



**TURUN
YLIOPISTO**

PAIKALLAANOLON JA LIIKKUMISEN JAKSOTTUNEISUUS KAKSI- JA
KOLMITILAISEN TILASIRTYMÄMALLIN AVULLA

LuK Mikko Suominen

Pro gradu -tutkielma
Toukokuu 2023

Tarkastajat:
Prof. Kari Auranen
VTT Henri Nyberg

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

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

TURUN YLIOPISTO

Matematiikan ja tilastotieteen laitos

MIKKO SUOMINEN: Paikallaanolon ja liikkumisen jaksottuneisuus kaksi- ja kolmitilaisen tilasiirtymämallin avulla

Pro gradu -tutkielma, 48 s., 7 liites.

Tilastotiede

Toukokuu 2023

Paikallaanolon ja liikkumisen jaksottuneisuutta tutkimalla voidaan kuvata ja luonnehtia yksilön aktiivisuuskäyttäytymistä. Päivän aikana voidaan saavuttaa sama minuuttimäärä aktiivisuutta eri pituisina jaksoina. Tällöin ollaan kiinnostuneita esimerkiksi siitä, miten eri tavat kerryttää aktiivisuutta ovat yhteydessä terveyteen ja toimintakykyyn.

Tässä tutkielmassa tutkitaan eläkeiän kynnyksellä olevien suomalaisten kunta-alalla työskentelevien aktiivisuuden jaksottuneisuutta tilasiirtymämallien avulla. Aineisto on kerätty FIREA-tutkimuksesta, ja se koostuu liikemittaus- ja terveys/toimintakykytiedoista. Tilasiirtymämallit määritellään tilojen ja niiden välisten siirtymien avulla. Mallit perustuvat Markov-oletukseen, jonka mukaan seuraavan tilan todennäköisyys riippuu vain mallin nykyisestä, mutta ei aiemmista tiloista. Oletetaan myös, että siirtymätodennäköisyydet eivät riipu ajasta. Tässä tutkielmassa tarkastellaan jo aiemmin tutkittua kaksitilaista mallia sekä laajennetaan tätä mallia siirtymällä kolmitilaiseen malliin. Kolmitilainen malli eroaa siinä, että siinä on eritelty kaksitilaisen mallin aktiivinen tila kahdeksi aktiiviseksi tilaksi, kevyeksi ja reippaaksi aktiivisuudeksi.

Klusteroinnissa tavoite on löytää samanlaisia, mutta muista erottuvia aktiivisuuskäyttämistä kuvaavia ryhmiä. Ryhmien muodostaminen perustuu kunkin yksilön erikseen arvioituihin tilasiirtymätodennäköisyyksiin. Terveys- ja toimintakykyä kuvaavia indikaattoreita on pyritty mallintamaan lineaarisella mallilla erilaisten liikkumisen jaksottuneisuutta kuvaavien indeksien tai klustereiden avulla. Uutta kolmitilaista mallia verrataan kaksitilaiseen malliin ja pyritään tunnistamaan keskeiset erot.

Tämän tutkielman perusteella voidaan todeta, että klusterointi pystyi tunnistamaan selvästi erottuvia aktiivisuuskäyttäytymistä kuvaavia ryhmiä. Lisäksi aktiivisuuden jaksottuneisuus oli yhteydessä hapenottokykyyn ja kymmeneen tuoliltanousuun kuuluneeseen aikaan. Kolmitilainen malli kuvasi eri tavalla aktiivisuuden jaksottuneisuutta kuin kaksitilainen malli.

Asiasanat: FIREA-tutkimus, tilasiirtymämalli, aktiivisuuden jaksottuneisuus, klusterointi.

Sisälllys

1	Johdanto	1
2	Tutkimusaineisto- ja ongelma	3
2.1	Havaintoaineisto	4
2.2	Päätavoitteet ja tutkimuskysymykset	5
3	Menetelmät ja teoria	7
3.1	Monitilaiset Markov-mallit	7
3.1.1	Diskreettiaikaiset Markovin ketjut	7
3.1.2	Tilasiirtymätodennäköisyyksien estimointi	10
3.1.3	Mallin oletusten ja sopivuuden arviointi	12
3.2	Analyysimallit	13
3.3	Klusterointi	13
3.3.1	Hierarkkinen klusterointi	14
3.3.2	Klusteroinnin hyvyys	17
3.4	Menetelmät ja ohjelmistot	18
4	Aktiivisuuden jaksottuneisuus – empiirinen analyysi	19
4.1	Aineiston kuvaus	19
4.2	Aktiivisuuden jaksottuneisuus kaksitilaisessa mallissa	21
4.2.1	Kaksitilaisen mallin klusterit	22
4.2.2	Mallin oletusten ja hyvyyden arviointi	27
4.2.3	Aktiivisuuden jaksottaisuus ja terveysvasteet	28
4.3	Kolmitilainen malli	32
4.3.1	Kolmitilaisen mallin klusterit	34
4.3.2	Mallin oletusten ja hyvyyden arviointi	38
4.3.3	Aktiivisuuden jaksottaisuus ja terveysvasteet	39
4.3.4	Klusterointien yhtäpitävyys	43
5	Pohdinta	44
	Liitteet	48
A	Aktiivisuusluokkien suhteelliset osuudet osa-aineistoissa	48
B	Aktiivisuusluokkien suhteelliset osuudet klustereissa	50
C	ASTP-indeksin/pirstoutuneisuusindeksin ja maksimaalisen kävely- nopeuden välinen yhteys	54

1 Johdanto

Ikääntymisen myötä päivittäinen fyysinen aktiivisuus harvenee ja pirstoutuu. Aktiivisuusjaksoista tulee lyhyempiä ja istuma- ja lepojaksosta vastaavasti pidempiä. Fyysistä aktiivisuutta voidaan tutkia siirtymämallien avulla, joilla kuvataan kuinka taajaan yksilö lopettaa aktiivisen liikkumisen ja siirtyy istumisen kaltaiseen passiiviseen tilaan. Mallien tarkoitus on lisätä tietoa kohdeväestön toiminnallisesta tilasta [1]. Ikääntyvän työvoiman ja kansallisten terveystavoitteiden edistämiseksi on tarpeen ymmärtää yksityiskohtaisesti työntekijöiden fyysistä aktiivisuutta ja liikkumisen yhteyttä erilaisiin terveystavoitteisiin [2].

Kokonaisvaltaisen fyysisen aktiivisuuden tarkastelu voi olla erityisen hyödyllistä iäkkäiden kohdalla. Aikaisempien tutkimusten perusteella vähäinen fyysisen aktiivisuus on yhteydessä väsymykseen, hitaaseen kävelyyn, tapaturmiin, kognitiiviseen heikkenemiseen sekä lisääntyneeseen kuolleisuuteen [2]. Erilaiset interventiot ovat todennäköisesti tehokkaimpia, jos ne toteutetaan varhaisessa vaiheessa. Päivittäistä fyysistä aktiivisuutta voidaan seurata erilaisilla fyysisen aktiivisuuden mittareilla. Kiinnostuksen kohteena voi tällöin olla aktiivisuuden pirstoutuneisuus eli jaksottuneisuus. Pirstoutuneisuudella tarkoitetaan yksilön päivittäisen aktiivisuuden jakautumista lyhyisiin osiin. [2]

Fyysisen aktiivisuuden heikentyessä aktiivisuusjaksot lyhentyvät asteittain ikääntymisen myötä. Samalla kun fyysinen aktiivisuus vähenee, paikallaanolon aika lisääntyy. Joku voi kävellä päivittäin 20-30 minuuttia niin, että kokonaisuus kertyy yhdestä pitkän toiminnan jaksosta. Joku toinen saattaa kerryttää saman minuuttimäärän useasta lyhyestä aktiivisuusjaksosta. Ensimmäinen esimerkki voisi edustaa fyysistä tervettä henkilöä, jolla on suhteellisen hyvä suorituskyky ja kestävyys. Toinen esimerkki on henkilö, joka jaksaa ylläpitää fyysistä toimintaa vain muutaman minuutin kerrallaan. Molemmat henkilöt voivat tyypillisen päivän tai viikon aikana kerätä yhtä paljon aktiivisuutta ja paikallaanoloa, mutta aktiivisen ja paikallaanolevien tilojen välisten siirtymien lukumäärät eroavat paljon toisistaan.

Erilaisten toimintamallien erottaminen päivän mittaan ja päivien välillä on tärkeää, koska edelleen tiedetään hyvin vähän siitä, mitkä toimintatavat ovat terveydelle hyödyksi. Fyysisen aktiivisuuden on laajalti raportoitu parantavan terveyttä ja vähentävän kroonisten sairauksien riskiä. On myös osoitettu, että fyysinen passiivisuus vähentää työn tuottavuutta sekä lisää sairauspoissaolojen, työkyvyttömyyseläkkeelle joutumisen ja työttömyyden riskiä. Nämä havainnot perustuvat kuitenkin itse ilmoitettuihin tietoihin. Näin ollen on hyödyksi tarkastella päivittäisiä aktiivisuusmalleja objektiivisesti erilaisten mittausten avulla ja tutkia niiden yhteyttä erilaisiin terveystavoitteisiin. [2]

Paikallaanolon jaksottumista on tähän mennessä tutkittu ns. ASTP-mallin avulla (ASTP = *active to sedentary transition probability*) [1]. Tämän mallin avulla arvioidaan todennäköisyyttä (per aika-askel) siirtyä aktiivisesta liikkumisesta paikallaanoloon. Todennäköisyys on sitä suurempi, mitä enemmän aktiivisuus on pirstoutunut. Tällaisen mallin muodostamiseen tarvitaan tieto henkilön fyysisen aktiivisuuden tilasta kullakin hetkellä. Mallin avulla voidaan pyrkiä löytämään ja luonnehtimaan erilaisia päivittäisen aktiivisuuden profileja. Voidaan myös tutkia ASTP-indeksin yhteyttä erilaisiin terveyden- ja toimintakyvyn indikaattoreihin, nk. ter-

veysvasteisiin.

Tässä tutkielmassa käytettävä aineisto on peräisin Turun yliopiston FIREA-tutkimuksesta ja koostuu kiihtyvyyksimittarilla mitatuista aktiivisuusluokkien ketjuista [3]. Ketju koostuu minuutin välein kerätyistä aktiivisuuden tasoista. Mittausaineiston lisäksi tutkittavista henkilöistä on kerätty erilaisten taustamuuttujien arvoja. Tutkielman tavoitteena on mallintaa tutkimushenkilöiden aktiivisuuden jaksottuneisuutta sekä tunnistaa sellaisia aktiivisuusryhmiä, joita yhdistää tietty piirre. Aiemmassa tutkimuksessa aktiivisuus on jaoteltu joko paikallaanoloon tai liikkeeseen [1]. Tässä tutkielmassa jaksottuneisuutta tutkitaan erikseen kaksi- ja kolmitilaisen tilasiirtymämallin avulla ja pyritään löytämään tärkeimmät erot näiden välillä. Kolmitilainen malli syntyy, kun aktiivisen liikkumisen tila jaotellaan kevyeen ja reippaaseen liikkumiseen. Mielenkiintoista on myös selvittää, kuinka erilaiset taustamuuttujat näkyvät jaksottuneisuudessa ja löydettyissä ryhmissä. Myös vapaa- ja työpäivien välisiä eroja pyritään luonnehtimaan.

Samaa FIREA-aineistoa on käsitelty aiemmin neliluokkaisena Turun yliopistossa 2021 järjestetyllä mallinnusprojektin kurssilla [4]. Hienojakoisemman nelitilaisen mallin (paikallaan – kevyt liike – reipas liike – rasittava liike) avulla selvitettiin keinoja mallintaa ja kuvailla aktiivisuuden ja paikallaanolon jaksottumista. Kyseinen projektityö on ollut tämän työn lähtökohtana, mutta nyt on tarkoitus tutkia paikallaanolon ja aktiivisuuden jaksottuneisuutta viiden sekunnin jaksojen sijaan minuutin jaksoissa ja karkeistaa tilojen määrä kahteen tai kolmeen.

Tutkielman rakenne on seuraavanlainen. Luvussa 2 esitellään tutkimusongelma ja siihen liittyvä havaintoaineisto sekä tutkielman päätavoitteet. Luvussa 3 esitellään tilasiirtymämalleja ja Markovin ketjujen teoriaa. Samoin tarkastellaan siirtymätodennäköisyyksien estimointia sekä mallin oletusten ja hyvyyden arviointia. Luvussa 3 esitellään myös terveysvasteiden mallintamiseen tarvittavia analyysimalleja ja aktiivisuusprofiilien luonnehtimiseen käytettävän klusteroinnin perusteita. Luvussa 4 esitellään kaksi- ja kolmitilaisiin tilasiirtymämalleihin perustuvien analyysien tuloksia. Luvussa tarkastellaan myös tilasiirtymätodennäköisyyksien tai klustereiden yhteyttä erilaisiin terveyden- ja toimintakyvyn indikaattoreihin. Lisäksi selvitetään, pystytäänkö kaksi- ja kolmitilaisille malleille löytämään ilmeisiä klustereita, ja miten vastaavat klusterit vertautuvat toisiinsa. Viimeisessä luvussa esitetään tutkielman tulosten yhteenveto ja pohdintaa.

2 Tutkimusaineisto- ja ongelma

Tässä luvussa esitellään tutkimusaineisto ja -ongelma. Aineisto on kerätty FIREA-tutkimuksessa, jonka päätarkoituksena on selvittää elintavoissa, terveydessä ja toimintakyvyssä tapahtuvia muutoksia sekä niihin vaikuttavia tekijöitä eläkkeellesiirtymisiässä [2]. FIREA-tutkimus on Turun yliopistossa vuonna 2013 aloitettu ikään-tyneiden pitkittäistutkimus, ja siinä on tällä hetkellä mukana noin 6700 osallistujaa [3]. Yksi tutkimuksen tavoitteista on tunnistaa kiihtyvyyksmittarilla mitatun fyysisen aktiivisuuden profiileja ja tutkia, miten fyysisen aktiivisuuden jaksottuneisuus liittyy ikääntyvän väestön terveyteen.

FIREA-tutkimukseen osallistuneisiin otettiin ensimmäisen kerran yhteyttä 18 kuukautta ennen arvioitua eläköitymispäivää lähettämällä laaja terveyttä, elintapoja ja työtä koskeva kysely. Kyselyyn vastanneet, vielä työssä olleet Varsinais-Suomessa asuvat suomenkieliset julkisen sektorin työntekijät, joiden arvioitu eläkkeelle jääminen ajoittui vuosien 2017 ja 2019 väliin, kutsuttiin osallistumaan kliiniseen osatutkimukseen ($N=773$). Heistä 290 osallistui tähän tutkimukseen. Näistä kolmella tutkittavalla ilmeni mittarivika, jonka vuoksi heiltä ei saatu luotettavaa aineistoa tallennettua. Lisäksi seitsemältä tutkittavalta ei saatu yhtään kelvollista mittauspäivää. Tässä yhteydessä kelvollisella mittauspäivällä tarkoitetaan sellaista päivää, jonka aikana mittaria on hereilläolon aikana pidetty kädessä vähintään kymmenen tuntia.

Koska tutkimuksessa keskityttiin hereilläoloajan fyysiseen aktiivisuuteen ja paikallaanoloon, nukkumisaajat poistettiin aineistosta [2]. Analysoitava aineisto voi sisältää eri yksilöillä eri määrän työ- ja vapaapäiviä, ja myöskin yö- tai vuorotyöläisiä. Analysoitava aineisto sisälsi kuitenkin vain kelvolliset mittauspäivät.

Kaikilta osallistujilta pyydettiin tietoinen suostumus. Osallistujia kehoitettiin pitämään mittaria kädessä koko ajan, myös vedessä tapahtuvan toiminnan aikana. Saunomista varten kehoitettiin mittari kuitenkin ottamaan pois. Osallistujia pyydettiin merkitsemään, onko mittauspäivä työ- vai vapaapäivä, ja työpäivän alkamis- ja päättymisaika niiltä työpäiviltä, kun mittari oli kädessä. Mittausjakson jälkeen osallistujat palauttivat lokin ja mittauslaitteen postitse. Aineistonkeruu suoritettiin lokakuun 2015 ja toukokuun 2018 välisenä aikana, ja aineisto sisältää tietoja kaikilta neljältä vuodenajalta (28 % kevät, 9 % kesä, 27 % syksy ja 37 % talvi). [2]

Kullekin tutkimukseen osallistuneelle annettiin mittauspäiväkirja, johon merkittiin tutkimusviikon aikana päivämäärä, nukkumaanmeno-aika ja heräämisaika, työpäivät ja työn alkamis- ja päättymisaika. Terveyteen liittyviä fyysisiä kuntomittauksia olivat kehon koostumus, kardiorespiratorinen kunto ja lihaskunto. Mittausviikolla järjestettiin fyysisten kuntomittausten lisäksi klinisiä mittauksia.

Fyysistä aktiivisuutta mitattiin 24 tuntia vuorokaudessa kolmiakselisella ActiGraph (wActiSleep-BT) -kiihtyvyyksmittarilla, jota pidettiin tutkittavan henkilön ei-dominoinvan käden ranteessa. Mittaria pidettiin kädessä vähintään seitsemän päivää ja yötä, ja tutkimusjaksoon sisältyi sekä työ- että vapaapäiviä. Kiihtyvyyksmittari alustettiin tallentamaan 80 hertsin taajuudella eli mittari keräsi värähtelyä 80 kertaa sekunnissa. Värähtely tulkittiin kolmeen eri suuntaan mitattuina kiihtyvyyksinä. Tämän jälkeen 80 Hz taajuudella kerätty raakadata ladattiin ActiLife-ohjelman versiolla 6.13 (ActiGraph) ja tätä dataa käsiteltiin R-ohjelman GGIR-pakettia käyttäen

[5]. GGIR-paketti tuottaa raakadatan uuden datan, jossa mittauksen jokainen 5-sekunttinen on luokiteltu aktiivisuusluokkiin (SED, LPA, MPA, VPA) kiihtyvyyden kynnyksarvojen perusteella:

- $SED < 30 \text{ mg}$
- $30 \text{ mg} \leq LPA < 100.6 \text{ mg}$
- $100.6 \text{ mg} \leq MPA < 428.8 \text{ mg}$
- $VPA \geq 428.8 \text{ mg}$.

MPA- ja VPA-luokat yhdistettiin, koska VPA-luokka oli harvinainen ja liikuntatutkimuksessa yleisesti yhdistetään MPA ja VPA eli *moderate* ("reipas") ja *vigorous* ("rasittava") *physical activity*. Yhdistetystä luokasta käytetään lyhennettä MVPA eli *moderate-to-vigorous* ("reipas/rasittava") *physical activity*. Paikallaanolosta käytetään lyhennettä SED (*sedentary*) ja kevyestä fyysisestä aktiivisuudesta lyhennettä LPA (*light physical activity*). Kiihtyvyyden yksikkö edellä on mg ($1 \text{ mg} = 0.00981 \text{ m s}^{-2}$).

Tutkimusyksilöiden demografiset tekijät kuten ikä, sukupuoli, ammatti sekä työn altist tiedot saatiin rekisteritiedoista. Tieto tuki- ja liikuntaelimestön kroonisista sairauksista saatiin tutkittavien täyttämästä kyselystä. Kehon painoindeksi, vyötärön ympäryys, maksimaalinen hapenottokyky, maksimaalinen kävelynopeus sekä kymmeneen tuoliltanousuun kulunut aika määritettiin klinisen tutkimuskäynnin ja kuntotestien yhteydessä.

2.1 Havaintoaineisto

Analysoitavan aineiston lopullinen koko oli 280 (yksilöä). Tutkimusjoukon keski-ikä oli 63 vuotta. Jokaiselta osallistujalta kerättiin aktiivisuustietoa kiihtyvyyksinä alunperin 80 kertaa sekunnissa. Tämän jälkeen jaksotus muutettiin viiden sekunnin mittaiseksi. Viiden sekunnin tarkkuudella olevaan kiihtyvyyksineen tehtiin kolmiluokkainen luokittelu edellisessä luvussa esiteltujen kiihtyvyyksien kynnyksarvojen mukaan. Tässä viiden sekunnin jaksotuksessa aktiivisuusluokkia oli kolme, paikallaanolo (SED) = 1, kevyt aktiivisuus (LPA) = 2 ja reipas aktiivisuus (MVPA) = 3.

Tämän tutkielman aineisto on saatu minuutin jaksotukseksi muutettuna niin, että kunkin minuutin aktiivisuustaso määriteltiin sen mukaan, mikä minuutin aikana oli yleisin viiden sekunnin jaksotuksella mitattu aktiivisuustaso. Minuutin tasolle karkeistamalla pyrittiin vähentämään kohinaa, joka johtuu todennäköisesti satunnaisista käden liikkeistä. Nämä satunnaiset liikkeet eivät ole koko kehon fyysistä aktiivisuutta, jota tässä tutkimuksessa ensijaisesti haettiin. Mukaan otettiin vain täydet mittausminuutit. Kaksitilaisessa mallissa kumpikin aktiivinen tila yhdistettiin tilaksi 2. Mikäli paikallaanoloa ja kevyttä/reipasta aktiivisuutta oli yhtä paljon, luokiteltiin minuutti paikallaanoloksi. Mikäli kevyttä ja reipasta aktiivisuutta oli yhtä paljon ja paikallaanoloa vähemmän, minuutti luokiteltiin kevyeksi aktiivisuudeksi. Minuuteista alle 3 % oli sellaisia, joissa oli yhtä suuri osuus kahta/kolmea aktiivisuustyyppiä. Esikäsitteilyn jälkeen aineistossa oli jäljellä kaikki kelvollisilta päiviltä saadut mittaukset. Tieto vapaa- ja työpäivän aktiivisuusluokista on eritelty aineistossa.

Taulukossa 1 on esitelty havaintoaineiston tärkeimmät muuttujat ja niiden merkitys. Taulukkoon on lisätty havaintoaineistosta johdettu muuttuja *class_min2*, jota tarvitaan myöhemmässä vaiheessa. Johdettu muuttuja kuvailee, mikä aktiivisuustaso on kaksitilaisessa mallissa yleisin minuuttitasolla.

Muuttujan nimi	Muuttujan merkitys
<i>id1</i>	Henkilön tunnistenumero
<i>e_timestamp_min</i>	Päivämäärä ja kellonaika pyöristettynä lähimpään minuuttiin
<i>e_workday</i>	Päiväkirjan mukaan työ/vapaapäivä
<i>class_min</i>	Mikä aktiivisuusluokka yleisin minuuttitasolla
<i>class_min2</i>	Kaksitilainen aktiivisuustaso, <i>class_min</i> muuttujasta johdettu
<i>sed_o</i>	Paikallaanolon osuus minuutista
<i>lpa_o</i>	Keveyen fyysisen aktiivisuuden osuus minuutista
<i>mvp_a_o</i>	Reippaan/rasittavan fyysisen aktiivisuuden osuus minuutista
<i>dur_d_min</i>	Mittarin pitoaika päivän hereilläoloaikana

Taulukko 1: Muuttujat ja niiden merkitys

Mittausaineisto jaettiin neljään yhtä suureen osa-aineistoon, joissa kussakin oli 70 tutkimushenkilön mittaukset. Jako tehtiin lähinnä teknisistä syistä, jotta aineistoa havainnoivat kuvat saatiin piirrettyä tarkemmin. Ensimmäisessä osa-aineistossa on tunnistenumeron mukaisessa järjestyksessä 70 ensimmäistä tutkimushenkilöä. Aktiivisuusluokkien suhteelliset osuudet osa-aineistoissa on esitetty luvun 4 tulosten yhteydessä sekä liitteessä A.

2.2 Päätavoitteet ja tutkimuskysymykset

Havainto- ja taustamuuttuja-aineiston perusteella pyritään löytämään vastauksia kaksi- ja kolmitilaisesta mallia koskeviin tutkimuskysymyksiin. Kolmitilainen malli mahdollistaa tarkemman aktiivisuuden analyysin, koska aktiivinen tila on eritelty kahdeksi tilaksi. Onko mahdollista, että tämä lisätieto toisi jotain sellaista, mikä selittäisi aktiivisuuden pirstoutumista kaksitilaisesta mallista paremmin? Toisaalta yksinkertainen malli voisi kuvata pirstoutumista ja kertymistä yksinkertaisemmin ja ymmärrettävästi. Kolmitilainen malli ei lähtökohtaisesti sisällä yhtä helposti valittavaa pirstoutumisen mittaa kuin kaksitilainen malli. Kolmitilaisessa mallissa pirstoutumisen mitaksi voidaan kaksitilaisen mallin tapaan valita jokin alaspäin tapahtuvan siirtymän todennäköisyys, mutta vaihtoehtoja on tässä tapauksessa kolme yhden sijaan. Alaspäin tapahtuvalla siirtymällä tarkoitetaan tässä yhteydessä siirtymää aktiivisemmasta tilasta vähemmän aktiiviseen tilaan, eli joko reippaasta aktiivisuudesta siirtymistä kevyeen aktiivisuuteen, reippaasta aktiivisuudesta siirtymistä paikallaanoloon tai kevyestä aktiivisuudesta siirtymistä paikallaanoloon. Painottamalla sopivasti jotain tiettyä alaspäin tehtyä siirtymää voidaan myös löytää pirstoutumiselle jokin järkevä mitta. Alustavasti ajateltuna siirtymätodennäköisyys kevyestä aktiivisuudesta paikallaanoloon voisi olla mielekäs vaihtoehto. Toisaalta näiden kolmen alaspäin tapahtuvan siirtymätodennäköisyyden summa voisi toimia pirstoutumisen mittana. Klusteroinnissa yhden tilan lisääminen saattaisi mahdollistaa helpommin erottuvien ryhmien löytymisen.

