



**TURUN
YLIOPISTO**

**RAKENNEYHTÄLÖMALLINNUS PUUTTUVUUTTA SISÄLTÄVÄLLÄ AINEISTOLLA -
TAPAUK DIGITALWELLS**

Joanna Carlsson

Pro Gradu -tutkielma
Kesäkuu 2023

Tarkastajat:
Prof. Kari Auranen
Lehtori Jouko Katajisto

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck-järjestelmällä.

JOANNA CARLSSON: Rakenneyhtälömallinnus puuttuvuutta sisältävällä aineistolla - tapaus
DigitalWells
Pro Gradu , 39 s., 7 liites.
Tilastotiede
Kesäkuu 2023

TIIVISTELMÄ

DigitalWells-ohjelma tutki ja kehitti vuosina 2019–2022 liikuntateknologiaa ja hyviä käytänteitä, joiden avulla pyrittiin ennaltaehkäisemään nuorten ikäihmisten (60–75-vuotiaiden) terveyden kannalta riittämättömän liikunnan aiheuttamia seurauksia. Osallistujat saivat käyttöönsä ilmaisen DigitalWells-sovelluksen, jonka avulla he pystyivät tallentamaan sekä seuraamaan omaa fyysistä aktiivisuuttaan. Ohjelman eräänä tutkimuksen kohteena oli uuden liikuntateknologian hyväksyminen ja käyttöaikomus. Tutkimusaineistoa kerättiin sekä DigitalWells-sovelluksen että kyselytutkimusten avulla.

Yhdistetty teoria teknologian käyttöönotosta ja käytöstä (UTAUT) on Venkateshin ja kollegoiden vuonna 2003 esittämä teoreettinen malli neljästä päätekijästä, jotka kuvastavat yksilön uutta teknologiaa koskevia ennakkokäsityksiä. Yhdessä nämä neljä päätekijää selittävät uuden teknologian käyttöönottoa ja täten myös sen käyttöä. Mobiiliteknologian nopean kehityksen myötä UTAUT-mallia on päivitetty ja laajennettu koskemaan myös kuluttajia, kun alkuperäistä vuoden 2003 teoriaa sovellettiin lähinnä työntekijän näkökulmasta.

Tutkielmassa tarkastellaan rakenneyhtälömallinnuksen sopivuutta analyysimenetelmänä, kun teoreettisena mallina on muunnettu UTAUT2-malli. Teoriaosassa esitetään rakenneyhtälömallinnuksen ensisijaiset ominaisuudet sekä estimointimenetelmät. Lisäksi tarkastellaan yleisiä käytäntöjä, joilla rakenneyhtälömallin sopivuutta arvioidaan. Teoriaosassa käydään myös läpi puuttuvan tiedon mekanismeja sekä esitetään rakenneyhtälömallinnukselle suotuisia menetelmiä puuttuvuuden hallinnointiin.

Tutkielman soveltavassa osassa DigitalWells-ohjelman kyselytutkimuksen osa-aineistoon sovitetaan UTAUT2-teoriaa hyödyntävä rakenneyhtälömalli. Rakenneyhtälömallinnuksen avulla tarkastellaan ohjelmaan osallistuneiden nuorten ikäihmisten uuden liikuntasovelluksen hyväksymistä. Tutkielmassa huomioidaan myös kyselyaineistossa esiintyvä puuttuvuus, jota korjataan rakenneyhtälömallinnukselle suotuisilla imputointimenetelmillä sekä valituilla analyysimenetelmillä. Lopuksi eri imputointi- ja analyysimenetelmillä tuotettujen rakenneyhtälömallien sopivuutta arvioidaan rakenneyhtälömallinnukselle ominaisilla indikaattoreilla.

Parametrisella ja epäparametrisella menetelmällä saadut tulokset ovat samansuuntaisia. Tutkielman muunnettu UTAUT2-malli sopii tutkittavaan DigitalWells-kyselyaineiston osa-aineiston kohtuullisesti.

Asiasanat: Rakenneyhtälömallinnus, puuttuvuus, UTAUT, nuoret ikäihmiset

Sisällys

1	Johdanto	1
2	Rakenneyhtälömallinnus	3
2.1	Mittamalli	3
2.2	Rakennemalli	3
2.3	Mallinnuksen kolme yhtälöä	4
2.4	Parametrien estimointi	7
2.5	Mallin sopivuustarkastelut	9
2.6	Yleiset sopivuusmittarit	10
3	Puuttuvuus	12
3.1	Puuttuvan tiedon lajit	12
3.2	Puuttuvan tiedon tunnistaminen	13
3.3	Puuttuvuuden hallinta	14
3.4	Rakenneyhtälömalleille suotuisat menetelmät	14
3.4.1	Ehdollistamaton keskiarvoimputointi	15
3.4.2	Moni-imputointi	15
3.4.3	Satunnaismetsä	15
3.4.4	FIML	16
3.4.5	PLS-SEM	16
3.5	Imputointimenetelmän valinta	17
4	DigitalWells-ohjelma	19
4.1	Tutkimusasetelma	19
4.2	Kyselytutkimuksen aineisto	20
5	Empiirinen osio	22
5.1	Tutkimuskysymykset ja -hypoteesit	24
5.2	Havaintoaineiston tarkastelu	25
5.3	Tarkasteltavat rakenneyhtälömallit	26
5.3.1	Mallit 1a ja 1b – täydelliset havaintorivit	27
5.3.2	Mallit 2a ja 2b – ehdollistamaton keskiarvoimputointi	27
5.3.3	Mallit 3a ja 3b – moni-imputointi	28
5.3.4	Mallit 4a ja 4b – FIML	28
5.3.5	Mallit 5a ja 5b – PLS-SEM	28
5.4	Mallien sopivuus aineistoon	31
5.4.1	Parametrisella menetelmällä saadut estimaatit	31
5.4.2	Epäparametrisella menetelmällä saadut estimaatit	32
6	Johtopäätökset ja pohdinta	35
	Lähdeluettelo	36
	Liite A	40

1 Johdanto

DigitalWells-ohjelma (2019–2022) tutki aktiivista liikuntaa tukevaa teknologiaa sekä loi hyviä käytäntöjä, joiden avulla pyrittiin tukemaan nuorten ikäihmisten (60–75-vuotiaiden) liikunnallista aktiivisuutta (DigitalWells, 2019). DigitalWells-ohjelman eräänä taustatekijänä oli kansallisesti ja globaalisti kasvava huolenaihe ihmisten terveyden kannalta liian vähäisestä liikkumisesta. Suomen väestöstä nuoria ikäihmisiä – ohjelman kohderyhmäläisiä – oli vuonna 2019 20,2 % eli 1,1 miljoonaa henkilöä (Tilastokeskus, 2022).

UKK-instituutin vuonna 2022 julkaisema liikuntaraportti osoittaa, että suomalaisten liikunnallisuus eri ikäryhmissä on vähentynyt verrattuna aiempien vuosien tuloksiin (Husu ym., 2022). Liikuntamotivaation nostamiseksi on esitetty erilaisia keinoja, esimerkiksi liikunnan pelillistämistä (Koivisto & Hamari, 2019) sekä älyllisempää liikuntateknologiaa (Carlsson ym., 2021).

Yhdistetty teoria teknologian hyväksymisestä ja käytöstä (*Unified Theory of Acceptance and Use of Technology*, UTAUT) on Venkateshin ja kollegoiden alunperin vuonna 2003 julkaisema ja sittemmin muokkaama teoreettinen malli (Venkatesh ym., 2012, 2016). UTAUT-malli yhdistää kahdeksan eri teoriaa, joissa jokaisessa tarkastellaan hieman eri tavalla (informaatio)teknologian hyväksymistä, aikomusta käyttää teknologiaa ja/tai itse teknologian käyttöä (Venkatesh ym., 2003). Kun tutkitaan DigitalWells-ohjelmaan osallistuneiden nuorten ikäihmisten liikuntateknologian hyväksymistä ja käyttöä, tarkastelu voidaan toteuttaa hyödyntämällä teoreettisen UTAUT-mallin uudempaa versiota eli UTAUT2-mallia (Makkonen ym., 2020; Carlsson ym., 2021).

Rakenneyhtälömallinnus (*Structural Equation Modeling*, SEM) käsitteenä tarkoittaa analyysimenetelmää, jonka avulla voidaan samanaikaisesti sekä tiivistää havaintoaineistoa että tarkastella muuttujien välisiä riippuvuussuhteita ja rakenteita. Perinteisesti rakenneyhtälömallinnukseksi mielletään vahvistava (*konfirmatorinen*) faktorianalyysi yhdessä *polkuanalyysin* kanssa. Käytännössä rakenneyhtälömallinnuksessa voidaan käyttää useita erilaisia monimuuttujamenetelmiä, esimerkiksi monen vasteen regressiota tai monen selittäjän regressiota sekä piiloluokka-, polku- ja faktorianalyysia. Valittuja analyysimenetelmiä sovelletaan niin, että mallissa mukana olevien muuttujien tarkastelu tapahtuu samanaikaisesti.

Rakenneyhtälömallinnuksessa verrataan teoreettisen mallin kovarianssimatriisia havaintoaineiston otoskovarianssimatriisiin. Vertailun voi myös tehdä korrelaatiomatriisien avulla, mutta yleisimmät rakenneyhtälömallinnuksen ohjelmat, kuten LISREL, MPlus ja AMOS, suosivat kovarianssimatriisia (Ramlall, 2017; Schumacker & Lomax, 2016). Mitä huolellisemmin rakenneyhtälömallinnuksen teoreettinen malli on luotu, sitä virheettömämmin sen pitäisi sopia havaintoaineistoon (Byrne, 2012; Halme ym., 2014; Vehkalahti & Everitt, 2019; Nordberg, 2006).

Analyysimenetelmänä rakenneyhtälömallinnusta voidaan käyttää monipuolisesti ja joustavasti, sillä se soveltuu niin yleisen regressiomallin kuin myös monimutkaisemman hierarkisen mallin tarkasteluun. Tehokkaimmillaan rakenneyhtälömallinnus sekä tiivistää havaintoaineistoa että tarkastelee mallin selittäviä ja selitettäviä *piilomuuttujia* yhdessä muuttujien mahdollisten mittausvirheiden kanssa (Vehkalahti, 2019; Halme ym., 2014). Piilomuuttuja eli *latentti* tai *faktori* on muuttuja, joka kuvaa sellaista käsitettä, jota ei voida suoraan mitata. Esimerkiksi kyselytutkimuksessa opiskelijan opiskelumotivaatiota voidaan tarkastella käyttämällä samanaikaisesti useampaa, mutta toisistaan hieman eriä-

vää, opiskelumotivaatiota käsittelevää väittämää. Yhdessä kyseiset väittämät muodostavat opiskelumotivaatiota kuvaavan faktorin.

UTAUT2-malli sisältää perättäisiä, osin jonomaisesti kuvattuja muuttujia, jotka ovat kaikki faktoreita. Mallin faktorit ovat joko selittäviä tai selitettäviä muuttujia ja mallin jonomaisesti, vasemmalta oikealle etenevää rakennetta sanotaan *polkumalliksi*. Polkumalli on muuttujien välisiä syy–seuraus-suhteita kuvaava malli, jonka yleisenä analyysimenetelmänä käytetään regressiopohjaista polkuanalyysia. Rakenneyhtälömallinnuksessa käytetään sekä faktori- että polkuanalyysia, joten UTAUT2-mallin analyysimenetelmänä rakenneyhtälömallinnus on hyvä vaihtoehto, etenkin kun muuttujien tarkastelu voidaan tehdä samanaikaisesti. Jos havaintoaineisto sisältää puuttuvuutta tai havaintoja on vähän, yleensä edellytetään tehokkaita (*robusteja*) menetelmiä. Mikäli oletus muuttujien normaali-jakautuneisuudesta ei toteudu, rakenneyhtälömallin analyysivaiheessa tämä rajoite täytyy huomioida.

Tässä tutkielmassa tarkastellaan rakenneyhtälömallinnuksen sopivuutta analyysimenetelmänä, kun aineistossa on puuttuvuutta. Tarkastelun yhteydessä aineiston puuttuvuutta käsitellään eri tavoin. Empiirisen osion aineistona käytetään vuosina 2019–2020 kerättyä DigitalWells-ohjelman toisen aallon kyselytutkimusaineistoa. Luvussa 2 esitetään rakenneyhtälömallinnuksen yleinen muoto sekä käytettävät estimointimenetelmät. Aineistossa esiintyvää puuttuvuutta ja sen käsittelyä tarkastellaan luvussa 3. Luvussa 4 esitetään DigitalWells-ohjelman tutkimusasetelma sekä tutkielman aineisto. Tutkielman aineiston analyysiin sopiva rakenneyhtälömalli tuotetaan sekä arvioidaan luvussa 5, ja tutkielman huomiot sekä johtopäätökset esitetään luvussa 6.

Tämä tutkielma rajoittuu rakenneyhtälömallin sopivuuden arvioimiseen, kun teoreettisena mallina on muokattu UTAUT2-malli ja käytettävissä oleva kyselyaineisto sisältää puuttuvuutta. Kyselyaineistoa ei ole tuotettu satunnaisotoksella eettisistä syistä johtuen. Täten tutkielmassa esitettyjä johtopäätöksiä ei sellaisenaan voida yleistää käsittämään suurempaa joukkoa, vaan ne koskevat ainoastaan DigitalWells-ohjelman toisessa aallossa mukana olleita nuoria ikäihmisiä. Kuitenkin saadut tulokset antavat tietoa, jota voidaan hyödyntää myöhemmissä tutkimuksissa.

2 Rakenneyhtälömallinnus

Analyysimenetelmänä rakenneyhtälömallinnus ei ole uusi, sillä jo 1930-luvulla Hotelling ja Thurstone edistivät pääkomponentti- ja faktorianalyysia tilastollisina menetelminä (Karimi & Meyer, 2014; Schumacker & Lomax, 2016; Westland, 2019). Trygve Haavelmo vei 1940-luvulla eteenpäin polkuanalyysin kehitystyötä, mikä hyödytti Karl Jöreskogia 1960-luvun loppupuolella *linear structural relations* eli LISREL-menetelmän – ja myöhemmin ohjelmiston – rakentamisessa. Herman Wold yhdisti 1960-luvulla osittaisen pienimmän neliösumman menetelmän polkuanalyysiin, jota seurasi myöhemmin varianssipohjainen rakenneyhtälömallinnus (*PLS-SEM*). Näin ollen rakenneyhtälömallinnus sekä käsittää että laajentaa regression ja faktorianalyysin menetelmiä (Bollen, 1989; Westland, 2019).

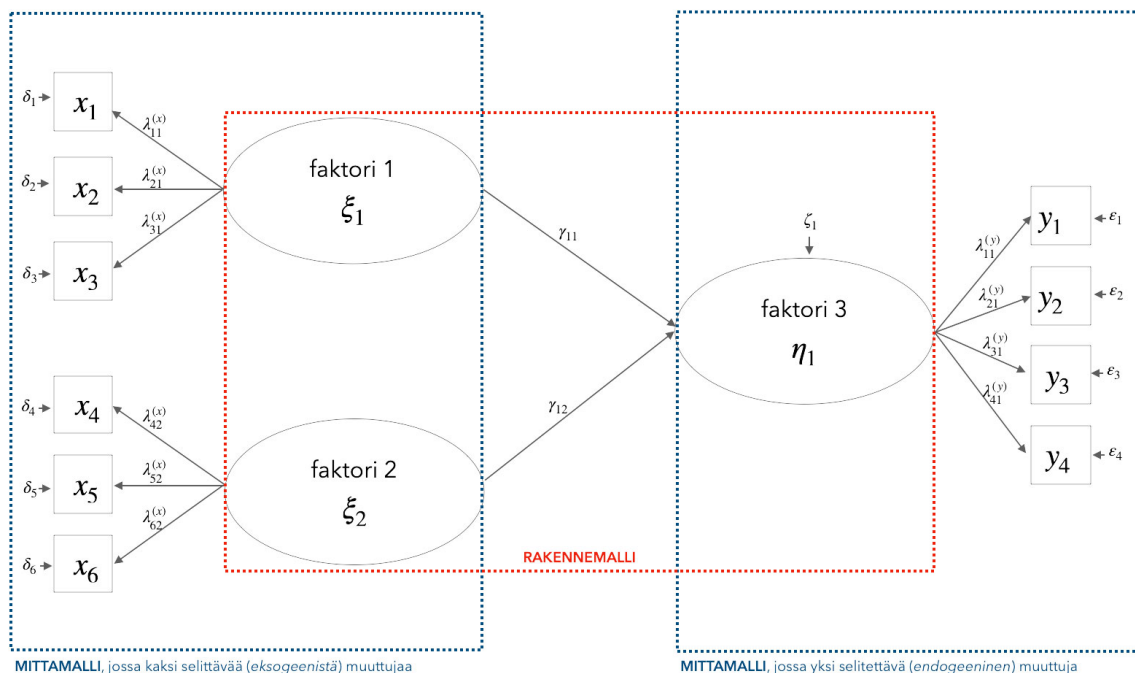
2.1 Mittamalli

Luvussa 1 todettiin, että rakenneyhtälömallinnusta käyttämällä saadaan havaintoaineisto muutettua kompaktimpaan muotoon faktorien avulla. Faktorin yhteyttä havattuihin muuttujiin kuvataan *mittamallin* avulla. Mittamallin graafisessa esitystavassa faktorit esitetään ellipseinä tai ympyröinä ja havaitut muuttujat puolestaan suorakulmioina tai neliöinä (kuva 1). Muuttujien väliset yhteydet kuvataan yksisuuntaisilla nuolilla siten, että faktorista lähtevän nuolen kärki osoittaa havaittua muuttujaa. Yksinkertaisimmillaan mallin jokainen havaittu muuttuja on yhteydessä vain yhteen faktoriin kerrallaan. Tätä esitystapaa tukee myös mittamallin yleisimmin käytetty analyysimenetelmä, joka on konfirmatorinen faktorianalyysi (*Confirmatory Factor Analysis*, CFA).

Konfirmatorinen faktorianalyysi Yleisemmin faktorianalyysi on kokeilevaa eli *eksploraatiivista*, jolloin havaintoaineistoa käsittelemällä, esimerkiksi rotatoimalla, saadaan esille taustalla vaikuttavat faktorit ja niiden lukumäärät. Eksploraatiivisesta menetelmästä poiketen konfirmatorisessa faktorianalyysissa faktoreiden yhteys havattuihin muuttujiin määritellään etukäteen, kuten myös tarkasteltavien faktorien lukumäärä. Tilastollisen analyysin tavoitteena on vahvistaa ennalta määrättyt faktorit ja täten rakenneyhtälömallin mittamalli(t). Konfirmatorinen faktorianalyysi on analyysimenetelmänä vahvistava.

