

Tartu Ülikool  
Loodus- ja täppisteaduste valdkond  
Matemaatika ja statistika instituut

Merili Konik

# Ankurpunktide mudeli tundlikkus andmetiheduse ja kellaaja kui sisendparameetri suhtes

Matemaatilise statistika eriala

Bakalaureusetöö (9 EAP)

Juhendajad: MSc Anne Selart

MSc Kaisa Vent

Tartu 2020

# **Ankurpunktide mudeli tundlikkus andmetiheduse ja kellaaja kui sisendparameetri suhtes**

Bakalaureusetöö

Merili Konik

Bakalaureusetöö eesmärk on testida Positium OÜ väljatöötatud ankurpunktide mudeli tundlikkust andmetiheduse ja ühe mudeli parameetri muutumise suhtes. Selleks parameetriks on keskmine kõnede alustamise aeg. Mudelis muudetakse vastava parameetri väärtusi ja luuakse neli erinevat stsenaariumit, sh baasstsenaarium. Töö käigus analüüsitakse mudeli põhjal arvutatud kodu ankurpunkte erinevate stsenaariumite lõikes kahe Eesti mobiilsideoperaatori korral. Töös antakse ülevaade inimese ruumilise käitumise mõjuteguritest, mobiilpositsioneerimisest ja Positiumi ankurpunktide mudelist. Praktilises osas esitatakse leitud ankurpunktide kirjeldav statistika ja viiakse läbi logistiline regressioon binaarsele tunnusele, mis näitab, kas määratud ankurpunkt langeb kokku kodule geograafiliselt lähima mobiilsideantenni asukohaga.

**CERCS teaduseriala:** P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika; S230 Sotsiaalne geograafia

**Märkõnad:** mobiilpositsioneerimine, ankurpunktide mudel, logistiline regressioon

# **The sensitivity of anchor point model to the data density and time as an input parameter**

Bachelor's thesis

Merili Konik

The aim of this bachelor's thesis is to test the sensitivity of the Positium's anchor point model to the data density and a change of one parameter. The parameter is called average record creation time. There are four different scenarios, including the base scenario, and each one has a different value of the named parameter. In this thesis the detected home anchor points of the described model are analyzed based on two Estonian mobile operators' data. The thesis gives an overview of the factors influencing human spatial behavior, mobile positioning and Positium's anchor point model. In the practical part of the thesis, the statistical analysis of the home anchor points is conducted and the logistic regression model is fitted to a binary dependent variable which shows if the assigned home anchor point coincides with the location of a cell that is geographically closest.

**CERCS research specialisation:** P160 Statistics, operations research, programming, financial and actuarial mathematics; S230 Social geography

**Key Words:** mobile positioning, anchor point model, logistic regression

# Sisukord

|  |           |
|--|-----------|
| <b>Sissejuhatus</b>  | <b>5</b>  |
| <b>1 Geograafiline taust</b>   | <b>7</b>  |
| 1.1 Inimese ruumiline käitumine . . . . .                            | 7         |
| 1.1.1 Tegevuskohad inimese ruumilises käitumises . . . . .           | 7         |
| 1.1.2 Variatiivsus inimese ruumilises käitumises . . . . .           | 8         |
| 1.2 Mobiilpositsioneerimine . . . . .                                | 9         |
| 1.2.1 Aktiivne ja passiivne mobiilpositsioneerimine . . . . .        | 9         |
| 1.2.2 Ankurpunkt geograafilisest distsipliinist tulenevalt . . . . . | 10        |
| 1.2.3 Probleemid seoses mobiilpositsioneerimisega . . . . .          | 10        |
| 1.3 Kodude tuvastamise mudelid . . . . .                             | 11        |
| <b>2 Positiumi meetod</b>  | <b>13</b> |
| 2.1 Alusandmed . . . . .   | 13        |
| 2.2 Kodu, tööaja ja multifunktsionaalne ankurpunkt . . . . .         | 13        |
| 2.3 Positiumi ankurpunktide mudeli kirjeldus . . . . .               | 14        |
| <b>3 Kasutatav metoodika</b>   | <b>18</b> |
| 3.1 Logistiline regressioon . . . . .                                | 18        |
| 3.2 Korduvmõõtmistega logistiline regressioon . . . . .              | 19        |
| <b>4 Andmete kirjeldus</b>   | <b>23</b> |
| 4.1 Stsenaariumid . . . . .  | 25        |
| <b>5 Tulemused</b>   | <b>26</b> |
| 5.1 Kirjeldav analüüs . . . . .                                      | 26        |
| 5.1.1 Mobiilsidetoimingute ülevaade . . . . .                        | 26        |
| 5.1.2 Määratud ankurpunktide ülevaade . . . . .                      | 26        |

|     |   |           |
|-----|---|-----------|
| 5.2 | Logistilise regressiooni mudeli rakendamine . . . . .                           | 35        |
| 5.3 | Korduvmõõtmistega logistilise regressiooni mudeli rakendamine . . . . .         | 36        |
|     | <b>Kokkuvõte</b>  | <b>38</b> |
|     | <b>Kirjandus</b>  | <b>40</b> |
|     | <b>Lisad</b>  | <b>43</b> |
|     | Lisa 1. Algandmete ülevaade . . . . .   | 43        |
|     | Lisa 2. Määratud ankurpunktide ülevaade . . . . .                               | 44        |
|     | Lisa 3. Logistilise regressiooni mudelite väljundid . . . . .                   | 49        |
|     | Lisa 4. Korduvmõõtmistega logistilise regressiooni mudelite väljundid . . . . . | 52        |
|     | Litsents . . . . .  | 55        |

# Sissejuhatus

Mobiilpositsioneerimise andmeid (MPD - *mobile positioning data*) on tänaseks Eestis ja maailmas kasutatud üle 10 aasta erinevates valdkondades, kus on oluline omada teadmisi inimeste liikumise ja paiknemise kohta, seda nii ajalises kui ruumilises dimensioonis.