Tässä tutkielmassa keskitytään erityisesti seuraaviin neljään päätavoitteeseen:

1. Tavoitteena on soveltaa tilasiirtymämatriiseja, käsittäen kaksi- ja kolmitilaiset mallit, aktiivisuuden jaksottumisen mallintamiseen.
2. Tavoitteena on tutkia, löydetäänkö kaksi- ja kolmitilaisilla malleilla eroja aktiivisuuden jaksottumisessa aktiivisuuden suhteen kahden hyvin erilaisen päivän, työ- ja vapaapäivän välillä.
3. Tavoitteena on soveltaa erilaisia klusterointitekniikoita erilaisten aktiivisuuden jaksottumisprofiilien löytämiseksi. Selvitetään löytääkö kaksi- ja/tai kolmitilainen malli ilmeisiä klustereita, ja miten näiden mallien klusterit vertautuvat toisiinsa. Löydetäänkö kolmitilaisella mallilla jotain sellaista mitä ei onnistuta löytämään kaksitilaisella mallilla?
4. Tavoitteena on tutkia, onko kaksitilaisen mallin ASTP-indeksi tai sen yleistykset kolmitilaisessa mallissa yhteydessä terveyden- ja toimintakyvyn indikaattoreihin kuten hapenottokykyyn ja maksimaaliseen kävelynopeuteen?

3 Menetelmät ja teoria

Tässä luvussa esitellään menetelmiä, joita myöhemmin tarvitaan aineiston analysointiin kaksi- tai kolmitilaisen mallin avulla. Kaksitilaisessa mallissa kuvataan siirtymiä aktiivisen ja paikallaanolevan tilan välillä. Siirtymiä aktiivisesta tilasta paikallaanoloon kuvataan ASTP-indeksillä (*active to sedentary transition probability*), joka määritellään aktiivisen tilan keskimääräisen keston käänteislukuna. Osallistujan SATP-indeksi (*sedentary to active transition probability*) on vastaavasti paikallaanolojakson keskimääräisen keston käänteisluku. Tässä tapauksessa (yhden) aktiivisuus-/paikallaanolojakson kestolla tarkoitetaan peräkkäisten minuutin pituisten aktiivisten/ei-aktiivisten ajanhetkien lukumäärää. Kolmitilainen malli toimii samalla tavalla kuin kaksitilainen malli, mutta siinä on yksi tila enemmän ja sen takia enemmän mahdollisia tilojen välisiä siirtymiä.

Luku 3.1 käsittelee monitilaisia Markov-malleja ja erityisesti aikahomogeenisia Markovin ketjuja, tilasiirtymätodennäköisyyksien estimointia sekä mallin oletusten ja sopivuuden arviointia. Luku 3.2 käsittelee luvussa 4 terveysvasteiden yhteydessä tarvittavia analyysimalleja. Terveysvasteita mallinnetaan lineaarisen regression avulla. Luvussa 3.3 esitellään menetelmiä, joilla havaintoja eli yksilöiden estimoituja tilasiirtymätodennäköisyyksien sisältämiä vektoreita ryhmitellään erilaisiin klustereihin hierarkkisen klusteroinnin avulla.

3.1 Monitilaiset Markov-mallit

Monitilamallit määritellään tilojen ja niiden välisten siirtymien avulla. Monitilaisia Markov-malleja voidaan käyttää arvioimaan yksilön hetkellistä riskiä (todennäköisyyttä) siirtyä tilasta toiseen sekä keskimääräistä oleskeluaikaa tietyssä tilassa. Markov-malleja käytetään useimmiten niiden yksinkertaisuuden vuoksi. Monitilamalli on Markov-malli, jos oletetaan, että uuteen tilaan siirtymisen todennäköisyys riippuu vain mallin nykyisestä tilasta. Tarkastellaan tässä tutkielmassa ainoastaan diskreettiaikaisia malleja, jolloin prosessia kutsutaan ketjuksi. Diskreetissä tapauksessa aikaväli on jaettu tasaisesti yhtä suuriin havaintoväleihin. Tämän luvun teoria perustuu pääosin lähteisiin [6] ja [7].

3.1.1 Diskreettiaikaiset Markovin ketjut

Diskreettiaikaista prosessia $\{X_k; k \in \mathbb{N}\}$, jossa $X_k \in S$ kaikilla k ja tila-avaruus $S = \{1, \dots, m\}$, kutsutaan Markovin ketjuksi, jos kaikille $i_0, i_1, \dots, i_{k-1}, i, j$, pätee

$$\mathbf{P}[X_{k+1} = j | X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0] = \mathbf{P}[X_{k+1} = j | X_k = i] = p_{ijk}.$$

Alaindeksi k viittaa ajanhetkeen. Todennäköisyyttä p_{ijk} kutsutaan tilasiirtymätodennäköisyydeksi. Se on ehdollinen todennäköisyys, että prosessin tila hetkellä $k+1$ on j ehdolla, että prosessi on tilassa i hetkellä k . Markov-ominaisuus tarkoittaa, että seuraavan tilan todennäköisyys riippuu vain ketjun nykyisestä mutta ei aiemmista tiloista. Jos todennäköisyydet p_{ijk} riippuvat ajanhetkestä k , ketjua kutsutaan epähomogeeniseksi. Muussa tapauksessa ketju on homogeeninen. Tässä tutkielmassa tarkastellaan vain homogeenisia Markov-ketjuja, joissa siis $p_{ijk} = p_{ij}$. Siirtymätodennäköisyydet eivät tällöin riipu ajasta.

Homogeeninen tilasiirtymätodennäköisyys p_{ij} toteuttaa seuraavat ehdot:

$$0 \leq p_{ij} \leq 1 \quad \text{ja} \quad \sum_{j \in S} p_{ij} = 1, \text{ jokaisella } i \in S.$$

Jälkimmäinen ehto tarkoittaa, että tilat ovat toisensa poissulkevat ja kollektiivisesti tyhjentävät eli aika-askeleella vähintään yhden siirtymän on tapahduttava. Siirtymä voi tällöin tarkoittaa myös nykyisessä tilassa pysymistä.

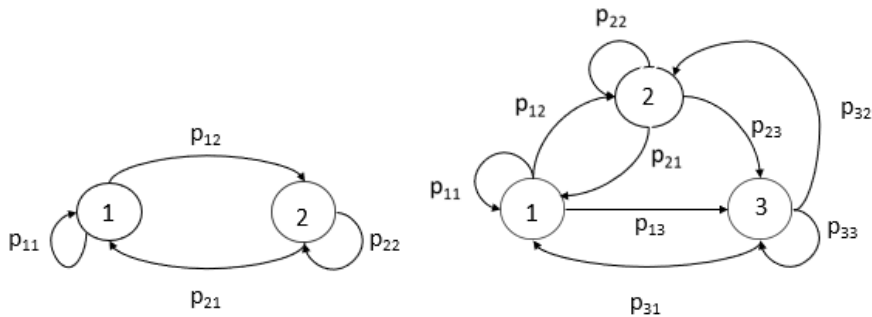
Siirtymätodennäköisyydet esitetään tyypillisesti $m \times m$ -neliömatriisina

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}.$$

Tilasiirtymämatriisi P on ns. stokastinen matriisi, sillä sen jokaiselle riville i pätee

$$\sum_{j \in S} p_{ij} = 1.$$

Tilasiirtymämalleja voidaan visualisoida kuvaajien avulla. Kuva 1 esittää kaksi- ja kolmitilaisen mallin mahdolliset siirtymät.



Kuva 1: Kaksi- ja kolmitilaisen mallin tilasiirtymät

Olkoon $p_{ij}^{(n)}$ ehdollinen todennäköisyys, että prosessi on tilassa j tasan n :n siirtymän jälkeen, ehdolla, että se lähtee tilasta i . Toisin sanoen

$$p_{ij}^{(n)} = \mathbf{P}[X_n = j | X_0 = i].$$

Tarkastellaan esimerkin vuoksi kahden askeleen päähän tehtävän siirtymän todennäköisyyttä $p_{ij}^{(2)}$, joka määritellään

$$p_{ij}^{(2)} = \mathbf{P}[X_2 = j | X_0 = i].$$

Tällöin

$$\begin{aligned}
p_{ij}^{(2)} &= \sum_{k \in S} \mathbf{P}[X_2 = j, X_1 = k | X_0 = i] \\
&= \sum_{k \in S} \mathbf{P}[X_2 = j | X_1 = k, X_0 = i] \mathbf{P}[X_1 = k | X_0 = i] \\
&= \sum_{k \in S} \mathbf{P}[X_2 = j | X_1 = k] \mathbf{P}[X_1 = k | X_0 = i] = \sum_{k \in S} p_{ik} p_{kj},
\end{aligned}$$

jossa kolmas yhtäpitävyys perustuu Markov-oletukseen. Viimeisessä lausekkeessa todennäköisyys aloittaa tilasta i ja olla tilassa j toisen siirtymän lopussa on todennäköisyys, että siirrymme ensin tilasta i välitilaan k ja sitten välitilasta k tilaan j . Summa ottaa huomioon kaikki mahdolliset välitilat k , jotka kuuluvat tila-avaruuteen S .

Ketjua kutsutaan pelkistymättömäksi (redusoimattomaksi), jos mille tahansa tilaparille (i, j) on olemassa sellainen $n > 0$, että $p_{ij}^{(n)} > 0$. Muuten ketju on pelkistytävä (redusoituva). Toisin sanoen pelkistymätön Markov-ketju voi päästä mistä tahansa tilasta mihin tahansa tilaan äärellisellä määrällä askelia positiivisella todennäköisyydellä.

Edellä määritellyssä mallissa voidaan osoittaa, että siirtymätodennäköisyydet $p_{ij}^{(n)}$ voidaan määrittää matriisin P^n alkiaina. Tällöin $p_{ij}^{(n)}$ on P^n -matriisin alkioiden (i, j) . Alkioiden $p_{ij}^{(n)}$ muodostamaa matriisia merkitään $P^{(n)}$:llä, jolloin siis

$$P^{(n)} = P^n$$

Määritellään $\mathbf{P}[X_0 = i]$ todennäköisyytenä, että prosessi on tilassa i ennen kuin se on tehnyt ensimmäistäkään askelta. Tällöin todennäköisyydet $\mathbf{P}[X_0 = i]$, $i = 1, \dots, m$, määrittävät ketjun alkutilan jakauman, kun $\sum_{i=1}^m \mathbf{P}[X_0 = i] = 1$.

Yhteistodennäköisyys $\mathbf{P}[X_{k+1}, X_k, \dots, X_0]$ voidaan laskea ketjusääntöä käyttämällä:

$$\begin{aligned}
&\mathbf{P}[X_{k+1} = j, X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0] \\
&= \mathbf{P}[X_{k+1} = j | X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0] \mathbf{P}[X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0] \\
&= \mathbf{P}[X_{k+1} = j | X_k = i] \mathbf{P}[X_k = i | X_{k-1} = i_{k-1}, \dots, X_0 = i_0] \mathbf{P}[X_{k-1} = i_{k-1}, \dots, X_0 = i_0] \\
&= \mathbf{P}[X_{k+1} = j | X_k = i] \mathbf{P}[X_k = i | X_{k-1} = i_{k-1}] \cdots \mathbf{P}[X_1 = i_1 | X_0 = i_0] \mathbf{P}[X_0 = i_0] \\
&= p_{ij} p_{i_{k-1}i} p_{i_{k-2}i_{k-1}} \cdots p_{i_0i_1} \mathbf{P}[X_0 = i_0].
\end{aligned} \tag{1}$$

Yhteistodennäköisyyden määräävät siis alkutilan X_0 todennäköisyys ja siirtymätodennäköisyydet.

Olkoon $\mathbf{P}[X(n) = j]$ todennäköisyys, että ketju on tilassa j ensimmäisen n siirtymän lopussa, jolloin m -tilaiselle prosessille pätee

$$\mathbf{P}[X(n) = j] = \sum_{i=1}^m \mathbf{P}[X_0 = i] p_{ij}^{(n)}.$$

Edellä määritellyssä mallissa voidaan osoittaa, että kun $n \rightarrow \infty$, n -askeleen siirtymätodennäköisyys $p_{ij}^{(n)}$ ei riipu i :stä [5]. Tämä tarkoittaa, että $\mathbf{P}[X(n) = j]$ lähestyy vakiota n :n lähestyessä ääretöntä.

Kaksitilaisessa mallissa tilassa 2 vietetyn ajan odotusarvo on $T_2 = 1/p_{21}$. Tulos seuraa siitä, että tilassa vietetty aika on geometrisen jakauman mukainen. Tämä tarkoittaa samalla sitä, että ASTP-indeksi, joka oli määritelty aktiivisessa tilassa vietetyn ajan käänteislukuna, on itse asiassa siirtymätodennäköisyys p_{21} . Vastaavasti SATP-indeksi on p_{12} . Yleisessä tapauksessa tilassa i vietetyn ajan odotusarvon lauseke on $T_i = 1/\sum p_{ij}$, missä indeksi j käy läpi kaikki muut tilat paitsi tilan i .

3.1.2 Tilasiirtymätodennäköisyyksien estimointi

Tilasiirtymämatriisi P on tuntematon ja se estimoidaan suoraan aineistosta. Parametrien lukumäärä on m^2 , jossa m on tilojen lukumäärä. Sidosehtojen vuoksi vapaita estimoitavia parametreja on $m(m-1)$. Estimointi tehdään jokaiselle aineiston yksilölle erikseen eli tilasiirtymämatriisin suurimman uskottavuuden estimaatit ovat yksilökohtaiset. Estimaattien voidaan ajatella olevan tilastollisesti hyvin tarkkoja, koska yksilöiden aktiivisuusjaksoista koostuvat ketjut ovat todella pitkiä. Tilastollinen tarkkuus ei kuitenkaan takaa sitä, että malli sopisi aineistoon hyvin. Tähän asiaan palataan myöhemmin.

Oletetaan, että yksilölle on havaittu ketju $x_0^n = (x_0, x_1, \dots, x_n)$. Merkintöjen yksinkertaistamisen vuoksi tässä vaiheessa ei merkitä yksilön indeksiä l näkyviin. Havaittu ketju on satunnaismuuttujan X_0^n realisaatio. Markov-oletuksen vallitessa realisaation todennäköisyys on kaavan (1) mukaisesti

$$\begin{aligned} \mathbf{P}(X_0^n = x_0^n) &= \mathbf{P}(X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= \mathbf{P}(X_0 = x_0) \prod_{k=1}^n \mathbf{P}(X_k = x_k | X_{k-1} = x_{k-1}) \\ &= \mathbf{P}(X_0 = x_0) \prod_{k=1}^n p_{x_{k-1}x_k}. \end{aligned}$$

Jättämällä alkutilan malli pois, saadaan siirtymätodennäköisyyksien $p = (p_{11}, p_{12}, \dots, p_{mm})$ uskottavuusfunktio:

$$L(p) = \prod_{k=1}^n p_{x_{k-1}x_k}.$$

Oletetaan, että yksilön havaitussa ketjussa tilasiirtymien $i \rightarrow j$ lukumäärät ovat n_{ij} , $i, j \in S$. Uskottavuusfunktio voidaan tällöin kirjoittaa muotoon

$$L(p) = \prod_{i=1}^m \prod_{j=1}^m p_{ij}^{n_{ij}}.$$

Havainto $n = (n_{11}, n_{12}, \dots, n_{mm})$ on siis tyhjentävä tunnusluku.

Suurimman uskottavuuden estimaatit saadaan maksimoimalla uskottavuusfunktio kaikkien tilasiirtymätodennäköisyyksien suhteen. Logaritminen uskottavuusfunktio on

$$\ell(p) = \log L(p) = \sum_{i,j} n_{ij} \log p_{ij}.$$

Ennen uskottavuusfunktion maksimointia on syytä huomata ehto, että jokaisesta tilasta siirtymisten todennäköisyyksien on summauduttava ykköseksi. Tämä tarkoittaa sitä, että tilasiirtymämatriisin vapaiden parametrin määrä ei ole m^2 vaan $m(m-1)$. Suoritetaan seuraavaksi parametrien estimointi eksplisiittisesti parametrien elimointia käyttäen. Toinen mahdollinen tapa olisi käyttää Lagrangen kertoimia.

Valitaan kullakin lähtötilalla i yksi siirtymätodennäköisyyksistä ja ilmaistaan se muiden suhteen. Olkoon tämä siirtymä $p_{i1} = 1 - \sum_{j=2}^m p_{ij}$. Seuraavaksi laskemme uskottavuuden osittaisderivaatat. Tällöin

$$\frac{\partial \ell(p)}{\partial p_{ij}} = \frac{n_{ij}}{p_{ij}} - \frac{n_{i1}}{p_{i1}}, \quad j = 1, \dots, m; \quad j \neq i.$$

Näiden asettaminen nolllaksi johtaa lausekkeisiin

$$\frac{n_{ij}}{n_{i1}} = \frac{\hat{p}_{ij}}{\hat{p}_{i1}}, \quad j = 1, \dots, m; \quad j \neq i.$$

Koska tämä pätee kaikille j , voidaan päätellä, että $\hat{p}_{ij} \propto n_{ij}$, ja näin ollen

$$\hat{p}_{ij} = \frac{n_{ij}}{\sum_{j=1}^m n_{ij}}, \quad i = 1, \dots, m \quad \text{ja} \quad j = 1, \dots, m. \quad (2)$$

On selvää, että siirtymätodennäköisyyden p_{i1} valinta oli täysin mielivaltainen. Samat estimaatit saadaan siis millä tahansa muulla valinnalla. Estimaatit voidaan tulkita tietyn siirtymän suhteellisina osuuksina kaikista ko. lähtötilasta mihin tahansa tilaan tehdyistä siirtymistä.

Tämän tutkielman aineistossa on havaittu kunkin ketju yhteensä L yksilöltä. Merkitään yksilöitä indeksillä l , jossa $l = 1, \dots, L$, ja olkoon yksilön l havaitun ketjun pituus n_l . Jokainen yksilö tuottaa havainnon eli tässä tapauksessa ketjun $x_0^{n_l} = (x_0^{(l)}, x_1^{(l)}, \dots, x_{n_l}^{(l)})$, ja tästä havaitusta ketjusta määräytyvän tilasiirtymien lukumääristä koostuvan vektorin $n^{(l)}$, joka on kahden tilan tapauksessa muotoa $n^{(l)} = (n_{11}^{(l)}, n_{12}^{(l)}, n_{21}^{(l)}, n_{22}^{(l)})$. Kolmen tilan tapauksessa vektori $n^{(l)}$ sisältää yhdeksän komponenttia.

Tutkielman aineiston analyysissä mahdollisten tilojen lukumäärä m on kaksi tai kolme. Kaksitilaiselle mallille lasketut tilasiirtymätodennäköisyyksien yksilökohtaiset tilasiirtymien estimaatit voidaan esittää yleisen lausekkeen (2) tapaan muodossa

$$\hat{p}_{ij}^{(l)} = \frac{n_{ij}^{(l)}}{\sum_{m=1}^2 n_{im}^{(l)}} = \frac{n_{ij}^{(l)}}{n_{i1}^{(l)} + n_{i2}^{(l)}}, \quad i = 1, 2, \quad j = 1, 2 \quad \text{ja} \quad l = 1, \dots, L. \quad (3)$$

Merkitään T :llä keskimääräistä tilassa vietettyä aikaa. Kaksitilaisen mallin yksilökohtainen ASTP-indeksi voidaan laskea edellisiä kaavoja hyväksi käyttäen muodossa

$$\text{ASTP}^{(l)} = \frac{1}{T_2^{(l)}} = \hat{p}_{21}^{(l)} = \frac{n_{21}^{(l)}}{\sum_{k=1}^2 n_{2k}^{(l)}} = \frac{n_{21}^{(l)}}{n_{21}^{(l)} + n_{22}^{(l)}}, \quad l = 1, \dots, L, \quad (4)$$

ja vastaavasti yksilökohtainen SATP-indeksi:

$$\text{SATP}^{(l)} = \frac{1}{T_1^{(l)}} = \hat{p}_{12}^{(l)} = \frac{n_{12}^{(l)}}{\sum_{k=1}^2 n_{1k}^{(l)}} = \frac{n_{12}^{(l)}}{n_{11}^{(l)} + n_{12}^{(l)}}, \quad l = 1, \dots, L. \quad (5)$$

Kolmitilaiselle mallille voidaan yhtä lailla laskea yksilökohtaisia pirstoutumista kuvaavia indeksejä. Tarkastellan kolmesta alaspäin tapahtuvasta siirtymästä koostuvaa tilasiirtymätodennäköisyyksien estimaattien summaa:

$$\hat{p}_{31}^{(l)} + \hat{p}_{32}^{(l)} + \hat{p}_{21}^{(l)} = \frac{1}{T_3^{(l)}} + \hat{p}_{21}^{(l)}, \quad l = 1, \dots, L. \quad (6)$$

Tätä indeksiä hyödynnetään luvussa 4 pirstoutumisen mittana vastaavalla tavalla kuin ASTP-indeksiä. Käytetään tästä jatkossa nimitystä *pirstoutuneisuusindeksi*.

3.1.3 Mallin oletusten ja sopivuuden arviointi

Tässä luvussa esitellään keinoja Markov-mallin oletusten ja mallin sopivuuden tarkistamiseen. Estimoitujen siirtymätodennäköisyyksien $\hat{p}_{ij}^{(l)}$ perusteella voidaan simuloida aikahomogeenisia Markovin ketjuja ja muodostaa näin erilaisia ennustejakaumia. Sopivasti valittujen testisuureiden ennustejakaumia voidaan tämän jälkeen verrata vastaavien suureiden havaittuihin arvoihin. Havaintojen poikkeama ennustejakaumasta voisi johtua siitä, että ketju on oikeasti epähomogeeninen ajan suhteen, ja/tai Markov-oletuksen paikkansapitämättömyydestä.

Kaksitilaisessa mallissa kunkin yksilön havaitusta ketjusta voidaan laskea esimerkiksi peräkkäisistä kakkosista koostuvien osaketjujen lukumäärä tai $(2 \rightarrow 2 \rightarrow 1)$ -osaketjujen lukumäärä. Vastaavat tarkastelut yleistyvät yleiseen m -tilaiseen ketjuun. Osaketjun $(2 \rightarrow 2 \rightarrow 1)$ siirtymä tilasta kaksi tilaan yksi ei mallin oletusten mukaan riipu siirtymää edeltävästä tilasta, mutta todellisuudessa näin saattaisi olla.

Ennustejakauma voidaan muodostaa seuraavalla algoritmilla:

1. Laske kullekin aineiston yksilölle $n^{(l)}$ -pituisen ketjun perusteella siirtymätodennäköisyyksien estimaatit $\hat{p}_{ij}^{(l)}$.
2. Simuloi kullekin yksilölle yksi $n^{(l)}$ -pituisen Markovin ketju havaitun ketjun siirtymätodennäköisyyksien estimaattien perusteella.
3. Vertaa testisuureen L :n yksilön havaittujen arvojen jakaumaa L :n simuloidun ketjun antamaan testisuureen empiiriseen jakaumaan.

3.2 Analyysimallit

Tässä tutkielmassa tutkitaan sekä ASTP-indeksin että sen kolmitilaiseen malliin tehdyn yleistyksen yhteyttä erilaisiin terveystasteisiin lineaarisen regression avulla. Samoin tutkitaan klustereiden yhteyttä terveystasteisiin.

Lineaarinen regressiomalli kuvaa selitettävän muuttujan y ja yhden tai useamman selittävän muuttujan Z välistä riippuvuutta. Selitettävää muuttujaa kutsutaan myös vastemuuttujaksi. Jatkuvia selittäviä muuttujia kutsutaan kovariaateiksi ja kategorisia selittäviä muuttujia faktoreiksi. Mallin jäännökset eli virhetermit ϵ_l ovat satunnaisia ja ei-havaittuja. Virhetermien oletetaan noudattavan normaalijakaumaa odotusarvolla nolla ja varianssilla σ^2 . Jäännösten ϵ_l varianssi σ^2 kuvaa yksittäisten havaintojen vaihtelua regressiosuoran määräämän odotusarvon ympärillä. Usean selittäjän lineaarisen regressiomalli voidaan kirjoittaa muodossa

$$y_l = \beta_0 + \beta_1 Z_{l1} + \beta_2 Z_{l2} + \dots + \beta_p Z_{lp} + \epsilon_l, \quad l = 1, 2, \dots, L.$$

Regressiokertoimien β estimaatit ratkaistaan suurimman uskottavuuden (SU) menetelmällä. Kun virhetermi on normaalisti jakautunut, parametrien SU-estimaatit vastaavat pienimmän neliösumman (PNS) estimaatteja.

