

PRO GRADU -TUTKIELMA

Tiina Hakanen

Aerobisen ja anaerobisen kynnyssykkeen mallintaminen MARS-
menetelmällä sekä liikunnan määrän ja laadun vaikutus kuntoarvioihin

TAMPEREEN YLIOPISTO
Informaatiotieteiden yksikkö
Tilastotiede
Maaliskuu 2011

Tampereen yliopisto

Informaatiotieteiden yksikkö

HAKANEN, TIINA: Aerobisen ja anaerobisen kynnyssykkeen mallintaminen MARS-menetelmällä sekä liikunnan määrän ja laadun vaikutus kuntoarvioihin

Pro gradu -tutkielma, 60 s., 8 liites.

Tilastotiede

Maaliskuu 2011

Tiivistelmä

Kestävyysvalmennuksessa aerobisen ja anaerobisen kynnyssykkeen tuntemus on urheilijan kehityksen kannalta merkittävä. Kynnyssykkeiden avulla urheilija löytää harjoittelutason, jolla hän voi saavuttaa optimaalisen suorituskyvyn. Samoista tiedoista on suuri hyöty myös kuntoilijoille. Aerobisen ja anaerobisen kynnyssykkeen määrittämiseen ei ole kehitetty yksiselitteistä kaavaa tai menetelmää. Tämän tutkimuksen lähtökohtana on kehittää tilastollisin menetelmin Testaus- ja kuntotutkimusaseman (TesKu ay) käyttämää syketasotestiä. Tavoitteena on määrittää aerobinen ja anaerobinen kynnyssyke hyödyntäen testattavan henkilön testituloksia sekä taustatietoja käyttämättä veren laktaattipitoisuuksia. Syketasojen määrittämisessä käytettävät tilastolliset mallit perustuvat regressiopohjaiseen MARS-menetelmään.

Lisäksi tutkielmassa tutkitaan liikuntakertojen sekä -lajien vaikutusta eri kestävyysalueiden kuntoarvioihin. Tässä tarkastelussa huomattiin liikuntakertojen ja kuntoarvioiden välillä tilastollisesti merkitsevä riippuvuus. Liikuntakerrat vaikuttavat selkeästi enemmän perus- kuin vauhti- tai maksimikestävyyyteen. Sen sijaan liikuntalajit vaikuttavat enemmän maksimikestävyyyteen. Merkittävää on tuloksista huomata, että palloilua harrastavat saivat selvästi heikompia arvioita perus- ja vauhtikestävyydestä kuin maksimikestävyydestä.

Myös testattavien liikuntatottumuksia tarkasteltiin tässä tutkielmassa. Testatuista yli kolmasosa harrastaa kävelyä. Naiset harrastavat miehiä enemmän mm. sauvakävelyä, kävelyä, jumppaa ja tanssia. Miehet mieltävät enemmän harrastuksikseen palloilulajit, golfin ja juoksun. Molemmat sukupuolet käyvät suunnilleen yhtä usein kuntosalilla. Myös liikuntatottumukset eroavat eri ikäluokissa: alle 40-vuotiaat harrastavat eniten pallopelejä, aerobicia ja käyvät useimmin kuntosalilla, yli 40-vuotiaat pelaavat golfia, jumppaavat ja kävelevät ilman tai sauvojen kanssa.

Tutkielmassa vertaillaan myös R-ohjelman eri funktioilla saatuja MARS-malleja sekä niiden tulosten eroavaisuuksia. Tässä yhteydessä havaittiin erään keskeisen kirjallisuuslähteen esimerkin virhe, mikä aiheuttaa MARS-mallin vääriä tulkintoja.

Asiasanat χ^2 -riippumattomuustesti, lineaarinen regressioanalyysi, paloittain lineaarinen malli, kuntotestaus, maksimisyke

Sisältö

1 Johdanto	7
1.1 Tutkimuksen tausta.....	8
1.2 Tutkimuksen tarkoitus	9
1.3 Tutkimuksen rakenne	9
2 Tutkimusaineiston kuvaus	10
2.1 Ikä.....	10
2.2 Rasvaprosentti	10
2.3 Aerobinen kynnyssyke	12
2.4 Anaerobinen kynnyssyke.....	13
2.5 Maksimisyke.....	15
2.6 Testattavien taustatietoja	16
3 Tutkimuksessa käytetyt menetelmät	17
3.1 χ^2 -riippumattomuustesti	17
3.2 Lineaarinen regressiomalli	18
3.2.1 Mallin estimointi.....	19
3.2.2 Regressiomallin selityssaste	20
3.2.3 Hypoteesien testaus lineaarisessa regressioanalyysissä.....	21
3.2.4 Regressiomalli ilman vakiotermiä	22
3.3 Paloittain lineaarinen malli.....	25
3.4 MARS-menetelmä.....	26
3.4.1 MARS-malli.....	26
3.4.2 MARS-mallin selityssaste	30
3.4.3 MARS-mallin muodostaminen R-ohjelman <i>earth</i> -funktiolla.....	32
4 Liikuntakertojen sekä liikuntalajien vaikutus kuntoarvioihin	34
4.1 Liikuntakertojen vaikutus eri kestävyysalueiden arvioihin.....	34
4.1.1 Liikuntakertojen vaikutus peruskestävyyteen.....	34
4.1.2 Liikuntakertojen vaikutus vauhtikestävyyteen	35
4.1.3 Liikuntakertojen vaikutus maksimikestävyyteen.....	36
4.2 Liikuntalajien vaikutus kestävyysalueiden kuntoarvioihin	37
4.2.1 Liikuntalajien vaikutus peruskestävyyteen	38
4.2.2 Liikuntalajien vaikutus vauhtikestävyyteen.....	40
4.2.3 Liikuntalajien vaikutus maksimikestävyyteen.....	41
5 Aerobisen ja anaerobisen kynnyksen syketason mallintaminen MARS-menetelmällä	43
5.1 Aerobinen kynnyssyke	43
5.2 Anaerobinen kynnyssyke.....	47
5.3 Mallien kehittämistä	51
5.3.1 Aerobisen kynnyssykkeen mallintaminen	51
5.3.2 Anaerobisen kynnyssykkeen mallintaminen.....	52
5.4 Aerobisen ja anaerobisen kynnyssykkeen mallintaminen MARS-menetelmällä <i>earth</i> - funktiota käyttäen	54
5.4.1 Aerobisen kynnyssykkeen mallintaminen	54
5.4.2 Anaerobisen kynnyssykkeen mallintaminen.....	55
6 Yhteenveto	57

Lähdeluettelo	59
Liite A: Taustakyselylomake ja testauspöytäkirja	61
Liite B: Muuttujaluettelo	63
Liite C: Kuviot mallien diagnostiikasta	65

1 Johdanto

Suomessa tehdään vuosittain noin 200 000 kuntotestiä, joilla tarkoitetaan ihmisen suorituskyvyn ja fyysisen kunnan osa-alueiden mittaamista sekä arviointia. Testejä toteutetaan terveydenhuollossa, kuntoliikunnassa, urheiluvalmennuksessa sekä työelämässä. Suurin kävijäjoukko kuntotestauksessa on työikäiset, joiden fyysiseen kuntoon työnantajat sekä ammattijärjestöt ovat kiinnittäneet viime vuosina selkeästi enemmän huomiota. (Helimäki, Keskinen, Alén, Komi & Takkala 2000.) Urheiluvalmennuksessa kuntotestausta käytetään apuna lajikohtaisen suorituskyvyn kehittämisessä. Testien avulla voidaan parhaiten seurata urheilijan kehitystä ja suunnata harjoittelua optimaalisen suorituskyvyn saavuttamiseen. Terveydenhuollossa rasituskokeilla ja kliinisillä testeillä arvioidaan liikunta-kelpoisuutta ja sairauden vaikeusastetta, seurataan ja diagnosoidaan sairauksia sekä vammoista parantumista.

Kuntotestin tavoitteena on antaa kuntoilijalle henkilökohtaiset ohjeet kunnan kohentamiseen. Kuntoilija saa testin jälkeen yksilöllisen liikuntaohjeen, jossa kerrotaan tarkasti, miten hänen pitää liikkua parantaakseen kuntoa. Kuntotesti herättää myös kiinnostusta ja tietoisuutta oman terveyden edistämiseen, mikä vaikuttaa luonnollisesti elämän laadun paranemiseen ja hyvinvoinnin lisääntymiseen. (Keskinen, Häkkinen & Kallinen 2004.)

Kunnan paranemisella on merkittävä osuus monien kroonisten sairauksien ehkäisyssä sekä hoidossa ja se auttaa myös uupumisongelmiin sekä stressihallintaan. Kuntoilija saa kuntotestistä hyvät tiedot omasta kunnostaan ja siitä millaisella liikunnalla hän voi parantaa omaa terveyttään, toimintakykyään ja hyvinvointiaan. Näin kunnan testaus voidaan nähdä kansanterveyttä edistävänä toimintana. Liikunnan on todettu vähentävän riskiä sairastua sydän- ja verisuonitauteihin, diabetekseen ja moneen syöpäsairauteen. Lihasvoiman vahvistamisella voidaan vähentää kaatumisia ja ylläpitää iäkkäimpien henkilöiden toimintakykyä (Vuori & Taimela 1999). WHO:n maailmanlaajuisen strategian ”Ravinto, liikunta ja terveys” tavoite on mm. vähentää liian vähäisen liikunnan aiheuttamaa sairastavuutta sekä kuolleisuutta ja lisätä tietoisuutta liikunnan vaikutuksesta terveyteen. Näitä tavoitteita myös kuntotestaus tukee, ja tämän strategian mukaan pitäisikin tukea mm. terveyttä edistäviä tutkimuksia.

Vähäinen liikunta maksaa vuodessa yhteiskunnalle noin 200 miljoonaa euroa. Liikkumattomuuden suurimmat terveystaloudelliset kustannukset on arvioitu tulevan työn tuottavuuden heikkenemisestä. Toiseksi suurimmat tulevat yleisimmistä kansantaudeista sekä lonkkamurtumista. Vähäisestä liikunnasta johtuvien sairauspäivien on arvioitu nousevan vuodessa noin kahteen miljoonaan. Tämä

tarkoittaa noin 400 miljoonan euron palkka- ja työnantajakuluja. Nämä kustannukset painottuvat eniten yksityissektorille (72 % työntekijöistä). Noin 40 % väestöstä liikkuu liian vähän, eli mikäli näistä saataisiin 100 000 uutta liikkumaan riittävästi, se vapauttaisi 20 miljoonaa euroa julkisia varoja. (Fogelholm, Paronen & Miettinen 2007.) Kuntotestauksen kansantaloudelliset vaikutukset voidaan arvioida edellä mainittujen arvioiden perusteella. Mikäli joka neljäs testatuista eli noin 50 000 alkaisi liikkua riittävästi, se vapauttaisi 10 miljoonaa euroa julkisia varoja käytettäväksi muuhun kuin liikkumattomuudesta aiheutuviin sairauksiin.

1.1 Tutkimuksen tausta

Tämän tutkimuksen tutkimusaineistona käytetään Testaus- ja kuntotutkimusasema TesKu ay:n peruskuntotestin tuloksia. TesKu ay on tehnyt kuntotestejä vuodesta 1994, ja sen testitarjontaan kuuluvat peruskuntotestin lisäksi mm. syketaso-, lihaskunto-, voima- ja nopeustestit. Peruskuntotestissä testattava suorittaa 3 - 7 kertaa 1 000 m siten, että räsitusaste nousee jokaisella kierroksella ja testin viimeinen kierros olisi testattavan maksimisuoritus. Testatuista 37 % on suorittanut 3 x 1 000 m ja 44 % 6 x 1 000 m. Testattavalla on testin aikana sykemittari, joka mittaa sykkeen ja ajan jokaiselta kierrokselta. Veren laktaattipitoisuus mitataan jokaisen kilometrin jälkeen ottamalla testattavasta verinäyte. Testin perusteella määritetään aerobisen ja anaerobisen kynnyksen sekä maksimaalisen aerobisen tehon syketasot, vauhti, hapenotto-kyky, veren laktaattipitoisuus sekä kuntoarviot.

Aerobisen ja anaerobisen kynnyksen tietäminen on tärkeää kuntoilijalle, sillä näiden avulla hän löytää sopivat harjoittelutasot, joilla hän voi kehittää perus- sekä vauhtikestävyystään. Kynnysten määrittämisessä tarkastellaan veren laktaattitasoa nousua suhteessa sykkeen sekä nopeuden nousuun. Suomessa kynnysten määrittämiseen käytetään tiettyjä kriteereitä. Esimerkiksi laktaattitestissä aerobinen ja anaerobinen kynnykset voidaan määrittää laktaatti- ja sykekäyrältä, jotka saadaan laktaattipitoisuuden lineaarisesti poikkeavista nousukohdista (Keskinen et al. 2004). Aerobisen ja anaerobisen kynnyksen määrittämiseen ei ole kehitetty yksiselitteistä kaavaa tai menetelmää, joten testaaja määrittää ne testituloksien avulla silmämääräisesti (Sovijärvi, Uusitalo, Länsimies & Vuori 1994). Tarkkojen kynnystasojen määrittäminen testatuille edellyttää hyvien testausolosuhteiden lisäksi myös testaajan kokemuksellista ammattitaitoa.

Aerobinen ja anaerobinen kynnys-aste voidaan määrittää myös ns. syketasotestillä. Tässä testissä ei mitata veren laktaattipitoisuutta, joten kynnys-asteita on hankalampi määrittää ja se on laktaattitestin verrattuna epätarkempi. Syketasotesti on myös hieman halvempi eikä yhtä työläs kuin peruskuntotesti, jossa mitataan veren laktaattipitoisuus. Tämän tutkimuksen yksi tavoite on löytää

sekä aerobiselle että anaerobiselle kynnyssykkeelle tilastollinen malli, jota voitaisiin käyttää syketasotestauksessa.

1.2 Tutkimuksen tarkoitus

Tutkimuksen lähtökohtana on TesKu ay:n käyttämän syketasotestin kehittäminen tilastollisin menetelmin. Tavoitteena on määrittää aerobinen ja anaerobinen kynnyssyke käyttämällä testattavan testituloksia sekä taustatietoja lukuun ottamatta veren laktaattipitoisuuksia. Syketasojen määrittämisessä käytetään MARS-menetelmää. Tutkimuksessa perehdytään myös poikkeaviin havaintoihin, sillä peruskuntotestin tuloksiin vaikuttavat testin suoritusajan lisäksi monet muut tekijät. Tutkimuksessa tarkastellaan myös liikuntakerran ja -lajin vaikutusta kestävyysalueiden kuntoarvioihin käyttäen χ^2 -riippumattomuustestiä. Tutkielman tilastolliset analyysit on tehty Tixel 8.48, STATISTICA 8.0 ja R 2.12.1 -ohjelmilla.

1.3 Tutkimuksen rakenne

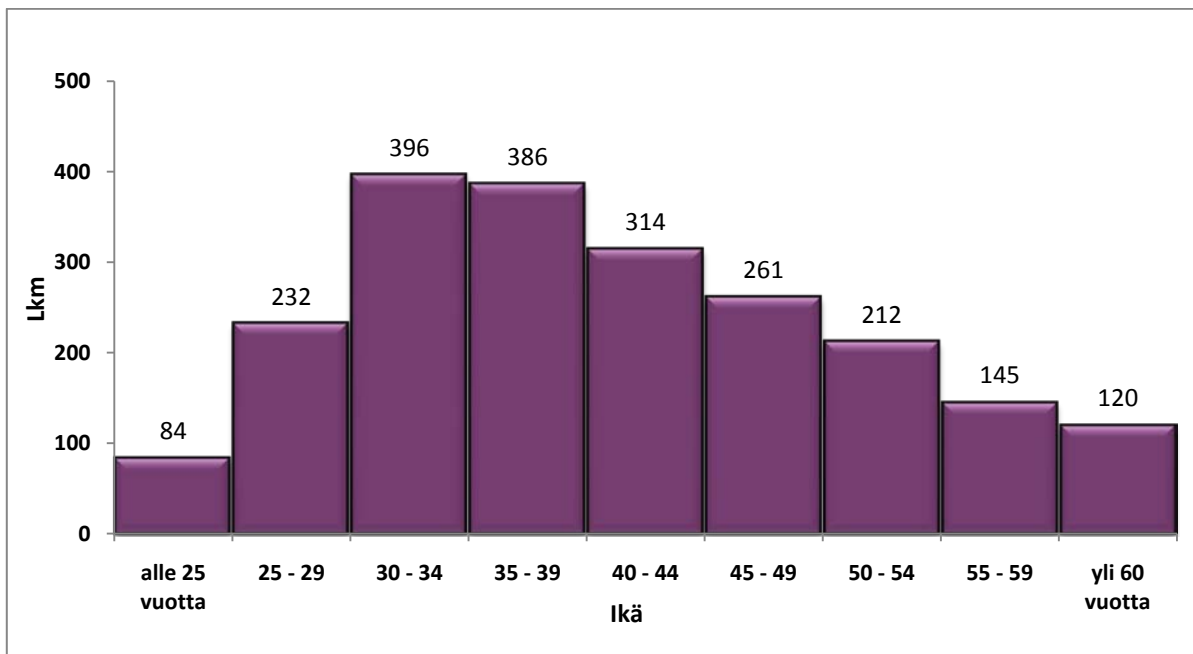
Johdantoluvun jälkeen luvussa 2 on kuvattu tutkimusaineisto ja sen tärkeimmät muuttujat. Aerobinen sekä anaerobinen kynnyssyke on esitetty tarkemmin kuin muut muuttujat, sillä ne ovat olennaisia tässä tutkimuksessa. Liitteessä B on esitetty lyhyesti aineiston kaikki muuttujat. Luvussa 3 esitellään aineiston analysointiin käytetyt χ^2 -riippumattomuustesti ja MARS-menetelmä. Tämän luvun jälkeen menetelmiä sovelletaan aineistoon. Ensin luvussa 4 tarkastellaan liikuntakertojen ja lajien vaikutusta kuntoarvioihin χ^2 -riippumattomuustestillä. Seuraavassa luvussa mallinnetaan MARS-menetelmällä aerobista ja anaerobista kynnyssykettä ja viimeisessä luvussa tehdään yhteenveto tutkimuksen tuloksista.

2 Tutkimusaineiston kuvaus

Tutkimusaineistona on 1 736:n henkilön testitulokset vuosilta 1995 - 2009. Havaintoja on yhteensä 2 456, joten jotkut henkilöt on testattu useammin kuin kerran. Havainnosta 875 on naisten ja 1 276 miesten testituloksia, 305:een tulokseen ei ole tallennettu sukupuolta. Aineistossa on yhteensä 59 muuttujaa, joista 15 muuttujaa on testattavan taustatietoja ja loput muuttujat ovat testistä saatuja. Tässä luvussa kuvataan lyhyesti tutkimuksen kannalta tärkeimmät muuttujat.

2.1 Ikä

Ikä-muuttuja on testattavan ikä testauspäivänä. Testattujen keski-ikä on 40,3 vuotta. Suurin osa (62 %) testatuista on 30 - 50 vuoden ikäisiä. Aineiston nuorin on 19- ja vanhin 71-vuotias. Iän jakaumasta (Kuvio 2.1) huomaamme, että alle 25-vuotiaita on vain 4 %. Kuntotestiin hakeutuneet ovat yleensä yli 30-vuotiaita. Naisten keski-ikä on 41,2 ja miesten 39,7 vuotta. Kuntotestissä on käynyt suhteessa enemmän yli 50-vuotiaita naisia kuin miehiä, sillä 26 % testatuista naisista on yli 50-vuotiaita, kun miehillä vastaava prosenttiluku on 19.

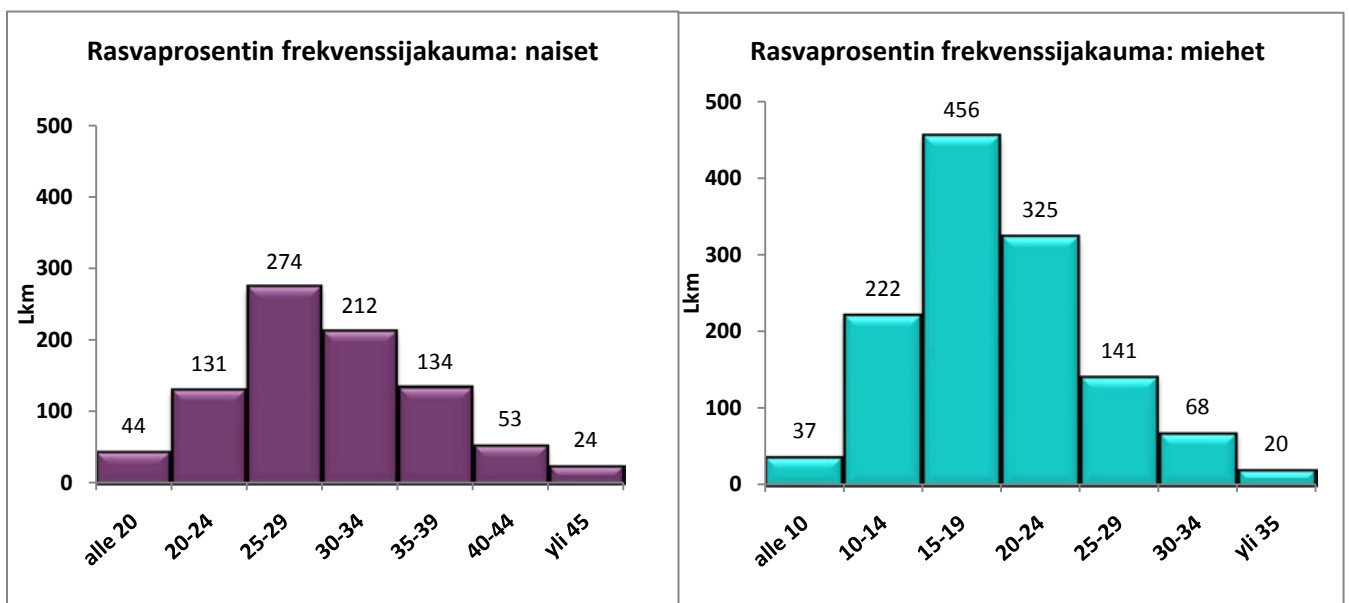


Kuvio 2.1. Iän frekvenssijakauma.

2.2 Rasvaprosentti

Rasvaprosentti kertoo kehossa olevan rasvan määrän prosentteina koko kehon painosta. Testatavalta mitataan ennen testiä rasvaprosentti ihopoimiumittauksella, jolla saadaan ihonalaisen rasvakudoksen paksuus. Se on melko luotettava kuva kehon rasvan osuudesta. Mittauksessa käytetään ihopoimiumittaria ja neljän pisteen mittausmenetelmää. Mittapaikat ovat olkavarren ojentaja, hauislihas, lapaluun alakärjen alapuoli ja suoliluun harjanteen yläpuoli. Saatujen neljän mittaustuloksen summan avulla arvioidaan rasvaprosentti. Ihopoimiumittauksen tekniikka vaikuttaa lopputulokseen, joten eri mittaajien saamat tulokset eivät ole täysin vertailukelpoisia. (Keskinen et al. 2004.) Kunto-testissä käyneiden henkilöiden rasvaprosentteja voidaan pitää vertailukelpoisina, sillä mittaajana on aina ollut sama henkilö.

Naisten kehossa on enemmän rasvaa kuin miesten, joten on syytä tarkastella niitä erikseen. Naisten havaintoja on yhteensä 872, joista saatu keskiarvo on 29,6 % ja keskihajonta 6,6. Pienin rasvaprosentti on 12,2 ja suurin 53. Kuviosta 2.2 nähdään, että naisten rasvaprosentti arvoista yli puolet kuuluu 25 - 34 välille (56 %). Molempien sukupuolten jakaumat ovat hieman oikealle vinoja. Miesten havaintoja on 1 269 ja keskiarvo on 19 % eli noin 10 prosenttiyksikköä pienempi kuin naisilla. Keskihajonta on 5,9. Pienin mitattu rasvaprosentti on vain 3,9 % ja suurin 41,3 %. Mitatuista miesten rasvaprosenttien arvoista 65 % on alle 20 %.