Mobiilpositsioneerimise meetodi puhul räägitakse aktiivsest ning passiivsest mobiilpositsioneerimisest. Aktiivse mobiilpositsioneerimise teel saadud andmestik on küll ruumiliselt täpsem, kuid vajab andmete kogumiseks mobiilikasutaja poolset luba. Passiivne mobiilpositsioneerimine eraldi luba ei vaja, võimaldades seeläbi kaasata uuringutesse suurt hulka populatsioonist. Samas on passiivse mobiilpositsioneerimise geograafiline täpsus mõnevõrra madalam. (Ahas *et al.*, 2010) Käesolevas töös kasutatakse passiivse mobiilpositsioneerimise teel kogutud andmeid. Kõige tavalisem passiivse mobiilpositsioneerimise teel kogutud andmestik on detailne kliendi toimingute logi, näiteks sissetulevad ja väljuvad kõned, SMS-id, MMS-id ning andmekasutus (Vanhoof *et al.*, 2018a). Logifailides on lisaks mobiilikasutaja unikaalsele identifikaatorile fikseeritud iga toimingu aeg sekundilise täpsusega ning toimingu ajal ühendatud mobiilsideantenni unikaalne identifikaator. Igale mobiilsideantennile on võimalik omistada antud identifikaatori alusel geograafiline asukoht.

Passiivse mobiilpositsioneerimise puhul eristatakse sageli CDR (*call detail record*) ja DDR (*data detail record*) andmetüüpe. CDR andmed on kõne, SMS ning MMS toiminguid ning DDR andmed on andmeside kasutamise seotud toiminguid. DDR andmestik on enamasti suuremahulisem, kuna andmekasutus on võrreldes kõnetoimingutega sagedasem (Tiru, 2014).

Mobiilpositsioneerimist kasutatakse peamiselt geograafilistes uuringutes ja uurimistöodes, kuna nendes soovitakse vaadelda inimeste paiknemist ajas ja ruumis (Ahas *et al.*, 2010). Passiivse mobiilpositsioneerimise andmestikke kasutatakse laialdaselt mobiilsusuuringutes, transpordi- ja linnauuringutes (Ahas *et al.*, 2008), turismiga seotud uuringutes ja turismistatistikas (Saluveer *et al.*, 2020), reisikäitumise uuringutes (Poom *et al.*, 2017) jne. Teadmisi kasutatakse liikuvuse planeerimisel, ruumilisel planeerimisel, turismi ja ürituste korraldamise parendamisel, keskkonna- ja tervishoiuvaldkonnas jm. Mobiilpositsioneerimise kasutamine informatsioonitehnoloogias võimaldab personaliseerida mobiilteenuseid. (Ahas *et al.*, 2010) Lisaks kasutatakse mobiilpositsioneerimise andmeid uurimaks näiteks etnilisi erinevusi ruumilises käitumises pühade ajal ja inimeste üldist käitumist pühade ajal (Mooses *et al.*, 2016). Otses- teks kasusajateks on nii riigi ja omavalitsuste tasand kui ka eraettevõtted ja seeläbi kogu ühiskond. Vastavad teenused leiavad pidevalt uusi kasutusvõimalusi ja potentsiaalseid turge üle kogu maailma. (Ahas *et al.*, 2010)

Mobiilpositsioneerimise andmetel on mitmeid eeliseid traditsiooniliste küsitlusuuringu vormis liikumis-, käitumis- ja loendusuringute ees. Näiteks on võimalus koguda andmeid pikkade

perioodide vältel, väiksem andmelünkade hulk ning minimaalsed valesti tõlgendamisest tulenevad vead, mis võivad tekkida osalejate intervjuerimise käigus (Poom *et al.*, 2017). Andmete kogumise tehniline ja rutiinne vorm nõuab osalejatelt vähest ajalist pühendumust ja pingutust (Poom *et al.*, 2017). Vastavate tehnikatega saab mobiilpositsioneerimise andmete põhjal kindlaks määrata inimeste peamised tegevuskohad - kodu ja tööajaga seonduvad ankurpunktid. See tuleneb asjaolust, et kohustuslikel tegevustel (tööl, koolis käimine) ja kodukohaga seotud tegevustel on ära tuntavad päevased ja nädalased mustrid (Mooses *et al.*, 2016). Ankurpunktide määramisel võib aga väljakutseks osutada mobiilsideantennide ebaühtlane paigutus ja mobiilsidelevi levimise omapärad. Samuti individuaalsel tasandil olevad erisused seoses tegevuskohtadega ja iseloomulikud ning eripärased käitumismustrid.

Bakalaureusetöö eesmärk on uurida Positium OÜ (edaspidi Positium) väljatöötatud ankurpunktide mudeli tundlikkust kodu ankurpunkti tuvastamise näitel. Töös soovitakse teada saada, kui tundlik on Positiumi mudel eeldatava kodus olemise aja lävendi nihutamise ning andmetiheduse suhtes.

