaksumiseen, niin miten tieteenharjoittaja pystyy tällöin näkemään, miltä maailma näyttää muiden tieteenalojen valossa?”. Peter L. Berger ja Thomas Luckmann (1995, 15) tuovat tähän liittyen esiin Pascalin toteamuksen: ”Se mikä on totta yhdellä puolella Pyreneitä, ei välttämättä päde toisella puolella”.

Clogg (1992) suosittelee, että tilastotieteilijöiden olisi kiinnitettävä enemmän huomiota tilastotieteen, matematiikan ja muiden tieteiden – sosiaalitieteet mukaan lukien –yhteistyöhön ja vuorovaikutukseen. Kvantitatiivista tutkimusta tekevän tutkijan odotetaan hallitsevan tilastolliset menetelmät, mutta häntä neuvovan tilastotieteilijän ei odoteta huomioivan yhteiskuntatieteellistä (tai jonkun muun tieteenalan) näkökulmaa. Sirkka Hirsjärvi (1984) kuvaa tutkijan ja tilastotieteilijän yhteistyötä seuraavasti: ”Kun tutkija hakee neuvoa menetelmäasiantuntijalta niin kuin nykyään tapahtuu, on apua odotettavissa teknisessä mielessä. On tosin ilmeistä, että teknisissäkin kysymyksissä asiantuntija voisi auttaa paremmin, jos hän tuntisi

sisällöllistä problematiikkaa.” Tilastotieteilijälle analyysimenetelmät ovat lähinnä tekniikoita, joiden avulla aineistosta haetaan rakenteita. Sosiologille ne eivät ole vain aineiston rakennetta eri tavoin etsiviä tekniikoita vaan ovat osa prosessia, jossa sosiologin teoreettinen perehtyneisyys ja aineiston parissa tehtävä työ nivoutuvat yhteen (Ronkainen 2014, 125–126). Sund (2000) toteaa, miten tilastollisten menetelmien suorittamisessa niin sanottu tekninen näpertely saa usein liian suuren vallan. Myös Stefan Collini (1998, 76) toteaa osuvasti, että humanististen alojen oppiaineiden ”luonnontieteellistä lukutaidottomuutta” on valiteltu, mutta vähintään yhtä vahingollista on luonnontieteen tutkijoiden historiallinen osaamattomuus.

Esa Läärä ja Seppo Lammi (1986, 38) toteavat tutkijoiden kääntyvän tilastotieteilijän puoleen usein liian myöhään eli vasta siinä vaiheessa, kun heillä on käsissään valmis aineisto. Heidän mukaansa tutkija saattaa kääntyä tilastotieteilijän puoleen myös ilman tutkimuskysymyksiä esittämällä kysymyksen: ”Minulla on tässä yksi tutkimusaineisto. Mitä sille pitäisi tehdä?” He toteavat, että jos suunnittelussa ja toteutuksessa on pahoin laiminlyöty eräitä tärkeitä tilastollisia periaatteita, on tutkimusta tässä vaiheessa vaikea pelastaa. Toisaalta Ronkainen ym. (2013, 125–126) toteavat analysoinnin olevan parhaimmillaan yrittämistä, erehtymistä ja erilaisten kokeilujen tekemistä, joista osa tuottaa käyttökelpoista tietoa ja osa ei, mutta kaikki kokeilut lisäävät tutkijan tietoa aineistostaan ja tutkimuskohteestaan. Heidän näkemyksensä mukaan paras ratkaisu syntyy suurelta osin sosiologiseen asiantuntemukseen perustuvista valinnoista, vaikka toki tilastolliset kriteeritkin on huomioitava.

Lazarsfeld (ks. Boudon 1993, 12) ja Sund (2000) vertaavat tilastotiedettä työkalupakkiin. Tutkimuksen suorittamista puolestaan voidaan verrata talon rakentamiseen (ks. Hox ym. 2008, 3; 15). Tilastotieteilijä osaa valita oikean työkalun eri tilanteeseen. Mutta työkalujen käyttötaito ei riitä siihen, että osaisi rakentaa kokonaisen talon. Työkalujen lisäksi talon rakentamiseen tarvitaan suunnitelma ja valvonta sen suhteen, että rakentaminen etenee suunnitelmien mukaan. Sund (2008, 20) tuokin esille tutkijan ymmärryksen merkityksen nimetessään tutkimusprosessin päävaiheiksi ilmiön ymmärtämisen, tutkimustavoitteen ymmärtämisen, aineiston esivalmistelun, mallinnuksen, arvioinnin ja raportin. Sosiologi Marshia (1982, 2-3; 148) lainaten ja hieman muuttaen voi todeta, että tilastolliset menetelmät ovat työkalu, ja työkaluna niitä on mahdollista käyttää monin tavoin (ks. myös Alastalo 2005, 19). Helenius (1992, i; ks. myös Sund 2008, 26) kuvaa tilastotieteen oppikirjassaan hyvän tilastollisen analyysin olevan kaavamaisuuden sijasta paremminkin lähellä taidetta.

Aineistoja analysoidessaan ja tilastollisia menetelmiä käyttäessään sosiologit joutuvat jatkuvasti tasapainottelemaan numeroiden ja niiden sisällöllisten merkitysten välissä. Tilastollisten menetelmien perustuessa matematiikkaan ovat monet niihin liittyvät kriteerit, säännöt ja suositukset matemaattisesti tarkkoja. Sosiologit joutuvat arvioimaan, miten aineiston muuttujat ja niille tehdyt analyysit tavoittavat ilmiötä. Muuttujia analysoidessaan he voivat asettua tilastomatemattiseen ajattelutapaan ja

käyttää tilastollisia menetelmiä matemaattisen tarkasti. Tai he voivat pysytellä sosiologisessa näkökulmassaan ja suhtautua tilastollisten menetelmien antamaan tietoon suuntaa antavana tietona, joka viimeistellään lopulliseen muotoonsa sosiologisen tulkinnan avulla. On kuitenkin mahdotonta arvioida, saati suositella sitä, kuinka paljon sosiologisessa kvantitatiivisessa analyysissä tulisi hyväksyä jouston varaa menetelmien matemaattisen tarkkuuden suhteen ja kuinka paljon antaa painoa sosiologiselle tulkinnalle. Osaltaan jouston vara riippuu käytetyistä menetelmistä. Näen faktorianalyysin, ryhmittelyanalyysin ja korrespondenssianalyysin tarjoavan hyvin joustavat mahdollisuudet sosiologiselle tulkinnalle. Myös monia menetelmiä sisällään pitävät menetelmät, kuten lineaariset rakenneyhtälömallit, mahdollistavat ja jopa edellyttävät sisällöllisen asiantuntemuksen käyttöä.

Ronkainen (2001, 28–29) toteaa, että tieteellinen tieto on aina sidoksissa tietäjäänsä, hänen instituutioonsa sekä tiettyyn kulttuuriin ja yhteiskunnalliseen ja historialliseen kontekstiin. Pystyäkseen ymmärtämään ja käsitteellistämään tutkimuskohdettaan on tutkijan sijoittauduttava johonkin keskusteluyhteisöön ja samalla antauduttava yhteisössä vallitsevalla ajattelutavalle. Mutta jos tutkija pystyy ottamaan etäisyyttä ja vaihtamaan tietämisen paikkaansa, hän pystyy myös refleктоimaan tietämistään ja tunnistamaan yhteisön pitämät itseäänselvytykset sekä tunnistamaan edustamansa tiedon rajallisuuden ja ehdollisuuden. (Ronkainen 2001, 28–29.)

9.4. Tulevaisuuden näkymät

Raftery (2000) totesi 2000-luvun alussa kansainvälisten tutkimusten olevan suuntautumassa täsmällisempiin hypoteeseihin, suurempiin ja yksityiskohtaisempiin aineistoihin sekä monimutkaisempiin tilastollisiin malleihin. Erola ja Räsänen (2014a, 246–247) arvelevat, että menetelmien merkitys sekä erityisesti kvantitatiivisen menetelmäosaamisen vaatimustaso tulevat kasvamaan myös suomalaisessa sosiologiassa. Erola ja Räsänen (2014b, 68) viittaavat uusiin teknologisiin tiedonkeräystapoihin ja big dataan. Big data tarkoittaa suurta järjestelemätöntä tietomassaa, jota kerääntyy muun muassa kauppojen asiakaskorttien käytön ja verkkovierailujen yhteydessä (Piela 2015, 25; Heikinheimo ja Ukkonen 2015; Hatakka 2015; Lazer ym. 2009). Semi Purhonen ja Arho Toikka (2016, 9) mainitsevat big dataan liittyen kolme v-kirjaimella alkavaa kuvausta: big data on valtavaa kooltaan, vauhdiltaan nopeaa ja vaihtelevaa muodoltaan. Heidän mukaansa yhteiskuntatieteilijät näkevät big datan houkuttelevana uusien aineistomuotojen ja menetelmien takia (ks. esim. Robertson ja Travaglia 2017). Uhkakuvana he näkevät big datan kykeytymisen tarkkailuun, valvontaan ja hallintaan. Big datan analysointi ei välttämättä onnistu perinteisillä tilastollisilla menetelmillä (Sund 2008, 19; 26). Sund (2015, 43) varoittaa big datan analyysiin – ja yleensäkin aineistolähtöiseen analyysiin – sisältyvistä

sudenkuopista. Vaikka teknologian kehitys mahdollistaakin yhä isompien datamäärien analysoinnin, ei teknologialla ratkaista tietoteoreettisia ongelmia (Sund 2015, 45). Aineiston suuruus ei takaa tilastollisen mallin onnistuneisuutta. Siitä voi Sundin (2005, 511) mukaan aiheutua jopa haittaa. Hän jatkaa, että yhteiskuntatieteellisessä tutkimuksessa tavoitteena ei ole mallin tilastotieteellinen täydellisyys vaan se, millä tavalla malli auttaa tutkijaa ymmärtämään tutkimaansa ilmiötä. Vuoden 2008 jälkeen valmistuneita sosiologian väitöskirjoja selaillessani kiinnitin huomiota rekisteriaineistojen yleistymiseen tutkimusaineistoina. Sund (2008; 2015) tuo esiin, että rekisteriaineistojen analysoinnissa on haasteita, joihin ei perinteisillä tilastotieteen keinoilla pystytä vastaamaan.

Erola ja Räsänen (2014c, 19) näkevät sosiologian nykyisellä monitieteisyyden aikakaudella pikemminkin näkökulmana ilmiöiden tutkimiseen kuin perinteisenä, yksiselitteisen tutkimuskohteen omaavana tieteenä. He toteavat yleisestikin tieteenalojen määritelmien olevan yhä monimutkaisempia. Heidän mukaansa usein kuulee sanottavan, että nykypäivänä kaikki tieteet ovat lähtökohtaisesti monitieteisiä. 2000-luvulla on yleistynyt mixed methods-tutkimustapa, jossa hyödynnetään sekä laadullisia ja määrällisiä menetelmiä (ks. esim. Kinnunen 2018 30–36). Myös tilastotiede on moninainen tiede eikä sille ole löydettävissä riittävän kattavaa määritelmää. Tilastotiedettä sovelletaan eri yhteyksissä monin eri tavoin. Voidaan kysyä, että onko jopa harhaanjohtavaa kutsua sitä kaikkea tilastotieteeksi.

Ylikoski (2000) on viitannut toimijaverkkoteorian yhteydessä valta-analyysiin. Ylikosken tapaan näen rajakohdemallin kauttakulkupisteiden valinnat eli kääntämiset kamppailuna tieteellisestä auktoriteetista. Tilastotiede on ollut vahvoilla tässä valtakamppailussa, koska tieteellisyyden ja matemaattisuuden välille on asetettu helposti yhtäläisyysmerkit (Gieryn 1983; Suoranta 1995, 33; Sund 2000; Bryman 1988, 12). Matematiikan hallitseminen auttaa sosiologia monimuuttujamenetelmien ja yleensäkin tilastollisen analyysin suorittamisessa (Bartholomew 1995a), mutta liiallinen keskittyminen matematiikkaan ei auta tutkimuskohteen haltuunotossa eikä takaa tutkimuksen onnistumista. Tilastolliset menetelmät paljastavat tutkijalle arvokasta tietoa aineistosta, mutta myös muunlaisen tiedon, erityisesti sosiologisen asiantuntemuksen käyttö ovat tarpeellisia analyysia tekeväälle tutkijalle.

Vesa A. Niskanen (2004, 34) suosittelee tutkijoita olemaan alistumatta pelkästään matemaattisen ajattelutavan piiriin. Hän toteaa, että vaikka matematiikka sinänsä on täsmällistä ja tarkkaa, ei sen käyttö tilastollisissa malleissa tee mallia täsmälliseksi tai tarkaksi. Niskanen toteaa, että matemaattinen mallinnus on erityisen ongelmallista ihmistieteissä, koska matemaattisten kaavojen vastaavuus maailmamme ilmiöiden kanssa on haasteellista. Tilastollisten analyysimenetelmien sosiologisessa tai yhteiskuntatieteellisessä käyttämisessä on kuitenkin tärkeitä käsitteellisiä ajattelutapojen eroja verrattuna ”perinteiseen” tilastotieteelliseen tapaan. Sosiologisen tutkimuksen tärkein työkalu on siksi sosiologinen asiantuntemus.

Lähteet

- Ahola, Anja (1993). Menetelmäkö määrää tulokset? Puhelin- ja käyntihaastattelun eroista hyvinvointikysymyksiin vastattaessa. *Sosiologia* 30:3. 200–211.
- Ahola, Anja (1998). Haastattelutilanne surveyvastausten rakentajana. Teoksessa Paananen, Seppo & Juntto Anneli & Sauli, Hannele (toim.) *Faktajuttu. Tilastollisen sosiaalitutkimuksen käytännöt*. Vastapaino. Tampere. 257–281.
- Ahola, Anja (2000a). Surveykysymysten tarkoitus ja vastausten tulkittavuus. *Hyvinvointikatsaus* 1/2000. 67–71.
- Ahola, Anja (2000b). Surveykysymysten tarkoitus ja haastattelutavan valinta. *Hyvinvointikatsaus* 4/2000. 64–70.
- Ahmavaara, Yrjö (1970). Transformaatioanalyysin salaisuus. *Sosiologia* 7:1. 55–60.
- Ahtiainen, Pekka & Tervonen, Jukka (1994). Traditio vai muutos? Historian ja sosiologian vaikea veljeys. Teoksessa Ahtiainen, Pekka & Rätty, Teuvo & Strömberg, John & Tervonen Jukka (toim.) *Historia, sosiologia ja Suomi. Hanki ja jää*. Helsinki. 7–18.
- Airaksinen Timo (1987). Tiedepolitiikan oikeudenmukaisuus: tutkimuksen vapaus, tutkijan oikeudet, tieteen laatu ja tuki. Teoksessa Airaksinen, Timo & Häyry, Heta & Häyry, Matti *Tiedepolitiikan oikeudenmukaisuus ja tutkijan vastuu*. Gaudeamus. Helsinki. 9–73.
- Alastalo, Marja (2005). Metodisuhdanteiden mahti. Lomaketutkimus suomalaisessa sosiologiassa 1947–2000. Vastapaino. Tampere.
- Alastalo, Marja (2014). Sosiologian menetelmät ja tutkimuksen infrastruktuuri. *Sosiologia* 51:3, 281–285.
- Alasuutari, Pertti (1989). Erinomaista, rakas Watson: johdatus yhteiskuntatutkimukseen. *Hanki ja jää*. Helsinki.
- Alasuutari, Pertti (1999). Laadullinen tutkimus. 3. uudistettu painos. Vastapaino. Tampere.
- Alasuutari, Pertti (2000). Hypoteettisesta universumista ja tilastollisista yhteyksistä johtolankoina. *Sosiologia* 37:2. 151–154.
- Alestalo, Matti & Rätty, Teuvo (1994). Sosiologian 1960- ja 1970-luvun kriisi väitöskirjojen valossa. Teoksessa Ahtiainen, Pekka & Rätty, Teuvo & Strömberg, John & Tervonen Jukka (toim.) *Historia, sosiologia ja Suomi. Hanki ja jää*. Helsinki. 217–243.
- Alkula, Tapani & Pöntinen, Seppo & Ylöstalo, Pekka (1994) *Sosiaalitutkimuksen kvantitatiiviset menetelmät*. WSOY. Helsinki.
- Allardt, Erik (1969). Sosiologisista selityksistä. *Sosiologia* 6:1, 11–22.
- Allardt, Erik (1973). Suomalaisen sosiologian tutkimisesta. Teoksessa Allardt, Erik & Eskola, Antti & Haatanen, Pekka & Haavio-Mannila, Elina & Toiviainen, Seppo (toim.) *Suomalaisen sosiologian juuret*. WSOY. Porvoo. 1–26.
- Allardt, Erik (1988). *Sosiologia I*. 4. painos. WSOY. Juva.
- Allardt, Erik & Alapuro, Risto & Alestalo, Matti (1992). Suomalaisen sosiologian historiasta. Teoksessa Alapuro, Risto & Alestalo, Matti & Haavio-Mannila, Elina (toim.) *Suomalaisen sosiologian historia*. WSOY, Porvoo. 13–25.
- Allardt, Erik & Jartti, Pentti & Jyrkilä, Faina & Littunen, Yrjö (1958). Nuorison harrastukset ja yhteisön rakenne. WSOY. Porvoo, Helsinki.