2.2 Rakennemalli

Kun rakenneyhtälömallinnuksessa esitetään faktorien välisiä riippuvuuksia, kuvataan *rakennemallia*. Rakennemallissa faktoreita esittävät ellipsit tai ympyrät ovat toisiinsa yhteydessä yksisuuntaisilla nuolilla. Nuolen lähtöpisteenä toimiva faktori on selittävä eli *eksogeeninen* faktori. Muuttuja, johon nuolen kärki kohdistuu, on puolestaan selitettävä eli *endogeeninen* faktori. Selitettävään faktoriin liittyy aina jäännösvirhe eli *residuaali*. Jäännösvirhettä kuvataan rakennemallissa pienellä nuolella, joka sijoitetaan selitettävän faktorin päälle (kuva 1). Havaittuun muuttujaan liittyy mittavirhe, joka esitetään myös pienellä nuolella. Mittavirheen piirtäminen mittamallissa olevan havaitun muuttujan yhteyteen on osin vapaaehtoista. Vaikka mittavirhettä ei rakenneyhtälömallin graafisessa esityksessä ole käytetty, mallitarkastelun yhteydessä se kuitenkin huomioidaan.



Kuva 1: Eräs rakenneyhtälömalli, joka käsittää kaksi mittamallia (sininen katkoviivallinen laatikko) sekä yhden rakennemallin (punainen katkoviivallinen laatikko). Selitettävä faktori 3 (η_1) sisältää aina jäännösvirheen (ζ_1). Yksinkertaisuuden vuoksi jokaista faktoria kohden on esitetty vain kolme tai neljä havaittua muuttujaa.

Polkuanalyysi Rakennemallin monimuuttujamenetelmänä käytetään yleensä polkuanalyysia (*Path Analysis*), joka on usean muuttujan lineaarinen regressioanalyysi. Polkuanalyysissa muuttujien väliset riippuvuussuhteet ovat sisäkkäisiä ja jonomaisia; niillä on etenemissuunta. Muuttujien välisiä yhteyksiä kuvaavat yksisuuntaiset nuolet kertovat syy–seuraus-suhteesta eli *kausaalisuudesta*. Rakenneyhtälömallinnuksessa polkuanalyysin avulla tarkastellaan faktorien välisiä kausaalisuuksia ja samalla ennakko-oletusten perusteella luodun rakennemallin todenmukaisuutta.

2.3 Mallinnuksen kolme yhtälöä

Rakenneyhtälömallinnuksen määrittävät kolme yhtälöä. Tyypillisesti nämä yhtälöt esitetään käyttämällä Jöreskogin (1972) kehittämän LISREL-ohjelman notaatiotapaa. Taulukossa 1 on esitetty rakenneyhtälömallinnuksessa käytettävien parametrien LISREL-notaatiot sekä kuvaukset.

Hyödyntämällä Vehkalahden (2020) ja Nordbergin (2006) opetusmateriaalia, voidaan havaittujen muuttujien sekä selittävien faktorien välinen yhteys kuvata yhtälön avulla

$$\mathbf{x} = \mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \quad (1)$$

jossa vektori $\mathbf{x}$ kuvaa havaittuja muuttujia, matriisi $\mathbf{\Lambda}_x$ faktorilatauksia, vektori $\boldsymbol{\xi}$ selittävien faktorien vektoria ja vektori $\boldsymbol{\delta}$ havaittujen muuttujien mittavirheitä. Edellytyksenä yhtälön (1) toteutumiselle on oletus, että mittavirheet δ_i eivät korreloi faktorien ξ_j kanssa. Lisäksi oletetaan, että odotusarvot $E(\boldsymbol{\xi}) = 0$ ja $E(\boldsymbol{\delta}) = 0$ toteutuvat (Bollen, 1989). Kuvan 1 vasemmanpuoleinen mittamalli on yhtälön (1) eräs graafinen esitystapa.

Vastaavasti voidaan muodostaa yhtälö kuvaamaan havaittujen muuttujien ja selitettävien faktorien välistä yhteyttä. Yhtälössä

$$\mathbf{y} = \mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (2)$$

vektori $\mathbf{y}$ kuvaa havaittuja muuttujia, matriisi $\mathbf{\Lambda}_y$ faktorilatauksia, vektori $\boldsymbol{\eta}$ selitettävien faktorien vektoria ja vektori $\boldsymbol{\varepsilon}$ havaittujen muuttujien mittavirheitä. Myös yhtälöön (2) liittyy oletus, että mittavirheet ε_i eivät korreloi faktorien η_j kanssa ja että molemmat odotusarvot $E(\boldsymbol{\eta}) = 0$, $E(\boldsymbol{\varepsilon}) = 0$ toteutuvat (Bollen, 1989). Yhtälöä (2) vastaava eräs graafinen esitystapa on kuvassa 1 oleva oikeanpuoleinen mittamalli.

Malleihin (1) ja (2) liittyvät lisäksi seuraavat oletukset (Bollen, 1989):

- a) Mittavirheet δ ja ε eivät korreloi keskenään.
- b) Mittavirheet δ eivät korreloi selitettävien faktorien $\boldsymbol{\eta}$ kanssa eivätkä mittavirheet ε korreloi selittävien faktorien $\boldsymbol{\xi}$ kanssa.
- c) Mittavirheet δ_i ja ε_i ovat homoskedastisia.

Rakenneyhtälömallinnuksen kolmas yhtälö kuvaa rakennemallia. Yhtälössä

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \quad (3)$$

vektori $\boldsymbol{\eta}$ käsittää selitettävät faktorit ja matriisi $\mathbf{B}$ selitettävien faktorien väliset regressiokertoimet. Matriisi $\mathbf{\Gamma}$ puolestaan sisältää selittävien ja selitettävien faktorien väliset regressiokertoimet. Vektori $\boldsymbol{\xi}$ on selittävien faktorien vektori ja vektori $\boldsymbol{\zeta}$ kuvaa selitettäviin faktoreihin liittyviä jäännösvirheitä eli *residuaaleja*. Kuvan 1 rakennemalli on yhtälön (3) eräs graafinen esitystapa.

Yhtälöön (3) kohdistuu myös oletuksia (Bollen, 1989):

- a) Jäännösvirheet ζ eivät korreloi faktorien $\boldsymbol{\xi}$ kanssa
- b) $E(\boldsymbol{\zeta}) = 0$
- c) ζ_i on homoskedastinen.

Yleensä termillä rakenneyhtälömalli viitataan malliin, jossa on sekä mittamalli että rakennemalli. Kuitenkin rakenneyhtälömallinnus käsitteenä on laaja, joten myös termiä rakenneyhtälömalli käytetään vaihtelevasti. Rakenneyhtälömalli voi myös sisältää yhden tai useamman *moderoivan* muuttujan. Tällöin kahden faktorin välinen yhteys ei ole vakio, vaan riippuu siihen vaikuttavan moderoivan muuttujan saamista arvoista.

Taulukko 1: LISREL-notaatio

Merkintä	Vektori Matriisi	Dimensio	Selite
$x_i, i = 1, \dots, q$	$\mathbf{x}$	$q \times 1$	Selittävään faktoriin ξ_j yhteydessä olevat havaitut muuttujat.
$y_i, i = 1, \dots, p$	$\mathbf{y}$	$p \times 1$	Selitettävään faktoriin η_j yhteydessä olevat havaitut muuttujat.
$\delta_i, i = 1, \dots, q$	$\boldsymbol{\delta}$	$q \times 1$	Havaittuun muuttujaan x_i liittyvä mittavirhe.
$\varepsilon_i, i = 1, \dots, p$	$\boldsymbol{\varepsilon}$	$p \times 1$	Havaittuun muuttujaan y_i liittyvä mittavirhe.
$\xi_j, j = 1, \dots, n$	$\boldsymbol{\xi}$	$n \times 1$	Selittävä faktori.
$\eta_j, j = 1, \dots, m$	$\boldsymbol{\eta}$	$m \times 1$	Selitettävä faktori.
$\zeta_i, i = 1, \dots, m$	$\boldsymbol{\zeta}$	$m \times 1$	Selitettävään faktoriin η_j liittyvä jäännösvirhe, <i>residuaali</i> .
$\lambda_{ij}^{(x)}$	$\boldsymbol{\Lambda}_x$	$q \times n$	Selittävän faktorin faktorilataukset.
$\lambda_{ij}^{(y)}$	$\boldsymbol{\Lambda}_y$	$p \times m$	Selitettävän faktorin faktorilataukset.
	$\boldsymbol{\Gamma}$	$m \times n$	Selittävien ja selitettävien faktorien väliset regressiokertoimet.
	$\boldsymbol{B}$	$m \times m$	Selitettävien faktorien väliset regressiokertoimet.
	$\boldsymbol{\Phi}$	$n \times n$	Selittävien faktorien kovarianssimatriisi.
	$\boldsymbol{\Theta}_\delta$	$q \times q$	Havaittujen muuttujien x_i mittavirheiden δ_i kovarianssimatriisi.
	$\boldsymbol{\Theta}_\varepsilon$	$p \times p$	Havaittujen muuttujien y_i mittavirheiden ε_i kovarianssimatriisi.
	$\boldsymbol{\Psi}$	$m \times m$	Jäännösvirheiden ζ_i kovarianssimatriisi.

2.4 Parametrien estimointi

Rakenneyhtälömallinnuksen keskeisin parametri on kovarianssi (Kline, 2016). Kovarianssi kuvaa kahden tai useamman satunnaismuuttujan yhteisvaihtelua ja ilmentää muuttujien välisiä riippuvuussuhteita. Tarkastellaan kuvassa 1 esiintyvien havaittujen muuttujien x_1 ja x_2 kovarianssia. Molemmat havaitut muuttujat latautuvat samalle faktorille ξ_1 . Täten

$$\begin{aligned}\text{Cov}(x_1, x_2) &= \text{Cov}[(\lambda_{11}\xi_1 + \delta_1), (\lambda_{21}\xi_1 + \delta_2)] \\ &= \lambda_{11} \cdot \text{Var}(\xi_1) \cdot \lambda_{21},\end{aligned}$$

kun yhtälön (1) oletukset toteutuvat. Vastaavasti saadaan kuvan 1 selittäviä faktoreita (ξ_1 ja ξ_2) käsittävän mittamallin kovarianssimatriisi yhtälön (1) avulla:

$$\begin{aligned}\text{Cov}(\mathbf{x}, \mathbf{x}) &= \text{E}(\mathbf{x}\mathbf{x}^T) \\ &= \text{E}[(\mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta})(\mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta})^T] \\ &= \text{E}[(\mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta})(\boldsymbol{\xi}^T \mathbf{\Lambda}_x^T + \boldsymbol{\delta}^T)] \\ &= \mathbf{\Lambda}_x [\text{E}(\boldsymbol{\xi} \boldsymbol{\xi}^T)] \mathbf{\Lambda}_x^T + \mathbf{\Lambda}_x [\text{E}(\boldsymbol{\xi} \boldsymbol{\delta}^T)] + [\text{E}(\boldsymbol{\delta} \boldsymbol{\xi}^T)] \mathbf{\Lambda}_x^T + \text{E}(\boldsymbol{\delta} \boldsymbol{\delta}^T) \\ &= \mathbf{\Lambda}_x \boldsymbol{\Phi} \mathbf{\Lambda}_x^T + \mathbf{\Lambda}_x [\text{E}(\boldsymbol{\xi} \boldsymbol{\delta}^T)] + [\text{E}(\boldsymbol{\delta} \boldsymbol{\xi}^T)] \mathbf{\Lambda}_x^T + \boldsymbol{\Theta}_\delta,\end{aligned}$$

josta saadaan edelleen

$$\boldsymbol{\Sigma}_{xx} = \mathbf{\Lambda}_x \boldsymbol{\Phi} \mathbf{\Lambda}_x^T + \boldsymbol{\Theta}_\delta, \quad (4)$$

kun yhtälön (1) oletukset toteutuvat.

Yhtälön (2) avulla saadaan muodostettua kuvan 1 selitettävän faktorin (η_1) sisältävän mittamallin kovarianssimatriisi:

$$\begin{aligned}\text{Cov}(\mathbf{y}, \mathbf{y}) &= \text{E}(\mathbf{y}\mathbf{y}^T) \\ &= \text{E}[(\mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon})(\mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon})^T] \\ &= \text{E}[(\mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon})(\boldsymbol{\eta}^T \mathbf{\Lambda}_y^T + \boldsymbol{\varepsilon}^T)] \\ &= \mathbf{\Lambda}_y [\text{E}(\boldsymbol{\eta} \boldsymbol{\eta}^T)] \mathbf{\Lambda}_y^T + \mathbf{\Lambda}_y [\text{E}(\boldsymbol{\eta} \boldsymbol{\varepsilon}^T)] + [\text{E}(\boldsymbol{\varepsilon} \boldsymbol{\eta}^T)] \mathbf{\Lambda}_y^T + \text{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) \\ &= \mathbf{\Lambda}_y [\text{E}(\boldsymbol{\eta} \boldsymbol{\eta}^T)] \mathbf{\Lambda}_y^T + \boldsymbol{\Theta}_\varepsilon,\end{aligned}$$

kun yhtälön (2) oletukset toteutuvat. Hyödyntämällä yhtälöä (3) saadaan ratkaistua selitettävien faktorien kovarianssi $\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\eta})$:

$$\begin{aligned}\text{E}(\boldsymbol{\eta} \boldsymbol{\eta}^T) &= \text{E}[((\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta}))((\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta}))^T] \\ &= \text{E}[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma} \boldsymbol{\xi} + \boldsymbol{\zeta})(\boldsymbol{\xi}^T \boldsymbol{\Gamma}^T + \boldsymbol{\zeta}^T)(\mathbf{I} - \mathbf{B})^{-T}] \\ &= (\mathbf{I} - \mathbf{B})^{-1} [\boldsymbol{\Gamma} \text{E}(\boldsymbol{\xi} \boldsymbol{\xi}^T) \boldsymbol{\Gamma}^T + \boldsymbol{\Gamma} \text{E}(\boldsymbol{\xi} \boldsymbol{\zeta}^T) + \text{E}(\boldsymbol{\zeta} \boldsymbol{\xi}^T) \boldsymbol{\Gamma}^T + \text{E}(\boldsymbol{\zeta} \boldsymbol{\zeta}^T)] (\mathbf{I} - \mathbf{B})^{-T} \\ &= (\mathbf{I} - \mathbf{B})^{-1} (\boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Gamma}^T + \boldsymbol{\Psi}) (\mathbf{I} - \mathbf{B})^{-T},\end{aligned}$$

kun yhtälön (3) oletukset toteutuvat. Lisäksi oletetaan, että matriisi $(\mathbf{I} - \mathbf{B})$ on epäsingulaarinen. Tällöin selitettäviä havaittuja muuttujia käsittävän mittamallin kovarianssimatriisi on muotoa:

$$\Sigma_{yy} = \Lambda_y(I - B)^{-1}(\Gamma\Phi\Gamma^T + \Psi)(I - B^T)^{-1}\Lambda_y^T + \Theta_\varepsilon. \quad (5)$$

Yhtälöiden (4) ja (5) mukaisesti saadaan mittamallien havaittujen muuttujien x ja y kovarianssimatriisi:

$$\Sigma_{xy} = \Lambda_x\Phi\Gamma^T(I - B^T)^{-1}\Lambda_y^T. \quad (6)$$

Havaittujen muuttujien x ja y kovarianssimatriisi voidaan esittää 2×2 lohkomatriisina (Vehkalahti, 2020):

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}, \quad (7)$$

Lohkomatriisin 7 oikean ylälohkon kovarianssimatriisi Σ_{yx} on kovarianssimatriisin Σ_{xy} transpoosi.

Rakenneyhtälömallinnuksen tavoitteena on tuottaa yhtälön 7 mukainen kovarianssimatriisi Σ . Jos rakenneyhtälömallinnuksessa käytettävä rakenneyhtälömalli on täysin oikea, mallin kovarianssimatriisi vastaa havaittujen muuttujien kovarianssimatriisia (Bollen, 1989; Byrne, 2012; Halme, 2014). Rakenneyhtälömallinnuksen keskeinen tavoite, kovarianssimatriisin mallintaminen joukolla parametreja θ voidaan ilmaista yhtälön (7) perusteella seuraavasti:

$$\Sigma = \Sigma(\theta) = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix},$$

jossa $\Sigma(\theta)$ kuvaa kovarianssimatriisia parametrin $\theta = (B, \Gamma, \Psi, \Phi)$ funktiona. Käytännössä rakenneyhtälömallinnuksen tarkastelu tapahtuu otoskovarianssimatriisin S avulla. Parametrien estimoinnin tavoitteena on tuottaa sellaisia arvoja, joiden avulla erotus:

$$S - \Sigma(\hat{\theta})$$

on pienin mahdollinen.

Estimointimenetelmät. Rakenneyhtälömallin estimoinnissa voidaan käyttää useita eri estimointimenetelmiä. Yleisin on suurimman uskottavuuden (SU) menetelmä (*Maximum Likelihood*, ML), joka löytyy jokaisesta SEM-tarkastelun mahdollistavasta kvantitatiivisen analyysin ohjelmistosta. SU-menetelmällä saadut estimaatit ovat johdonmukaisia, harhattomia, skaalattomia sekä normaalijakautuneita, edellyttäen että oletus muuttujien multinormaalijakautuneisuudesta toteutuu (Schumacker & Lomax, 2016). Derivoinnin avulla minimoidaan sovitusfunktio:

$$F_{ML} = \log |\Sigma(\theta)| + \text{tr}[S\Sigma^{-1}(\theta)] - \log |S| - (p + q).$$

Mikäli oletus muuttujien multinormaalijakautuneisuudesta ei toteudu, on syytä valita robustimpi estimointimenetelmä. Tällainen menetelmä on esimerkiksi yleistetty pienimmän neliösumman menetelmä (GLS). GLS-menetelmässä painotetaan residuaaleja painomatriisiin W avulla. Yleisin valinta painomatriisiksi W on S^{-1} . Tavoitteena on minimoida $S - \Sigma$ välistä erotusta:

$$F_{GLS} = \left(\frac{1}{2} \right) \text{tr} \left\{ [I - \Sigma(\theta) S^{-1}]^2 \right\}.$$

Vaihtoehtoisesti, kun oletus muuttujien normaaliuudesta ei päde, estimointimenetelmänä voidaan käyttää painottamatonta pienimmän neliösumman menetelmää (*Unweighted Least Squares*, ULS). ULS-menetelmässä ei tehdä oletuksia havaittujen muuttujien jakaumista. Myös ULS-menetelmä minimoi sovituskfunktion:

$$F_{ULS} = \left(\frac{1}{2} \right) \text{tr} \left\{ [S - \Sigma(\theta)]^2 \right\}.$$

Satorra–Bentlerin korjausmenetelmän avulla saadaan tuotettua robusteja keskivirheitä sekä estimaatteja, vaikka estimointimenetelmänä käytettäisiin SU-menetelmää, kun oletus muuttujien normaaliuudesta ei toteudu. Satorra–Bentlerin skaalaus huomioi tarkastelun kohteena olevan rakenneyhtälömallin, estimointimenetelmän sekä otoksen kurtoottisuuden (Byrne, 2012; Satorra ja Bentler, 2010). Skaalattua khi-neliösuuretta kutsutaan Satorra–Bentlerin khi-neliösuureeksi (S-B χ^2) ja se saadaan:

$$\text{S-B}\chi^2 = \frac{\chi^2}{c}, \text{ jossa } c \text{ on skaalaustekijä.}$$

Satorra–Bentlerin korjausmenetelmän sisältävää robustia SU-menetelmää kutsutaan MLM-menetelmäksi (*maximum likelihood mean adjusted*, MLM) Mplus-ohjelmistossa sekä lavaan-paketissa.