Regressiokertoimille raportoidaan suurimman uskottavuuden estimaatit ja 95 %:n luottamusvälit. Esitetään tulosten yhteydessä myös mallin selitysaste R^2 (*R-squared*) ja korjattu selitysaste R_{adj}^2 (*adjusted R-squared*). Selitysaste kuvaa havaintojen (regressiosuoran avulla) selitetyn vaihtelun osuutta kokonaisvaihtelusta, ja se rajoittuu välille $[0,1]$ eli voidaan ilmaista prosentteina. Selitysaste on mallin hyvyyden mitta siitä, kuinka lähellä aineisto on sovitettua regressiosuoraa. Korjattu selitysaste mukautuu mallin selittäjien lukumäärän ja aineiston koon mukaan. Sen arvo kasvaa vain, kun uusi selittäjä parantaa mallin sopivuutta enemmän kuin pelkästään sattumalta odotettiin. Korjattu selitysaste itse asiassa pienenee, kun selittäjä ei paranna mallin sopivuutta riittävästi. [8]

Tässä tutkielmassa vastemuuttujina käytetään yksilön toimintakykyä mittaavia *hapenottokykyä, maksimaalista kävelynopeutta ja kymmeneen tuolittanousuun kulunutta aikaa*. Selittäviä muuttujia ovat vaihtoehtoisesti ASTP-indeksi, pirstoutuneisuusindeksi tai klusteri (ks. luku 3.3). ASTP-indeksiä käsitellään vaihtoehtoisesti sekä jatkuvana- että kategorisena selittäjänä. Kategorisointi tapahtuu niin, että ASTP-indeksi jaetaan kolmeen tertiiliin. Mallit vakioidaan demografisten tietojen (ikä, sukupuoli ja työn kuormittavuus), painoindeksin ja kroonisten sairauksien historian mukaan.

3.3 Klusterointi

Klusterointi pyrkii jakamaan aineiston yksilöt keskenään samankaltaisiin, mutta muista erottuviin ryhmiin. Samankaltaisten havaintojen ryhmää kutsutaan klusteriksi. Klusterin muodostamisen määrää valittu kriteeri, jonka perusteella mahdollisimman lähellä toisiaan olevat havainnot yhdistetään. Ryhmien muodostamisen jälkeen tavoitteena on tunnistaa ne erityispiirteet, jotka kutakin klusteria luonnehtivat.

Sopivan menetelmän löytäminen on klusteroinnin vaikeimpia vaiheita, koska eri vaihtoehdot tuottavat erilaisia ryhmäjakoja. Klusterointi kuuluu ns. ohjaamattoman

oppimisen menetelmiin. Ohjaamattomalla oppimisella pyritään kuvaamaan aineiston rakennetta oppimalla aineistosta oleellisia piirteitä. Vastemuuttujan arvoja eli todellisia ryhmäjakoja ei useinkaan tiedetä. Tässä tutkielmassa rajoittaudutaan ainoastaan hierarkkiseen klusterointiin. Päälähteenä käytetään tällöin Trevor Hastien ja muiden kirjaa *The Elements of Statistical Learning* ja sen lukua 14 [9].

3.3.1 Hierarkkinen klusterointi

Hierarkkinen klusterointi on yksi koneoppimisen suosituimmista klusterointitekniikoista. Menetelmä voidaan jakaa kahteen vaihtoehtoon sen perusteella, suoritetaan-ko ryhmien jakaminen ns. alhaalta ylös vai ylhäältä alas. Vaihtoehdoista käytetään nimiä kokoava (*agglomerative*) ja jakava (*divisive*). Kirjallisuudessa jakavia menetelmiä ei ole tutkittu läheskään yhtä laajasti kuin kokoavia [9]. Tästä syystä tutkielman pääpaino on kokoavissa menetelmissä. Molemmilla menetelmillä on $N - 1$ hierarkian tasoa, missä N on havaintojen lukumäärä. Molemmissa menetelmissä hierarkian alimmalla tasolla jokainen klusteri sisältää yksittäisen havainnon. Ylimmällä hierarkian tasolla yksi klusteri sisältää aineiston kaikki havainnot. Hierarkkisen klusteroinnin hyvänä puolena voidaan pitää sitä, että klustereiden määrää ei tarvitse etukäteen päättää. Huonona puolena taas on, että menetelmä on herkkä kohinalle ja poikkeaville havainnoille.

Oletetaan, että yksittäisten havaintojen välinen etäisyys lasketaan euklidisena etäisyytenä. Tässä tutkielmassa yksi havainto on kunkin yksilön siirtymätodennäköisyyksien estimaattien sisältämä vektori, jolloin euklidinen etäisyys yksilön l havainnon ja yksilön h havainnon välillä lasketaan seuraavasti:

$$d(l, h) = \sqrt{\sum_{i,j} w_{ij} (\hat{p}_{ij}^{(l)} - \hat{p}_{ij}^{(h)})^2}. \quad (7)$$

Lausekkeessa w_{ij} on tilasiirtymätodennäköisyyden estimaattiin \hat{p}_{ij} liittyvä painokerroin. Lähtökohtaisesti painokerroin on kaikille alkiolle ykkösen suuruinen. Tämä antaa mahdollisuuden painottaa esimerkiksi pirstoutumisindeksin (6) jokaista komponenttia sopivasti valitulla ykköistä suuremmalla painolla. Tässä tutkielmassa klusteroinnissa käytettävät painot ovat kuitenkin kaikki ykkösiä.

Kokoavassa menetelmässä jokaista havaintoa pidetään aluksi omana klusterinaan. Menetelmä tuottaa seuraavaksi korkeammalle tasolle ryhmittelyn, jossa on ylempään tasoon nähden yksi klusteri vähemmän. Yhdistettäväksi pariiksi valitaan kaksi klusteria, joilla on pienin klusterien välinen etäisyys. Jokaisen iteraation jälkeen samankaltaiset klusterit siis yhdistyvät toisiinsa, kunnes muodostuu haluttu lukumäärä k klustereita. Avainasemassa ovat kahden klusterin eroavaisuuden määrittely ja klusterien määrän valinta.

Tarkastellaan eri määritelmiä laskea klustereiden välinen etäisyys. Olkoot G ja H kaksi klusteria. Klustereiden G ja H välinen etäisyys $d(G, H)$ lasketaan niiden parittaisten havaintojen etäisyyksien $d(i, i')$ perusteella, kun havainto i kuuluu klusteriin G ja havainto i' kuuluu klusteriin H . Lähimmän naapurin kokoavassa klusterointimenetelmässä (*single linkage agglomerative clustering*, SL) klusterien välinen

eroavaisuus on lähimpänä toisiaan olevan parin välinen etäisyys:

$$d_{SL}(G, H) = \min(d(i, i')), \quad \text{jossa } i \in G \text{ ja } i' \in H. \quad (8)$$

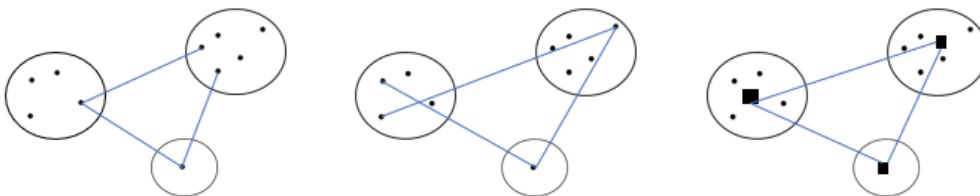
Kaukaisimman naapurin kokoavassa klusteroinnissa (*complete linkage agglomerative clustering*, CL) klusterien välinen eroavaisuus on kauimpana toisistaan olevan parin välinen etäisyys:

$$d_{CL}(G, H) = \max(d(i, i')), \quad \text{jossa } i \in G \text{ ja } i' \in H. \quad (9)$$

Naapurikeskiarvon menetelmä (*group average*, GA) käyttää klusterien välisen eroavaisuuden mittana keskiarvoa:

$$d_{GA}(G, H) = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{j \in H} d(i, j), \quad (10)$$

jossa N_G ja N_H ovat klustereiden G ja H havaintojen lukumäärät. Kuva 2 esittää kolme edellä esitettyä tapaa määrittää klusterien eroavaisuus.



Kuva 2: Yleisimmät klusterien väliset eroavaisuuden mitat. Vasemmalta oikealle: lähin naapuri, kaukaisin naapuri ja naapurikeskiarvo. Kuva piirretty mukailien lähettä [10].

On muitakin tapoja ryhmien erilaisuuden määrittämiseksi, mutta edellä esitetyt kolme tapaa ovat yleisimmin käytettyjä. Jos jokainen klusteri on kompakti ja hyvin erillään muista, kaikki kolme menetelmää tuottavat samanlaisia tuloksia. Klusterit ovat kompakteja, jos kaikki niissä olevat havainnot ovat suhteellisen lähellä toisiaan verrattuna havaintoihin eri klustereissa. Lähimmän naapurin menetelmä edellyttää vain, että yksi parittainen eroavaisuus $d(i, i')$ on pieni kahden ryhmän G ja H katsomiseksi lähelle toisiaan, riippumatta muista ryhmien välisistä havaintojen eroista. Menetelmä on herkkä ”ketjutukseksi” kutsutulle vaikutukselle, jolloin huonosti erotetut mutta erilliset klusterit yhdistetään varhaisessa vaiheessa. Ketjuttumista pidetään menetelmän puutteena. Määritellään klusterin halkaisija D_G suurimmaksi eroiksi klusterin sisällä:

$$D_G = \max(d(i, i')), \quad \text{jossa } i \in G \text{ ja } i' \in G.$$

Lähimmän naapurin menetelmä voi tuottaa klusterita, joilla on erittäin suuri halkaisija.

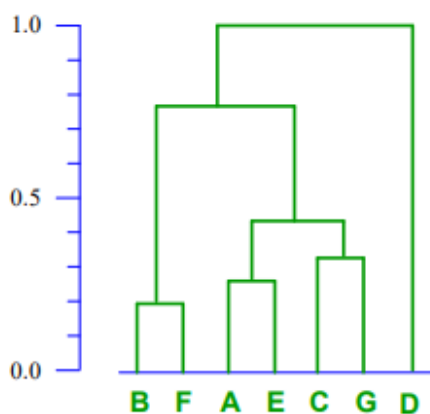
Kaukaisimman naapurin menetelmä edustaa päinvastaista ääripäätä. Kaksi ryhmää H ja G katsotaan läheisiksi, jos kaikki parittaiset havaintojen väliset eroavaisuudet ovat suhteellisen samanlaisia. Menetelmällä on taipumus tuottaa pienen hal-

kaisijan kompakteja klustereita. Mahdollista on myös, että klusteriin osoitetut havainnot voivat olla paljon lähempänä muiden klustereiden jäseniä kuin joitain oman klusterin jäseniä. Tällöin ei voida puhua kompakteista klustereista.

Naapurikeskiarvon menetelmä edustaa kompromissia kahdesta ääripäästä, lähimmän ja kaukaisimman naapurin menetelmästä. Se yrittää tuottaa suhteellisen kompakteja klustereita, jotka ovat suhteellisen kaukana toisistaan. Tulokset kuitenkin riippuvat numeerisesta asteikosta, jolla havaintojen etäisyydet $d(i, i')$ mitataan. Soveltamalla keskiarvomittaan (10) monotonisesti kasvavaa muunnosta $h(\cdot)$, voivat tulokset muuttua. Etäisyysmitat (8)–(9) puolestaan riippuvat vain parittaisten etäisyyksien järjestyksestä ja ovat siten invariantteja tällaisille monotonisille muunnoksille. Tätä ominaisuutta käytetään usein argumenttina lähimmän- tai kaukaisimman naapurin menetelmien puolesta.

Dendrogrammi tarjoaa helposti tulkittavissa olevan graafisen esityksen hierarkkisesta klusteroinnista. Dendrogrammia pidetään usein yhteenvetona aineistosta eikä algoritmin tulosten kuvauksena. Tulosten tulkintaan on kuitenkin syytä suhtautua varoen, koska erilaiset hierarkkiset menetelmät sekä pienet muutokset aineistossa voivat johtaa melko erilaisiin dendrogrammeihin.

Kuva 3 esittää esimerkkiä seitsemästä havainnosta (**A-G**) koostuvasta dendrogrammista. Kuvasta nähdään, että havainnot **B** ja **F** ovat valitun eroavaisuudeltaan mielessä lähimpänä toisiaan, jolloin ne muodostavat aluksi oman klusterin. Pysty-akselilta luetaan havaintojen/klustereiden väliset eroavaisuudet. Havaintojen **B** ja **F** välinen eroavaisuus on 0.2. Seuraavassa vaiheessa havainnot **A** ja **E** yhdistyvät ja näiden välinen eroavaisuus on 0.25. Neljännessä vaiheessa klusterit (**A** ja **E**) ja (**C** ja **G**) yhdistyvät ja klustereiden välinen eroavaisuus on hieman yli 0.4. Aivan viimeisenä havainto **D** yhdistyy kuuden kokoiseen klusteriin. Haluttu määrä k klusteria voidaan valita kuvan perusteella niin, että piirretään vaakasuora viiva sopivaan kohtaan. Esimerkiksi, jos halutaan kolme klusteria, voidaan piirtää vaakasuora viiva kohtaan 0.5. Tällöin saadaan kooltaan 2, 4 ja 1 olevat klusterit.



Kuva 3: Havainnoista (**A-G**) koostuva dendrogrammi. Pystyakseli kuvaa havaintojen/klustereiden välisiä eroavaisuuksia. Kuva lainattu lähteestä [11].

Luvussa 4 muodostetaan jokaisella yksilöllä omat 2×2 ja 3×3 -tilasiirtymämatriisit ja asetetaan kaikki mahdolliset estimoidut siirtymätodennäköisyydet peräkkäin riveiksi, yksilöiden tunnistenumeron mukaiseen järjestykseen. Kukin rivi siis koostuu yhden yksilön tilasiirtymien vektorista. Tällä tavalla muodostetut 280×4 ja 280×9 -matriisit käsittävät kaikkien yksilöiden estimoitujen siirtymätodennäköisyyksien arvot. Havainnollistetaan N yksilön tilasiirtymien sisältävää matriisia kaksitilaisen mallin tapauksessa:

$$\begin{array}{c}
 id1 \\
 id2 \\
 \vdots \\
 idL
 \end{array}
 \begin{pmatrix}
 1 \rightarrow 1 & 1 \rightarrow 2 & 2 \rightarrow 1 & 2 \rightarrow 2 \\
 \hat{p}_{11}^{(1)} & \hat{p}_{12}^{(1)} & \hat{p}_{21}^{(1)} & \hat{p}_{22}^{(1)} \\
 \hat{p}_{11}^{(2)} & \hat{p}_{12}^{(2)} & \hat{p}_{21}^{(2)} & \hat{p}_{22}^{(2)} \\
 \vdots & \vdots & \ddots & \vdots \\
 \hat{p}_{11}^{(L)} & \hat{p}_{12}^{(L)} & \hat{p}_{21}^{(L)} & \hat{p}_{22}^{(L)}
 \end{pmatrix}
 \quad (11)$$

Kolmitilaiselle mallille vastaava matriisi on samanlainen, mutta jokainen rivi sisältää nyt yhdeksän komponenttia aiemmin esitetyn neljän sijaan. Matriisin rivejä käytettiin klusteroinnissa havaintoina eli matriisin jokainen rivi on yksi havaintopiste euklidisessa 4- tai 9-ulotteisessa avaruudessa. Riippumatta valitusta eroavaisuuden mitasta, ensimmäinen vaihe on laskea havaintopisteiden väliset euklidiset etäisyydet (7), jolloin syntyy ensimmäisen vaiheen etäisyysmatriisi d . Valittu etäisyysmitta (8)–(10) tulee tarkasteluun vasta kun kaksi lähintä havaintopistettä on yhdistetty.

3.3.2 Klusteroinnin hyvyys

Kaikki klusterointitekniikat ovat luonteeltaan aineistolähtöisiä, kuvailevia tilastollisia menetelmiä. Lopputuloksena saadun klusteroinnin laadun ja sopivuuden arviointi on tämän vuoksi tärkeää. Klusterointialgoritmit pyrkivät löytämään aineistoon sopivan luokkarakenteen, vaikkei sellaista oikeasti olisi. Lopputuloksena syntynyt ryhmittely ei useinkaan ole tulkinnallisesti mielekäs. Algoritmin tuottaman ryhmittelyn tarkoitus on tiivistää aineistoa ja löytää havaintojen samankaltaisuuksia. Jos ryhmittely tuottaa useita luokkia, voi hyvin olla mahdollista, että aineistossa ei ole olemassa selvää ryhmärakennetta. Oikean menetelmän valinta onkin lähes aina subjektiivista ja tutkijan itse päätettävissä [10]. Lopputulokseen vaikuttaa pisteiden välisen etäisyyden laskemiseen käytettävä metriikka, eroavaisuuden mitan valinta sekä sopiva ryhmien määrän valinta. Jokaisella näistä päätöksistä voi olla voimakas vaikutus lopputulokseen, joten on syytä kokeilla useita eri vaihtoehtoja ja etsiä niistä sopivin ja tulkittavissa oleva ratkaisu.

Esitellään yksi mahdollinen tapa määrittää klustereiden määrä k . Tiedot perustuvat pääosin lähteeseen [12]. Klusterianalysissä kyynärpämenetelmä (*elbow method*) on heuristiikka, jota käytetään aineiston klustereiden lukumäärän määrittämiseen. Menetelmässä kuvataan klustereiden sisäinen neliösumma (WSS = *Within Cluster Sums of Squares*) klustereiden määrän funktiona. Klusterin sisäinen neliösumma on kaikkien klustereiden kunkin havainnon l etäisyyksien neliösumma niiden painopisteisiin \bar{C} :

$$\text{WSS} = \sum_{i=1}^k \sum_{l \in C_i} d(l, \bar{C}_i)^2. \quad (12)$$

Tavoitteena on löytää neliösumman kuvaajasta kohta, jonka jälkeen uuden klusterin lisääminen ei ole enää mielekäs. Tätä kohtaa kutsutaan ”kynärpääksi”. Sama menetelmä on yleisesti käytetty k-keskiarvon klusteroinnissa ja sitä on mahdollista hyödyntää päätöksenteon tukena dendrogrammin rinnalla. Luvussa 4 tätä menetelmää käytetään dendrogrammin tukena klustereiden määrää valittaessa.

3.4 Menetelmät ja ohjelmistot

Aineisto saatiin SAS7bdat-tiedostomuodossa ja sitä jatkokäsiteltiin R-studiossa. Analyysit toteutettiin R-ohjelmiston versiolla 4.1.2 [13]. Tärkeimmät tutkielman aineiston muokkaamiseen sekä analysointiin käytetyt paketit olivat *tidyverse* ja *factoextra*.

4 Aktiivisuuden jaksottuneisuus – empiirinen analyysi

Tässä luvussa esitellään tutkielman analyysi ja tulokset. Luku 4.1 esittelee aineiston yleispiirteitä. Luku 4.2 kattaa kaksitilaiseen malliin liittyvät tulokset ja analyysit. Luku 4.3 käsittelee kolmitilaisen mallin tuloksia ja analyysiä. Lisäksi lukuun 4.3 sisältyy klusteroinnin yhtäpitävyyden tutkimista.

4.1 Aineiston kuvaus

Analysoitavassa aineistossa oli yhteensä 280 osallistujaa. Kaiken kaikkiaan aineistossa oli rivejä reilut 2.26 miljoonaa. Tutkimusjoukon keski-ikä oli 63 vuotta. Suurin osa oli naisia (83 %). Koko tutkimusjoukosta 17 % teki fyysisesti raskasta työtä. Ammattinimekkeen perusteella tutkimusjoukko jakautui neljään ryhmään, erityisasiantijoihin ($N = 99$), asiantuntijoihin/toimistotyöntekijöihin ($N = 97$), palvelualan työntekijöihin ($N = 63$) ja fyysistä työtä tekeviin ($N = 21$). Tupakointi oli harvinaista (5 %). Taulukkoon 2 on koottu aineistosta laskettuja terveys- ja toimintakykyyn liittyviä tunnuslukuja.

	Kaikki	Miehet ($N = 47$)	Naiset ($N = 233$)
Tuki- ja liikuntaelämistön kroonisten sairauksien lukumäärä			
Ei lainkaan (%)	103 (37 %)	19 (40 %)	84 (36 %)
Yksi (%)	108 (39 %)	20 (43 %)	88 (38 %)
Kaksi tai kolme (%)	57 (20 %)	8 (17 %)	49 (21 %)
Puuttuvia arvoja (%)	12 (4 %)	0 (0 %)	12 (5 %)
Painoindeksi (kg/m^2)			
Keskiarvo (hajonta)	26.2 (4.5)	26.9 (3.6)	26.1 (4.7)
Mediaani [min.,maks.]	25.7 [17.0,44.3]	26.5 [17.0,35.7]	25.5 [17.0,44.3]
Puuttuvia arvoja	2	0	2
Vyötärön ympärys (cm)			
Keskiarvo (hajonta)	91.6 (12.9)	101.2 (11.2)	89.7 (12.4)
Mediaani [min.,maks.]	90.0 [65.1,139.8]	99.5 [73.1,123.9]	88.5 [65.1,139.8]
Hapenottokyky, VO₂ (ml/kg/min)			
Keskiarvo (hajonta)	28.7 (5.3)	32.3 (6.5)	27.9 (4.6)
Mediaani [min.,maks.]	28.0 [14.6,48.5]	30.9 [20.6,48.5]	27.3 [14.6,42.6]
Puuttuvia arvoja	45	4	41
Maksimaalinen kävelynopeus (m/s)			
Keskiarvo (hajonta)	1.7 (0.3)	1.8 (0.3)	1.7 (0.3)
Mediaani [min.,maks.]	1.6 [0.9,2.9]	1.7 [1.3,2.7]	1.6 [0.9,2.9]
Kymmeneen tuoliltanousuun kulunut aika (s)			
Keskiarvo (hajonta)	21.3 (4.7)	20.1 (4.7)	21.2 (4.6)
Mediaani [min.,maks.]	20.1 [12.8,52.7]	19.7 [14.8,42.0]	20.2 [12.8,52.7]
Puuttuvia arvoja	3	0	3

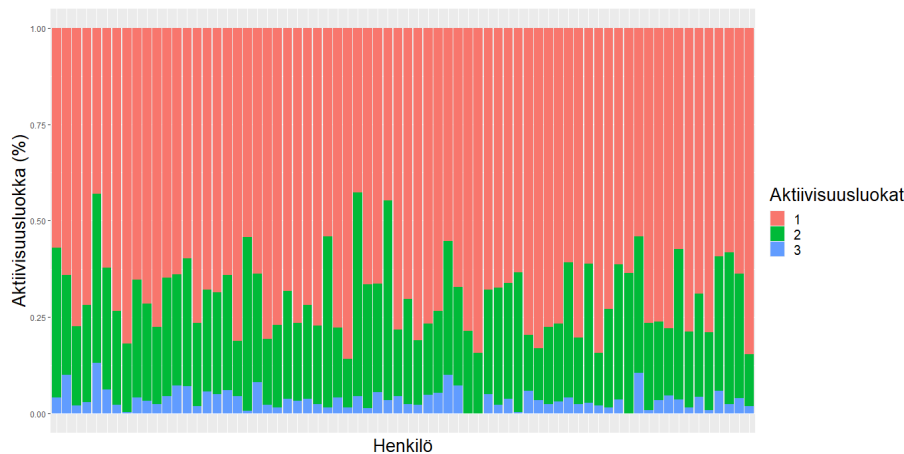
Taulukko 2: Tutkimusjoukon terveys- ja toimintakykytietoja. Tuki- ja liikuntaelämistön kroonisten sairauksien lukumäärä on summattu kysymyksistä: ”Onko lääkäri todennut jonkin seuraavista sairauksista? 1. Nivelkuluma, 2. Osteoporoosi, 3. Iskias (lannenikaman välilevyn pullistuma), 4. Nivelrikko.” Yhdelläkään yksilöllä ei ollut enempää kuin kolme tuki- ja liikuntaelämistön kroonista sairautta.

	Kaikki	Miehet ($N = 47$)	Naiset ($N = 233$)	Työ/vapaapäivä
Mittarin pitoaika (minuuttia/päivä) hereilläoloaikana				
Keskiarvo (hajonta)	946 (108)	955 (108)	949 (102)	984 (89)/917 (105)
Mediaani [min.,maks.]	957 [604,1351]	972 [604,1265]	958 [607,1351]	995 [604,1351]/919 [612,1265]
Seurantapäivien lukumäärä				
Keskiarvo (hajonta)	8.5 (1.8)	8.4 (1.5)	8.6 (1.9)	4.4 (1.6) /4.2 (1.8)
Mediaani [min.,maks.]	9 [2,12]	8 [4,11]	9 [2,12]	5 [0,9]/4 [0,10]

Taulukko 3: Mittauksen kesto.

Taulukkoon 3 on koottu mittauksen keston liittyviä tietoja. Liikemittaria pidettiin kädessä keskimäärin lähes 16 tuntia (per päivä), ja työpäivinä mittaria pidettiin keskimäärin hieman yli tunnin kauemmin kuin vapaapäivinä. Tutkittavat pitivät mittaria yötä päivää, mutta näistä analyyseistä uniaika on nyt poistettu. Työpäivinä tulee enemmän pitoaikaa hereilläoloaikana, koska työpäivinä nukutaan vähemmän kuin vapaapäivinä. Seurantapäivien lukumäärä oli keskimäärin 8.5 per tutkimusyksilö keskihajonnan ollessa 1.8. Työpäiviä oli keskimäärin 4.4 ja vapaapäiviä 4.2.