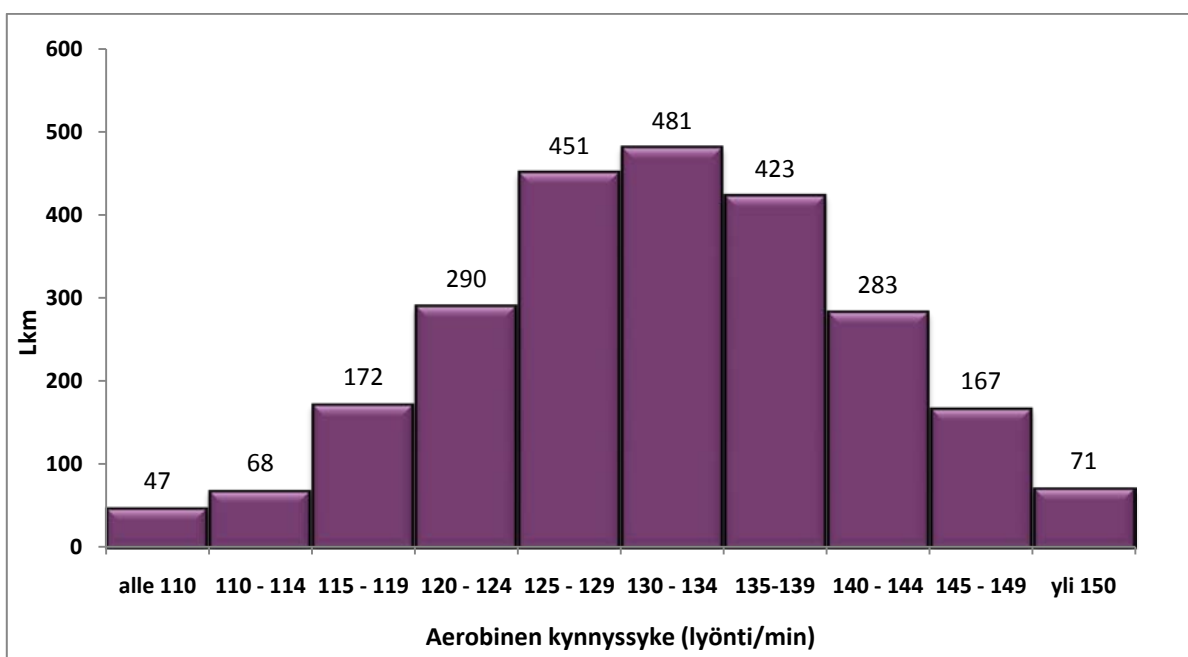
Kuvio 2.2. Rasvaprosentin frekvenssijakaumat.

2.3 Aerobinen kynnyssyke

Fyysisessä rasituksessa lihaksiin ja vereen alkaa muodostua laktaattia. Ennen aerobisen kynnyksen saavuttamista ihmisen kehon lihakset saavat riittävästi happea, jolloin veren laktaattipitoisuus ei nouse lepotasoa korkeammalle. Kun aerobinen kynnyks on ylitetty, kehoon alkaa muodostua maitohappoa. Aerobinen kynnyks tarkoittaa urheiluvalmennuksessa peruskestävyyden ja vauhtikestävyyden välistä rajaa (Keskinen et al. 2004).

Kynnyksen avulla pystytään arvioimaan testattavan peruskestävyys. Aerobisen kynnyssykkeeseen tietäminen on kuntoilijalle hyvin tärkeää, sillä sen avulla löydetään peruskestävyyttä parantava optimaalinen harjoittelutaso. Aerobinen kynnyssyke laskee iän myötä, koska sydämen maksimaalinen syketaajuus pienenee mm. sairauksien takia. Tosin terveen sydämen pumppauskyky säilyy vanhetessa hyvin ja onkin todettu, että urheilua koko elämänsä jatkavilla sydämen pumppauskyky säilyy muita paremmin (Vuori et al. 1999).

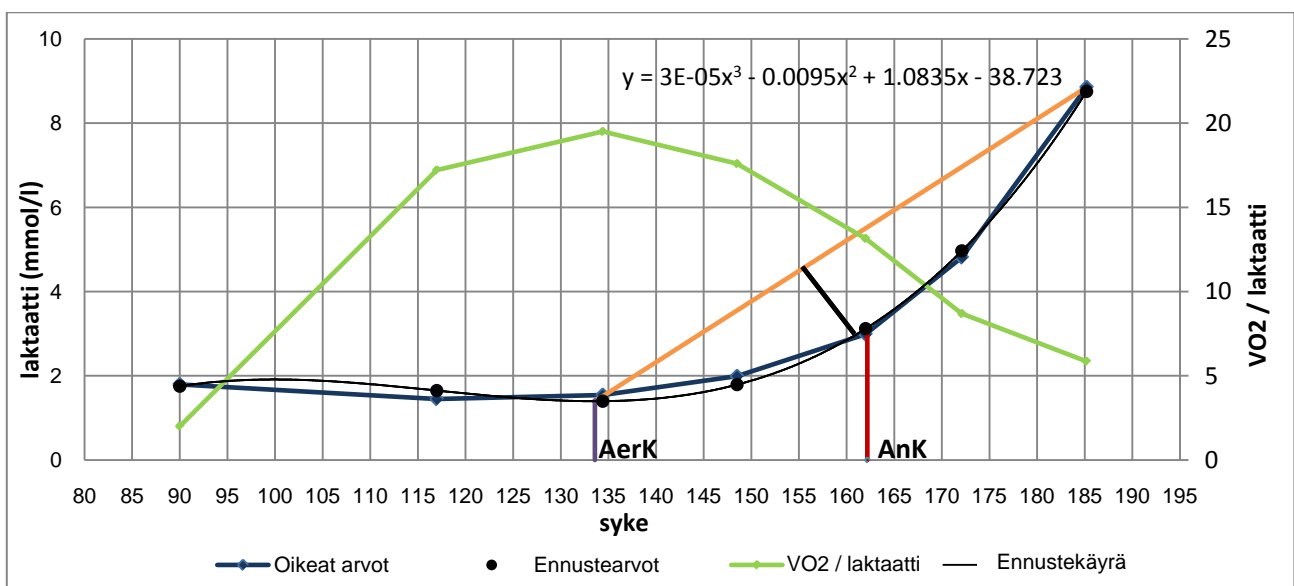
Kuviossa 2.3 on aerobisen kynnyssykkeeseen frekvenssijakauma, mikä on lähes normaalijakauman muotoinen. Testattavien kynnyssykkeiden keskiarvoksi saadaan 131 ja keskihajonnaksi 10,04. Testattujen arvoista 67 % on välillä 120 - 140. Alhaisin aerobinen kynnyssyke on 84 ja korkein 159. Kun tarkastellaan iän vaikutusta aerobiseen kynnyssykkeeseen, huomataan ikäryhmien kynnyssykkeiden keskiarvojen laskevan, sillä alle 30-vuotiaiden keskiarvo on 138 (n = 332) ja yli 60-vuotiaiden 119 (n = 129).



Kuvio 2.3. Aerobisen kynnyssykkeeseen frekvenssijakauma.

Kuntotestistä saatujen tuloksien perusteella aerobinen kynnyks määritetään Kuvion 2.4 avulla. Kuviossa x-akselilla on syke ja y-akselilla veren laktaattipitoisuus sekä hapenottokyvyn ja veren laktaattipitoisuuden suhde. Sininen käyrä kuvaa testin jokaisen kierroksen veren laktaattipitoisuuden arvoa ja vihreä käyrä hapenottokyvyn ja laktaatin suhdetta. Kuvioon on piirretty 3.asteen polynomisovite jokaisen kierroksien veren laktaattipitoisuuden ja syke arvioiden välille (musta käyrä).

Aerobinen kynnyssyke saadaan määritettyä polynomikäyrän avulla, sillä aerobinen kynnyks on laktaattikäyrän ensimmäinen nousukohta eli polynomikäyrän minimikohta. Hapenottokyvyn ja laktaatin suhdekäyrän avulla pystytään myös arvioimaan aerobinen kynnyks. Kyseinen suhde on suurimmillaan aerobisessa kynnystasossa, koska tämän kynnyksen jälkeen veren laktaattipitoisuus alkaa rasituksen aikana kohota perustasostaan. (Keskinen et al. 2004.)

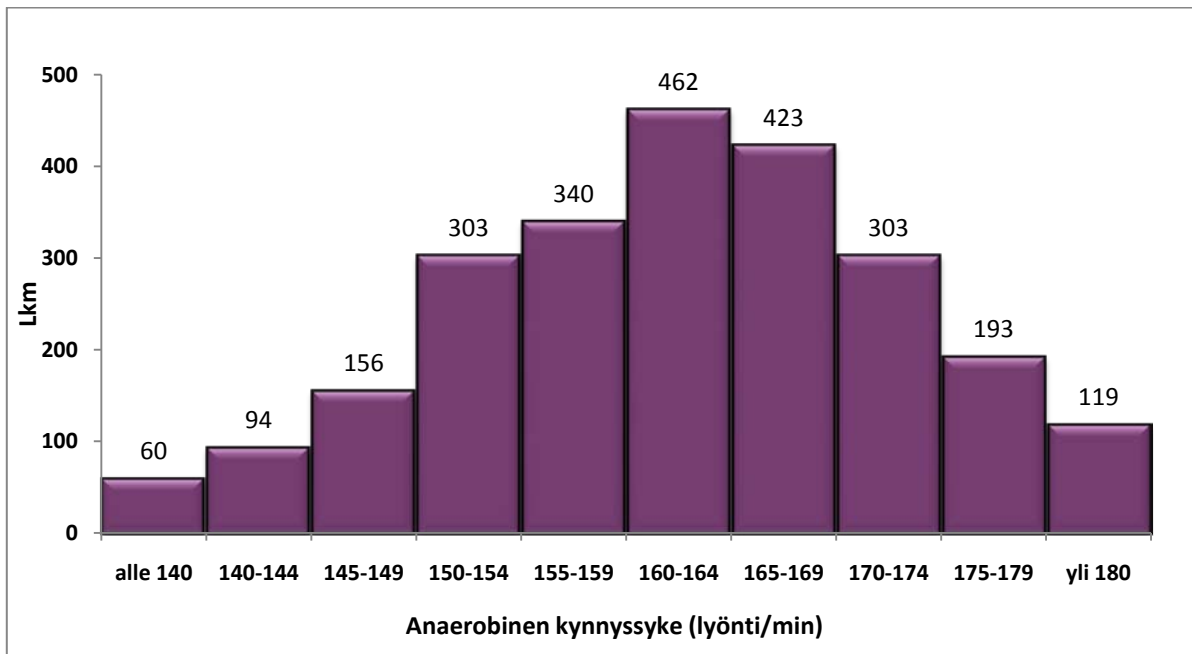


Kuvio 2.4. Esimerkki aerobisen ja anaerobisen kynnyksen määrittämisestä.

2.4 Anaerobinen kynnyssyke

Rasituksen edetessä alkaa veren laktaattipitoisuus kasvaa, ja siksi lihasten työteho heikkenee sekä suoritusteho laskee. Kun laktaatin tuotto ja sen poisto eivät pysy enää tasapainossa, on anaerobinen kynnyks ylitetty. Anaerobisen kynnyksen on siis määritelty olevan korkein mahdollinen työteho ja energiakulutuksen taso, jossa laktaatin tuotto ja sen poisto pysyvät vielä tasapainossa. Anaerobisen kynnyksen jälkeen veren laktaattipitoisuus alkaa nousta nopeasti, joten lihastyötä ei enää pystytä ylläpitämään, ja tämä johtaa lopulta uupumiseen. Anaerobista kynnystä käytetään erityisesti kestävyysurheilijoiden valmennuksessa. Sen avulla löydetään sopiva harjoittelutaso, mikä kehittää urheilijan vauhtikestävyyttä. Veren laktaattipitoisuus on anaerobisella kynnyksellä noin 2,5 - 5,5 mmol/l. (Keskinen et al. 2004.)

Anaerobisen kynnyssykkeen keskiarvo on 162 ja keskihajonta 11,23. Testattavien arvoista 62 % on välillä 155 - 174. Alhaisin anaerobinen kynnyssyke on 102 ja korkein 199. Kuviossa 2.5 on anaerobisen kynnyssykkeen frekvenssijakauma, mikä on hieman vasemmalle vino. Anaerobinen kynnyssyke laskee iän myötä. Tämän voi todeta selkeästi ikäryhmien kynnyssykkeiden keskiarvoista, sillä alle 30-vuotiaiden keskiarvo on 171 (n = 332) ja yli 60-vuotiaiden 147 (n = 129).



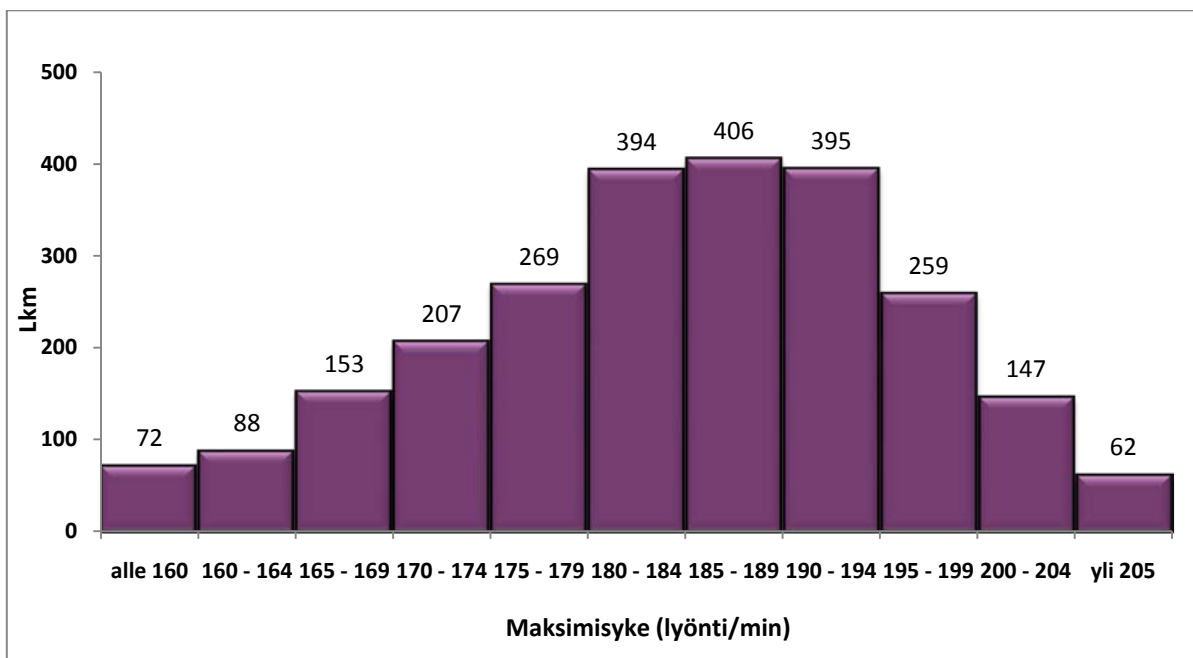
Kuvio 2.5. Anaerobisen kynnyssykkeen frekvenssijakauma.

Kuntotestaaja määrittää myös anaerobisen kynnyksen Kuvion 2.4:n avulla. Laktaatti-sykekäyrällä anaerobinen kynnyks löydetään laktaattipitoisuuden toisen lineaarisen nousun kohdasta. Laskennallisesti käytetään apuna 3. asteen polynomisovitetta. Anaerobinen kynnyks löydetään polynomikäyrän avulla niin, että yhdistetään suoralla viivalla polynomien minikohta sekä maksimikohta (Kuviossa 2.4 oranssiviiva) ja tämän jälkeen määritetään polynomikäyrältä kohta, jossa kohtisuora etäisyys suoran ja polynomikäyrän välillä on suurin. Löydetty piste on anaerobinen kynnyks. Tätä menetelmää kutsutaan Dmax-analyysiksi, mitä käytetään enimmäkseen määrittettäessä anaerobista kynnyksstä laktaatti-työtehokäyrän avulla. Sitä voidaan käyttää myös tilanteessa, jossa määritetään anaerobinen kynnyks laktaatti-sykekäyrän avulla (Australian Sports Commission 2000).

2.5 Maksimisyke

Maksimisyke on sydämen suurin lyöntitiheys. Maksimisyke laskee keskimäärin yhden lyönnin jokaista ikävuotta kohden. Kuntotestistä saatu maksimisyke on viimeisen kierroksen syke, mikä on korkein testissä saatu raskautustaso. Sydän- ja verisuonitautien hoitoihin käytettävät lääkkeet vaikuttavat maksimisykkeeseen, mutta myös yksilöllisesti syke vaihtelee suuresti (Vuori et al. 1999).

Maksimisykkeen keskiarvo on 184 ja keskihajonta 12,31. Alhaisin maksimisyke on 128 ja korkein 217. Alle 30-vuotiaiden maksimisykkeen keskiarvo on 193 (n = 332) ja yli 60-vuotiaiden 168 (n = 130). Kuviossa 2.6 maksimisykkeen frekvenssijakaumasta huomataan, että testatuista lähes puolella maksimisyke on 180:n ja 194:n välillä.



Kuvio 2.6. Maksimisykkeen frekvenssijakauma.

2.6 Testattavien taustatietoja

Testattava täyttää ennen testiä taustatietolomakkeen (Liite A), missä on kysytty mm. sairauksista, liikuntatottumuksista, tupakoinnista. Tässä alaluvussa tarkastellaan joitakin taustatietomuuttujia.

Vastanneista 34 % kulkee työmatkan joko kävellen tai pyörällä (n = 2 315), ja keskimäärin aikaa työmatkaan menee 32,6 minuuttia. Naiset kulkevat työmatkansa kävellen tai pyörällä ja harrastavat hyötyliikuntaa enemmän kuin miehet. Hyötyliikuntaa vastanneista harrastaa suurin osa (74 %).

Hengitys- ja verenkiertosairauksia on 14 %:lla, tuki- ja liikuntaelimestön vaivoja 19 %:lla ja jatkuva lääkitys 16 %:lla vastanneista. Vanhemmilla ikäryhmillä esiintyy näitä edellä mainittuja enemmän kuin nuorilla, ja naisten keskuudessa nämä ovat yleisempiä kuin miesten. Vain 10 % on vastannut tupakoivansa, mikä on vähän, sillä joka viides suomalainen on vastannut tupakoivansa THL:n postikyselytutkimuksessa ”Suomalaisen aikuisväestön terveyskäyttäytyminen ja terveys”. Sukupuolella ei ole vaikutusta tupakointiin, sillä vastanneista naisista tupakoi 10 % ja miehistä vain prosentin enemmän. Edellä mainitussa tutkimuksessa miehistä tupakoi joka neljäs ja naisista 18 %, joten peruskuntotestin sukupuolien välinen ero on hyvin pieni verrattuna THL:n tutkimuksesta saatuihin tuloksiin. Kun verrataan testin kävijöiden ja koko väestön sairauksien ja tupakoivien määrää, on testatuilla selkeästi terveellisemmät elämäntavat kuin väestössä keskimäärin.

Liikuntakerrat, -lajit ja kestävyysalueiden kuntoarviot ovat tutkimuksen kannalta tärkeitä muuttujia. Niitä tarkastellaan yksityiskohtaisesta luvussa 4, jossa samalla on verrattu muuttujien välistä riippuvuutta.

3 Tutkimuksessa käytetyt menetelmät

3.1 χ^2 -riippumattomuustesti

Muuttujien välistä tilastollista riippuvuutta $I \times J$ -frekvenssitaulukossa voidaan testata Pearsonin χ^2 -testisuureen avulla. Testisuure on muotoa

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - e_{ij})^2}{e_{ij}},$$

missä I = rivien lukumäärä, J = sarakkeiden lukumäärä, O_{ij} = havaitut frekvenssit, e_{ij} = odotetut frekvenssit. Nollahypoteesina (H_0) on oletus, että muuttujien välillä ei ole riippuvuutta. Odotetut eli teoreettiset frekvenssit lasketaan olettaen, että muuttujien välillä ei ole riippuvuutta. Havaittuja frekvenssejä verrataan odotettuihin frekvensseihin. Mitä enemmän havaitut ja odotetut frekvenssit eroavat toisistaan, sitä suuremman arvon χ^2 -testisuure saa. Nollahypoteesin vallitessa χ^2 -testisuure noudattaa χ^2 -jakaumaa vapausastein $df = (I - 1) \times (J - 1)$. (Agregsti 2007.)

Havaittua p-arvoa verrataan valittuun testin merkitsevyystasoon. Yleisin käytetty merkitsevyystaso on 0,05. Mikäli merkitsevyystasoksi on valittu 0,05, ja havaittu p-arvo on alle 0,05, niin nollahypoteesi hylätään.

χ^2 -riippumattomuustestiä suositellaan käytettäväksi vain silloin, kun kaikki odotetut frekvenssit ovat suurempia kuin yksi ja korkeintaan 20 % odotetuista frekvensseistä on pienempiä kuin 5. Testiä käytettäessä on syytä huomioida, että testi kertoo vain muuttujien välisen riippuvuuden olemassaolon, ei sen tyyppiä. Tästä muuttujien välisestä riippuvuuden voimakkuudesta voidaan saada käsitys kontingenssikertoimen avulla, joka lasketaan kaavalla

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}},$$

missä χ^2 on χ^2 -testisuureen arvo ja n on havaintojen lukumäärä. Kontingenssikerroin voi saada arvoja nollan ja yhden välillä. Muuttujien välillä on sitä enemmän riippuvuutta, mitä lähempänä yhtä kontingenssikertoimen arvo on. Kertoimen arvo riippuu kuitenkin luokkien lukumäärästä. Esimerkiksi 2×2 -taulukossa suurin mahdollinen kontingenssiarvo on noin 0,7 ja 3×3 -taulukossa 0,8. Tämän takia kontingenssikerroin ei ole vertailukelpoinen erikokoisten taulukkojen välillä. (Heikkilä 2001.)

3.2 Lineaarinen regressiomalli

Lineaarinen regressiomalli pyrkii selittämään muuttujien välistä lineaarista riippuvuutta. Merkitään selitettävää muuttujaa y :llä ja selittäviä muuttujia x_1, x_2, \dots, x_k . Tällöin malli on muotoa

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k,$$

missä β_0 on vakiokerroin, β_j on selittäjän x_j regressiokerroin ja ε_i on virhetermi. Malli on yhden selittäjän lineaarinen regressiomalli, kun $j = 1$. Useamman selittäjän tapauksessa ($j > 1$) malli on usean selittäjän lineaarinen regressiomalli, jota kutsutaan yleiseksi lineaariseksi malliksi. (Mellin 2008.)

Malli voidaan esittää myös matriisimuodossa

$$(3.1) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

missä

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix},$$

(3.2)

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Seuraavat oletukset ovat mallin (3.1) standardioletuksia:

- (i) Selittäjämatrisin \mathbf{X} alkioit ovat ei-satunnaisia muuttujia
- (ii) Selittäjämatrisin \mathbf{X} sarakevektorit ovat lineaarisesti riippumattomia
- (iii) Virhetermin odotusarvo on nolla $E(\boldsymbol{\varepsilon}) = \mathbf{0}$
- (iv) $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$
- (v) $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$
- (vi) $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$

Standardioletuksissa vektori $\mathbf{0}$ on n -ulotteinen nollavektori $\mathbf{0} = (0, 0, \dots, 0)'$ ja matriisi \mathbf{I} on $n \times n$ yksikkömatriisi.

3.2.1 Mallin estimointi

Sopivan mallin löytämiseksi parametrit estimoidaan pienimmän neliösumman menetelmällä (PNS-menetelmä). Menetelmässä residuaalien neliösumma minimoidaan mallin parametrien suhteen. Merkitään residuaalien neliösummaa $S(\boldsymbol{\beta})$:lla, joka on muotoa

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} - \mathbf{X}'\boldsymbol{\beta}'\mathbf{y} + \mathbf{X}'\boldsymbol{\beta}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \end{aligned}$$

Kun ratkaistaan $S(\boldsymbol{\beta})$ -funktion derivaatan nollakohta $\boldsymbol{\beta}$:n suhteen, saadaan normaaliyhtälö

$$(3.3) \quad \frac{\partial S}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}.$$

Derivaatan nollakohta vastaa residuaalien neliösumman minimiä, sillä $\frac{\partial^2 S}{\partial \boldsymbol{\beta}^2} = 2\mathbf{X}'\mathbf{X}$ on positiivisesti definiitti. Yhtälöstä (3.3) voidaan ratkaista $\boldsymbol{\beta}$:n PNS-estimaattori

$$(3.4) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Koska selittävät muuttujat oletetaan lineaarisesti riippumattomiksi, niin matriisi $(\mathbf{X}'\mathbf{X})^{-1}$ on olemassa. Estimoitu malli on nyt muotoa

$$(3.5) \quad \hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y},$$

missä $n \times n$ -matriisia $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ kutsutaan hattu-matriisiksi. (Montgomery & Peck 1991.)