Bakalaureusetööl on praktiline väärtus: Positium arendab vastavat ankurpunktide mudelit ning soovib teada saada, millised tunnused mõjutavad mudelit ning kuidas. Käesolev töö on sisend Positiumi ankurpunktide mudeli arendamiseks. Lisaks annab töö üldisemalt teavet, milline on uuritud parameetrite mõju ankurpunktide mudeli tulemustele.

Käesoleva töö esimeses osas antakse lühike ülevaade inimese ruumilisest käitumisest ja tegevuskohtadest ning variatiivsusest inimese ruumilises käitumises. Samuti kirjeldatakse mobiilpositsioneerimise ja ankurpunkti mõistet ning tuuakse välja peamised probleemid seoses mobiilpositsioneerimisega. Teises osas seletatakse Positiumi ankurpunktide mudeli töökäiku. Töö kolmandas osas antakse ülevaade töös kasutatavast metoodikast, mida kasutatakse ankurpunkti mudeli tundlikkuse analüüsil ja sisendparameetri muutumise mõju väljaselgitamiseks.

Töö neljas ja viies peatükk on praktilised osad. Nendes kirjeldatakse kasutatavaid andmestikke ning analüüsitakse Positiumi ankurpunktide mudeli põhjal määratud ankurpunkte. Samuti koostatakse mudelid, et uurida püstitatud eesmärki ning hinnata Positiumi ankurpunktide mudeli tundlikkust.

Töö on koostatud kasutades tekstitöötlusprogrammi L<sup>A</sup>T<sub>E</sub>X. Analüüsiv osa on läbiviidud andmebaasisüsteemis PostgreSQL ja andmetöötlusprogrammis R (R Core Team, 2019). Töös kasutatavad tegeliku elukoha asukoha andmed pärinevad Eesti Statistikaameti uuringust, mille eesmärgiks oli olemasoleva ankurpunktide mudeli tulemuste valideerimine. Uuringu valimis olevate isikute mobiilpositsioneerimise andmed pärinevad kahelt Eesti mobiilsideoperaatorilt.

Autor tänab juhendajaid Anne Selartit ja Kaisa Venti täpsustavate kommentaaride ja täiendavate nõuannete eest.

# 1 Geograafiline taust

## 1.1 Inimese ruumiline käitumine

Inimese ruumiline käitumine on keerukas radade ja tegevuste muster ajas ja ruumis. Seda mõjutavad ühiskondlikud tavad, isiku elukeskkond ja sotsiaalne struktuur - peamiselt sotsio-ökonomilised tegurid. Samuti on mõjuteguriteks ka inimese individuaalsed vajadused, soovid, harjumused ja võimalused. (Järv *et al.*, 2013)

Ruumilise käitumise uurimiseks kasutatakse erinevaid meetodeid. Traditsioonilised meetodid on reisipäevikud, küsitlused ja intervjuud. (Griffin *et al.*, 2009) Küsitluspõhised uuringud on tavaliselt lühiajalised ja katavad vaid väga väikest ajalist perioodi. Pikaajaliste uuringute jaoks kasutatakse reisipäevikuid (Axhausen *et al.*, 2002, Crane *et al.*, 1998). Lisaks traditsioonilistele meetmetele on tänu arenevale tehnoloogiale järjest rohkem hakatud kasutama globaalse positsioneerimissüsteemi (GPS) funktsioone näiteks GPS jälgimist (Paršova, 2019), mobiilpositsioneerimist jne. Need andmekogumismeetodid võimaldavad samuti läbi viia pikaajalisi uuringuid. Lisaks aitab täpse anduritehnoloogia, näiteks mobiilpositsioneerimise, GPSi ning Bluetooth- või kiipkaarditehingute, rakendamine paremini mõista ja kaardistada inimeste liikuvust (Järv *et al.*, 2013).

Traditsioonilisi ruumilise käitumise uurimise meetodeid kasutatakse küll palju, aga neil on mitmeid puudusi. Näiteks on reisipäevikute meetod kallis ja aeganõudev. Samuti nõuavad traditsioonilised meetodid (küsitlused) vastajatelt suurt koormust, mistõttu on valimimahud väikesed ja andmestikud ebatäpsed. Lisaks võib tekkida suur mittevastamise protsent. GPS-andmete eeliseks on nende täpsus. Küll aga võib GPS-seadmete kasutamine andmete kogumisel inimesele pikema aja jooksul tülikas olla, kuna vastavat seadet peab pidevalt kaasas kandma. Seda aspekti aitab parandada mobiiltelefonides vastavate rakenduste kasutamine. Probleemseks kohaks on aga kasutatavate rakenduste suur energiakulu, mis vähendab aku tööaega. (Vent, 2014)

Mobiilpositsioneerimise andmete puhul on hea see, et inimesed ei pea nende kogumiseks ise vaeva nägema, kuid probleemseid kohti on MPD juures siiski. Näiteks on nendes vähem personaalset informatsiooni. Seetõttu võib nende andmete analüüsimine olla töömahukam, kuna vajaliku informatsiooni kätte saamine suurest hulgast andmetest on keerulisem. (Vent, 2014)

### 1.1.1 Tegevuskohad inimese ruumilises käitumises

Inimese ruumiline käitumine moodustab tema tegevusruumi. Tegevusruum koosneb tegevuskohadest. Tegevuskohad on inimeste külastatavad kohad või külastatav ruum (Ahas *et al.*,

2010). Need võib jagada kolmeks peamiseks osaks: 1) ruumiline käitumine kodus ja kodu lähimbruses; 2) regulaarselt külastatavatesse tegevuskohtadesse liikumine ja nendes olemine (töö, kool, pood jne); 3) ruumiline käitumine regulaarselt külastatavate tegevuskohtade lähimbruses (Vent, 2014). Tegevuskohad võib jagada olulisteks tegevuskohtadeks (eelnevalt mainitud punktid 1) ja 2)) ja juhuslikeks tegevuskohtadeks (eelnevalt mainitud punkt 3)). Isiku olulise tegevuskoha saab defineerida kui tihedalt külastatav koht, millel on oluline tähendus isiku igapäevaelus (Ahas *et al.*, 2010).