- Alwin, Duane F. & Beattie, Breet A. (2016). The KISS Principle in Survey design: Question Length and Data Quality. *Sociological Methodology* 46:1. 121–152. Osoitteessa: <https://doi.org/10.1177%2F0081175016641714>. Luettu 15.4.2017.
- Aro, Jari (1999). *Sosiologia ja kielenkäyttö. Retoriikka, narratiivi, metafora*. Tampereen yliopisto. Tampere.
- Aro, Jari & Jokivuori, Pertti (2015). *Klassinen sosiologia ja moderni maailma*. Docendo Oy. Jyväskylä.
- Asp, Erkki (1965). *Lappalaiset ja lappalaisuus: Sosiologinen tutkimus Suomen nykylappalaisista*. Turun yliopisto. Turku
- Babchuk, Nicholas & Keith, Bruce & Peters, George (1999). Collaboration in sociology and other scientific disciplines. *The American Sociologist* 30:3. 5–21. Osoitteessa: <https://www.jstor.org/stable/27698918>. Luettu 19.5.2017.
- Bandalos, Deborah L. & Boehm-Kaufman, Meggen R. (2010). Four Common Misconceptions in Exploratory Factor Analysis. Teoksessa Lance, Charles E & Vandenberg, Robert J. (toim.) *Statistical and Methodological Myths and Urban Legend. Doctrine, Verity and Fable in the Organizational and Social Sciences*. Routledge. New York, London. 61–88.
- Bartholomew, David J. (1995a). What is Statistics? *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 158:1. 1–20.
- Bartholomew, David J. (1995b). Spearman and the origin and development of factor analysis. *British Journal of Mathematical and Statistical Psychology*. 48:2. 211–220. Osoitteessa: DOI: <https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1111%2Fj.2044-8317.1995.tb01060.x>. Luettu 12.6.2015.
- Bellhouse, D. R. (2000). Survey sampling theory over the twentieth century and its relation to computing technology. *Survey Methodology* 26:1. 11–20. *Statistics Canada, Catalogue No. 12–001*.
- Berger, Peter L. & Luckmann, Thomas (1995). *Todellisuuden sosiaalinen rakentuminen: tiedonsosiologinen tutkielma. 2. painos*. Gaudeamus. Helsinki
- Biemer, Paul P. & Christ, Sharon L. (2008). Weighting survey data. Teoksessa de Leeuw, Edith D & Hox, Joop J. & Dillman, Don A. (toim.) *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. New York. 317–341.
- Bird, Alexander (2001). Thomas Kuhn. ProQuest Ebook Central <https://ebookcentral-proquest-com.ezproxy.ulapland.fi>. Luettu 29.1.2020.
- Blaxter, Loraine & Hughes, Christina & Tight, Malcolm (1996). *How to research*. Open University Press. Buckingham.
- Blomgren, Jenni (2005). Huono-osaisuus Suomen kaupunkiseutukunnissa: alue-erot ja sosiaalisen ympäristön vaikutukset 1990-luvulla. Helsingin yliopiston sosiologian laitoksen tutkimuksia nro 24. Helsingin yliopisto. Helsinki.
- Blumer, Herbert (1956). Sociological Analysis and the "Variable". *American Sociological Review* 21:6. 683–690.
- Bollen, Kenneth, A. & Long, Scott, J. (toim.) (1993). *Testing Structural Equation Models*. SAGE Publications. Newbury Park, London, New Delhi.
- Boudon, Raymond (1993). *On social research and its language*. University of Chicago Press. Chicago.
- Brown, James D. (2009). Principal components analysis and exploratory factor analysis - Definitions, differences, and choices. *Shiken: JALT Testing & Evaluation SIG Newsletter* 13:1. 26 – 30.
- Bryant, Christopher G. A (1985). *Positivism in Social Theory and Research*. MacMillan. London.
- Bryman, Alan (1988). *Quantity and Quality in Social Research*. Unwin Hyman. London.

- Bryman, Alan & Cramer, Duncan (1990). *Quantitative Data Analysis for Social Scientists*. Routledge. London and New York.
- Bryman, Alan & Cramer, Duncan (2004). Constructing Variables. Teoksessa Hardy, Melissa & Bryman, Alan (toim.) *Handbook of data analysis*. SAGE. London. 17–34.
- Byrne, David (2001). What Is Complexity Science? Thinking as a Realist About Measurement and Cities and Arguing for Natural History. *Emergence* 3:1. 61–76.
- Byrne, David (2002). *Interpreting quantitative data*. Sage Publications Ltd London.
- Byrne, Dave (2004). Complex and contingent causation – the implications of complex realism for quantitative modelling. The case of housing and health. Teoksessa Carter, Bob & New, Caroline (toim.) *Making realism Work. Realist Social theory and Empirical Research*. Routledge, Oxfordshire. New York. 50–66.
- Byrne, David (2011). *Applying social science. The role of social research in politics, policy and practice*. Policy Press. Bristol.
- Carlile, Paul R. (2002). A Pragmatic View of Knowledge and Boundaries. *Boundary Objects in New Product Development, Organization Science*, 13:4. 442–455.
- Champkin, Julian (2014). The timeline of statistics. Osoitteessa: <https://www.statslife.org.uk/history-of-stats-science/1190-the-timeline-of-statistics>. Luettu 8.10.2018.
- Clogg, Clifford C. (1992). The Impact of Sociological Methodology on Statistical Methodology. *Statistical Science* 7:2. 183–196.
- Collini, Stefan (1998). Johdanto. Teoksessa Snow, C. P. *Kaksi kulttuuria. Suom. Kimmo Pietiläinen. Terra Cognita*. Helsinki. 7–91.
- Cook, Thomas D. & Cooper, Harris & Cordray, David S. & Hartmann, Heidi & Hedges, Larry V & Light, Richard J. & Louis, Thomas A. & Mosteller, Frederick (1992). *Meta-Analysis for Explanation. A Casebook*. Russell Sage Foundation. New York.
- Cook, Colleen & Heath, Fred & Thompson, Russel, L (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement* 60:6. 821–836.
- Cornesse, Carina & Bosnjak, Michael (2018). Is there an association between survey characteristics and representativeness? A meta-analysis. *Survey Research Methods* 12:1. 1–13.
- Costello, Anna B. & Osborne, Jason (2005). Best practices in in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation* 10:7. Osoitteessa: DOI: <https://doi.org/10.7275/jyj1-4868>. Luettu 12.5.2020.
- Cresswell, John W. (1994). *Research Design, Qualitative and Quantitative Approaches*. Sage Publications. London.
- de Leeuw, Edith, Desirée (2008). Choosing the Method of Data Collection. Teoksessa de Leeuw Edith D. & Hox Joop J. & Dillman Don A. (toim.) *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. New York. 113–35.
- Desrosières, Alain (2002). *The Politics of Large Numbers. A History of Statistical Reasoning*. Harvard University Press. Cambridge.
- Ekholm, Anders (1969). Toiviaisien ilmapallot. *Sosiologia* 6:2. 86.
- Engman, Athena (2013). Is there life after $P < 0.05$? Statistical Significance and quantitative sociology. *Quality and Quantity* 47:1. 257–270. Osoitteessa <https://doi.org/10.1007/s11135-011-9516-z>. Luettu 17.5.2019.
- Erola, Jani & Räsänen, Pekka (2014a). Sosiologian haasteet 2000-luvulla. Teoksessa Erola, Jani & Räsänen, Pekka (toim.) *Johdatus sosiologian perusteisiin*. Gaudeamus. Helsinki. 241–250.
- Erola, Jani & Räsänen, Pekka (2014b). Sosiologinen tutkimus ja menetelmät. Teoksessa Erola, Jani & Räsänen, Pekka (toim.) *Johdatus sosiologian perusteisiin*. Gaudeamus. Helsinki. 60–79.
- Erola, Jani & Räsänen, Pekka (2014c). Sosiologinen lähestymistapa. Teoksessa Erola, Jani & Räsänen, Pekka (toim.) *Johdatus sosiologian perusteisiin*. Gaudeamus. Helsinki. 11–27.

- Ervasti, Heikki & Saari, Juho (2011). Onnellisuus hyvinvointivaltiossa. Teoksessa Saari, Juho (toim.) Hyvinvointi. Suomalaisen yhteiskunnan perusta. Gaudeamus. Helsinki. 191–218.
- Eskola, Antti (1961). Social Influence and Power in Two-person Groups. Transactions of the Westermarck Society 6. Turku.
- Eskola, Antti (1962). Sosiologian tutkimusmenetelmät I. WSOY. Porvoo, Helsinki.
- Eskola, Antti (1967). Sosiologian tutkimusmenetelmät II. WSOY. Porvoo, Helsinki.
- Eskola, Antti (1971). Faktorianalyysi suomalaisessa sosiologiassa. Sosiologia 8:6. 321–325.
- Eskola, Antti (1973). Suomalaisen sosiologian uudistuminen. Teoksessa Alapuro, Risto & Alestalo, Matti & Allardt, Erik & Eskola, Antti & Haatanen, Pekka & Haavio-Mannila, Elina & Toiviainen, Seppo (toim.) Suomalaisen sosiologian juuret. WSOY. Helsinki. 269–317.
- Eskola, Antti (1977). Sosiologia. Johdatus perusteisiin. Neljäs painos. WSOY. Porvoo.
- Eskola, Antti (1992). Sosiologian uudistuminen 1950-luvulla. Teoksessa Alapuro, Risto & Alestalo, Matti & Haavio-Mannila Elina (toim.) Suomalaisen sosiologian historia. WSOY, Porvoo. 241–285.
- Eskola, Antti (1994). Sosiaalitieteen muuttuvat tekstit ja käytännöt. Teoksessa Weckroth, Klaus & Tolkki-Nikkonen, Mirja (toim.) Jos A niin... Vastapaino. Tampere. 13–53.
- Eskola, Antti (1997). Jäähvyäisluentoja. Tammi. Helsinki.
- Eskola, Jari & Suoranta, Juha (1995). Kvalitatiivista kasvatustutkimusta jäljittämässä – kvalitatiivisten tutkimusmenetelmien tulosta suomalaiseen kasvatustieteeseen. Teoksessa Eskola, Jari & Mäkelä, Jukka & Suoranta, Juha (toim.) Ihmistieteiden 1990-luvun metodologiaa etsimässä. Keskustelua kasvat- ja sosiaalitieteiden 1990-luvun metodologiasta. Lapin yliopiston kasvatustieteellisiä julkaisuja C 8 Katsauksia ja puheenvuoroja. 89–107.
- Forsman, Maria (2016). Julkaisut ja tieteen mittaaminen. Bibliometriikan käännekohtia. Enostone. Tampere.
- Gieryn, Thomas F. (1983). Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. American Sociological Review 48:6. 781–795. Osoitteessa: <https://doi-org.ezproxy.ulpland.fi/10.2307/2095325>. Luettu 8.12.2015.
- Glass, Gene V & McGaw, Barry & Smith, Mary Lee (1987). Meta-analysis in social research. Sage publications. London.
- Gorard, Stephen (2016). Damaging Real Lives through Obstinacy: Re-Emphasising Why Significance Testing is Wrong. Sociological Research Online 21:1. 102–115. Osoitteessa: <https://doi.org/10.5153/sro.3857>. Luettu 24.4.2020.
- Gorard, Stephen (2018). Do we really need confidence intervals in the new statistics? International Journal of Social Research Methodology 22:3. 281–291. Osoitteessa: <https://doi.org/10.1080/13645579.2018.1525064>. Luettu 24.4.2020.
- Gorard, Stephen & Gorard, Jonathan (2015). What to do instead of significance testing? Calculating the ‘number of counterfactual cases needed to disturb a finding’. International Journal of Social Research Methodology 19:4. 481–490. Osoitteessa: <https://doi.org/10.1080/13645579.2015.1091235>. Luettu 24.4.2020.
- Gronow, Jukka & Noro, Arto & Töttö, Pertti (1997). Johdanto teoksessa Gronow, Jukka & Noro, Arto & Töttö, Pertti Sosiologian klassikot. Gaudeamus. Helsinki.
- Haaparanta, Leila & Niiniluoto, Ilkka (1986). Johdatus tieteelliseen ajatteluun. Helsingin yliopiston filosofian laitoksen julkaisuja 3.
- Haatanen Pekka (1992). Sosiaalishistoria. Teoksessa Alapuro, Risto & Alestalo, Matti & Haavio-Mannila, Elina (toim.) Suomalaisen sosiologian historia. WSOY, Porvoo. 149–199.
- Hacking, Ian (1990). The taming of chance. Cambridge University Press. Cambridge.

- Hague, Paul (1993). *Questionnaire design*. Kogan Page. London.
- Hair, Joseph F. Jr. & Anderson, Rolph E. & Tatham, Ronald L. & Black, William C. (1998). *Multivariate data analysis*. Fifth edition. Prentice Hall. New Jersey.
- Hair, Joseph F. Jr. & Anderson, Rolph E. & Tatham, Ronald L. & Black, William C. (2005). *Multivariate data analysis*. Sixth edition. Prentice Hall. New Jersey.
- Hand, David A. (1994). Deconstructing Statistical Questions. *Journal of Royal Statistical Society. Series A (Statistics in Society)* 157:3. 317–356. Osoitteessa: <https://doi-org.ezproxy.ulapland.fi/10.2307/2983526>. Luettu 12.6. 2015.
- Hand, David A. (2009). Modern Statistics: The Myth and the Magic. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 172:2. 287–306. Osoitteessa: <https://www.jstor.org/stable/20622500>. Luettu 12.6.2015.
- Harman, Harry, H. (1976). *Modern factor analysis*. Third Edition Revised. The University of Chicago Press. Chicago and London.
- Hatakka, Tarja (2015). Soveltuvatko kaupan kanta-asiakasrekisterit kulutuksen tilastointiin? *Tieto & Trendit* 2. 34–38.
- Heikinheimo, Hannes & Ukkonen, Antti (2015). Ison datan alkulähteillä. *Tieto & Trendit* 2. 28–33.
- Heikkilä, Matti (1990). Köyhyys ja huono-osaisuus hyvinvointivaltiossa: Tutkimus köyhyydestä ja hyvinvoinnin puutteiden kasautumisesta Suomessa. Sosiaalivaltio. Helsinki.
- Heiskanen, Markku (1998). *Surveypäiväkirja*. Teoksessa Paananen, Seppo & Juntto, Anneli & Sauli, Hannele (toim.) *Faktajuttu. Tilastollisen sosiaalitutkimuksen käytännöt*. Vastapaino. Tampere. 185–205.
- Heiskanen, Markku (2002). Väkivalta, pelko ja turvattomuus. *Surveytutkimusten näkökulmia suomalaisten turvallisuuteen*. Tilastokeskus.
- Helenius, Hans (1992). *Tilastollisten menetelmien perustiedot*. Paino-S Ky. Tampere.
- Hirsjärvi, Sirkka (1984). Milloin metodi on pätevä. *Suomen kasvatustieteellinen aikakauskirja Kasvatus* 15:5. 279–286.
- Hokka, Johanna (2016). Illusiot suomalaisen sosiologian kentällä. *Sosiologia* 53:3. 239-257.
- Honkala, Kauko (1971). Sosiaalinen eriarvoisuus maaseudulla: Neljässä varsinaissuomalaisessa kylässä suoritettu sosiologinen tutkimus. Turku.
- Hox, Joop J. (2008). Accomodating measurement errors. Teoksessa de Leeuw, Edith D & Hox, Joop J & Dillman, Don A. (toim.) *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. New York. 387–402.
- Hox, Joop J.& de Leeuw, Edith D. & Dillman, Don A. (2008). The cornerstones of survey research. Teoksessa de Leeuw, Edith D & Hox, Joop J & Dillman, Don A. (toim.) *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. New York. 1–17.
- Hunter, Laura & Leahey, Erin (2008). Collaborative Research in Sociology: Trends and Contributing Factors. *The American Sociologist* 39:4. 290–306. Osoitteessa: <https://doi-org.ezproxy.ulapland.fi/10.1007/s12108-008-9042-1>. Luettu 28.8.2018.
- Hyhkö, Heikki (2013). Johdatus monimuuttujamenetelmiin. Osoitteessa <http://www.mv.helsinki.fi/hyhko/mmm13ks/monim.pdf>. Luettu 28.8.2018.
- Hämäläinen, Seppo & Leskinen Esko (1985). LISREL-mallien soveltaminen seuranta-aineistoon analysoitaessa ryhmittelyratkaisujen tehokkuutta peruskoulun yläasteen englanninkielen oppimisessa. Selosteita ja tiedotteita. Kasvatustieteiden tutkimuslaitos. Jyväskylä.
- Jokivuori, Pertti (2002). Sitoutuminen työorganisaatioon ja ammattijärjestöön: kilpailevia vai täydentäviä? Jyväskylän yliopisto. Jyväskylä.
- Jokivuori, Pertti & Hietala, Risto (2007). Määrällisiä tarinoita. Monimuuttujamenetelmien käyttö ja tulkinta. WSOY. Helsinki.

- Joliffe, IT & Morgan, Bjt (1992). Principal component analysis and exploratory factor analysis. *Statistical Methods in Medical Research* 1:1. 69–95. Osoitteessa: <https://doi.org/10.1177/096228029200100105>. Luettu 12.5.2020.
- Jyrinki, Erkki (1974). Kysely ja haastattelu tutkimuksessa. Gaudeamus. Helsinki.
- Kainulainen, Sakari (1998). Elämäntapahtumat ja elämään tyytyväisyys eri sosiaaliluokissa: Life events and satisfaction with life in different social classes. Kuopion yliopisto. Kuopio.
- Kalimo, Esko (1969). Lääkintäpalvelusten käyttöön vaikuttavat tekijät: Aikuisväestön sairauskäyttäytyminen ja sen vaikuttimet Suomessa ennen sairausvakuutusta korrelatiivisen monimuuttujatarkastelun nojalla. Sosiaaliturvan tutkimuslaitos. Helsinki.
- Kandolin, Irja (1997). Gender, worklife and family responsibilities in Finland and Estonia. Effects on economic and mental well-being. Finnish Institute of Occupational Health. Helsinki.
- Kankainen, Tomi (2007). Yhdistykset, instituutiot ja luottamus. Jyväskylän yliopisto. Jyväskylä.
- Karvonen Sakari (1997). The Regional Context of Health Behaviour among Finnish Adolescents. Stakes. Helsinki.
- Ketokivi, Mikko (2009). Tilastollinen päättely tieteellisenä argumenttina. Gaudeamus. Helsinki.
- Kiikeri Mika & Ylikoski, Petri (2004). Tiede tutkimuskohteena. Filosofinen johdatus tieteentutkimukseen. Gaudeamus. Helsinki.
- Kinnunen, Maarit (2018). Total festival experience: a mixed methods research approach to consumer experiences in Finnish cultural festivals. *Acta Universitatis Lapponiensis*. Lapin yliopisto. Rovaniemi.
- Kuhn Thomas S. (1994). Tieteellisten vallankumousten rakenne. Art House. Juva.
- Kutvonen, Heikki (2004). Oi atk. Osoitteessa http://www.fujitsu.fi/historia/oi_atk/. Luettu 19.7.2017.
- Kuusela Vesa (2000). Tilastografikan perusteet. Edita. Helsinki.
- Kuusela, Vesa (2010). Tilastonteko on yhtä vanhaa perua kuin kirjoitustaito. *Hyvinvointikatsaus* 2010:4. 55–58.
- Kuussaari, Kristiina (2006). Näkemysten kirjo, sirpaloitunut tieto: Terveystieteiden, sosiaalitoimen ja nuorisotoimen työntekijöiden käsityksiä huumeiden käyttäjien hoitopalvelujärjestelmästä. Stakes. Helsinki.
- Kämäräinen, Kauko (1966). Kaupunkijärjestelmän alueellinen eriytyminen ja poikkeavuus. WSOY. Porvoo, Helsinki.
- Köppä, Tapani (1979). Viljelijäperhe, yhteisö ja yhteistoiminta. Tutkimus mautilojen välisen yhteistoiminnan kehitykseen ja omaksumiseen vaikuttavista sosiaalisista ja taloudellisista tekijöistä. Pellervon taloudellinen tutkimuslaitos. Helsinki.
- Laaksonen, Seppo (2013). *Surveymetodiikka*. Aineiston kokoamisesta puhdistamisen kautta analyysiin. Osoitteessa <https://bookboon.com/fi/surveymetodiikka-ebook>. Luettu 6.11.2018.
- Lakomäki, Sami & Latvala, Pauliina & Laurén, Kirsi (2011). Menetelmien jäljillä. Teoksessa Lakomäki, Sami & Latvala, Pauliina & Laurén, Kirsi *Tekstien rajoilla*. Monitieteisiä näkökulmia kirjoitettuihin aineistoihin. Suomalaisen Kirjallisuuden Seura. Helsinki. 7–27.
- Lammi-Taskula, Johanna (2007). Parental leave for fathers? Gendered conceptions and practices in families with young children in Finland. Stakes. Helsinki.
- Latour, Bruno (1987). *Science in Action: how to follow scientists and engineers through society*. Harvard University Press Cambridge, Massachusetts.
- Latour, Bruno (1990). Postmodern? No, simply a modern! Steps towards an anthropology of Science. *Studies in History and Philosophy of Science Part A*. 21:1. 145–171. Osoitteessa: [https://doi.org/10.1016/0039-3681\(90\)90018-4](https://doi.org/10.1016/0039-3681(90)90018-4). Luettu 12.6.2017.
- Latour, Bruno (1993). *We Have Never Been Modern*. Transl. Catherine Porter. Harvard University Press Cambridge, Massachusetts.