Rakenneyhtälömallinnuksessa on yleisenä oletuksena muuttujien jakaumien normalisuus sekä faktorien välisten yhteyksien lineaarisuus. Yleisesti oletetaan myös, että aineistossa ei ole puuttuvuutta eikä oudokkeja (Bollen, 1989; Metsämuuronen, 2008). Aineistossa esiintyvää puuttuvuutta voidaan kuitenkin hallita käyttämällä sekä aineiston että tarkasteltavan mallin kannalta sopivia korjaus- ja imputointimenetelmiä (Allen, 2003; Newman, 2014; Schumacker & Lomax, 2016). Näitä korjaus- ja imputointimenetelmiä tarkastellaan luvussa 3.3.

2.5 Mallin sopivuustarkastelut

Rakenneyhtälömalli on yksilöityvä eli *identifioituva*, mikäli kaikki mallin estimoitavat parametrit voidaan ratkaista yksiselitteisesti. Rakenneyhtälömallinnuksessa oletetaan yleensä, että jokainen havaittu muuttuja on yhteydessä vain yhteen faktoriin kerrallaan. Eräs keino vähentää mallin estimoitavien parametrien määrää on kiinnittää yksi jokaisen tarkasteltavan faktorin latauksista ($\lambda_{ij}^{(x)}$, $\lambda_{ij}^{(y)}$) esimerkiksi ykköseksi. Samalla varmistetaan, että mallin faktorilatauksilla on sama mittayksikkö.

Mallin identifioitavuuden tarkastelussa voidaan käyttää ns. *t-sääntöä*. Menetelmä on nopea ja tarpeellinen, mutta ei yksistään riittävä (Bollen, 1989). Kuvassa 1 on kymmenen havaittua muuttujaa, joista kuusi muuttujaa ($x_1, \dots, x_6$) on yhteydessä selittäviin faktoreihin (ξ_1, ξ_2) ja neljä muuttujaa ($y_1, \dots, y_4$) puolestaan selitettävään faktoriin (η_1). Mallin ei-estimoitavien eli *kiinnitettyjen* parametrien kokonaismäärä saadaan laskemalla mallin selittäviin faktoreihin liittyvien havaittujen muuttujien lukumäärä $p = 6$ sekä selitettäviin faktoreihin liittyvien havaittujen muuttujien lukumäärä $q = 4$, jolloin

$$\frac{1}{2}(p+q)(p+q+1) = \frac{1}{2}(6+4)(6+4+1) = 55 \quad (8)$$

Yhtälön 8 avulla tuotettu kiinnitettyjen parametrien lukumäärä on yläraja mallin estimoitavien, *vapaiden* parametrien lukumäärälle. Kun mallin vapaiden parametrien lukumäärä t on pienempi (tai yhtäsuuri) kuin kiinnitettyjen muuttujien

$$t \leq \frac{1}{2}(p+q)(p+q+1),$$

malli on identifioituva (Bollen, 1989; Metsämuuronen, 2008).

2.6 Yleiset sopivuusmittarit

Mallin sopivuutta tarkastellaan erilaisten indikaattorien avulla. Yleisimmin käytetyt indikaattorit ja niiden suositellut kynnsarvot ovat (Halme ym., 2014; Byrne, 2012):

- *Khi-neliötesti* – Mikäli saatu testisuure χ^2 ei ole tilastollisesti merkitsevä, ero otoskovarianssimatriisin ja teoreettisen mallin tuottaman kovarianssimatriisin välillä on minimaalinen.
- *Khi-neliösuure:vapausaste* – Suhdeluvun tulisi olla korkeintaan 5.

$$\frac{\chi^2}{df_{malli}} \leq 5, \text{ df} = \text{vapausaste.}$$

- *Akaiken informaatiokriteeri*, AIC – Vertailtaessa kahta mallia keskenään, pienempi AIC-arvo indikoi parempaa mallia. Eri tilastolliset ohjelmat laskevat AIC-arvon hieman eri tavalla. R-kielen *lavaan*-paketti tuottaa AIC-arvon seuraavasti:

$$AIC = -2 \cdot (\text{mallin log-uskottavuus}) + 2 \cdot (\text{estimoitavien parametrien lkm})$$

- *Bayesin informaatiokriteeri*, BIC – Mitä pienempi BIC-arvo, sitä parempi malli. R-kielen *lavaan*-paketti tuottaa BIC-arvon seuraavasti:

$$BIC = -2 \cdot (\text{mallin log-uskottavuus}) + \log(N) \cdot (\text{estimoitavien parametrien lkm})$$

- *Vertaileva sopivuusindeksi (Comparative fit index)*, CFI – Arvojen tulisi olla vähintään 0,90. Nollamalli on malli, jossa muuttujien välillä ei ole lainkaan vuorovaikutusta. CFI-indeksi on

$$CFI = 1 - \frac{\chi_{malli}^2 - df_{malli}}{\chi_{nollamalli}^2 - df_{nollamalli}}, \text{ df} = \text{vapausaste.}$$

- *Tucker-Lewis-indeksi*, TLI – Hieman konservatiivisempi indeksi kuin CFI. Arvojen tulisi myös olla vähintään 0,90.

$$TLI = \frac{\frac{\chi_{nollamalli}^2}{df_{nollamalli}} - \frac{\chi_{malli}^2}{df_{malli}}}{\frac{\chi_{nollamalli}^2}{df_{nollamalli}} - 1}, \text{ df} = \text{vapausaste.}$$

- *Keskineliövirheen neliöjuuri (Root Mean Square Error of Approximation), RMSEA*
– Mitä pienempi arvo, sen parempi malli. Arvon tulisi olla alle 0,06.

$$\text{RMSEA} = \sqrt{\frac{\chi^2_{\text{malli}} - \text{df}_{\text{malli}}}{N - \text{df}_{\text{nollamalli}}}}, \text{ N = otoskoko ja df = vapausaste.}$$

- *Keskimääräinen jäännöskorrelaatio (Standardised Root Mean Square Residual), SRMR* – Otokorrelaation ja estimoidun korrelaation standardoitu ero. Saatu tunnusluku on suhteutettu otokorrelaatioon, ja sen tulisi olla korkeintaan 0,08.

Kunkin indikaattorin antamaa tietoa tulisi tarkastella yhdessä muiden indikaattorien kanssa. Edellä esitettyjen indikaattorien lisäksi mallin sopivuutta tarkastellaan modifikaatioindeksien (MI) avulla. Kiinnitetyn parametrin modifikaatioindeksin arvo kertoo mahdollisuudesta parantaa mallin sopivuutta vapauttamalla kyseinen kiinnitetty parametri estimoitavaksi. Malliin mahdollisesti tehtävien muutosten on oltava alkuperäisen mallin teorian ja siihen liittyvien tutkimushypoteesien mukaisia (Haydek, 2014).

3 Puuttuvuus

Kun aineistossa on puuttuvia havaintoja, analyysin tuloksiin voi liittyä puuttuvuuden tuomaa harhaa. Jotta teoreettisen mallin sopivuuden tarkastelu onnistuisi parhaalla mahdollisella tavalla, on pyrittävä selvittämään, miksi havaintoaineistossa on puuttuvaa tietoa. Puuttuva tieto on heterogeenistä, ja puuttuvuutta aiheuttavat mekanismit voivat olla erilaisia. Rubin (1976) loi puuttuvalle tiedolle luokittelumenetelmän, jonka avulla havaintoaineistossa esiintyvää puuttuvuutta voidaan luokitella kolmen eri pääkategorian avulla. Kun puuttuvan tiedon luokka eli laji on tunnistettu, puuttuvuutta voidaan korjata valitsemalla sekä havaintoaineistolle että puuttuvuuden lajille sopiva estimointi- tai imputointimenetelmä.

3.1 Puuttuvan tiedon lajit

Puuttuva tieto on *täysin satunnaista puuttuvuutta* (*Missing Completely at Random*, MCAR) silloin, kun puuttuvuus ei riipu aineiston muista havainnoista eikä itse puuttuvista tiedoista. Puuttuva tieto on täysin satunnaista puuttuvuutta esimerkiksi silloin, kun tutkimuksessa käytettävä mittalaite vikaantuu ja täten mittatulosta ei saada taltioitua (van Buuren, 2012).

Kun puuttuvuus on riippuvaista aineiston havaituista arvoista, mutta ei puuttuvista tiedoista, kyseessä on *satunnainen puuttuvuus* (*Missing at Random*, MAR) (van Buuren, 2012). Oletetaan, että joukolta 12–20-vuotiaita kysytään energijaumien viikoittaisen kulutuksen määrää. Koska energijaumia ei suositella alle 15-vuotiaille, kyselyssä 12–14-vuotiaat voivat olla haluttomia vastaamaan kysymykseen. Puuttuva tieto riippuu tällöin muuttujasta ikä, mutta ei muuttujasta, joka mittaa energijauman viikottaisen kulutuksen määrää.

Kolmas puuttuvuuden laji on *ei-satunnainen puuttuvuus* (*Missing Not at Random*, MNAR). Tällöin puuttuvuus riippuu puuttuvista arvoista. Oletetaan, että samaiselta joukolta 12–20-vuotiaita kysytään hammaskarieksen määrästä, joka todennettiin määrääikaisen hammaslääkärikäynnin yhteydessä. Henkilö, jolla todettiin olevan hampaissa yli viisi reikää, voi tietoisesti jättää vastaamatta kysymykseen. Puuttuva tieto riippuu täten muuttujasta, jolla määritetään vastaajan hammaskarieksen määrää.

Van Buuren (2012) esittää puuttuvan tiedon lajit ja niihin liittyvät ehdolliset jakaumat Littlea ja Rubinia (2002) lainaten, mutta yksinkertaisemmin. Merkitään havaintoaineistoa matriisilla $\mathbf{Y} \in \mathbb{R}^{n \times p}$, jossa on p muuttujaa n havaintoyksiköltä. Indikaattorimatriisi $\mathbf{R} \in \mathbb{R}^{n \times p}$ kuvaa, puuttuuko havaintoyksiköltä jonkin satunnaismuuttujan arvo vai ei. Merkitään $\mathbf{Y} = (y_{ij})$ sekä $\mathbf{R} = (r_{ij})$, jossa $i = 1, \dots, n$, ja $j = 1, \dots, p$. Kun havainto y_{ij} ei puutu, $r_{ij} = 1$ ja vastaavasti $r_{ij} = 0$, kun havainto y_{ij} puuttuu. Merkitään ei-puuttuvia havaintoja matriisilla $\mathbf{Y}_{\text{havaittu}}$ ja puuttuvia havaintoja matriisilla $\mathbf{Y}_{\text{puuttuva}}$. Matriisi $\mathbf{Y}_{\text{puuttuva}}$ käsittää kaikki ne alkiot y_{ij} , joissa $r_{ij} = 0$. Merkinnällä $\mathbf{Y} = (\mathbf{Y}_{\text{havaittu}}, \mathbf{Y}_{\text{puuttuva}})$ kuvataan täten koko havaintoaineistoa. Matriisi $\mathbf{R}$ sisältää puuttuvien havaintojen sijainnit, ja sen jakauma voi riippua matriisista $\mathbf{Y}$. Mekanismi, joka kuvaa matriisien $\mathbf{R}$ ja $\mathbf{Y}$ välistä riippuvuutta, määrittää puuttuvuuden lajin. Merkitään puuttuvuutta aiheuttavan mekanismin ominaisuuksia parametrilla ϕ , jolloin saadaan yleinen, puuttuvuutta kuvaava todennäköisyys

$$P(\mathbf{R} | \mathbf{Y}_{\text{havaittu}}, \mathbf{Y}_{\text{puuttuva}}, \phi). \quad (9)$$

Täysin satunnainen puuttuvuus (MCAR) tarkoittaa, että

$$P(\mathbf{R} = 0 | \mathbf{Y}_{\text{havaittu}}, \mathbf{Y}_{\text{puuttuva}}, \phi) = P(\mathbf{R} = 0 | \phi). \quad (10)$$

Ehdossa (10) esitetty puuttuvuus riippuu ainoastaan puuttuvuutta aiheuttavasta mekanismista. Mekanismi voi olla keinotekoisesti tuotettu tai sattumanvarainen.

Satunnainen puuttuvuus (MAR) tarkoittaa, että:

$$P(\mathbf{R} = 0 | \mathbf{Y}_{\text{havaittu}}, \mathbf{Y}_{\text{puuttuva}}, \phi) = P(\mathbf{R} = 0 | \mathbf{Y}_{\text{havaittu}}, \phi). \quad (11)$$

Ehdon (11) mukaan todennäköisyys, että havainto puuttuu, on sidoksissa ei-puuttuviin havaintoihin sekä puuttuvuutta aiheuttavaan mekanismiin.

Kolmas puuttuvan tiedon laji eli ei-satunnainen puuttuvuus (MNAR) tarkoittaa, että:

$$P(\mathbf{R} = 0 | \mathbf{Y}_{\text{havaittu}}, \mathbf{Y}_{\text{puuttuva}}, \phi). \quad (12)$$

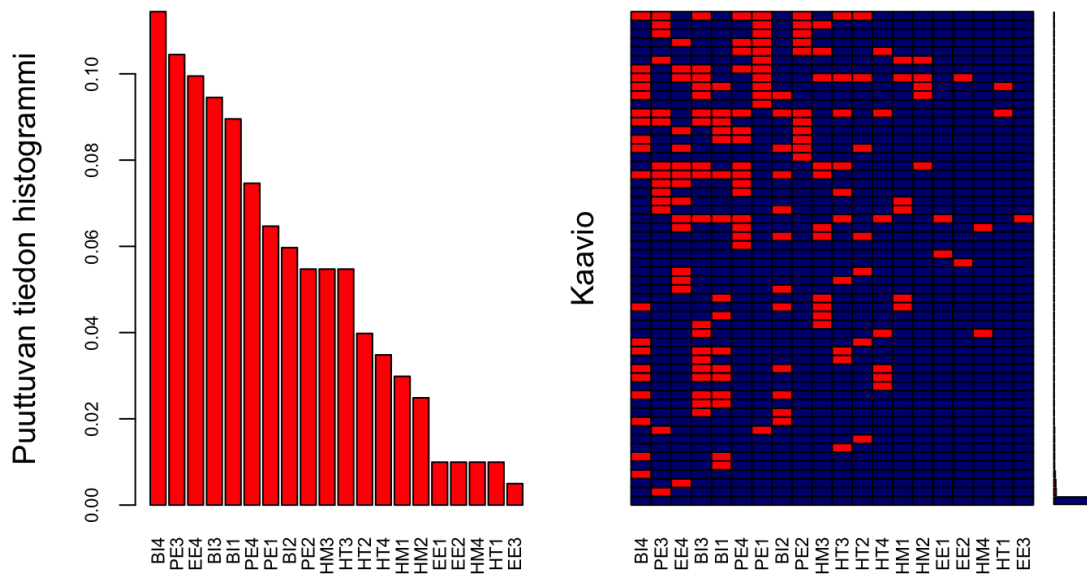
Ehdosta (12) nähdään, että puuttuvuuden määrittävät sekä ei-puuttuvat että puuttuvat havainnot yhdessä puuttuvuutta aiheuttavan mekanismin kanssa.

3.2 Puuttuvan tiedon tunnistaminen

Havaintoaineiston puuttuvuutta voidaan tarkastella myös puuttuvan tiedon tasojen kautta. Kun havaintoyksiköltä eli vastaajalta puuttuu havaintoarvo yksittäisen selittävän muuttujan kohdalla, kyseessä on muuttujatason puuttuvuus (*item-level missingness*). Mikäli vastaajalta puuttuu havaintoarvo useamman selittävän muuttujan osalta tai yksikin havaintoarvo selitettävän muuttujan kohdalla, puhutaan ns. rakenteellisesta puuttuvuudesta (*construct-level missingness*). Yksilötason puuttuvuus puolestaan (*person-level missingness*) tarkoittaa sitä, että vastaajan kaikki havaintoarvot puuttuvat tarkasteltavien muuttujien osalta. Puuttuvuuden tasorakenne havaintoaineistossa tulee myös huomioida käsiteltäessä puuttuvaa tietoa. (Newman, 2014).

Hyödyntämällä tilastollista R-ohjelmointikieltä voidaan tuottaa havaintoaineistossa esiintyvän puuttuvan tiedon kaavio. Tarkastelemalla kaaviota on mahdollista nähdä, mikäli puuttuvuus on säännönmukaista. Kuvan 2 vasemmanpuoleisesta histogrammista nähdään, että tarkasteltavan havaintoaineiston muuttujassa BI4 esiintyy eniten puuttuvaa tietoa (11,4 %). Kuvan 2 oikeanpuoleisessa kaaviossa on havaintoaineistossa esiintyvää puuttuvaa tietoa merkitty punaisilla suorakaiteilla. Punaiset ja siniset suorakaiteet muodostavat kaaviossa rivikohtaisesti erilaisia kuvioita. Kaaviossa olevien rivien lukumäärä kertoo, miten monella eri tavalla havaintoaineistossa esiintyy puuttuvaa tietoa. Kaavion jokainen rivi kuvaa yhtä yksilöllistä puuttuvan tiedon mallikuviota (*pattern*), joka esiintyy tarkasteltavassa havaintoaineistossa.

Kuvan 2 kaaviossa esiintyy jonkistaista yhdenmukaisuutta, mutta ei varsinaista säännönmukaisuutta. Puuttuvan tiedon luokittelua voidaan täydentää *Littlen testillä* (Little, 1988; Li, 2013), joka on lähinnä suuntaa antava. Littlen testi vertailee havaintoaineiston puuttuvuuden kuvioita keskenään ja testaa eroavaisuuden osalta. Mikäli kuvioiden eroavaisuus on tilastollisesti merkitsevä, oletus täysin satunnaisesta puuttuvuudesta ei saa tukea (Little, 1988; Li, 2013). Littlen testin voi tehdä hyödyntämällä esimerkiksi R-ohjelman *mcartest*-pakettia. Kuvassa 2 esiintyvä puuttuvuus ei Littlen testin ($H_0 = \text{MCAR}$) mukaan ole täysin satunnaista ($p < 0.001$).



Kuva 2: Tarkasteltavassa havaintoaineistossa ilmenevä puuttuvuus histogrammina ja kaaviona. Histogrammi (vasemmalla) esittää muuttujakohtaisen puuttuvan tiedon osuuden. Kaavio (oikealla) kuvaa havaintoaineistossa esiintyviä puuttuvuuden eri mallikaavioita. Havaintoaineisto on DigitalWells-ohjelman kyselytutkimuksen osa-aineisto sisältäen 20 havaittua muuttujaa. $N = 201$.