Uuringute põhjal on välja tulnud, et inimesed veedavad 80% ajast siseruumides (kodus, tööl, koolis jm). Tööl käivad inimesed veedavad 60% oma ajast kodus ehk üle poole ööpäevast veedetakse kodus (Farrow *et al.*, 1997). Seega võib võtta kodu kui kõige tihedamini külastatavat ja kõige olulisemat tegevuskohta (Hu *et al.*, 2015).

Andmetes tekitavad tihedasti ja regulaarselt külastatavad kohad mustreid. Seetõttu on isiku olulisi tegevuskohti automaatsete meetoditega lihtsam uurida. Juhuslike tegevuskohtade puhul regulaarsus puudub ja seega võib nende tuvastamine olla peaaegu võimatu.

### 1.1.2 Variatiivsus inimese ruumilises käitumises

Arenevas ühiskonnas on inimese ruumiline käitumine erinev ning kujunevad välja kindlad rütmid seoses erinevat tüüpi tegevuskohtadega. Näiteks on elamurajoon ja töö- ning ärikeskused ajaliselt vastupidised, viimastes on toiminguid peamiselt tehtud n-ö töö kellaegadel ja esimeses on toiminguid peamiselt tehtud väljaspool tööaega, st töönädal ollakse päevasel ajal töökohas ja öösel kodus (Ahas *et al.*, 2015). Uuringute käigus on leitud, et vaatamata inimeste liikumisvajadusele püsivad individuaalsed ruumilise käitumise ööpäevased rütmid pika perioodi vältel stabiilsed. Üldisemalt mõjutavad suurekaalulised ja pikaajalised otsused isiku igapäevaseid rütme. (Järv *et al.*, 2013)

Inimeste ruumilise käitumise nädalased rütmid on ruumiliselt ja ajaliselt samuti stabiilsed. Individuaalsel tasemel vaadatuna on tööperioodid rutiinsed, nädalavahetused või vabad päevad on tegevuste ja ruumilise ulatuse osas mitmekülgsemad. Ruumilise käitumise kuiseid rütme on peamiselt uuritud seoses sesoonsusega, kindlaid rütme üle mitmete uuringute pole täheldatud. (Järv *et al.*, 2013)

Lisaks mõjutavad loodus ja ilmastik ning sotsiaalne keskkond isiku päevaseid, nädalasi või isegi kuiseid ning aastaseid ruumilise käitumise rütme. Näiteks mõjutab sesoonsus inimeste ruumilist käitumist läbi (ühis)transpordi kasutuse: tihedamini kasutatakse seda märtsist maini ja kasutamise madalhooaeg on septembrist novembrini. Suvekuudel võib inimeste igapäevasesse rütmi sisse tulla ka teiste kodude (maakohtade) külastamine (Silm *et al.*, 2010).

## 1.2 Mobiilpositsioneerimine

Mobiilpositsioneerimine on asukohatuvastuse meetod, mille abil on võimalik teatud täpsusega kindlaks määrata telefonikasutaja asukoht. (Ahas *et al.*, 2010)

### 1.2.1 Aktiivne ja passiivne mobiilpositsioneerimine

Mobiilpositsioneerimine jaguneb kaheks: aktiivne ja passiivne. Aktiivset mobiilpositsioneerimist kasutatakse mobiiltelefoni asukoha kindlaks määramiseks spetsiaalse päringu abil, mis kasutab raadiolaineid (Ahas *et al.*, 2007). Asukoha päringu saamiseks on vaja spetsiaalset keskkonda ja telefonivaldaja luba (Ahas *et al.*, 2008). Aktiivset mobiilpositsioneerimist kasutatakse hädaabikõnede puhul, „leia sõber“ ja paljude teiste asukohapõhiste teenuste rakenduste korral (Ahas *et al.*, 2008). Võrreldes passiivse mobiilpositsioneerimisega on aktiivne mobiilpositsioneerimine täpsem. Passiivse mobiilpositsioneerimise andmed salvestuvad logifailidena mobiiloperaatori süsteemis. Neid andmeid kasutab operaator klientidele arvete koostamiseks, aga ka võrgutegevuse monitoorimiseks, optimeerimiseks ja planeerimiseks. Kõige tavalisem meetod vastavate andmete saamiseks on kõnelogi (*billing log*), milleks on mobiiltelefoni iga aktiivne kasutus - väljaminev kõne, SMS või MMS ja mobiilne andmeside. (Ahas *et al.*, 2010)

Passiivse mobiilpositsioneerimise andmed salvestatakse mobiilsideantenni täpsusega, seega andmete täpsus sõltub mobiilsideantennide paigutusest: mida hõredamalt antennid paiknevad, seda ebatäpsemaks muutub positsioneerimine. Iga mobiilsideantenniga saab seostada tema geograafilise asukoha, mis on tavaliselt esitatud punktasukohana. Samuti on igal mobiilsideantennil konkreetne geograafiline katvusala ja unikaalne identifitseerimisnumber.