- Lazarsfeld, Paul, F. (1993a). *Sociology of Social Research. The Sociology of Empirical Social Research*. Teoksessa Boudon, Raymond (ed.) *On Social Research and its Language*. The University of Chicago Press. Chicago and London. 257–274.
- Lazarsfeld, Paul F. (1993b). *The Relevance of Methodology. Methodological Problems in Empirical Social Research*. Teoksessa Boudon, Raymond (ed.) *On Social Research and its Language*. The University of Chicago Press. Chicago and London. 236–254.
- Lazer, David M. & Pentland, Alex & Adamic, Lada & Aral, Sinan & Barabási, Albert-László & Brewer, Devon & Christakis, Nicholas & Contractor, Noshir & Fowler, James & Gutman, Myron & Jebara, Tony & King, Gary & Macy, Michael & Roy, Deb & Van Alstyne, Marshall (2009). *Computational Social Science*. Osoitteessa <http://nrs.harvard.edu/urn-3:HUL.InstRepos:4142693>. Luettu 25.10.2018.
- Lehto, Anna-Maija (1998). *Laatua surveytutkimukseen*. Teoksessa Paananen, Seppo & Juntto, Anneli & Sauli, Hannele (toim.) *Faktajuttu. Tilastollisen sosiaalitutkimuksen käytännöt*. Vastapaino. Tampere. 207–232.
- Leimu, Heikki (1983). *Työntekijäin työasema ja työpaikkaliikkuvuus erikokoisissa teollisuusyrityksissä. Osa 1: Teoreettiset lähtökohdat ja pienten sekä suurten teollisuusyritysten työntekijäin vertailua*. Turun yliopiston julkaisuja. Sarja C Scripta lingua Fennica edita 40. Turun yliopisto. Turku.
- Lepenes, Wolf (1988). *Between literature and science. The rise of Sociology*. Cambridge University Press. Cambridge.
- Leskinen, Esko (1987). *Faktorianalyysi: konfirmatoristen faktorimallien teoria ja rakentaminen*. Jyväskylän yliopiston tilastotieteen laitoksen julkaisuja; 1. Jyväskylän yliopisto. Jyväskylä.
- Leskinen, Esko (2005). *Pitkittäisaineistojen analyysistä latenttien muuttujien malleilla*. *Psykologia* 40:5–6. 476–488.
- Leskinen, Esko ja Kuusinen, Jorma (1991). *Faktorianalyysin käytöstä kasvatustieteellisessä tutkimuksessa*. *Suomen kasvatustieteellinen aikakauskirja Kasvatus* 22:4. 289–297.
- Lohr, Sharon, L. (2008). *Coverage and sampling*. Teoksessa de Leeuw, Edith D & Hox, Joop J & Dillman, Don A. (toim.) *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. New York. 97–112.
- Lundberg, George A. (1943). *Social Research. A Study in Methods of Gathering Data*. (Re-write of all but four chapters of Lundberg 1929.) Longmans, Green and CO. New York.
- Läärä, Esa & Lammi, Seppo (1986). *Tilastotieteen perusteet lääketiedettä ja lähialoja varten*. Kuopion yliopiston ylioppilaskunta. Kuopio.
- Läärä, Esa (1991). *Tilastomenetelmät kehittyvät – terävöityykö epidemiologinen analyysi?* *Sosiaalilääketieteellinen Aikakauslehti* 28. 404–422.
- Manninen, Pentti (1974). *Tilastotiedettä yhteiskuntatieteilijöille*. Gaudeamus. Helsinki.
- Marsh, Catherine (1982). *The Survey Method. The Contribution of Surveys to Sociological Explanation*. George Allen & Unwin. London.
- Marin, Marjatta (1970). *Tiedekunnat opintoympäristöinä. Tutkimus opiskelusta ja opiskelijoista Helsingin yliopistossa vuosina 1962–1968*. Helsinki.
- Markkanen, Touko (1964). *Sosiaaliset ja psykologiset faktorit väkijuomien käytön selittäjinä. Tutkimus faktoritulosten sisällöllisestä ryhmittämisestä ja vertailusta*. Alkoholipoliittisen tutkimuslaitoksen tutkimusseloste no. 12. Alkoholipoliittinen tutkimuslaitos. Helsinki.
- McElreath, Richard (2016). *Statistical Rethinking. A Bayesian Course with Examples in R and Stan*. CRP Press. Taylor & Francis Group. Boca Raton. London. New York.
- Miettunen, Jouko (2004). *Statistical methods in psychiatric research, with special reference on factor analysis*. Oulu University Press. Oulu.

- Mikkeli, Heikki & Pakkasvirta, Jussi (2007). Tieteiden välissä? Johdatus monitieteisyyteen, tieteidenvälisyyteen ja poikkitieteisyyteen. WSOY Oppimateriaalit. Helsinki.
- Molenaar, Ivo W. (1992). [The Impact of Sociological Methodology on Statistical Methodology]: Comment: The Fence Between Statistics and Social Research. *Statistical Science* 7:2, 203–205. doi:10.1214/ss/1177011358. Osoitteessa: <https://projecteuclid.org/euclid.ss/1177011358>. Luettu 12.4.2015.
- Moran, Joe (2002). *Interdisciplinarity*. Routledge. London and New York.
- MOTV (2018). Menetelmäopetuksen tietovaranto. Osoitteessa <https://www.fsd.uta.fi/menetelmaopetus/>. Luettu 6.11.2018.
- Mäkelä, Jukka (1991). Sunnuntaina sataa aina - tutkimus tilastollisen ajattelun siirtymisestä osaksi empiiristä sosiaalitutkimusta. Lapin yliopiston yhteiskuntatieteellisiä julkaisuja B. Tutkimusraportteja ja selvityksiä 13. Rovaniemi.
- Mäkelä, Jukka (1995). Tilastollisen päättelyn, kvantitatiivisen menetelmän ja sosiaalitutkimuksen suhteesta. Teoksessa Eskola, Jari & Mäkelä, Jukka & Suoranta, Juha (toim.) *Ihmistieteiden 1990-luvun metodologiaa etsimässä*. Keskustelua kasvatusta- ja sosiaalitieteiden 1990-luvun metodologiasta. Lapin yliopiston kasvatustieteellisiä julkaisuja C 8 Katsauksia ja puheenvuoroja. 59–73.
- Mäkelä, Jukka (1996). Menneisyyden nykyisyys. Kvalitatiivisen ja kvantitatiivisen tuolla puolen. *Acta Universitatis Lapponiensis* 13. Lapin yliopisto. Rovaniemi.
- Mäkelä, Pia (1999). Alkoholiin liittyvät kuolemat. *Stakes*. Helsinki.
- Mäkelä, Pia & Härkönen, Janne (2017). Suomalaisten alkoholinkulutus on vähentynyt: keiden ja millainen kulutus? *Yhteiskuntapolitiikka* 82:5. 513–526.
- Niemelä, Jukka (1996). Ammattirajoista tiimityöskentelyyn: Työnjaon ja työelämän suhteiden muutos Suomen telakoilla 1980- ja 1990-luvulla. Turun yliopisto. Turku.
- Niemi, Ilppo (1978). Organisaatio uskonnollisten yhteisöjen säilyttäjänä ja sopeuttajana. Helsingin yliopisto. Helsinki.
- Niiniluoto, Ilkka (2003). Totuuden rakastaminen. Tieteenfilosofisia esseitä. Otava. Helsinki.
- Niiniluoto, Ilkka & Koskinen, Heikki J. (2002). (toim.). *Wienin piiri*. Gaudeamus. Helsinki.
- Niskanen, Vesa A. (2004). Sumea logiikka. Kirkasta älyä ja mallinnusta. WSOY. Helsinki.
- Nummenmaa, Lauri (2004). Käyttäytymistieteiden tilastolliset menetelmät. Tammi. Helsinki.
- Nummenmaa Tapio (1980). Kokeellinen menetelmä psykologiassa. Tampereen yliopiston psykologian laitoksen tutkimuksia 121. Tampere.
- Nummenmaa, Tapio & Konttinen, Raimo & Kuusinen, Jorma & Leskinen, Esko (1997). *Tutkimusaineiston analyysi*. WSOY. Porvoo.
- Nyssölä, Kari (1994). Nuoret ja työmarkkinoiden joustavuus: Nuorten joustava työllistyminen työvoima- ja koulutuspoliittisena ongelmana. Turun yliopisto. Turku.
- Olkinuora, Erkki (1974). Norm socialization: The formation of personal norms. Jyväskylän yliopisto. Jyväskylä
- Oppenheim, A. N. (1983). *Questionnaire design and attitude measurement*. Heinemann. London.
- Orjasniemi, Tarja (1997). ”Viinakaupan jälkeeseen...”: panelitutkimus alkoholipoliittisen mielipide- ja asenneilmaston sekä alkoholin käytön kulttuurisen ankuroitumisen muutoksista Ranualla ja Posiolla 1980-luvulla. Yhteiskuntatieteiden tiedekunta, Lapin yliopisto. Rovaniemi.
- Orjasniemi, Tarja (2005). Ottaako vai ei.. raittiuden ja alkoholin käytön sukupolvittaiset muutokset moraalivallan murtumisen ilmentymänä pohjoisella maaseudulla 1980–2000-luvulla. *Acta Universitatis Lapponiensis*. Lapin yliopisto. Rovaniemi.
- Owen, D.B. & Minton, Paul D. & Pratt, John, W. (toim.) (1976). *On the History of Statistics and Probability*. Marcel Dekker, Inc. New York.

- Paananen, Seppo & Juntto, Anneli & Sauli, Hannele (1998) (toim.) Faktajuttu. Tilastollisen sosiaalitutkimuksen käytännöt. Vastapaino. Tampere.
- Partonen Monika (2004). Osoitteessa: http://www.fujitsu.fi/historia/oi_atk/seurasaaari.htm. Luettu 19.7.2017.
- Patty, John W. & Penn, Elizabeth Maggie (2015). Analyzing Big Data: Social Choice and Measurement. *PS, Political Science & Politics*; Washington 48:1. 95–101. Osoitteessa DOI:10.1017/S1049096514001814. Luettu 2.2.2020.
- Piela, Pasi (2015). Iso data tilastotoimessa. *Tieto & Trendit* 2. 23–27.
- Piepponen Paavo (1960). Harrastusten valinta. Tutkimus vapaa-ajan aktiivisuudesta ja henkilökohtaisista vaikutussuhteista. WSOY. Porvoo, Helsinki.
- Platt, Jennifer (1976). *Realities of Social Research. An empirical Study of British Sociologists.* Chatto and Windus for the Sussex University Press. Lontoo.
- Platt, Jennifer (1996). *A history of Sociological Research Methods in America.* Cambridge University Press. Cambridge.
- Porter, Theodore, M. (1986). *The Rise of Statistical Thinking 1820–1900.* Princeton University Press, Princeton.
- Porter, Theodore, M. (1995). *Trust in numbers. The pursuit of objectivity in science and public life.* Princeton University Press. Princeton.
- Purhonen, Semi & Toikka Arho (2016). ”Big datan” haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät. Esimerkkitapauksena aiheanalyysi tasavallan presidenttien uudenvuodenpuheista 1935–2015. *Sosiologia* 53:1. 6–27.
- Puuronen, Vesa (2005). Suomalaisen sosiologian paradigmoja 1960-luvulta nykypäivään. *Sosiologia* 42:1. 50–62.
- Puuronen, Vesa (2015). Tutkijan ja tutkimuskohteen suhteesta sosiologiassa. Teoksessa Salonen Toivo & Sotasaari, Seppo, I (toim.) *Ajatuksia tutkimiseen: metodisia lähtökohtia.* Lapin yliopisto. Rovaniemi. 108–144.
- Pöntinen, Seppo (2004). Sosiologian tutkimusmenetelmiä. Teoksessa: Kantola, Ismo & Koskinen, Keijo & Räsänen, Pekka (toim.) *Sosiologisia karttalehtiä.* 3. täysin uudistettu painos. Vastapaino. Tampere. 71–91.
- Pöntinen, Seppo (2011). Mitä mielekäs kvantitatiivisen sosiologian väitöskirjan tekeminen vaatii. Esitelmä Metodifestivaaleilla 30.8.2011. Tampere
- Raftery, Adrian E. (2000). Statistics in sociology, 1950–2000. *Journal of the American Statistical Association* 95:450. 654–661.
- Raftery, Adrian E. (2005). *Quantitative Research Methods.* Teoksessa Calhoun, Craig & Rojek, Chris & Turner, Bryan (eds.) *The SAGE Handbook of Sociology.* SAGE Publication Ltd. London. 15–39.
- Raiski, Seppo (1974a). Mittauksen ongelmista. *Sosiologia* 11:4. 186–192.
- Raiski, Seppo (1974b). Operationaalistamisen harhat. Tutkimus variaabelianalyysin tietoteoreettisesta asemasta. Tampereen yliopisto, sosiologian ja sosiaalipsykologian laitoksen tutkimuksia 7.
- Raiski Seppo (1976). Metodisia oppikirjoja. *Sosiologia* 13:1. 57–60.
- Randell, Seppo (1965). *Varusmiespalvelus ja asennemuutokset.* Yhteiskunnallinen korkeakoulu. Tampere.
- Ranta, Esa & Rita, Hannu & Kouki, Jari (1997). *Biometria. Tilastotiedettä ekologeille.* Kuudes painos. Yliopistopaino. Helsinki.
- Rantalaiho, Liisa (1968). *Sydäninfarktin sairastaneiden kuntoutuminen.* Tampereen yliopiston tutkimuslaitos. Tampere.

- Rantalaiho, Liisa (1972). *Metodi ja teoria*. *Sociologia* 9:1. 3–11.
- Rantalaiho, Liisa (1998). Paljaita faktoja? Teoksessa Paananen, Seppo & Juntto Anneli & Sauli, Hannele (toim.) *Faktajuttu. Tilastollisen sosiaalitutkimuksen käytännöt*. Vastapaino. Tampere. 11–21.
- Rao, J. N. K. (2006). *Interplay Between Sample Survey Theory and Practice: An Appraisal*. *Survey Methodology* 31:2. Statistics Canada, Catalogue No. 12–002. 117–138. Osoitteessa: <https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2005002/article/9040-eng.pdf?st=stLnquRd>. Luettu 28.8.2018.
- Rao, J. N. K. ja Fuller, W. A. (2017). *Samplesurvey theory and methods. Past, present, and future directions*. *Survey Methodology, Statistics Canada, Catalogue No. 12-001-X. Vol 43, No. 2*. Osoitteessa <http://www.statcan.gc.ca/pub/12-001-x/2017002/article/54888-eng.htm>. Luettu 28.8.2018.
- Raunio, Kyösti (1999). *Positivismi ja ihmistiede. Sosiaalitutkimuksen perustat ja käytännöt*. Gaudeamus. Helsinki.
- Rautio, Pertti (1985). *Tutkimuslaitoksen neljä vuosikymmentä*. Teoksessa Töttö, Pertti (toim.) *Ovi raollaan. Kurkistuksia yhteiskuntaan*. Tampereen yliopisto. Tampere. 151–164.
- Reunanen, Esa & Suikkanen, Risto (1998). *Latent Class Analysis: samoilua latentissa tilassa*. *Media & Viestintä* 21:4. 50–75.
- Riihinen, Olavi (1965). *Teollistuvan yhteiskunnan alueellinen erilaistuneisuus*. WSOY. Porvoo. Helsinki.
- Robertson, Hamish & Travaglia, Joanne (2017). *Big Data Sociology: Preparing for the Brave New World*. Osoitteessa <https://www.thesociologicalreview.com/blog/big-data-sociology-preparing-for-the-brave-new-world.html>. Luettu 25.10.2018.
- Rolin, Kristiina (2006). *Voiko soveltava yhteiskuntatiede olla arvovapaata?* Teoksessa Rolin, Kristiina & Kakkuri-Knuuttila, Marja-Liisa & Henttonen, Elina (toim.) *Soveltava yhteiskuntatiede ja filosofia*. Gaudeamus. Helsinki. 16–35.
- Rolin, Kristiina & Kakkuri-Knuuttila, Marja-Liisa & Henttonen, Elina (2006). *Johdanto: soveltava tutkimus ja tutkimuksen soveltaminen*. Teoksessa Rolin, Kristiina & Kakkuri-Knuuttila, Marja-Liisa & Henttonen, Elina (toim.) *Soveltava yhteiskuntatiede ja filosofia*. Gaudeamus. Helsinki. 7–15.
- Ronkainen, Suvi (1998). *Kaikuva empiirisuus – surveyn epistemologiset mahdollisuudet*. Teoksessa Paananen, Seppo & Juntto, Anneli & Sauli, Hannele (toim.) *Faktajuttu. Tilastollisen sosiaalitutkimuksen käytännöt*. Vastapaino. Tampere. 231–255.
- Ronkainen, Suvi (1999). *Ajan ja paikan merkitsemät. Subjektiviteetti, tieto ja toimijuus*. Gaudeamus. Helsinki.
- Ronkainen, Suvi (2001). *Monitieteellisyiden käytäntöjä*. Teoksessa Leskelä, Maarit (toim.) *Puheenvuoroja monitieteisyydestä. Kulttuurisen vuorovaikutuksen ja integraation tutkijakoulun julkaisuja 6*. Turun yliopisto. 9–35.
- Ronkainen, Suvi (2002). *Vallan metodit ja metodien valta*. *Sociologia* 39:2. 129–135.
- Ronkainen, Suvi (2014). *Goals, Tasks and Uses of Knowledge: Mixing Methodologies in Interdisciplinary Research*. Teoksessa Brusila, Riitta & Vanhanen, Hannu (toim.) *Integrated media in change*. Lapland University Press. Rovaniemi. 13–34.
- Ronkainen, Suvi & Karjalainen Anne (toim.) (2008). *Sähköä kyselyyn! Web-kysely tutkimuksessa ja tiedonkeruussa*. Lapin yliopiston menetelmätieteiden laitoksen tutkimuksia 1. Lapin yliopisto. Rovaniemi.
- Ronkainen, Suvi & Pehkonen, Leila & Lindblom-Ylänne, Sari & Paavilainen, Eija (2013). *Tutkimuksen voimasanat*. Sanoma Pro Oy. Helsinki.