3.3 Puuttuvuuden hallinta

Mikäli aineistossa ilmenevä puuttuvuus on hyvin vähäistä sekä täysin satunnaista, voidaan mahdollisesti käyttää *täydellisten havaintorivien analyysia* (*list-wise deletion*). Tällöin havaintoaineistosta poistetaan ne havaintoyksiköt, joilla on puuttuvaa tietoa jonkun muuttujan kohdalla. Menetelmä saattaa kuitenkin pienentää aineistokokoa huomattavasti, mikä puolestaan aiheuttaa harhaa tilastollisen analyysin yhteydessä. Lisäksi puuttuva tieto on harvoin täydellisesti jotain tiettyä yhtä lajia; se voi esimerkiksi olla jatkumo satunnaisen ja ei-satunnaisen puuttuvuuden välillä (Newman, 2014; Graham, 2009). Vaikka täydellisten havaintorivien analyysi on menetelmänä suosittu, yleisesti suositellaan käyttämään jotain vaihtoehtoisia menetelmiä puuttuvan tiedon käsittelyyn (Newman, 2014; Graham, 2009; van Buuren, 2012).

Imputointimenetelmät perustuvat pääsääntöisesti aina tilastolliseen malliin, mutta on myös poikkeuksia: puuttuva tieto voidaan korvata muilta havaintoyksiköiltä saatavilla havaintoarvoilla eli ns. *hot deck*-menetelmällä (Härkänen, 2023).

3.4 Rakenneyhtälömalleille suotuisat menetelmät

Rakenneyhtälömallinnuksessa parametrien estimointimenetelmänä käytetään yleensä SU-menetelmää, kun oletus muuttujien multinormaalijakaumasta toteutuu. Multinormaalijakauman määräävät sen parametrit eli odotusarvovektori μ sekä kovarianssimatriisi Σ .

Rakenneyhtälömallinnuksessa parametrit (μ, Σ) ovat yleensä muunnoksia pienemmästä parametrijoukosta (Allison, 2003).

Kun puuttuvaa tietoa on useammassa tarkasteltavassa muuttujassa samanaikaisesti, kyse on *moniuloitteisesta puuttuvuudesta*. Näin on yleensä rakenneyhtälömallien kohdalla. Allison (2003) esittää mahdollisia menetelmiä puuttuvan tiedon käsittelyyn, kun malliparametrien estimointi pohjautuu SU-menetelmään.

3.4.1 Ehdollistamaton keskiarvoimputointi

Imputointimenetelmänä *ehdollistamaton keskiarvoimputointi* on yksinkertainen. Muuttujan puuttuvaa tietoa paikataan laskemalla muuttujan ei-puuttuvien arvojen keskiarvo. Vaikka menetelmänä ehdollistamaton keskiarvoimputointi on suosittu, niin menetelmä tuottaa usein liian pieniä variansseja (van Buuren, 2012; Härkänen, 2023).

3.4.2 Moni-imputointi

Moni-imputointi (*Multiple imputation*, MI) on iteratiivinen menetelmä, jossa puuttuva tieto voidaan korvata esimerkiksi Markovin ketju Monte Carlo (*MCMC*) -algoritmin avulla tuotetulla eli simuloidulla arvolla. Simulointi tapahtuu valittua tilastollista mallia käyttäen, esimerkiksi lineaarista regressiomallia. Moni-imputoinnissa puuttuvan tiedon korvaava arvo tuotetaan simuloimalla aineistoa monta kertaa. Suosituksena on tuottaa vähintään viisi erillistä aineistoa, jolloin saadut estimaatit ovat yleensä harhattomia (Shin ym., 2022). Lopullinen korvaava arvo saadaan simuloitujen arvojen keskiarvona (Lee & Shi, 2021).

Puuttuvaa tietoa voidaan myös korvata jollain aineistosta havaitulla arvolla. Tällöin aineistosta valitaan tilastolliseen malliin, esimerkiksi regressiomalliin, pohjautuen sellaiset havaintoyksiköt eli *luovuttajat*, joilla ei ole puuttuvaa tietoa. Valitut luovuttajat ovat samankaltaisia niiden havaintoyksiköiden kanssa, joilla on puuttuvuutta. Eräs tällainen käytetty menetelmä on *ennustekeskiarvojen kaltaistus*, (PMM). Imputointimenetelmänä ennustekeskiarvojen kaltaistus on *hot deck* -menetelmä (Härkänen, 2023).

Moni-imputointimenetelmiä on useita, ja jotkut menetelmistä ovat hyvin monimutkaisia. R:n *mi* -paketti sisältää useita eri menetelmiä.

3.4.3 Satunnaismetsä

Koneoppimisessa käytetty satunnaismetsä (*Random Forest*, RF) on imputointimenetelmänä melko uusi. Rakenneyhtälömallinnuksessa puuttuvan tiedon paikkaamista satunnaismetsän avulla voidaan lähestyä kolmella eri tavalla:

1. Imputoidaan ensiksi aineistoa, jonka jälkeen lisätään päätöspuita ("kasvatetaan metsää"). Päivitetään aiemmin imputoituja arvoja saatujen estimaattien avulla. Toistetaan, jotta saadaan tarkempia tuloksia.
2. Imputoidaan ja samanaikaisesti lisätään päätöspuita ("kasvatetaan metsää"). Toistetaan, jotta saadaan parempia tuloksia.
3. Imputoidaan ensiksi aineistoa, jonka jälkeen lisätään päätöspuita ("kasvatetaan metsää") käyttämällä muuttujia, joissa on puuttuvuutta. Käytetään saatuja ennusteita puuttuvuuden paikkaamiseen. Toistetaan, jotta saadaan parempia tuloksia.

Imputointimenetelmänä satunnaismetsä on joustava, sillä aineiston muuttujien ei tarvitse olla normaalijakautuneita eikä faktorien välisten yhteyksien lineaarisia. Lisäksi muuttujien väliset riippuvuudet voivat olla hyvinkin monimutkaisia. Shin ja kollegat (2022) ovat saaneet mielenkiintoisia tuloksia käyttämällä satunnaismetsää imputointiin.

3.4.4 FIML

Eräs SU-menetelmän muunnos on *Full Information ML*, FIML. Verrattuna yleiseen SU-menetelmään, FIML-menetelmä tuottaa jokaiselle havaintoyksikölle uskottavuusfunktion hyödyntämällä kaikkea käytettävissä olevaa informaatiota, myös puuttuvuutta (Lee & Shi, 2021). SU-menetelmän muunnoksena FIML on yleensä luotettava. Menetelmä on suosittu, sillä sitä voi käyttää suoraan analyysivaiheessa. Aineistolle tehtävää erillistä imputointia ei tarvita.

3.4.5 PLS-SEM

Osittaisen pienimmän neliösumman rakenneyhtälömallinnus (*Partial Least Squares SEM*, PLS-SEM) eroaa ”tavallisesta” kovarianssiin pohjautuvasta rakenneyhtälömallinnuksesta (CB-SEM) siten, että PLS-SEM on iteratiivinen, epäparametrinen menetelmä. PLS-SEM-menetelmässä rakenneyhtälömallin sopivuustarkastelu pohjautuu siihen, missä määrin selitettävän faktorin vaihtelua voidaan selittää mallin avulla (Astrachan ym. (2014)). Lisäksi PLS-SEM-menetelmässä rakenneyhtälömalli ositetaan ja tarkastelu tapahtuu faktori kerrallaan. ”Tavallisessa” rakenneyhtälömallinnuksessa analyysi tehdään koko rakenneyhtälömallille samalla kertaa (*full information approach*).

PLS-SEM-menetelmässä mittamalliin viitataan myös käsitteellä *ulkoinen malli* (*outer model*) ja rakennemalliin puolestaan *sisäinen malli* (*inner model*). ”Tavallisesta” rakennemallinnuksesta poiketen PLS-SEM-menetelmässä mittamallin havaitut muuttujat oletetaan virheettömiksi; ainoastaan selitettävän faktorin yhteydessä mallinnetaan havaittujen muuttujien mittavirheet (Hair ym., 2017; Tenenhaus, 1998). Mittamalli voi olla *formatiivinen* (*formative*) tai *reflektiivinen* (*reflective*). Formatiivisessa mittamallissa havaittujen muuttujien ja faktorin väliset yhteydet kuvataan yksisuuntaisella nuolella siten, että havaitut muuttujat osoittavat selittävää faktoria. Reflektiivinen mittamalli vastaa ”tavallisen” rakenneyhtälömallin mittamallin esitystapaa (kuva 1).

Kun rakenneyhtälömalli sisältää ainakin yhden selitettävän faktorin, kyseisestä rakenneyhtälömallista käytetään nimitystä *malli A*. Konfirmatorista faktorianalyysia vastaava malli on puolestaan *malli B* (Tenenhaus, 1998; Hair ym., 2017).

PLS-SEM-menetelmän analyysi etenee osittaisen pienimmän neliösumman menetelmän (*Partial Least Squares*, PLS) avulla, joka on painotettu regressiomenetelmä (Saari, 2008). PLS-algoritmi etenee seuraavasti:

1. Approksimoidaan ulkoisen mallin faktorilataukset.
2. Estimoidaan sisäisen mallin painotukset.
3. Approksimoidaan sisäisen mallin faktorilataukset.
4. Estimoidaan ulkoisen mallin painotukset.

Vaihtoehtoisesti voidaan valita painotuksia, jotka ovat faktoripainotuksia tai pääkomponenttianalyysiin (*principal component analysis*, PCA) pohjautuvia. Painotusten lisäksi valitaan menetelmä, jolla puuttuvuutta korjataan. Yleinen menetelmä on keskiarvoimputointi.

PLS-SEM-menetelmän yhteydessä tarkastellaan mallin jokaisen faktorin reliabiliteettiä eli miten luotettavasti ja toistettavasti faktori mittaa kyseistä ilmiötä. Faktorin reliabiliteetti ilmaistaan arvolla ρ_C (*composite reliability*, CR), jonka tulisi olla vähintään 0,7 (Brown, 2015). Mallin sopivuustarkastelun yhteydessä tutkitaan myös faktorin valideettiä eli miten hyvin faktori mittaa sitä ilmiön ominaisuutta, jota sen on määrä mitata. Faktorin valideetin arvioimisessa käytetään kahta ehtoa, joiden molempien tulisi täytyä. Ensimmäisen ehdon mukaan faktorin suhteellisen selitysosuuden (*average variance extracted*, AVE) tulisi olla vähintään 50 % (Hair ym., 2019). Toisen ehdon mukaan faktorin suhteellisen selitysosuuden neliöjuuren tulisi olla suurempi kuin faktorin korrelaatio muiden faktorien kanssa (Hair ym., 2019).

Faktorin reliabiliteetin ja valideetin lisäksi tulisi tarkastella mallin jokaisen yksittäisen havaitun muuttujan faktorilatausta, jonka tulisi olla vähintään 0,7 (Hair ym., 2019). Myös mahdollinen *multikollineaarisuus* tulee tarkistaa. Multikollineaarisuudella tarkoitetaan tilannetta, jossa faktorin havaittujen muuttujien välillä on hyvin suuria riippuvuuksia. Hair ja kollegat (2019) antavat multikollineaarisuutta ilmaisevan *varianssin inflaatiotekijän* (*variance inflated factor*) eli VIF-kertoimen ideaaliseksi viitearvoksi alle 3. Käytännössä alle viitearvon 5 oleva VIF-kerroin kertoo, että multikollineaarisuutta ei esiinny. VIF-kerroin saadaan seuraavasti:

$$\text{VIF} = \frac{1}{1 - R^2}, \text{ jossa } R^2$$

kuvaa muuttujien välisen tilastollisen riippuvuussuhteen voimakkuutta.

PLS-SEM käyttää bootstrap-menetelmää parametriestimaattien luottamusvälien sekä estimaattien keskivirheiden tuottamiseen. Täten t-testin avulla on mahdollista tarkastella saatujen estimaattien tilastollista merkitsevyyttä valitulla merkitsevyystasolla.

Molemmat rakenneyhtälömallinnustavat, ”tavallinen” SEM sekä PLS-SEM, eivät ole toisiaan poissulkevia vaan täydentäviä. Kun tarkasteltava aineisto sisältää puuttuvuutta, aineistokoko on pieni tai oletus muuttujien multinormaalisuudesta ei toteudu, niin silloin PLS-SEM on hyvä vaihtoehto (Hair ym., 2016, 2017, 2018, 2023).

3.5 Imputointimenetelmän valinta

Newman (2014) esittää viisi käytännön toimenpidettä puuttuvan tiedon käsittelyyn:

1. Käytä koko havaintoaineisto, joten unohda täydellisten havaintorivien analyysi.
2. Älä käytä ehdollistamatonta keskiarvoimputointia.
3. Käytä SU-estimoinnin ja moni-imputoinnin menetelmiä, kun puuttuvuus on rakenteellista ja sitä on yli 10 % havaintoaineistosta. Käytä imputointimenetelmissä hyödyksi havaintoaineistosta saatavia apumuuttujia eli muuttujia, jotka ovat varsinaisen mallitarkastelun ulkopuolella, mutta voivat tuoda lisäinformaatiota imputointiin.
4. Kun puuttuvuus on muuttujatason puuttuvuutta, yksi havaintoarvo on riittävä ”edustamaan” kahdesta tai useammasta muuttujasta koostettua yhdistemuuttujaa.

5. Mikäli yksilötason puuttuvuuden seurauksena vastausprosentti on alle 30 %, on syytä tehdä herkkyysanalyysi.

Newmanin (2014) käytännön toimenpiteitä sovelletaan tutkielman empiirisessä osiossa.

4 DigitalWells-ohjelma

Tilastokeskuksen (2021) mukaan Suomen väestö ikääntyy vauhdilla ja tulevina vuosina on odotettavissa, että väestöllinen huoltosuhde heikkenee. Täten ikääntyvän väestön terveyttä tukevaa, liikunnallista aktiivisuutta sekä huonoa kuntoa ennaltaehkäisevää toimintaa on tuettu erilaisin projektein (Husu ym., 2022).

Eläkeliiton, Eläkkeensaajien Keskusliiton sekä Svenska Pensionärsförbundetin tuema ja Kansaneläkelaitoksen (Kelan) rahoittama DigitalWells-ohjelma toteutettiin vuosina 2019–2022. Kaiken kaikkiaan ohjelmaan osallistui noin 850 eläkeläisliittojen jäsentä. Heidän lisäksi DigitalWells-ohjelmassa oli mukana kaksi erillistä ryhmää, Peurungan kuntoutusasiakkaat (noin 50 henkilöä) sekä Tanhuvaaran ALLi 60+ -asiakkaat (noin 105 henkilöä). Covid-19-pandemian myötä aiheutuneet yhteiskunnan eristystoimet vuoden 2020 keväällä ja syksyllä loivat kolmannen erillisen ryhmän eli virtuaaliryhmän. Tämän virtuaaliryhmän noin 30 jäsentä toimi täysin ohjelmaa varten räätälöidyn DigitalWells-liikuntasovelluksen (myöhemmin DW-liikuntasovellus) varassa.

Ohjelman keskeisenä tavoitteena oli luoda sellaisia rutiineja ja toimenpiteitä, jotka edesauttaisivat ja ylläpitäisivät nuorten ikäihmisten (60–75-vuotiaiden) terveyttä edistävää liikunnallista aktiivisuutta. Ohjelmaan osallistuneiden henkilöiden enemmistö sitoutui ohjelmaan vähintään 24 kuukauden ajaksi.

Ennen Covid-19-pandemian aiheuttamia rajoituksia, DigitalWells-ohjelmaan osallistujille järjestettiin alueyhdistyskohtaisia lähtötapaisuuksia. Aloitustapaamisessa esiteltiin ohjelman tavoitteet, sisältö sekä toimintaperiaatteet. DigitalWells-ohjelmaan osallistuvilta kerättiin kirjallinen suostumus tieteelliseen tutkimukseen osallistumisesta sekä tutkimusaineiston käytöstä. Osallistuminen ohjelmaan oli täysin vapaaehtoista ja osallistujalla oli mahdollisuus keskeyttää milloin tahansa. Ainoa edellytys ohjelmaan osallistumiselle oli älypuhelimien käyttömahdollisuus ohjelmakauden ajan.

DigitalWells-ohjelma toteutettiin sekä suomeksi että ruotsiksi. Ryhmätapaamisissa pidettyjen esitysten sekä ohjelman verkkosivujen kautta jaetun materiaalin lisäksi osallistujilla oli käytössä koko ohjelmakauden ajan omaa liikunnallista aktiivisuutta tukeva DW-liikuntasovellus (kuva 3). Jokaisella osallistuvalla alueyhdistyksellä oli oma kenttätutkija, johon osallistujilla oli mahdollisuus ottaa aina tarvittaessa yhteyttä. Lisäksi jokainen osallistuja kuului oman alueyhdistyksensä WhatsApp-ryhmään. Täten viestintä ohjelmaan osallistujien kanssa oli monikanavaista ja -kielistä.

4.1 Tutkimusasetelma

Ohjelmakauden aikana osallistujilta kerättiin heidän liikuntasuoritustensa lisäksi mielenpitoja DigitalWells-ohjelmaa varten kehitettyjen kyselylomakkeiden avulla. Kysely toteutettiin viidessä eri aikapisteessä. Aloitustapaamisen yhteydessä osallistujilta kerättiin kyselylomakkeiden avulla taustatietoja sekä liikunnallisen aktiivisuuden lähtötasoa. Seurantakyselyjen avulla kerättiin liikunta-aktiivisuuden tason lisäksi näkemyksiä ja kokemuksia uuden teknologian eli DW-liikuntasovelluksen käyttöönotosta. Seurantakysely toteutettiin niin, että ensimmäinen kyselylomake lähetettiin osallistujille neljä kuukautta ohjelmassa aloittamisen jälkeen. Toinen seurantakysely toteutettiin 6 kuukauden, kolmas 12 kuukauden ja neljäs 24 kuukauden kuluttua ohjelman aloitushetkestä. Kesken ohjelmakautta lopettaneille osallistujille lähetettiin erillinen lopetuskysely.

DigitalWells-ohjelman tutkimusalue käsitti koko Manner-Suomen sekä Ahvenanmaan. Johtuen osittain käytettävissä olevien kenttätutkijoiden rajallisesta määrästä, osallistujat aloittivat DigitalWells-ohjelman kolmessa eri aallossa: kesällä 2019 (1. aalto), vuoden 2019 lopussa ja vuoden 2020 alussa (2. aalto) sekä vuoden 2020 loppupuoliskolla (3. aalto). Jokaiselle eri aallolle tuotettiin oma tietokanta, sillä käytetyt kyselylomakkeet olivat osittain eriäviä. Tietokannat rakennettiin käyttäen IBM SPSS -ohjelmaa, joka oli DigitalWells-ohjelman tutkijoille entuudestaan tuttu. Kyselytutkimukset toteutettiin vuoden 2019 aikana paperilomakkeiden avulla, mutta keväällä 2020 paperilomakkeet korvattiin sähköisillä LimeSurvey-lomakkeilla.