Mobiilsidevõrk põhineb tugijaamade komplektil, kus ühe asukohaga võib seostada sageli enam kui ühte mobiilsideantenni. Mobiilsideantenni leviala ja mobiilsidevõrgu suurus pole fikseeritud. Mobiiltelefon ühendub selle mobiilsideantenniga, mille poolt saadetakse signaal on tugevaim ning millel on vaba ressursi uute kasutajate vastuvõtmiseks. Mobiilsideantennide levialad varieeruvad suuresti: linna keskosades võib minimaalne leviala olla vaid mõnisada ruutmeetrit, maapiirkondades võib see ulatuda kuni sadade ruutkilomeetriteni (Tiru *et al.*, 2010). Maksimalse teoreetilise leviala ulatus on 35 kilomeetrit. Vähem asustatud ja rannikualadel kasutatakse võimendatud mobiilsideantenne, mis levivad suurematel aladel. Seetõttu saab eeldada, et maapiirkondades, kus on vähem asustust, on mobiilsideantennid hõredamalt ja seega võib passiivse mobiilpositsioneerimise täpsus ulatuda mitmete kilomeetriteni. Samas kui linnades, kus asustus on tihe, asuvad mobiilsideantennid vaid mõnesaja meetri kaugusel üksteisest. Sellest järeldub, et mobiilside võrgu katvus on ebahühtlane ja sõltub mobiilsideantennide paigutusest, mis omakorda sõltub populatsiooni tihedusest ja transpordi infrastruktuurist. (Ahas *et al.*, 2010)

### 1.2.2 Ankurpunkt geograafilisest distsipliinist tulenevalt

Inimesed on väga mobiilsed ning nende igapäevaelu hõlmab erinevate kohtade külastamist ja nendes viibimist. Võib öelda, et mida rohkem inimene mingis kohas aega veedab, seda olulisem on vastav koht tema jaoks. Seega kohti, kus inimesed tihti aega veedavad, võib nimetada ankurpunktideks. Ankurpunktid on olulised, kirjeldamiseks inimese oluliste tegevuskohtade paiknemist ajas ja ruumis. (Ahas *et al.*, 2010)

Ankurpunktid võib tähenduse põhjal jagada kaheks: üldised ja personaalsed. Üldised ankurpunktid on olulised kohad keskkonnas, mida üldiselt tuntakse ja külastatakse, näiteks park, kaubanduskeskus. Personaalsed ankurpunktid on seotud isiku ruumilise käitumise ja tegevusega. Personaalsed ankurpunktid, peamiselt kodu, töö ja muud kohad, leitakse külastatavuse sageduse ja regulaarsuse alusel. (Ahas *et al.*, 2010)

Kasutatavates andmestikes, MPD puhul, on informatsiooni ainult inimeste mobiilikasutuse kohta ning puudub muu teave nende tegevuste ja inimeste enda kohta (sugu, vanus, elukoht jne). Seega sõltub passiivse mobiilpositsioneerimise andmete edasine kasutus ja usaldusväärsus suuresti andmetele rakendatud mudelite ning meetodite täpsusest ja usaldusväärsest. (Ahas *et al.*, 2010)

### 1.2.3 Probleemid seoses mobiilpositsioneerimisega

Mobiilpositsioneerimise andmetega seoses on peamine probleem privaatsusnõuded. Andmed, mida mobiilioperaatorid välja annavad, on pseudonümiseeritud, mistõttu puudub faktiline informatsioon mobiiliomaniku kohta - pole teada isiku sugu, vanus, elukoht jne (Ahas *et al.*, 2010). Täiendava isikliku informatsiooni puudumine võtab ära võimaluse kasutada mudeleid ja lähenemisi, mis põhinevad individuaalsel informatsioonil. (Vanhoof *et al.*, 2018a)

Samuti on kasutatavaid mudeleid keeruline valideerida, kuna mudeli valideerimiseks on vaja läbi viia küsitlusuuring. Seejärel peab uuringust saadud andmed siduma mobiilioperaatoritelt saadud andmetega. Meetodi või mudeli väljatöötamise positiivseks küljeks on see, et enamasti enamasti tehakse üks uuring ning üldjuhul rohkem pole vaja. Kui mudel on kord juba edukalt valideeritud, siis saab seda kasutada ilma täiendava isikustatud informatsioonita. Valideerimise tulemusena on mudel täpsem ja usaldusväärsem. (Vanhoof *et al.*, 2018b)

Ankurpunkti määramise juures on probleemiks mobiilsidevõrgu ebaühtlane jaotumine. Mobiilsideantennide hulk on väiksem maapiirkondades ja piirkondades, kus asustustihedus on väike. Seetõttu paiknevad mobiilsideantennid hõredamalt ja nende vahelised vahemaad on suuremad. Sellega kaasneb ebatäpsem mobiilpositsioneerimine ja määratud ankurpunktid võivad paikneda tegelikest isikule olulistest asukohtadest kaugel. (Ahas *et al.*, 2010) Samuti võib esi-

neda riigis piirkondi, mida mobiilsidelevi ei kata. Eestis sellega probleeme pole riigi väiksuse tõttu, aga suuremates riikides võib esineda vähem asustatud alasid, kus mobiililevi puudub. (Tiru *et al.*, 2010) Lisaks on mobiilpositsioneerimise andmestikud väga suured ning nende töötlemine on tehniliselt keeruline, aeganõudev ja kallis (Vanhoof *et al.*, 2018a).