3.2.2 Regressiomallin selitysaste

Estimoidun mallin hyvyttä voidaan mitata mallin selitysasteen avulla. Selitysaste kertoo kuinka hyvin malli selittää selitettävän muuttujan y kokonaisvaihtelua. Selitysasteen kaava on muotoa

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$$

missä SST on kokonaisneliösumma, SSE jäännöseliösumma ja SSR regressioneliösumma. Kokonaisneliösumma kuvaa selitettävän muuttujan y arvojen kokonaisvaihtelua, jäännöseliösumma mallin residuaalien vaihtelua ja regressioneliösumma estimoitujen \hat{y} arvojen vaihtelua. SST voidaan hajottaa kahteen osaan; jäännöseliösummaan (SSE) ja regressioneliösummaan (SSR). Tätä kutsutaan varianssianalyysihajotelmaksi. SSR kuvaa kokonaisneliösummasta sitä osaa selitettävän muuttujan y vaihtelusta, jonka malli on selittänyt ja SSE osaa, jota malli ei ole selittänyt. (Mellin 2008.)

Neliösummien SSE:n ja SSR:n kaavat ovat muotoa

$$(3.6) \quad SSE = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|(\mathbf{I} - \mathbf{H})\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}},$$

$$(3.7) \quad SSR = \|\hat{\mathbf{y}} - \bar{\mathbf{y}}\|^2 = \|(\mathbf{H} - \mathbf{J})\mathbf{y}\|^2 = \mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y} = \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} - n\bar{y}^2,$$

missä $\mathbf{J} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$.

Nyt SSE:n ja SSR:n summa on SST:

$$\begin{aligned} SST &= \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} + \mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y} \\ &= \mathbf{y}'\mathbf{I}\mathbf{y} - \mathbf{y}'\mathbf{H}\mathbf{y} + \mathbf{y}'\mathbf{H}\mathbf{y} - \mathbf{y}'\mathbf{J}\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y} \end{aligned}$$

eli

$$(3.8) \quad SST = \mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y} = \mathbf{y}'\mathbf{y} - n\bar{y}^2.$$

Selitysaste R^2 voidaan esittää myös

$$(3.9) \quad R^2 = 1 - \frac{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}}{\mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y}} = \frac{\mathbf{y}'(\mathbf{H} - \mathbf{J})\mathbf{y}}{\mathbf{y}'(\mathbf{I} - \mathbf{J})\mathbf{y}}.$$

3.2.3 Hypoteesien testaus lineaarisessa regressioanalyysissä

Mallin selitysasteen lisäksi voidaan tutkia mallin selittävien muuttujien merkitsevyyttä mallissa. Kun testataan riippuuko selitettävä muuttuja y lineaarisesti selittävästä muuttujasta x_j , ovat hypoteesit muotoa

$$H_0: \beta_j = 0, j = 0, 1, 2, \dots, k$$

$$H_1: \beta_j \neq 0.$$

Mikäli selitettävän ja selittävän muuttujan välillä ei ole tilastollista riippuvuutta, jää nollahypoteesi voimaan ja tällöin on voimassa

$$t = \frac{\hat{\beta}_j}{s(\hat{\beta}_j)} \sim t(n - k - 1),$$

missä $s(\hat{\beta}_j)$ on estimoidun parametrin hajonta, n on havaintojen lukumäärä ja k mallin selittäjien lukumäärä. Nollahypoteesi H_0 hylätään riskitasolla α , jos laskettu t -arvo on suurempi kuin $t_{\frac{\alpha}{2}; n-2}$ (Mellin 2008).

Useamman selittäjän mallissa tulee tarkastella myös selittäjien yhteisvaikutusta. Tällöin tutkitaan, kuinka hyvin kaikki tarkasteltavat selittäjät yhdessä selittävät y :n kokonaisvaihtelua. Hypoteesit ovat tällöin muotoa

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \beta_k \neq 0 \text{ jollain } k: n \text{ arvolla.}$$

Jos nollahypoteesi H_0 jää voimaan, niin silloin selitettävä muuttuja y ei riipu lineaarisesti selittäjästä. Tällöin

$$F = \frac{MSR}{MSE} = \frac{n - k - 1}{k} \cdot \frac{SSR}{SSE} \sim F(k, n - k - 1),$$

missä k on mallin selittäjien lukumäärä ja n havaintojen lukumäärä. Nollahypoteesi hylätään, jos laskettu F on suurempi kuin $F_{\alpha; k, n-k-1}$ riskitasolla α . (Puntanen 1999.) Yhden selittäjän mallissa t - ja F -testillä testataan samaa hypoteesia $H_0: \beta_1 = 0$. Tällöin $t^2 = F$.

3.2.4 Regressiomalli ilman vakiotermiä

Kun halutaan muodostaa regressiomalli ilman vakiotermiä, silloin malli on muotoa

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k,$$

ja matriisimuodossa

$$(3.10) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Nyt matriiseissa \mathbf{X} ja $\boldsymbol{\beta}$ ei ole huomioitu vakiotermiä

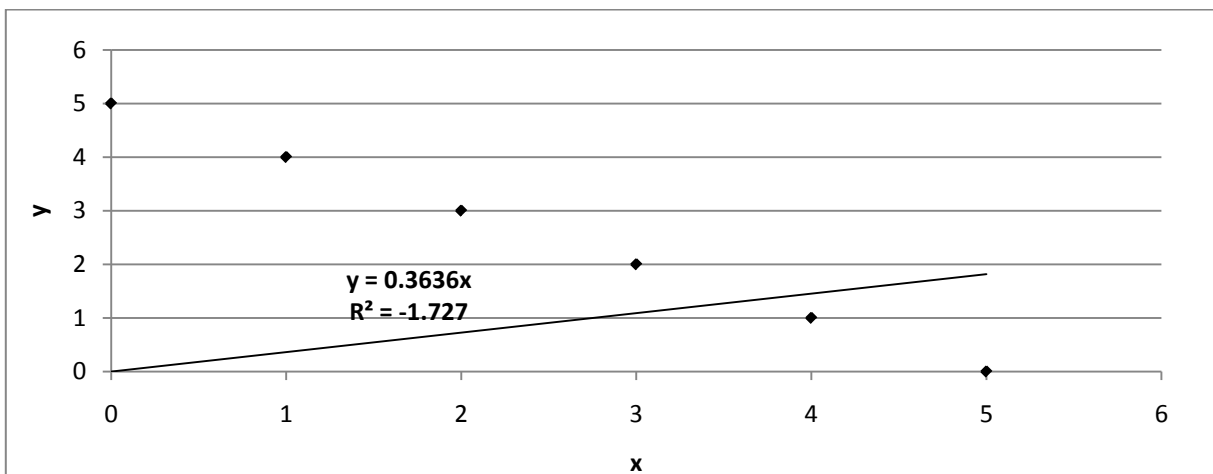
$$(3.11) \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}.$$

Estimoitu malli on muotoa

$$(3.12) \quad \hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y},$$

missä $\hat{\boldsymbol{\beta}}$ saadaan laskettua kaavasta (3.4), kun matriisit ovat muotoa (3.11).

Monissa tilanteissa vakiotermitön regressiomalli on käytännöllinen, esimerkiksi kemiallisissa ja muissa tuotantoprosesseissa. Useimmissa tilanteissa ei kuitenkaan ole syytä pakottaa regressiota kulkemaan origon kautta. Esimerkki tällaisesta tilanteesta on kuvattu Kuviossa 3.1. Kuvioista nähdään x :n ja y :n välinen negatiivinen riippuvuus. Kun tähän aineistoon estimoidaan vakiotermitön regressiomalli, on regressiosuoran kulmakerroin tällöin positiivinen. Tämä esimerkki on ääritapaus tilanteesta, jolloin ei ole järkevää pakottaa regressiota kulkemaan origon kautta. (Hahn 1977.)



Kuvio 3.1. Regressiosuoran sovittaminen origon kautta.

Kun regressiosuora pakotetaan kulkemaan origon kautta, niin yleensä suora ei mene pisteen (\bar{x}, \bar{y}) kautta. Mikäli vakiotermitön regressiomalli selittää huonosti selitettävää muuttujaa, voi silloin regressiosuoran ympärillä olla enemmän vaihtelua kuin \bar{y} :n ympärillä eli $SSE > SST$. Näin ollen mallin selitysaste voi olla negatiivinen, mikäli käytetään selitysasteen kaavaa (3.9) (Hahn 1977). Vakiotermitömän mallin varianssianalyysihajotelma eroaa laskennallisesti vakiotermitillisen mallin varianssianalyysihajotelmasta, sillä se on muotoa

(3.13)

$$\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Yhtälöstä (3.13) saadaan, että vakiotermitömän mallin tapauksessa $SST = \sum_{i=1}^n y_i^2$, $SSR = \sum_{i=1}^n \hat{y}_i^2$ ja SSE ei muutu vaan se lasketaan kaavan (3.6) mukaan (Rawlings 1988). Matriisimuodossa SST ja SSR ovat muotoa

$$(3.14) \quad SSR = \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$(3.15) \quad SST = \mathbf{y}'\mathbf{y}.$$

Näin ollen selitysasteen kaava on

$$(3.16) \quad R^2 = \frac{\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}}{\mathbf{y}'\mathbf{y}}.$$

Tätä selitysastetta ei voida verrata vakiollisen regressiomallin selitysasteen (kaava 3.9) kanssa. Myös mallien (3.5) ja (3.12) F-testisuureita ei voida verrata keskenään. Vakiotermitömän regressiomallin F-testisuuretta laskettaessa vapausasteet ovat k ja $n - k$. Näin ollen F-testisuureen kaava on muotoa

$$F = \frac{MSR}{MSE} = \frac{n - k}{k} \cdot \frac{SSR}{SSE} \sim F(k, n - k).$$

Eri tilasto-ohjelmat laskevat selitysasteen vakiotermitömän regressiomallin tapauksessa joko kaavan (3.9) tai (3.16) mukaan (Eisenhauer 2003). Esimerkiksi Excel-ohjelma laskee kaavan (3.9) mukaan, jolloin selitysaste voi olla negatiivinen, kuten Kuvion 3.1 tapauksessa. SPSS 16.0 ja R -ohjelmisto käyttävät taas kaavaa (3.16). SPSS-ohjelma on ainut näistä kolmesta ohjelmistosta, joka

antaa varoitusviestin, että tätä vakiotermitöman mallin selityssastetta ei voi verrata vakiollisen mallin selityssasteeseen. Tutkijan on oltava tarkkana vakiotermitöman regressiomallin hyvyyden tulkinnassa, sillä kaava (3.16) antaa useissa tapauksessa suuremman arvon kuin kaava (3.9). Mikäli halutaan verrata malleja (3.5) ja (3.12), on suositeltavaa vertailla keskenään esimerkiksi mallien jäännöskehajontoja.

3.3 Paloittain lineaarinen malli

Kun tutkitaan kahden muuttujan välistä riippuvuutta, ei yksinkertainen lineaarinen malli pysty aina riittävästi selittämään muuttujien välistä riippuvuutta, silloin paloittain lineaarinen malli on eräs luonnollinen vaihtoehto. Malli on muotoa

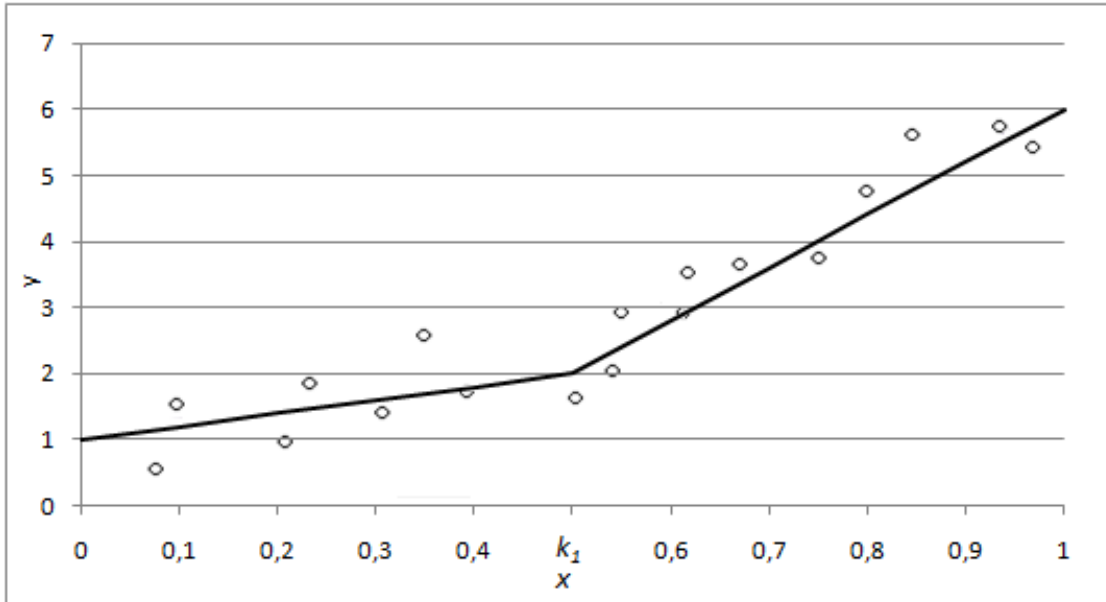
$$g(x) = \sum_{j=0}^p B_j f_j(x),$$

missä $f_j(x)$ ovat kantafunktioita, jotka ovat esimerkiksi muotoa

$$f_0(x) = 1, \quad f_1(x) = x, \quad f_2(x) = (x - k_m)_+,$$

missä $(x - k_m)_+ = (x - k_m)$, jos $x > k_m$ muuten $(x - k_m)_+ = 0$. Kantafunktiossa k_m tarkoittaa solmukohtaa. Niitä voi olla useampia havaintovälillä $[a, b]$. Solmukohdat k_1, \dots, k_m jakavat välin pienempiin osaväleihin $a < k_1 < \dots < k_m < b$. (Hastie, Tibshirani & Friedman 2009.)

Kuviossa 3.2 on esitetty paloittain lineaarinen malli, joka on muotoa $g(x) = B_0 + B_1x + B_2(x - k_1)_+$. Mallin kantafunktiot ovat siis $f_0(x) = 1$, $f_1(x) = x$, $f_2(x) = (x - k_1)_+$. Funktio $(x - k_1)_+$ on nolla, kun $x < k_1$, muuten $(x - k_1)_+ = (x - k_1)$.



Kuvio 3.2. Paloittain lineaarinen malli, jonka solmukohta on $k_1 = 0,5$.

3.4 MARS-menetelmä

MARS on lyhenne sanoista Multivariate Adaptive Regression Splines. Jerome Friedman esitteli MARS menetelmän vuonna 1991. Se on joustava regressiomenetelmä, joka sopii hyvin suuriin aineistoihin, joissa on useita selittäviä muuttujia. MARS-menetelmä perustuu paloittain lineaariseen malliin. Menetelmän etu on siinä, että MARS määrittää kantafunktioiden solmukohdat automaattisesti suoraan datasta. (Friedman 1991.)

3.4.1 MARS-malli

Tarkastellaan tässä aluvussa MARS-mallin muodostamista. Mallin systemaattinen osa on muotoa

(3.17)

$$g(x) = B_0 + \sum_{m=1}^M B_m f_m,$$

missä B_0 on vakiotermi, ja f_m :t ovat kantafunktioita. Kertoimet B_m estimoidaan pienimmän neliösumman menetelmällä eli samalla tavalla kuin lineaarisessa regressioanalyysissä (Faraway 2006). Kantafunktiot valitaan kokoelmasta C , joka on muotoa

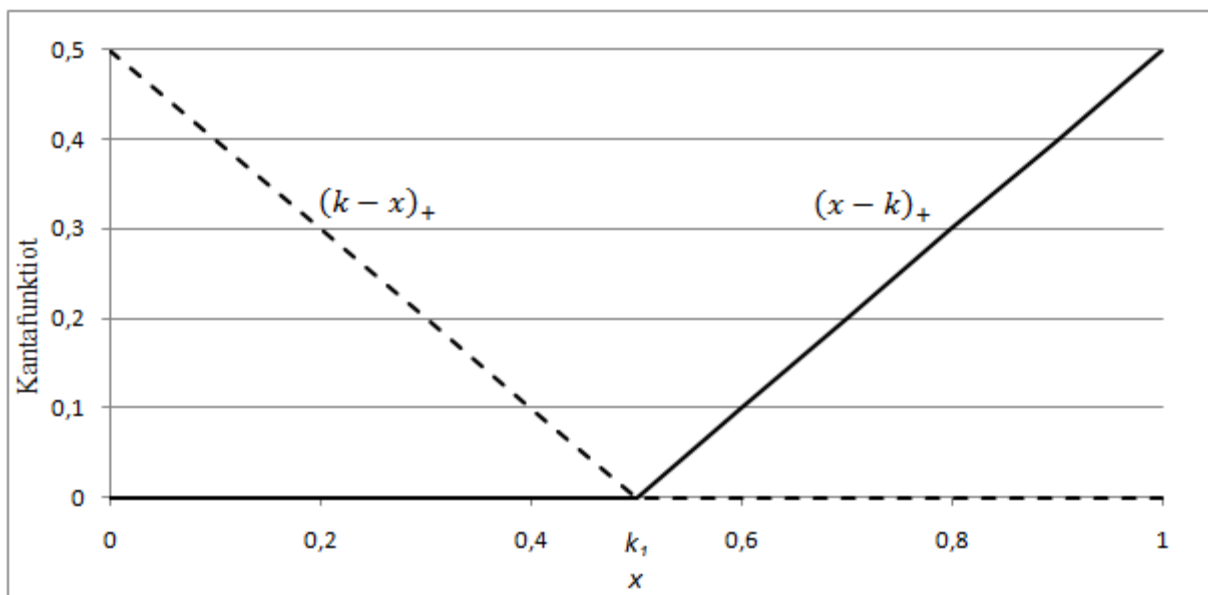
$$C = \{(x_j - k)_+, (k - x_j)_+\}_{k \in \{x_{1j}, x_{2j}, \dots, x_{Nj}\}, j=1,2,\dots,p}$$

missä k on muuttujan x_j solmukohta, p on selittäjien lukumäärä ja N mahdollisten solmukohtien lukumäärä. Paloittain lineaariset selittäjät ovat muotoa

$$(x_j - k)_+ = \begin{cases} x_j - k, & \text{kun } x_j - k > 0 \\ 0, & \text{muulloin} \end{cases} \quad \text{ja} \quad (k - x_j)_+ = \begin{cases} k - x_j, & \text{kun } k - x_j > 0 \\ 0, & \text{muulloin.} \end{cases}$$

Funktio f_m voi olla myös joukon C kahden tai useamman funktion tulo.

MARS-mallin muodostamisessa tärkeä osa on solmukohtien määrän ja paikan valinta, sillä mikäli solmukohdat on huonosti valittu, voidaan menettää sovitteesta tärkeitä yksityiskohtia. MARS käy läpi kaikkien muuttujien mahdolliset solmukohdat ja näin löytää parhaiten sopivan kantafunktion malliin. Näitä kantafunktioita $(x - k)_+$ ja $(k - x)_+$ kutsutaan reflektiivisiksi pareiksi. Kuviossa 3.3 on esitetty reflektiiviset parit $(x - k)_+$ ja $(k - x)_+$, kun k on 0,5. Solmukohdan k kohdalla $(x - k)_+$ -funktio on nolla sekä kun $x < k$. Samoin $(k - x)_+$ -funktio on nolla k :n kohdalla ja, kun $x > k$.



Kuvio 3.3. Reflektiiviset parit, kun $k = 0,5$.

MARS-malli rakennetaan vaiheittain ja mallin valinta tehdään samaan tapaan kuin eteenpäin askeltavassa regressioanalyysissä. Alkutilanteessa on mukana vain vakiotermi. Seuraavaksi etsitään kaikista muuttujista mahdolliset solmukohdat ja valitaan kantafunktio, joka sopii parhaiten aineistoon. Tämän jälkeen toistetaan samaa prosessia ja valitaan paras kantafunktio, kun valitut kantafunktiot ovat jo mukana mallissa. (Hastie et al. 2009; Faraway 2006.)

Kun kaikki kantafunktiot on valittu, saatu malli on yleensä ylisovitettu. Tämän jälkeen mallista poistetaan muuttuja siten, että ensiksi poistetaan se muuttuja, jonka poistaminen pienentää selitystasetta vähiten. Tätä toistetaan, kunnes löydetään malli, jossa on sopiva määrä kantafunktioita. MARS-menetelmä valitsee kantafunktioiden lukumäärän malliin yleistetyn ristiinvalidointikriteerin (GVC) avulla. (Hastie et al. 2009.) Optimaalinen malli etsitään minimoimalla yleistetyn ristiinvalidointikriteerin arvo. Yleistetyn ristiinvalidointikriteerin kaava on muotoa

$$GVC(M) = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{g}(x_i))^2}{\left(1 - \frac{C(M)}{N}\right)^2}.$$

Kaavassa y_i :t ovat selitettävän muuttujan y havaitut arvot, $\hat{g}(x_i)$ mallin (3.17) antamat sovitearvot, N havaintojen lukumäärä ja $C(M)$ malliin $\hat{g}(x)$ sovitettujen parametrien lukumäärä. Yleistetty ristiinvalidointikriteeri ei ota huomioon vain residuaalien neliöiden keskiarvoa vaan myös mallin monimutkaisuuden. (Friedman 1991.)

Kantafunktioiden lukumäärien valintaa voidaan rajoittaa erilaisilla säännöillä. Muuttujien yhdysvaikutuksia voidaan myös säännellä, mm. voidaan sallia yhdysvaikutuksia tai estää ne kokonaan. R 2.12.1 ja STATISTICA 8.0 -ohjelmat antavat käyttäjän valita kuinka monta termiä saa malliin maksimissaan tulla, ja käyttäjä saa itse valita sallitaanko muuttujien yhdysvaikutuksia vai ei. R-ohjelmassa MARS-malli voidaan sovittaa aineistoon *mars*-funktiolla, joka löytyy *mda*-kirjastosta. Havainnollistetaan MARS-mallinnusta seuraavassa esimerkissä.

Esimerkki 3.1.

Sovitetaan *ozone*-aineistoon MARS-malli, jossa selitettävänä muuttujana on otsonipitoisuus ja mahdollisia selittäjiä ovat lämpötila, tuulen voimakkuus ja auringonsäteily. Aluksi ladataan *mda*-kirjasto [`library(mda)`], josta *mars*-funktio löytyy.