Omaa liikunnallista aktiivisuutta osallistujat seurasivat maksuttoman DW-liikunta-soveluksen avulla (kuva 3). Älypuheliin asennettavan sovelluksen avulla osallistujat pystyivät tallentamaan liikuntasuorituksiaan manuaalisesti tai automaattisesti esimerkiksi urheilurannekkeen toimintoa tukevan sovelluksen avulla. Osallistujilla oli myös mahdollisuus viedä DW-liikuntasovellukseen tallennetut liikuntasuoritteet Kelan Omatietovarantoon.



Kuva 3: DigitalWells-liikuntasovellus. Liikuntasovellukseen tallennettiin liikuntasuorituksia manuaalisesti tai toisen sovelluksen avulla automaattisesti.

4.2 Kyselytutkimuksen aineisto

Tässä tutkielmassa hyödynnetään DigitalWells-ohjelman toisen aallon kyselytutkimusaineistoa ensimmäisen seurantajakson osalta eli noin neljä kuukautta ohjelmassa aloittamisen jälkeen. Täten seurantajakso sijoittuu ajalle toukokuu – syyskuu 2020, sillä osallistujat aloittivat ohjelmassa porrastetusti vuoden 2019 lopulla ja vuoden 2020 alussa. Tutkielmassa käytettävään ensimmäisen seurantajakson aineistoon on lisätty osallistujien aloituskyselyssä kerättyjä taustamuuttujia. Käytetyt taustamuuttujat ovat sukupuoli, ikä, koulutusaste, maakunta, jossa vastaajan vakituinen osoite sijaitsee, käytössä olevan älypuhelimien merkki sekä kokemus älypuhelinsovellusten käytöstä. Taustamuuttujaksi lisättiin myös tieto kyselylomakkeen kieliversiosta, jonka tutkimuksen osallistuja palautti (suomi tai ruotsi).

Tässä tutkielmassa käytettävässä aineistossa on huomioitu ainoastaan ne osallistujat, jotka ovat vastanneet sekä aloituskyselyyn että ensimmäiseen seurantakyselyyn. Kaiken

kaikkiaan 60 ohjelmassa aloittanutta ja aloituskyselyn täyttäneitä osallistujaa ei vastannut lainkaan ensimmäiseen seurantakyselyyn. Täten heidät rajattiin aineistosta pois. Lisäksi kolme osallistujaa oli jättänyt täyttämättä aloituskyselyn, mutta oli vastannut ensimmäiseen seurantakyselyyn. Myös heidät rajattiin aineistosta pois. Täten tutkielmassa käytettävä aineisto sisältää 201 henkilöä.

Liitteen taulukosta T1 nähdään, että DigitalWells-ohjelman toisen aallon osallistujista enemmistö oli naisia (60,7 %). Osallistujien suurin ikäryhmä oli 70–74-vuotiaat (45,3 %) ja toiseksi suurin oli 65–69-vuotiaat (35,3 %). Toisen aallon osallistujista 10,4 % oli suorittanut yliopisto- tai korkeakoulututkinon, mutta koulutusasteista yleisin oli ammattikoulu (75,1 %). Osallistujat olivat pääosin kotoisin Keski-Suomesta (30,3 %), Varsinais-Suomesta (27,9 %) sekä Etelä-Karjalasta (19,9 %).

Toisen aallon osallistujien keskuudessa suosituin älypuhelimien merkki oli Samsung (43,8 %) ja toisella sijalla oli Huawei (23,9 %) (taulukko T2). Kolmanneksi suosituin älypuhelinmerkki oli iPhone (15,4 %). Älypuhelinsovelluksia toisen aallon nuoret ikäihmiset olivat pääosin käyttäneet 3–5 vuotta (30,8 %) tai 6–10 vuotta (28,9 %). Osallistujista 18,9 % oli käyttänyt erilaisia älypuhelinsovelluksia yli kymmenen vuoden ajan. Täten yli puolella osallistujista (78,6 %) oli vähintään kolmen vuoden käyttökokemus älypuhelinsovelluksista. Lähes 80 %:lle DigitalWells-ohjelman toisen aallon nuorista ikäihmisistä älypuhelin tietojärjestelmänä ei ollut täysin uusi kokemus.

5 Empiirinen osio

Yksi DigitalWells-ohjelman keskeisistä tutkimuskysymyksistä tarkasteli ohjelmalle räätöidyn DW-liikuntasovelluksen (kuva 3) hyväksymistä ja sen käyttöä. Venkateshin ja kollegoiden (2012) teoreettinen UTAUT2-malli kuvaa kuluttajien uuden teknologian hyväksymistä ja sen käyttöä, joten tutkimuskysymys mallinnetaan käyttämällä UTAUT2-mallia, mutta muunnettuna. Kuvassa 4 on esitetty sekä Venkateshin ja kollegoiden (2012) alkuperäinen UTAUT2-malli että muunnettu UTAUT2^{**}-malli. Venkateshin ja kollegoiden (2012) alkuperäinen UTAUT2-malli käsittää seitsemän selittävää ja kaksi selitettävää faktoria. Selittävät faktorit, *suorituskykyodotukset* (PE), *vaivattomuusodotukset* (EE), *sosiaalinen vaikutus* (SI), *mahdollistavat olosuhteet* (FC), *hedonistinen motivaatio* (HM), *hint* (PV) ja *tottumus* (HT), selittävät teknologian *käyttöaikomusta* (BI) ja käyttöaikomuksen kautta myös varsinaista *käyttöä* (UB). Alkuperäisessä UTAUT2-mallissa on myös kolme moderoivaa muuttujaa: ikä, sukupuoli ja kokemus. Kuvassa 4 moderoivia muuttujia ei kuitenkaan ole esitetty, koska niitä ei sovelleta UTAUT2^{**}-malliin.

Myös Makkonen ja kollegat (2020) käyttivät myös muunnettua UTAUT2-mallia tutkiessaan Digital-Wells-ohjelman DW-liikuntasovelluksen käyttöaikomusta. Kohderyhmänä oli ohjelman kesällä 2019 aloittaneet nuoret ikäihmiset eli ohjelman ensimmäisen aallon osanottajat. Makkosen ja kollegoiden (2020) muunnettu UTAUT2-malli käsitti kuvan 4 viisi selittävää faktoria PE, EE, HM, HT ja SI, jotka yhdessä selittivät DW-liikuntasovelluksen käyttöaikomusfaktoria BI.

Makkosen ja kollegoiden (2020) muunnetusta UTAUT2-mallista poiketen tässä tutkielmassa DW-liikuntasovelluksen käyttöaikomusta BI selitetään ainoastaan neljällä faktorilla: PE, EE, HM ja HM. DW-liikuntasovelluksen käyttöä UB ei tarkastella lainkaan. Moderoivia muuttujia ei myöskään käytetä. Kuva 4 esittää muunnetun UTAUT2^{**}-mallin siten, että mallitarkastelun ulkopuolella olevat faktorit on kuvattu harmaana.

Tutkielman UTAUT2^{**}-malli ei sisällä mahdollistavien olosuhteiden faktoria FC, sillä jokaista ohjelmaan osallistunutta nuorta ikäihmistä opastettiin DW-liikuntasovelluksen käyttöön otossa. Ongelmatilanteissa osallistuja sai ohjelman tutkijoilta teknistä tukea. DW-liikuntasovellus oli osallistujille maksuton, joten faktori PV (hint) on rajattu pois.

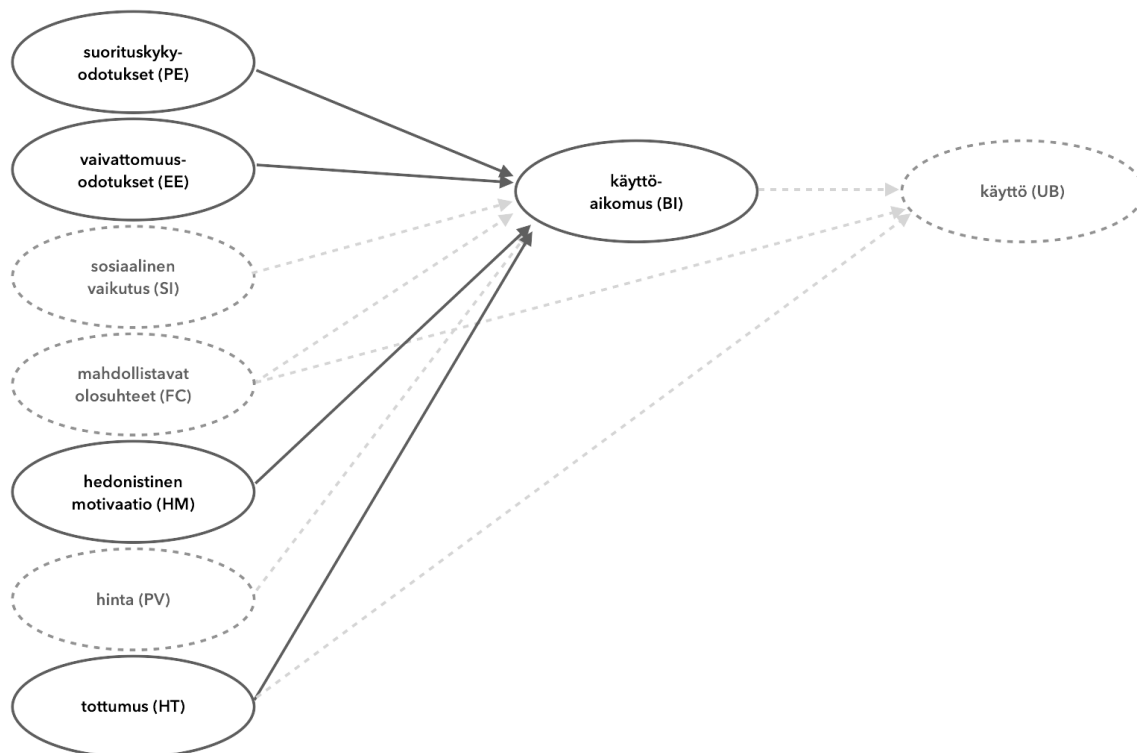
Sosiaalista vaikutusta kuvaava faktori SI on myös rajattu mallitarkastelun ulkopuolelle. Keväällä 2020 DigitalWells-ohjelman osanottajat käyttivät DW-liikuntasovellusta omien liikuntasuoritusten taltioimiseen lähinnä ohjelman tutkijoiden kannustamana, ei niinkään lähipiirin tai ystävien vuorovaikutuksesta. Covid-19-pandemiasta johtuvat rajoitustoimet eivät mahdollistaneet luonnollista vuorovaikutusta ystävien ja tuttavien kanssa, joten aitoa sosiaalista vaikutusta ei tapahtunut. Näin ollen seurantatutkimuksen sosiaalista vaikutusta mittaavat kysymykset koettiin keväällä 2020 osin epätarkkoina. Toisen aallon osallistujista lähes kolmannes vastasi jokaiseen neljään, sosiaalista vaikutusta mittaavaan kysymykseen vastausvaihtoehdolla ”en osaa sanoa”.

Tutkielmaa varten rakennetun UTAUT2^{**}-mallin faktorit ja niihin liittyvät havaitut muuttujat on esitetty taulukossa 2. Jokaisen faktorin kohdalla on listattuna neljä havaittua muuttujaa, joista yksi on merkitty tähdellä. Luvussa 5.3 tarkasteltavat rakenneyhtälömallit on muodostettu siten, että malliversioissa *a* jokaista faktoria kohden on kolme havaittua muuttujaa; taulukossa 2 tähdellä merkityt havaitut muuttujat on pudotettu pois. Malliversioissa *b* on puolestaan neljä havaittua muuttujaa jokaista UTAUT2^{**}-mallin faktoria kohden.

Taulukko 2: UTAUT2**-mallin faktorit ja havaitut muuttujat

PE	Suorituskykyodotukset (<i>performance expectancy</i>): ”Miten järjestelmän käyttö(aikomus) auttaa parantamaan työsuoritusta”
PE1	DW-sovellus on hyödyllinen jokapäivittäisessä elämässäni.
PE2*	Saavutan liikunnalliset tavoitteeni varmemmin käyttämällä DW-sovellusta.
PE3	Saavutan liikunnalliset tavoitteeni nopeammin käyttämällä DW-sovellusta.
PE4	Saavutan liikunnalliset tavoitteeni tehokkaammin käyttämällä DW-sovellusta.
EE	Vaivattomuusodotukset (<i>effort expectancy</i>): ”Missä määrin järjestelmän käyttö(aikomus) on vaivatonta, helppoa”
EE1	On helppo oppia käyttämään DW-sovellusta.
EE2*	DW-sovelluksen käyttäminen on selkeää ja helposti ymmärrettävää.
EE3	DW-sovelluksen käyttäminen on helppoa.
EE4	On helppoa tulla taitavaksi DigitalWells-sovelluksen käyttäjäksi.
HM	Hedonistinen motivaatio (<i>hedonistic motivation</i>): ”Missä määrin järjestelmän käyttö(aikomus) on hauskaa, tuottaa hupia”
HM1	DW-sovelluksen käyttäminen on hauskaa.
HM2*	DW-sovelluksen käyttäminen on ilahduttavaa.
HM3	DW-sovelluksen käyttäminen on viihdyttävää.
HM4	DW-sovelluksen käyttäminen on mukavaa.
HT	Tottumus (<i>habit</i>): ”Missä määrin järjestelmän käyttö(aikomus) nojaa aiemmin opittuun, tapaan”
HT1	DW-sovelluksen käyttämisestä on tullut tavanomaista.
HT2	Tunnen pakonomaista tarvetta käyttää DW-sovellusta.
HT3	Koen, että minun täytyy käyttää DW-sovellusta.
HT4*	DW-sovelluksen käyttämisestä on tullut luontevaa.
BI	Käyttöaikomus (<i>behavioural intention</i>): ”Järjestelmän käyttö(aikomus)”
BI1	Aion tulevaisuudessa jatkaa DW-sovelluksen käyttämistä.
BI2	Yritän jatkaa DW-sovelluksen käyttämistä jokapäiväisessä elämässäni.
BI3	Suunnittelen jatkavani DW-sovelluksen säännöllistä käyttämistä.
BI4*	Tulen todennäköisesti jatkamaan DW-sovelluksen käyttöä tulevaisuudessa.

Huomautus: *-merkki = havaittu muuttuja ei ole mukana malliversioissa *a*



Kuva 4: Teoreettinen UTAUT2**-malli, jossa on mukana Venkateshin ja kollegoiden (2012) alkuperäisen UTAUT2-mallin faktoreista vain viisi (mustat ellipsit). Mallitarkastelun ulkopuolella olevat faktorit on kuvattu harmailla, katkoviivaisilla ellipseillä. Faktorien väliset regressiopolut on kuvattu yksisuuntaisilla nuolilla, joista mustat ovat mallitarkastelussa mukana ja harmaat katkoviivalliset tarkastelun ulkopuolella.

5.1 Tutkimuskysymykset ja -hypoteesit

Luvun 5.3 mallitarkastelujen avulla pyritään löytämään paras valituista puuttuvuutta korjaavista menetelmistä siten, että tuotettu UTAUT2**-rakenneyhtälömalli sopii korjattuun havaintoaineistoon mahdollisimman hyvin. Saatu tulos vastaa ensimmäiseen tutkimuskysymykseen:

1. Onko jokin valituista viidestä havaintoaineiston puuttuvuutta korjaavista menetelmistä riittävän hyvä UTAUT2**-rakenneyhtälömallinnukseen tarkasteltaessa yleisiä sopivuusmittareita?

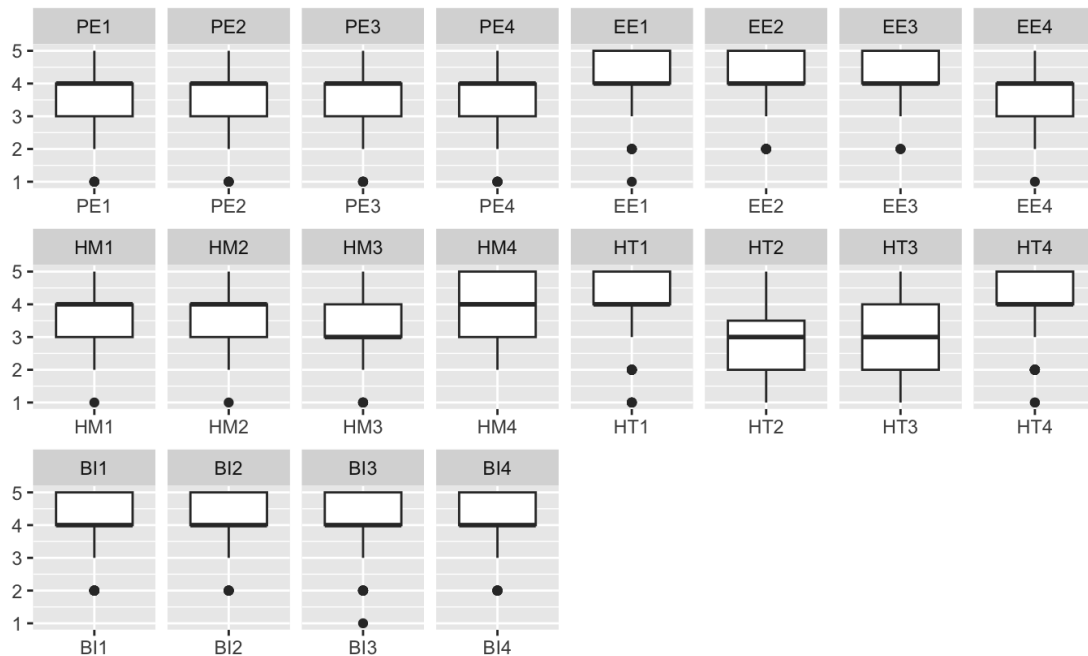
UTAUT2**-rakenneyhtälömallin avulla mallinnetaan DigitalWells-ohjelman nuorten ikäihmisten uuden liikuntasovelluksen käyttöaikomusta. Suorien yhteyksien testaamiseksi asetettiin seuraavat tutkimushypoteesit:

2. $H_{01} : \gamma_{11} = 0$; $H_{11} : \gamma_{11} \neq 0$ DW-liikuntasovelluksen käyttöaikomus edellyttää, että suorituskykyodotukset täyttyvät.
3. $H_{02} : \gamma_{12} = 0$; $H_{12} : \gamma_{12} \neq 0$ DW-liikuntasovelluksen käyttöaikomus edellyttää, että vaivattomuusodotukset täyttyvät.

4. $H_{03} : \gamma_{13} = 0$; $H_{13} : \gamma_{13} \neq 0$ DW-liikuntasovelluksen käyttöaikomus edellyttää riittävää hedonistista motivaatiota.
5. $H_{04} : \gamma_{14} = 0$; $H_{14} : \gamma_{14} \neq 0$ DW-liikuntasovelluksen käyttöaikomus edellyttää riittävää tottumusta.

5.2 Havaintoaineiston tarkastelu

Rakenneyhtälömallinnuksessa keskeisenä oletuksena on, että muuttujien jakaumat ovat normaalisia. Tarkastelu on hyvä aloittaa rakenneyhtälömallinnuksessa käytettävien havaittujen muuttujien jakaumien graafisella tarkastelulla. Taulukosta 2 saatujen 20 havaittujen muuttujien jakaumat on esitetty kuvassa 5. Oletus muuttujien jakaumien normaalisuudesta näyttää toteutuvan vain yhden havaitun muuttujan (HT3) kohdalla.