- Rosenthal, Robert (1986). *Meta-analytic procedures for social research*. Sage publications. Beverly Hills.
- Ruskoaho, Juho & Vänskä, Jukka & Heikkilä, Teppo & Hyppölä, Harri & Halila, Hannu & Kujala, Santero & Virjo, Irma & Mattila, Kari (2010). Postitse vai sähköisesti? Näkemyksiä tiedonkeruun menetelmän valintaan *Lääkäri 2008 – tutkimuksen pohjalta*. *Sosiaalilääketieteellinen aikakauslehti* 47. 279–285.
- Räsänen, Pekka & Anttila, Anu-Hanna & Melin, Harri (2005). *Tutkimus menetelmien pyörteisä. PS-kustannus*. Jyväskylä.
- Rässler, Susanne & Rubin, Donald B. & Schenker, Nathaniel (2008). Incomplete data: Diagnosis, imputation and estimation. Teoksessa de Leeuw, Edith D & Hox, Joop J & Dillman, Don A. (toim.) *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. New York. 370–386.
- Räty Teuvo (1988). Professoritulpasta tutkijapulaan? *Suomen Akatemian julkaisuja* 3/1988. Helsinki.
- Räty Teuvo (1989). Väittelyiän kasvu sosiaali- ja valtiotieteissä – tehottomuutta vai riman nousua? *Sosiologia* 26:4. 286–297.
- Saari, Tiina & Melin, Harri (2020). Työn voimavarat ja organisaatioon sitoutuminen suomalaisyritysten yksiköissä Suomessa ja Venäjällä. *Työelämän tutkimus* 18:1. 44–62.
- Salonen, Toivo (2004). *Tieteenfilosofia. Lapin yliopiston menetelmätieteellisiä tutkimuksia nro 1*. Lapin Yliopistopaino. Rovaniemi.
- Sanaksenaho, Sanna (2006). *Eriarvoisuus ja luottamus 2000-luvun taitteen Suomessa: Bourdieulainen näkökulma*. Jyväskylän yliopisto. Jyväskylä.
- Sariola, Sakari (1956). *Sosiaalitutkimuksen menetelmät*. WSOY. Porvoo, Helsinki.
- Saris, Willem E. & Gallhofer, Irmaud N. (2014). *Design, Evaluation and Analysis of Questionnaires for Survey Research*. John Wiley & Sons. Hoboken, New Jersey.
- Selkälä, Arto (2013). *Kansalaisuuden hallinta suomalaisessa sosiaali- ja terveystieteissä*. Lapin yliopistokustannus. Rovaniemi.
- Selvin, Hanan C. (1958). Durkheim's *Suicide* and Problems of Empirical Research. *American Journal of Sociology* 63:6. 607–619. Osoitteessa: <http://www.jstor.org/stable/2772991>. Luettu 30.3.2019.
- Silén, Marianne (2008). Tilastotiede tutkimuksessa ja opetuksessa. *Suomen kasvatustieteellinen aikakauskirja Kasvatus* 39:2. 145–154.
- Silén, Marianne ja Ronkainen, Suvi (2013). The Effect of Data-Collection Methods in Studying the Finnish People's Attitudes toward Sexual Violence. Osoitteessa http://www.esa11thconference.eu/skins/default/static/ESA_2013_Abstract_Book.pdf. Luettu 21.7.2018.
- Simpura, Jussi ja Melkas, Jussi (2013). *Tilastot käyttöön! Opas tilastojen maailmaan*. Gaudeamus. Helsinki.
- Sintonen, Matti (2005). Tieteiden välisyys: suhteellisen itsenäiset tieteet – ja mitä on olla niiden välissä? Teoksessa Rydman, Jan (toim.) *Suhteellista? Yliopistopaino*. Tieteellisten seurain valtuuskunta. Helsinki. 251–265.
- Snow, C. P. (1998). *Kaksi kulttuuria*. Johdanto Stefan Collini. Suomentanut Kimmo Pietiläinen. Terra Cognita.
- Sobel, Michael, E. (1995) *Causal Inference in the Social and Behavioral Sciences*. Teoksessa Arminger, Gerhard & Clogg, Clifford C. & Sobel, Michael (toim.) *Handbook of Statistical Modelling for the Social and Behavioral Sciences*. Plenum Press. New York and London. 1–38.
- Stapleton, Laura M. (2008). Analysis of data from complex surveys. Teoksessa de Leeuw, Edith D. & Hox, Joop J. & Dillman & Don A. (toim.) *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates. New York. 342–369.

- Star, Susan Leigh ja Griesemer, James R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907 – 39. *Social Studies of Science* 19:3. 387–420.
- Star, Susan Leigh (2010). This is not a boundary-object. Reflections on the origin of the concept. *Science, technology, and human values* 35:5. 601–607.
- Stolte-Heiskanen Veronica (1967). Social Structure. Family and Interpersonal Influence. *Transactions of the Westermarck Society* 14. Helsinki.
- Stolte-Heiskanen Veronica (1992). Sosiologia vakiintuneena tieteenä. Teoksessa Alapuro, Risto & Alestalo, Matti & Haavio-Mannila Elina (toim.) *Suomalaisen sosiologian historia*. Porvoo-Helsinki-Juva. 287–292.
- Stoop, Ineke, A. L. (2005). The hunt for the last Respondent. Nonresponse in sample surveys. Osoitteessa: https://www.scp.nl/english/Publications/Publications_by_year/Publications_2005/The_Hunt_for_the_Last_Respondent. Luettu 6.11.2018.
- Sulkunen, Pekka (1972). Surveyanalyysistä ja teoreettisesta mielivallasta. *Sosiologia* 9:4. 157–166.
- Sund, Reijo (2000). Mitä on tilastotiede? *Tyypiarvo* 4/2000. Osoitteessa http://www.helsinki.fi/jarj/moodi/taas_tilastotiede.html. Luettu 31.5.2012.
- Sund, Reijo (2005). Terveystieteenhuoltojärjestelmän vaikuttavuuden arvioinnista. *Yhteiskuntapolitiikka* 70:5. 507–513.
- Sund, Reijo (2008). Methodological Perspectives for Register-Based Health System Performance Assessment. Developing a Hip Fracture Monitoring System in Finland. *Stakes*. Helsinki.
- Sund, Reijo (2015). Miksi isoon dataan hukutaan? *Tieto & Trendit* 2. 40–45.
- Sund, Reijo & Nylander, Olli & Palonen, Tuula (2004). Raa'asta rekisteriaineistosta terveystieteellisesti relevanttiin informaatioon. *Yhteiskuntapolitiikka* 69:4. 372–379.
- Suominen, Jaakko (2000). Osoitteessa <http://www.tuug.fi/~jaakko/tutkimus/kronologia.html>. Luettu 18.7.2017.
- Suoranta, Juha (1995). Metodologia muukalaisuutena Teoksessa Eskola, Jari & Mäkelä, Jukka & Suoranta, Juha (toim.) *Ihmistieteiden 1990-luvun metodologiaa etsimässä*. Keskustelua kasvatustieteiden ja sosiaalitieteiden 1990-luvun metodologiasta. *Lapin yliopiston kasvatustieteellisiä julkaisuja C 8 Katsauksia ja puheenvuoroja*. 27–40.
- Suoranta, Juha & Eskola, Jari (1995). Empiirinen tutkimus teoreettisen tutkimuksen vaiheista. Teoksessa Eskola, Jari & Mäkelä, Jukka & Suoranta, Juha (toim.) *Ihmistieteiden 1990-luvun metodologiaa etsimässä*. Keskustelua kasvatustieteiden ja sosiaalitieteiden 1990-luvun metodologiasta. *Lapin yliopiston kasvatustieteellisiä julkaisuja C 8 Katsauksia ja puheenvuoroja*. 185–199.
- Sänkiäho, Risto (1974). Tempu ja kuinka ne tehdään. Monimuuttujamenetelmät Kansan palvelijoina. 4. uudistettu painos. *Kasvatustieteiden tutkimuslaitos*. Jyväskylä.
- Taanila, Aki (2019). Kato. Akin menetelmäblogi. Osoitteessa: <https://tilastoapu.wordpress.com/tag/vastausprosentti/>. Luettu 30.7.2019.
- Taanila, Aki (2020). Korrelaatio – lisätarkastelu. Akin menetelmäblogi. Osoitteessa: <https://tilastoapu.wordpress.com/tag/kovarianssi/>. Luettu 8.1.2020.
- Tacq, Jacques (1997). *Multivariate analysis techniques in social science research. From problem to analysis*. Sage Publications. London.
- Takala, Hannu (1973). Miten faktoripistemäärillä arvioidaan sopeutumattoman nuoren huollontarve, kun isä juo ja äiti huoraa. *Sosiologia* 10:2, 94–95.
- Talikka, Annikki (1969). Frustraatiotilanteet ja demokratia: Sosiologinen tutkimus frustraatiotilanteista sisäpoliittisten konfliktien strategioiden ennustajana. Tampere.
- Tarkkonen, Lauri (1997). Lauri Tarkkonen haastattelu. *Tyypiarvo* 4/97. Osoitteessa http://www.helsinki.fi/jarj/moodi/taha_tarkkonen.html. Luettu 25.3.2018

- Tarkkonen Lauri (2007). 40 years of multivariate analysis. Teoksessa Niemelä, Jarmo & Puntanen, Simo & Liski, Erkki P. (toim.) Abstracts of the Annual Conference of Finnish Statisticians 2007, "Multivariate Methods". Report A 375, Dept. of Mathematics, Statistics and Philosophy, University of Tampere. 12–16.
- Tarkkonen, Lauri & Vehkalahti, Kimmo (2005). Measurement errors in multivariate measurement scales. *Journal of Multivariate Analysis* 96:1. 172–189. Osoitteessa: <https://doi.org/10.1016/j.jmva.2004.09.007>. Luettu 31.1.2020.
- Tauriainen, Juhani (1970). Kehitysalueiden muuttuva maatalous. Tutkimus Itä-, Sisä- ja Pohjois-Suomen maatalouden uudenaikaistumisesta. Pellervo-Seuran Markkinatutkimuslaitoksen julkaisuja 1. Helsinki.
- Teddlie, Charles & Tashakkori, Abbas (2009). *Foundations of Mixed Methods Research. Integrating Quantitative and Qualitative in the Social and Behavioral Sciences*. SAGE. Los Angeles.
- Thurstone, Louis, L. (1935). *The Vectors of Mind: Multiple-factor analysis for the isolation of primary traits*. University of Chicago Press. Osoitteessa: <https://doi.org/10.1037/10018-000>. Luettu 6.6.2017.
- Thurstone, Louis, L. (1947). *Multiple-factor analysis; a development and expansion of The Vectors of Mind*. University of Chicago Press.
- Tilastokeskus (2018). Osoitteessa http://www.tilastokoulu.stat.fi/verkkokoulu_v2.xql?page_type=esim&course_id=tkoulu_tlkt&lesson_id=3&subject_id=4&example_id=1. Luettu 6.8.2018.
- Toiviainen, Seppo (1969). Lauseita sosiologiasta. *Sociologia* 6:2. 84–85.
- Toiviainen, Seppo (1970). Yhteiskunnalliset ja kulttuuriset ristiriidat – musikologisten osakulttuurien sosiologista tarkastelua. *Acta Universitatis Tamperensis ser. A vol. 43*. Tampereen yliopisto. Tampere.
- Toiviainen, Seppo (1998). Sosiaalisen elämän alkeisluvut. Teoksessa Paananen, Seppo & Junto, Anneli & Sauli, Hannele (toim.) *Faktajuttu. Tilastollisen sosiaalitutkimuksen käytännöt*. Vastapaino. Tampere. 313–335.
- Toivonen, Timo (1999). Empiirinen sosiaalitutkimus. *Filosofia ja metodologia*. WSOY. Porvoo.
- Traag, Vincent & Franssen, Thomas (2016). Revealing the quantitative-qualitative divide in sociology using bibliometric visualization. Osoitteessa: <https://www.cwts.nl/blog?article=n-q2v294&title=revealing-the-quantitative-qualitative-divide-in-sociology-using-bibliometric-visualization>. Luettu 18.1.2020.
- Töttö Pertti (1997a). Auguste Comte – positivismin isä. Teoksessa Gronow, Jukka & Noro, Arto & Töttö Pertti. *Sociologian klassikot*. Gaudeamus. Helsinki. 61–88.
- Töttö Pertti (1997b). Pirullinen positivismi. Kysymyksiä laadulliselle tutkimukselle. *Kampus Kustannus, Jyväskylä*.
- Töttö, Pertti (2000a). Kuvitteellisesta perusjoukosta. *Sociologia* 37:2. 148–151.
- Töttö, Pertti (2000b). Pirullisen positivismin paluu. Laadullisen ja määrällisen tarkastelua. *Vastapaino*. Tampere.
- Töttö, Pertti (2004). Syvällistä ja pinnallista. Teoria, empiria ja kausaalisuus sosiaalitutkimuksessa. *Vastapaino*. Tampere.
- Töttö, Pertti (2012). Paljonko on paljon? Luvuilla argumentoinnista empiirisessä tutkimuksessa. *Vastapaino*. Tampere.
- Uitto, Yrjö (1964). Kaupan sosiaaliset arvot ja roolit. Vähittäiskaupan itsenäisen ja palkatun ammatin sosiologinen vertailu. Helsinki.
- Ullman, Jodie B. & Bentler, Peter M. (2004). *Structural Equation Modeling*. Teoksessa Hardy, Melissa & Bryman, Alan (toim.) *Handbook of data analysis*. SAGE. London. 431–458.

- Vahervuo, Toivo & Ahmavaara, Yrjö (1958). Johdatus faktorianalyysiin. WSOY. Porvoo.
- Valkonen, Tapani (1981). Haastattelu- ja kyselyaineiston analyysi sosiaalitutkimuksessa. Seitsemäs painos. Gaudeamus. Helsinki.
- Varto, Juha (1995). Keskustelu ihmistieteiden metodologiasta 1900-luvulla. Teoksessa Eskola, Jari & Mäkelä, Jukka & Suoranta, Juha (toim.) Ihmistieteiden 1990-luvun metodologiaa etsimässä. Keskustelua kasvatus- ja sosiaalitieteiden 1990-luvun metodologiasta. Lapin yliopiston kasvatustieteellisiä julkaisuja C 8 Katsauksia ja puheenvuoroja. 13–25.
- Vasama, Pyry-Matti & Vartia, Yrjö (1980). Johdatus tilastotieteeseen. Osa I. Neljäs korjattu painos. Gaudeamus. Helsinki.
- Vehkalahti, Kimmo (2002). Monimuuttujamenetelmät. Osoitteessa <http://www.helsinki.fi/~kvehkala/mmm/moniste.pdf>. Luettu 28.8.2018.
- Vuorijärvi, Petri (2014). Moniulotteisen korrespondenssianalyysin soveltamisesta sosiaalitieteissä. Osoitteessa <http://urn.fi/URN:NBN:fi:ula-201511061357>. Luettu 28.8.2018.
- Wasserstein, Ronald L. & Schirm, Allen L. & Lazar, Nicole A. (2019). Moving to a World Beyond “ $p < 0.05$ ”. *The American Statistician* 73:1. 1–19. Osoitteessa <https://doi.org/10.1080/00031305.2019.1583913>. Luettu 30.12.2019.
- Williams, Brett & Onsmann, Andry & Brown, Ted (2010). Exploratory factor analysis: A five-step guide for novices. *Journal of Emergency Primary Health Care (JEPHC)* 8:3. 1–13.
- Winship, Christopher & Sobel, Michael (2004). Causal Inference in Sociological Studies. Teoksessa Hardy, Melissa & Bryman, Alan (toim.) *Handbook of data analysis*. SAGE. London. 481–504.
- Wolf, Fredric M. (1986). *Meta-analysis. Quantitative methods for research synthesis*. Sage publications. Beverly Hills.
- Ylikoski, Petri (2000). Bruno Latour ja tieteen tutkimus. *Tiede & Edistys* 25:4. 296–310. Osoitteessa: <http://urn.fi/URN:NBN:fi:ELE-1487691>. Luettu 13.6.2015.
- Yong, An Gie & Pearce, Sean (2013). A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology* 9:2. 79–94. Osoitteessa: <http://www.tqmp.org/RegularArticles/vol09-2/p079/>. Luettu 12.5.2020.

LIITE 1 Bibliografia: suomalaiset sosiologian väitöskirjat 1960–2018.