MARS-malli, johon asetetaan termien lukumääräksi neljä ($nk = 4$), ja ei sallita yhdysvaikutuksia, saadaan muodostettua *mars*-funktiolla:

```
m<- mars(ozone[, - 1], ozone[, 1], nk=4)
```

Malliin tulevat muuttujat saadaan seuraavalla komennolla:

```
m$factor[m$Ss, ]
      radiation  temperature  wind
[1, ]         0             0     0
[2, ]         0             1     0
[3, ]         0            -1     0 .
```

Otsonipitoisuutta selittää *temperature*-muuttuja. Tulostuksessa ensimmäinen rivi vastaa vakiotermejä. Toisella rivillä *temperature*-muuttujan kohdalla ”1” tarkoittaa, että reflektiivisistä pareista positiivinen osa ($temperature - k$)₊ tulee malliin ja kolmannen rivin ”-1” tarkoittaa negatiivista osaa ($k - temperature$)₊. Muuttujan solmukohta *k* saadaan komennolla:

```
m$cuts[m$Ss, ]
      [, 1] [, 2] [, 3]
[1, ]    0    0    0
[2, ]    0   85    0
[3, ]    0   85    0 .
```

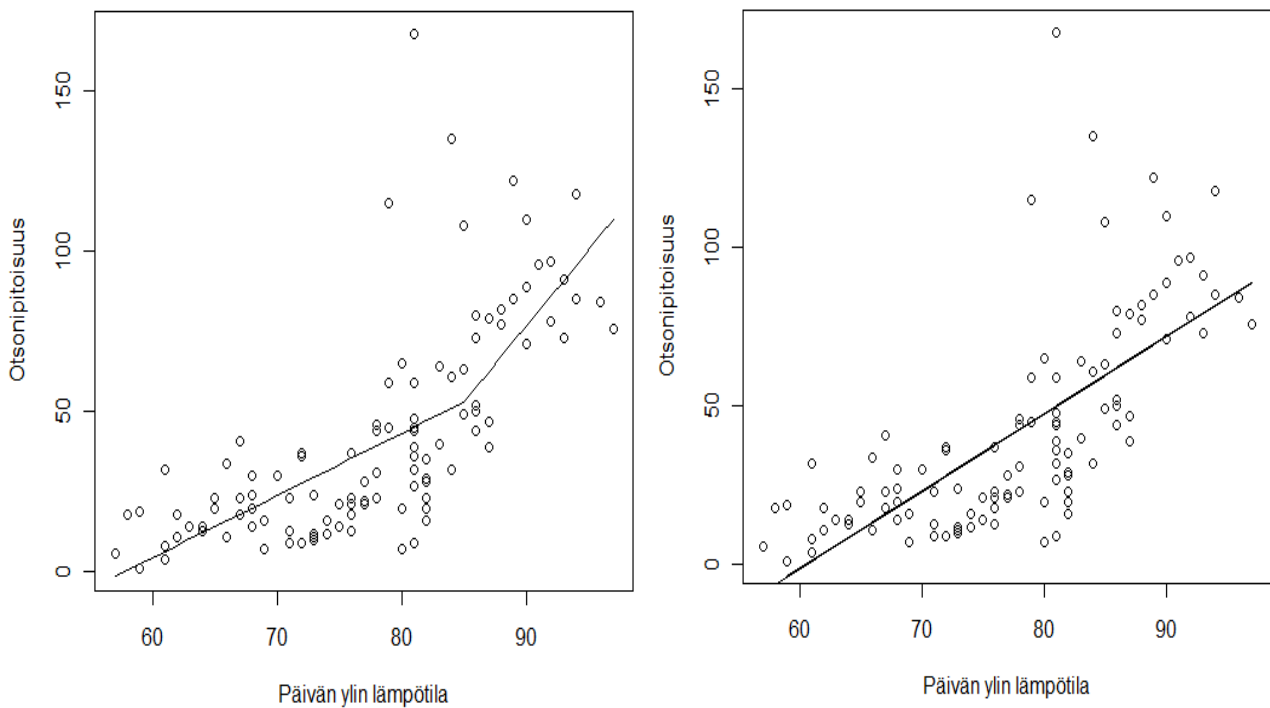
Temperature-muuttujan solmukohdaksi MARS-menetelmä saa 85. Mallin kertoimien estimaatit ovat muotoa

```
m$coef
      [, 1]
[1, ] 52.978290
[2, ]  4.738286
[3, ] -1.942790.
```

Saatu malli on siis muotoa

$$(3.18) \text{ ozone} = 52,98 + 4,74 \times (\text{temperature} - 85)_+ - 1,94 \times (85 - \text{temperature})_+$$

Temperature-muuttuja lisää selkeästi enemmän otsonipitoisuutta, kun sen arvo on yli 85. Kun päivän ylin lämpötila on 90 astetta, otsonipitoisuuden arvoksi saadaan 76,68, ja kun lämpötila on 10 astetta vähemmän, otsonipitoisuus on 43,28. MARS-mallin residuaalien hajonta on 23,38, joka on hieman pienempää kuin tavallisen regressiomallin (23,92), missä otsonikerrosta selitetään *temperature*-muuttujalla. Kuviosta 3.4 nähdään MARS-mallin sovitekäyrän ja regressiomallin regressiosuoran eroavaisuudet.



Kuvio 3.4. MARS-mallin ja regressiomallin vertailua.

3.4.2 MARS-mallin selitysaste

MARS-mallin selitysaste määritetään samalla kaavalla kuin regressioanalyysissä (kaava 3.9). Myös jäännös-, kokonais- ja regressioneliösumma lasketaan (3.6) - (3.8) kaavojen mukaan. Tässä aluvussa keskitytään MARS-mallin selitysasteen määrittämiseen R-ohjelmalla.

Tarkastellaan esimerkin 3.1 MARS-mallin selitystettä Farawayn (2006) kirjan sivun 272 esimerkin mukaan. Kirjan mukaan MARS-mallin yhteenvedon saa R-ohjelmassa *lm*-funktion avulla niin, että selitetään mallin selitettävää muuttujaa *y*:tä MARS-mallin kantafunktioilla. Muodostetaan mallin (3.18) yhteenvedo kyseisen esimerkin mukaan. Tällöin R-ohjelman komennot ovat muotoa:

```
marsmalli <- lm(data[, 1] ~ m$x - 1)  
summary(marsmalli)
```

Mallin F-testisuureksi saadaan 158,2 vapausasteiden ollessa 3 ja 108, sekä selitysasteeksi 0,81. On syytä epäillä, että tämä komento ei anna tarkoitettua mallin oikeaa F-testi tulosta, sillä vapausasteiden pitäisi tässä tapauksessa olla 2 ja 108. Kun tutkitaan tarkemmin R-ohjelman komentoa, huomataan virhe, joka vaikuttaa vapausasteiden lukumäärän suuruuteen. Matriisin *m*\$x ensimmäinen sarake on yksikkövektori, joka vastaa mallin vakioselittäjää, toinen sarake on reflektiivisistä pareista positiivinen osa (*temperature* - 85)₊ ja kolmas negatiivinen osa (85 - *temperature*)₊. Selittävinä muuttujina ovat nyt vakiotermi ja nämä reflektiiviset parit, minkä takia vapausaste on kolme eikä kaksi.

Farawayn kirjan esimerkissä saatu MARS-mallin selitysaste on hyvin korkea, ja kirjan mukaan tämä malli selittää hyvin selitettävän muuttujan vaihtelua. Faraway myös vertaa MARS-mallin ja additiivisen mallin selitysasteita toisiinsa ja tekee johtopäätöksen, että MARS-malli sopii paremmin aineistoon kuin additiivinen malli. Kun tarkastellaan vielä tarkemmin R-ohjelman komentoa, huomataan toinen virhe, jonka takia mallille saadaan korkea selitysaste. Tämä johtuu komennossa "**lm**(data[, 1] ~ **m**\$x - 1)" olevasta -1:stä, joka tekee regressiomallin ilman vakiotermiä. Näin ollen selitysaste lasketaan kaavan (3.16) mukaan, sillä nyt on kyseessä regressiomalli, joka kulkee origon kautta. Tämän takia saatu selitysaste on korkea ja sitä ei voi verrata selitysasteeseen (3.9) (kts. luku 3.2.4).

MARS-mallin oikean selityksasteen ja F-testisuureen saa R-ohjelmalla niin, että muodostetaan malli, missä *ozone*-muuttujaa selitetään vain $(temperature - 85)_+$ ja $(85 - temperature)_+$ funktioilla. Esimerkin 3.1 MARS-mallin yhteenveto saadaan ao. komennolla, jossa tämä pieni R-koodivirhe `m$X-1` on korjattu `m$X[, -1]`:

```
marsmalli <- lm(data[, 1] ~ m$X[, -1])  
summary(marsmalli)
```

Nyt saadaan mallin selityksasteeksi 0,51, eli noin 0,3 yksikköä pienempi kuin Farawayn kirjan esimerkin mukaan saatu selityksaste. Myös F-testisuure on selkeästi pienempi 57,42. Tämän MARS-mallin selityksaste on hieman suurempi kuin tavallisen regressiomallin (0,48), missä otsonikerrosta selitetään *temperature*-muuttujalla.

Tätä Farawayn kirjaa monet ovat käyttäneet lähdekirjana. Esimerkkejä MARS-mallin muodostamisesta R-ohjelmalla ei juuri muista kirjoista löydy, joten tämän takia on voitu tulkita MARS-mallin paremmuutta virheellisesti. Kirjan hyvin pienen R-koodivirheen takia voi vahingossa tulkita mallin selityksasteen jopa kaksi kertaa suuremmaksi kuin mitä se on todellisuudessa.

3.4.3 MARS-mallin muodostaminen R-ohjelman *earth*-funktiolla

R-ohjelmassa MARS-mallin voi muodostaa myös *earth*-funktiolla, joka löytyy *earth*-paketista R-ohjelman 2.4.7 versiosta lähtien. Funktio on johdettu *mars*-funktion koodista ja siihen on lisätty muutamia erityistoimintoja. *Earth*-funktio antaa käyttäjälle vapaammat kädet MARS-mallia muodostaessa kuin *mars*-funktio. Muodostetaan MARS-malli *earth*-funktoita käyttäen, jossa selitettävänä muuttujana on otsonipitoisuus. Mallin termien lukumääräksi sallitaan neljä ($nk = 4$), mutta yhdysvaikutuksia ei lainkaan kuten esimerkissä 3.1.

MARS-mallin muodostaminen *earth*-funktiolla:

```
m2 <- earth(data[, -1], data[, 1], nk=4)
summary(m2)
Call: earth(x=data[, -1], y=data[, 1], nk=4)

              coefficients
(Intercept)      20.921014
h(temperature-76) 4.305435

Selected 2 of 3 terms, and 1 of 3 predictors
Importance: temperature, radiation-unused, wind-unused
Number of terms at each degree of interaction: 1 1 (additive model)
GCV 518.7273    RSS 54508.43    GRSq 0.5357548    RSq 0.552483
```

Malli on muotoa

$$(3.19) \quad \textit{ozone} = 20,92 + 4,31 \times (\textit{temperature} - 76)_+$$

Malli eroaa jonkin verran mallista (3.18), jossa *temperature*-muuttujan solmukohta on 85 ja mallissa (3.18) on reflektiivisistä pareista mukana myös negatiivinen osa. Mallin (3.19) selitysaste (0,55) on hieman parempi kuin mallin (3.18). Myös mallin residuaalien hajonta 22,36 on pienempi kuin mallin (3.18).

Funktiot *mars* ja *earth* toteuttavat MARS-mallin muodostamisen erilailla, josta johtuvat mallien (3.18) ja (3.19) eroavaisuus. Trevor Hastie ja Robert Tibshirani ovat tehneet *mars*-funktion R-ohjelman versioon 0.4.1. Tämä *mars*-funktio valitsee muuttujan solmukohdat hieman erilailla miten Jerome Friedman (1991) aikkelissaan "Multivariate Adaptive Regression Splines". STATISTICA 8.0-ohjelma muodostaa MARS-mallin samalla tavalla kuin *mars*-funktio. *Earth*-funktio muodostaa MARS-mallin lähes kokonaan Friedmanin artikkelin mukaan.

Tehdään vielä *earth*-funktiolla MARS-malli, jossa MARS-menetelmä ei karsi alkuperäisiä valittuja termejä vaan pitää ne mallissa. Sallitaan kuitenkin malliin termien lukumääräksi vain neljä ($nk = 4$), mutta ei sallita yhdysvaikutuksia. Tällainen malli saadaan muodostamalla *earth*-funktiolla:


```

m3 <- earth(data[, -1], data[, 1], nk=4, pmethod="none")
summary(m3)
Call: earth(x=data[, -1], y=data[, 1], nk=4, pmethod="none")

              coefficients
(Intercept)      24.4440638
h(temperature-76) 4.0008773
h(76-temperature) -0.6477525

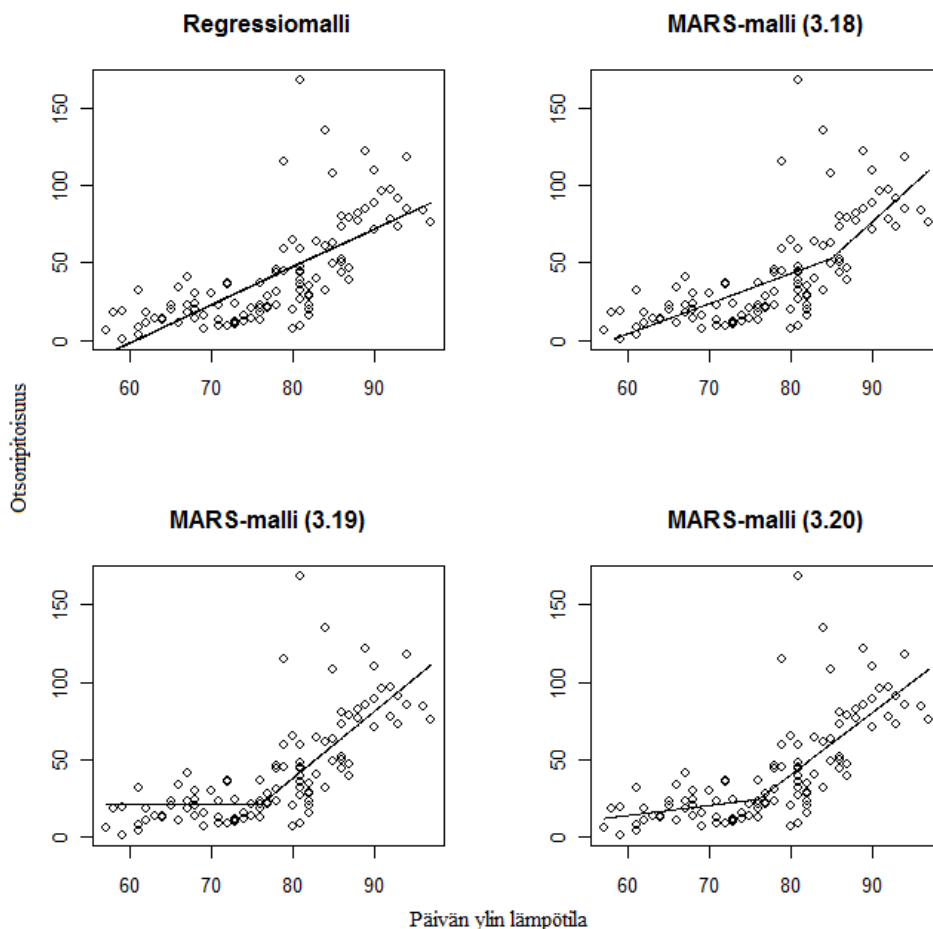
Selected 3 of 3 terms, and 1 of 3 predictors
Importance: temperature, radiation-unused, wind-unused
Number of terms at each degree of interaction: 1 2 (additive model)
GCV 529.6009    RSS 53608.97    GRSq 0.5260233    RSq 0.5598676

```

Tämä malli on muotoa

$$(3.20) \text{ ozone} = 24,44 + 4,00 \times (\text{temperature} - 76)_+ - 0,65 \times (76 - \text{temperature})_+$$

Mallin selitysaste on sama kuin mallin (3.19) eli 0,55, mutta residuaalien hajonta on hieman pienempi (22,28). Kuvista 3.5 nähdään MARS-mallien sovitekäyrien ja regressiomallin regressiosuoran eroavaisuudet. Kuvien perusteella on vaikea sanoa tarkasti, mikä MARS-malleista sopii parhaiten aineistoon. Silmämääräisesti voidaan kuitenkin arvioida, että malli (3.20) mallintaisi parhaiten otsonipitoisuutta.



Kuvio 3.5. MARS-mallien ja regressiomallin vertailua.

4 Liikuntakertojen sekä liikuntalajien vaikutus kuntoarvioihin

Kestävyysharjoittelu kehittää hengitys- ja verenkiertoelimistön kuntoa sekä parantaa lihasten energiantuottoa ja hermoston toimintaa. Hyvä aerobinen kestävyys vähentää sairastumisriskiä mm. sydän- ja verisuonitauteihin sekä saattaa pidentää terveen elinajan odotetta. Kestävyysliikuntaa tulisi harrastaa 3 - 5 kertaa viikossa noin 20 - 60 minuuttia kerrallaan, jotta se edistäisi terveyttä. (Keskinen et al. 2004.) Jo edellä mainitun THL:n tutkimuksen mukaan suomalaisista miehistä 29 % ja naisista 34 % harrastaa liikuntaa vapaa-ajallaan vähintään neljä kertaa viikossa, joka on kuntoliikuntasuositus.

Tässä luvussa tarkastellaan liikuntakertojen sekä liikuntalajien vaikutusta testistä saatuihin kuntoarvioihin käyttäen Pearsonin χ^2 -testiä. Testin nollahypoteesiksi asetetaan oletus, että liikuntakerroilla tai liikuntalajeilla ei ole vaikutusta kuntoarvioon. Testistä saadaan kuntoarviot kestävyysalueiden eri osaluueille, jotka ovat perus-, vauhti- ja maksimikestävyys. Nollahypoteesi hylätään kaikilla näillä osaluueille (kaikki p-arvot = 0). Voidaan siis todeta, että liikuntakerroilla sekä liikuntalajeilla on vaikutusta perus-, vauhti- ja maksimikestävyysalueeseen. Jokaista eri osa-aluetta tarkastellaan vielä seuraavissa alaluvuissa.

4.1 Liikuntakertojen vaikutus eri kestävyysalueiden arvioihin

Testissä kävijät liikkuvat keskimäärin 1 - 2 kertaa viikossa. Näistä 20 % ei liiku lainkaan tai vain satunnaisesti. Lähes puolet testissä kävijöistä harrastaa liikuntaa enemmän kuin kolme kertaa viikossa.

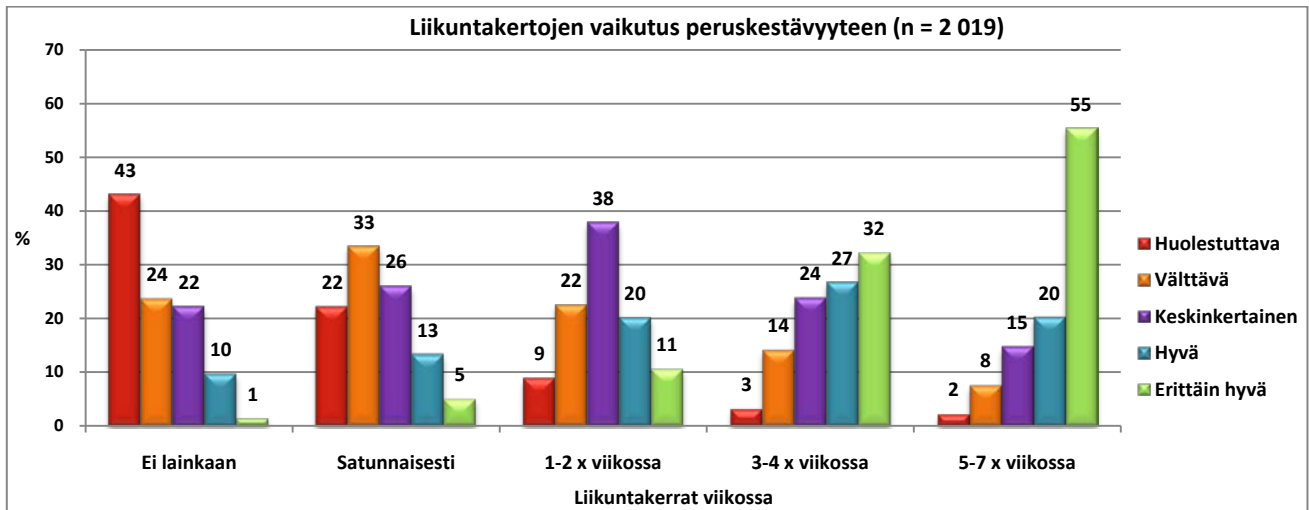
Pearsonin χ^2 -testin perusteella sukupuolella ei ole vaikutusta liikuntakertoihin (p-arvo = 0,19) eli naiset ja miehet harrastavat liikuntaa yhtä usein viikossa, kun taas iällä on tilastollisesti merkittävä vaikutus liikuntakertoihin (p-arvo = 0). Alle 30-vuotiaista 54 % ja yli 50-vuotiaista 46 % liikkuu yli kolme kertaa viikossa.

4.1.1 Liikuntakertojen vaikutus peruskestävyyteen

Tarkastellaan ensin liikuntakertojen vaikutusta peruskestävyyteen. χ^2 -arvon avulla saadaan lasketua kontingenssikerroin, joka kuvaa muuttujien välistä riippuvuutta. Kertoimeksi saatiin 0,47, eli liikuntakerroilla ja peruskestävyydellä on kohtalainen riippuvuus. Testatuista 22 % sai erittäin hyvän, 21 % hyvän, 27 % keskinkertaisen, 20 % välttävän ja 10 % huolestuttavan peruskestävyyden arvon.

Kuviosta 4.1 huomataan, että suurin osa testatuista sai vastaavan arvion peruskestävyydestään, kun otetaan huomioon liikunnan määrä viikossa. Liikuntaa harrastamattomista 67 % ja viikossa 1 - 2 kertaa tai vähemmän liikkuvista 41 % sai kuntoarviokseen enintään välttävän.

Enemmän kuin viisi kertaa viikossa liikkuvista 75 % sai testissä arviokseen hyvän tai erittäin hyvän, kun viikossa 3 - 4 kertaa tai enemmän liikuntaa harrastavista 64 % sai vastaavat arviot. Näistä prosenttiluvuista voidaan todeta liikuntakertojen suuri vaikutus peruskestävyyden kuntoarviioihin.



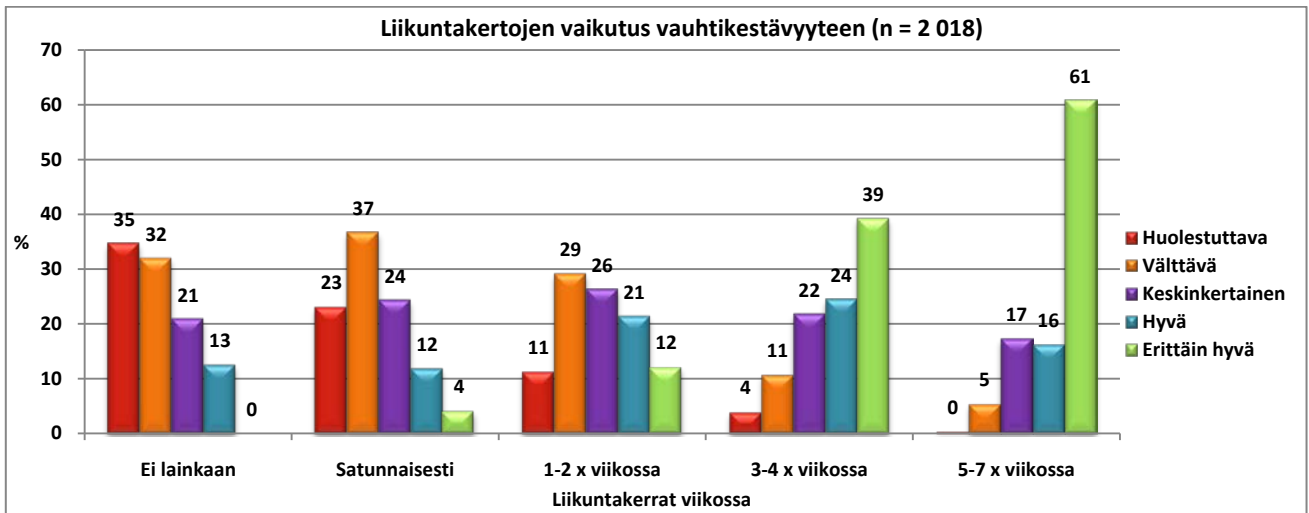
Kuvio 4.1. Liikuntakertojen vaikutus peruskestävyyteen.

4.1.2 Liikuntakertojen vaikutus vauhtikestävyyteen

Kontingenssikertoimeksi saadaan nyt 0,48, joten riippuvuus liikuntakertojen ja vauhtikestävyyden välillä on hieman suurempaa kuin liikuntakertojen ja peruskestävyyden välinen riippuvuus. Testatuista 26 % sai vauhtikestävyyden arviokseen erittäin hyvän, 20 % hyvän, 23 % keskinertaisen, 21 % välttävän ja 10 % huolestuttavan.

Liikuntakertojen vaikutus vauhtikestävyyteen ja peruskestävyyteen näyttäisi olevan lähes sama, kun verrataan Kuvioita 4.1 ja 4.2. Liikuntaa harrastamattomista 67 % sai kuntoarviokseen enintään välttävän. Viikossa 1 - 2 kertaa tai vähemmän liikuntaa harrastavista 48 % sai vauhtikestävyyden arviokseen huolestuttavan tai välttävän, kun saman verran liikkuvista 41 % sai peruskestävyydestä vastaavat arviot.

Jopa 77 % testatuista, jotka liikkuvat enemmän kuin viisi kertaa viikossa, sai testissä arviokseen hyvän tai erittäin hyvän, kun viikossa 3 - 4 kertaa tai enemmän liikkuvista 68 % sai samat arviot. Näistä prosenttiluvuista voidaan todeta, että liikuntakertojen määrä vaikuttaa myös vauhtikestävyyteen.