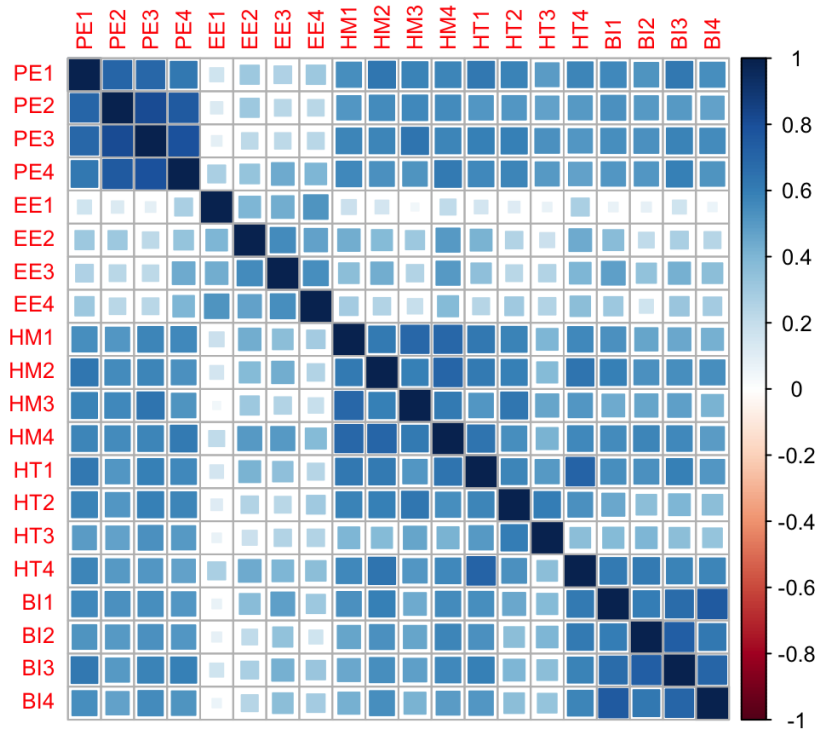
Kuva 5: Tarkasteltavan havaintoaineiston muuttujien laatikko–jana-kuviot. Oletus muuttujien normaalijakautuneisuudesta näyttää toteutuvan vain yhden muuttujan (HT3) kohdalla. Havaintoaineisto on DigitalWells-ohjelman kyselytutkimuksen osa-aineisto sisältäen 20 havaittua muuttujaa. $N = 201$.

Rakenneyhtälömallinnuksessa käytettävän havaintoaineiston faktorirakennetta voidaan yksinkertaisimmillaan tarkastella korrelaatiokaavion kautta. Korrelaatiokaaviosta nähdään mitattujen suureiden välinen korrelaatorakenne. Kuvassa 6 on esitettyä kaikki taulukossa 2 listattua 20 havaittua muuttujaa. Neljän peräkkäisen havaitun muuttujan vahva korrelaatorakenne kertoo taustalla vaikuttavasta faktorista. Täten korrelaatiokaaviosta halutaan löytyvän tummia, erillisiä tiivistymiä.

Kuvasta 6 nähdään myös, että faktoriin suorituskyky (PE) liittyvät havaitut muuttujat PE1, PE2, PE3 ja PE4 korreloivat vahvasti keskenään. Vastaavasti havaittujen muuttujien

BI1, BI2, BI3 ja BI4 välinen korrelaattiorakenne on vahva. Sen sijaan havaittujen muuttujien EE1, EE2, EE3, ja EE4 keskinäiset korrelaatiot ovat suhteellisen vaatimattomia.

Taulukossa 2 esitetyt 20 havaittua muuttujaa on mitattu käyttämällä 5-portaista Likert-arvoasteikkoa, jossa vastausvaihtoehto ”täysin eri mieltä” vastaa arvoa 1 ja ”täysin samaa mieltä” arvoa 5. Vaikka Likertin skaalalla mitatut muuttujat ovat ordinaalisia, tämän tutkielman yhteydessä ne tulkitaan jatkuvien muuttujien approksimaatioiksi tiedostaen tulkintaan liittyvät ongelmat (Tarka, 2017). Likertin arvoasteikon lisänä oli myös vastausvaihtoehto ”en osaa sanoa”, joka aineiston käsittelyn yhteydessä muutettiin puuttuvaksi tiedoksi.



Kuva 6: Tarkasteltavan havaintoaineiston muuttujien korrelaatiokaavio. Neljän peräkkäisen havaitun muuttujan vahva korrelaattiorakenne kertoo taustalla vaikuttavasta faktorista. Havaintoaineisto on DigitalWells-ohjelman kyselytutkimuksen osa-aineisto sisältäen 20 havaittua muuttujaa. $N = 201$.

5.3 Tarkasteltavat rakenneyhtälömallit

Kuten luvussa 5 mainittiin, Makkonen ja kollegat (2020) mallinsivat DW-liikuntasovelluksen käyttöaikomusta BI viidellä selittävällä faktorilla: PE (suorituskykyodotukset), EE (vaivattomuusodotukset), SI (sosiaalinen vaikutus), HM (hedonistinen motivaatio) ja HM (tottumus). Moderoivia tekijöitä ei käytetty eikä DW-liikuntasovelluksen käyttöä UB mallinnettu. Jokainen faktori oli yhteydessä kolmeen havaittuun muuttujaan, ja jokaista havaittua muuttujaa selitti vain yksi faktori. Rakenneyhtälömallinnus toteutettiin käyttämällä menetelmää PLS-SEM ja puuttuvaa tietoa paikattiin analyysivaiheessa keskiarvoimputoinnilla. Makkosen ja kollegoiden (2020) rakenneyhtälömalli toimii lähtökohtana tutkielmassa tarkasteltaville rakenneyhtälömalleille sekä valituille puuttuvuutta korjaaville

menetelmille. Jokainen tarkasteltava rakenneyhtälömalli luodaan ensiksi niin, että jokainen faktori selittää kolmea havaittua muuttujaa, ja jokaista havaittua muuttujaa selittää vain yksi faktori. Tällaisia rakenneyhtälömalleja merkitään kirjaimella *a*. Tarkastelut toistetaan käyttämällä neljää havaittua muuttujaa jokaista faktoria kohden. Näitä rakenneyhtälömalleja merkitään kirjaimella *b*. Taulukossa 2 on merkitty tähdellä ne muuttujat, jotka puuttuvat kolmen muuttujan faktorirakenteesta eli malliversioista *a*. Rakenneyhtälömallien analyysivaiheessa tarkastellaan saatuja modifikaatioindeksejä, mutta tässä tutkielmassa niitä ei sovelleta tarkasteltaviin rakenneyhtälömalleihin. Täten varmistetaan, että tarkasteltavat rakenneyhtälömallit ovat keskenään vertailukelpoisia.

Rakenneyhtälömallinnusta varten havaintoaineistosta muodostettiin osa-aineisto, joka sisälsi ainoastaan yksilöä kuvaavan numeroinnin sekä taulukossa 2 esitetyt 20 havaittua muuttujaa. Osa-aineisto tarkistettiin oudokkien sekä jakaumien osalta. Muodostettua osa-aineistoa käytettiin myös eri imputointimenetelmien yhteydessä; imputoinnoissa ei hyödynnetty taustamuuttujia. Liitteen taulukko T3 esittää osa-aineiston 20 havaitun muuttujan tunnusluvut sekä puuttuvan tiedon osuuden.

Tässä tutkielmassa rakenneyhtälömallinnukset, rakenneyhtälömallien analysointi sekä puuttuvan tiedon paikkaus on toteutettu R-kielillä (versio "Already Tomorrow" (4.3.0)), R-Studiolla (versio "Cherry Blossom" (6e31ffc3)) sekä käyttöjärjestelmäversiolla MacOS Ventura (Versio 13.4 Beta (22F5037d)). Rakenneyhtälömallinnukseen käytettiin lapaan-pakettia, eri imputointimenetelmien toteuttamisessa *mi*ce-pakettia ja PLS-SEM-menetelmän implementoinnissa puolestaan *semr*nr-pakettia. Työtä sujuvoitettiin *tidyverse*-paketilla. Kyselytutkimuksen tietokantatyö tehtiin IBM SPSS -ohjelmistolla (versio 28.0.0).

5.3.1 Mallit 1a ja 1b – täydelliset havaintorivit

Rakenneyhtälömallinnus toteutettiin kuvan 4 mukaisesti siten, että mallissa 1a jokainen faktori selitti kolmea havaittua muuttujaa ja mallissa 1b jokainen faktori selitti neljää havaittua muuttujaa. Jokaisen faktorin ensimmäinen havaittu muuttuja kiinnitettiin ykköseksi. Rakenneyhtälön analysointi toteutettiin täydellisillä havaintoriveillä. Täten malli 1a käsitti 134 yksilöä ja malli 1b puolestaan 127 yksilöä.

Aineistotarkastelun yhteydessä (kuva 5) ilmeni, että oletus muuttujien normaalijakautuneisuudesta ei toteudu. Näin ollen mallin estimoinnissa käytettiin robustia MLM-menetelmää.

Mallien 1a ja 1b sopivuustarkastelut toteutettiin käyttämällä osiossa 2.6 esitettyjä yleisiä sopivuusmittareita. Sopivuustarkastelun mallikohtaiset robustit tulokset on esitetty taulukossa 3.

5.3.2 Mallit 2a ja 2b – ehdollistamaton keskiarvoimputointi

Rakenneyhtälömallinnus toteutettiin kuvan 4 mukaisesti siten, että mallissa 2a jokainen faktori selitti kolmea ja mallissa 2b jokainen faktori selitti neljää havaittua muuttujaa. Jokaisen faktorin ensimmäinen havaittu muuttuja kiinnitettiin ykköseksi. Aineiston puuttuvaa tietoa paikattiin ehdollistamattomalla keskiarvoimputoinnilla. Imputointi suoritettiin ennen rakenneyhtälön analysointia.

Mallin estimoinnissa käytettiin robustia MLM-menetelmää. Sopivuustarkastelun mallikohtaiset robustit tulokset on esitetty taulukossa 3.

5.3.3 Mallit 3a ja 3b – moni-imputointi

Rakenneyhtälömallinnus toteutettiin kuvan 4 mukaisesti siten, että mallissa 3a jokainen faktori selitti kolmea ja mallissa 3b jokainen faktori selitti neljää havaittua muuttujaa. Jokaisen faktorin ensimmäinen havaittu muuttuja kiinnitettiin ykköseksi. Aineiston puuttuvaa tietoa paikattiin moni-imputoinnin avulla käyttämällä pmm-menetelmää. Simulointi toistettiin sekä viisi että kymmenen kertaa, ja tuotetuista aineistoista saatuja keskiarvoja käytettiin puuttuvan tiedon paikkaamiseen. Imputointi suoritettiin ennen rakenneyhtälön analysointia. Kymmenkertaisella simuloinnilla saatiin tuotettua parempia puuttuvuuden korjaamisessa käytettyjä estimaatteja. Täten mallien 3a ja 3b sopivuustarkastelun robustit tulokset pohjautuvat kymmenkertaiseen simulointiin. Sopivuustarkastelun robustit tulokset on esitetty taulukossa 3. Mallin estimoinnissa käytettiin robustia MLM-menetelmää.

5.3.4 Mallit 4a ja 4b – FIML

Rakenneyhtälömallinnus toteutettiin kuvan 4 mukaisesti siten, että mallissa 4a jokainen faktori selitti kolmea ja mallissa 4b jokainen faktori selitti neljää havaittua muuttujaa. Jokaisen faktorin ensimmäinen havaittu muuttuja kiinnitettiin ykköseksi. Aineiston puuttuvaa tietoa ei paikattu ennen rakenneyhtälön analysointia, vaan valittu FIML-menetelmän algoritmi käsitteli puuttuvuutta analyysin yhteydessä. Mallin estimoinnissa käytettiin yleistä SU-menetelmää. Sopivuustarkastelun mallikohtaiset tulokset ovat taulukossa 3.

5.3.5 Mallit 5a ja 5b – PLS-SEM

Rakenneyhtälömallinnus toteutettiin kuvan 4 mukaisesti siten, että mallissa 5a jokainen faktori selitti kolmea ja mallissa 5b neljää havaittua muuttujaa. Aineiston puuttuvaa tietoa ei paikattu ennen rakenneyhtälön analysointia, vaan valittu PLS-SEM-menetelmän algoritmi käsitteli puuttuvuutta analyysin yhteydessä keskiarvoimputoinnin avulla. Hair ja kollegat (2017) ohjeistavat poistamaan analyysistä muuttujat, joissa on puuttuvaa tietoa yli 15 %. Taulukosta T3 nähdään, että yhdenkään rakenneyhtälömallinnuksessa käytetyn muuttujan puuttuvuuden osuus ei ylittänyt mainittua ”kynnysarvoa”.

Rakenneyhtälömallinnus toteutettiin Makkosen ja kollegoiden (2020) PLS-SEM-menetelmävalintoja seuraten. Mallien 5a ja 5b painotukset perustuivat regressiopolkujen painotuksiin ja aloituspainokertoimeksi valittiin +1. Bootstrap-menetelmän avulla aineistosta tuotettiin 2500 osa-aineistoa.

Sopivuustarkastelun mallikohtaiset tulokset on esitetty erillisessä taulukossa 4, sillä PLS-SEM-menetelmä tarkastelee rakenneyhtälömallin sopivuutta havaintoaineistoon eri tavalla (luku 3.4.5) kuin ”tavallinen”rakenneyhtälömallinnus. Taulukossa 4 faktorin reliabiliteetti on ilmaistu sarakkeessa ρ_C . Molemmissa malleissa 5a ja 5b jokaisen faktorin reliabiliteetti ρ_C on yli 0,7.

Taulukosta 4 saadaan jokaisen faktorin valideettia kuvaava suhteellinen selitysosuus AVE (*average variance extracted*), jonka pitäisi olla vähintään 50 % (Hair ym., 2019). Faktorin suhteellisen selitysosuuden neliöjuuri on ilmaistu taulukossa 4 diagonaalisesti ja lihavoituna. Suureen tulisi olla suurempi kuin faktorin korrelaatio muiden faktorien kanssa (Hair ym., 2019). Molemmat ehdot toteutuvat jokaisen tarkasteltavan faktorin osalta malleissa 5a ja 5b.

Taulukko 3: Osa-aineiston ja UTAUT2**-mallien yhteensopivuus I.

MLM	N	$S - B\chi^2$	df	p-arvo	CFI	TLI	AIC	RMSEA	SRMR
malli 1a	134	165.28	80	<0.01	0.93	0.91	4443.63	0.09	0.06
S- $B\chi^2$ /df	2.07								
skaalaustekijä	1.08								
malli 1b	127	295.0	160	<0.01	0.92	0.90	5315.67	0.09	0.06
S- $B\chi^2$ /df	1.84								
skaalaustekijä	1.10								
malli 2a	201	140.53	80	<0.01	0.96	0.94	6836.06	0.07	0.05
S- $B\chi^2$ /df	1.76								
skaalaustekijä	1.20								
malli 2b	201	272.96	160	<0.01	0.95	0.94	8712.40	0.06	0.06
S- $B\chi^2$ /df	1.71								
skaalaustekijä	1.16								
malli 3a	201	153.31	80	<0.01	0.96	0.94	6830.90	0.07	0.05
S- $B\chi^2$ /df	1.92								
skaalaustekijä	1.11								
malli 3b	201	326.29	160	<0.01	0.94	0.92	8699.49	0.08	0.06
S- $B\chi^2$ /df	2.04								
skaalaustekijä	1.11								
SU	N	χ^2	df	p-arvo	CFI	TLI	AIC	RMSEA	SRMR
malli 4a	201	148.37	80	<0.01	0.98	0.94	6504.59	0.07	0.05
χ^2 /df	1.85								
skaalaustekijä	-								
malli 4b	201	302.53	160	<0.01	0.94	0.93	8268.63	0.07	0.06
χ^2 /df	1.89								
skaalaustekijä	-								

MLM = robusti SU-menetelmä; taulukon arvot robusteja arvoja, joissa Satorra–Bentler skaalaus
 SU = SU-menetelmä; taulukon arvot ei-skaalattuja

Mallien 5a ja 5b muuttujilla ei esiintynyt multikollinearisuutta (taulukko 5). Taulukossa 5 sarakenumero yhdessä rivin kirjainlyhenteen kanssa muodostaa muuttujan nimen, kuten PE1. Mallien 5a ja 5b jokaisen yksittäisen havaitun muuttujan faktorilataus on esitettyä liitteen taulukossa T4. Ainoastaan muuttujan EE1 faktorilataus mallissa 5b jäi alle 0,7 (arvo 0,66). Muuttujaa EE1 ei poistettu mallitarkastelusta, sillä mallivertailut haluttiin tehdä mahdollisimman yhteneväisesti.

Taulukko 4: Osa-aineiston ja UTAUT2**-mallien yhteensopivuus II. N = 201.

PLS-SEM	Faktori	ρ_C	AVE	PE	EE	HM	HT	BI
malli 5a (3 muuttujaa)								
	PE	0.92	0.79	0.89				
	EE	0.85	0.66	0.32	0.81			
	HM	0.91	0.77	0.71	0.34	0.88		
	HT	0.86	0.67	0.74	0.33	0.69	0.82	
	BI	0.91	0.77	0.68	0.36	0.61	0.63	0.87
malli 5b (4 muuttujaa)								
	PE	0.94	0.79	0.89				
	EE	0.86	0.61	0.36	0.78			
	HM	0.92	0.74	0.72	0.43	0.86		
	HT	0.88	0.64	0.75	0.43	0.76	0.80	
	BI	0.92	0.75	0.68	0.39	0.64	0.66	0.87

ρ_C = faktorin reliabiliteetti; tulisi olla vähintään 0,7

AVE = faktorin valideetti; tulisi vähintään 0,5

Taulukko 5: Multikollineaarisuutta ilmaiseva VIF-kerroin. N = 201.