Passiivse mobiilpositsioneerimise andmeid kasutava Positiumi ankurpunktide mudeli puhul on probleemiks see, et kui isikul on rohkem kui üks töökoht, liikuv töö või kui isik töötab kodus, siis on mudelil seda raske tuvastada kui isiku tööaja ankurpunkti (Ahas *et al.*, 2010). Mitme kodu olemasolul ei pruugi vähem külastatav elukoht kodu ankurpunktina välja tulla.

### 1.3 Kodude tuvastamise mudelid

Selles punktis antakse ülevaade erinevatest kodutuvastuse algoritmidest (HDA - *home detection algorithm*).

Isikute kodude tuvastamine on üks olulisemaid aspekte mõistmaks inimese elu ja keskkonda. Teades inimeste kodude paiknevust, saavad teadurid näiteks modelleerida populatsiooni hajuvust, uurida inimese käitumismustreid, teha järeldusi erinevate elustiilide kohta ning mõista seoseid kodu paiknemise ja muude eluliste aspektide, näiteks tervises seisundi, haiguste leviku ja keskkonnamuutuste vahel. (Hu *et al.*, 2015)

Passiivse MPD alusel on ankurpunktide tuvastamise ja hindamise metoodikat kasutades võimalik isiku ajalise ja ruumilise käitumismustri alusel välja arvutada tema peamine elukoht (kodu või kodu-töö ankurpunkt) ja peamine tööaja tegevuskoht (selle olemasolul). Kui inimene paikneb ka tööajal koduankurpunktis, siis nimetatakse selline ankurpunkt kodu-töö ankurpunktiks.

Lisaks mobiilpositsioneerimise andmetele on kodu ankurpunkte võimalik tuvastada ka GPS andmete põhjal. Nende puhul võib tegevuskohtade ja ankurpunktide identifitseerimist võtta kui klasterdamise ülesannet. Klassikaline klasterdamise viis tuvastamiseks ankurpunkte vaatab üldist külastatud punktide jada ning määrab selle järgi kaugusel ja ajal põhinevad parameetrid klastrite tuvastamiseks. Klastrite moodustavad punktid, mis on tihedalt üksteise lähedal, st ei paikne üksteisest kaugemal kui määratud kaugus. Algoritm testib vaatlusi iteratiivselt, et kindlaks määrata lähestikku asetsevad punkte. Samuti on klastrite määramisel oluline aja parameeter, kui aeg esimesena ja viimasena vaadatud punkti vahel on suurem määratud parameetri suurus, siis moodustatakse klaster eelnevalt vaadeldud punktide põhjal. Tegevuskoha asukohaks määratakse klastrite keskpunkt. (Thierry *et al.*, 2013)

GPS-andmetest tegevuskohtade tuvastamiseks kasutatakse ka Kerneli tihedusfunktsiooni. See lähenemine ei analüüsi andmepunkte järjestikku, vaid kasutab GPS punkte, et luua Kerneli

tiheduse pind. Nimetatud meetodil on vaja määrata kaugusel ja ajal põhinevad parameetrid. (Thierry *et al.*, 2013)

Kodusid on võimalik tuvastada ka reisiväikute alusel või kombineerides GPS-andmeid ja reisiväikute sissekandeid. Viimasest kombinatsioonist saadud andmete peal läbi viidud analüüsist on tulnud välja, et täpseima kodukohta saab tuvastada indikaatori alusel, mis näitab päeva viimast sihtkohta (Krumm, 2007).

Tänapäeva maailmas on sotsiaalmeedial suur roll ja paljudel platvormidel on kasutajatel postituse juurde võimalik märkida ka geograafiline asukoht. Sotsiaalmeediast saadavaid geoinformatsiooniga andmeid on võimalik kasutada ka kodude tuvastamiseks. Selliste andmete põhjal on võimalik tuvastada kodulinna, -maakonda või ajatsooni. Kodulinna määramise täpsused varieeruvad, parim saavutatud täpsus on olnud 82% (Pontes *et al.*, 2012). Kodu kaudset asukohta on võimalik määrata ka isiku sotsiaalmeedia *check-in*-ide põhjal, arvutades nende kohtade geograafiliselt keskmise punkti. (Hu *et al.*, 2015)

Seega kodude tuvastamise mudeleid on mitmeid ja saadud tulemuste informatiivsus sõltub mitmeti ka kasutatud andmetest. Järgmises peatükis tutvustatakse Positiumi ankurpunktide mudelit, mis kasutab passiivse mobiilpositsioneerimise andmeid ning mille abil on võimalik tuvastada isiku kodu, tööaja ja sekundaarseid ankurpunkte.

## 2 Positiumi meetod

### 2.1 Alusandmed

Passiivse mobiilpositsioneerimise andmed salvestatakse automaatselt mobiilioperaatorite logifailides. Positiumi ankurpunkti mudel kasutab ankurpunktide määramisel just logifailidesse salvestatud andmeid. Vastavad failid sisaldavad kõnetoimingute (CDR) ja andmeside (DDR) andmeid. CDR andmed on isiku tehtud kõne, SMS või MMS. DDR andmed salvestatakse iga andmeside kasutuskorra puhul. Mobiilioperaatoritelt tulevad passiivse mobiilpositsioneerimise andmed on pseudonümiseeritud.