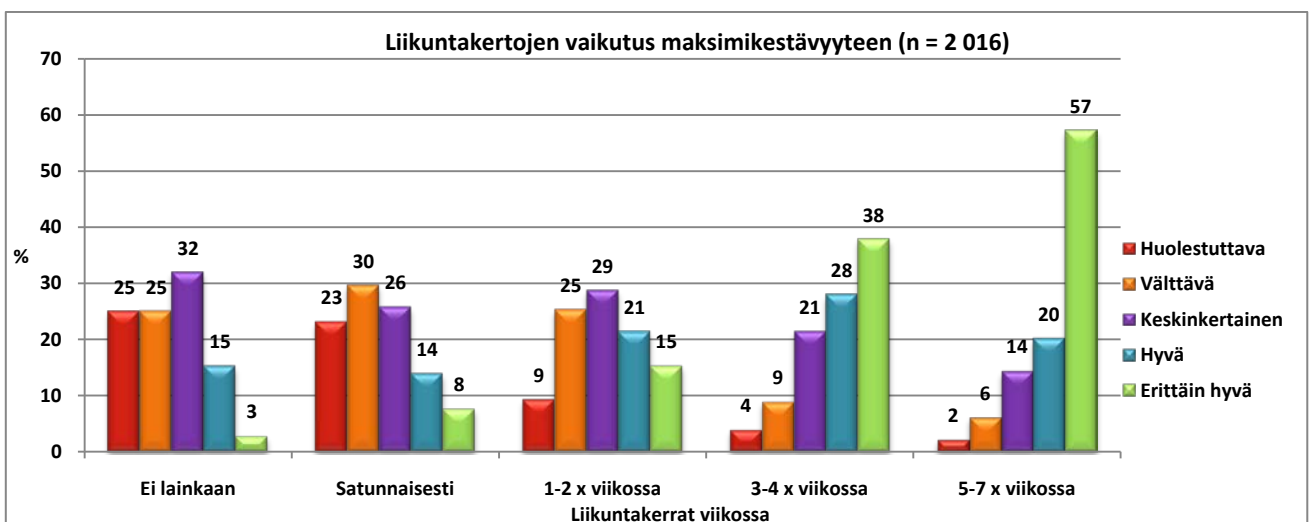


Kuvio 4.2. Liikuntakertojen vaikutus vauhtikestävyteen.

4.1.3 Liikuntakertojen vaikutus maksimikestävyteen

Liikuntakertojen ja maksimikestävyden välillä on myös kohtalaista riippuvuutta, sillä kontingenssikertoimeksi saadaan 0,44. Testatuista 27 % sai maksimikestävyden arviokseen erittäin hyvän, 22 % hyvän, 24 % keskinkertaisen, 18 % välttävän ja 9 % huolestuttavan.

Kuvion 4.3 perusteella näyttää, että 1 - 2 kertaa viikossa tai sitä vähemmän liikkuvilla liikuntakertojen määrä ei juuri vaikuta maksimikestävyden arvioihin. Vain puolet liikuntaa harrastamattomista sai kuntoarvioksi huolestuttavan tai välttävän. Viikossa 1 - 2 kertaa tai vähemmän liikuntaa harrastavista jopa 31 % sai maksimikestävyden arviokseen hyvän tai erittäin hyvän, kun saman verran liikkuvista 26 % sai peruskestävyydestä vastaavat arviot. Enemmän kuin viisi kertaa viikossa liikuntaa harrastavista saivat 77 % arviokseen hyvän tai erittäin hyvän. Kuvioita 4.1 - 4.3 vertailtaessa voidaan todeta, että liikuntakerrat eivät vaikuta maksimikestävyteen niin paljon kuin ne vaikuttavat perus- ja vauhtikestävyteen.

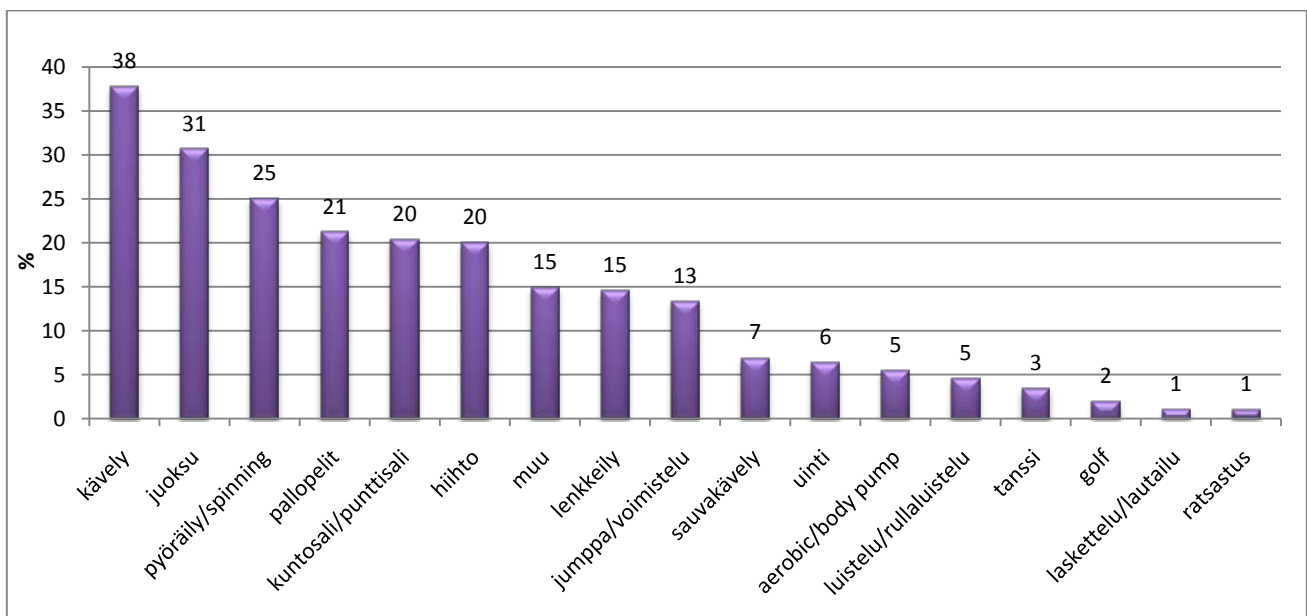


Kuvio 4.3. Liikuntakertojen vaikutus maksimikestävyteen.

4.2 Liikuntalajien vaikutus kestävyysalueiden kuntoarvoihin

Tutkimusaineistoon testatun liikuntalaji piti muuttua numeeriseksi muuttujaksi, sillä taustatietolomakkeessa testatun liikuntatottumukset on kysytty avoimilla kysymyksillä. Liikuntalajimuuttuja on siis uudelleen koodattu numeeriseksi muuttujaksi siten, että yksi vastaa kävelyä, kaksi pyöräilyä jne. Osa testatuista on voinut ilmoittaa useammankin liikuntalajin, joten liikuntalajimuuttuja on aineistossa multiresponse-tyyppinen muuttuja, eli se voi saada enemmän kuin yhden arvon. Osa testattavista on saattanut ilmoittaa harrastaneensa jopa kuutta eri liikuntalajia. Keskimäärin testattavat harrastavat 2 - 3 eri liikuntalajia. Testatuista 35 % on ilmoittanut harrastavansa vain yhtä liikuntalajia, 32 % harrastaa kahta ja kolmea eri liikuntalajia harrastaa 21 %. Kuviosta 4.4 nähdään eri lajien peittoprosentit. Peittoprosentti tarkoittaa tässä tilanteessa sitä, kuinka suuri osa testatuista harrastaa kyseistä liikuntalajia.

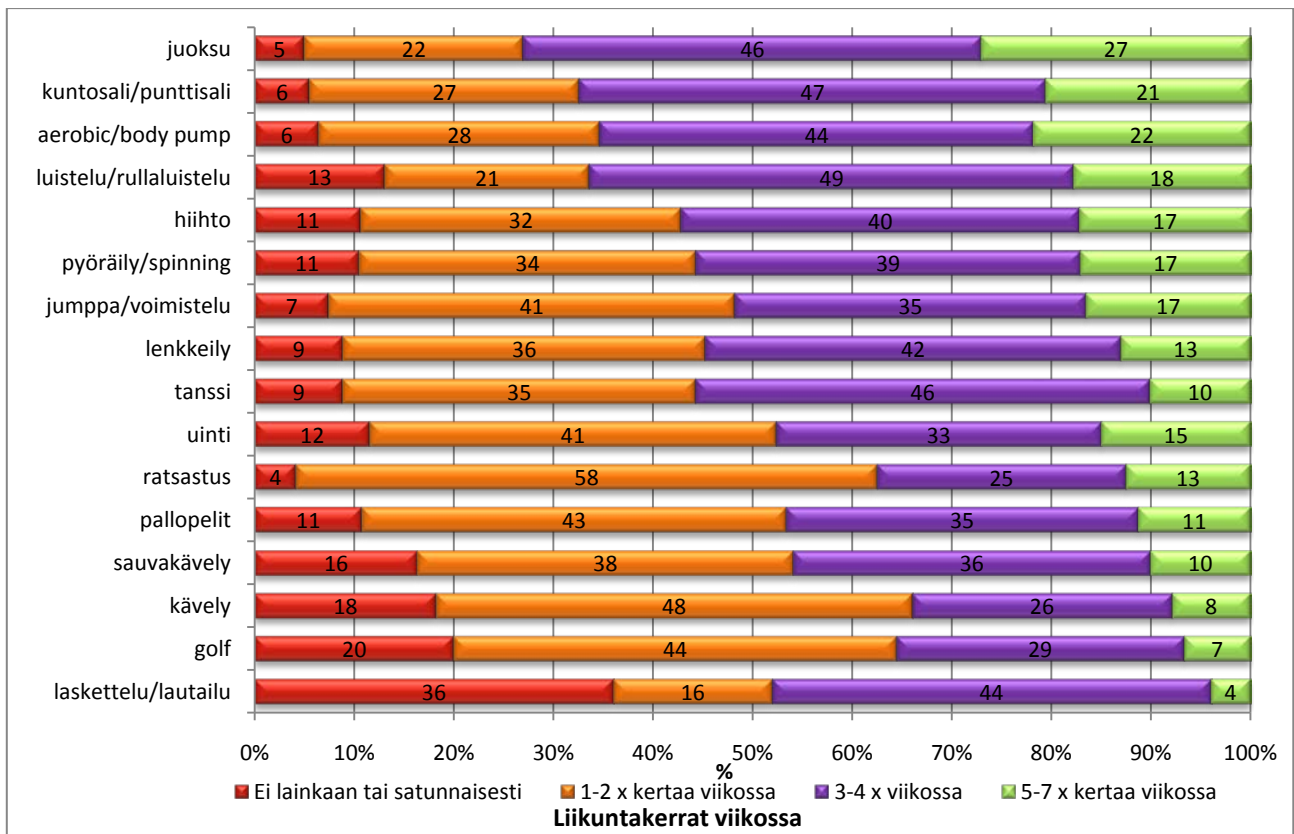
Suosituin liikuntalaji testattavien joukossa on kävely (38 %). Kuviosta huomataan, että juoksu ja lenkkeily on määritelty erikseen. Tämä johtuu siitä, että jollekin testattavalle lenkkeily voi tarkoittaa reipasta kävelyä ja juoksu reipasta juoksua, joten päätettiin pitää nämä kaksi liikuntalajia erillään.



Kuvio 4.4. Liikuntalajien suosio testissä kävijöiden keskuudessa.

Pearsonin χ^2 -testin perusteella sukupuolella on tilastollisesti erittäin merkitsevä vaikutus liikuntalajeihin (p -arvo = 0) ja kontingenssikerroin on 0,37, joten muuttujien välinen riippuvuus on kohtalaista. Naiset harrastavat selkeästi enemmän kävelyä, sauvakävelyä, tanssia, jumppaa, sekä aerobicia kuin miehet, joiden lajit ovat pallopelit, golf ja juokseminen. Myös iällä on tilastollisesti erittäin merkitsevä vaikutus liikuntalajeihin (p -arvo = 0). Alle 40-vuotiaat harrastavat eniten pallopelejä,

aerobicia, ratsastusta ja käyvät selkeästi eniten kuntosalilla, kun taas yli 40-vuotiaat golfaavat, jumppaavat, sauvakävelevät sekä kävelevät. Pearsonin χ^2 -testin perusteella liikuntalajien ja liikuntakertojen välillä on erittäin merkitsevä tilastollinen riippuvuus (p-arvo = 0). Muuttujien välinen riippuvuus ei ole kovin voimakas, sillä saatu kontingenssikerroin on vain 0,28. Kuviosta 4.5 nähdään, että juoksua harrastavat harjoittelevat eniten keskimäärin 3 - 4 kertaa viikossa ja kävelyä harrastetaan eniten 1 - 2 kertaa viikossa tai vähemmän (66 %).



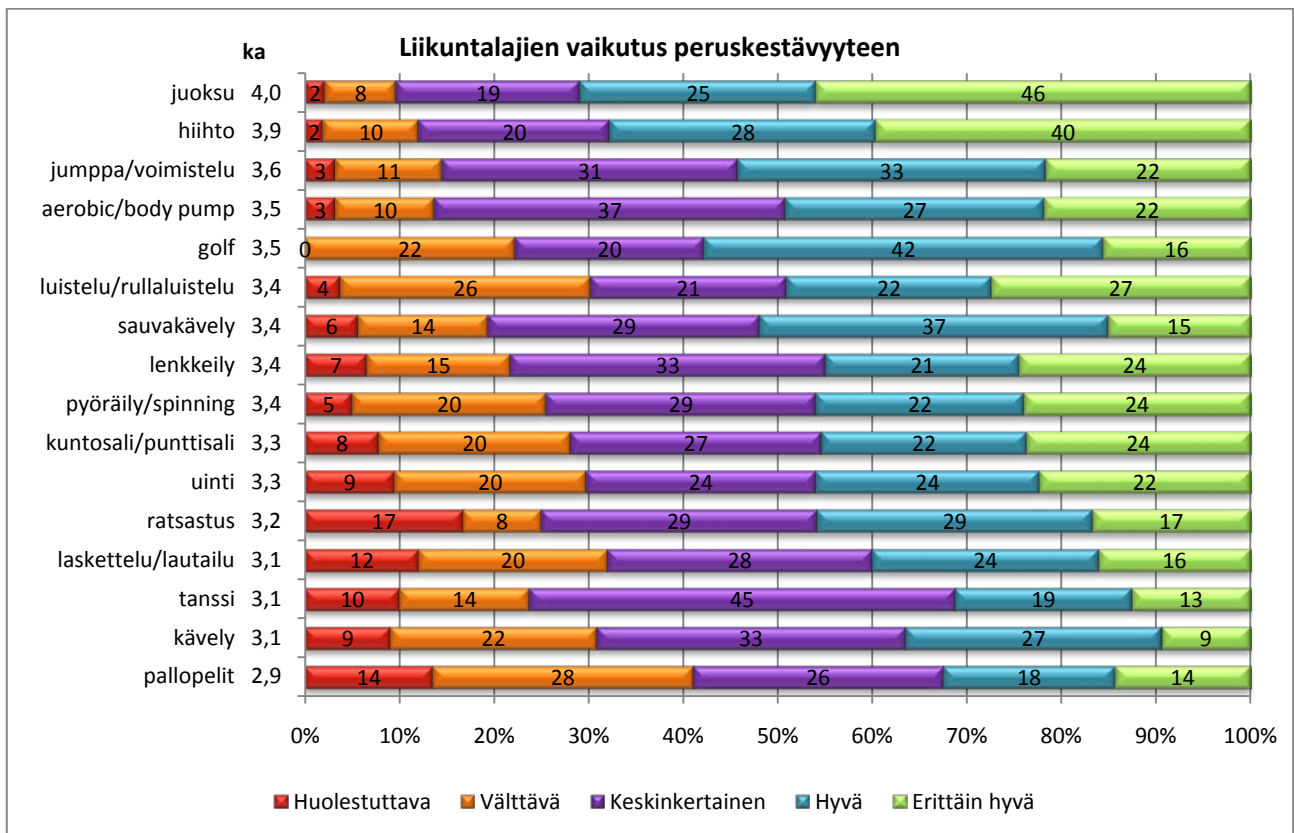
Kuvio 4.5. Liikuntakertojen vaikutus liikuntalajeihin.

4.2.1 Liikuntalajien vaikutus peruskestävyyteen

Liikuntalajit vaikuttavat tilastollisesti erittäin merkitsevästi peruskestävyyteen ($\chi^2 = 618,28$; $df = 68$, p -arvo = 0) ja niiden välinen riippuvuus on kohtalaista (kontingenssikerroin = 0,33). Kuviota 4.6:tta tarkasteltaessa on hyvä muistaa, että testattava voi harrastaa useampaa kuin yhtä liikuntalajia. Eli mikäli testattava harrastaa juoksua ja jumppaa, hänen arvionsa on laskettu mukaan sekä juoksun että jumpan prosenttiosuuksiin. Kuviossa 4.6 on esillä myös peruskestävyysarvion keskiarvo (Huolestuttava = 1, ... , Erittäin hyvä = 5), minkä avulla lajit ovat järjestetty kuviossa.

Juoksua harrastavat saavat jopa 71 %:n todennäköisyydellä hyvän tai erittäin hyvän arvion, kun pallopelejä harrastavista vain 32 % saa vastaavat arviot. Pallopelejä harrastavat saavat todennäköisimmin välttävän tai huolestuttavan arvion (42 %) verrattuna muihin lajeihin. Suosituimmat lajiyh-

distelmät ovat kävely ja pyöräily sekä kävely ja jumppa. Molempia lajiyhdistelmiä aineistossa on 62 kappaletta. Kävely ja pyöräily -lajiyhdistelmän peruskestävyyden kuntoarvion keskiarvo on 2,7, ja näitä molempia lajeja harrastava saa 16 %:n todennäköisyydellä hyvän tai erittäin hyvän arvion. Toisen suosituksen lajiyhdistelmän kävelyn ja jumpan peruskestävyyden kuntoarvion keskiarvo on 3,2, ja jopa 39 % testatuista, jotka ovat ilmoittaneet harrastavansa kävelyä sekä jumppaa, sai kuntoarviokseen hyvän tai erittäin hyvän.



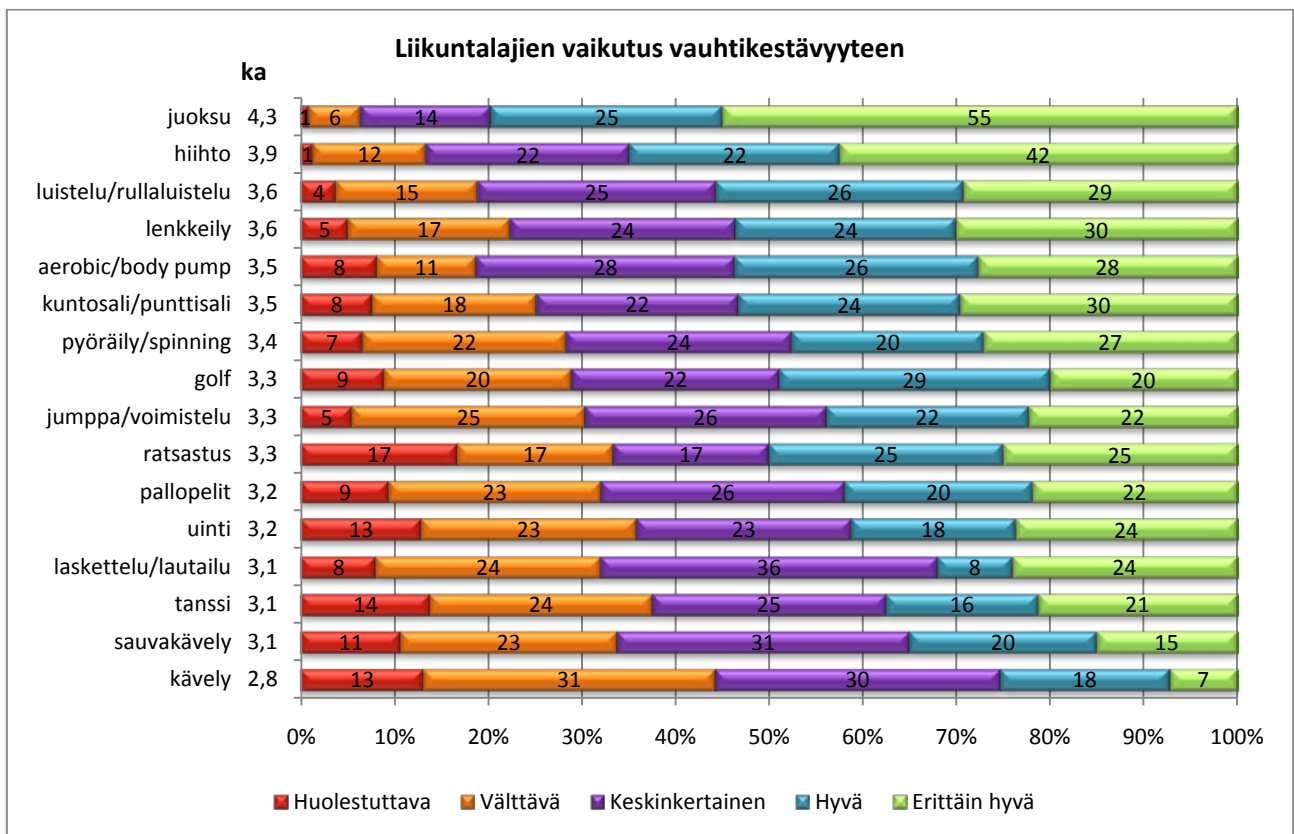
Kuvio 4.6. Liikuntalajien vaikutus peruskestävyyteen.

Tarkastellaan vielä vain yhtä lajia harrastavien peruskestävyyden kuntoarvioita. Vain kävelyä harrastaa 260 testattavaa, joista neljäsosa sai peruskestävyyden kuntoarviokseen hyvän tai erittäin hyvän, kun taas juoksua harrastavista (n = 70) jopa 72 % saivat vastaavat arviot. Testattavat jotka käyvät vain kuntosalilla (n = 27) saavat 70 %:n todennäköisyydellä peruskestävyyden arviokseen huolestuttavan tai välttävän. Peräti 77 % vain pallopelejä harrastavista (n = 47) saivat vastaavat arviot.

4.2.2 Liikuntalajien vaikutus vauhtikestävyteen

Liikuntalajit vaikuttavat tilastollisesti erittäin merkitsevästi vauhtikestävyteen ($\chi^2 = 760,14$; $df = 68$, p -arvo = 0). Liikuntalajien ja vauhtikestävyden arvion välinen riippuvuus on hieman suurempaa (kontingenssikerroin = 0,36) kuin edellä peruskestävyyden arvion ja liikuntalajien välinen riippuvuus. Tarkasteltaessa alla olevaa Kuviota 4.7 nähdään, että kahden ensimmäisen lajin harrastajat saivat myös peruskestävyyden puolella parhaimmat arviot verrattuna muihin lajeihin. Tämän jälkeen lajit ovatkin täysin erijärjestyksessä kuin mitä ne olivat Kuviossa 4.6. Selkeästi myös juoksua ja lenkkeilyä harrastavat saavat todennäköisemmin paremmat arviot vauhti- kuin peruskestävyydestä.

Juoksua harrastavat saavat 81 %:n todennäköisyydellä vauhtikestävyden arviokseen hyvän tai erittäin hyvän. Pallopelejä harrastavat saavat paremman arvion vauhti- kuin peruskestävyydestä, sillä he saavat 42 %:n todennäköisyydellä hyvän tai erittäin hyvän (peruskestävyyden puolella tämä sama prosenttiluku oli 32 %). Kävelyä harrastavat saavat todennäköisimmin välttävän tai huolestuttavan arvion (43 %) verrattuna muihin lajeihin, ja arvioiden keskiarvokin on selkeästi pienin. Kuvion 4.6 ja 4.7 perusteella voidaan päätellä, että kävelyllä voidaan ylläpitää paremmin perus- kuin vauhtikestävyttä.