PLS-SEM	1	2	3	4
malli 5a (3 muuttujaa)				
PE	2.09	-	2.99	2.27
EE	1.54	-	1.51	1.56
HM	2.43	-	1.89	2.10
HT	1.43	1.71	1.52	-
BI	2.20	1.77	2.50	-
malli 5b (4 muuttujaa)				
PE	2.24	3.38	3.72	2.49
EE	1.55	1.62	1.84	1.61
HM	2.56	2.26	1.98	2.48
HT	2.19	1.77	1.52	2.09
BI	2.81	1.83	2.69	2.71

huomautus: PE + 1 = PE1 havaittu muuttuja

5.4 Mallien sopivuus aineistoon

Mikäli mallien 1a–4b sopivuustarkastelut tehdään khi-neliösuuretta χ^2 käyttäen, mikään malleista ei ole aineistoon sopiva. Jos saatu khi-neliöarvo suhteutetaan mallin vapausasteisiin, mallien 1a–4b kohdalla kynnysarvo 5 ei ylity (Halme, ym., 2014). Sopivuusmittarit CFI (vertaileva sopivuusindeksi) ja TLI (Tucker–Lewis-indeksi) ovat jokaisen tarkasteltavan mallin kohdalla yli 0,90. Ainoastaan mittarille RMSEA (keskineliövirheen neliöjuuri) asetettu ehto, että saadun arvon tulisi olla alle 0,06, ei täyty yhdenkään mallin (1a–4b) kohdalla. Alin saatu arvo on 0,06, ja se toteutuu mallin 2b kohdalla. Lisäksi mallissa 2b saadut arvot CFI = 0,95 ja TLI = 0,94 viestivät mallin hyvästä sopivuudesta aineistoon. Ehdollistamaton keskiarvoimputointi mallissa 2b onnistui paikkaamaan aineistossa ollutta puuttuvuutta hyvin. Mallit 4a ja 4b toimivat myös hyvin, vaikka arvo RMSEA = 0,07 on molemmilla malleilla yli toivotun kynnysarvon. Saadut CFI- ja TLI-arvot ovat molemmissa malleissa korkeat, kynnysarvon 0,90 ylittäviä (taulukko 3). Lisäksi mallin 4b AIC-arvo on pienempi kuin mallin 2b. Kokonaisuutta tarkasteltaessa sekä mahdollisia jatkoanalyysejä ajatellen FIML-menetelmä on robustimpi kuin ehdollistamaton keskiarvoimputointi.

Perustuen mallien 4a ja 4b sopivuusmittareihin molemmat kyseiset mallit näyttävät sopivan, mutta korkeintaan kohtalaisesti. Mallien 4a ja 4b muuttujien mahdollinen multikollineaarisuus voidaan tarkistaa korrelaatiotaulukoiden avulla (liitteen taulukot T5 ja T6). Hyvin voimakkaita korrelaatioita muuttujien välillä ei ole nähtävissä, joten multikollineaarisuutta ei esiinny. Voidaan todeta, että mallien 4a ja 4b UTAUT2**-rakenneyhtälömallit sopivat molemmat tarkasteltavaan aineistoon kohtalaisesti.

Sopivuustarkastelun perusteella myös mallit 5a ja 5b (taulukko 4) sopivat tarkasteltavaan aineistoon. Kuitenkin mallissa 5b muuttujan EE1 faktorilataus oli hieman alle 0,7 (0,663). Malli 5b ei näin ollen ole erityisen sopiva. Voidaan siis todeta, että mallien 5a ja 5b UTAUT2**-rakenneyhtälömallit sopivat molemmat DigitalWells-aineistoon kohtalaisesti, kun tarkastelu toteutettiin epäparametrisella PLS-SEM-menetelmällä.

Kun tarkastellaan taulukkojen 3 ja 4 yleisiä sopivuusmittareita, nähdään jokaisen tarkasteltavan mallin yhtäläinen sopivuus DigitalWells-kyselyaineistosta muodostettuun osa-aineistoon. Aineistossa esiintyvä puuttuvuus on häiriötekijä, mutta se ei yksistään selitä UTAUT2**-rakenneyhtälömallien kohtalaista sopivuutta tarkasteltavaan osa-aineistoon. Varsinainen ongelma on UTAUT2**-mallissa, ja etenkin UTAUT2-mallin muodostavissa havaituissa muuttujissa. Havaitut muuttujat ovat joukko melko standardoituja väittämiä, joiden käyttöä edellytetään UTAUT2-mallintamisessa. Kuitenkin kyseiset väittämät eivät välttämättä sovi hyvin nykyiseen teknologiakehikkoon, joten havaittujen muuttujien mittaamiseen liittyy epävarmuutta. Täten seuraava askel olisi tarkastella, kuinka Venkateshin ja kollegoiden (2012) alkuperäistä UTAUT2-mallia tulisi kehittää ja päivittää.

5.4.1 Parametrisella menetelmällä saadut estimaatit

Tarkastellaan seuraavaksi mallien 4a ja 4b parametrien estimaatteja. Edellä luvussa 5.4 todettiin, että sopivuustarkastelun perusteella molemmat rakenneyhtälömallit 4a ja 4b sopivat tarkasteltavaan aineistoon kohtalaisesti. Kuvasta K1 nähdään mallien 4a ja 4b residuaalien korrelaatiomatriisit. Jos itseisarvoltaan 0,1 suurempia arvoja on havaittavissa, residuaalien korreloivia muuttujia on syytä tarkastella yksityiskohtaisemmin. Kuvasta K1 voidaan nähdä, että esimerkiksi molemmissa malleissa 4a ja 4b muuttujien HM3 ja EE1 välinen korrelaatio on suurempaa kuin muilla muuttujilla. Täten tarkasteltaviin

rakenneyhtälömalleihin 4a ja 4b mahdollisesti kannattaisi tehdä muunnos.

Mallien 4a ja 4b estimaatit on esitetty taulukossa 6. Taulukosta 6 nähdään, että molemmissa malleissa 4a ja 4b ainoastaan suorituskykyodotusten (PE) vaikutus DW-liikuntasovelluksen käyttöaikomukseen (BI) oli tilastollisesti merkitsevä. Saatu tulos tukee päätelmää siitä, että UTAUT2**-malli sopii tarkasteltavaan DigitalWells-kyselyaineiston osa-aineistoon korkeintaan kohtalaisesti. Kuvassa 7 nähdään UTAUT2**-malli toteutettuna mallin 4a mukaisesti.

5.4.2 Epäparametrisella menetelmällä saadut estimaatit

Taulukosta 7 nähdään epäparametrisella PLS-SEM-menetelmällä saadut estimaatit. Mallien 5a ja 5b estimaatit ovat hyvin samanlaisia, mutta mallin 5b selitysaste on hieman korkeampi kuin mallin 5a. Molemmissa malleissa 5a ja 5b suorituskykyodotusten (PE) sekä tottumuksen (HT) vaikutukset DW-liikuntasovelluksen käyttöaikomukseen (BI) ovat molemmat tilastollisesti merkittäviä. Kuitenkin rakenneyhtälömallin melko alhainen selitysaste R^2 yhdessä mallien 5a ja 5b estimaattien kanssa tukee päätelmää siitä, että UTAUT2**-malli sopii tarkasteltavaan DigitalWells-kyselyaineiston osa-aineistoon korkeintaan kohtalaisesti. Mallin 5b graafinen esitys on nähtävissä kuvassa 7.

Taulukkojen 6 ja 7 sopivuusmittareiden perusteella tuotetut kymmenen UTAUT2**-rakenneyhtälömallia sopivat jokainen DigitalWells-kyselyaineiston osa-aineistoon kohtuullisesti. Havaintoaineiston puuttuvuuden korjaamiseksi valittu menetelmä ei varsinaisesti tuota erityisen hyvää sopivuutta minkään mallin kohdalla. Täten vastaus luvun 5.1 tutkimuskysymykseen on, että *yleisten sopivuusmittarien mukaan mikään valituista viidestä havaintoaineiston puuttuvuutta korjaavista menetelmistä ei ole riittävän hyvä UTAUT2**-rakenneyhtälömallinnuksessa*. Täten $\Sigma = \Sigma(\theta)$ ei toteudu.

Taulukko 6: Mallien 4a ja 4b estimaatit. N = 201.

MLM		estimaatti	keskivirhe	z	CI _{2.5%}	CI _{97.5%}	std. est.
malli 4a (3 muuttujaa)							
H ₀₁ :	PE → BI	0.51	0.24	2.12	0.04	0.97	0.55
H ₀₂ :	EE → BI	0.19	0.11	1.71	-0.03	0.41	0.13
H ₀₃ :	HM → BI	0.02	0.19	0.08	-0.36	0.39	0.02
H ₀₄ :	HT → BI	0.19	0.36	0.54	-0.51	0.90	0.20
malli 4b (4 muuttujaa)							
H ₀₁ :	PE → BI	0.42	0.12	3.46	0.18	0.66	0.44
H ₀₂ :	EE → BI	0.15	0.12	1.23	-0.09	0.40	0.09
H ₀₃ :	HM → BI	0.01	0.19	0.04	-0.36	0.37	0.01
H ₀₄ :	HT → BI	0.32	0.21	1.54	-0.08	0.72	0.34

std. est = standardoitu estimaatti

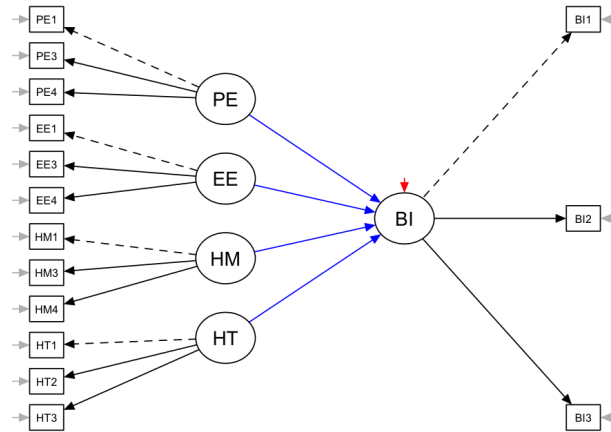
Taulukko 7: Mallien 5a ja 5b standardoidut estimaatit. Bootstrap_{otos} = 2500. N = 201.

PLS-SEM		estimaati	ka _{boot}	sd _{boot}	t	CI _{2.5%}	CI _{97.5%}
malli 5a (3 muuttujaa)							
H ₀₁ :	PE → BI	0.37	0.36	0.08	4.41	0.20	0.53
H ₀₂ :	EE → BI	0.11	0.12	0.06	1.86	0	0.24
H ₀₃ :	HM → BI	0.17	0.17	0.09	1.83	-0.01	0.35
H ₀₄ :	HT → BI	0.21	0.21	0.09	2.27	0.04	0.40
R ²	0.52						
korjattu R ²	0.51						
malli 5b (4 muuttujaa)							
H ₀₁ :	PE → BI	0.35	0.34	0.09	3.99	0.16	0.50
H ₀₂ :	EE → BI	0.09	0.10	0.07	1.33	-0.04	0.23
H ₀₃ :	HM → BI	0.17	0.17	0.10	1.61	-0.04	0.36
H ₀₄ :	HT → BI	0.24	0.25	0.10	2.29	0.06	0.47
R ²	0.54						
korjattu R ²	0.53						
ka _{boot} = keskiarvo; bootstrap-otosten avulla, otokset palautuksella							
sd _{boot} = keskihajonta; bootstrap-otosten avulla, otokset palautuksella							

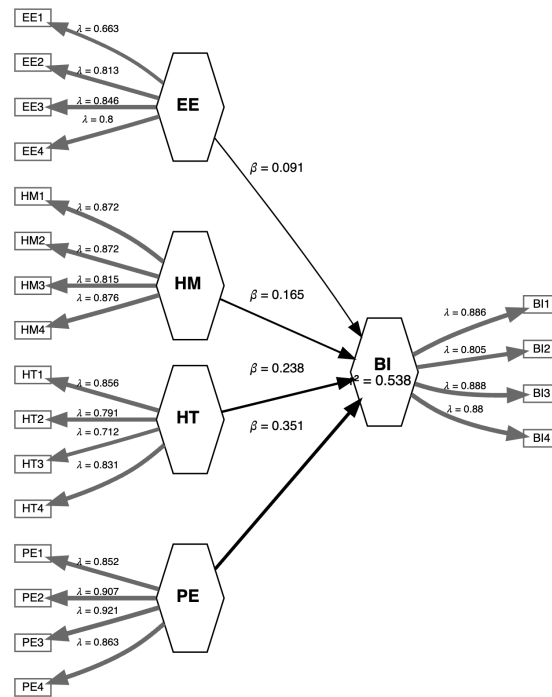
Taulukko 8: Asetetut tutkimushypoteesit ja PLS-SEM-menetelmällä tuotetun mallin 5b standardoidut estimaatit. N = 201.

	Tutkimushypoteesit	std. est (t-arvo)	p-arvo
H ₁₁ :	PE → BI (γ ₁₁ ≠ 0)	0.35 (3.99)	< 0.01
H ₁₂ :	EE → BI (γ ₁₂ ≠ 0)	0.09 (1.33)	0.09
H ₁₃ :	HM → BI (γ ₁₃ ≠ 0)	0.17 (1.61)	0.05
H ₁₄ :	HT → BI (γ ₁₄ ≠ 0)	0.24 (2.29)	0.01
std. est = standardoitu estimaatti; α = 0.05			
mallin 5b R ² = 0.54			

Tarkastellaan seuraavaksi asetettuja tutkimushypoteeseja ja mallin 5b standardoituja estimaatteja. Taulukosta 8 nähdään, että DigitalWells-kyselyaineiston osa-aineisto tuki tutkimushypoteesia H₁₁, jonka mukaan 'DW-liikuntasovelluksen käyttöaikomus edellyttää, että suorituskyydötukset täyttyvät'. Havaintoaineisto tuki myös tutkimushypoteesia H₁₄ eli 'DW-liikuntasovelluksen käyttöaikomus edellyttää riittävää tottumusta'. Sen sijaan tutkimushypoteesi H₁₂, 'DW-liikuntasovelluksen käyttöaikomus edellyttää, että väitvattomuusodotukset täyttyvät' ja H₁₃, 'DW-liikuntasovelluksen käyttöaikomus edellyttää riittävää hedonistista motivaatiota', eivät saaneet vahvistusta aineistosta.



(a) malli 4a



(b) malli 5b

Kuva 7: UTAUT2*-mallin mukainen rakenneyhtälömalli, jossa ylempi kuva (a) esittää mallia, jossa puuttuvuutta on korjattu FIML-menetelmän avulla. Jokaiselle faktorille latautuu kolme havaittua muuttujaa. Alemmassa kuvassa (b) on PLS-SEM-menetelmän avulla tuotettu rakenneyhtälömalli, jossa jokaiselle faktorille latautuu neljä havaittua muuttujaa. Alemmassa kuvassa (b) on nähtävillä myös tuotetut estimaatit. Havaintoaineistona on DigitalWells-kyselyaineiston osa-aineisto. $N = 201$.

6 Johtopäätökset ja pohdinta

Analyysimenetelmänä rakenneyhtälömallinnusta voidaan käyttää monipuolisesti ja joustavasti, sillä se soveltuu niin yleisen regressiomallin kuin myös monimutkaisemman hierarkisen mallin tarkasteluun. Tehokkaimmillaan rakenneyhtälömallinnus sekä tiivistää havaintoaineistoa että tarkastelee mallin selittäviä ja selitettäviä piilomuuttujia yhdessä muuttujien mahdollisten mittausvirheiden kanssa.

Tässä tutkielmassa tarkasteltiin rakenneyhtälömallinnuksen sopivuutta analyysimenetelmänä, kun aineistossa on puuttuvuutta. Tutkielmassa tuotettiin kymmenen erilaista rakenneyhtälömallia, joiden sopivuutta tarkasteltiin erilaisten sopivuusmittarien avulla. Lisäksi vertailtiin mallien tuottamia estimaatteja ja niiden sopivuutta. Tutkimusaineistona käytettiin vuosina 2019–2020 kerättyä DigitalWells-ohjelman toisen aalloon kyselytutkimusaineistoa.

Parametrisena menetelmänä luotettavimmat tulokset saatiin käyttämällä FIML-menetelmää. Lisäksi epäparametrinen menetelmä PLS-SEM toimi kohtuullisesti. Tutkielmassa tuotettu UTAUT2**-malli sopi tutkittavaan DigitalWells-kyselyaineiston osa-aineiston kohtuullisesti. Vaikka tarkasteltavassa aineistossa oli puuttuvuutta, varsinainen ongelma on kuitenkin UTAUT2**-mallissa. UTAUT2-malli tuotetaan joukolla melko standardoituja väittämiä, jotka eivät sovi hyvin nykyiseen teknologiakehikkoon. Täten havaittujen muuttujien mittaamiseen liittyy epävarmuutta. Seuraava askel olisikin tarkastella, että miten Venkateshin ja kollegoiden (2012) alkuperäistä UTAUT2-mallia tulisi kehittää ja päivittää.

Puuttuvan tiedon imputoinnissa käytettiin ainoastaan tarkastelun kohteena ollutta osa-aineistoa. Mahdollisesti taustamuuttujien käyttäminen imputoinnin yhteydessä olisi tuonut parempia tuloksia. Lisäksi käytettäessä satunnaismetsää imputointimenetelmänä on saatu mielenkiintoisia tuloksia. Tulevaisuudessa olisi hyvä tarkastella myös satunnaismetsää imputointimenetelmänä.