MPD sisaldab järgmisi parameetreid:

- aktiivse tegevuse täpne kuupäev ja kellaaeg;
- juhuslik ja unikaalne ID number iga telefoninumbri jaoks, mis jääb samaks kogu vastava numbri kasutamise ajaks;
- mobiilsideantenni ID koos geograafiliste koordinaatidega.

Näide andmestiku reast: Time September 8. 2007. 22:03:11; ID 64353; Location E27-44-39.00 N59-25-49.00. (Ahas *et al.*, 2010)

### 2.2 Kodu, tööaja ja multifunktsionaalne ankurpunkt

Ankurpunktide leidmine muudab MPD uuringute jaoks paremini kasutatavaks - see annab andmetele juurde väärtust ning muudab andmed uurijatele informatiivsemaks (Ahas *et al.*, 2010). Ankurpunktide leidmiseks kasutatakse mudeleid, mis määravad igale isikule kuised kodu ja tööaja ankurpunktid või multifunktsionaalse ankurpunkti ning sekundaarsed ankurpunktid.

Kodu ankurpunkt tuvastatakse kodutuvastusalgoritmidega (HDA). Indiviidi kodu ankurpunktiks loetakse ühte mobiilsideantenni või lähestikku paiknevate mobiilsideantennide gruppi, mille levialas on etteantud parameetritega indiviidi tegeliku kodu kõige tõenäolisem asukoht. HDA usaldusväärsust näitab tõenäosus, kui tõenäoliselt asub indiviidi tegelik elukoht identifitseeritud kodu ankurpunkti levialas - näiteks võrreldes algoritmi määratud ankurpunkte indiviidi tegeliku elukohaga või võrreldes määratud ankurpunktide osakaalu mingis piirkonnas sama piirkonna populatsiooni osakaaluga. (Vanhoof *et al.*, 2018a)

Kodu ankurpunktide tuvastus on oluline mitmel juhul. Isikute elukohtade paiknemise teadmine ja sellest lähtuvalt isikute liikumise määramine teiste ankurpunktide vahel on aluseks

mobiilsusuuringutele. Näiteks on selle informatsiooni põhjal võimalik uurida ja planeerida transpordiühendust (Ahas *et al.*, 2008).

Indiviidi tööaja ankurpunkt on igapäevane ankurpunkt ehk üks mobiilsideantenn või lähestikku paiknevate mobiilsideantennide grupp, mille levialas asub indiviidi töökoht kõige tõenäolisemalt. Ankurpunkti kutsutakse tööaja ankurpunktiks, sest ei ole võimalik eristada tööd, kooli ja muid isiku regulaarseid asukohti, kus isik veedab enamuse aja päevast n-ö töötundidel ühe kuu jooksul. (Ahas *et al.*, 2010)

Indiviidi multifunktsionaalne ankurpunkt on igapäevane ankurpunkt ehk üks mobiilsideantenn või lähestikku paiknevate mobiilsideantennide grupp, mille levialas indiviidi tegelik kodu ja töökoht kõige suurema tõenäosusega asuvad. Multifunktsionaalne ankurpunkt määratakse indiviidile, kui ei ole võimalik eristada kodu- ja tööaja ankurpunkte. (Ahas *et al.*, 2010) Need ankurpunktid on segatunnustega ja ideaalsel juhul määratakse isikule, kelle kodu ja töö asuvad samade mobiilsideantennide levialas, kuid esineb ka erandeid.

## 2.3 Positiumi ankurpunktide mudeli kirjeldus

Mudel, mis tuvastab isiku olulised asukohad või personaalsed ankurpunktid, koosneb viiest põhisammust, mis omakorda koosnevad mitmest alamsammust. Algoritm vaatab korraga ühte kuud ning isikule tuvastatavad ankurpunktid määratakse vastava kuu lõikes.

Esimese sammuna arvutatakse igale telefonikasutajale mobiilsideantennide kasutusstatistika s.t kui palju andmekirjeid on iga antenniga seotud. Järgmisena järjestatakse iga isiku puhul mobiilsideantennid kahanevalt vastavalt mobiilsideantenni kasutuse päevade arvule ja toimingute rohkusele. Seejärel grupeeritakse vastavalt kasutuse tihedusele naabrussuhete algoritmi alusel lähestikku paiknevad antennid kokku. Põhimõte seisneb selles, et mobiilsideantennid, mille levialad on ülekattes, satuvad ühte mobiilsideantenni gruppi. Üks antenn saab kuuluda vaid ühte gruppi korraga. Põhjus, miks mobiilsideantenne grupeeritakse, on mobiililevi vahetumine mobiilsideantennide vahel, st ühendus võib liikuda mobiilsideantennide vahel isegi kui isik asub samas kohas või toas. See on tavaline juhul, kui mobiiltelefon asub levialade piiril ning mobiililevi kantakse suurema koormusega antennilt väiksema koormusega antennile või mobiilsideantennile, mis on paremini nähtav. Antennigruppe saab vaadata kui potentsiaalseid ankurpunkte. Selle sammu lõpuks on tuvastatud esialgsed ankurpunktid, millel pole veel kindlat tähendust.

Teise sammuna arvutatakse iga kasutaja igale antennigrupile gruppi iseloomustavad näitajad. Arvutatavad näitajad on järgmised:

1. toimingute arv;

2. päevade arv, mil toiminguid on tehtud;
3. unikaalsete mobiilsideantennide arv;
4. keskmine kõnede alustamise aeg;
5. kõnede alustamise aja standardhälve;
6. nädalavahetuse päevade hulk.