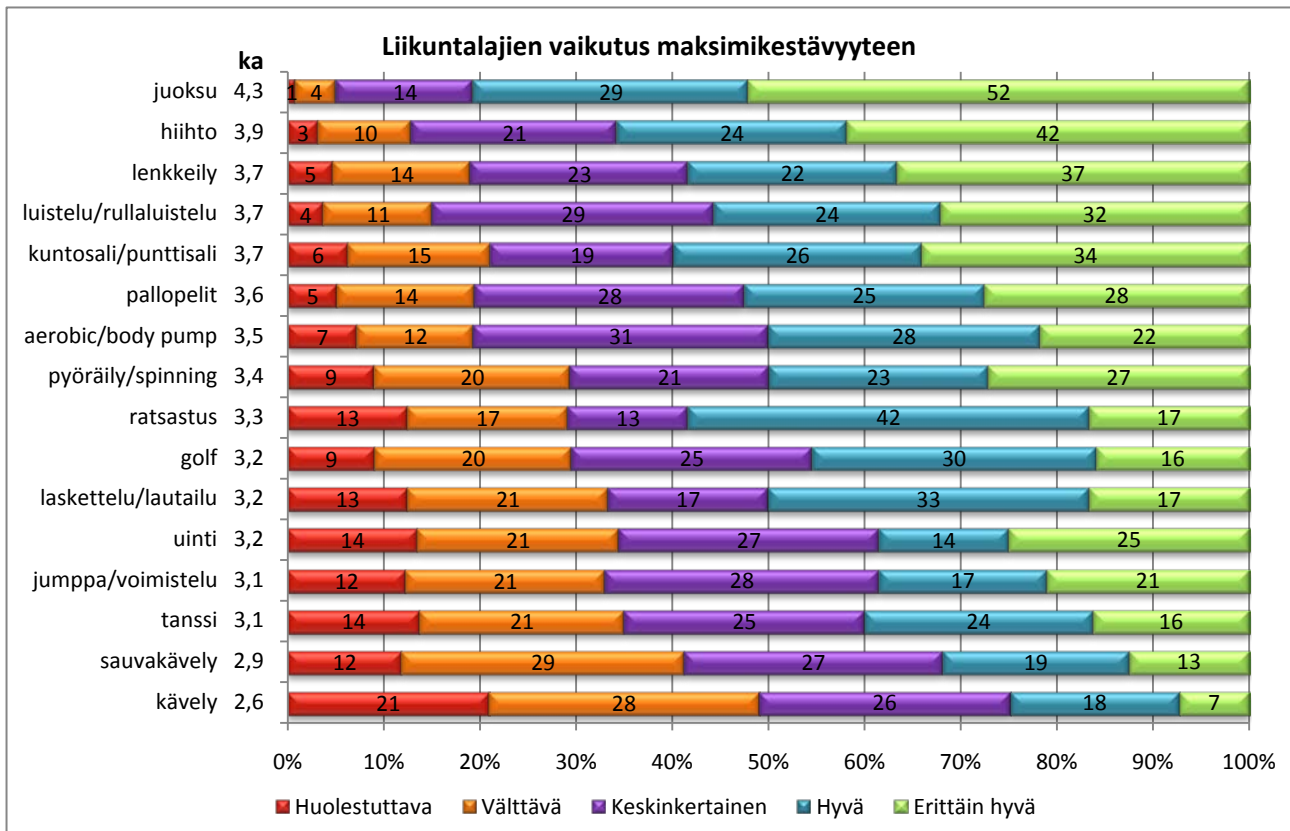
Kuvio 4.7. Liikuntalajien vaikutus vauhtikestävyteen.

Kun verrataan yhtä lajia harrastavien peruskestävyyden ja vauhtikestävyyden arvioita, huomataan kävelijöiden kuntoarvioiden huononevan. Nyt vain kävelyä harrastavat saavat 11 %:n todennäköisyydellä vauhtikestävyyden arviokseen hyvän tai erittäin hyvän, joten kävelyä harrastavat saavat selkeästi huonompia arvioita kuin mitä he saivat peruskestävyydestä. Juoksijoiden vauhtikestävyyden arviot ovat vastaavasti paremmat kuin peruskestävyydessä, sillä heistä 78 % sai vauhtikestävyyden arviokseen hyvän tai erittäin hyvän. Testattavien, jotka pelaavat vain pallopelejä tai käyvät kuntosalilla, vauhtikestävyyden arviot ovat selkeästi parempia kuin peruskestävyydessä.

4.2.3 Liikuntalajien vaikutus maksimikestävyyteen

Liikuntalajit vaikuttavat tilastollisesti erittäin merkitsevästi myös maksimikestävyyteen ($\chi^2 = 883,01$; $df = 68$, p -arvo = 0, kontingenssikerroin = 0,38). Liikuntalajien ja maksimikestävyyden arvion välinen riippuvuus on suurin, kun verrataan muiden kestävyysalueiden ja liikuntalajien välistä riippuvuutta. Verrattaessa Kuvioita 4.7 ja 4.8 huomataan, että samat lajit saavat parhaimmat ja huonoimmat arviot sekä vauhti- että maksimikestävyydestä. Pallopelejä harrastavat saavat selkeästi paremmat arviot maksimikestävyydestä kuin vauhti- tai peruskestävyydestä, toisin kuin muut lajit, jotka saavat joko samoja tai huonompia arvioita maksimikestävyydestä.

Juoksua harrastavien arviot ovat lähes samoja maksimi- ja vauhtikestävyydessä. Pallopelejä harrastavilla on selkeästi paremmat arviot maksimi- kuin perus- tai vauhtikestävyydessä. Kävelijät saavat todennäköisimmin välttävän tai huolestuttavan arvion (43 %) verrattuna muihin lajeihin ja arvioiden keskiarvokin on kävelyssä selkeästi pienin. Suositun lajiyhdistelmän kävelyn ja pyöräilyn maksimikestävyyden kuntoarvion keskiarvo on 2,1, eli paljon huonompi kuin peruskestävyyden kuntoarvion keskiarvo (2,6). Myös toisen yleisen lajiyhdistelmän, kävelyn ja jumpan, maksimikestävyyden kuntoarvion keskiarvo 2,4 on selkeästi pienempi kuin peruskestävyyden vastaava keskiarvo (3,2). Vain 16 % kävelyä sekä jumppaa harrastavista testatuista saivat kuntoarviokseen hyvän tai erittäin hyvän.



Kuvio 4.8. Liikuntalajien vaikutus maksimikestävyyteen.

Maksimikestävyyden kuntoarviot ovat vauhtikestävyydsarvioita paremmat niillä, jotka harrastavat vain juoksua, pallopelejä tai käyvät kuntosalilla. Nyt vain 34 % pallopelejä pelaavista sai maksimikestävyyden arviokseen huolestuttavan tai välttävän, kun peruskestävyydsarvioissa vastaava prosenttiosuus oli 78 %. Kuitenkaan kukaan vain pallopelejä pelaavista tai kuntosalilla kävijöistä ei saa maksimikestävyyden arviokseen erittäin hyvää. Kävelijöiden maksimikestävyyden arviot ovat suunnilleen samat kuin vauhtikestävyyden arviot, sillä heistä 12 % sai maksimikestävyyden arviokseen hyvän tai erittäin hyvän.

5 Aerobisen ja anaerobisen kynnyksen syketason mallintaminen MARS-menetelmällä

Testiaineistoon sovitetaan MARS-malli, jossa selitettävänä muuttujana on ensiksi aerobinen kynnyssyke. Tämän jälkeen alaluvussa 5.2 mallinnetaan anaerobista kynnyssykettä. Aineistossa on paljon puuttuvia arvoja ja R-ohjelman *mars*-funktio ei näitä hyväksy. Pystyäkseen muodostamaan MARS-mallit *mars*-funktioita käyttäen, on poistettava puuttuvia arvoja sisältävät havainnot. Tällöin aineistoon jää 1 967 havaintoa 1 414 eri henkilöltä.

5.1 Aerobinen kynnyssyke

Sovitetaan aineistoon MARS-malli, johon sallitaan mahdollisiksi selittäjiksi kaikki muuttujat lukuun ottamatta veren laktaattipitoisuusmuuttujia, aerobisen ja anaerobisen kynnyksien syketasoja, vauhtia, hapenottokykyä sekä kuntoarvioita. Taulukosta 5.1 nähdään saadut MARS-mallit, kun termien lukumääräksi on sallittu joko 3 - 4, 5 - 6 tai 7 - 8. Jokaiseen malliin on sallittu toisen asteen yhdysvaikutukset.

Taulukko 5.1. MARS-mallit, kun selitettävänä muuttujana on aerobinen kynnyssyke.

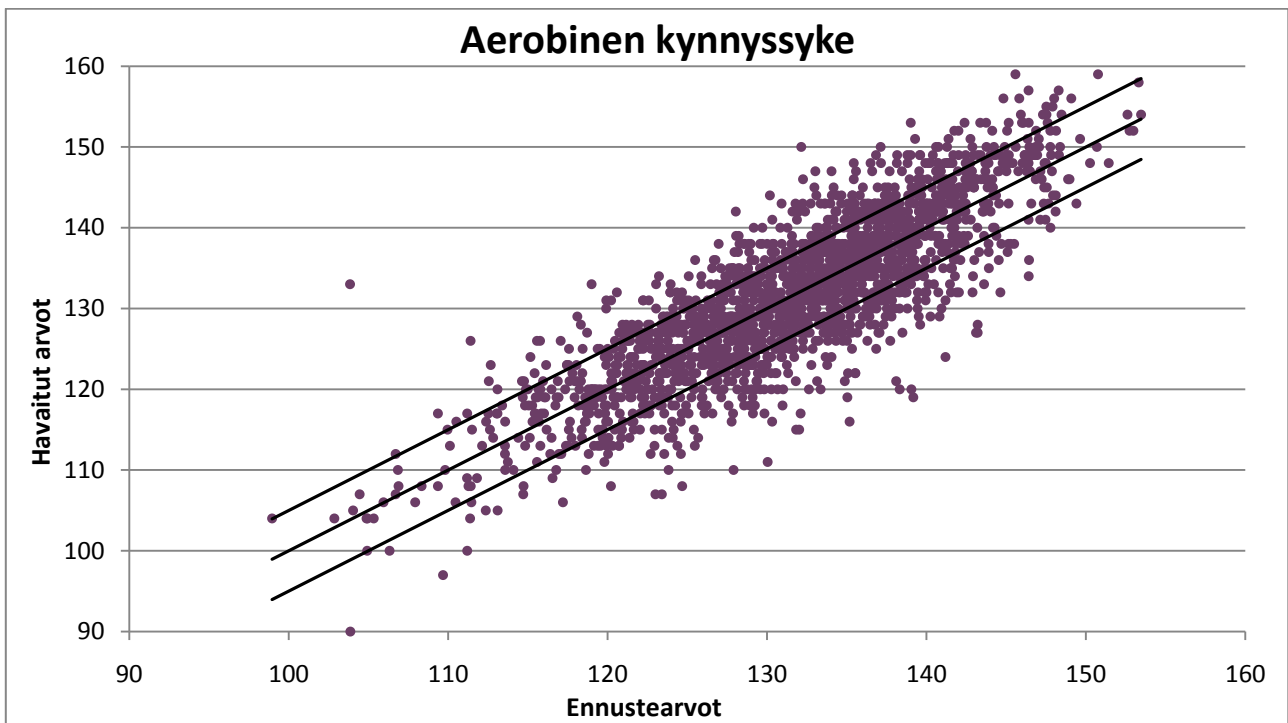
	Termien lkm	SOLMUKOHDAT				
		SykeKA	Syke3	Kestävyysepisteet		
Malli (5.1)	3-4	152,83 152,83			Selityssaste: GCV: Residuaalien hajonta:	0,54 46,16 6,78
Malli (5.2)	5-6	152,83 152,83	160 160		Selityssaste: GCV: Residuaalien hajonta:	0,69 31,73 5,61
Malli (5.3)	7-8	152,83 152,83	160 160	8 8	Selityssaste: GCV: Residuaalien hajonta:	0,71 29,79 5,42

Taulukossa ensimmäisellä sarakkeella on mallin termien lukumäärä. Toisesta sarakkeesta neljänten sarakkeeseen on merkitty mallin reflektiiviset parit siten, että tummennettu solmukohta vastaa positiivista osaa $(x - k)_+$. Viimeisessä sarakkeessa on tärkeimpiä mallin riippuvuutta kuvaavia tilastollisia tunnuslukuja.

Kaikkien mallien selittäjät ovat malleissa erittäin merkitseviä (kaikki p-arvot ≈ 0). Missään näissä kolmessa mallissa ei yhdysvaikutuksia ole, vaikka toisen asteen yhdysvaikutukset sallittiin malleja määriteltäessä. Aerobista kynnyssykettä selittää pelkästään sykkeiden keskiarvo, kun malliin sallitaan kolme tai neljä termiä. Tämän mallin selitysaste on 0,54 eli vähän yli puolet aerobisen kynnyssykkeen vaihtelusta pystytään selittämään tällä kyseisellä mallilla. Kun malliin sallitaan viisi tai kuusi termiä, selittää sykkeiden keskiarvon lisäksi kolmannen kierroksen sykearvo aerobista kynnyssykettä. Tämän mallin selitysaste on 0,15 yksikköä suurempi kuin ensimmäisen mallin selitysaste. Malliin (5.3) on sallittu maksimissaan kahdeksan termiä. Tällöin aerobista kynnyssykettä selittää sykkeiden keskiarvon ja kolmannen kierroksen sykkeen lisäksi *kestävyyspisteet*-muuttuja (kts. muuttujaluettelo Liite B). Tämän muuttujan lisääminen ei kuitenkaan paranna mallia merkitsevästi, sillä selitysaste nousee vain 0,02:lla ja residuaalien hajonta laskee 5,42:een eli vain noin 0,2:lla. Valitaan näiden tuloksien perusteella parhaaksi malliksi malli (5.2), missä selittävinä muuttujina ovat vain sykkeiden keskiarvo ja kolmannen kierroksen sykearvo.

Mallissa (5.2) sykkeiden keskiarvon solmukohta on 152,8 ja kolmannen kierroksen sykkeen 160. Kun tarkastellaan mallin kertoimien arvoja, huomataan, että muuttujien solmukohdat eivät juuri vaikuta aerobiseen kynnyssykkeeseen. Aerobinen kynnyssyke kasvaa sykkeiden keskiarvon kasvaessa. Kolmannen kierroksen syke laskee hieman vähemmän aerobista kynnyssykettä, kun sen arvo on yli 160 kuin sen ollessa alle 160.

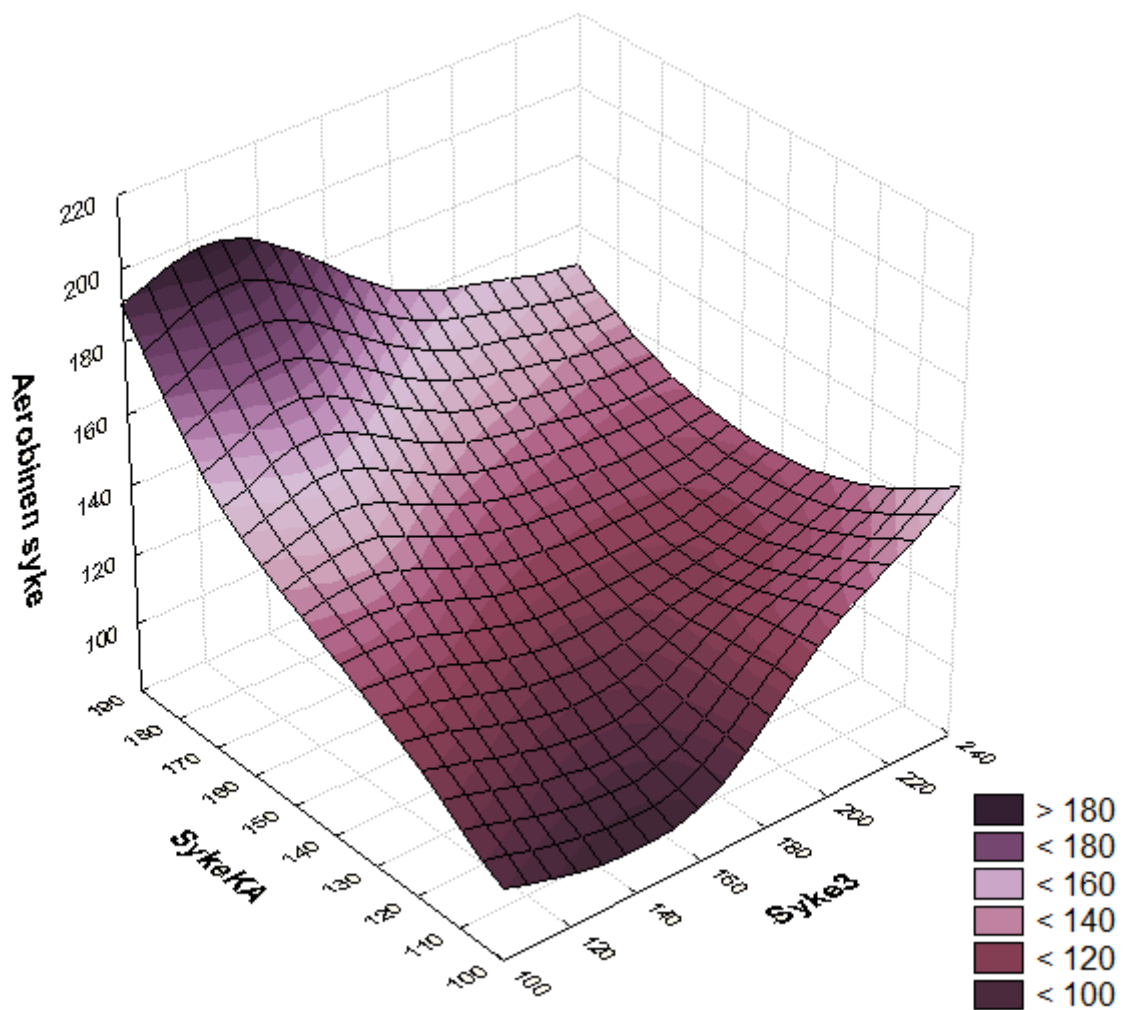
Tarkastellaan vielä mallin (5.2) hyvyttä alla olevan Kuvion 5.1 avulla. Kuntotestistä saatujen kynnyssykkeiden virhemarginaali ei saisi olla enempää kuin ± 5 yksikköä. Ennustearvoista 65 % on tämän virhemarginaalin välissä. Kuviosta huomataan, että malli antaa herkemmin ennustearvoksi havaittua arvoa suuremman.



Kuvio 5.1. Mallin (5.2) ennustearvojen eroavaisuus havaituista arvoista.

Kuviosta 5.2 nähdään sykkeiden keskiarvon sekä kolmannen kierroksen sykearvon vaikutus aerobiseen kynnyssykkeeseen.

Mikäli selitettäisiin aerobista kynnyssykettä vain testattavan taustatiedoilla (kts. muuttujaluettelo Liite B), tulisi parhaaseen malliin selittäjiksi testattavan paino, kestävyyspisteet sekä aerobisen kynnyssykkeeseen ikäryhmän keskiarvo ja kaikista muuttujista tulisi reflektiivisistä pareista vain positiiviset osat. Mallin residuaalien hajonta on 7,89 ja selitysaste 0,38 eli lähes 40 % aerobisen kynnyssykkeeseen vaihtelusta pystytään selittämään testattavan taustatiedoilla. Mallissa aerobisen kynnyssykkeeseen ikäryhmän keskiarvon solmukohta on 118,8, mikä on yli 60-vuotiaiden aerobisen kynnyssykkeeseen keskiarvo. Mikäli testattava on alle 60-vuotias, selittää testattavan ikäryhmän aerobinen kynnyssyke positiivisesti aerobista sykettä, sillä alle 60-vuotiaiden ikäryhmien kynnyssykkeiden keskiarvot ovat suurempia kuin 118,8. Kestävyysliikunnan harjoittaminen (kts.muuttujaluettelo Liite B) nostaa aerobista kynnyssykettä, kun sitä harjoitetaan enemmän kuin kohtalaisesti (kestävyyspisteiden tulee olla > 8 pistettä). Testattavan paino vaikuttaa negatiivisesti aerobiseen kynnyssykkeeseen, mikäli hän painaa enemmän kuin 59,6 kiloa; muulloin paino ei vaikuta aerobisen kynnyssykkeeseen arvoon.



Kuvio 5.2. Sykkeiden keskiarvon sekä kolmannen kierroksen sykearvon vaikutus aerobiseen kynnyssykkeeseen.

5.2 Anaerobinen kynnyssyke

Muodostetaan MARS-malli, jossa selitettävänä muuttujana on tällä kertaa anaerobinen kynnyssyke. Sallitaan mallin mahdollisiksi selittäjiksi samat muuttujat kuin mallintaessa aerobista kynnyssyketä. Taulukon 5.2 ensimmäisessä MARS-mallissa termien lukumääräksi on sallittu maksimissaan neljä, toisessa mallissa kuusi ja kolmannessa kahdeksan. Jokaiseen malliin on sallittu toisen asteen yhdysvaikutukset.

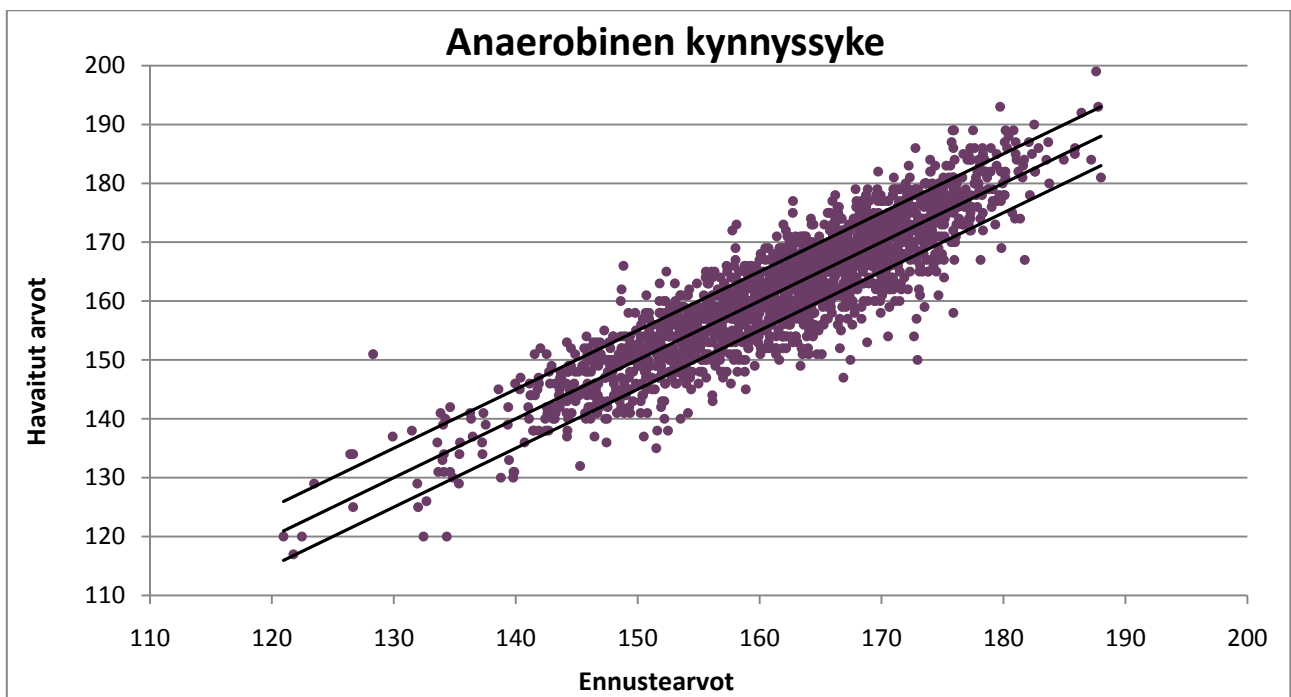
Taulukko 5.2. MARS-mallit, kun selitettävänä muuttujana on anaerobinen kynnyssyke.