Aktiivisuusluokkien suhteelliset osuudet ensimmäisessä osa-aineistossa ($N = 70$) koko seuranta-ajan aikana on esitetty kuvassa 4. Kolmen muun osa-aineiston vastaavat tulokset on esitetty liitteen kuvissa A1–A3. Keskimäärin aktiivisuusluokkien suhteelliset osuudet ovat prosentin tarkkuudella samat myös kolmessa muussa osa-aineistossa. Koko aineistossa paikallaanoloa (tila 1) on selvästi eniten (keskimäärin 69 %) ja reipasta aktiivisuutta (tila 3) vähiten (keskimäärin 4 %). Yksilöiden välinen vaihtelu on suurinta paikallaanolon (vaihteluväli [32 %, 96 %]) ja kevyen aktiivisuuden (tila 2, vaihteluväli [4 %, 63 %]) osuuksissa. Reippaan aktiivisuuden vaihteluväli on [0 %, 16 %].



Kuva 4: Aktiivisuusluokkien jakautuminen ensimmäisessä osa-aineistossa ($N=70$). Tässä osa-aineistossa aktiivisuusluokkien 1, 2 ja 3 suhteelliset osuudet ovat keskimäärin 69 %, 27 % ja 4 %.

4.2 Aktiivisuuden jaksottuneisuus kaksitilaisessa mallissa

Kaksitilaisen mallin tila-avaruus on $S = \{1, 2\} = \{\text{paikallaan, aktiivinen}\}$. Kaikkien henkilöiden keskimääräinen 2×2 -tilasiirtymämatriisi kuvaa koko tutkimusjoukon yleistä aktiivisuuskäyttäytymistä. Jokaiselle yksilölle laskettiin ensin lausekkeen (3) mukaiset tilasiirtymien estimaatit ja näistä laskettiin sen jälkeen keskiarvo yli kaikkien tutkimushenkilöiden. Kaksitilaisen mallin tapauksessa keskimääräinen tilasiirtymämatriisi on (yksikkö ‘per minuutti’)

$$\hat{P}_2 = \frac{1}{280} \begin{bmatrix} 1 - \sum_{l=1}^{280} \hat{p}_{12}^{(l)} & \sum_{l=1}^{280} \hat{p}_{12}^{(l)} \\ \sum_{l=1}^{280} \hat{p}_{21}^{(l)} & 1 - \sum_{l=1}^{280} \hat{p}_{21}^{(l)} \end{bmatrix} = \begin{bmatrix} 0.89 & 0.11 \\ 0.26 & 0.74 \end{bmatrix}.$$

Keskimääräinen SATP-indeksi on siis 0.11 ja ASTP-indeksi 0.26. Matriisin \hat{P}_2 diagonaali-alkiot ovat selvästi suurempia kuin diagonaalien ulkopuoliset alkio. Tämä tarkoittaa sitä, että keskimäärin on todennäköisempää (per minuutti) pysyä sen hetkessä tilassa kuin siirtyä toiseen tilaan. Henkilö todennäköisimmin siis jatkaa paikallaanoloa tai vaihtoehtoisesti aktiivista toimintaa.

Tutkitaan vielä, miten todennäköisyydet eroavat, kun otetaan huomioon erikseen työ- ja vapaapäivät:

$$\hat{P}_{2,(\text{työ})} = \begin{bmatrix} 0.89 & 0.11 \\ 0.28 & 0.72 \end{bmatrix} \quad \text{ja} \quad \hat{P}_{2,(\text{vapaa})} = \begin{bmatrix} 0.89 & 0.11 \\ 0.24 & 0.76 \end{bmatrix}.$$

Työ- ja vapaapäivien välille muodostuu ero aktiivisesta tilasta siirtymisen todennäköisyydessä. Vapaapäivänä jatketaan työpäivää todennäköisemmin liikkeessä minuutin jakson aikana.

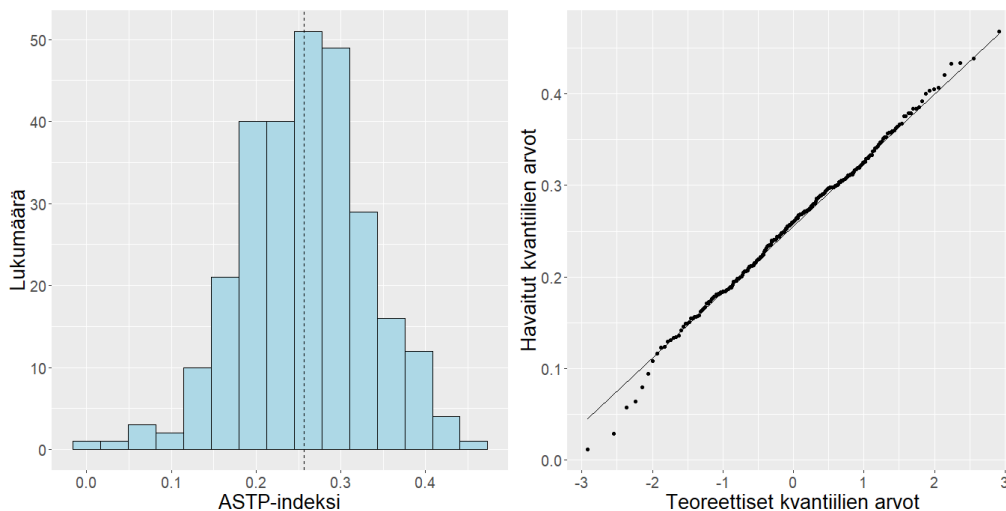
Taulukko 4 kokoaa yhteen kaksitilaisen mallin perusteella estimoidut SATP- ja ASTP-indeksit sekä keskimääräiset kummassakin tilassa yhtäjaksoisesti vietetyt ajat. Keskilukujen (keskiarvo ja mediaani) mukaan miehillä on naisia pienempi todennäköisyys siirtyä paikallaanolosta aktiiviseen tilaan ja suurempi todennäköisyys siirtyä aktiivisesta tilasta paikallaanoloon.

Paikallaanolojakson arvioitu keskimääräinen kesto on noin 12 minuuttia ja aktiivisen jakson reilu neljä ja puoli minuuttia. Miehillä on keskimäärin pidemmät paikallaanolo- ja aktiivisuusjaksot. Paikallaanolon kohdalla tämä selityy osin sillä, että miehillä on selvästi naisia suurempi havaintojen välinen vaihtelu ja suuria poikkeavia havaintoja. Keskiarvojen vertailun epävarmuus johtuu siis suuresta hajonnasta ja siitä, että miehiä on naisiin verrattuna suhteellisen vähän. Paikallaanolojakson keskimääräinen kesto asettuu reiluun 12 minuuttiin sekä työ- että vapaapäivinä. Työpäivinä aktiivisen tilan keskimääräisen jakson kesto on noin neljä ja puoli minuuttia ja vapaapäivinä hieman yli viisi minuuttia.

	Kaikki	Miehet ($N = 47$)	Naiset ($N = 233$)	Työ/vapaapäivä
SATP-indeksi				
Keskiarvo (hajonta)	0.11 (0.04)	0.09 (0.04)	0.11 (0.04)	0.11 (0.05)/0.11 (0.04)
Mediaani [min.,maks.]	0.11 [0.01,0.26]	0.09 [0.01,0.18]	0.11 [0.01,0.26]	0.11[0.00,0.30]/0.11 [0.00,0.25]
Paikallaanolojakson kesto				
Keskiarvo (hajonta)	11.77 (15.04)	17.17 (26.13)	10.68 (11.38)	12.15 (17.95)/12.18 (14.14)
Mediaani [min.,maks.]	9.47 [3.82,174.95]	10.83 [5.48,174.95]	9.13 [3.82,161.90]	9.19 [3.32,222.31]/9.34 [4.02,150.44]
ASTP-indeksi				
Keskiarvo (hajonta)	0.26 (0.07)	0.28 (0.09)	0.25 (0.07)	0.28 (0.08)/0.24 (0.08)
Mediaani [min.,maks.]	0.26 [0.01,0.47]	0.29 [0.01,0.44]	0.25 [0.06,0.47]	0.28 [0.01,0.55]/0.23 [0.01,0.50]
Aktiivisuusjakson kesto				
Keskiarvo (hajonta)	4.68 (5.57)	6.22 (13.03)	4.37 (1.75)	4.55(7.47)/5.12 (4.92)
Mediaani [min.,maks.]	3.84 [2.14,87.58]	3.42 [2.28,87.58]	3.91 [2.14,17.45]	3.55 [1.82,120.23]/4,29 [2.00,70.60]

Taulukko 4: Kaksitilaisen mallin tärkeimpiä tuloksia. Kaikki taulukon tiedot laskettiin estimoitujen siirtymätodennäköisyyksien avulla. Erityisesti SATP-indeksi on tällöin siirtymän $1 \rightarrow 2$ (paikallaanolo \rightarrow aktiivisuus) todennäköisyys ja keskimääräinen paikallaanolon yhtäjaksoinen kesto tämän todennäköisyyden käänteisluku; vastaavasti ASTP-indeksi. Kestojen yksikkö on minuutti.

ASTP- ja SATP-indeksin jakaumien luonnehdinta. ASTP-indeksin yksilöiden välinen vaihtelu on $[0.01,0.47]$ keskihajonnan ollessa 0.07. Vastaavasti SATP-indeksin yksilöiden välinen vaihteluväli on $[0.01,0.26]$ ja keskihajonta 0.04. Työpäivien ASTP-indeksin yksilöiden välinen vaihtelu asettuu välille $[0.01,0.55]$ keskihajonnan ollessa 0.08. Vastaava vapaapäivien vaihteluväli on $[0.01,0.50]$ ja keskihajonta 0.08. Koko tutkimusjoukon ASTP-indeksin frekvenssijakauma ja kvantiilikuvio on esitetty kuvassa 5. Jakaumaa voidaan pitää kuvan perusteella likimain normaalisena.

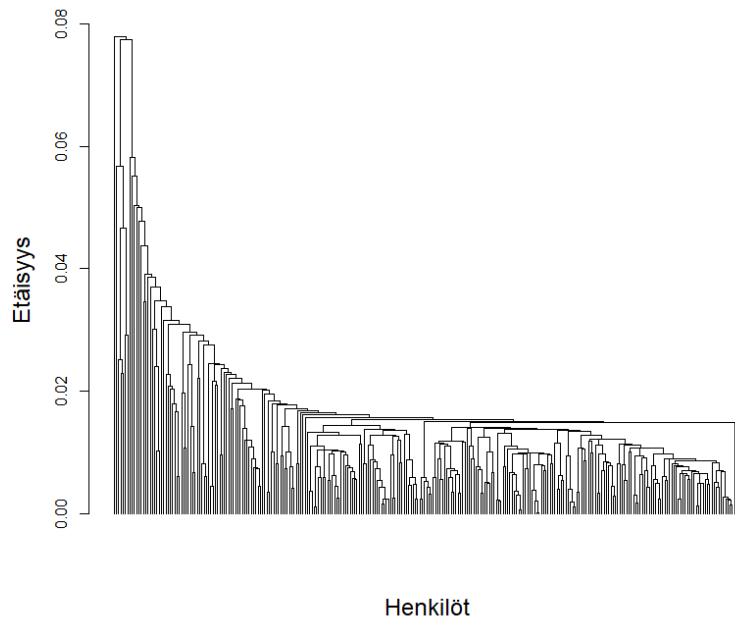


Kuva 5: ASTP-indeksin frekvenssijakauma ja kvantiilikuvio. ASTP-indeksin keskiarvo (katkoviiva) on 0.26.

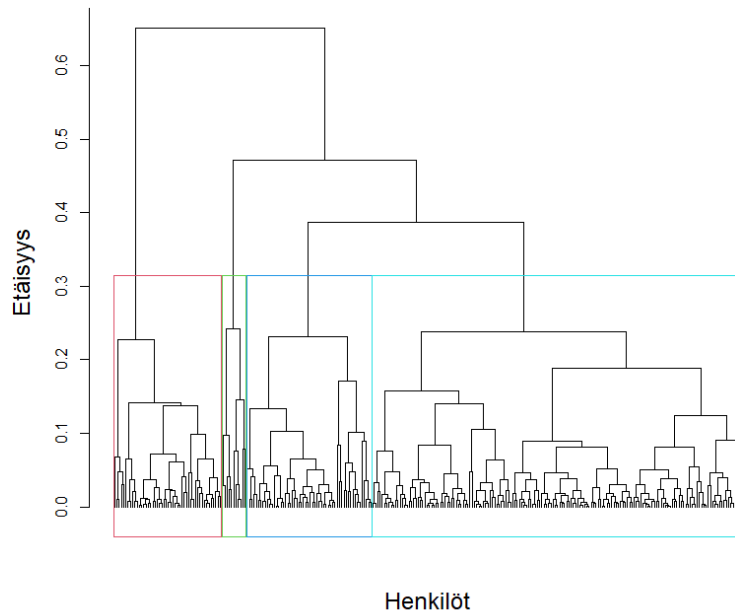
4.2.1 Kaksitilaisen mallin klusterit

Matriisia (11) käytettiin klusteroinnissa etäisyysmatriisina. Ryhmien muodostaminen aloitettiin soveltamalla luvussa 3 esiteltyjä kokoavan menetelmän tekniikoita. Klusterointi suoritettiin käyttäen kaikkia kolmea eroavaisuuden mittaa (8)–(10). Dendrogrammien perusteella näistä valittiin parhain tarkempaa analyysiä varten.

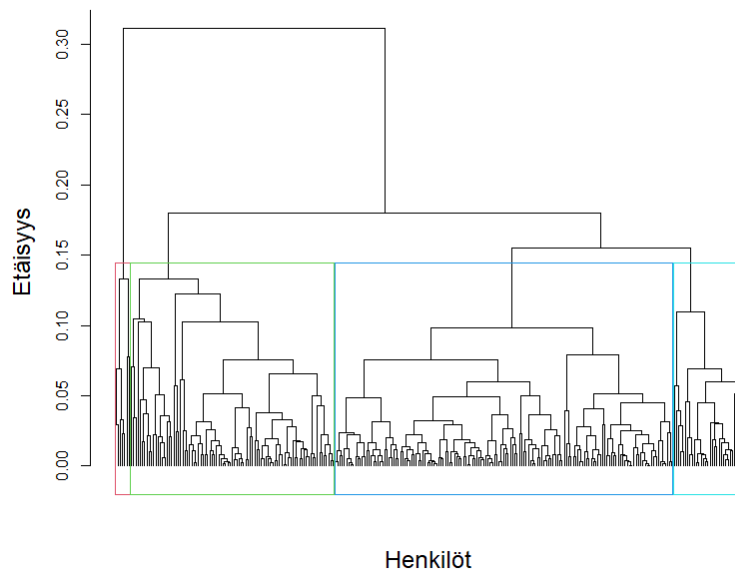
Kuva 6 esittää kaksitilaiselle mallille muodostettua lähimmän naapurin menetelmän dendrogrammia. Kuvasta on erittäin vaikea havaita mitään järkevää ryhmärakennetta. Valinnoilla $k = 2, 3, 4, 5$ yksi klusteri sisältää minimissään 272 yksilöä. Vastaava tulos saatiin nelitilaiselle mallille projektityössä [4]. Menetelmä siis luo yhden lähes koko aineiston käsittävän ryhmän ja loput ryhmistä ovat kooltaan todella pieniä. Kuvien 7 ja 8 perusteella huomataan, että aineisto voidaan kaukaisimman- ja naapurikeskiarvon menetelmän dendrogrammin perusteella jakaa neljään klusteriin.



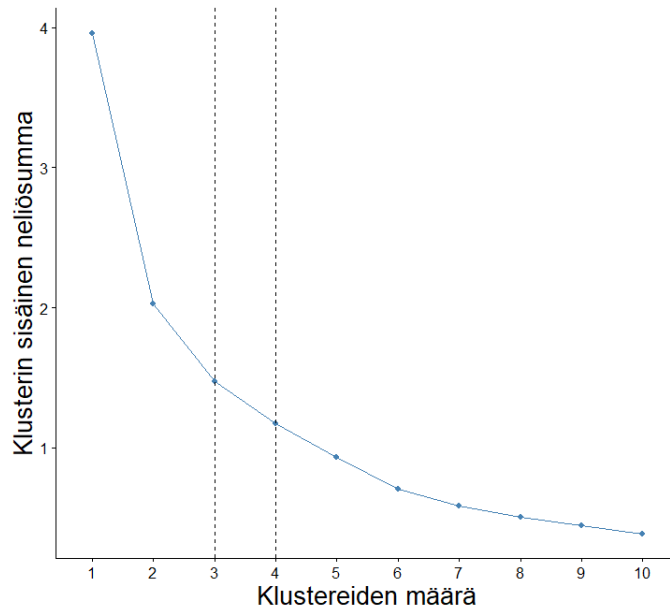
Kuva 6: Lähimmän naapurin menetelmän tuottama dendrogrammi kaksitilaiselle mallille. Kuvasta on vaikea erottaa järkevää ryhmärakennetta. Yksi klusteri on kooltaan lähes koko aineiston kokoinen järkevän pienillä klusterimäärän k valinnoilla.



Kuva 7: Kaukaisimman naapurin menetelmän tuottama dendrogrammi kaksitilaiselle mallille. Kuvassa on eroteltu eri värien avulla neljä klusteria (2A, 2B, 2C, 2D). Klustereiden koot ovat 48, 11, 56 ja 165.



Kuva 8: Naapurikeskiarvon menetelmän tuottama dendrogrammi kaksitilaiselle mallille. Kuvassa on eroteltu eri värien avulla neljä klusteria. Klustereiden koot ovat 7, 91, 151 ja 31.



Kuva 9: Klustereiden sisäinen neliösumma klusterimäärän k funktiona. Valinta $k = 3$ tai $k = 4$ vaikuttaisi järkevältä. Kuva antaa tukea kaukaisimman naapurin ja naapurikeskiarvon menetelmän dendrogrammien perusteella tehdyille valinnalle $k = 4$.

Dendrogrammien perusteella vaikuttaisi järkevältä edetä joko kaukaisimman naapurin- tai naapurikeskiarvon menetelmän mukaisen ryhmäjaon perusteella. Tarkastellaan tässä kaukaisimman naapurin menetelmän mukaisia klustereita ja raportoidaan liitteissä vastaavat naapurikeskiarvon menetelmän tulokset. Kyynärpäämenetelmä (kuva 9) perustuu kaavaan (12) ja suosittaa kolmea tai neljää klusteria, joten valitaan neljä dendrogrammiin vedoten.

Jaetaan siis henkilöt kuvan 7 perusteella neljään klusteriin (koot 48, 11, 56 ja 165). Klusterikohtaisesti lasketut keskimääräiset tilasiirtymätodennäköisyysmatriisit ovat

$$\hat{P}_{2,(2A)} = \begin{bmatrix} 0.91 & 0.09 \\ 0.36 & 0.64 \end{bmatrix}, \quad \hat{P}_{2,(2B)} = \begin{bmatrix} 0.96 & 0.04 \\ 0.10 & 0.90 \end{bmatrix}$$

$$\hat{P}_{2,(2C)} = \begin{bmatrix} 0.84 & 0.16 \\ 0.18 & 0.82 \end{bmatrix} \quad \text{ja} \quad \hat{P}_{2,(2D)} = \begin{bmatrix} 0.89 & 0.11 \\ 0.26 & 0.74 \end{bmatrix}.$$

Seuraavaksi vertaillaan ja luonnehditaan näitä neljää klusteria sekä yksilökohtaisten tilasiirtymätodennäköisyyksien (3) että yllä esitettyjen keskimääräisten siirtymätodennäköisyyksien avulla. Huomioidaan vertailussa ja luonnehdinnassa taulukkoon 5 koottuja klusterikohtaisia terveys- ja toimintakykytietoja sekä aktiivisuusluokkien suhteellisia osuuksia koko seuranta-ajan aikana. Aktiivisuusluokkien suhteelliset osuudet on esitetty liitteen B kuvissa B1–B4.

Ensimmäinen klusteri (2A) on jo dendrogrammin perusteella kaikkein erilaisin, koska se yhdistyy viimeisenä muihin. Klusterissa on suhteellisesti enemmän miehiä (27 %) kuin naisia. Yksilöillä on keskimäärin pienin todennäköisyys (64 %) jatkaa

aktiivisessa tilassa ja toiseksi suurin todennäköisyys (91 %) pysyä paikallaan minuutin jakson aikana. Klusterin aktiivisuuskäyttäytymistä voisi luonnehtia passiiviseksi, koska paikallaanolojaksot ovat verrattain pitkiä ([7min, 27min]) ja aktiivisuusjaksot lyhyitä (kahdesta minuutista vähän yli kolmeen minuuttiin). Klusterin yksilöt viettävät keskimäärin 81 % hereilläoloajastaan paikallaan. Terveys- ja toimintakykytiedot ovat toisen klusterin kanssa samankaltaiset (taulukko 5). Kolmanteen ja neljänteen klusteriin verrattuna erot huomaa selvimmin painoindeksissä ja vyötärön ympärysmittassa.

Toinen klusteri (2B) on kooltaan selvästi pienin ($N = 11$) ja siinä on suhteellisesti enemmän miehiä (36 %) kuin naisia. Klusteriin päätyy yksilöitä, joilla on kaikilla yli 90 %:n todennäköisyys (vaihteluväli [91 %, 99 %]) pysyä paikallaan minuutin jakson aikana. Kun nämä yksilöt pääsevät aktiiviseen tilaan, he myös hyvin todennäköisesti (83-99 %) pysyvät siinä minuutin jakson aikana. Näitä yksilöitä voisi luonnehtia paikallaanpysyviksi, joilla paikallaanolo- ja aktiivisuusjakson odotettu kesto on verrattain suuri. Toisaalta aktiivisuusjaksojen lukumäärä on verrattain vähäinen. Klusteria voisi luonnehtia jossain mielessä poikkeukselliseksi. Pienimmillään aktiivisuusjakson odotettu kesto tässä klusterissa on kuusi minuuttia ja korkeimmillaan melkein 88 minuuttia. Klusterin yksilöt viettävät keskimäärin 71 % hereilläoloajastaan paikallaan eli 10 %-yksikköä vähemmän kuin klusterin 2A yksilöt. Mediaanin mielessä klusterissa on yksilöitä, joilla on kolmeen muuhun klusteriin verrattuna keskimäärin korkeampi painoindeksi sekä vyötärön ympärysmitta. Klusterilla on keskimäärin parhain hapenottokyky (30.0) eli voisi ajatella klusterin käsittävän henkilöitä, jotka todennäköisesti harrastavat kuntoliikuntaa, mutta eivät ehkä muuten liiku juurikaan.

Kolmas klusteri (2C) koostuu suurimmaksi osaksi naisista (95 %) ja jakautuu ammattiaseman perusteella tasan asiantuntijoiden (erityisasiantuntijat ja asiantuntijat/toimistotyöntekijät) ja palvelu- ja fyysistä työtä tekevien kesken. Klusterin yksilöillä on muihin klustereihin verrattuna keskimäärin korkeampi todennäköisyys siirtyä paikallaanolosta aktiiviseen tilaan ja toiseksi korkein todennäköisyys siirtyä aktiivisesta tilasta paikallaanoloon. Tätä klusteria voisi luonnehtia aktiivisimmaksi alati liikkujien klusteriksi, jolla paikallaanolonjakson keskimääräinen kesto on kaikkein pienin. Klusterin yksilöt viettävät keskimäärin 53 % hereilläoloajastaan paikallaan eli toisaalta ovat melkein puolet tästä ajasta liikkeessä. Terveys- ja toimintakykytietojen perusteella kolmannen klusterin tulokset ovat selvästi kauttaaltaan parhaimmat.

Neljäs klusteri (2D) on kooltaan selvästi suurin ($N = 165$). Aktiivisessa tilassa pysymisen todennäköisyys vaihtelee klusterin sisällä välillä [67 %, 82 %]. Nämä kaikki aktiivisessa tilassa pysymisen todennäköisyydet ovat pienempiä kuin yksikään toisen klusterin vastaava todennäköisyys. Toisaalta yksilöt viettävät aktiivisessa tilassa keskimäärin suhteellisesti yhtä paljon aikaa (29 %) kuin 2B klusterin yksilöt. Huomataan, että näiden kahden klusterin aktiivisuusprofiilit poikkeavat toisistaan jaksottuneisuuden suhteen mutta ovat likipitään samanlaisia aktiivisuuden kokonaismäärän suhteen. Dendrogrammin perusteella neljäs klusteri on lähimpänä kolmatta klusteria ja kauimpana ensimmäisestä klusterista. Terveys- ja toimintakykytiedot ovat lähes samansuuntaisia kolmannen klusterin kanssa, mutta pieni ero on havaittavissa kolmannen klusterin eduksi. Neljäs klusteri osuu kahden ensimmäisen

ja kolmannen klusterin välimaastoon keskimääräisten siirtymätodennäköisyyksien sekä terveys- ja toimintakykytietojen perusteella. Ainoastaan toisen klusterin korkea aktiivisessa tilassa pysymisen todennäköisyys rikkoo tätä tulkintaa. Klusteria voisi luonnehtia keskivertoliikkujien klusteriksi.