Mitä tahansa tilastollisia menetelmiä sisältävät väitöskirjat esitetään keltaisella taustalla. Monimuuttujamenetelmiä sisältävät väitöskirjat on lihavoitu. Yliopistot on lyhennetty seuraavasti:

HY = Helsingin yliopisto

JoY = Joensuun yliopisto

JY = Jyväskylän yliopisto

LY = Lapin yliopisto

KY = Kuopion yliopisto

TaY = Tampereen yliopisto

TY = Turun yliopisto

UEF = Itä-Suomen yliopisto

ÅA = Åbo Academi

| | | |
|----------------------------------|---|------------|
| 1960 | | |
| Jyrkilä Vieno (Faina) | Society and adjustment to old age. | TaY |
| Piepponen Paavo | Harrastusten valinta. | HY |
| 1961 | | |
| Eskola Antti | Social influence and power in two-person groups. | HY |
| Vehmas Raimo | Suomen romaniväestön ryhmäluonne ja akkulturoituminen. | TaY |
| 1962 | | |
| Laaksonen Oiva | Suomen liike-elämän johtajisto. | HY |
| Luoma Väinö | Valtion virkamiesten järjestäytyminen Suomessa. | TY |
| Piha Kalevi | Asuntoyhteisö, sen muodostuminen ja rakenne. | TY |
| 1963 | | |
| Haavio Ari | Evankelinen liike. | TY |
| 1964 | | |
| Uitto Yrjö | Kaupan sosiaaliset arvot ja roolit. | HY |
| 1965 | | |
| Asp Erkki | Lappalaiset ja lappalaisuus. | TY |
| Randell Seppo | Varusmiespalvelus ja asennemuutokset. | TaY |
| 1966 | | |
| Hirvas Juhani | Identity and mental illness. | HY |
| Koiranen Vilho A. | Suomalaisten siirtolaisten sulautuminen Ruotsissa. | HY |
| Kämäräinen Kauko | Kaupunkijärjestelmän alueellinen eriytyminen ja poikkeavuus. | TaY |
| 1967 | | |
| Stolte-Heiskanen Veronica | Social structure, family and interpersonal influence. | HY |
| 1968 | | |
| Koskelainen Osmo | Maallistuva suurkaupunki. | HY |
| Rantalaiho Liisa | Sydäninfarktin sairastaneiden kuntoutuminen. | TaY |
| Uusitalo Paavo | Vankila ja työsiirtola rangaistuksena. | HY |
| 1969 | | |
| Kalimo Esko | Lääkintäpalvelusten käyttöön vaikuttavat tekijät. | HY |
| Kiviranta Pekka | Alcoholism syndrome in Finland. | HY |
| Kolari Risto | Über ideologische und nationale Werte. Eine typologische Analyse mit ergänzender Betrachtung des Isomorphieproblems. | HY |
| Talikka Annikki | Frustraatiotilanteet ja demokratia. | TaY |
| 1970 | | |
| Blom Raimo | Luottamus oikeuslaitukseen. | TaY |
| Marin Marjatta | Tiedekunnat opintoympäristöinä. | HY |
| Nurmi Raimo | Conformity with group in a serial judgement situation. | TY |
| Salonen Erkki | Suomalaisen kulttuurin säätelyn järjestelmä. | HY |
| Tauriainen Juhani | Kehitysalueiden muuttuva maatalous. | HY |
| Toiviainen Seppo | Yhteiskunnalliset ja kulttuuriset ristiriidat – musikologisten osakulttuurien sosiologista tarkastelua. | TaY |
| Valkonen Tapani | Yhteisön vaikutukset yksilöön ja niiden kvantitatiivinen tutkiminen. | HY |
| 1971 | | |
| Honkala Kauko | Sosiaalinen eriarvoisuus maaseudulla. | ÅÅ |
| Koskinen Tarmo | Jokapäiväinen leipämme. | TaY |
| Mäkelä Klaus | Alkoholitutkimus ja alkoholipolitiikka. | HY |

| | | |
|------------------------------|--|------------|
| 1972 | | |
| Lampikoski Kari | Ammattiyhdistys ja ammattikunnan status. Sosiologinen tutkimus ammattiyhdistyksen toiminnasta ammattikuntana statuksen kohottamiseksi organisaatiossa. Tutkimustapauksena Suomen konepäällystöliiton toiminta merikonemestarien statuksen kohottamiseksi laivaorganisaatiossa. | TY |
| Pietilä Veikko | Lehdistökirjoittelu ja mielipiteenmuodostus | TaY |
| Toivonen Timo | Lakot ja yhteiskunnan muutos. | TY |
| 1973 | | |
| Alapuro Risto | Akateeminen Karjala-Seura. Ylioppilasliike ja kansa 1920- ja 1930-luvuilla. | HY |
| 1974 | | |
| Elovainio Päivi | Korkeakoululaitoksen rakenne ja yhteiskunnan muutos. | HY |
| Olkinuora Erkki | Norm socialization. | HY |
| 1975 | | |
| Aalto Ritva | Nuorten sosiaalistuminen ja sosiaalityö. | HY |
| Ahlström-Laakso Salme | Drinking Habits among Alcoholics. | HY |
| Suolinna Kirsti | Uskonnollisten liikkeiden asema sosiaalisessa muutoksessa. | HY |
| Uusitalo Hannu | Income and welfare. | HY |
| Sandlund Tom | Social Classes, Ethnic Groups and Capitalist Development. An Outline of a Theory. | ÅÅ |
| 1977 | | |
| Laitinen Ahti | Lääkkeiden käyttö yhteiskunnallisena kysymyksenä. Sosiologinen tutkimus lääkkeiden käytön yhteiskunnallisista yhteyksistä, erityisesti silmällä pitäen lääkkeiden mainontaa ja tuotannon, mainonnan ja käytön valvontaa. | TY |
| 1978 | | |
| Niemi Ilppo | Organisaatio uskonnollisten yhteisöjen säilyttäjänä ja sopeuttajana. | HY |
| 1979 | | |
| Eskola Katariina | Suomalaiset kirjanlukijoina. | HY |
| Järvikoski Timo | Vesien säännöstely ja paikallisyhteisö. Tutkimus valtakunnallisen päätöksenteon seurauksista esimerkkitaupauksena Lokan ja Porttipahdan tekojärvet. | TY |
| Köppä Tapani | Viljelijäperhe, yhteisö ja yhteistoiminta. Tutkimus maatalojen välisen yhteistoiminnan kehitykseen ja omaksumiseen vaikuttavista sosiaalisista ja taloudellisista tekijöistä. | TY |
| Parjanen Matti | Learning, earning and withdrawing. | TaY |
| Vuorela Pertti | Kaupunkimaan omistus ja maankäyttörakenne Suomessa. | HY |
| Vuorinen Juha | Korkeakoulujen itsehallinto. Tutkimus Suomen korkeakoulujen itsehallinnon toteutumisesta suhteessa korkeakoululaitoksen keskushallintoon toisen maailmansodan jälkeen. | TY |
| 1980 | | |
| Miemois Karl-Johan | Språk, kultur och service: undersökningar av finlandssvenskarnas predikament. | HY |
| Sulkunen Pekka | Alkoholin kulutus ja elinolojen muutos toisen maailmansodan jälkeen. | HY |
| 1981 | | |
| Mäki Martti | Alkoholijuomien käyttö tieliikenteen turvallisuusongelmana. | HY |
| 1982 | | |

| | | |
|-----------------------|---|-----------|
| Kasvio Antti | Teollisuustyö ja elämäntapa. Tutkimus kahden suuren teollisuuslaitoksen ammattitaitoisten työntekijöiden työn ja työn ulkopuolisen elämän kehityspiirteistä 1980-luvun suomalaisessa yhteiskunnassa. | TaY |
| Moilanen Paavo | Social environment and adaptation. | TY |
| 1983 | | |
| Jallinoja Riitta | Suomalaisen naisasialiikkeen taistelukaudet: naisasialiike naisten elämäntilanteen muutoksen ja yhteiskunnallis-aatteellisen murroksen heijastajana. | HY |
| Leimu Heikki | Työntekijän asema ja työpaikkaliikkuvuus erikokoisissa teollisuusyrityksissä. | TY |
| Pöntinen Seppo | Social Mobility and Social Structure. A comparison of Scandinavian countries. | HY |
| Takala Tuomas | Oppivelvollisuuskoulu ja yhteiskunnalliset intressit. Tutkimus kvalifikaatioihin ja koulutuskustannuksiin kohdistuvista intresseistä Suomen oppivelvollisuuskoulua koskevan koulutuspoliittisen päätöksenteon määreinä. | TaY |
| 1984 | | |
| Ilmonen Kaj | Jäsenet ja jäsenten liike. Tutkimus kaupan ja jäsenten suhteesta E-osuuskauppaliikkeessä osuustoiminnan alkuajoilta 1980 - luvulle. | TaY |
| Jaakkola Magdalena | Siirtolaiselämää. Tutkimus ruotsinsuomalaisista siirtolaisyhteisönä. | HY |
| 1986 | | |
| Alestalo Matti | Structural Change, Classes and the State. Finland in a Historical and Comparative Perspective. | HY |
| Gronow Jukka | On the Formation of marxism. Karl Kautsky's Theory of Capitalism, the Marxism of the Second International and Karl Marx's Critique of Political Economy. | HY |
| Lipén Keijo | Työvoimatoimistot työvoimapolitiikan toteuttajina. | JY |
| Saralehto Sampsa | Teknologian kansainvälinen siirto kehitysmaiden teollistumisprosessiin. | TY |
| Siisiäinen Martti | Intressit, yhdistyslaitos ja poliittisen järjestelmän vakaisuus. | JY |
| 1987 | | |
| Kosonen Pekka | Hyvinvointivaltioiden haasteet ja pohjoismaiset mallit. | HY |
| Moilanen Olavi | Elämän puhuteltavana. Tutkimus ahdistuksen käsittelystä herännäisyyden alkuvaiheissa ja tämän vuosisadan Kainuussa. | JY |
| 1988 | | |
| Holmila Marja | Wives, husbands and alcohol. | HY |
| Kivinen Osmo | Koulutuksen järjestelmäkehitys: Peruskoulutus ja valtiollinen kouludoktriini Suomessa 1800- ja 1900-luvuilla. | TY |
| Lillrank Paul | Organization for Continuous Improvement. | HY |
| Sabour M'hammed | Homo Academicus Arabicus. | JoY |
| 1989 | | |
| Granberg Leo | Valtio maataloustulojen tasaajana ja takaajana. | HY |
| Kivinen Markku | The New Middle Classes and the Labour Process: Class Criteria Revisited. | HY |
| Kortteinen Timo | Agricultural Alcohol and Social Change in the Third World. | HY |
| Lahelma Eero | Unemployment, re-employment and mental well-being: a panel survey of industrial job-seekers in Finland. | HY |

| | | |
|------------------------|---|-----------|
| Notkola Irma-Leena | Luonnollisesta hedelmällisyydestä lapsirajoitukseen. Mikrodemograafinen tutkimus hedelmällisyyden transitiosta. | HY |
| Rannikko Pertti | Metsätyö - pienviljelijäkylä. Tutkimus erään yhdyskuntatyyppin noususta ja tuhosta. | JoY |
| Urponen Helka | Varhain sairastuneen ja vammaisen sosiaalinen selviytyminen. | TY |
| 1990 | | |
| Alasoini Tuomo | Tuotannolliset rationalisoinnit ja teollisuuden työvoiman käyttötapojen muutos. | HY |
| Alasuutari Pertti | Desire and craving: studies in a cultural theory of alcoholism. | TaY |
| Alkula Tapani | Work orientations in Finland. | HY |
| Heikkilä Matti | Köyhyys ja huono-onnisuus hyvinvointivaltiossa. | HY |
| Järvinen Margaretha | Prostitution i Helsingfors. En studie i kvinnokontroll. | ÅA |
| Kaukonen Erkki | Theory, Dynamics and Policy of Science. Science Studies from a Small Country Perspective. | TaY |
| Luukkonen Terttu | Citations in the Rhetorical, Reward, and Communication Systems of Science. | TaY |
| Raiski Seppo | Laboratorio, teoria ja parempi elämä: "modernin" sosiologiaa episteemisen historian edellytyksiltä. | TaY |
| Siiskonen Pirjo | Emännän ja isännän roolin muutos maatalouden uudenaikaistuksessa. | HY |
| 1991 | | |
| Alanen Ilkka | Miten teoretisoida maatalouden pientuotantoa. | JY |
| Alestalo Marja | Science and the Politico-Economic System. Social Change, Transformation of Political Structures, and the Social Value of Science. | HY |
| Kangas Olli | The politics of social rights: studies on the dimensions of sickness insurance in OECD countries. | HY |
| K'Okul Richard N.O. | Maternal and child health in Kenya. A study of poverty, disease and malnutrition in Samia. | HY |
| Konttinen Esa | Perinteisesti moderniin. Profiessoiden yhteiskunnallinen synty Suomessa. | JY |
| Kontula Osmo | Sukupuolielämän aloittamisen yhteiskunnallisista ehdoista. | HY |
| Murto Kari | Towards the well functioning community. The development of Anton Makarenko and Maxwell Jones' communities. | JY |
| Noro Arto | Muoto, moderniteetti ja 'kolmas'. Tutkielma Georg Simmelin sosiologiasta. | HY |
| Partanen Juha | Sociability and intoxication. Alcohol and drinking in Kenya, Africa, and the modern world. | HY |
| Salmi Minna | Ansio työ kotona - toiveuni vai painajainen? Kotiansio työ Suomessa työntekijän arkipäivän kannalta. | HY |
| Sundback Susan | Utträdet ur Finlands lutherska kyrka. | ÅA |
| Töttö Pertti | Werner Sombart ja kiista kapitalismin hengestä. | TaY |
| 1992 | | |
| Alanen Leena | Modern childhood? Exploring the 'child question' in sociology. | JY |
| Ehrnrooth Jari | Sanan vallassa, vihan voimalla. Sosialistiset vallankumousopit ja niiden vaikutus Suomen työväenliikkeessä 1905–1914. | HY |
| Falk Pasi | The corporeality of consumption. | HY |
| Hakkarainen Pekka | Suomalainen huumeekysymys: huumeaineiden yhteiskunnallinen paikka Suomessa toisen maailmansodan jälkeen. | TY |

| | | |
|-----------------------|--|-----------|
| Kauppinen Timo | Suomen työmarkkinamallin muutos. | HY |
| Kortteinen Matti | Kunnian kenttä: suomalainen palkkatyö kulttuurisena muotona. | HY |
| Saarinen Aino | Feminist research - an intellectual adventure? A research autobiography and reflections on the development, state and strategies of change of feminist research. | OY |
| Silius Harriet | Den kringgårdade kvinnligheten. Att vara kvinnlig jurist i Finland. | ÅÅ |
| 1993 | | |
| Hoikkala Tommi | Katoaako kasvatus, himmeneekö aikuisuus? Aikuistumisen puhe ja kulttuurimallit. | HY |
| Hytti Helka | Työkyvyttömyyseläkkeelle siirtymisen yhteiskunnalliset taustatekijät. | HY |
| Kaartinen Timo | Kuinka kulttuuria käännetään: arvon tuottaminen ja antropologinen käytäntö. | HY |
| Koski Leena | Tieteen tahtomana, yliopiston tekemänä: yliopiston sisäiset symboliset järjestykset. | JoY |
| Kyllönen Timo | The golden handshake of the first world: development research, developmentalism and take-off into development studies. | HY |
| Kyntäjä Timo | Tulopoliitikka Suomessa: tulopoliittinen diskurssi ja instituutiot 1960- luvulta 1990-luvun kynnykselle. | TY |
| Mannila Simo | Työhistoria ja syrjäytyminen: vaikeasti työllistyvien terveysongelmaisten elämäkulusta. | HY |
| Mjelde Liv | Apprenticeship. From practice to Theory and Back Again. | JoY |
| Mäki-Kulmala Airi | Initiaatio ja alakulttuuri. | TaY |
| Mäkinen Heikki | Yhteiskunnan tieto. | TaY |
| Pirttilä Ilkka | Me ja maailman mallit: tiedonsosiologian ydintä etsimässä. | JoY |
| Ruonavaara Hannu | Omat kodit ja vuokrahuoneet: sosiologinen tutkimus asunnonhallinnan muodoista Suomen asutuskeskuksissa 1920–1950. | TY |
| Salminen Kari | Pension Schemes in the Making. A Comparative Study of the Scandinavian Countries. | HY |
| Arosalo Sirkka | Poliittisen väkivallan yhteiskunnallisista edellytyksistä: punainen ja valkoinen väkivalta Suomessa vuonna 1918. | TaY |
| Eklund Erland | Kustfiskare och kustfiske i Finland under den industriella epoken. Studier i en yrkesgrupps yttre villkor, sociala skiktning och organisation 1860–1970. | ÅÅ |
| Kivivuori Janne | Understand and punish: the culture of juvenile transgressions in the context of a welfare state. | HY |
| 1994 | | |
| Martelin Tuija | Differential Mortality at Older Ages: Sociodemographic Mortality Differences Among the Finnish Elderly. | HY |
| Nyysölä Kari | Nuoret ja työmarkkinoiden joustavuus: nuorten joustava työllistyminen työvoima- ja koulutuspoliittisena ongelmana. | TY |
| Siurala Lasse | Nuoriso-ongelmat modernisaatioperspektiivissä. | HY |
| Strandell Harriet | Sociala mötesplatser för barn: aktivitetsprofiler och förhållingskulturer på daghem. | HY |
| Uusitalo Eero | Maaseutupoliitiikan keinot: elinkeinojen edistäminen maaseudun kehittäjäyhteisöissä. | TY |
| Vikat Andres | Family Formation in Estonia. | HY |
| 1995 | | |

| | | |
|------------------------|---|-----------|
| Häkkinen Antti | Rahasta - vaan ei rakkaudesta: prostituutio Helsingissä 1867–1939. | HY |
| Jokinen Pekka | Tuotannon muutokset ja ympäristöpolitiikka. Ympäristösosiologinen tutkimus suomalaisesta maatalouden ympäristöpolitiikasta vuosina 1970–1994. | TY |
| Jyrkämä Jyrki | "Rauhaisesti alas illan lepoon"? Tutkimus vanhenemisen sosiaalisuudesta neljässä paikallisyhteisössä. | TaY |
| Näre Sari | Etnopsykoanalyttisia näkökulmia sukupuolikulttuuriin. | HY |
| Peltola Pekka | The Lost May Day: Namibian Workers Struggle for Independence. | HY |
| Puuronen Vesa | Nuoret maailmansa tuottajina? Tutkimus nuorisososiologian metodologiasta ja nuorten keskusteluista. | JoY |
| 1996 | | |
| Itkonen Hannu | Kenttien kutsu: tutkimus liikuntakulttuurin muutoksesta. | JoY |
| Joutsenoja Markku | The calling of sociology: early Talcott Parsons and the construction of a disciplinary ship. | LY |
| Kerkelä Heikki | Vanhan maailman peilissä: modernin yhteiskunnan synty ja pohjoinen aineisto. | LY |
| Lagerspetz Mikko | Constructing Post-communism. A Study in the Estonian Social Problems Discourse. | TY |
| Lahti Vesa-Matti | Riskiyhteiskunta tuli kylään: sosiologinen tutkimus vesijohtoveden saastumisen seurauksista ihmisten elämässä. | HY |
| Melin Harri | Suunnitelman varjossa: tutkimus yritysjohtajista Neuvostoliitossa ja Venäjällä. | TaY |
| Mäkelä Jukka | Menneisyyden nykyisyys: kvalitatiivisen ja kvantitatiivisen tuolla puolen. | LY |
| Niemelä Jukka | Ammattirajoista tiimityöskentelyyn: työnjaon ja työelämän suhteiden muutos Suomen telakoilla 1980- ja 1990-luvulla. | TY |
| Nurmela Juha | Kotitaloudet ja energia vuonna 2015: tutkimus kotitalouksien rakennemuutoksen vaikutuksesta energiankulutukseen. | HY |
| Salo Markku | Sietämisestä solidaarisuuteen: mielisairaalariformit Italiassa ja Suomessa. | TaY |
| Tapaninen Anna-Maria | Kansan kodit ja kaupungin kadut: etnografinen tutkimus eteläitalialaisesta kaupungista. | HY |
| 1997 | | |
| Assmuth Laura | Women's Work, Women's Worth: Changing Lifecourses in Highland Sardinia. | HY |
| Gould Jeremy | Localizing Modernity: Action, Interests and Association in Rural Zambia. | HY |
| Heiskala Risto | Society as semiosis: neostructuralist theory of culture and society. | HY |
| Helen Ilpo | Äidin elämän politiikka: naissukupuolisuus, valta ja itsetuho Suomessa 1880-luvulta 1960-luvulle. | HY |
| Jerman Helena | Between Five Lines. The Development of Ethnicity in Tanzania with Special Reference to the Western Bagamoyo District. | HY |
| Jokinen Kimmo | Suomalaisen lukemisen maisemahanteet. | JY |
| Kandolin Irja | Gender, worklife and family responsibilities in Finland and Estonia. | HY |
| Kangas Ilka | Naisten vaihdevuosien ja vanhenemisen tulkinnat. | HY |
| Karvonen Sakari | The regional context of health behaviour among Finnish adolescents. | HY |