Lähdeluettelo

1. Astrachan, C.B., Patel, V.K., Wanzenried, G. (2014). A comparative study of CB-SEM and PLS-SEM for theory development in family firm research. *Journal of Family Business Strategy*, 5(1, s. 116–128). <https://doi.org/10.1016/j.jfbs.2013.12.002>. [Viitattu 11.5.2023]
2. Allison, P.D. (2003). Missing Data Techniques for Structural Equation Modeling. *Journal of Abnormal Psychology*, 112(4, s. 545–557). <https://pubmed.ncbi.nlm.nih.gov/14674868/>. [Viitattu 11.5.2023]
3. Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Yhdysvallat: Wiley.
4. Brown, T.A. (2015). *Confirmatory Factor Analysis for Applied Research*. (2. painos, s. 208, 309, 316). Yhdysvallat: The Guilford Press.
5. Byrne, B.M. (2012). *Structural Equation Modeling with Mplus*. (s. 3–124). Yhdysvallat: Routledge.
6. Carlsson, C., Walden, P., Kari, T., Makkonen, M., Frank, L. (2021). Forming Sustainable Physical Activity Programs Among Young Elderly : A Combined ELM & UTAUT Approach. Teoksessa: A. Pucihar, M. Kljajic Borstnar, R. Bons, H. Cripps, A. Sheombar, D. Vidmar (Eds.) *34th Bled eConference : Digital Support from Crisis to Progressive Change*. (s. 421–438). <https://aisel.aisnet.org/bled2021/34/>. [Viitattu 20.5.2023]
7. DigitalWells. (2019). DigitalWells-ohjelma. <https://digitalwells.fi>. [Viitattu 11.5.2023]
8. Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*(60, s. 549–576).
9. Hair, J., Hollingsworth, C.L., Randolph, A.B., Chong, A.Y.L. (2016). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems*, 117(3, s. 442–458). <https://doi.org/10.1108/IMDS-04-2016-0130>. [Viitattu 11.5.2023]
10. Hair, J.F., Jr., Sarstedt, M., Ringle, C.M., Gudergan, S.P. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. (s. 1–224). Yhdysvallat: Sage.
11. Hair, J.F., Jr., Sarstedt, M., Ringle, C.M., Gudergan, S.P. (2018). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. (s. 1–37). Yhdysvallat: Sage.
12. Hair, J.F., Risher, J.J. Sarstedt, M., Ringle, C.M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1, s. 2–24). <https://doi.org/10.1108/EBR-11-2018-0203>. [Viitattu 11.5.2023]

13. Hair, J., Hult, T., Ringle, C.M., Sarstedt, M., Danks, N., Ray, S. (2023). Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R. *GitHub*. <https://sem-in-r.github.io/seminr/>. [Viitattu 11.5.2023]
14. Halme, N., Kanste, O., Nummi, T., Perälä, M-L. (2014). Rakenneyhtälömallin kehittäminen ja arviointi – tutkimuksen kohteena avun antaminen lasten ja perheiden palveluissa. *Sosiaalilääketieteellinen Aikakauslehti*, 51(4, s. 272–288). <http://ojs.tsv.fi/index.php/SA/article/view/48474>. [Viitattu 11.5.2023]
15. Hayduk, L.A. (2014). Shame for disrespecting evidence: the personal consequences of insufficient respect for structural equation model testing. *BMC Medical Research Methodology*, 14(124). <https://doi.org/10.1186/1471-2288-14-124>. [Viitattu 11.5.2023]
16. Husu, P., Tokola, K., Vähä-Ypyä, H., Vasankari, T. (2022). Liikuntaraportti: Suomalaisen mitattu liikkuminen, paikallaanolo ja fyysinen kunto 2018–2022. *Opetus- ja kulttuuriministeriön julkaisuja*(33). <http://urn.fi/URN:ISBN:978-952-263-808-3>. [Viitattu 11.5.2023]
17. Härkänen, T. (2023). Puuttuvat havainnot. *Kvantitatiivisen tutkimuksen verkkokäsikirja*. Tampere: Yhteiskuntatieteellinen tietoaarkisto. <https://www.fsd.tuni.fi/fi/palvelut/metodelmaopetus>. [Viitattu 11.5.2023]
18. Jöreskog, K.G., van Thillo, M. (1972). LISREL: A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. *ETS Research Bulletin Series*, 1972(2, s. 1–72). <https://doi.org/10.1002/j.2333-8504.1972.tb00827.x>. [Viitattu 11.5.2023]
19. Kline, R.B. (2016). *Principles and Practice of Structural Equation Modeling*. (4. painos, s. 7–163, 188–211, 231–366, 452–468). Yhdysvallat: The Guilford Press.
20. Lee, T., Shi, D. (2021). A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data. *Psychol Methods*, 26(4, s. 466–485). <https://doi.org/10.1037/met0000381>. [Viitattu 11.5.2023]
21. Li, C. (2013). Little's test of missing completely at random. *The Stata Journal*, 13(4, s. 795–809). <https://journals.sagepub.com/doi/pdf/10.1177/1536867X1301300407>. [Viitattu 11.5.2023]
22. Little, R.J.A., Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2. painos, s.3–23). Yhdysvallat: Wiley.
23. Little, R.J.A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83(404, s. 1198–1202). <https://doi.org/10.2307/2290157>. [Viitattu 11.5.2023]

24. Makkonen, M., Kari, T., Frank, L. (2020). Applying UTAUT2 To Explain The Use of Physical Activity Logger Applications Among Young Elderly. *BLED 2020 Proceedings*(5). <https://aisel.aisnet.org/bled2020/5/>. [Viitattu 11.5.2023]
25. Metsämuuronen, J. (2008). *Monimuuttujamenetelmien perusteet* (2. painos, s. 59–84). Jyväskylä, Suomi: Gummerus.
26. Newman, D.A. (2014). Missing Data: Five Practical Guidelines. *Organizational Research Methods*, 17(4, s. 372–411). <https://doi.org/10.1177/1094428114548590>. [Viitattu 11.5.2023]
27. Nordberg, L. (2006). *Structural Equation Modeling (SEM)*. Opetusmateriaali. Åbo Akademi. Tilastotiede.
28. Ramlall, I. (2017). *Applied Structural Equation Modelling for Researches and Practitioners*. Yhdistynyt kuningaskunta: Emerald.
29. Saari, T-H. (2008). PLS-regressio kemometrian kalibrointiongelmassa. *Mat-2.108 Sovelletun matematiikan erikoistyöt*. Teknillinen korkeakoulu. http://salserver.org.aalto.fi/vanhat_sivut/Opinnot/Mat-2.4108/pdf-files/esaa08.pdf. [Viitattu 11.5.2023]
30. Satorra, A, Bentler, P.M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*(75, s. 243–248). <https://doi.org/10.1007/s11336-009-9135-y>. [Viitattu 11.5.2023]
31. Schumacker, R.E., Lomax, R.G. (2016). *A Beginner's Guide to Structural Equation Modeling* (4. painos, s. 1–140, 238–248). New York, NY: Routledge.
32. Shin, T., Long, J.D., Davison, M.L. (2022). An evaluation of methods to handle missing data in the context of latent variable interaction analysis: multiple imputation, maximum likelihood, and random forest algorithm. *Japanese Journal of Statistics and Data Science*(5, s. 629—659). <https://doi.org/10.1007/s42081-022-00176-w>. [Viitattu 11.5.2023]
33. Tenenhaus, M. (1998). *La Régression PLS - Théorie et Pratique*. (s. 233–244). Ranska: Éditions Technip.
34. Tarka, P. (2017). The comparison of estimation methods on the parameter estimates and fit indices in SEM model under 7-point Likert scale. *Archives of Data Science, Series A (Online First)*, 2(1). <https://doi.org/10.5445/KSP/1000058749/10>. [Viitattu 11.5.2023]
35. Tilastokeskus. (2021). Suomen virallinen tilasto (SVT): Väestöennuste. Helsinki: Tilastokeskus. http://www.stat.fi/til/vaenn/2021/vaenn_2021_2021-09-30_tie_001_fi.html. [Viitattu 11.5.2023]
36. van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Chapman & Hall / CRC Press. <https://stefvanbuuren.name/fimd/>. [Viitattu 11.5.2023]

37. Vehkalahti, K., Everitt, B.S. (2018). *Multivariate Analysis for the Behavioral Sciences* (2. painos, s. 319–340). Boca Raton, FL: CRC Press.
38. Vehkalahti, K. (2020). *Structural Equation Models*. Opetusmateriaali. Helsingin yliopisto. Valtiotieteellinen tiedekunta.
39. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3, s. 425–478). <https://doi.org/10.2307/30036540>. [Viitattu 11.5.2023]
40. Venkatesh, V., Thong, J. Y. L., Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1, s. 157–178). <https://doi.org/10.2307/41410412>. [Viitattu 11.5.2023]
41. Venkatesh, V., Thong, J. Y. L., Xu, X. (2016). Unified Theory of Acceptance and Use of Technology: A synthesis and the Road Ahead. *Journal of the Association for Information Systems*, 17(5). <https://doi.org/10.17705/1jais.00428>. [Viitattu 11.5.2023]
42. Westland, J.C. (2019). *Structural Equation Models* (2. painos). Sveitsi: Springer.

Liite A

Taulukko T1: Taustamuuttujien frekvenssit (N = 201)

	n	%
Kieli		
suomi	192	95.52
ruotsi	9	4.48
Sukupuoli		
mies	79	39.30
nainen	122	60.69
Ikä		
alle 60-vuotias	-	-
60–64-vuotias	16	7.96
65–69-vuotias	71	35.32
70–74-vuotias	91	45.27
75-vuotias ja yli	22	10.95
* puuttuva tieto	1	0.50
Ylin koulutusaste		
yleissivistävä	15	7.46
ammattikoulu	151	75.12
ammattikorkeakoulu	9	4.48
yliopisto tai korkeakoulu	21	10.45
muu	3	1.49
* puuttuva tieto	2	1.00
Maakunta		
Etelä-Karjala	40	19.90
Etelä-Savo	2	1.00
Keski-Suomi	61	30.35
Kymenlaakso	30	14.93
Pohjanmaa	9	4.48
Uusimaa	1	0.50
Varsinais-Suomi	56	27.86
* puuttuva tieto	2	1.00

Taulukko T2: Teknologiaan liittyvien taustamuuttujien frekvenssit (N = 201)

	n	%
Älypuhelimien merkki		
Doro	2	1.00
Honor	4	1.99
Huawei	48	23.88
iPhone	31	15.42
Nokia	18	8.96
One Plus	2	1.00
Samsung	88	43.78
Sony	4	1.99
Other	3	1.49
* puuttuva tieto	1	0.50
Sovellusten käyttökokemus		
alle 1 vuosi	8	3.98
1–2 vuotta	31	15.42
3–5 vuotta	62	30.85
6–10 vuotta	58	28.86
yli 10 vuotta	38	18.91
* puuttuva tieto	4	1.99

Taulukko T3: UTAUT2**-mallin havaittujen muuttujien tunnusluvut ja puuttuvuus.

Faktori		keskiarvo	keskihajonta	puuttuu (n)	puuttuu (%)
PE					
	PE1	3.52	1.09	13	6.47
	PE2	3.48	1.08	11	5.47
	PE3	3.32	1.07	21	10.45
	PE4	3.46	1.09	15	7.46
EE					
	EE1	4.21	0.85	2	1.00
	EE2	4.19	0.84	2	1.00
	EE3	4.34	0.75	1	0.50
	EE4	3.80	0.94	20	9.95
HM					
	HM1	3.70	0.97	6	2.99
	HM2	3.72	0.93	5	2.49
	HM3	3.33	1.11	11	5.47
	HM4	3.87	0.89	2	1.00
HT					
	HT1	3.82	1.12	2	1.00
	HT2	2.62	1.06	8	3.98
	HT3	3.04	1.20	11	5.47
	HT4	3.88	1.07	7	3.48
BI					
	BI1	3.94	0.97	18	8.96
	BI2	4.06	0.94	12	5.97
	BI3	3.99	0.99	19	9.45
	BI4	4.01	0.91	23	11.44
Huomautus:		N = 201			

Taulukko T4: Faktorilataukset | mallit 5a ja 5b

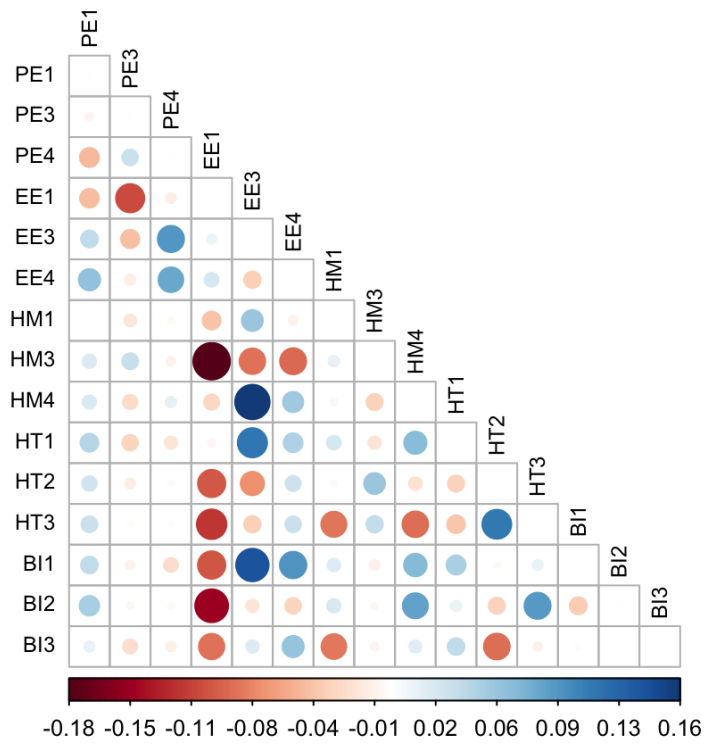
	PE	EE	HM	HT	BI		PE	EE	HM	HT	BI
Malli 5a (3 muuttujaa)				-		Malli 5b (4 muuttujaa)					
PE1	0.87	-	-	-	-	PE1	0.85	-	-	-	-
PE2	-	-	-	-	-	PE2	0.91	-	-	-	-
PE3	0.92	-	-	-	-	PE3	0.92	-	-	-	-
PE4	0.87	-	-	-	-	PE4	0.86	-	-	-	-
EE1	-	0.70	-	-	-	EE1	-	0.66	-	-	-
EE2	-	-	-	-	-	EE2	-	0.81	-	-	-
EE3	-	0.86	-	-	-	EE3	-	0.85	-	-	-
EE4	-	0.86	-	-	-	EE4	-	0.80	-	-	-
HM1	-	-	0.90	-	-	HM1	-	-	0.87	-	-
HM2	-	-	-	-	-	HM2	-	-	0.87	-	-
HM3	-	-	0.84	-	-	HM3	-	-	0.82	-	-
HM4	-	-	0.89	-	-	HM4	-	-	0.88	-	-
HT1	-	-	-	0.83	-	HT1	-	-	-	0.86	-
HT2	-	-	-	0.84	-	HT2	-	-	-	0.79	-
HT3	-	-	-	0.79	-	HT3	-	-	-	0.71	-
HT4	-	-	-	-	-	HT4	-	-	-	0.83	-
BI1	-	-	-	-	0.88	BI1	-	-	-	-	0.89
BI2	-	-	-	-	0.84	BI2	-	-	-	-	0.81
BI3	-	-	-	-	0.91	BI3	-	-	-	-	0.89
BI4	-	-	-	-	-	BI4	-	-	-	-	0.88

Taulukko T5: Korrelaatiomatriisi | malli 4a

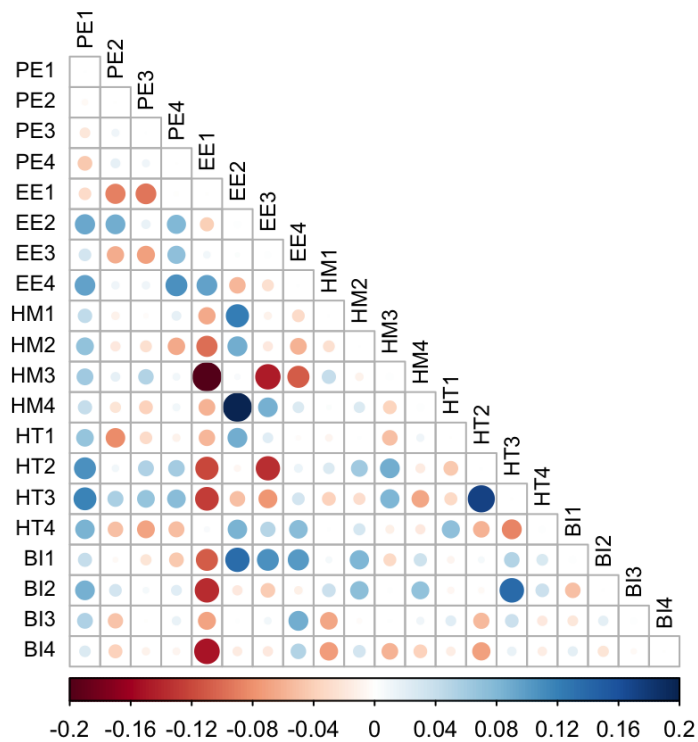
	PE1	PE3	PE4	EE1	EE3	EE4	HM1	HM3	HM4	HT1	HT2	HT3	BI1	BI2	BI3
PE1	1.00														
PE3	0.75	1.00													
PE4	0.68	0.74	1.00												
EE1	0.20	0.22	0.20	1.00											
EE3	0.23	0.25	0.23	0.49	1.00										
EE4	0.24	0.26	0.23	0.50	0.58	1.00									
HM1	0.59	0.65	0.59	0.24	0.28	0.28	1.00								
HM3	0.53	0.58	0.52	0.22	0.25	0.25	0.66	1.00							
HM4	0.56	0.61	0.55	0.23	0.26	0.27	0.70	0.62	1.00						
HT1	0.58	0.63	0.57	0.22	0.25	0.25	0.58	0.52	0.55	1.00					
HT2	0.56	0.61	0.55	0.21	0.24	0.25	0.56	0.50	0.53	0.56	1.00				
HT3	0.48	0.52	0.47	0.18	0.20	0.21	0.48	0.43	0.45	0.48	0.46	1.00			
BI1	0.56	0.61	0.56	0.24	0.28	0.28	0.52	0.46	0.49	0.50	0.49	0.41	1.00		
BI2	0.52	0.57	0.51	0.22	0.25	0.26	0.48	0.43	0.45	0.47	0.45	0.38	0.68	1.00	
BI3	0.59	0.64	0.58	0.25	0.29	0.29	0.54	0.48	0.51	0.53	0.51	0.43	0.76	0.71	1.00

Taulukko T6: Korrelaatiomatriisi | malli 4b

	PE1	PE3	PE3	PE4	EE1	EE2	EE3	EE4	HM1	HM2	HM3	HM4	HT1	HT2	HT3	HT4	BI1	BI2	BI3	BI4
PE1	1.00																			
PE2	0.73	1.00																		
PE3	0.76	0.82	1.00																	
PE4	0.68	0.74	0.77	1.00																
EE1	0.19	0.20	0.21	0.19	1.00															
EE2	0.22	0.24	0.25	0.23	0.45	1.00														
EE3	0.25	0.27	0.28	0.25	0.50	0.60	1.00													
EE4	0.21	0.23	0.24	0.22	0.43	0.52	0.57	1.00												
HM1	0.56	0.60	0.63	0.56	0.26	0.32	0.35	0.30	1.00											
HM2	0.54	0.59	0.61	0.55	0.26	0.31	0.34	0.30	0.70	1.00										
HM3	0.49	0.53	0.56	0.50	0.24	0.28	0.31	0.27	0.64	0.62	1.00									
HM4	0.55	0.59	0.61	0.55	0.26	0.31	0.34	0.30	0.70	0.69	0.624	1.00								
HT1	0.56	0.61	0.63	0.57	0.26	0.32	0.35	0.30	0.62	0.60	0.55	0.60	1.00							
HT2	0.48	0.52	0.54	0.49	0.23	0.27	0.30	0.26	0.53	0.52	0.47	0.52	0.57	1.00						
HT3	0.40	0.43	0.45	0.40	0.19	0.23	0.25	0.21	0.43	0.39	0.46	0.43	0.47	0.40	1.00					
HT4	0.56	0.61	0.63	0.57	0.27	0.32	0.35	0.31	0.62	0.61	0.55	0.61	0.66	0.57	0.47	1.00				
BI1	0.55	0.60	0.62	0.56	0.24	0.29	0.31	0.27	0.53	0.52	0.47	0.52	0.55	0.47	0.39	0.55	1.00			
BI2	0.49	0.53	0.55	0.50	0.21	0.25	0.28	0.24	0.47	0.46	0.42	0.46	0.49	0.42	0.34	0.49	0.69	1.00		
BI3	0.55	0.60	0.62	0.56	0.24	0.29	0.31	0.27	0.53	0.52	0.47	0.52	0.55	0.47	0.39	0.55	0.78	0.69	1.00	
BI4	0.55	0.60	0.62	0.56	0.24	0.29	0.31	0.27	0.53	0.52	0.47	0.52	0.55	0.47	0.39	0.55	0.78	0.69	0.77	1.00



(a) malli 4a



(b) malli 4b

Kuva K 1: Residuaalien korrelaatiokaavio rakenneyhtälömalleille 4a (ylempi kuva (a)) ja 4b (alempi kuva (b)). Tarkasteltava havaintoaineisto on DigitalWells-kyselyaineiston osa-aineisto. $N = 201$.