	Klusteri 2A (N = 48)	Klusteri 2B (N = 11)	Klusteri 2C (N = 56)	Klusteri 2D (N = 165)
Sukupuoli				
Miehet Lkm. (%)	13 (27 %)	4 (36 %)	3 (5 %)	27 (16 %)
Naiset Lkm. (%)	35 (73 %)	7 (64 %)	56 (95 %)	138 (84 %)
Ammattiasema				
Asiantuntijat Lkm. (%)	39 (81 %)	9 (82 %)	28 (50 %)	120 (73 %)
Palvelu- ja fyysinen työ Lkm. (%)	9 (19 %)	2 (18 %)	28 (50 %)	45 (27 %)
Painoindeksi (jatkuva, kg/m²)				
Keskiarvo (hajonta)	28.7 (5.5)	27.9 (4.2)	25.0 (3.2)	25.8 (4.4)
Mediaani [min.,maks.]	27.3 [17.0,44.2]	28.2 [21.5,37.5]	24.9 [18.3,33.1]	25.4 [17.0,44.3]
Puuttuvia arvoja	1	0	0	1
Vyötärönympäryys (cm)				
Keskiarvo (hajonta)	99.9 (14.2)	97.3 (13.2)	87.8 (9.6)	90.1 (12.4)
Mediaani [min.,maks.]	97.4 [67.2,139.8]	98.3 [74.7,123.3]	86.2 [73.5,122.2]	89.0 [65.1,123.2]
Hapenottokyky (V02, ml/kg/min)				
Keskiarvo (hajonta)	26.7 (4.8)	30.0 (6.6)	28.8 (4.0)	29.0 (5.6)
Mediaani [min.,maks.]	27.1 [14.6,37.9]	28.8 [21.5,41.1]	28.1 [21.3,37.3]	28.4 [17.1,48.5]
Puuttuvia arvoja	13	1	5	26
Maksimaalinen kävelynopeus (m/s)				
Keskiarvo (hajonta)	1.65 (0.33)	1.64 (0.31)	1.72 (0.32)	1.66 (0.29)
Mediaani [min.,maks.]	1.55 [1.14,2.68]	1.54 [1.14,2.17]	1.69 [1.23,2.67]	1.60 [0.89,2.86]
Kymmeneen tuoltanousuun kulunut aika (s)				
Keskiarvo (hajonta)	22.6 (5.8)	20.9 (4.6)	20.3 (4.4)	20.8 (4.3)
Mediaani [min.,maks.]	20.2 [14.9,42.0]	20.3 [15.2,30.1]	19.9 [12.8,44.7]	20.1 [13.5,52.7]
Puuttuvia arvoja	1	0	0	2
Tuki- ja liikuntaelimestön kroonisten sairauksien lukumäärä				
Ei lainkaan (%)	18 (37 %)	3 (27 %)	23 (41 %)	59 (36 %)
Yksi (%)	19 (40 %)	6 (55 %)	20 (36 %)	63 (38 %)
Kaksi tai kolme (%)	10 (21 %)	2 (18 %)	10 (18 %)	35 (21 %)
Puuttuvia arvoja (%)	1 (2 %)	0 (0 %)	3 (5 %)	8 (5 %)

Taulukko 5: Klusterikohtaisesti laskettuja terveys- ja toimintakykytietoja. Klusterointi suoritettiin kaksitilaiselle mallille kaukaisimman naapurin menetelmällä. Asiantuntija-asemassa olevat käsittävät taulukon tapauksessa erityisasiantuntijat ja asiantuntijat/toimistotyöntekijät.

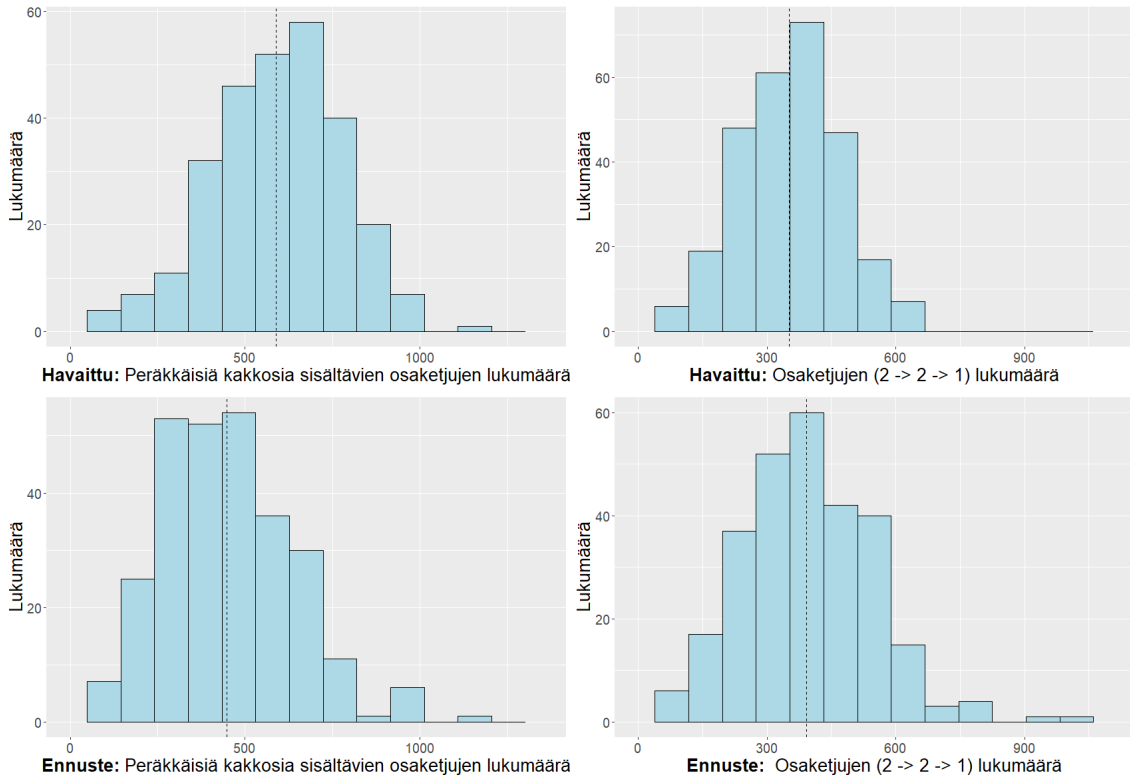
4.2.2 Mallin oletusten ja hyvyyden arviointi

Arvioidaan kaksitilaisen mallin oletuksia ja hyvyyttä luvun 3.1.3 keinojen avulla. Lasketaan ensin kaikille yksilöille aktiivisuusjaksojen lukumäärä eli peräkkäisiä kakosia (tila 2) sisältävien osaketjujen lukumäärä ja verrataan tätä jakaumaa vastaavan aikahomogeenisen Markovin ketjun ennustejakaumaan. Ennustejakauma on muodostettu vastaavan aikahomogeenisen Markov-ketjun estimoitujen tilasiirtymätodennäköisyyksien avulla. Lasketaan kullekin yksilölle yksi ennuste ja tarkastellaan miten jakaumien tunnusluvut eroavat toisistaan.

Toiseksi lasketaan jokaiselle yksilölle osaketjujen (2 → 2 → 1) lukumäärä. Osaketjun (2 → 2 → 1) siirtymä tilasta kaksi tilaan yksi ei mallin oletusten mukaan riipu siirtymää edeltävästä tilasta.

Kuva 10 esittää havainto- ja ennustejakaumia edellä esitetyille testisuureille. Aktiivisuusjaksojen lukumäärä oli havaitussa aineistossa keskimäärin 589, kun taas ennustetussa tapauksessa 448. Ennuste siis keskimäärin aliarvioi aktiivisuusjaksojen lukumäärää noin 34 %:lla. Osaketjujen (2 → 2 → 1) lukumäärä oli havaitussa aineistossa keskimäärin 352, kun taas ennustetussa tapauksessa 392. Nyt ennuste yliarvioi osaketjujen lukumäärää 11 %:lla. Jakaumat ovat havaitun ja ennusteen mielessä muodoiltaan hyvin samanlaisia etenkin osaketjujen (2 → 2 → 1) suhteen.

Oletettavasti $2 \rightarrow 1$ siirtymien lukumäärän malli kykenee ennustamaan hyvinkin tarkasti. Ennuste vain katkaisee hieman aikaisemmin peräkkäisistä aktiivisista tiloista koostuvat ketjut eli ei pysty täydellisesti kuvaamaan aktiivisuuden jaksottaisuutta. Näiden tulosten perusteella kuitenkin voidaan sanoa, että havaitut ketjut ovat kohtuullisen lähellä aikahomogeenista Markovin ketjua keskimääräisesti.



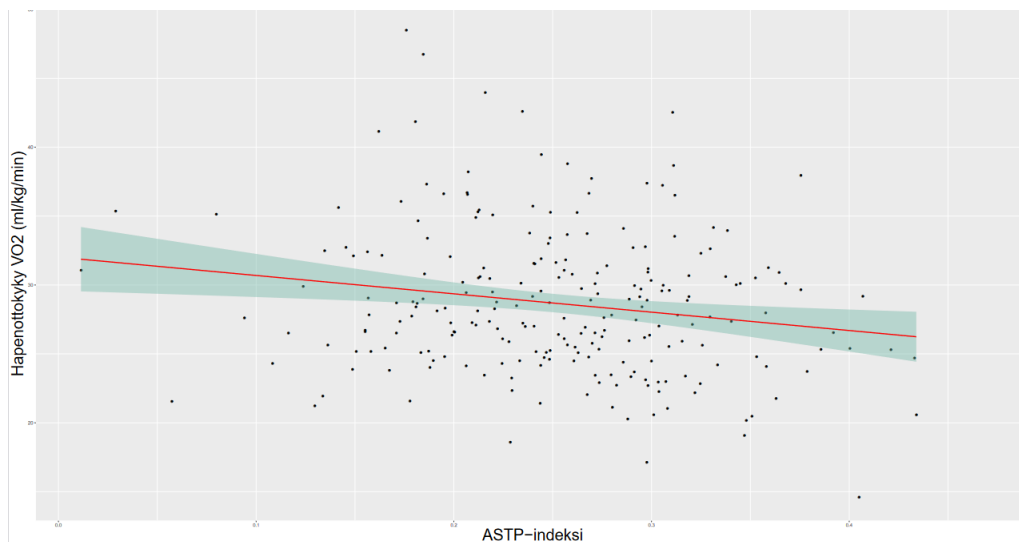
Kuva 10: Havainto- ja ennustejakaumat kaksitilaiselle mallille. Vasemmalla ylhäällä havaittu aktiivisuusjaksojen lukumäärän jakauma (keskiarvo 589) ja oikealla osaketujen ($2 \rightarrow 2 \rightarrow 1$) havaittu lukumäärän jakauma (keskiarvo 352). Vasemmalla alhaalla ennustettu aktiivisuusjaksojen lukumäärän jakauma (keskiarvo 448) ja oikealla osaketujen ($2 \rightarrow 2 \rightarrow 1$) ennustettu lukumäärän jakauma (keskiarvo 392).

4.2.3 Aktiivisuuden jaksottaisuus ja terveystvasteet

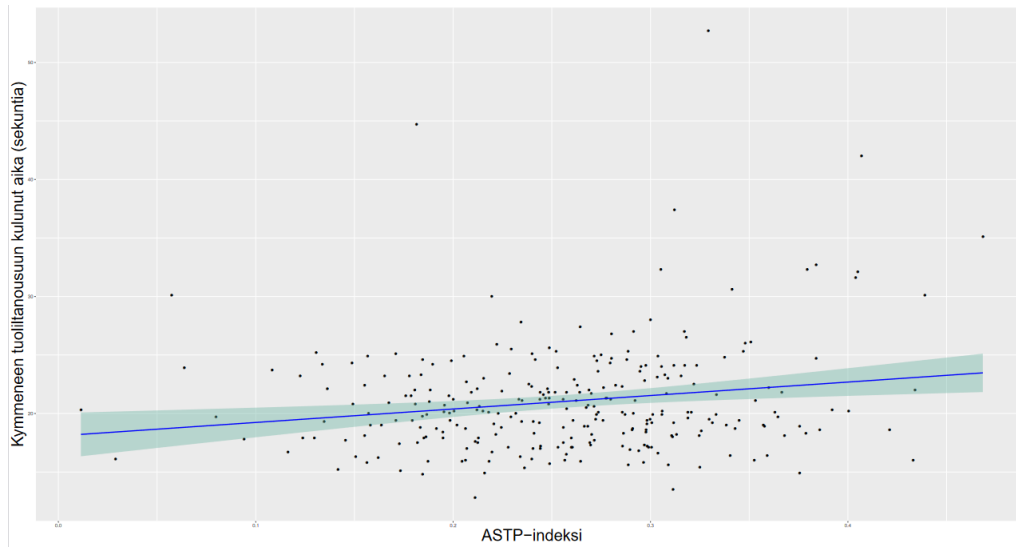
Tässä luvussa esitellään kaksitilaiseen malliin liittyvät analyysimallien tulokset. Esi- tellään aluksi kuvailevia tuloksia. Kuvailevat tulokset poikkeavat taulukon 6 tulok- sista, koska taulukon malleissa tulokset on vakioitu tiettyjen taustamuuttjien suh- teen.

ASTP-indeksi ja terveystvasteet. Kuvat 11 ja 12 esittävät ASTP-indeksin ja hapenottokyvyn sekä ASTP-indeksin ja kymmeneen tuoliltanousuun kuluneen ajan välistä yhteyttä. Molempiin kuviin on piirretty pienimmän neliösumman (PNS) suo- ra. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Selvästi huomataan, että hapenottokyky on keskimäärin sitä parempi, mitä pienempi on ASTP-indeksi. PNS-suoran kulmakerroin on -13.3 eli 0.1 yksikön lisäys ASTP-indeksissä vaikut- taa hapenottokykyyn -1.33 yksikköä. Toisaalta kymmeneen tuoliltanousuun kuuluu

keskimäärin sitä vähemmän aikaa mitä pienempi ASTP-indeksi on. Tällöin 0.1 yksikön lisäys ASTP-indeksissä vaikuttaa kymmeneen tuoliltanousuun kuluneeseen aikaan noin 1.2 sekuntia. Molemmissa tapauksissa luottamusväli ei sisällä arvoa nolla. Kolmas terveystilasto kuvaa maksimaalista kävelynopeutta (m/s). ASTP-indeksin ja maksimaalisen kävelynopeuden välillä ei näytä olevan yhteyttä. PNS-suoran kulmakerroin on -0.03 ja 95 %:n luottamusväli [-0.51,0.44] (liite C: kuva C1).



Kuva 11: ASTP-indeksin ja hapenottoikyyn VO₂ (ml/kg/min) välinen yhteys. Pistejoukkoon sovitettu pienimmän neliösumman suora. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Pienimmän neliösumman suoran kulmakerroin on -13.3 ja 95 %:n luottamusväli [-22.6,-4.0]. Tieto hapenottoikyvystä puuttuu 45 yksilöltä.



Kuva 12: ASTP-indeksin ja kymmeneen tuoliltanousuun kuluneen ajan (s) välinen yhteys. Pistejoukkoon sovitettu pienimmän neliösumman suora. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Pienimmän neliösumman suoran kulma-kerroin on 11.5 ja 95 %:n luottamusväli [4.2,18.8]. Tieto kymmeneen tuoliltanousuun kuluneesta ajasta puuttuu kolmelta yksilöltä.

Taulukkoon 6 on koottu lineaarisen regression tulokset kaikille kolmelle terveysteelle, kun selittävänä muuttujana on joko ASTP-indeksi tai klusteri. ASTP-indeksiä käsitellään joko jatkuvana tai kategorisena selittäjänä. Klusteri on luonnollisesti kategorinen selittäjä.

Vastemuuttuja	Hapenottokyky (<i>ml/kg/min</i>)	Maksimaalinen kävelynopeus (<i>m/s</i>)	Kymmeneen tuoliltanousuun kulunut aika (s)
Jatkuva ASTP (per 0.1 yksikköä)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	-1.21 [-1.92,-0.51]	-0.01 [-0.06,0.04]	0.81 [0.08,1.55]
R^2/R_{adj}^2	0.51/0.49	0.06/0.04	0.09/0.07
Kategorinen ASTP			
Alatertiili ($N = 94$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	referenssiluokka	referenssiluokka	referenssiluokka
Keskitertiili ($N = 93$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	-0.76 [-1.99,0.47]	-0.003 [-0.09,0.09]	-0.02 [-1.30,1.27]
Ylätertiili ($N = 93$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	-1.81 [-3.08,-0.54]	-0.01 [-0.10,0.08]	1.23 [-0.10,2.57]
R^2/R_{adj}^2	0.50/0.48	0.06/0.03	0.09/0.07
Klusteri			
Klusteri 2A ($N = 48$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	referenssiluokka	referenssiluokka	referenssiluokka
Klusteri 2B ($N = 11$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	3.06 [0.34,5.78]	-0.01 [-0.21,0.19]	-1.24 [-4.12,1.64]
Klusteri 2C ($N = 56$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	1.48 [-0.27,3.24]	0.10 [-0.03,0.22]	-1.52 [-3.35,0.31]
Klusteri 2D ($N = 165$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	1.52 [0.06,2.98]	0.03 [-0.07,0.13]	-1.04 [-2.53,0.45]
R^2/R_{adj}^2	0.50/0.48	0.07/0.04	0.09/0.06

Taulukko 6: Lineaarisen regression tulokset. Vastemuuttujia ovat hapenottokyky, maksimaalinen kävelynopeus ja kymmeneen tuoliltanousuun kulunut aika. ASTP-indeksiä käsitellään vaihtoehtoisesti joko jatkuvana tai kategorisena selittäjänä. Klusteri on luonnollisesti kategorinen selittäjä. Mallit vakioitiin demografisten tietojen (ikä, sukupuoli ja työn kuormittavuus), painoindeksin ja kroonisten sairauksien historian mukaan.

Tulosten perusteella 0.1 yksikön lisäys ASTP-indeksissä vaikuttaa hapenottokyydyn -1.21 (*ml/kg/min*), maksimaaliseen kävelynopeuteen -0.01 (*m/s*) ja kymmeneen tuoliltanousuun kuluneeseen aikaan 0.81 (*s*). Parametrien 95 %:n luottamusvälit eivät sisällä nollaa, kun vastemuuttujana on hapenottokyky tai kymmeneen tuoliltanousuun kuluneeseen aika. Hapenottokyky mallin vasteena tuottaa selvästi parhaimman selityksasteen R^2 (0.51). Tällöin ASTP-indeksi ja muut valitut selittäjät muuttujat pystyvät selittämään 51 % hapenottokyvyn vaihtelusta.

Kategorisen ASTP-indeksin alatertiili toimii vertailuryhmänä. Keskitertiilin ryhmän hapenottokyky on -0.76 pienempi kuin vertailuryhmällä, mutta parametrin 95 %:n luottamusväli sisältää arvon nolla. Toisaalta ylätertiilin eli suurimman ASTP-indeksin omaava ryhmä poikkeaa selvästi vertailuryhmästä. Tällöin ryhmien välinen ero on -1.81 eli vertailuryhmällä on parempi hapenottokyky. Muiden terveysteille

den tapauksessa luottamusvälit kattoivat arvon nolla.

Klusteri 2A toimii nyt vertailuryhmänä. Kaikilla kolmella klusterilla 2B, 2C ja 2D on parempi hapenottokyky kuin klusterilla 2A, mutta vain klustereiden 2B ja 2D kohdalla vaikutuksen 95 %:n luottamusväli ei sisällä nollaa. Tulos vahvistaa klusterikohtaisten keskimääräisten tilasiirtymätodennäköisyyksien ja taulukon 5 tulkintoja. Maksimaalinen kävelynopeus puolestaan ei kykene erottelemaan klustereita toisistaan. Huomataan myös, että kolme muuta klusteria saavuttavat vertailuryhmään nähden parempia tuoliltanousuun liittyviä tuloksia, vaikka luottamusvälit sisältävätkin arvon nolla. Arvo nolla on kuitenkin selvästi luottamusvälien keskikohdan oikealla puolella. Näyttää siltä, että klusterointi pystyi tunnistamaan jonkinlaisen riskiryhmän (klusteri 2A) eli ryhmän, jolla on huonoimmat terveys- ja toimintakykytiedot sekä eniten/vähiten paikallaanoloa/aktiivisuutta.

4.3 Kolmitilainen malli

Kolmitilaisen mallin tila-avaruus on $S = \{1, 2, 3\} = \{\text{paikallaan, kevyt aktiivisuus, reipas aktiivisuus}\}$ ja koostuu kaksitilaiseen malliin verrattuna kahdesta aktiivisesta tilasta. Kaikkien yksilöiden yli laskettu keskimääräinen 3×3 -tilasiirtymämatriisi on

$$\hat{P}_3 = \begin{bmatrix} 0.89 & 0.10 & 0.01 \\ 0.28 & 0.68 & 0.04 \\ 0.15 & 0.34 & 0.51 \end{bmatrix}.$$

Diagonaalialkiot ovat selvästi suurempia verrattuna vastaavan rivin muihin alkioihin. Minuutin jakson aikana on siis todennäköisintä pysyä senhetkisessä tilassa. On selvää, että useimmin on hyvin epätodennäköistä siirtyä paikallaanolosta suoraan reippaaseen aktiivisuuteen. Kevyestä aktiivisuudesta reippaaseen aktiivisuuteen siirtymisen keskimääräinen todennäköisyys minuutin jakson aikana on myös hyvin pieni, vain 4 %. Reippaasta aktiivisuudesta siirrytään todennäköisemmin kevyeen aktiivisuuteen kuin paikallaanoloon.

Tutkitaan tässäkin tilanteessa eroa työ- ja vapaapäivien välillä:

$$\hat{P}_{3,(\text{työ})} = \begin{bmatrix} 0.89 & 0.10 & 0.01 \\ 0.30 & 0.64 & 0.06 \\ 0.17 & 0.36 & 0.47 \end{bmatrix} \quad \text{ja} \quad \hat{P}_{3,(\text{vapaa})} = \begin{bmatrix} 0.89 & 0.10 & 0.01 \\ 0.25 & 0.71 & 0.04 \\ 0.14 & 0.33 & 0.53 \end{bmatrix}.$$

Vapaapäivinä keskimääräiset tilasiirtymätodennäköisyydet aktiivisista tiloista paikallaanoloon ovat pienempiä kuin työpäivinä.

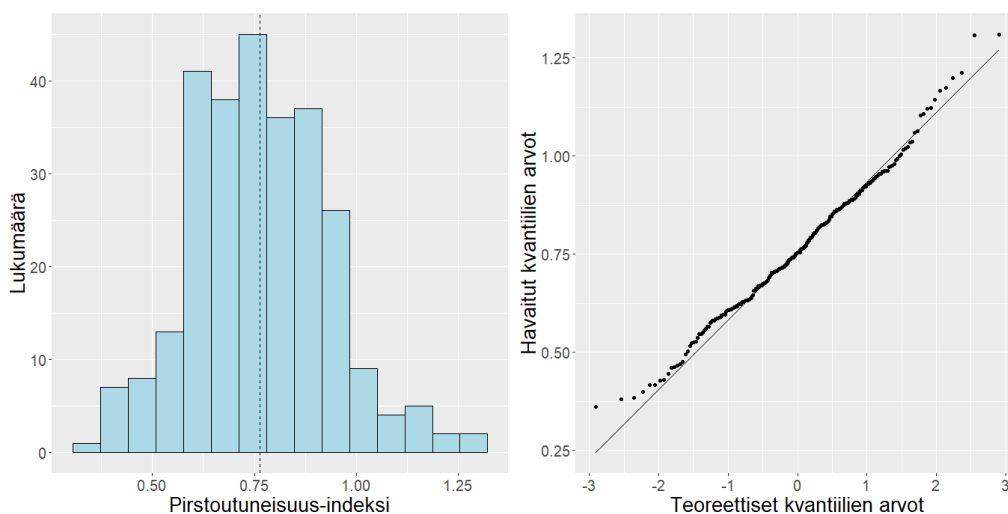
Taulukon 7 perusteella huomataan, että missä tahansa tilassa vietetään keskimäärin pidempään vapaapäivinä kuin työpäivinä. Toisaalta keskimääräisten tilasiirtymämatriisien perusteella vapaapäivinä siirrytään epätodennäköisemmin kevyestä aktiivisuudesta reippaaseen aktiivisuuteen, mutta kun siirrytään pysytään tilassa keskimäärin kauemmin. Miehillä paikallaanolojaksot ovat keskilukujen (keskiarvo ja mediaani) perusteella pidempiä kuin naisilla. Miesten ryhmässä myös kevyet ja reippaat aktiivisuusjaksot ovat keskimäärin pidempiä kuin naisilla. Miehillä on selvästi naisia suurempi havaintojen välinen vaihtelu ja suuria poikkeavia havaintoja.