| | | |
|------------------------------|--|-----------|
| Lavikka Riitta | Big Sisters. Spacing Woman Workers in the Clothing Industry. A Study on Flexible Production and Flexible Women. | TaY |
| Lehtiö Jouko | Suomalaisten vammaisten elämä varsinaisessa elämäkerta- ja vammaisten elämää keskeisesti käsittelevässä kaunokirjallisuudessa. | TY |
| Maula Johanna | Small Scale Production of Food and Traditional Alcohol Beverages in Benin and Tanzania: Implications for the Promotion of Female Entrepreneurship. | HY |
| Nikula Jouko | From State-dependency to Genuine Worker Movement? The Working Class in Socialism and Post-socialism. | TaY |
| Piispa Matti | Valistus, holhous, suojele: tutkimus alkoholipoliittisesta ja tupakkapoliittisesta mielipideilmastosta. | TaY |
| Stenström Marja-Leena | Educational and gender equality in vocational education: the case of commercial education in Finland. | .JY |
| Stålström Olli | Homoseksuaalisuuden sairausleiman loppu. | KY |
| Temkina Anna | Russia in transition: the cases of new collective actors and new collective actions. | HY |
| Tenhunen Sirpa | Secret Freedom in the City. Women's Wage Work and Agency in Calcutta. | HY |
| Topo Päivi | Dissemination of climacteric and postmenopausal hormone therapy in Finland : an example of the social shaping of a medical technology | HY |
| 1998 | | |
| Arminen Ilkka | Therapeutic interaction: a study of mutual help in the meetings of Alcoholics Anonymous. | HY |
| Honkatukia Päivi | Sopeutuvat tytöt? Sukupuoli, sosiaalinen kontrolli ja rikokset. | HY |
| Kainulainen Sakari | Elämäntapahtumat ja elämään tyytyväisyys eri sosiaaliluokissa. | KY |
| Kaskinen Juha | Tuhatta ja sataa Turusta Pietariin: Eurooppatie E18:n tiensuunnittelu ympäristövaikutusten arvioinnin käytäntönä. | TY |
| Linko Maaria | Aitojen elämysten kaipuu: yleisön kuvataiteelle, kirjallisuudelle ja museoille antamat merkitykset. | HY |
| Manderbacka Kristiina | Questions on survey questions on health. | HY |
| Oksa Jukka | Syrjäkylä muutoksessa. | JoY |
| Valli Raine | Koulun alkoholikasvatus keskioluen vähittäismyynnin vapautuessa. | LY |
| Veijola Soile | Liikkuvat subjektit, paikallinen tieto: tutkimuksia urheilusta, turismista ja sosiologiasta: väitöskirjan yhteenvedo. | HY |
| 1999 | | |
| Aapola Sinikka | Murrosikä ja sukupuoli: julkiset ja yksityiset ikämäärittelyt. | HY |
| Aro, Jari | Sosiologia ja kielenkäyttö: retoriikka, narratiivi, metafora. | TaY |
| Engeström Ritva | Toiminnan moniäänisyys: tutkimus lääkärinvastaanottojen keskusteluista. | HY |
| Kevätsalo Kimmo | Jäykät joustot ja tuhlatut resurssit. | HY |
| Koskinen Ilpo | Managerial evaluations at the workplace: an ethnographically situated conversation analytic study of evaluations in a research institute. | HY |
| Kurvinen Arja | Tilinteon aika: tutkimus pankista työnsä menettävien naisten identiteetin uudelleen arvioinneista. | JoY |
| Kuula Arja | Toimintatutkimus: kenttätöitä ja muutospyrkimyksiä. | TaY |
| Lehtonen Turo-Kimmo | Rahan vallassa: ostoksilla käyminen ja markkinatalouden arki. | HY |

| | | |
|-----------------------|---|-----------|
| Lonkila Markku | Social networks in Post-Soviet Russia: continuity and change in the everyday life of St. Petersburg teachers. | HY |
| Mäkelä Pia | Alkoholiin liittyvät kuolemat. | HY |
| Nisula Tapio | Everyday Spirits and Medical Interventions. Ethnographic and Historical Notes on Therapeutic Conventions in Zanzibar Town. | TaY |
| Pietilä Tuulikki | Gossip, markets and gender. The dialogical construction of morality in Kilimanjaro. | HY |
| Pohjola Kirsi | Naisten salaiset maailmat. | JY |
| Saloniemi Antti | Työn tekemisen turvallisuus, rakenteet ja rakentaminen. Sosiologisia näkökulmia rakennustyöhön ja työsuojeluun. | TaY |
| Sintonen Teppo | Etninen identiteetti ja narratiivisuus: Kanadan suomalaiset miehet elämänsä kertojina. | JY |
| Törrönen Jukka | Juomisen vapaus ja vastuu: sosiosemiotittainen analyysi alkoholipoliittisesta liberalismista maallikkoajattelussa. | HY |
| 2000 | | |
| Arppe Tiina | Pyhä ja kirottu. Pahan ongelma ranskalaisessa yhteiskuntateoriassa. | HY |
| Berg Mari-Anna | Ravitsemussuosiin liittyvien ruokatottumusten alue-erot ja niiden muutokset Suomessa. | HY |
| Harinen Päivi | Valmiiseen tulleet: tutkimus nuoruudesta, kansallisuudesta ja kansalaisuudesta. | JoY |
| Kilpeläinen Arja | Teknologiavälitteisyys kylälaisten arjessa: tutkimus ikääntyvien sivukylien teknologiavälitteisyydestä ja sen rajapinnoista maaseutusosiaalityöhön. | JY |
| Kilpinen Erkki | The Enormous Fly-wheel of Society. Pragmatism's Habitual Conception of Action and Social Theory. | HY |
| Laine Kaarlo | Koulukuvia: koulu nuorten kokemistilana. | JY |
| Lehtinen Anja-Riitta | Lasten kesken: lapset toimijoina päiväkodissa. | JY |
| Lepola Outi | Ulkomaalaisesta suomenmaalaiseksi: monikulttuurisuus, kansalaisuus ja suomalaisuus 1990-luvun maahanmuuttopoliittisessa keskustelussa. | HY |
| Lähteenmäki Jaana | Myöhäismoderni nuorisokulttuuri. Tulkintoja ryhmistä ja ryhmiin kuulumisten ulottuvuuksista. | HY |
| Nouko-Juvonen Susanna | Kehitysvammopalvelut ja muutoksen vuosikymmen. | TY |
| Nupponen Terttu | Arkkitehdit, sota ja yhdyskuntasuhteiden hallinta. Alvar Aallon Kokemäkjokilaakson aluesuunnitelma tilansääätelyprojektina. | JY |
| Rubin Anita | Growing up in social transition: in search of late-modern identity. | TY |
| Saaristo Kimmo | Avoin asiantuntijuus: ympäristökysymys ja monimuotoinen ekspertisi. | HY |
| Teittinen Antti | Miten tutkia vammaisuutta yhteiskunnallisena kysymyksenä. | JY |
| Valtonen Pekka | The politics of Agrarian Transformation in Mexico. | TaY |
| Värtö Petteri | "Mies vastaa tekosistaan ... siinä missä nainenkin": maskuliinisuuden rakentaminen päiväkodissa. | KY |
| 2001 | | |
| Castrén Anna-Maija | Family and Work in Helsinki and St. Petersburg: Social Circles and Societies in the social Networks of Teachers. | HY |
| Hansson Leeni | Networks Matter. The Role of Informal Social Networks in the Period of Socio-Economic Reforms of the 1990s in Estonia. | JY |

| | | |
|-----------------------------------|---|-----------|
| Hohenthal-Antin Leonie | Luvan ottaminen. Ikäihmiset teatterin tekijöinä. | JY |
| Kangas Risto | Yhteiskunta. Tutkielmia yhteiskunnasta, yhteiskuntakäsitteestä ja sosiologiasta. | HY |
| Kautto Mikko | Diversity among welfare states. Comparative studies on welfare state adjustment in Nordic countries. | HY |
| Kinnunen Merja | Luokiteltu sukupuoli. | TaY |
| Komonen Katja | Koulutusyhteiskunnan marginaalissa? Ammatillisen koulutuksen keskeyttäneiden nuorten yhteiskunnallinen osallisuus. | JoY |
| Kumpulainen Riitta-Helena | Timber and Herring: Modernisation and Mobility in Finnish Lapland and the Western Islands of Scotland, 1770–1970. | HY |
| Litmanen Tapio | The Struggle over Risk. The Spatial, Temporal and Cultural Dimensions of Anti-Nuclear Protest. | JY |
| Lojander-Visapää Catharina | Med rätt att välja : språkval och språkstrategier i språkligt blandade hushåll i Helsingfors. | HY |
| Lövdahl Ulrica | The social construction of mental health in Nordic psychotropic drug advertising. | ÅÅ |
| Mae Vanessa | Lone Motherhood in Finnish Women's Life Stories Creating Meaning in a Narrative Context. | ÅÅ |
| Oinas Elina | Making Sense of the Teenage Body. Sociological Perspectives on Girls, Changing Bodies, and Knowledge. | ÅÅ |
| Rantala Kati | "Ite pitää keksii se juttu" - Tutkimus kuvataidekasvatuksen ja kasvatettavan kohtaamisesta. | HY |
| Suopajarvi Leena | Vuotos- ja Ounasjokikamppailujen kentät ja merkitykset Lapissa. | LY |
| Tolonen Tarja | Nuorten äänet ja tilat. Sukupuolen järjestykset koulun arjessa. | HY |
| Tuulentie Seija | Meidän vähemmistömmä. Valtaväestön retoriikat saamelaisten oikeuksista käydyissä keskusteluissa 2001. | LY |
| Wrede Sirpa | Decentering Care for Mothers. The Politics of Midwifery and the Design of Finnish Maternity Services. | ÅÅ |
| 2002 | | |
| Bergman Solveig | The Politics of Feminism Autonomous Feminist Movements in Finland and West Germany from the 1960s to the 1980s. | ÅÅ |
| Hautamäki Tiina | Tuberkuloosin ja aidsin kulttuuriset merkitykset. | TaY |
| Heiskanen Markku | Väkivalta, pelko ja turvattomuus. Surveytutkimusten näkökulmia suomalaisten turvallisuuteen. | HY |
| Huttunen Laura | Kotona, maanpaossa, matkalla: kodin merkitykset maahanmuuttajien omaelämäkerroissa. | TaY |
| Jallinoja Piia | Genetics, negotiated ethics and the ambiguities of moral choices. | HY |
| Jokivuori Pertti | Sitoutuminen työorganisaatioon ja ammattijärjestöön. | JY |
| Kouvonen Anne | Part-time work and deviant behaviour among Finnish adolescents. | HY |
| Louhimaa Eila | Luonnon sosiaalinen konstruointi, ympäristödiskurssit ja ympäristöön orientoiva kasvatustutkimus. Tutkimus institutionaalisen ympäristökasvatuksen yhteiskunnallisista rakenne-ehdoista ja kulttuuristen mahdollisuuksien kentistä. | OY |
| Luoma Pentti | Vihreät viirit. Muutos ja pysyvyys Maa- ja metsätaloustuottajain Keskusliiton ympäristöpoliittisissa näkemyksissä vuosina 1980 – 2000 Maataloustuottaja -lehden valossa. | OY |
| Mäkelä Johanna | Syömisen rakenne ja kulttuurinen vaihtelu. | HY |

| | | |
|------------------------|---|-----------|
| Pohjanen Jorma | Mitä kello on? Kello modernissa yhteiskunnassa ja sen sosiologisessa teoriassa. | JY |
| Reijonen Merja | Tila ja kulttuurinen identiteetti - entisen Neuvostoliiton alueelta tulleiden paluumuuttajien kulttuurisen identiteetin muodostuminen Suomessa asutun vuoden aikana. | HY |
| Ruuska Petri | Kuviteltu Suomi: Globalisaation, nationalismin ja suomalaisuuden punos julkisissa sanoissa 1980–90-luvuilla. | TaY |
| Suurpää Leena | Erilaisuuden hierarkiat. Suomalaisia käsityksiä maahanmuuttajista, suvaitsevaisuudesta ja rasismista. | HY |
| Toiskallio Kalle | Taksimatkan seurallisuus, liikenteen neuvottelut, automainosten näkökulmat. | HY |
| Vahtio Eeva-Leena | Rekrytointi, ikä ja ageismi. | JY |
| 2003 | | |
| Eriksson-Piela Susan | Tunnetta, tietoa vai hierarkiaa? Sairaanhoidon moninainen ammatillisuus. | TaY |
| Houtsonen Jarmo | Sopimus talouden instituutiona. | JoY |
| Högbacka Riitta | Naisten muuttuvat elämänmuodot maaseudulla. | HY |
| Lehtonen Jukka | Seksuaalisuus ja sukupuoli koulussa – näkökulmana heteronormatiivisuus ja ei-heteroseksuaalisten nuorten kertomukset. | HY |
| Lempiäinen Kirsti | Sosiologian sukupuoli. | TaY |
| Melkas Tuula | Sosiaalisesta muodosta toiseen - Suomalaisten yksityiselämän sosiaalisuuden tarkastelua vuosilta 1986 ja 1994. | HY |
| Mikkola Teija | Muuttuvat arvot ja uusi keskiluokka. | HY |
| Nakari Maija | Työilmapiiri, työntekijöiden hyvinvointi ja muutoksen mahdollisuus. | JY |
| Palosuo Hannele | Health and Well-Being in Moscow and Helsinki. | HY |
| Palukka Hannele | Johtotähdet. Lennonjohtajien ammatti-identiteetin rakentuminen ryhmähaastatteluissa. | TaY |
| Pensola Tiina | From past to present: The effect of the lifecourse on mortality, and social class differences in mortality in middle adulthood". | HY |
| Ponnikas Jouni | Globaali käsite paikallisessa kontekstissa: Kestävän kehityksen paikallistuminen Pohjois-Pohjanmaan ja Kainuun EU:n aluepolitiikan toimijoiden mielipideilmastossa vuosina 1995–2002. | OY |
| Poursadigh Amir | The Determinants of the Revolutionary Disintegration of the State in Iran. | TaY |
| Poutanen Seppo | Kuinka päättää geenitutkimuksiin osallistumisesta? Normatiivinen ja sosiologinen analyysi bayesilaisen menetelmän kelpoisuudesta avuksi. | TY |
| Ruuskanen Petri | Verkostotalous ja luottamus. | JY |
| Ruutsoo Rein | Civil Society and Nation Building in Estonia and the Baltic States. Impact of Traditions on Mobilization 1986-2000: Historical and Sociological Study. | LY |
| Räsänen Pekka | In the Twilight of Social Structures: A Mechanism-Based Study of Contemporary Consumer Behaviour. | TY |
| Valkonen Jarno | Lapin luontopolitiikka: analyysi vuosien 1946–2000 julkisesta keskustelusta. | TaY |
| 2004 | | |
| Alho Arja | Silent Democracy, Noisy Media. | HY |

| | | |
|------------------------------|--|------------|
| Dahlgren Susanne | Contesting Realities. Morality, propriety and the Public Sphere in Aden, Yemen. | HY |
| Grenier Alain | The Nature of Turism. | LY |
| Helander Miika | Fenomenologisk globalisering. Vardaglig nationalism bland transnationellt aktiva svenska och finländska företagsledare. | HY |
| Honkanen Antti | Menneisyyden tulevaisuus. Postmodernit matkailuteoriat ja vapaa-ajan matkailun muutokset eräissä Euroopan unionin jäsenvaltioissa vuosina 1985 ja 1997. | TY |
| Järvinen Riitta | Ammatillisen maahanmuuttotyön kulttuuri: erilaisuus sosiaali- ja terveydenhuollon jäsenyksissä. | TaY |
| Kallinen Timo | Some Chiefs Are "More Under" than Others. Kinship, ritual, and the concept of political hierarchy among the Asante. | HY |
| Kauppinen Timo | Asuinalueen ja perhetaustan vaikutukset helsinkiläisnuorten keskiasteen tutkintojen suorittamiseen. | HY |
| Kovero Seppo | Puhdas, soveltava ja empiirinen. Ferdinand Tönniesin "erityisen sosiologian" järjestelmä: mies, teoria ja tulkinnat. | JoY |
| Lettinen Eva | Saaristolaisen kokemus omasta kulttuuristaan. Eläminen merestä ja meren kanssa. | TY |
| Länsman Anni-Siiri | Väärtisuhteet Lapin matkailussa. Kulttuurianalyysi suomalaisten ja saamelaisten kohtaamisesta. | LY |
| Oinonen Eriikka | Finnish and Spanish Families in Converging Europe. | TaY |
| Palmroth Aino | Käännösten kautta kollektiiviin – tuuliosuuskunnat toimijaverkkoina. | JY |
| Peuhkuri Timo | Tiedon roolit ympäristökiistassa. Saaristomeren rehevöityminen ja kalankasvatus julkisen keskustelun ja päätöksenteon kohteena. | TY |
| Rotko Tuulia | Social Aspects of Air Pollution. | HY |
| Sankari Anne | Ikääntyviä tietokoneen äärellä. Tutkimus kulttuurisista ajattelutavoista ja ikääntymisen sosiaalisesta tilasta. | JY |
| Sillander Kenneth | Acting Authoritatively: How Authority is Expressed through Social Action among the Bentian of Indonesian Borneo | HY |
| Tervo Hannu | Rakkaudesta teknologiaan. Toimijaverkostonäkökulma Oulun IT-keskittymän rakentumiseen ja toimintaan. | LY |
| Tuunainen Juha | Hybrid practices: the dynamics of university research and emergence of a biotechnology company. | HY |
| Vaaranen Heli | Kaaharipoikia ja rappioromantiikkaa - tutkimus korttelirallin ja kaahauskulttuurin elämänilosta sekä tuhoisuudesta. | HY |
| Virkki Tuija | Viha muutosvoimana. Toimijuus, sukupolvet ja sukupuoli vihakertomuksissa. | JY |
| 2005 | | |
| Akhlaq Ahmad | Huono-osaisuus Suomen kaupunkiseutukunnissa: alue-erot ja sosiaalisen ympäristön vaikutukset 1990-luvulla. | HY |
| Alastalo Marja | Metodisuhdanteiden mahti. Lomaketutkimus suomalaisessa sosiologiassa 1947–2000. | TaY |
| Anttila Anu-Hanna | Loma tehtaan varjossa. Teollisuustyöväestön loma- ja vapaa-ajan moraalisaätely Suomessa 1930–1960-luvuilla. | TY |
| Blomgren Jenni | Huono-osaisuus Suomen kaupunkiseutukunnissa: alue-erot ja sosiaalisen ympäristön vaikutukset 1990-luvulla. | HY |
| Hilden-Paajanen Tarja | Pahojen poikien piiri: Joensuun skinien epäilyt rikollisuus 1995–1998. | TaY |