	Termien lkm	SOLMUKOHDAT				
		Maxsyke	Kierrokset	SykeKA		
Malli (5.4)	3-4	188 188			Selitysaste:	0,70
					GCV:	38,79
					Residuaalien hajonta:	6,21
Malli (5.5)	5-6	188 188	3		Selitysaste:	0,76
					GCV:	31,46
					Residuaalien hajonta:	5,59
Malli (5.6)	7-8	188 188	3	144,6 144,6	Selitysaste:	0,80
					GCV:	26,22
					Residuaalien hajonta:	5,10

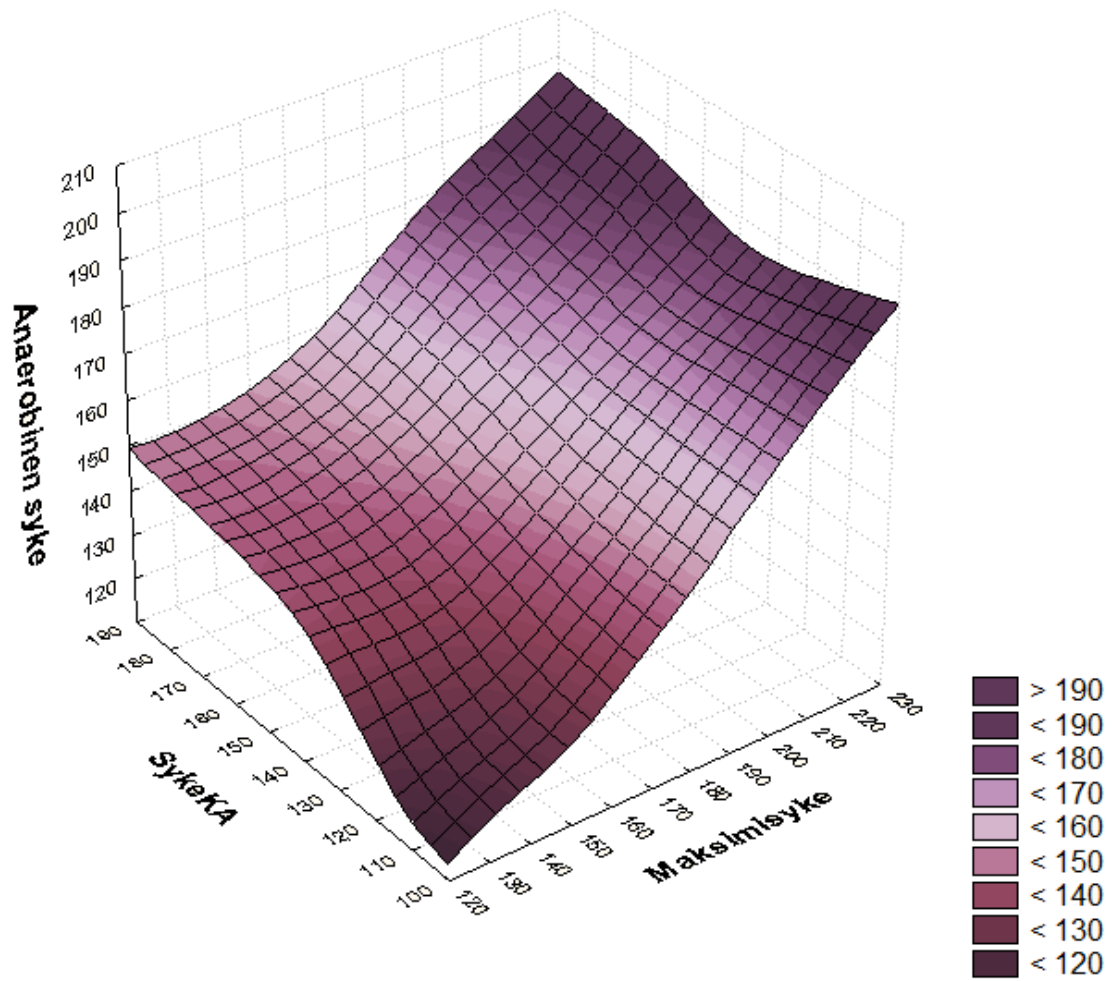
Kaikissa malleissa jokainen selittävä muuttuja on erittäin merkitsevä (kaikki p-arvot ≈ 0). Myöskään mihinkään näihin malleihin ei tule muuttujien yhdysvaikutuksia. Kun sallitaan malliin maksimissaan neljä termiä, niin anaerobista kynnyssyketä selittää pelkästään maksimisyke. Tämän mallin selitysaste on jopa 0,70 eli maksimisyke yksin pystyy selittämään erinomaisesti anaerobisen kynnyssykkeen vaihtelua. Mallissa (5.5) maksimisykkeen lisäksi malliin tulee *kierrokset*-muuttujan reflektiivisistä pareista positiivinen osa. Tämän mallin selitysaste on 0,06 yksikköä suurempi kuin mallin (5.4) selitysaste. Malliin (5.6) sallitaan maksimissaan kahdeksan termiä, jolloin anaerobista kynnyssyketä selittää maksimisykkeen ja kierroksien lisäksi sykkeiden keskiarvo. Tämän muuttujan reflektiivisten parien lisääminen parantaa mallia, sillä selitysaste nousee 0,04:llä ja residuaalien hajonta laskee 5,10:een eli noin 0,5:llä. Valitaan siis parhaaksi malliksi malli (5.6), missä selittävinä muuttujina ovat maksimisyke, kierrokset ja sykkeiden keskiarvo.

Mallissa (5.6) maksimisykkeen solmukohta on 188, kierroksien 3 ja sykkeiden keskiarvon 144,6. Kun tarkastellaan maksimisykkeen reflektiivisten parien kertoimien arvoja, huomataan, että muuttujan solmukohta ei juuri vaikuta anaerobiseen kynnyssykkeeseen. Anaerobinen kynnyssyke kasvaa sitä mukaan, kun maksimisyke kasvaa. Sykkeiden keskiarvo kasvattaa hieman enemmän anaerobista kynnyssyketä, kun sen arvo on yli 144,6. Kun testissä juostujen kierrosten määrä on enemmän kuin kolme, se nostaa mallista saadun anaerobisen kynnyssykkeen arvoa, muulloin kierrosten lukumäärä ei vaikuta kynnyssykkeeseen.

Tarkastellaan vielä mallin (5.6) hyvyttä alla olevan Kuvion 5.3 avulla. Mallista saaduista ennuste-arvoista 69 % on ± 5 havaittujen arvojen välissä. Kuviossa 5.4 on esitetty kolmiulotteisella kuviolla maksimisykkeen sekä sykkeiden keskiarvon vaikutus anaerobiseen kynnyssykkeeseen.



Kuvio 5.3. Mallin (5.6) ennustearvojen eroavaisuus havaituista arvoista.



Kuvio 5.4. Maksimisykkeen sekä sykkeiden keskiarvon vaikutus anaerobiseen kynnyssykkeeseen.

Mikäli selitettäisiin anaerobista kynnyssyykettä vain testattavan taustatiedoilla (kts. muuttujaluettelo Liite B), tulisi parhaaseen mallin selittäjiksi muuttujien paino ja ikä reflektiiviset parit. Tämän mallin residuaalien hajonta on 8,63 ja selitysaste 0,42 eli 42 % anaerobisen kynnyssyykkeen vaihtelusta pystytään selittämään testattavan painon sekä iän perusteella. Kun testattava on vanhempi kuin 47,6, laskee silloin anaerobinen kynnyssyyke hieman enemmän kuin testattavan ollessa nuorempi. Mallin kertoimien tarkastelussa huomaamme, että yli 80,8 kiloilla paino vaikuttaa aerobiseen kynnyssyykkeeseen enemmän alentavasti kuin alle 80,8 kiloilla.

5.3 Mallien kehittämistä

Kuntotestin tuloksiin vaikuttavat testattavan taustatietojen lisäksi testin suoritustapa sekä olosuhteet. Esimerkiksi mikäli testattava on ohjeista huolimatta lähtenyt ensimmäisen kierroksen jälkeen liian kovaa ja saavuttanut nopeasti maksimikuormituksen, testaajan on silloin vaikeampi löytää hänelle kynnyssykkeet kuin testattavalle, joka osaa nostaa kuormitusta annettujen ohjeiden mukaisesti. Myös tietyt lääkeaineet, tupakka, alkoholi, syöminen sekä rasitus ennen testiä vaikuttavat kuntotestin tulokseen. Kun näinkin moni asia vaikuttaa, on hyvin vaikeaa löytää kynnyssykkeille mallit, jotka sopisivat kaikille testattaville.

Tässä alaluvussa tarkastellaan mallin (5.2) ja mallin (5.6) poikkeavia havaintoja. Molemmista tarkastellaan noin 30 poikkeavaa havaintoa, joiden residuaalit ovat suurimmat. Huomataan kahdeksan havaintoa, jotka kuuluvat molempien mallien kolmenkymmenen poikkeavan havainnon joukkoon eli yhteensä aineistosta tarkastellaan 52 eri havaintoa. Tarkasteltaessa näiden testattavien taustatietoja huomataan, että yli puolet näistä henkilöistä käyttää jotain tiettyä lääkeainetta tai tupakoivat. Tämä on suhteessa enemmän, kun verrataan koko aineistoon, sillä 23 % testattavista tupakoi tai käyttää tiettyä lääkeainetta. Kun taas tarkastellaan havaintojen testituloksia, huomataan, että lähes kaikissa tapauksissa kuormitusta on lisätty testattavan omaan kuntoon nähden 1. tai 2. kierroksen jälkeen liikaa. Muutamassa tapauksessa kuormitus on jäänyt selkeästi liian alhaiseksi. Voidaan päätellä, että mallit sopivat huonosti testatuille, jotka eivät ole nostaneet kuormitusta annettujen ohjeiden mukaisesti. Poistetaan nämä poikkeavat havainnot, lukuun ottamatta kuutta havaintoa, joista ei löytynyt testin suoritustavasta virhettä. Mallinnetaan seuraavaksi uudet mallit käyttäen aineistoa, josta on poistettu kaiken kaikkiaan 46 poikkeavaa havaintoa.

5.3.1 Aerobisen kynnyssykkeiden mallintaminen

Tehdään MARS-malli, jossa selitettävänä muuttujana on aluksi aerobinen kynnyssyke. Asetetaan termien ja interaktioiden lukumääräksi samat kuin mallia (5.2) muodostaessa eli termejä viisi ja sallitaan toisen asteen yhdysvaikutukset. Saatu malli nähdään Taulukosta 5.3. Mallin selityssaste on 0,73 ja residuaalien hajonta 5,21. Mallista saaduista ennustearvoista 65 % on ± 5 havaittujen arvojen välissä.

Taulukko 5.3. MARS-malli, jossa selitettävänä muuttujana on aerobinen kynnyssyke, ja kun aineistosta on poistettu poikkeavat havainnot.

	Selittäjien lkm	SOLMUKOHDAT			
		SykeKA	Syke3		
Malli (5.7)	5-6	152,25	174	Selitysaste:	0,73
		152,25			GCV:
			174	Residuaalien hajonta:	5,21

Vertailtaessa malleja (5.2) ja (5.7) huomataan, että *SykeKA*-muuttujan solmukohta on laskenut erittäin vähän, mutta *Syke3*-muuttujan solmukohta on noussut 160:stä 174:ään. Mallien kertoimissa ei ole suuria eroja. Mallin (5.7) residuaalien hajonta on pienempää kuin mallin (5.2) ja residuaalit ovat paremmin normaalisti jakautuneita (Liite C) mallissa (5.7) kuin mallissa (5.2). Mallin (5.7) selitysaste on hieman parempi kuin mallin (5.2). Näiden tuloksien perusteella voidaan tehdä johtopäätös, että poikkeavien havaintojen poisto ei paranna mallia olennaisesti.

5.3.2 Anaerobisen kynnyssykkeen mallintaminen

Mallinnetaan anaerobista kynnyssykettä MARS-menetelmällä aineistosta, josta on poistettu 46 poikkeavaa havaintoa. Mallissa (5.6) termien lukumääräksi asetettiin seitsemän sekä sallittiin toisen asteen yhdysvaikutukset, joten asetetaan samat muodostettaessa mallia (5.8). Saadun mallin reflektiiviset parit ja mallin riippuvuutta kuvaavat tilastolliset tunnusluvut nähdään Taulukosta 5.4. Mallin selitysaste on 0,83 ja residuaalien hajonta 4,70. Mallista saaduista ennustearvoista 71 % on ± 5 havaittujen arvojen välissä.

Taulukko 5.4. MARS-malli, jossa selitettävänä muuttujana on anaerobinen kynnyssyke, ja kun aineistosta on poistettu poikkeavat havainnot.

	Termien lkm	SOLMUKOHDAT				
		Maxsyke	Kierrokset	SykeKA		
Malli (5.8)	7-8	178	3	144,67	Selitysaste:	0,83
		178				GCV:
				144,67	Residuaalien hajonta:	4,70

Mallin (5.8) selittävät muuttujat ovat samat kuin mallissa (5.6) ja solmukohdat ovat muuten samat, paitsi maksimisykkeen solmukohta on laskenut kymmenen yksikköä. Mallin (5.8) vakiokerroin on pienempi kuin mallin (5.6), mutta muut kertoimet ovat lähes samat. Residuaalien hajonta on selkeästi pienempää ja residuaalit ovat paremmin normaalisti jakautuneita (Liite C) mallissa (5.8) kuin mallissa (5.6). Poikkeavien havaintojen poisto vaikuttaa suunnilleen yhtä paljon anaerobisen kynnyssykkeen mallintamiseen kuin aerobisen kynnyssykkeen, sillä molempien selitysaste paranee sekä residuaalien hajonta laskee suunnilleen saman verran.

5.4 Aerobisen ja anaerobisen kynnyssykkeen mallintaminen MARS-menetelmällä *earth*-funktiota käyttäen

Muodostetaan tässä alaluvussa MARS-malli aerobiselle sekä anaerobiselle kynnyssykkeelle käyttäen R-ohjelman *earth*-funktiota (kts. luku 3.4.3). Tarkastellaan miten *earth*-funktiolla saadut MARS-mallit eroavat edellä olevista malleista. Sallitaan näihin malleihin kaikki samat muuttujat kuten malleissa (5.1) - (5.6). Jotta voidaan vertailla *earth*-funktiolla sekä *mars*-funktiolla muodostettuja malleja, sovitetaan myös nyt MARS-mallit aineistoon, josta on poistettu puuttuvia mittauksia sisältävät havainnot.

5.4.1 Aerobisen kynnyssykkeen mallintaminen

Mallinnetaan ensin aerobista kynnyssykettä MARS-menetelmällä käyttäen *earth*-funktiota. MARS-mallit nähdään Taulukosta 5.5, kun termien lukumääräksi on sallittu joko 3 - 4, 5 - 6 tai 7 - 8. Jokaiseen malliin on sallittu toisen asteen yhdysvaikutukset.

Taulukko 5.5. MARS-mallit *earth*-funktiolla, kun selitettävänä muuttujana on aerobinen kynnyssyke.

	Termien lkm	SOLMUKOHDAT				
		SykeKA	Syke3	Kestävyysepisteet		
Malli (5.9)	3-4	152,83 152,83			Selitysaste: GCV: Residuaalien hajonta:	0,54 46,16 6,78
Malli (5.10)	5-6	152,83 152,83	165 165		Selitysaste: GCV: Residuaalien hajonta:	0,69 31,73 5,61
Malli (5.11)	7-8	152,83 152,83	165 165	7 7	Selitysaste: GCV: Residuaalien hajonta:	0,71 29,79 5,42

Malleihin (5.9) - (5.11) tulee samat selittävät muuttujat kuin malleihin (5.1) - (5.3). Mallien (5.10) - (5.11) maksimisykkeen ja kestävyyspisteiden solmukohdat sekä vakio- että regressiokertoimet eroavat hieman malleista (5.2) - (5.3). Mallien selitysasteet, GCV-arvo ja residuaalien hajonta ovat samat kuin *mars*-funktiolla muodostetut MARS-mallit. Nämä *earth*-funktiolla muodostetut MARS-mallit eivät eroa malleista (5.1) - (5.3) niin paljon kuin mitä luvussa kolme esitetyt mallit (3.18) ja (3.19) eroavat keskenään.

5.4.2 Anaerobisen kynnyssykkeen mallintaminen

Muodostetaan malleja (5.4) - (5.6) vastaavat MARS-mallit käyttäen *earth*-funktiota. Sovitetaan aineistoon MARS-malli, jossa selitettävänä muuttujana on nyt anaerobinen kynnyssyke. Taulukosta 5.6 nähdään saadut MARS-mallit, kun termien lukumääräksi on asetettu joko 3 - 4, 5 - 6 tai 7 - 8, ja toisen asteen yhdysvaikutukset on sallittu.

Taulukko 5.6. MARS-mallit *earth*-funktiolla, kun selitettävänä muuttujana on anaerobinen kynnyssyke.

	Termien lkm	SOLMUKOHDAT				
		Maxsyke	Kierrokset	SykeKA		
Malli (5.12)	3-4	185 185			Selitysaste:	0,70
					GCV:	38,78
					Residuaalien hajonta:	6,21
Malli (5.13)	5-6	185 185	4 4		Selitysaste:	0,76
					GCV:	31,46
					Residuaalien hajonta:	5,59
Malli (5.14)	7-8	185 185	4 4		Selitysaste:	0,80
				144,8 144,8	GCV:	26,28
					Residuaalien hajonta:	5,10

Vertailtaessa Taulukkojen 5.2 sekä 5.6 malleja nähdään, että mallit ovat lähes samanlaiset. Selittävien muuttujien solmukohdat eroavat malleissa, varsinkin maksimisykkeen solmukohta. Kun sallitaan maksimissaan kahdeksan termiä, tulee malliin selittäväksi muuttujaksi kaikkien kolmen muuttujan reflektiiviset parit. Tämä malli (5.14) eroaa siis hieman mallista (5.6), johon tuli *kierrokset*-muuttujan reflektiivisistä pareista vain positiivinen osa. Mallien (5.12) - (5.14) vakio- sekä regressiokertoimet eroavat myös jonkin verran malleista (5.4) - (5.6), mutta mallien selityksasteet, GCV-arvo sekä residuaalien hajonta ovat samat kuin *mars*-funktiolla muodostettujen MARS-mallien. Nämä mallit eroavat selkeästi enemmän *mars*-funktiolla muodostetuista malleista kuin aerobista kynnyssykettä *earth*-funktiolla mallintaessa saadut mallit (5.9) - (5.11).

6 Yhteenveto

Tutkielman tarkoituksena oli mallintaa MARS-menetelmällä aerobista sekä anaerobista kynnyssykettä sekä tarkastella liikunnan määrän ja laadun vaikutusta eri kestävyysalueiden kuntoarvioihin. Tutkielmaa tehdessä huomattiin yhdessä lähdekirjassa virhe, joka vaikutti suuresti MARS-mallien hyvyden tulkintaan. Onneksi virhe huomattiin alkuvaiheessa, ja näin tutkielmaan ei jäänyt MARS-malleista vääränlaisia tulkintoja.

Tutkielmassa tarkasteltiin liikuntatottumuksien lisäksi kuntotestissä kävijöiden taustatietoja. Kuntotestiin hakeutuu enemmän yli 30-vuotiaita sekä enemmän miehiä kuin naisia. Testattavien elämäntavat ovat selkeästi terveellisemmät ja heillä on vähemmän ylipainoa kuin keskimäärin koko väestössä. Testissä kävijät liikkuvat suunnilleen yhtä paljon kuin suomalaiset normaalistikin, vaikka voisi kuvitella, että testattavat harrastaisivat liikuntaa selkeästi enemmän. Kuntotestissä käyneet naiset ja miehet harrastavat liikuntaa yhtä usein viikossa, mutta alle 30-vuotiaat harrastavat enemmän kuin yli 50-vuotiaat.

Tutkimuksen mukaan liikunnan määrä vaikuttaa perus-, vauhti-, ja maksimikestävyysalueeseen. Liikunta-kerroilla ei ole vaikutusta maksimikestävyysalueeseen niin paljon kuin perus- ja vauhtikestävyysalueeseen. Sen sijaan liikuntalajit vaikuttavat enemmän maksimikestävyysalueeseen. Liikuntalajeja on tutkittu multiresponse-muuttujan avulla, joten tulokset liikuntalajien vaikutuksesta kestävyysalueisiin ovat suuntaan antavia, sillä monet testissä kävijät harrastavat enemmän kuin yhtä liikuntalajia. Juoksua harrastavat saivat parhaimmat kuntoarviot kaikista kestävyysalueista, ja kävelyä harrastavat selkeästi huonoimmat. Noin joka kymmenes vain kävelyä harrastaneista sai vauhti- ja maksimikestävyysalueen arviokseen hyvän tai erittäin hyvän. Kun taas selkeä enemmistö (n. 80 %) testatuista, jotka harrastavat vain juoksua, saivat vauhti- ja maksimikestävyysalueesta vastaavat arviot. Eri lajien vaikutusta kestävyysalueisiin tarkasteltaessa huomataan mielenkiintoinen tulos: palloilua harrastavat saivat paljon parempia arvioita maksimi- kuin perus- ja vauhtikestävyysalueesta.

Tutkimuksen tärkein tavoite oli määrittää MARS-menetelmällä aerobisen ja anaerobisen kynnyssyke käyttämällä testattavan testituloksia sekä taustatietoja käyttämättä veren laktaattipitoisuuksia. Aerobisen kynnyssyksen vaihtelusta pystytään noin 70 prosenttia selittämään kolmannen kierroksen sykkeellä sekä sykkeiden keskiarvolla. Pelkästään testattavan taustatiedoilla pystytään selittämään noin 40 prosenttia aerobisen sekä anaerobisen kynnyssyksen vaihtelusta. Anaerobisen kynnyssyksen tärkein selittäjä on testattavan maksimisyke, koska pelkästään sillä voidaan

ennustaa jopa 70 prosenttia kynnysykykkeen vaihtelusta. Tämä tulos ei ole yllättävä, sillä aikaisempien tutkimusten perusteella maksimisykkeen avulla voidaan arvioida anaerobinen syke.

Toisaalta monien lähteiden mukaan myös aerobinen kynnysykyke saadaan maksimisykkeen avulla. Kuitenkin tässä tutkimuksessa MARS-menetelmä löytää aineistosta tärkeimpiä muuttujia, jotka selittävät aerobista kynnysykykettä paremmin kuin maksimisykyke. MARS-menetelmällä pystytään selittämään anaerobisen kynnysykykkeen vaihtelua paremmin kuin aerobisen, sillä 80 prosenttia tästä vaihtelusta pystytään selittämään maksimisykkeen, sykkeiden keskiarvon sekä testissä kierrettyjen kierroksien lukumäärän avulla. Aerobista ja anaerobista kynnysykykettä voidaan siis mallintaa hyvinkin yksinkertaisilla malleilla. Merkittävä huomio on sekin, että MARS-menetelmän perusteella testitulokset ovat tärkeämpiä muuttujia kuin esimerkiksi testattavan ikä, sukupuoli tai antropometriset muuttujat.

Tutkielman analyysien tuloksien perusteella voidaan todeta, että aerobista sekä anaerobista kynnysykykettä onnistuttiin mallintamaan hyvin tarkasti, kun otetaan huomioon kuntotestin tuloksiin vaikuttavat erilaiset häiritsevät tekijät. Näitä saatuja malleja on käytetty syketasotestissä ja ne ovat mallintaneet mainiosti kynnysykykkeitä. Syketasotestistä voitaisiin saada vieläkin tarkempi, mikäli erilaisille ryhmille, kuten esimerkiksi kilpakuntoilijoilla ja kuntoileville, muodostettaisiin omat mallit.

Tutkielman tekovaiheessa lähdekirjasta löydetyn virheen ansioista tutkielmassa tutkittiin vakiotermin millisen ja -termittömän regressiomallien tulkintojen eroavaisuuksia, ja huomattiin, että tutkijan on oltava hyvin tarkkana vakiotermitöntömän regressiomallin hyvyyden tulkinnassa, sillä eri tilasto-ohjelmat laskevat selitysasteen eri kaavoja käyttäen. Lisäksi vertailtiin R-ohjelman kahdella eri funktiolla muodostettuja MARS-malleja, jotka aluksi näyttivät antavan selvästi erilaiset mallit, mutta tutkimusaineistoa analysoitaessa funktiot antoivat lähes samanlaiset mallit.

Lopuksi haluan esittää erityiskiitokset Testaus- ja kuntotutkimusasema TesKu ay:n testauspäällikölle Matti Jääskeläiselle loistavasta mahdollisuudesta tehdä mielenkiintoisesta ja urheiluhenkisestä aiheesta Pro gradu -tutkielma. Professori Erkki Liskille sekä tutkimuspäällikölle Jyrki Ollikaiselle kiitokset hyvistä neuvoista ja kärsivällisyydestä. Kiitos myös eteläafrikkalaiselle Stephen Milborrowille, joka jaksoi vastaila kysymyksiini MARS-mallin muodostamisesta R-ohjelmalla.