	Kaikki	Miehet ($N = 47$)	Naiset ($N = 233$)	Työ-/vapaapäivä
Paikallaanolojakson kesto				
Keskiarvo (hajonta)	11.8 (15.0)	17.2 (26.1)	10.7 (11.4)	12.1 (18.0)/12.2 (14.1)
Mediaani [min.,maks.]	9.5 [3.8,175.0]	10.8 [5.5,175.0]	9.1 [3.8,161.9]	9.2 [3.3,222.3]/9.3 [4.0,150.4]
Kevyen aktiivisuusjakson kesto				
Keskiarvo (hajonta)	3.8 (5.6)	5.5 (13.2)	3.5 (1.6)	3.7 (7.5)/4.2 (4.9)
Mediaani [min.,maks.]	3.1 [1.9,87.6]	2.8 [2.0,87.6]	3.2 [1.9, 17.3]	2.7 [1.5,120.2]/3.5 [1.6,70.6]
Reippaan aktiivisuusjakson kesto				
Keskiarvo (hajonta)	2.3 (1.1)	2.4 (1.8)	2.3 (0.9)	2.1 (0.8)/2.7 (2.0)
Mediaani [min.,maks.]	2.1 [1.0,11.0]	2.0 [1.1,11.0]	2.1 [1.0,8.6]	1.9 [1.0,6.8]/2.2 [1.0,18.6]
Pirstoutuneisuus-indeksi				
Keskiarvo (hajonta)	0.76 (0.17)	0.82 (0.18)	0.75 (0.16)	0.84 (0.18)/0.72 (0.21)
Mediaani [min.,maks.]	0.75 [0.36,1.31]	0.85 [0.38,1.31]	0.74 [0.36,1.31]	0.85 [0.37,1.40]/0.69 [0.17,1.36]
Puuttuvia arvoja	6	3	3	19/10

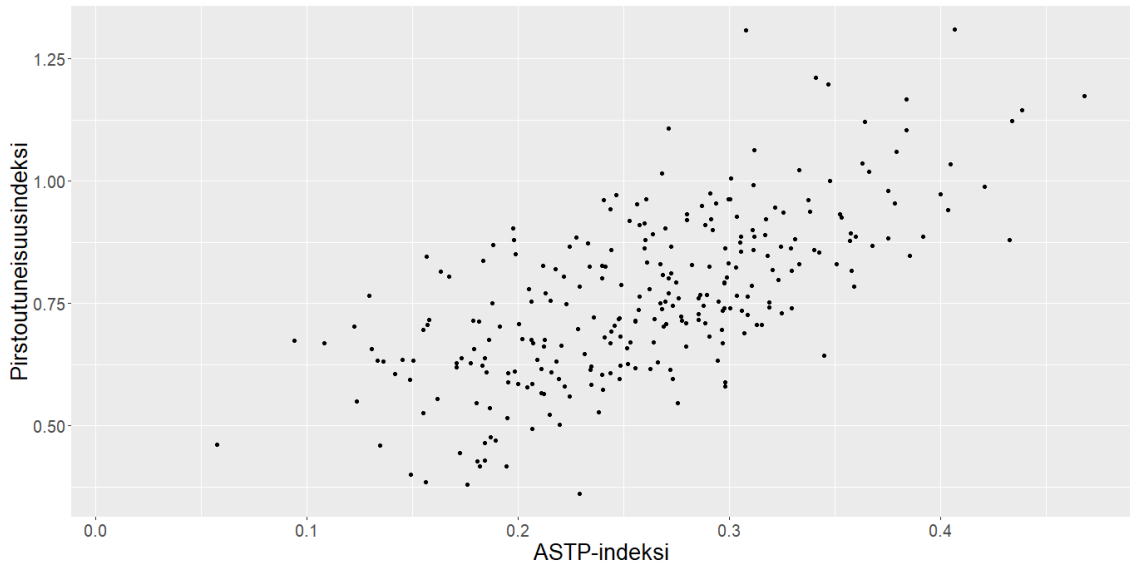
Taulukko 7: Kolmitilaisen mallin tärkeimpiä tuloksia. Puuttuviksi arvoiksi luetaan, jos yksikin alaspäin estimoitu siirtymätodennäköisyys puuttuu. Kestojen yksikkö on minuutti.

Keskiarvojen vertailu on epävarmaa, mikä johtuu suuresta hajonnasta ja ryhmien välisestä kokoerosta samoin kun kaksitilaisen mallin kohdalla. Reippaan aktiivisuusjakson keskimääräinen kesto on noin 2-3 minuuttia riippumatta sukupuolesta ja työ- ja vapaapäivistä.

Tarkastellaan yksilöiden aktiivisuuden pirstoutumista hyödyntäen luvussa 3 esiteltyä pirstoutuneisuuden mitta (6). Kuva 13 esittää pirstoutumis-indeksin jakauman ja kvantiilikuvion. Jakauman keskiarvo on 0.76. Kuvan perusteella jakauma näyttää likimain normaaliselta. Kuva 14 esittää kaksitilaisen mallin ASTP-indeksin ja pirstoutuneisuusindeksin välistä sirontakuviota. Muuttujien välinen Pearsonin korrelaatio on 0.67, mikä kertoo vahvasta positiivisesta korrelaatiosta indeksien välillä. Mitä enemmän aktiivisuus on pirstoutunut kaksitilaisessa mallissa, sitä enemmän se on pirstoutunut kolmitilaisessa mallissa kolmen eri alaspäin tapahtuvan siirtymätodennäköisyyden suhteen.



Kuva 13: Pirstoutuneisuusindeksin frekvenssijakauma ja kvantiilikuvio. Jakauman keskiarvo (katkoviiva) on 0.76.

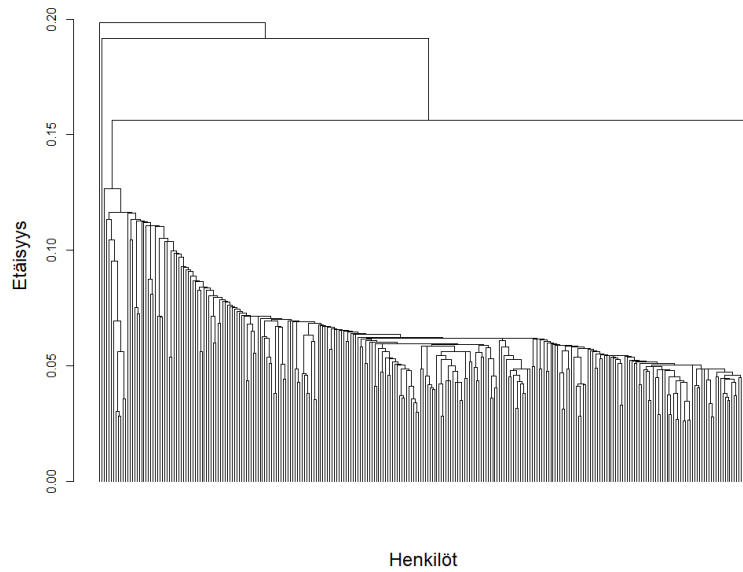


Kuva 14: ASTP-indeksin ja pirstoutuneisuusindeksin välinen sirontakuvi. Muuttujien välinen Pearsonin korrelaatio on 0.67, mikä kertoo vahvasta positiivisesta korrelaatiosta.

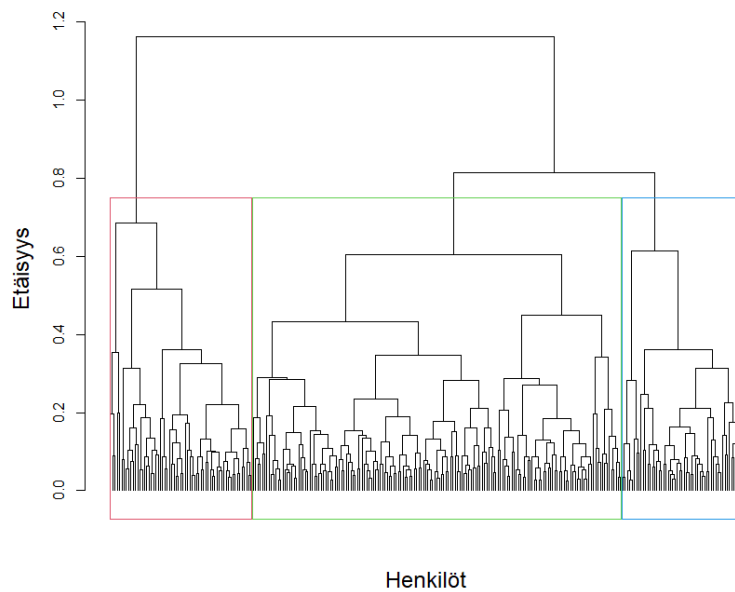
4.3.1 Kolmitilaisen mallin klusterit

Muodostetaan luvun 4.1.1 tapaan jokaisella yksilöllä oma 3×3 -tilasiirtymämatriisi ja asetetaan kaikki yhdeksän tilasiirtymätodennäköisyyden estimaattia peräkkäin riveiksi, yksilöiden tunnistenumeron mukaiseen järjestykseen. Tällä tavalla muodostuu yhdistetty kaikkien yksilöiden tilasiirtymien sisältämä 280×9 -matriisi. Aloitetaan ryhmien muodostaminen soveltamalla luvussa 3 esiteltyjä kokoavan menetelmän tekniikoita. Sovelletaan aluksi lähimmän naapurin menetelmää. Kuvassa 15 nähdään tilanteesta muodostettu klusterijakoa kuvaava dendrogrammi. Kuva on hyvin samanlainen kun kaksitilaisen mallin tapauksessa. Nytkään ei tarkastella lähimmän naapurin menetelmän tuottamaa klusterijakoa tämän enempää. Sen sijaan kaukaisimman naapurin menetelmä tuottaa selvästi kuvan 16 perusteella kolme klusteria, joista kaksi ovat lähes samankokoisia ja yksi selvästi suurempi. Ensimmäinen klusteri on näistä kaikkein erilaisin, koska se yhdistyy vasta viimeisenä kahteen muuhun klusteriin. Sama erilaisuus nähdään naapurikeskiarvon menetelmän dendrogrammistä kuvasta 17. Myös kaksi klusteria ovat lähes samankokoisia, mutta hieman pienempiä kuin kaukaisimman naapurin vastaavat klusterit.

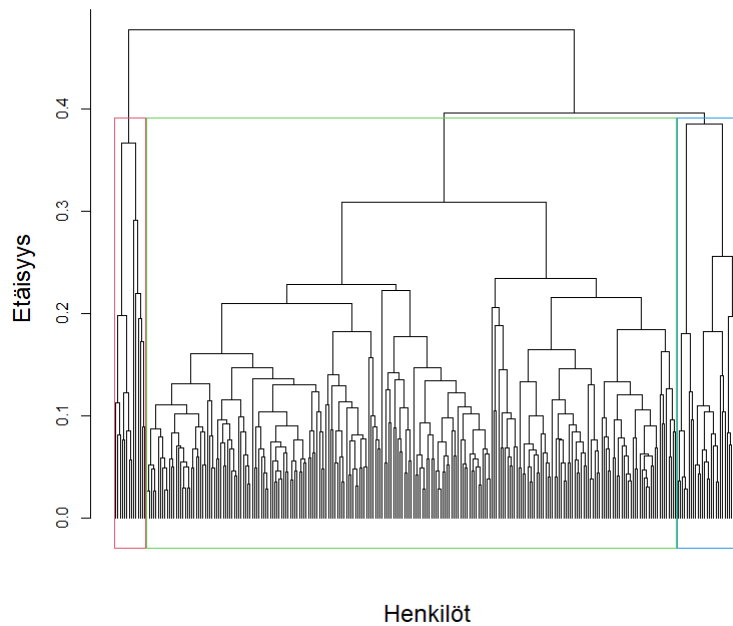
Kaukaisimman naapurin klusterijako vaikuttaa dendrogrammin perusteella mielekkäimmältä valinnalta, joten tarkastellaan sitä samalla tavalla kuin edellä kaksitilaiselle mallille. Kyynärpäämenetelmä (kuva 18) suosittaa kuvasta katsottuna kolmea tai neljää klusteria.



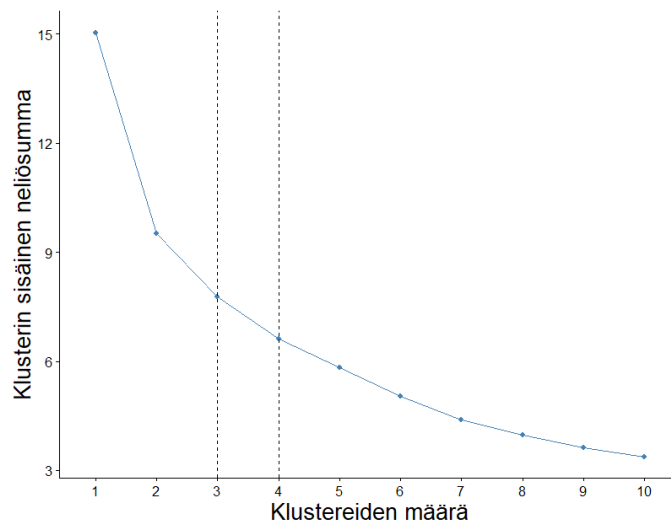
Kuva 15: Lähimmän naapurin menetelmän tuottama dendogrammi kolmitilaiselle mallille. Kuvasta on vaikea nähdä ryhmärakennetta. Yksi klusteri on kooltaan lähes koko aineiston kokoinen järkevän pienillä klusterimäärän k valinnoilla.



Kuva 16: Kaukaisimman naapurin menetelmän tuottama dendogrammi kolmitilaiselle mallille. Kuvasta on eroteltu eri värien avulla kolme klusteria (3A, 3B, 3C). Klustereiden koot ovat 63, 163 ja 54. Ensimmäinen klusteri on kaikkein erilaisin.



Kuva 17: Naapurikeskiarvon menetelmän tuottama dendogrammi kolmitilaiselle mallille. Kuvasta on eroteltu eri värien avulla kolme klusteria. Ensimmäinen klusteri on kaikkein erilaisin.



Kuva 18: Klustereiden sisäinen neliösumma klusterimäärän k funktiona. Valinta $k = 3$ tai $k = 4$ vaikuttaisi järkevältä. Kuva antaa tukea dendrogrammien perusteella tehdyille valinnalle $k = 3$.

Jaetaan henkilöt kuvan 16 perusteella kolmeen klusteriin (koot 63, 163 ja 54). Tutkitaan aluksi klustereiden keskimääräisiä tilasiirtymätodennäköisyyksiä. Tämän

avulla saadaan karkea käsitys ryhmien välisistä eroista:

$$\hat{P}_{3,(3A)} = \begin{bmatrix} 0.89 & 0.10 & 0.01 \\ 0.33 & 0.63 & 0.04 \\ 0.19 & 0.47 & 0.34 \end{bmatrix}, \quad \hat{P}_{3,(3B)} = \begin{bmatrix} 0.89 & 0.10 & 0.01 \\ 0.28 & 0.67 & 0.05 \\ 0.14 & 0.26 & 0.60 \end{bmatrix}$$

$$\text{ja } \hat{P}_{3,(3C)} = \begin{bmatrix} 0.86 & 0.13 & 0.01 \\ 0.18 & 0.79 & 0.03 \\ 0.13 & 0.42 & 0.45 \end{bmatrix}.$$

Kaksi ensimmäistä klusteria ovat keskiarvon mielessä samanlaisia ensimmäisen ja toisen rivin siirtymätodennäköisyyksissä. Kolmas klusteri eroaa jo selvästi edellisistä, ja klusteriin kuuluvat pysyvät todennäköisemmin kevyen aktiivisuuden tilassa. Toisen klusterin henkilöt pysyvät keskimäärin todennäköisimmin reippaassa aktiivisuuden tilassa. Kolmanteen klusteriin päätyi kuusi yksilöä, joilta ei löytynyt lainkaan siirtymiä reippaaseen tilaan eikä näin ollen sieltä poiskaan. Nämä kaikki kuusi yksilöä kuuluvat kaksitilaisen mallin toiseen klusteriin. Kaikkia kuutta yksilöä yhdistää erittäin korkea todennäköisyys (96–99 %) pysyä paikallaan. Toisen ja kolmannen klusterin yksilöiltä löytyy kaikilta siirtymiä reippaaseen aktiivisuuteen. Klusterikohtaiset aktiivisuusluokkien suhteelliset osuudet on esitetty liitteen B kuvissa B5–B7. Klusterin 3A yksilöt viettävät keskimäärin 75 % hereilläoloajastaan paikallaan, 23 % kevyessä liikkeessä ja 2 % reippaassa liikkeessä. Klusterin 3B yksilöt viettävät 71 % paikallaan, 24 % kevyessä liikkeessä ja 5 % reippaassa liikkeessä. Klusterissa 3C taas vietetään selvästi vähemmän aikaa paikallaan (59 %) ja toisaalta enemmän aikaa kevyessä liikkeessä (37 %).

Taulukkoon 8 on koottu klusterikohtaisesti terveys- ja toimintakykytietoja eri tunnuslukujen suhteen. Klusterissa 3A on suhteellisesti enemmän miehiä (24 %) kuin naisia. Kaikki klusteri 3A:n terveys- ja toimintakykytiedot ovat keskimäärin heikommat kuin kahden muun klusterin. Klusterin 3B yksilöt näyttäisivät olevan keskimäärin hieman paremmassa kunnossa kuin klusterin 3C yksilöt. Tarkempia eroja tutkitaan seuraavassa alaluvussa lineaarisen regression yhteydessä. Klusterikohtaiset aktiivisuusluokkien suhteellisia osuuksia esittävät kuvat on esitetty tutkielman liitessä B.

	Klusteri 3A (<i>N</i> = 63)	Klusteri 3B (<i>N</i> = 163)	Klusteri 3C (<i>N</i> = 54)
Sukupuoli			
Miehet Lkm. (%)	15 (24 %)	25 (15 %)	7 (13 %)
Naiset Lkm. (%)	48 (76 %)	138 (85 %)	47 (87 %)
Ammattiasema			
Asiantuntijat Lkm. (%)	40 (63 %)	126 (77 %)	30 (56 %)
Palvelu- ja fyysinen työ Lkm. (%)	23 (37 %)	37 (23 %)	24 (44 %)
Painoindeksi (<i>kg/m²</i>)			
Keskiarvo (hajonta)	28.7 (5.7)	25.5 (3.8)	25.6 (3.9)
Mediaani [min.,maks.]	28.1 [17.0,44.3]	25.4 [17.0,39.1]	25.2 [17.2,37.5]
Puuttuvia arvoja	1	1	0
Vyötärönympäryys (<i>cm</i>)			
Keskiarvo (hajonta)	98.4 (15.1)	89.4 (11.5)	90.4 (11.5)
Mediaani [min.,maks.]	97.2 [67.2,139.8]	88.5 [68.2,119.3]	89.0 [65.1,123.3]
Hapenottokyky (V02 , <i>ml/kg/min</i>)			
Keskiarvo (hajonta)	26.7 (5.1)	29.6 (5.5)	27.9 (4.2)
Mediaani [min.,maks.]	25.8 [14.6,38.8]	28.8 [17.1,48.5]	27.6 [21.2,39.5]
Puuttuvia arvoja	16	20	9
Maksimaalinen kävelynopeus (<i>m/s</i>)			
Keskiarvo (hajonta)	1.66 (0.33)	1.67 (0.29)	1.69 (0.32)
Mediaani [min.,maks.]	1.60 [1.06,2.86]	1.60 [0.89,2.67]	1.62 [1.14,2.67]
Kymmeneen tuoliltanousuun kulunut aika (<i>s</i>)			
Keskiarvo (hajonta)	22.7 (5.5)	20.2 (4.1)	21.5 (4.7)
Mediaani [min.,maks.]	21.2 [15.6,42.0]	19.6 [12.8,52.7]	20.8 [15.7,44.7]
Puuttuvia arvoja	0	3	0
Tuki- ja liikuntaelimestön kroonisten sairauksien lukumäärä			
Ei lainkaan (%)	19 (30 %)	63 (39 %)	21 (39 %)
Yksi (%)	27 (43 %)	60 (37 %)	21 (39 %)
Kaksi tai kolme (%)	14 (22 %)	35 (21 %)	8 (15 %)
Puuttuvia arvoja (%)	3 (5 %)	5 (3 %)	4 (7 %)

Taulukko 8: Klusterikohtaisesti terveys- ja toimintakykytietoja. Klusterointi suoritettiin kolmitilaiselle mallille kaukaisimman naapurin menetelmällä. Asiantuntija-asemassa olevat käsittävät taulukon tapauksessa erityisasiantuntijat ja asiantuntijat/toimistotyöntekijät.

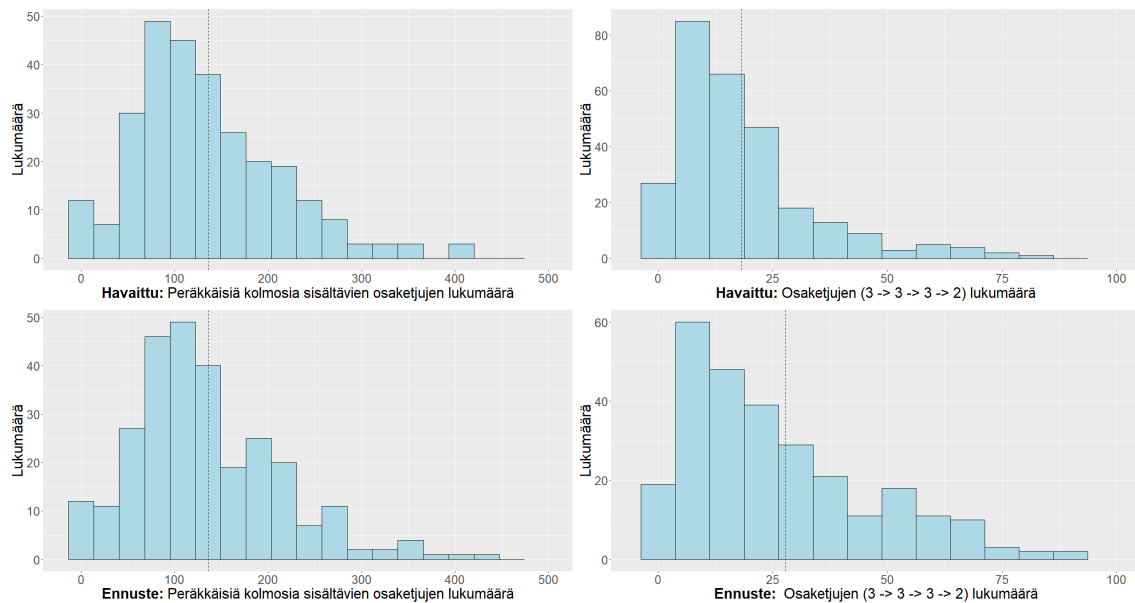
4.3.2 Mallin oletusten ja hyvyyden arviointi

Arvioidaan kolmitilaisen mallin oletuksia ja hyvyyttä luvun 3.1.3 keinojen avulla. Lasketaan ensin kaikille yksilöille reippaiden aktiivisuusjaksojen lukumäärä eli peräkkäisiä kolmosia (tila 3) sisältävien osaketjujen lukumäärä ja verrataan tätä jakaumaa vastaavan aikahomogeenisen Markovin ketjun ennustejakaumaan. Ennustejakauma on muodostettu vastaavan aikahomogeenisen Markov-ketjun estimoitujen tilasiirtymätodennäköisyyksien avulla.

Toiseksi lasketaan jokaiselle yksilölle osaketjujen ($3 \rightarrow 3 \rightarrow 3 \rightarrow 2$) lukumäärä eli tarkastellaan kaksitilaisen malliin verrattuna yhtä askelta pidempää osaketjua.

Osaketjun ($3 \rightarrow 3 \rightarrow 3 \rightarrow 2$) siirtymä tilasta kolme tilaan kaksi ei mallin oletusten mukaan riipu siirtymää edeltävistä tiloista.

Kuva 19 esittää havainto- ja ennustejakaumia edellä esitetuille testisuureille. Aktiivisuusjaksojen lukumäärä oli havaitussa aineistossa keskimäärin 136 eli täysin sama kuin ennustetussa tapauksessa. Ennuste siis löysi reipasta aktiivisuutta kuvaajan jaksottuneisuuden profiilin tämän tunnusluvun näkökulmasta. Osaketjujen ($3 \rightarrow 3 \rightarrow 3 \rightarrow 2$) lukumäärä oli havaitussa aineistossa keskimäärin 18, kun taas ennusteen tapauksessa 28. Ennuste siis yliarvioi osaketjujen lukumäärää kymmenellä. Jakau- mat ovat havaitun ja ennusteen mielessä muodoiltaan hyvin samanlaisia molemmissa tapauksissa. Näiden tulosten perusteella voidaan sanoa, että havaitut ketjut kuvaavat melko tarkasti reippaan aktiivisuuden jaksottuneisuutta eli ovat tässä suhteessa lähellä aikahomogeenista Markovin ketjua ainakin keskimääräisesti. Toki pitää huo- mioida reippaan aktiivisuuden vähäinen suhteellinen osuus hereilläoloaikana.