| | | |
|-------------------------------------|---|------------|
| Hokkanen Tiina | Äitinä ja isänä eron jälkeen: yhteishuoltajavanhemmuus arjen kokemuksena. | JY |
| Ilvesviita Pirjo | Paaluraudoista kotkansuojeluun. Suomalainen metsästyspolitiikka 1865–1993. | LY |
| Itäpuisto Maritta | Kokemuksia alkoholiongelmaisten vanhempien kanssa eletystä lapsuudesta. | KY |
| Karttunen Marie-Louise | Making a communal world. English merchants in imperial St. Petersburg. | HY |
| Kurki Eija | Näkyvä ja näkymätön. Nainen Suomen helluntailiikkeen kentällä. | JY |
| Luostarinen Matti | Ekologinen klusteri ja innovaatiopolitiikka. | TY |
| Mäenpää Pasi | Narkissos kaupungissa. Tutkimus kuluttajakaupunkilaisesta ja julkisesta tilasta. | HY |
| Nieminen Ari | Towards a European Society? Integration and Regulation of Capitalism. | HY |
| Nieminen Mika | Academic Research in Change. Transformation of Finnish University Policies and University Research during the 1990s. | TaY |
| Nyman Jopi | Imagining Englishness. Essays on the Representations of National Identity in Modern British Culture. | |
| Turunen Maija | Faith in the heart of Russia: the religiosity of post-Soviet university students. | HY |
| Vainionpää Kirsi | “Male menopause”. The birth of a new illness? | LY |
| Yrjälä Ann | Public health and Rockefeller wealth. Alliance strategies in the early formation of Finnish public health nursing. | ÅÅ |
| 2006 | | |
| Aaltonen Sanna | Tytöt, pojat ja sukupuolinen häirintä. | HY |
| Ahlbeck-Rehn Jutta | Diagnostisering och disciplinering: Medicinsk diskurs och kvinnligt vansinne: på Sjäälö hospital 1889–1944. | ÅÅ |
| Allaste Airi-Alina | Drug cultures in Estonia. Contexts, meanings and patterns of illicit drug use. | HY |
| Bäckman Johan | Itämafia - Uhkakuvapolitiikka, rikosilmiöt ja kulttuuriset merkitykset. | HY |
| Ganskau Elena | Trust and safe choices. Coping with health-related risks in food consumption in St. Petersburg. | HY |
| Heikkilä Lydia | Reindeer Talk. Sámi Reindeer Herding and Nature Management. | LY |
| Karjalainen Timo | The environment in context: Environmental concern in the Komi Republic (Russia). | OY |
| Kutinlahti Pirjo | Universities approaching market. Intertwining scientific and entrepreneurial goals. | HY |
| Kuussaari Kristiina | Näkemyksen kirjo, sirpaloitunut tieto. Työntekijöiden käsityksiä huumeiden käyttäjistä ja käyttäjien hoidosta. | TY |
| Käyhkö Mari | Siivoojaksi oppimassa. Etnografinen tutkimus työläistyöistä puhdistuspalvelualan koulutuksessa. | JoY |
| Leinonen Anu | Vanhusneuvoston funktioita jäljittämässä. Tutkimus maaseutumaisten kuntien vanhusneuvostoista. | JY |
| Malin Maili | Biovalta, toiseus ja naisten toimijuus hedelmöityshoidoissa. | HY |
| Martiskainen de Koenigswarter Heini | Discursive matrixes of motherhood. Cultivating decency and emotion in Finnish and French mother-talk. | HY |
| Mattila Pekka | Toiminta, valta ja kokemus organisaation muutoksessa. Tutkimus kolmesta suuryrityksestä. | HY |
| Peteri Virve | Mediaksi kotiin. Tutkimus teknologioiden kotouttamisesta. | TaY |

| | | |
|------------------------------|---|------------|
| Pirinen Riitta | Urheilleva Nainen lehtiteksteissä. | TaY |
| Pyöriä Pasi | Understanding Work in the Age of Information. Finland in Focus. | TaY |
| Rasmus Ari | Uudet liikkeet. Radikaali kansalaisaktivismi 1990-luvun Suomessa. | TaY |
| Salmi Anna-Maria | Social Networks and Everyday Practices in Russia. | HY |
| Sanaksenaho Sanna | Eriarvoisuus ja luottamus 2000-luvun taitteen Suomessa. Bourdieulainen näkökulma. | JY |
| Shemeikka Riikka | Fertility in Namibia - Changes in fertility levels in North-Central Namibia 1960–2001, including an assessment of the impact of HIV. | HY |
| Weckroth Antti | Valta ja merkitysten tuottaminen korvaushoidossa. Etnografinen tutkimus huumehoitolaitoksesta. | |
| Vilkkä Hanna | Keho. Omaa sukupuolta koskevan tiedon ja ymmärryksen muotoutumisen perusta transsukupuolisilla. | HY |
| 2007 | | |
| Bauvois Gwenaëlle | Pensseli ja mitali. Albert Edelfeltin yhteistyöverkostot Ranskan taidekentällä 1874–1905. | JoY |
| Jalovaara Marika | The Effects of Marriage Partners' socio-economic Positions on the Risk of Divorce in Finland. | HY |
| Jauho Mikko | Kansanterveysongelman synty. Tuberkuloosi ja terveydenhallinta Suomessa ennen toista maailmansotaa. | HY |
| Kankainen Tomi | Yhdistykset, instituutiot ja luottamus. | JY |
| Kantola Ismo | The role of knowledge in energy politics. A sociological study. | TY |
| Karsikas Leevi | Metsästyksen ongelmapuhe. | OY |
| Lammi-Taskula Johanna | Parental leave for fathers? Gendered conceptions and practices in families with young children in Finland. | TaY |
| Määttä Mirja | Yhteinen verkosto? Tutkimus nuorten syrjäytymistä ehkäisevistä poikkihallinnollisista ryhmistä. | HY |
| Purhonen Semi | Sukupolvien ongelma. Tutkielmia sukupolven käsitteestä, sukupolvi-tietoisuudesta ja suurista ikäluokista. | HY |
| Salmenniemi Suvi | Growing Flower in the Frost: Citizenship and Gender in Contemporary Russia. | HY |
| Silvennoinen Piia | Ikä, identiteetti ja ohjaava koulutus. Ikääntyvät pitkäaikaistyöttömät oppimisyhteiskunnan haasteena. | JY |
| Tammi Tuukka | Medicalising prohibition : harm reduction in Finnish and international drug policy | HY |
| Tikka Tiina | Taitavaksi hoitajaksi, arvostetuksi kasvattajaksi – ammatillinen sosiaalisatio perhepäivähoitajan työssä. | JoY |
| 2008 | | |
| Bäckman Sonja | Työkykyarviokokous moniammatillisen työryhmän päätöksentekotilanteena. Tutkimus vuorovaikutuksesta ja potilaan osallisuudesta. | HY |
| Jokiranta Kari | Konkretisoitua uhka. Ilkka-lehden huumekirjoitukset vuosina 1970–2002. | TaY |
| Kauppinen Ilkka | Tiedon omistaminen on valtaa. Globalisoituvan patenttijärjestelmän poliittinen moraalitalous ja globaali kapitalismi. | JY |
| Kauranen Ralf | Seriedebatt i 1950-talets Finland. En studie i barndom, media och reglering. | ÅÅ |

| | | |
|-------------------------|--|------------|
| Kestilä Laura | Pathways to health. Determinants of health, health behaviour and health inequalities in early adulthood. | HY |
| Kinnunen Aarne | Kriminaalipolitiikan paradoksi. Tutkimuksia huumausainerikollisuudesta ja sen kontrollista Suomessa. | HY |
| Koivumäki Jaakko | Työyhteisöjen sosiaalinen pääoma. Tutkimus luottamuksen ja yhteisöllisyyden rakentumisesta ja merkityksestä muuttuvissa valtion asiantuntijaorganisaatioissa. | TaY |
| Kokkonen Jouko | Kansakunta kilpasilla. urheilu nationalismin kanavana ja lähteenä Suomessa 1900–1952. | JY |
| Kontula Anna | Punainen eksodus. Tutkimus seksityöstä Suomessa. | TaY |
| Pelkonen Antti | The Finnish Competition State and Entrepreneurial Policies in the Helsinki Region. | HY |
| Rannikko Ulla | Yhteinen ja erillinen lapsuus. Sisarusten sosiaalistava merkitys. | TaY |
| Sourabh Naresh Chandra | The Culture of Women's Housework. A Case Study of Bihar, India. | HY |
| Tupasela Aaro | Consent Practices and Biomedical knowledge Production in Tissue Economies. | HY |
| 2009 | | |
| Haapakorpi Arja | Sulkeuman ja neuvottelun ehdoilla - asiantuntija-aseman rakentuminen työelämässä. | HY |
| Hakala Johanna | Academic Cultures in the Finnish Mass Research University. Change and continuity. | TaY |
| Kivimäki Riikka | Vanhemmuus työpaikoilla. | TaY |
| Kjaernes Unni | Regulating Food Consumption. Studies of change and variation in Europe. | HY |
| Leinikki Sikke | Pelon ja toivon välissä. Pätkätyöläisen urakerronta. | HY |
| Liljeroth Pia | Rädsla inför förlossningen. Ett uppenbart kliniskt problem? Konstruktionen av förlossningsrädsla som medicinsk kategori. | ÅÅ |
| Pyyhtinen Olli | Bringing the social alive. essays on Georg Simmel's social theory. | TY |
| Sallila Seppo | Hyvinvointivaltion optimaalisen köyhyyspolitiikan määrittelyä. | HY |
| Snell Karoliina | Social responsibility in developing new biotechnology. interpretations of responsibility in the governance of Finnish biotechnology. | HY |
| Tuori Salla | The politics of multicultural encounters. Feminist postcolonial perspectives. | ÅÅ |
| Vesikansa Sari-Liisa | Kuka kasvattaa, kuka opettaa? Genealoginen tutkimus perheen ja koulun välisen kasvatusvastuun politiikasta. | HY |
| Vihavainen Rosa | Homeowners' Associations in Russia after the 2005 Housing Reform. | HY |
| 2010 | | |
| Auvinen Tero | On money. | LY |
| Einiö Elina | Determinants of institutional care at older ages in Finland. | HY |
| Hemminki Arja | Kertomuksia avioerosta ja parisuhteen päättymisestä. Suomalainen eropuhe pohjalaisten kirjoituksissa ja naistenlehdissä. | JY |
| Herttua Kimmo | The effects of the 2004 reduction in the price of alcohol on alcohol-related harm in Finland. A natural experiment based on register data. | HY |
| Kalanti Timo | Ruumis ja rauta. Esseitä esineiden sosiaalisuudesta. | HY |

| | | |
|----------------------------|--|------------|
| Kaskiharju Eija | Koteja ja kodinomaisuutta. Tutkimus vanhenemisen paikoista valtiopäiväpuheissa 1950–2005. | JY |
| Ketokivi Kaisa | The Relational Self, the Social Bond and the Dynamics of Personal Relationships. A Sociological Analysis. | HY |
| Korkala Siru | Luottamuksen ilmeneminen alueellisissa yhteistyöverkostoissa. | TY |
| Lillbroända-Annala Sanna | Från kåk till kulturarv. En etnologisk studie av omvärderingen av historiska trädadsområden i Karleby och Ekenäs. | ÅA |
| Luomanen Jari | Living with the Media. Analysing Talk about Information and Communication Technology. | TaY |
| Maksimainen Jaana | Parisuhde ja ero. Sosiologinen analyysi terapeuttisesta ymmärryksestä. | HY |
| Mäki Netta | Not in all Walks of Life? Social Differences in Suicide Mortality. | HY |
| Perho Sini | Rasistisuus nuorten yhteisöissä. Tutkimus vuosituuhannen vaihteen Joensuusta. | UEF |
| Saksela-Bergholm Sanna | Immigrant Associations in the Metropolitan Area of Finland. Forms of Mobilisation, Participation and Representation. | HY |
| Shahedul Halim Kazi | Making Ends Meet - Entrepreneurship in a Bangladeshi Village. | TaY |
| Vehmas Hanna | Liikuntamatkalla Suomessa. Vapaa-ajan valintoja jälkimodernissa yhteiskunnassa. | JY |
| Voutilainen Liisa | Emotional experience in psychotherapeutic interaction. Conversation analytical study on cognitive psychotherapy. | HY |
| Yliskylä Satu | Kuka suunnittelee mainokset? Tutkimus mainosalan kulisseista, yhteistyökäytännöistä ja ristiriidoista. Mainonnan suunnittelijan näkökulma. | TY |
| 2011 | | |
| Halonen Katri | Kulttuurituottajat taiteen ja talouden risteyskohdassa. | JY |
| Jungar Katarina | Long live! South African HIV-activism, knowledge and power. | ÅA |
| Katainen Anu | Tupakka, luokka ja terveyskäyttämisen ongelma. | HY |
| Kekkonen Kari | HYVÄÄ VAUHTIA METSÄTÖILLE. Puunkorjuu ja Suomi muutoksessa. | TaY |
| Korhonen Sirpa | Veteraanien sotakokemukset ja hyvinvointi. Seurantatutkimus pohjoissavolaisista sodan kokeneista miehistä vuosina 1993 ja 2006. | UEF |
| Mankkinen Teija | Palomiehen ammatti työnä ja elämäntapana. | TaY |
| Mattson Kristin | Att skriva det finlandssvenska. Etnicitet i kvinnors livshistorier. | ÅA |
| Notko Marianne | Väkivalta, vallankäyttö ja vahingoittuminen naisten perhesuhteissa. | JY |
| Nummela Tuija | Asiakkaan asema ja oikeuksien toteutuminen aikuissosiaalityössä. | UEF |
| Paju Petri | Koulua on käytävä. Etnografinen tutkimus koululuokasta sosiaalisena tilana. | UEF |
| Perälä Jussi-Pekka | "Miksi lehmät pitää tappaa?" Etnografinen tutkimus 2000-luvun alun huumemarkkinoista Helsingissä. | HY |
| Selin Jani | Hallinnan näkökulmia huumeriippuvuuden hoitoon Suomessa vuosina 1965–2005. | JY |
| Smolej Mirka | News Media, Crime and Fear of Violence. | HY |
| Souto Anne-Mari | Pelko ja rasismi koulussa | UEF |
| Strauss Hannah | For the Good of Society. Public participation in the siting of nuclear and hydro power projects in Finland. | OY |
| Tuomi Margaret Trotta | Human Dignity in the Learning Environment. | JY |

| | | |
|----------------------------|--|-----------|
| Voipio Timo | From Poverty Economics to Global Social Policy. A Sociology of Aid for Poverty Reduction. | UEF |
| Yli-Mäyry Soile | Esteettinen kokemus kulttuurisiltana. Soile Yli-Mäyrän taide kiinalaisen, japanilaisen ja suomalaisen yleisön kokemana. | HY |
| 2012 | | |
| Aurélie Mary | The Illusion of the Prolongation of Youth - Transition to Adulthood among Finnish and French Female University Students. | TaY |
| Autto Janne | Päivähoitopolitiikka kamppailuna hyvinvointivaltiosta. Kentät, subjektiasemat ja oikeutukset. | LY |
| Habti Driss | Highly Skilled Mobility and Migration from MENA Region to Finland. A Socio-analytic Approach. | UEF |
| Haikkola Lotta | Monipaikkainen nuoruus. Toinen sukupolvi, transnationaalisuus ja identiteetit. | HY |
| Jonker-Hoffrén Paul | The Finnish Paper Workers' Union at a crossroads. Labor union representativeness in a changing environment, 1980–2008. | TY |
| Kamila Marjo | Katsojana ja katsottuna. Opettajan kontrolloitu ulkoasu. | JY |
| Kanckos Lise | Barnets bästa i politikens främsta rum. Finlands riksdags debatt om assisterad befruktning. | ÅÅ |
| Kekki Tuula | Taparikollisuuden kulttuuriset ja rakenteelliset ulottuvuudet. Huumeiden käyttäjien ja poliisin tulkinnat taparikollisuudesta. | HY |
| Kivelä Päivi | Oman paikan taju. Keski-ikäiset perheenäidit paikallisilla kentillä. | JY |
| Koskela Jari | Avioero suomalaisen miehen kokemana. Ero ja erosta toipuminen prosessina. | UEF |
| Leppo Anna | Precarious Pregnancies. Alcohol, drugs and the regulation of risks. | HY |
| Näre Lena | Moral Economies of Reproductive Labour. An Ethnography of Migrant Domestic and Care Labour in Naples, Italy. | HY |
| Perälä Riikka | Haittojen vähentäminen suomalaisessa huumehoidossa. Etnografinen tutkimus huumeongelman yhteiskunnallisesta hallinnasta 2000-luvun Suomessa. | HY |
| Pärnä Katariina | Kehittävä moniammatillinen yhteistyö prosessina. Lapsiperheiden varhaisen tukemisen mahdollisuudet. | TY |
| Ranssi-Matikainen Hanna | Yhdessä sittenkin. Kokemukset erokriisistä, vertaistuesta ja muutoksesta parisuhteessa. | HY |
| Remes Hanna | Social determinants of mortality from childhood to early adulthood. | HY |
| Salminen Veli-Matti | Verkostot, paikallisuus ja eriarvoisuus. | JY |
| Sarno Irina | Learning Managers in a Transforming Economy - The Case of Russia 1999–2006. | TaY |
| Tarvainen Merja | Production, reproduction and citizenship. Key discourses of social policy in the turn of the 20th century. | UEF |
| Teräväinen Tuula | The politics of energy technologies. Debating climate change, energy policy, and technology in Finland, United Kingdom, and France. | HY |
| Tuomi Marja | Diffusion of social innovations across the borders social sector cooperation with the Republic of Karelia. | UEF |
| Turunen Teemu | Työorientaatiot muutoksessa? Suomalaisten palkansaajien työhön ja organisaatioon sitoutuminen sekä työhön kohdistuvat odotukset eurooppalaisessa vertailussa. | HY |