Lähdeluettelo

- Agresti, A. (2007), "An Introduction to Categorical Data Analysis", *Hoboken: John Wiley & Sons, Inc.*
- Australian Sports Commission (2000), "Physiological tests for elite athletes", *Australia: Human Kinetics.*
- Eisenhauer, J. G. (2003), "Regression through the Origin", *Teaching Statistics* (Vol. 25, Num. 3), 76–80.
- Faraway, J. J. (2006), "Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models", *Boca Raton: Chapman & Hall/CRC.*
- Fogelholm, M., Paronen, O. & Miettinen, M. (2007), "Liikunta - hyvinvointipoliittinen mahdollisuus", *Helsinki: Sosiaali- ja terveystieteiden tutkimuskeskus.*
- Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines", *Annals of Statistics* (Vol. 19, Num. 1), 1–67.
- Hahn, G. J. (1977), "Fitting Regression Models with No Intercept Term", *Journal of Quality Technology* (Vol. 9, Num. 2), 56–61.
- Hastie, T., Tibshirani, R. & Friedman, J. H. (2009), "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", *California: Springer Series in Statistics* (2nd ed.).
- Heikkilä, T. (2001), "Tilastollinen tutkimus", *Helsinki: Edita.*
- Helakorpi, S., Paavola, M., Prättälä, R. & Uutela, A. (2009), "Suomalaisen aikuisväestön terveyskäyttäytyminen ja terveys, kevät 2008", *Helsinki: Terveystieteiden tutkimuskeskus.* Saatavilla Internetistä: <http://www.thl.fi/thl-client/pdfs/dcb684e6-d94f-4724-96d1-9f382492ac54>
- Helimäki, E., Keskinen, K. L., Alén, M., Komi, P. V. & Takala, T. E. S. (2000), "Kuntotestaus Suomessa", *Jyväskylä: Jyväskylän yliopisto, liikuntabiologian laitos.* Saatavilla Internetistä: http://www.kuntotestaus.net/filearc/3_Kuntotestaus%20Suomessa.pdf?PHPSESSID=9sm0amf0u949ntl56huieoc9q4
- Keskinen, K., Häkkinen, K. & Kallinen, M. (2004), "Kuntotestauksen käsikirja", *Helsinki: Liikuntatieteellinen Seura ry.*
- Mellin, I. (2008), "Tilastolliset menetelmät: Lineaarinen regressioanalyysi". Saatavilla Internetistä: <http://www.math.tkk.fi/opetus/sovtoda/oppikirja/Regranal.pdf>
- Montgomery, D. C. & Peck, E. A. (1992), "Introduction Linear Regression Analysis", *Canada: A Wiley-Interscience Publication* (2nd ed.).
- Puntanen, S. (1999), "Regressioanalyysi 1", *Tampere: Tampereen yliopisto, matematiikan ja tilastotieteen laitos.*

- Rawlings, J. O. (1988), "Applied Regression Analysis: A Research Tool", *California : Wadsworth & Brooks*.
- Sovijärvi, A., Uusitalo, A., Länsimies, E. & Vuori, I. (1994), "Kliininen fysiologia", *Helsinki: Kustannus Oy Duodecim*.
- World Health Organization (2005), "WHO:n maailmanlaajuinen strategia 2005; Ravinto, liikunta ja Terveys". Saatavilla Internetistä: http://www.ktl.fi/attachments/suomi/julkaisut/muut_ohjeet/who_ravintoliikuntastrategia.pdf
- Vuori I. & Taimela S. (1999), "Liikuntalääketiede", *Helsinki: Kustannus Oy Duodecim*.

Liite A: Taustakyselylomake ja testauspöytäkirja

Nimi _____ Synt. aika _____ / _____ 19 _____ Pituus _____ paino _____

Osoite _____ Postinumero ja
paikka _____

Sähköpostiosoite _____ @ _____ Puhelin _____

On tärkeää, että tiedämme terveydentilastasi ja aiemmista liikuntatottumuksistasi, ennen kuin testaamme Sinut. Pyydämme ystävällisesti Sinua vastaamaan seuraaviin kysymyksiin.

- | | kyllä | ei |
|---|--------|--------|
| 1. Onko Sinulla todettu hengitys- sydän- tai verenkierto-elimistön sairauksia tai verenpainetta? Mitä ? _____ | x | x |
| 2. Onko Sinulla lääkkeitä säännöllisesti käytössäsi? Mitä? _____ | x | x |
| 3. Onko Sinulla ollut rintakiputuntemuksia? a) levossa b) rasituksessa | x
x | x
x |
| 4. Onko Sinulla ollut tuki- ja liikuntaelinten pitkäaikaisia tai usein toistuvia vaivoja ? | x | x |
| 5. Oletko sairastanut viimeisen 2 viikon aikana jotain tulehdustautia(kuume)? | x | x |
| 6. Onko Sinulla ollut jotain erityisiä tuntemuksia viimeisen kuukauden aikana ? (kohonnut syke, painon nousu/lasku, vatsan toimintahäiriö, väsymys, uni, päänsärky, ärtyisyys) | x | x |
| 7. Oletko nauttinut alkoholia viimeisen 24 tunnin aikana? | x | x |
| 8. Tupakoitko ? (lopettanut _____ vuotta sitten) | x | x |
| 9. Kuntoilijat: Kuinka usein viikoittain olet harrastanut liikuntaa viimeisen puolen vuoden aikana?
a) en lainkaan b) satunnaisesti c) 1 – 2 x viikossa d) 3 – 4 x viikossa e) 5 – 7 x viikossa | | |
| Mitä kestävyystyypistä kuntoilua harrastat? _____ km tai h _____ / viikko | | |
| Mitä lihaskuntoa teet ? _____ | | |
| Muuta , mitä ? _____ | | |
| Urheilijat: laji _____, harjoitusvuodet _____, kerrat _____/ viikko, viikossa _____ km / tuntia | | |
| 10. Onko Sinulla tavoitteena joku tapahtuma / aika (esim. maraton)? _____ | | |
| 11. Oletko osallistunut johonkin tapahtumaan aiemmin? _____
Aiemmat ennätykset / saavutukset _____ | | |
| 11. Mitä hyötyliikuntaa harrastat? _____
(kalastus, marjastus, metsästys, puutarhatyöt, polttopuiden teko, lumen luonti väh. 30 min. yms.) | | |
| 12. Nykyinen työsi tai ammattisi? _____ | | |
| 13. Miten kuljet työmatkasi _____, ajassa _____ min / päivä? (muu kuin auto, juna) | | |

Tunnen testaustavan ja kykenen osallistumaan siihen omalla vastuullani.

_____ / _____ 20 _____

Nimi						Test.pvm. / 20				
						Synt.pvm. / 19				
Pituus	cm.	Paino	kg.	F	%.	S	T	B	I	Yht.
HB		Hkr	Olosuhteet			Lihaskunto				
Aika/Jakso		Nopeus/Työ	Suoritus/ml	Syke	Maitohappo	Isometr. Selkä kg				
0-1		XXXXXXXX	XXXXXXXX			Selkä / oma paino				
1-4	1.					Isometr. Vatsa kg				
4-7	2.					Vatsa / oma paino				
7-10	3.					Vatsa / selkä suhde				
10-13	4.					Puristusvoima				
13-16	5.					Käsi: oikea kg				
16-19	6.					Käsi oikea / op				
19-22	7.					Käsi: vasen kg				
22-25	8.					Käsi vasen / op				
25-28	9.					Jalkatyö w				
28-31	10.					jalkatyö w / op				
31-34	11.					jalkatyö aika				
Palautuminen		Syke	La	La-poisto					Arvio	
1 min				Hyötysuhde			AerT			
4 min							AnT			
7 min							Max.			
10 min				Keskiarvo					Arvio	
TULOKSET		Aerob. kynnys		Anaerob. kynnys		Max. VO2				
%maksimista										
O2-kulutus L/min										
ml / kg / min										
suoritus / työ w										
syke										
maitohappo										
Arvio										

Liite B: Muuttujaluettelo

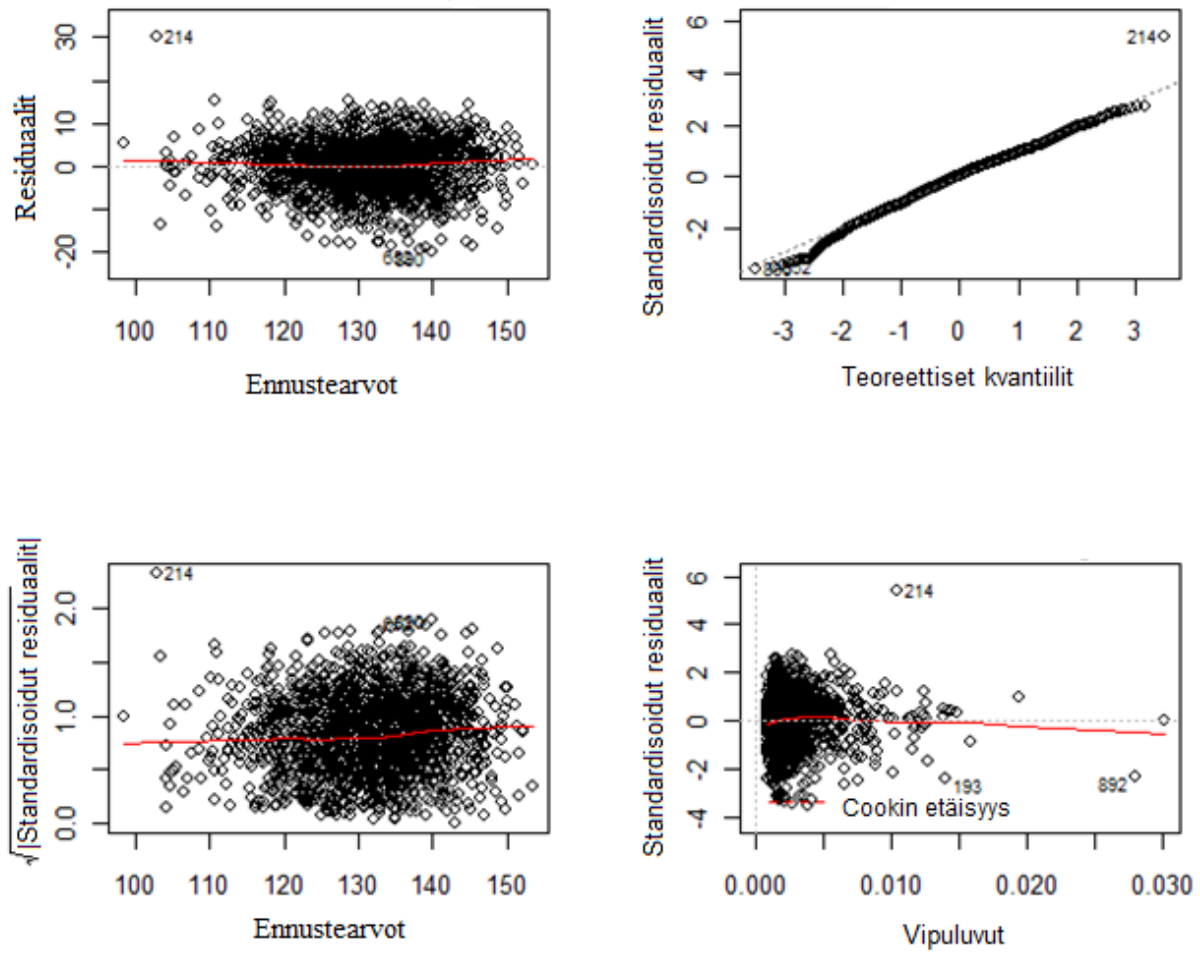
Testattavan taustatiedot:

Sex	=	Testattavan sukupuoli
Ikä	=	Testattavan ikä testipäivänä
Pituus	=	Testattavan pituus (cm)
Paino	=	Testattavan paino (kg)
Hengitys-/verenkierto-elimistön sairaudet	=	Dikotominen muuttuja. 1 = sairaus, 0 = ei sairautta
Lääkitys	=	Dikotominen muuttuja. 1 = lääkitys, 0 = ei lääkitystä
Tuki- ja liikuntaelimistön vaivat	=	Dikotominen muuttuja. 1 = tukielimistön vaivoja, 0 = ei vaivoja
Tupakointi	=	Dikotominen muuttuja. 1 = tupakoi, 0 = ei tupakoi
Hyötyliikunta	=	Dikotominen muuttuja. 1 = harrastaa hyötyliikuntaa, 0 = ei harrasta
Työmatkaliikunta	=	Dikotominen muuttuja. 1 = jalkaisin tai pyörällä, 0 = ei lainkaan
Työmatka aika/päivä	=	Työmatkaan menevä aika päivässä (min)
Liikuntakerrat	=	Testattavan liikuntakerrat viikossa viimeisen puolen vuoden aikana
Liikuntalaji	=	Liikuntalaji(t), mitä testattava on sanonut harrastavansa. Multiresponse-muuttuja.
Kestävyyspisteet	=	Kestävyyspisteet on saatu testattavan liikuntakertojen ja liikuntalajin avulla. Pisteitä muodostaessa liikuntalajit on luokiteltu 0-7 niin, että 0-luokassa ovat ne lajit, jotka vaikuttavat negatiivisesti aerobiseen kuntoon ja 7-luokassa lajit, jotka vaikuttavat positiivisesti aerobiseen kuntoon. Kestävyyspisteet saadaan siis kertomalla testattavan suurin liikuntalaji-luokka liikuntakerroilla.
Rasva%	=	Rasvaprosentti
IkäryhmänAEsyke	=	Ikäryhmän aerobisen kynnyssykkeen keskiarvo (lyönti/min)
IkäryhmänANSyke	=	Ikäryhmän anaerobisen kynnyssykkeen keskiarvo (lyönti/min)
BMI	=	Painoindeksi (kg/m ²)

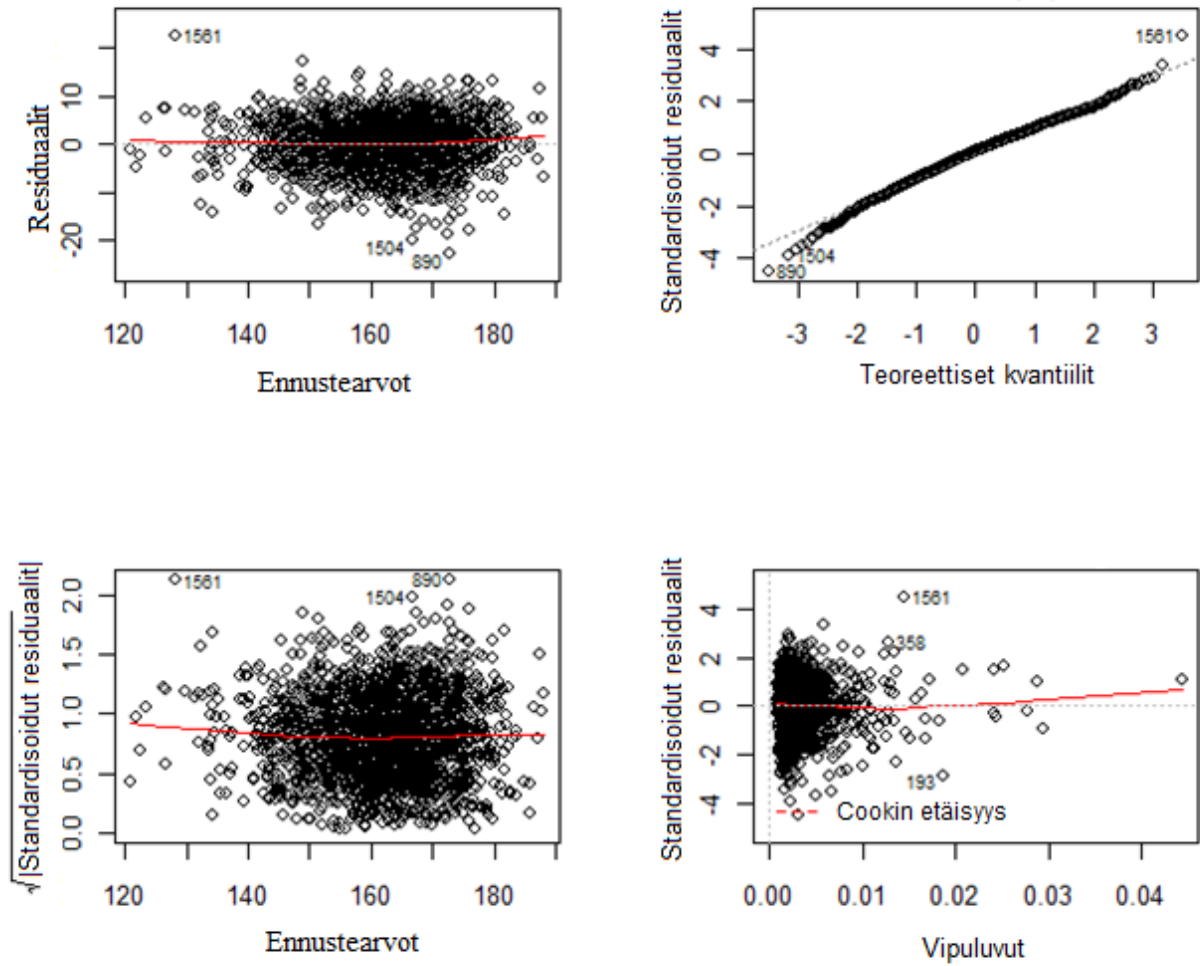
Testattavan testitulokset:

Kierrokset	=	Testissä juostujen kierroksien lukumäärä
k1, k2, ..., k7	=	Kierroksien kilometriaika (min)
s1, s2, ..., s7	=	Syketaajuus (lyönti/min)
l0	=	Veren laktaattipitoisuus ennen testiä (mmol)
l1, l2, ..., l7	=	Veren laktaattipitoisuus (mmol)
SykeKA	=	Kierros sykkeiden keskiarvo (lyönti/min)
Aerobmax	=	Aerobisen kynnyksen osuus maksimaalisesta hapenottokyvystä (%)
Amlkgmi	=	Aerobisen kynnyksen hapenottokyky (ml/kg/min)
Aminkm	=	Kilometriaika aerobisella kynnyksellä
Asyke	=	Aerobinen kynnyssyke (lyönti/min)
Ala	=	Veren laktaattipitoisuus aerobisella kynnyksellä (mmol)
AEarvio	=	Peruskestävyyden kuntoarvio
Anaermax	=	Anaerobisen kynnyksen osuus maksimaalisesta hapenottokyvystä (%)
Anmlkgmi	=	Anerobisen kynnyksen hapenottokyky (ml/kg/min)
Anminkm	=	Kilometriaika anaerobisella kynnyksellä
Ansyke	=	Anaerobinen kynnyssyke (lyönti/min)
Anla	=	Veren laktaattipitoisuus anaerobisella kynnyksellä (mmol)
ANarvio	=	Vauhtikestävyyden kuntoarvio
Mmlkgmin	=	Maksimaalinen hapenottokyky (ml/kg/min)
Mminkm	=	Nopein kilometriaika
Msyke	=	Maksimisyke (lyönti/min)
Mla	=	Suurin mitattu veren laktaattipitoisuuden arvo (mmol)
MAXarvio	=	Maksimikestävyyden kuntoarvio

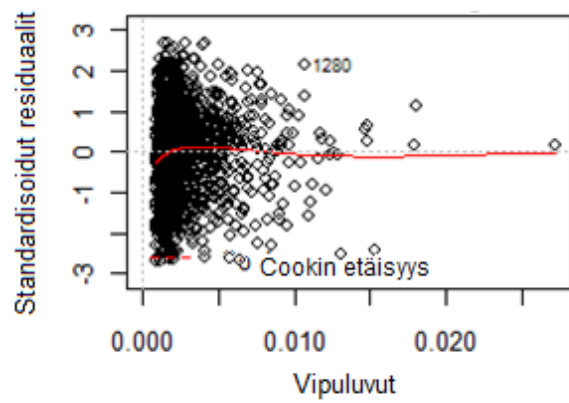
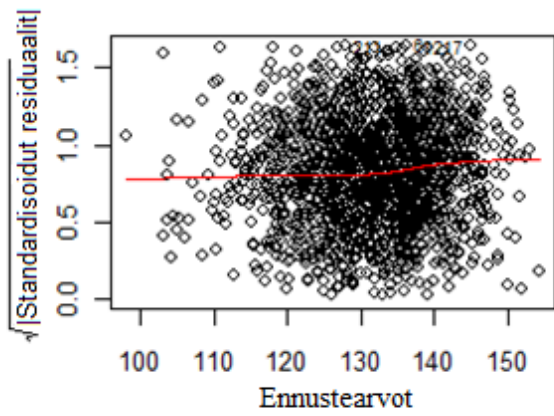
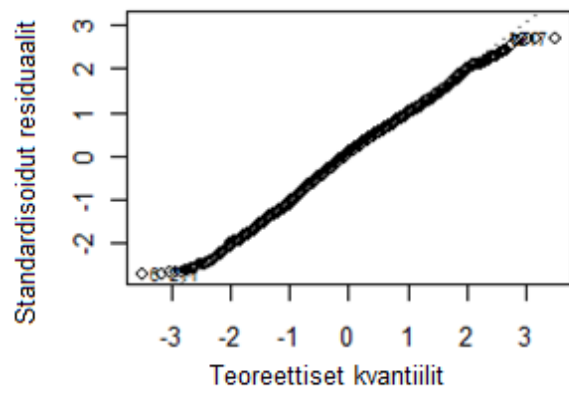
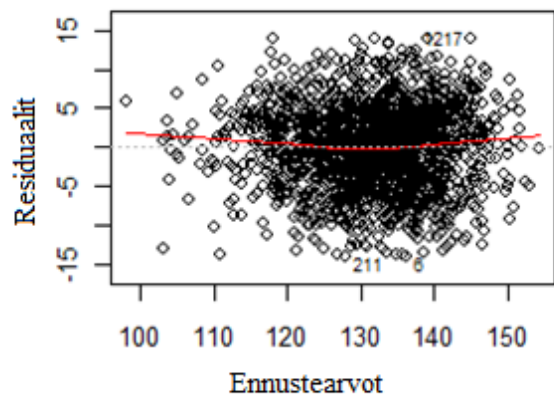
Liite C: Kuviot mallien diagnostiikasta



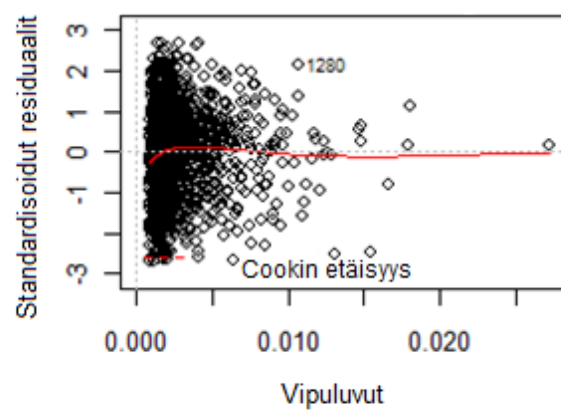
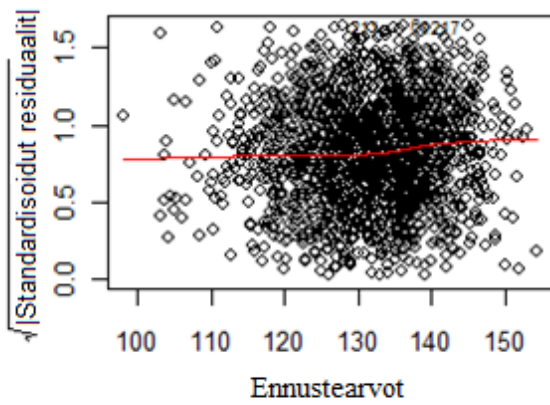
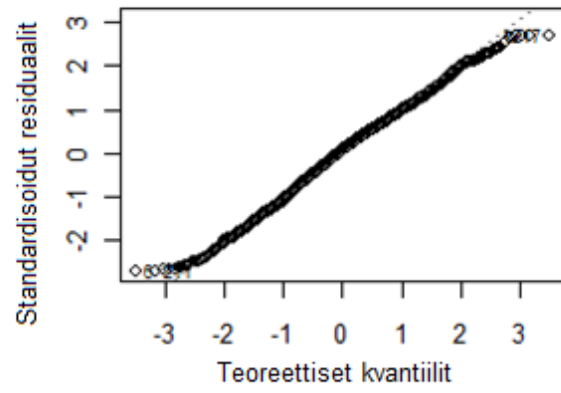
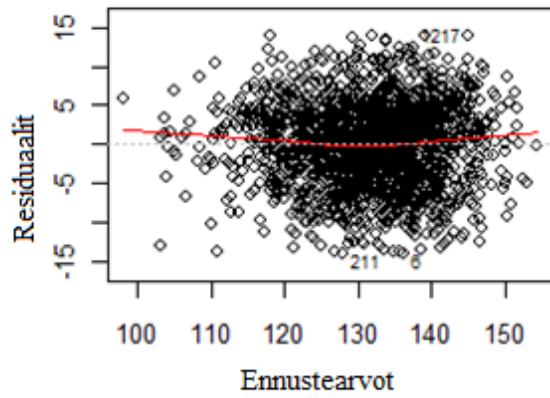
Kuvio 6.1. Mallin (5.2) diagnostiikkaa



Kuvio 6.2. Mallin (5.6) diagnostiikkaa



Kuvio 6.3. Mallin (5.7) diagnostiikkaa



Kuvio 6.4. Mallin (5.8) diagnostiikkaa