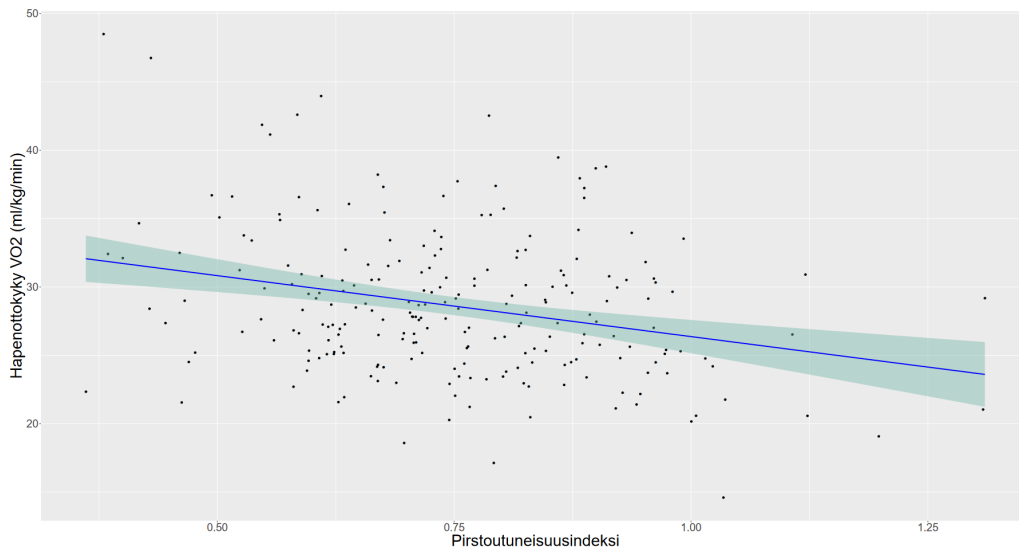
Kuva 19: Havainto- ja ennustejakaumat kolmitilaiselle mallille. Vasemmalla ylhäällä havaittu reippaiden aktiivisuusjaksojen lukumäärän jakauma (keskiarvo 136) ja oikealla osaketjujen ($3 \rightarrow 3 \rightarrow 3 \rightarrow 2$) havaittu lukumäärän jakauma (keskiarvo 18). Vasemmalla alhaalla ennustettu reippaiden aktiivisuusjaksojen lukumäärän jakauma (keskiarvo 136) ja oikealla osaketjujen ($3 \rightarrow 3 \rightarrow 3 \rightarrow 2$) ennustettu lukumäärän jakauma (keskiarvo 28).

4.3.3 Aktiivisuuden jaksottaisuus ja terveysvasteet

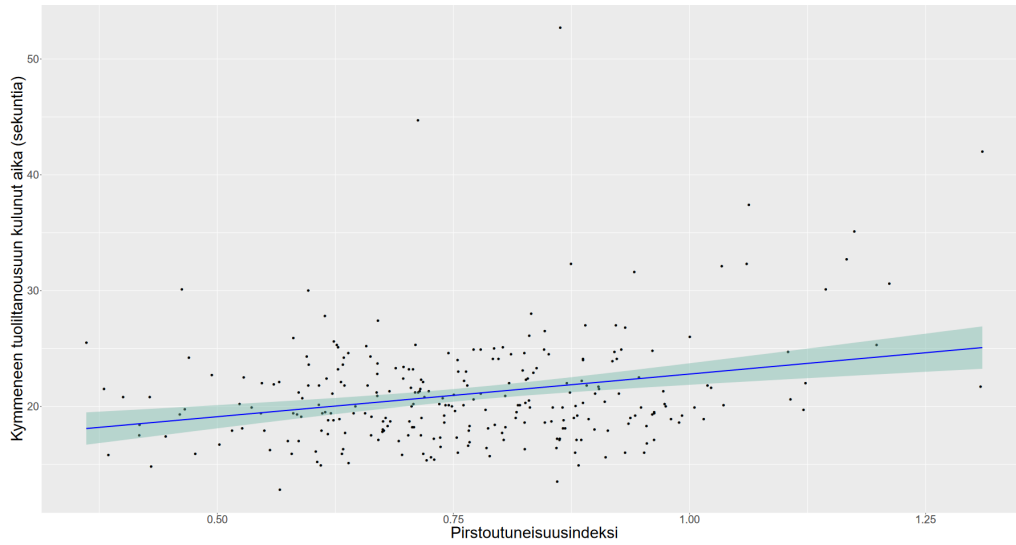
Tässä luvussa esitellään kolmitilaiseen malliin liittyvät analyysimallien tulokset. Esitellään aluksi pirstoutuneisuusindeksin ja terveysvasteiden välisiä kuvailevia tuloksia. Kuvailevat tulokset poikkeavat taulukon 9 tuloksista, koska taulukossa on mal- leihin lisätty vakioivia selittäjiä.

Pirstoutuneisuusindeksi ja terveysvasteet. Kuvat 20 ja 21 esittävät pirstoutuneisuusindeksin ja hapenottokyvyn sekä pirstoutuneisuusindeksin ja kymme- neen tuoliltanousuun kuluneen ajan välisiä yhteyksiä. Molempiin kuviin on piirretty

pienimmän neliösumman (PNS) suora. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Vastaavat tulokset pirstoutuneisuusindeksin ja maksimaalisen kävelynopeuden välillä on esitetty liitteessä B. Yksilöltä ei voida määrittää pirstoutuneisuusindeksiä, jos yksilöllä ei ole lainkaan siirtymiä reippaan aktiivisuuden tilaan. Kuudelta yksilöltä puuttuu tieto pirstoutuneisuusindeksistä. Selvästi huomataan, että hapenottokyky on sitä parempi, mitä pienempi on pirstoutuneisuusindeksi. PNS-suoran kulmakerroin on -8.9 eli keskimäärin 0.1 yksikön lisäys pirstoutuneisuusindeksissä vaikuttaa hapenottokykyyn -0.89 yksikköä. Toisaalta kymmeneen tuoliltanousuun kuluu sitä vähemmän aikaa mitä pienempi pirstoutuneisuusindeksi on. Tällöin 0.1 yksikön lisäys pirstoutuneisuusindeksissä vaikuttaa kymmeneen tuoliltanousuun kuluneeseen aikaan noin puoli sekuntia. Molemmissa tapauksissa luottamusväli ei sisällä arvoa nolla. Kolmas terveystieteiden kuvaa maksimaalista kävelynopeutta (m/s). Pirstoutuneisuusindeksin ja maksimaalisen kävelynopeuden välillä ei ole yhteyttä. PNS-suoran kulmakerroin on tässä tapauksessa -0.02 ja 95 %:n luottamusväli $[-0.23, 0.20]$ (liite C: kuva C2).



Kuva 20: Pirstoutuneisuusindeksin ja hapenottokyvyn välinen yhteys. Pistejoukkoon sovitettu pienimmän neliösumman suora. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Pienimmän neliösumman suoran kulmakerroin on -8.9 ja vastaava 95 %:n luottamusväli $[-13.0, -4.9]$. Tieto pirstoutuneisuusindeksistä puuttuu kuudelta yksilöltä. Tieto hapenottokyvystä puuttuu 45 yksilöltä.



Kuva 21: Pirstoutuneisuusindeksin ja kymmeneen tuoliltanousuun kuluneen ajan välinen yhteys. Pistejoukkoon sovitettu pienimmän neliösumman suora. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Pienimmän neliösumman suoran kulmakerroin on 7.4 ja vastaava 95 %:n luottamusväli [4.2,10.6]. Tieto pirstoutuneisuusindeksistä puuttuu kuudelta yksilöltä. Kymmeneen tuoliltanousuun kulunut aika puuttuu kolmelta yksilöltä.

Taulukkoon 9 on koottu lineaarisen regression tulokset kaikille kolmelle terveysvasteelle, kun selittävänä muuttujana on joko pirstoutuneisuusindeksi tai klusteri. Pirstoutuneisuusindeksiä käsitellään jatkuvana selittäjänä ja klusteri on luonnollisesti kategorinen selittäjä

Vastemuuttuja	Hapenottokyky (<i>ml/kg/min</i>)	Maksimaalinen kävelynopeus (<i>m/s</i>)	Kymmeneen tuolilta- nousuun kulunut aika (s)
Jatkuva pirstoutuneisuus-			
indeksi (per 0.1 yksikköä)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	-0.74 [-1.06,-0.43]	-0.00 [-0.02,0.27]	0.50 [0.16,0.83]
R^2/R_{adj}^2	0.53/0.51	0.06/0.04	0.11/0.09
Klusteri			
Klusteri 3A ($N = 63$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	referenssiluokka	referenssiluokka	referenssiluokka
Klusteri 3B ($N = 163$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	1.86 [0.54,3.18]	-0.01 [-0.11,0.08]	-1.42 [-2.77,-0.66]
Klusteri 3C ($N = 54$)			
$\hat{\beta}$ ja $\hat{\beta}$:n 95 %:n lv.	0.41 [-1.22,2.05]	0.02 [-0.10,0.13]	-0.01 [-1.69,1.66]
R^2/R_{adj}^2	0.51/0.49	0.06/0.04	0.10/0.07

Taulukko 9: Lineaarisen regression tulokset. Vastemuuttujia ovat hapenottokyky, maksimaalinen kävelynopeus ja kymmeneen tuoliltanousuun kulunut aika. Pirstoutuneisuusindeksiä käsitellään jatkuvana ja klusteria kategorisena selittäjänä. Mallit vakioitiin demografisten tietojen (ikä, sukupuoli ja työn kuormittavuus), painoindeksin ja kroonisten sairauksien historian mukaan.

Tulosten perusteella 0.1 yksikön lisäys pirstoutuneisuusindeksissä vaikuttaa hapenottokykyyn -0.74 (ml/kg/min), maksimaaliseen kävelynopeuteen -0.00 (m/s) ja kymmeneen tuoliltanousuun kuluneeseen aikaan 0.50 (s). Parametrien 95 %:n luottamusvälit eivät sisällä nollaa kun vastemuuttujana on hapenottokyky tai kymmeneen tuoliltanousuun kulunut aika. Hapenottokyky mallin vasteena tuottaa selvästi parhaimman selityksasteen R^2 (0.53). Tällöin pirstoutuneisuusindeksi ja muut valitut selittäjät muuttujat pystyvät selittämään 53 % hapenottokyvyn vaihtelusta.

Klusteri 3A toimii vertailuryhmänä. Klusterilla 3B on selvästi parempi hapenottokyky kuin vertailuryhmällä. Tulos vahvistaa klusterikohtaisten keskimääräisten tilasiirtymätodennäköisyyksien ja taulukon 8 tulkintoja. Maksimaalinen kävelynopeus puolestaan ei kykene erottelemaan klustereita toisistaan. Huomataan myös, että klusteri 3B saavuttaa myös vertailuryhmään nähden parempia tuoliltanousuun liittyviä tuloksia. Näyttää kaksitilaisen mallin tapaan siltä, että klusterointi pystyi tunnistamaan jonkinlaisen riskiryhmän (klusteri 3A) eli ryhmän, jolla on huonoimmat terveys- ja toimintakykytiedot sekä eniten/vähiten paikallaanoloa/aktiivisuutta.

4.3.4 Klusterointien yhtäpitävyys

Tarkastellaan lopuksi kaksi- ja kolmitilaisen mallin tuottamien klusterien yhtäpitävyyttä. Tässä yhteydessä yhtäpitävyyttä voidaan tutkia vertaamalla, ovatko samat yksilöt samoissa klustereissa. Ensin luokitellaan jokainen yksilö kaksitilaisen mallin mukaisesti luokkiin 2A-2D. Tämän jälkeen luokitellaan yksilöt kolmitilaisen mallin antamiin kolmeen luokkaan 3A-3C. Kukin luokka määräytyy vastaavan mallin dendrogrammin mukaisesti vasemmalta oikealle. Muodostetaan 4×3 -matriisi, jossa jokainen alkio kertoo parittaisten klustereihin kuuluvien henkilöiden lukumäärän. Kolmitilainen malli on yhteensopiva kaksitilaisen mallin kanssa, jos kunkin rivin yksi alkio sisältää suurimman osan sen rivin kokonaismäärästä. Tällöin samat henkilöt kuuluvat samoihin klustereihin.

Yhteensopivuutta kuvaava matriisi \mathbf{C} voidaan kirjoittaa muodossa

$$\mathbf{C}_{4 \times 3} = \begin{bmatrix} n_{(2A,3A)} & n_{(2A,3B)} & n_{(2A,3C)} \\ n_{(2B,3A)} & n_{(2B,3B)} & n_{(2B,3C)} \\ \vdots & \vdots & \vdots \\ n_{(2D,3A)} & n_{(2D,3B)} & n_{(2D,3C)} \end{bmatrix} = \begin{bmatrix} 25 & 23 & 0 \\ 0 & 2 & 9 \\ 1 & 25 & 30 \\ 37 & 113 & 15 \end{bmatrix}.$$

Pienennetään kaksitilaisen mallin klustereiden määrää yhdellä niin, että kaksitilaisen mallin kolmas ja neljäs klusteri yhdistyvät eniten samanlaisina. Tällöin matriisi \mathbf{C} on:

$$\mathbf{C}_{3 \times 3} = \begin{bmatrix} 25 & 23 & 0 \\ 0 & 2 & 9 \\ 38 & 138 & 45 \end{bmatrix}.$$

Ensimmäinen havainto on, että klusteri 2A sisältää yksilöitä klustereista 3A ja 3B lähes samassa suhteessa. Toisaalta 2B klusteri sisältää yksilöitä klustereista 3B ja 3C, niin että 82 % on lähtöisin klusterista 3C. Viimeinen eli tässä tapauksessa lukumäärältään suurimman klusterin 2C ($N = 221$) yksilöt jakautuvat kaikkiin kolmeen klusteriin prosenttiosuuksilla 17 %, 63 % ja 20 %.

Kolmitilaisen mallin kolmas tila selvästi siis jakaa yksilöitä eri klustereihin kuin kaksitilainen malli. Nähdään, että kolmas tila tuo klusterointiin aidosti jotain lisää, mutta on vaikea sanoa, onko tästä tiedosta jotain hyötyä. Matriisien kahta nollasolua voidaan pitää merkinä siitä, että klusterit eivät sekoitu kuitenkaan täydellisesti keskenään.

5 Pohdinta

Tässä tutkielmassa oli tarkoitus selvittää paikallaanolon ja liikkumisen jaksottuneisuutta eläkkeelle siirtyvien julkisen sektorin työntekijöiden joukossa. Aineistona käytettiin FIREA-tutkimuksen liikemittaus- ja terveys/toimintakykytietoja. Liikemittarin käyttö tieteellisessä ja kliinisessä tutkimuksessa tarjoaa uusia mahdollisuuksia ymmärtää fyysisen aktiivisuuden jakautumista pienempiin osiin ja jaksottuneisuuden yhteyttä erilaisiin terveyttä- ja toimintakykyä kuvaaviin indikaattoreihin. Tässä tutkielmassa jaksottuneisuutta tarkasteltiin erikseen kaksi- ja kolmitilaisen tilasiirtymämallin avulla.

Tilasiirtymämallit määriteltiin diskreetin aikahomogeenisen Markovin ketjun perusteella. Mallissa seuraavan tilan eli aktiivisuuden intensiteettiä kuvaavan tason todennäköisyys riippuu vain edellisestä tilasta ja todennäköisyyteen ei vaikuta tarkasteltava ajanhetki. Tutkielmassa määriteltiin jokaiselle yksilölle erikseen omat tilasiirtymätodennäköisyyksien estimaatit. Näitä estimaatteja käytettiin jaksottuneisuutta kuvaavien indeksien ja paikallaanolo/aktiivisuusjaksojen kestojen kuvaamiseen sekä hierarkkisen klusteroinnin yhteydessä etäisyysmatriisin muodostamiseen. Hierarkkisessa klusteroinnissa päädyttiin käyttämään kokoavan menetelmän tekniikoita. Yksilöitä pyrittiin ryhmittelemään keskenään samankaltaisiin, mutta muista erottuviin ryhmiin. Terveys- ja toimintakykyä kuvaavia indikaattoreita mallinnettiin lineaarisen regression keinoin niin, että selittävinä muuttujina olivat vaihtoehtoisesti jokin jaksottuneisuutta kuvaava indeksi tai klusteri.

Tutkielmassa keskityttiin erityisesti neljään päätavoitteeseen. Tärkein tavoite oli tilasiirtymämatriisien soveltaminen aktiivisuuden jaksottumisen tutkimiseen käyttämällä sekä kaksi- että kolmitilaisia malleja. Toisena tavoiteena tutkittiin, löydetäänkö kaksi- ja kolmitilaisilla malleilla eroja aktiivisuuden jaksottumisessa aktiivisuuden suhteen kahden hyvin erilaisen päivän, työ- ja vapaapäivän välillä. Kolmas tavoite oli soveltaa erilaisia klusterointitekniikoita erilaisten aktiivisuuden jaksottumisprofiilien löytämiseksi ja vertailla mallien klustereiden yhtäpitävyyttä. Neljäs tavoite oli tutkia, onko kaksitilaisen mallin ASTP-indeksi tai sen yleistyksen kolmitilaisessa mallissa yhteydessä erilaisiin terveyden- ja toimintakyvyn indikaattoreihin.

Kaksitilaisen mallin yhteydessä aktiivisuuden jaksottuneisuutta kuvattiin todennäköisyydellä (ASTP) siirtyä aktiivisesta liikkeestä istumisen kaltaiseen paikallaanoloon. Vastaavasti paikallaanolon jaksottuneisuutta kuvattiin todennäköisyydellä (SATP) siirtyä paikallaanolosta aktiiviseen liikkeeseen. Aktiivisuus oli työpäivinä keskimäärin enemmän jaksottunut kuin vapaapäivinä, ja aktiivisuusjaksot olivat pidempiä työpäivinä ja paikallaanolojaksot suunnilleen yhtä pitkiä kumpanakin päivänä. ASTP-indeksi oli likimain normaalisti jakautunut.

Kaksitilainen malli löysi kaukaisimman naapurin menetelmällä selvästi neljä erottuvaa klusteria. Klustereille löydettiin mielekkäät tulkinnat. Yksi klustereista koostui passiivisista paikallaanoliijoista, toinen poikkeuksellisista yksilöistä, joilla oli pitkät paikallaanolo- ja aktiivisuusjaksojen kestot sekä ryhmistä korkein hapenottokyky. Kaksi muuta klusteria olivat selvästi aktiivisempia: kolmas klusteri piti sisällään alati liikkuvia yksilöitä ja neljäs klusteri keskivertoliikkujia. Klustereiden sisällä aktiivisuusluokkien suhteelliset osuudet olivat lähellä toisiaan ja toisaalta erottuivat muista klustereista. Lähimmän naapurin menetelmä yhdisti suurimman osan ha-

vainnoista samaan klusteriin, koska lähes jokaisen vaiheen jälkeen seuraava havainto oli lähempänä yhdistettyä klusteria kuin yksittäisiä havaintoja. Naapurikeskiarvon menetelmän klusterijako muistutti dendrogrammin perusteella kaukaisimman naapurin menetelmän klusterijakoa.

Aktiivisuuden jaksottuneisuus näkyi myös aktiivisessa tilassa vietetyn ajan suhteellisissa osuuksissa. Klusterin 2A yksilöillä oli keskimäärin pienin todennäköisyys (64 %) jatkaa aktiivisessa tilassa minuutin jakson aikana, ja he viettivät myös suhteellisesti vähiten aikaa aktiivisessa tilassa (keskimäärin 19 % hereilläoloajasta). Toisaalta klusterin 2C yksilöillä oli keskimäärin pieni todennäköisyys lopettaa aktiivinen liike minuutin jakson aikana (82 %), ja he viettivät selvästi eniten aikaa aktiivisessa tilassa (keskimäärin 47 %). Klustereiden 2B ja 2C yksilöt viettivät keskimäärin yhtä suuren osan hereilläoloajastaan liikkeessä (29 %), mutta aktiivisuuden jaksottumista kuvaavat profiilit olivat selvästi erilaiset.

Lineaarisen regression tuloksena ASTP-indeksi jatkuvana muuttujana oli yhteydessä hapenottokykyyn sekä kymmeneen tuoliltanousuun kuluneeseen aikaan. Lineaarinen trendi oli havaittavissa suuresta yksilöiden välisestä hajonnasta huolimatta. Ainoastaan ASTP-indeksin ja maksimaaliseen kävelynopeuden välillä ei löydetty minkäänlaista yhteyttä. Kategorisen ASTP-indeksin tapauksessa vaikutus nähtiin vertailtaessa alatertiiliä ylätertiiliin. Näytti siltä, että ylätertiilin eli suurimpien ASTP-indeksien ryhmässä vaikutus näkyi myös jatkuvan ASTP-indeksin regressiosuoran kulmakertoimessa positiivisesti. Klusteri mallin selittäjänä löysi selvästi passiivisten paikallanolioiden riskiryhmän (2A). Klusterit 2B ja 2D olivat hapenottokyvyn perusteella paremmissa kunnossa olevia kuin referenssiluokka 2A. Maksimaalinen kävelynopeus ei kyennyt erottelamaan klustereita lainkaan toisistaan.

Kaksitilaisen mallin oletukset tuntuivat sopivan havaintoaineiston aktiivisuusketjujen kuvaamiseen. Malli valittiin lähinnä, koska oletukset olivat varsin yksinkertaisia tulkinnallisesti. Oletusten perusteella aktiivisuuden jaksottumisen profiilia on hyvin vaikea mallintaa täydellisesti näin yksinkertaisella mallilla, mutta toisaalta ilmiön kuvaamiseen yksinkertainen aktiivisuuden jaksottuneisuutta kuvaava todennäköisyys oli varsin hyvä. Hereilläolon aikaista aktiivisuuden jaksottuneisuutta olisi voinut mallintaa vaihtoehtoisella mallilla, mikä ottaisi huomioon vuorokaudenajan ja riippuvuuden useammasta kuin yhdestä aikaisemmasta tilasta, koska todennäköisesti aktiivinen toiminta herkästi jatkuu, kun on päässyt siinä alkuun. Vaihtoehtona voisi sovittaa epähomogeenisen Markov-mallin. Kaikesta huolimatta mallin sopivuuden tarkastelu osoitti, että homogeeninenkin malli kuvasi aktiivisuuden todellista jaksottuneisuutta kohtuullisen hyvin.

Kolmitilaista mallia tarkasteltiin uutena vaihtoehtona kaksitilaisen mallin rinnalle. Tässä mallissa aktiivisuuden jaksottuneisuutta kuvattiin nk. pirstoutuneisuusindeksillä, mikä oli kolmen alaspäin tapahtuneen siirtymätodennäköisyyden summa. Myös pirstoutuneisuusindeksi oli likimain normaalisti jakautunut.

Kolmitilainen malli löysi kaukaisimman naapurin menetelmällä kolme klusteria. Ensimmäiseen klusteriin luokiteltiin jonkilaiseen riskiryhmään kuuluvia passiivisia yksilöitä. Toinen klusteri oli kooltaan suurin, ja siihen klusterointi luokitteli keskimäärin terveys- ja toimintakykytietojen valossa hyväkuntoisia yksilöitä. Kolmas klusteri koostui lähes yhtä hyväkuntoisista yksilöistä kuin toinen klusteri, mutta tässä klusterissa paikallaanoloa oli suhteellisesti kolmesta klusterista vähiten ja kevyttä

aktiivisuutta eniten. Reippaan aktiivisuuden suhteen suuria eroja ei havaittu keskimääräisissä reippaan aktiivisuusluokan suhteellisissa osuuksissa. Klusteroinnissa olisi voinut vaihtoehtoisesti painottaa esimerkiksi pirstoutuneisuusindeksin jokaista komponenttia mielekkäästi valitulla painolla ja tutkia miten klusterointi muuttaa ryhmärakennetta ja kahden eri mallin klustereiden yhtäpitävyyttä. Nyt yhtäpitävyys kaksitilaisen mallin kanssa ei ollut mitenkään intuitiivista, koska yksilöt sekoittuivat osittain eri klustereihin. Toki tästä voidaan päätellä, että kolmas tila toi klusterointiin aidosti jotain lisää, mutta on vaikea sanoa onko tästä tiedosta jotain hyötyä. Yllättävintä oli, että kolmitilainen malli löysi kaksitilaiseen malliin verrattuna yhden klusterin vähemmän, vaikka siirtymätodennäköisyyksiä oli kolmitilaisessa mallissa enemmän.

Lineaarisen regression tuloksena pirstoutuneisuusindeksi jatkuvana muuttujana oli yhteydessä hapenottokykyyn sekä kymmeneen tuoliltanousuun kuluneeseen aikaan. Lineaarinen trendi oli havaittavissa suuresta yksilöiden välisestä hajonnasta huolimatta. Pirstoutuneisuusindeksin ja maksimaaliseen kävelynopeuden välillä ei löydetty minkäänlaista yhteyttä. Klusteri 3B erosi selvästi referenssiluokasta 3A hapenottokyvyn ja kymmeneen tuoliltanousuun kuluneen ajan suhteen. Maksimaalinen kävelynopeus ei kyennyt tässäkään tapauksessa erottelemaan klustereita lainkaan toisistaan, joten voidaan ajatella, ettei tämä terveyttä kuvaava mittari ollut optimaalinen valinta toimintakyvyn kuvaamiseen.

Kolmitilaista mallia koskevat oletukset osoittivat aikahomogeenisen Markov-mallin sopivan varsin hyvin reippaan aktiivisuuden jaksottuneisuuden kuvaamiseen. Ennustejakaumat olivat yllättävänkin lähellä havaittuja aineistosta laskettuja testisuureen jakaumia. Toki reippaan aktiivisuuden tilassa vietettiin selvästi vähemmän aikaa verrattuna kahdessa muussa tilassa vietettyyn aikaan, jolloin reippaan aktiivisuuden jaksottuneisuuden ennustaminen aikahomogeenisella Markov-mallilla voisi ajatella olevan tarkempaa. Testisuureen valinta oli täysin subjektiivista, jolloin tulosten tulkintaan on syytä suhtautua varauksella.