| | | |
|--------------------------|--|-----------|
| Ylä-Outinen Tuulikki | Ikäihmisten arki. Kotona asuvien palvelutaloon muuttaneiden ikäihmisten kertomuksia. | UEF |
| 2013 | | |
| Alanko Anu | Osallisuusryhmä demokratiakasvatuksen tilana. Tapaustutkimus lasten ja nuorten osallisuusryhmien toiminnasta Oulussa. | OY |
| Autti Outi | Valtavirta muutoksessa - vesivoima ja paikalliset asukkaat Kemijoella. | OY |
| Eräranta Kirsi | Taloudellinen, sukupuolittunut ja transnationaali sosiaalinen. Hyvinvointihallinnan muutos ja kysymys työn ja perheen yhteensovittamisesta. | HY |
| Husu Hanna-Mari | Social movements and Bourdieu. Class, embodiment and the politics of identity. | JY |
| Härkönen Janne | Not a wet generation but a wet nation. The dynamics of change and stasis in the Finnish drinking culture from 1968-2008. | HY |
| Janhunen Kirsi-Marja | Kouluhyvinvointi nuorten tulkitsemana. | UEF |
| Koikkalainen Saara | Making it abroad. Experiences of highly skilled Finns in the European Union labour markets. | LY |
| Koivula Riitta | Muistisairaana ihmisen omaisena terveyskeskuksen pitkäaikaisosastolla. Tutkimus toimijuudesta. | JY |
| Kotanen Riikka | Näkymättömästä näkökulmaksi. Parisuhdeväkivallan uhrit ja oikeudellisen sääntelyn muutos Suomessa. | HY |
| Le Vecchia-Mikkola Vanja | Longing to Return and Spaces of Belonging. Iraqis' Narratives in Helsinki and Rome. | TY |
| Lähteenmäki Liisa | Keskusteluja vuokratyöstä. Vuokratyön diskursiivinen rakentuminen Suomessa. | TY |
| Löfblom Katariina | Nuorten kokemien elämysten yksilölliset ja yhteisölliset merkitykset. | TY |
| Mäkelä Pekka | Collective Responsibility. Against Collectivism. | JY |
| Pirkkalainen Päivi | Transnational responsibilities and multi-sited strategies. Voluntary associations of Somali diaspora in Finland. | JY |
| Pääkkönen Leena | Nuorten musisointiprosessi koulussa toteutetussa konserttiprojektissa. Musiikkiluokkalaisten kertomukset yhdessä tekemisestä. | OY |
| Pöllänen Pirjo | Hoivan rajat. Venäläiset maahanmuuttajanaiset ja ylitajainen perhehoiva. | UEF |
| Raito Petteri | Toimittajien ja keskiöjohdon kokemukset ja selviytyminen työttömyydessä 1990-luvun alkupuolen laman aikana. Ammattiryhmiä vertaileva pitkittäistutkimus. | HY |
| Rautalin Marjaana | Domestication of International Comparisons. The role of the OECD Programme for International Student Assessment (PISA) in Finnish education policy. | TaY |
| Rekonen Petri | Raha, aika, itsekuri. Modernin maailman synty paikallisena ilmiönä. Modernisaatio Savossa ja Etelä-Pohjanmaalla 1870-1917. | TaY |
| Selkälä Arto | Kansalaisuuden hallinta suomalaisessa sosiaali- ja terveyspolitiikassa. | LY |
| Simonen Jenni | Lähentyvätkö naisten ja miesten juomatavat? Kvalitatiivinen tutkimus sukupuolesta ja juomisesta. | HY |
| Stevanovic Melisa | Deontic rights in interaction. A conversation analytic study on authority and cooperation. | HY |
| Takala Tuomo | Tutkimuksia karismasta eri konteksteissaan. | LY |

| | | |
|-------------------------|---|-----------------|
| Vuolanto Pia | Boundary-work and the Vulnerability of Academic Status. The Case of Finnish Nursing Science. | TaY |
| Ylilahti Minna | Itsestä kiinni. Etnografinen tutkimus työikäisten laitoskuntoutuksesta. | JY |
| 2014 | | |
| Aaltojärvi Inari | Making Domestic Technology Meaningful. From purification to emotions. | TaY |
| Auranen Otto | University research performance. Influence of funding competition, policy steering and micro-level factors. | TaY |
| Brennan Christopher | Precarious Working Tourists. Working Holiday Makers in Australia. | TaY |
| Choroswicz Marta | Managing competitiveness in pursuit of a legal career. Women attorneys in Finland and Poland. | UE F |
| Jokinen Kimmo | Suomalaisen lukemisen maisemaihanteet. | JY |
| Kabatilo Fadi | Democratization processes and the democratic culture in Arab countries. The case of Jordan. | UEF |
| Kinni Riitta-Liisa | Gerontologinen sosiaalityö ja ihmisten työstäminen. Kategorisointia sairaalan moniammatillisessa työssä. | UEF |
| Kouvo Antti | Luottamuksen lähteet. Vertaileva tutkimus yleistynyt luottamusta synnyttävistä mekanismeista. | TY |
| Lehtonen Päivi | Koulutuksen järjestäjien ja työelämäedustajien yhteistyö sosiaali- ja terveysalalla. Tapaustutkimukseen pohjautuva substantiivinen teoria. | TaY |
| Leinonen Taina | Disability retirement. Socioeconomic differences and health outcomes. | HY |
| Lyytikäinen Laura | Performing Political Opposition in Russia. The Case of the Youth Movement Oborona. | HY |
| Lähdesmäki Tuuli | Kuohahdus Suomen kansan sydäimestä. Henkilömonumentti diskursiivisena ilmiönä 1900-luvun lopun Suomessa. | UEF |
| Maunu Antti | Yöllä yhdessä. Yökerhot, biletyt ja suomalainen sosiaalisuus. | HY |
| Muhonen Reetta | Opinteitä ikääntyvien yliopistoon. Sotien sukupolvi ja yliopiston ylistys. | JY |
| Neophytou Andri | Conceptions of "us" and "them". The attitude of Greek and Turkish Cypriot university students towards the "other". | UEF |
| Nurmi Johanna | Let Go and Remember. How Finnish Communities Experienced and Memorialized School Shootings. | TY |
| Rantala Päivi | Erlaisia tapoja käyttää kylähullua. Kalkkimaan pappi aatteiden ja mentaliteettien tulkkina 1800-luvulta 2000-luvulle. | LY |
| Rautajoki Hanna | Hyvät "kanssalaiset"! Yhteiskunnallistamisen strategiat televisiokeskustelun vuorovaikutuksessa. | TaY |
| Rouhinen Sauli | Matkalla mallimaaksi? Kestävän kehityksen juurtuminen Suomessa. | UEF |
| Saari Tiina | Psykologiset sopimukset ja organisaatioon sitoutuminen tietotyössä. | TaY |
| Salo Miikka | Uusiutuva energia ja energiajärjestelmän konsensus. Mekanismin näkökulma liikenteen biopoltoaineiden ja syöttötariffin käyttöönottoon Suomessa. | JY |
| Talsi Noora | Kodin koneet. Teknologioiden kotouttaminen, käyttö ja vastustus. | UEF |
| Toivanen Mari | Negotiating home and belonging. Young Kurds in Finland. | TY |

| | | |
|-----------------------|--|-------------|
| Tuukkanen Terhi | A framework for children's participation in online environments. | JY |
| Wang Li | The Learning Experiences of Chinese Postgraduate Students at the University of Eastern Finland. | TaY |
| Villanen Heli | Our place, my future and their project. Reflecting children's lifeworld in education for sustainable development. | OY |
| Vuorinen Jukka | Parasitic Order Machine. A Sociology and Ontology of Information Securing. | TY |
| Åkerblad Leena | Epävarmuuden tuolla puolen. Muuttuvat työmarkkinat ja prekaari toimijuus. | UEF |
| 2015 | | |
| Anttila Erkki | Esikaupunkien vuosisata. Paikallinen yhteisöllisyys Helsingin seudun vanhoissa työväen esikaupungeissa 1900-luvulla. | HY |
| Bergholm Tapio | Kaksoissidoksen synty. Suomen työmarkkinasuhteiden muotoutuminen 1944-1969. | UEF |
| Berlin Jenni | Assimilated individuals and segregated communities. A comparative study of the housing and living related well-being of Finnish roma and housed gypsies and travellers in England. | UEF |
| Berntsen Lisa E. | Agency of labour in a flexible pan-European labour market. A qualitative study of migrant practices and trade union strategies in the Netherlands. | JY |
| Buchert Ulla | Maahanmuuttajuuden institutionaaliset kategoriat. | HY |
| Jaakkola Eevi | Yhteisöllisyys syrjäisellä maaseudulla asuvien ikääntyneiden arjessa. | LY |
| Kari-Björbacka Irmeli | Alueen ja työn sukupuoli. | LY |
| Kaukoma Timo | Facial expressions as an interactional resource in everyday face-to-face conversation. | HY |
| Kivijärvi Antti | Etnisyyden merkityksiä nuorten vertaissuhteissa. Tutkimus maahanmuuttajataustaisten ja kantaväestön nuorten kohtaamisista nuorisotyön kentillä. | UEF |
| Kulmalainen Taru | Tyttöjen kaverisuosio. Etnografinen tutkimus yläkoulun epävirallisista järjestyksistä. | UEF |
| Kuokkanen Anna | Johtamisen ihmissuhdekoulukunta Suomessa. Työntekijäkeskeiset johtamisopit suomalaisen työelämän muutoksessa. | HY |
| Könönen Jukka | Tilapäinen elämä, joustava työ. Rajat maahanmuuton ja työvoiman prekarisaation mekanismina. | UEF |
| Marionneau Virve | Socio-cultural contexts of gambling. A comparative study of Finland and France. | HY |
| Mäenpää Elina | Socio-economic homogamy and its effects on the stability of cohabiting unions. | HY |
| Ojajärvi Anni | Terve sotilas! Etnografinen tutkimus varusmiesten terveystajusta sosiaalisena ilmiönä. | HY |
| Peiponen Leila | Sosiaalisesti koettu ja jaettu työuupumus. Työuupumus puolison kokemuksena. | UEF |
| Pyrhönen Niko | The True Colors of Finnish Welfare Nationalism. Consolidation of Neo-Populist Advocacy as a Resonant Collective Identity through Mobilization of Exclusionary Narratives of Blue-and-White Solidarity. | HY |
| Ruuskanen Petri | Verkostotalous ja luottamus. | JY |
| Soininen Tiina | Changing expectations and realities of employment stability. Longitudinal analysis of tenures in Finland. | UE F |

| | | |
|-------------------|---|-----|
| Venäläinen Marina | Oikeusturvan rajat. Julkinen oikeusapu ja sovittelu siviilioikeudellisissa asioissa Venäjällä. | UEF |
| 2016 | | |
| Achté Päivi | Asiakaspalautteet osana arjen työtä sairaalassa. Kategorisointi ja selonteot kirurgian alalla. | TaY |
| Eranti Veikko | Individuals Doing Politics. Urban participation, social media campaigning and online nano-politics. | HY |
| Hart Linda | Relational Subjects. Family relations, law and gender in the European Court of Human Rights. | HY |
| Hautamäki Lotta | Movements of Moods. Interplay Between Science, Clinical Practice and Patient in Psychiatry. | HY |
| Issakainen Mervi | Youth depression. Young people's distress in relation to the cultural conceptions of depression. | UEF |
| Kortesoja Matti | Power of Articulation. Imagery of Society and Social Action in Structural Marxism and Its Critique. | TaY |
| Kosonen Toni | Opiskeleva ammattimies, yhteiskuntaluokka ja sukupuoli. Tutkimus ammatilliseen aikuiskoulutukseen osallistuvista työläismiehistä. | UEF |
| Krutova Oxana | Entailing Temporality. Longitudinal Analysis of Labor Market Integration of Immigrants in Finland. | TaY |
| Lattu Emmi | Naisten tekemä väkivalta. | TaY |
| Lundbom, Pia | Eläinten puolustajat. Suomalaisen eläinoikeusaktivismiin muuttuva poliittinen tyyli ja toiseus. | JY |
| Nisén Jessica | Education and fertility. A study on patterns and mechanisms among men and women in Finland. | HY |
| Rautio, Susanna | Neuvolan perhetyö vanhemmuuden varhaisena tukena ja yhteistyönä. | JY |
| Sirniö Outi | Constrained life chances. Intergenerational transmission of income in Finland. | HY |
| Syväterä Jukka | Creation and domestication of global policy trends. The case of national bioethics committees. | TaY |
| Tarkiainen Lasse | Income and Mortality - The Dynamics of Disparity. A Study on the Changing Association Between Income and Mortality in Finland. | HY |
| Virkki Tuija | Vihan voima. Toimijuus ja muutos vihakertomuksissa. | JY |
| Vähä-Savo Valteri | Sektoritutkimuksen genealogia. Komissioista ja toimistoista valtiolliseen tutkimusjärjestelmään. | TaY |
| 2017 | | |
| Alanko Anna | Improving mental health care. Finnish mental health policy rationale in the era of dehospitalisation. | HY |
| Alho Sirkka | Perheiden tukena. Tutkimus kasvatus- ja perheneuvolatyöstä ja kollektiivisesta identiteetistä asiantuntijoiden tulkitsemina. | UEF |
| Aro Riikka | Living standards and changing expectations. Investigating domestic necessity and environmental sustainability in an affluent society. | JY |
| Barzoki Meysam | Sexual Consumer Culture. Explanation of shopping and self-sexualizing behavior in the light of sexual objectification. | UEF |
| Kinnunen Veera | Tavarat tiellä. Sosiologinen tutkimus esinesuhteista muutossa. | LY |

| | | |
|--------------------------------|---|-------------|
| Obstbaum-Federley Yaira | From the social sector to selective individualized prison practices? A study on substance abuse among prisoners and its treatment. | HY |
| Ohisalo Maria | Murusia hyvinvointivaltion pohjalla. Leipäjonot, koettu hyvinvointi ja huono-osaisuus. | UE F |
| Otonkorpi-Lehtoranta Katri | Sotilaat, työ ja perhe. | TaY |
| Ranta Sirkku | Ammattiin kasvattamisen aikakuri. Kokin perustutkintokoulutuksen aika työläiseksi opettamisen tilana. | UEF |
| Saarikkomäki Elsa | Trust in public and private policing. Young people's encounters with the police and private security guards. | HY |
| Simonen Mika | A Conversation Analysis of Social Interaction in the Functional Capacity Interview. | HY |
| Ylä-Anttila Tuukka | The Populist Toolkit. Finnish Populism in Action 2007–2016. | HY |
| 2018 | | |
| Kuoppamäki Sanna | The role of age and life course stage in digital consumption. | JY |
| Kuukkanen Mari | Anarkistien keinot ja päämäärät. Tutkimus suomalaisesta anarkistiliikkeestä 2010-luvun alussa. | HY |
| Laihiala Tuomo | Kokemuksia ja käsityksiä leipäjonoista. Huono-osaisuus, häpeä ja ansaitsevuus. | UE F |
| Laitinen Hanna | Miten ja kenelle valtakunnalliset nuorisjärjestöt legitimoivat toimintansa? | JY |
| Liukkonen Marjo | Hennalan naismurhat 1918. | LY |
| Rannikko Anni | Kamppailua kunnioituksesta. Vaihtoehtoliikunnan alakulttuurien moraaliset järjestykset. | UEF |
| Riekkinen-Tuovinen Sointu | Sosiokulttuurista vanhustyötä paikantamassa. Tutkimus erilaisissa vanhusten asumisympäristöissä. | UEF |
| Selander Kirsikka | Työhyvinvoinnin paradoksit kolmannen sektorin palkkatyössä. | JY |
| Tuomivaara Salla | Searching for the roots of exclusion. Animals in the sociologies of Westermarck and Durkheim. | TaY |

LIITE 2 Liitetaulukot

Liitetaulukko 1 Faktorointimenetelmien esiintyminen tilastollisten ohjelmien mukaan

| | laskettu itse | keskustietokone | GLIM | SAS | SPSS | R | ei tietoa | Yhteensä |
|--------------------------------|---------------|-----------------|----------|----------|-----------|----------|-----------|-----------|
| sentroidi | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| pääakseli | 1 | 9 | 0 | 2 | 3 | 0 | 4 | 19 |
| pääkomponentti | 0 | 5 | 0 | 1 | 5 | 0 | 4 | 15 |
| pääakseli ja pääkomponentti | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 2 |
| ei mainittu | 1 | 6 | 1 | 0 | 5 | 0 | 3 | 16 |
| PNS | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 2 |
| Yhteensä | 4 | 21 | 1 | 3 | 14 | 1 | 12 | 56 |

Liitetaulukko 2 Suorien ja vinojen rotaatioiden esiintyminen eri aikoina tilastollisten ohjelmien mukaan

| | 1960 - 1985 | | 1986 - 2018 | | | | ohjelmasta ei tietoa | Väitöskirjat yhteensä |
|-----------------|---------------|-----------------|-------------|----------|-----------|----------|-------------------------|--------------------------|
| | laskettu itse | keskustietokone | GLIM | SAS | SPSS | R | | |
| suora | 1 | 14 | 1 | 2 | 7 | 1 | 4 | 30 |
| vino | 3 | 1 | 0 | 1 | 5 | 0 | 2 | 12 |
| molempia | 0 | 5 | 0 | 0 | 1 | 0 | 1 | 7 |
| ei tietoa | 0 | 1 | 0 | 0 | 1 | 0 | 5 | 7 |
| Yhteensä | 4 | 21 | 1 | 3 | 14 | 1 | 12 | 56 |

Liitetaulukko 3 Faktoriaalyysin jatkotoimenpiteet eri aikoina tilastollisen ohjelmien mukaan

| | 1960 - 1985 | | 1986 - 2018 | | | Ohjelmasta ei tietoa | Väitöskirjat yhteensä |
|---|------------------|-----------------|-------------|-----------|----------|-------------------------|--------------------------|
| | laskettu itse | keskustietokone | GLIM/SAS | SPSS | R | | |
| transformaatioanalyysi | 1 | 1 | 0 | 0 | 0 | 2 | 4 |
| faktoripistemäärät | 0 | 10 | 1 | 2 | 1 | 4 | 18 |
| summamuuttujat | 0 | 5 | 3 | 9 | 0 | 5 | 22 |
| summamuuttujat ja faktoripistemäärät | 0 | 0 | 0 | 2 | 0 | 0 | 2 |
| ei mainittu | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| ei jatkoa | 3 | 5 | 0 | 0 | 0 | 1 | 9 |
| Yhteensä | 4 | 21 | 4 | 14 | 1 | 12 | 56 |