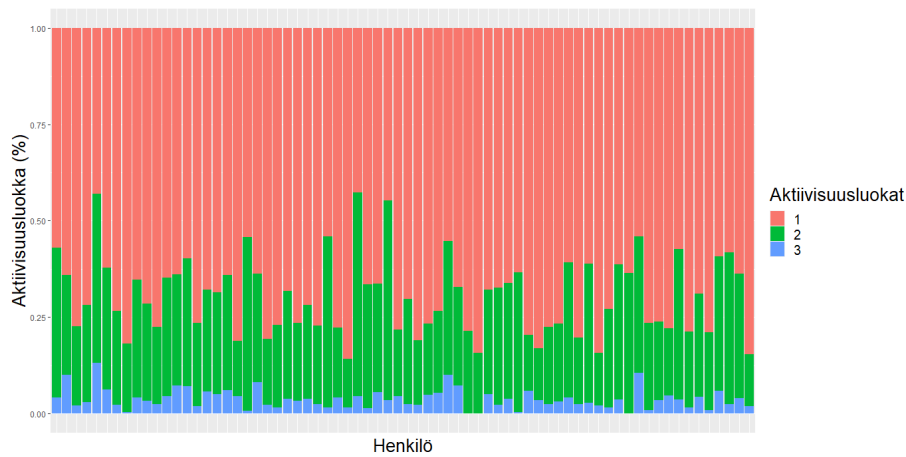
Kaksitilaisessa mallissa korostuu se, kuinka usein paikallaanoloa katkotaan ja kuinka pitkään liikettä ylläpidetään. Kolmitilainen malli taas ottaa huomioon yhden lisäulottuvuuden siinä mielessä, että katsotaan myös sitä, kuinka eri kuormitustason aktiivisuudet (kevyt ja reipas) jaksottuvat. Tällöin kyse ei enää ole vain paikallaanolon katkomisesta ja liikkeen ylläpitämisestä.

Jatkotutkimuksena olisi mahdollista tutkia aktiivisuuden jaksottaisuutta niin, että kullekin yksilölle otettaisiin analyyseja varten sama määrä työ- tai vapaapäiviä, esimerkiksi molempia kaksi. Tämän tutkielman aineiston tapauksessa tähän ei kiinnitetty huomiota vaan otettiin yksilön kaikki mittauspäivät huomioon vaikka päiviä oli hyvinkin eri määrä yksilöiden välillä. Tämä tietysti muuttaa analyysejä, koska kahden työ- ja vapaapäivän jälkeen ylimääräiset mittauspäivät pitäisi poistaa sopivasti valitulla tavalla, jolloin osa mittaushistorian tuomasta informaatiosta hukkuu. On tietenkin mahdollista myös osittaa analyysi erikseen työ- ja vapaapäiville, jolloin mitään tietoa ei heitetä hukkaan.

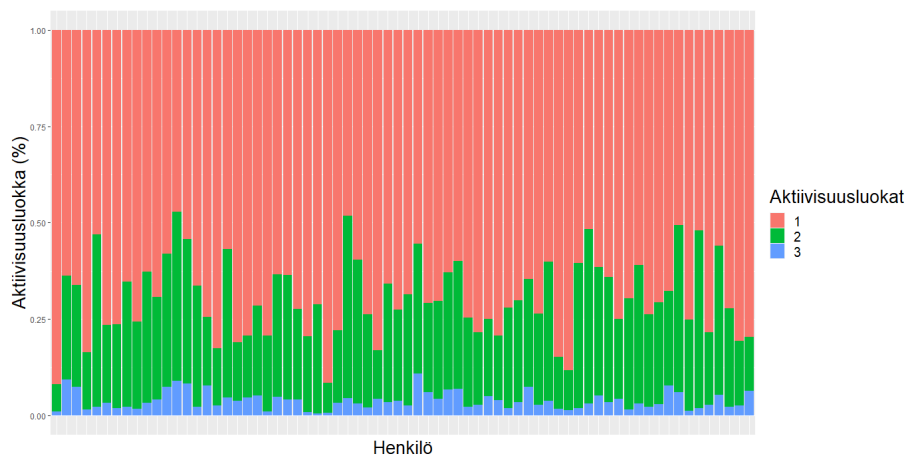
Viitteet

- [1] Shrack, J., Pei-Lun, K., Wanigatunga, A., Di, J., Simonsick, E., Spira, A., Ferrucci, L., Zipunnikov, V. (2019) *Active-to-Sedentary Behavior Transitions, Fatigability, and Physical Functioning in Older Adults*. *Journals of Gerontology: Medical Sciences*, 74(4), 560-567.
- [2] Stenholm, S., Pulakka, A., Leskinen, T., Pentti, J., Heinonen, O.J., Koster, A., Vahtera, J. (2020) *Daily Physical Activity Patterns and Their Association With Health-Related Physical Fitness Among Aging Workers— The Finnish Retirement and Aging Study*. *Journals of Gerontology: Medical Sciences*, 76(7), 1242–1250.
- [3] Finnish Retirement and Aging Study. URL <https://sites.utu.fi/firea/firea-hankkeesta/>. (luettu 9.2.2023).
- [4] Turun yliopisto, matematiikan ja tilastotieteen laitos: Holopainen, S. & Kilpi, J. (2021) *Paikallaanolon ja liikkumisen jaksottuneisuus: Mallinnusprojektin loppuraportti*.
- [5] R: *Accelerometer data processing with GGIR*. URL <https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html>.
- [6] C Ibe, Oliver (2013). *Markov Processes for Stochastic Modeling*. ISBN: 978-0124077959.
- [7] Grimmett, Geoffrey R. & David R. Stirzaker (2001). *Probability and Random Processes*. ISBN: 978-0198572220.
- [8] The Investopedia Team (2022). *R-Squared vs. Adjusted R-Squared: What's the Difference?* (luettu 8.4.2023).
- [9] Hastie T., Tibshirani R., Friedman J. (2009) *The Elements of Statistical Learning (2nd edition)*. ISBN: 978-0387848570.
- [10] Nummenmaa, Lauri (2021). *Tilastotieteen käsikirja*. ISBN: 978-9520401382.
- [11] Greenacre, M. (2008). *Hierarchical cluster analysis*. Chapter 7. URL <http://www.econ.upf.edu/michael/stanford/maeb7.pdf>. (luettu 4.3.2023).
- [12] Jeong, Jiwon (2019). *Who is your Golden Goose?: Cohort Analysis*. URL <https://towardsdatascience.com/who-is-your-golden-goose-cohort-analysis-50c9de5dbd31>. (luettu 1.4.2023).
- [13] R: *The R Project for Statistical Computing*. URL <https://www.r-project.org/>.

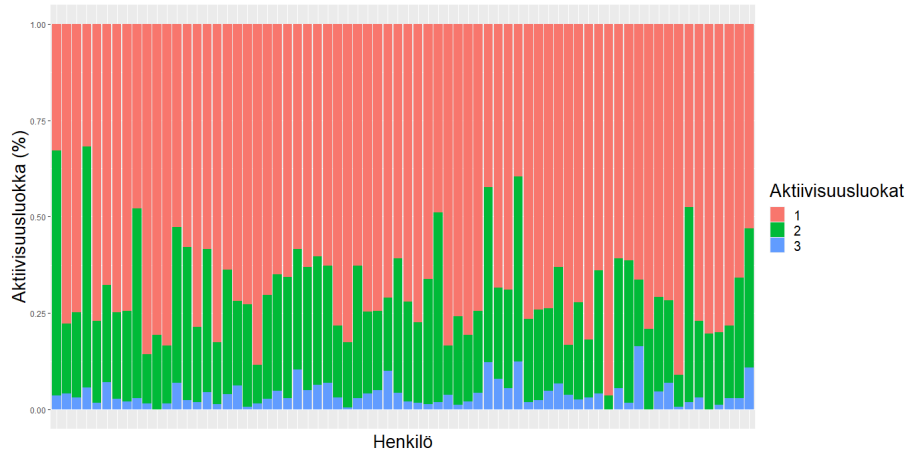
Liite A Aktiivisuusluokkien suhteelliset osuudet osa-aineistoissa



Kuva A1: Aktiivisuusluokkien jakautuminen toisessa osa-aineistossa. Tässä osa-aineistossa aktiivisuusluokkien 1, 2 ja 3 suhteelliset osuudet ovat keskimäärin 69 %, 27 % ja 4 %.



Kuva A2: Aktiivisuusluokkien jakautuminen kolmannessa osa-aineistossa. Tässä osa-aineistossa aktiivisuusluokkien 1, 2 ja 3 suhteelliset osuudet ovat keskimäärin 69 %, 27 % ja 4 %.



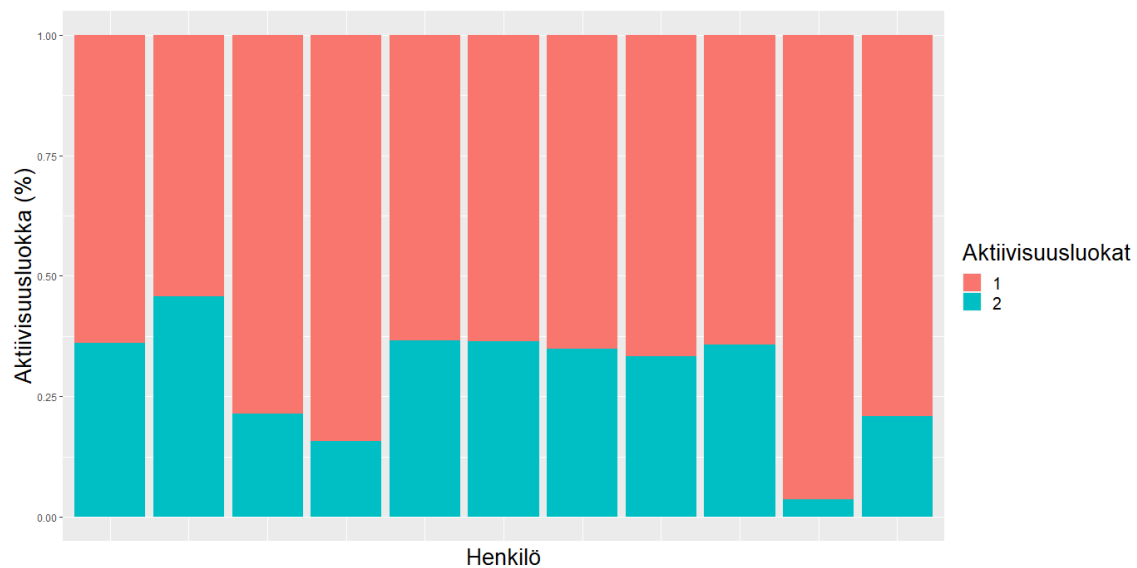
Kuva A3: Aktiivisuusluokkien jakautuminen neljännessä osa-aineistossa. Tässä osa-aineistossa aktiivisuusluokkien 1, 2 ja 3 suhteelliset osuudet ovat keskimäärin 69 %, 27 % ja 4 %.

Liite B Aktiivisuusluokkien suhteelliset osuudet klusterissa

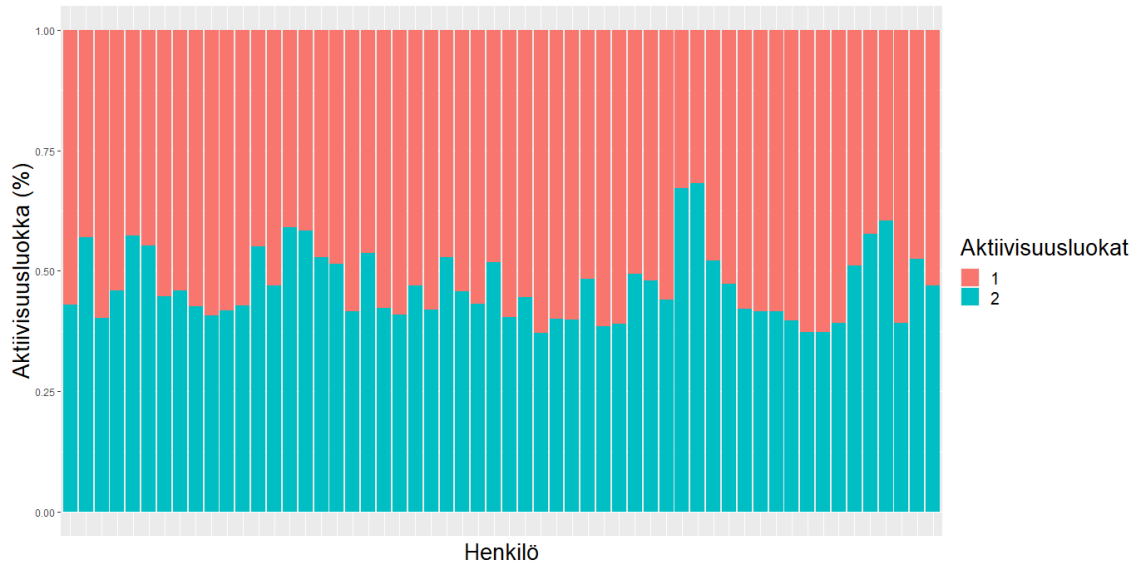
Kaksitilaisen mallin klusterit



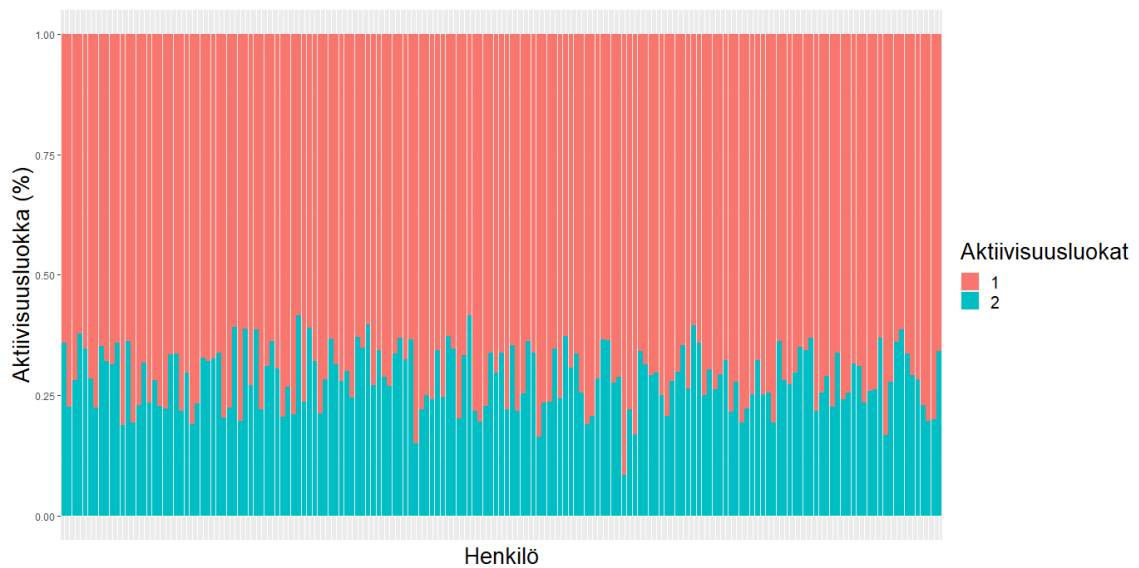
Kuva B1: Aktiivisuusluokkien suhteelliset osuudet klusterissa 2A. Tässä klusterissa aktiivisuusluokkien 1 ja 2 suhteelliset osuudet ovat keskimäärin 81 % ja 19 %.



Kuva B2: Aktiivisuusluokkien suhteelliset osuudet klusterissa 2B. Tässä klusterissa aktiivisuusluokkien 1 ja 2 suhteelliset osuudet ovat keskimäärin 71 % ja 29 %.

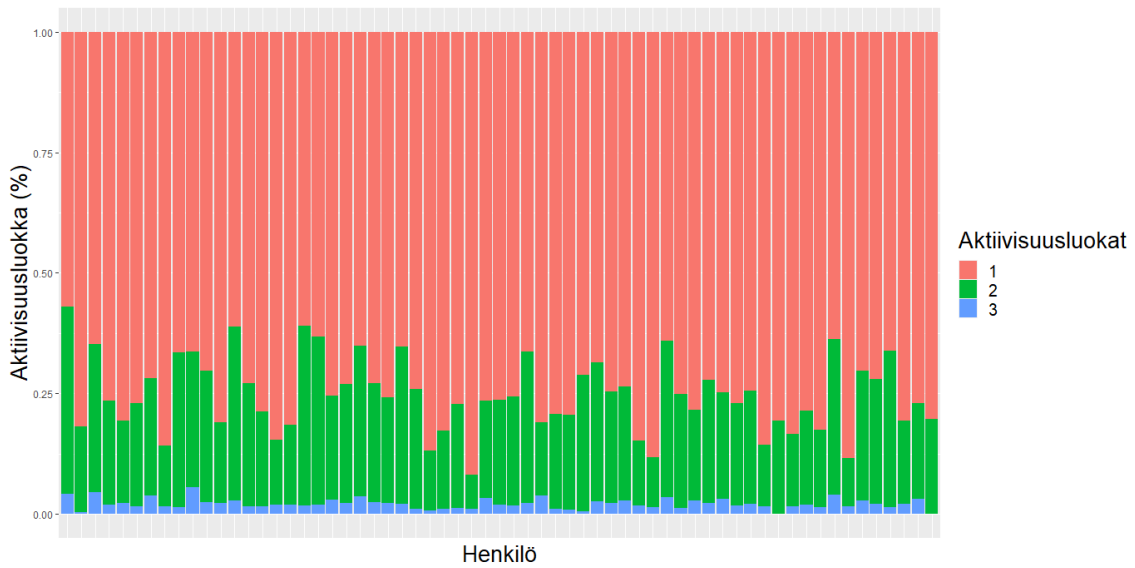


Kuva B3: Aktiivisuusluokkien suhteelliset osuudet klusterissa 2C. Tässä klusterissa aktiivisuusluokkien 1 ja 2 suhteelliset osuudet ovat keskimäärin 53 % ja 47 %.



Kuva B4: Aktiivisuusluokkien suhteelliset osuudet klusterissa 2D. Tässä klusterissa aktiivisuusluokkien 1 ja 2 suhteelliset osuudet ovat keskimäärin 71 % ja 29 %.

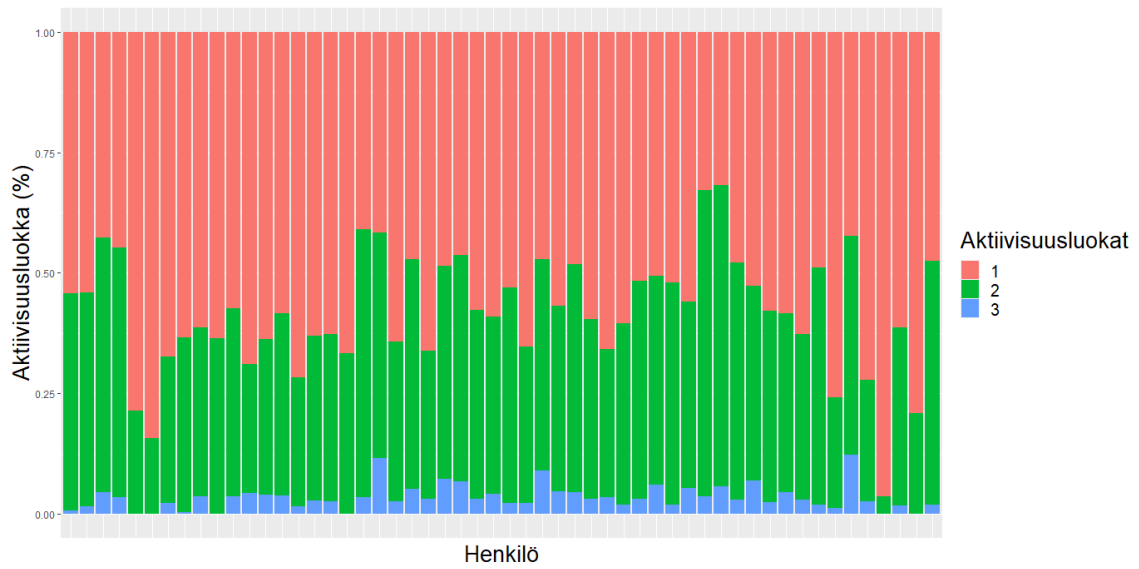
Kolmitilaisen mallin klusterit



Kuva B5: Aktiivisuusluokkien suhteelliset osuudet klusterissa 3A. Tässä klusterissa aktiivisuusluokkien 1, 2 ja 3 suhteelliset osuudet ovat keskimäärin 75 %, 23 % ja 2 %.

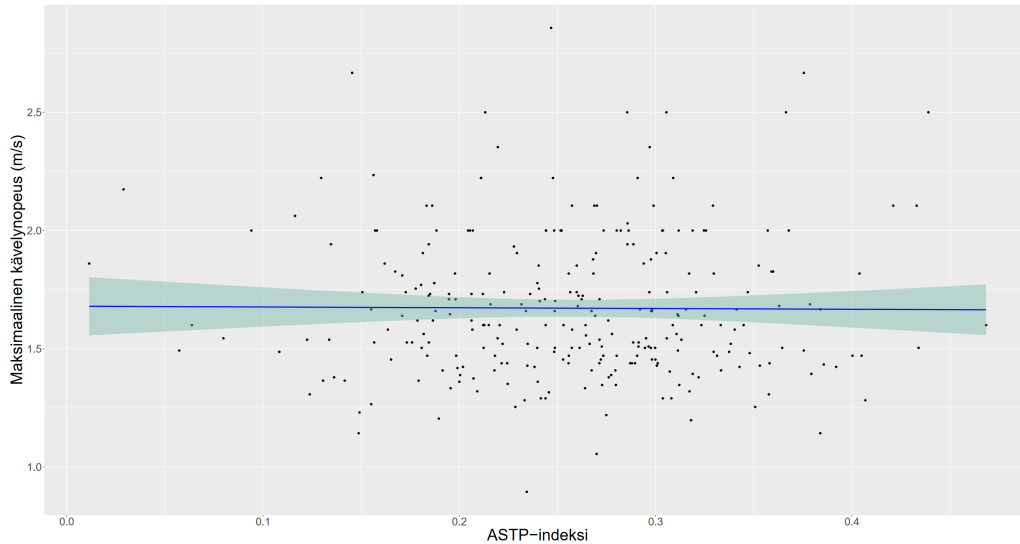


Kuva B6: Aktiivisuusluokkien suhteelliset osuudet klusterissa 3B. Tässä klusterissa aktiivisuusluokkien 1, 2 ja 3 suhteelliset osuudet ovat keskimäärin 71 %, 24 % ja 5 %.

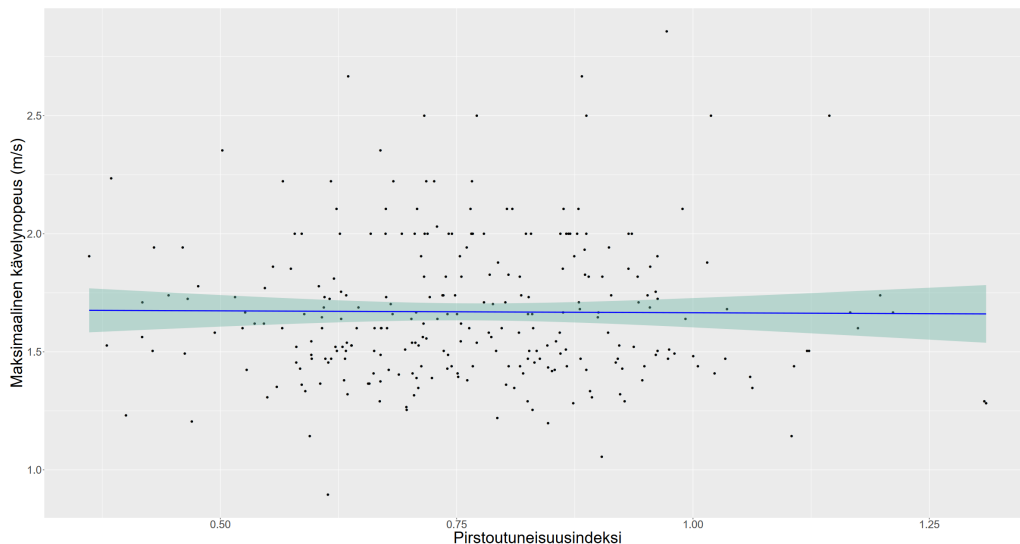


Kuva B7: Aktiivisuusluokkien suhteelliset osuudet klusterissa 3C. Tässä klusterissa aktiivisuusluokkien 1, 2 ja 3 suhteelliset osuudet ovat keskimäärin 59 %, 37 % ja 4 %.

Liite C ASTP-indeksin/pirstoutuneisuusindeksin ja maksimaalisen kävelynopeuden välinen yhteys



Kuva C1: ASTP-indeksin ja maksimaalisen kävelynopeuden (m/s) välinen yhteys. Pistejoukkoon sovitettu pienimmän neliösumman suora. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Pienimmän neliösumman suoran kulmakerroin on -0.03 ja 95 %:n luottamusväli [-0.51,0.44].



Kuva C2: Pirstoutuneisuusindeksin ja maksimaalisen kävelynopeuden (m/s) välinen yhteys. Pistejoukkoon sovitettu pienimmän neliösumman suora. Varjostettu alue esittää 95 %:n pisteittäistä luottamusväliä. Pienimmän neliösumman suoran kulmakerroin on -0.02 ja 95 %:n luottamusväli [-0.23,0.20]. Tieto pirstoutuneisuusindeksistä puuttuu kuudelta yksilöltä.