Iga ankurpunkti asukoht on defineeritud tugevaima mobiilsideantenni järgi tervest antennigrupist. Lisaks on igal ankurpunktis loend mobiilsideantennidest, mis on antud ankurpunktiga seotud. Ankurpunktis peab olema vaatluse all oleva kuu jooksul tehtud toiminguid vähemalt kahel erineval päeval.

Lisaks järjestatakse antennigrupid päevade arvu ja andmekirjete hulga alusel. Esimese koha saab antennigrupp, milles on tehtud toiminguid kõige rohkematel päevadel, teise koha saab antennigrupp, mille puhul päevade arv suuruselt teisel kohal jne. Juhul kui mitmel antennigrupil on päevade arv sama, siis järjestatakse need vastavalt kõnetoimingute arvule.

Kolmanda sammuna teostatakse andmete puhastamine erandlikest kasutajatest. Välja jäetakse kasutajad, kellel on esimesel kohal olevas ankurpunktis tehtud toiminguid vähem kui neljal päeval ühes kuus. Liiga väikese toimingute arvu põhjal ei ole võimalik isikule ankurpunkte määrata (Ahas *et al.*, 2010). Samal põhimõttel eemaldatakse ka liiga suure toimingute arvuga ID-d, kuna tõenäoliselt on need organiseeritud kõneprotseduurid (teeninduskeskused) või GSM-võrku kasutatav tehniline seade (Ahas *et al.*, 2010). Liiga suur kõnetoimingute arv võib andmestike lõikes varieeruda - see määratakse andmestiku esmase analüüsi põhjal. Üldjuhul on see mitu tuhat kõnetoimingut kuus kahe kõige sagedamini kasutatava mobiilsideantenni peale kokku.

Neljanda sammuna lisatakse ankurpunktidele unikaalsed identifitseerimisnumbrid, et järgnevate arvutuste käigus siduda kuiseid ankurpunkte omavahel.

Viienda sammuna antakse saadud ankurpunktidele tähendus. See koosneb kümnest alamsammust:

1. Alustatakse kodu ankurpunkti otsinguga antennigruppide hulgast, mille järjekorranumber ei ole suurem kui 2.
2. Kodu ankurpunkti silt „h“ määratakse ankurpunktile, millest on toiminguid tehtud vähemalt 4 päeval kuus ja keskmine kõnede alustamise aeg on peale 17:00.
3. Tööaja ankurpunkti silt „w“ määratakse järgnevatele ankurpunktidele:

- puudub silt;
- koha number pole madalam kui 2;
- toiminguid on tehtud vähemalt 4 päeval kuus;
- kõnede alustusaja standardhälve on väiksem kui või võrdne 0,175;
- nädalavahetuste päevade osakaal kõikide päevade suhtes on väiksem kui 0,2.

Eelnevad analüüsid on näidanud, et kronoloogiline varieeruvus on väiksem kõnedel, mis on tehtud tööajal. Töökohas oldud aeg on kompaktsem ja üldiselt inimesed väljaspool tööaega töökohas aega ei veeda. (Ahas *et al.*, 2010) Sellest tulenevat on valitud ka tööaja ankurpunkti määravad parameetrid, näiteks nädalavahetuste päevade osakaal kõikide päevade suhtes ja kõnede alustusaja standardhälve. Standardhälbe kasutamine aitab mõista ankurpunkti ajakasutuse kompaktsust - tööl ollakse enamasti vaid osa päevast - näiteks 8 tundi, mitte terve päev nagu see võib olla kodude puhul. Standardhälbe arvulist väärtust saab interpreteerida järgmiselt: kõnede algusaja standardhälve on 1, kui toimingute algusajad on jaotatud võrdselt 24 h vahel. Standardhälve 0,175 märgib 4,2 tundi ehk toiminguid on tehtud keskmisest kõnede alustamise ajast mõlemale poole kokku 8,4 tundi. Selle põhjal saab öelda, et inimene viibib antud ankurpunktis (tööaja ankurpunktis) päeva jooksul maksimaalselt üle kolmandiku ajast.

4. Juhul, kui eelneva kolme sammuga ei määratud kahe kasutatavama ankurpunkti seast kodu või tööaja ankrut, siis määratakse sildistamata ankurpunktidele (nende kahe seast) kodu ankurpunkti silt „h“ (eeldusel, et nendes on toimingud tehtud vähemalt 4 päeval kuus).
5. Kui suhe - teise kõige kasutatava ankurpunkti (koht 2) summaarne unikaalsete päevade arv jagatud esimese kõige kasutatavama ankurpunkti (koht 1) summaarse unikaalsete päevade arvuga - on väiksem kui või võrdne 0,25, siis muudetakse kõige külastatavam ankurpunkt (esimesel kohal olev ankurpunkt) multifunktsionaalseks ankurpunktiks ja lisatakse sellele silt „m“. Sisuliselt tähendab see, et kui ühe ankurpunkti osakaal on teistega võrreldes palju suurem, siis loetakse see multifunktsionaalseks ankurpunktiks.
6. Tuvastatakse isikud, kellel on eelnevate punktide järel kõik ankurpunktid tuvastatud ainult kui kodu ankurpunktid või ainult kui tööaja ankurpunktid. Seejärel valitakse järgmine sildistamata ankurpunkt ja proovitakse sellele määrata silt „h“ või „w“ põhinedes keskmisele kõnede alustusajale, maksimaalsele lubatud standardhälbe väärtusele (tööaja tüüp), nädalavahetuste päevade osakaal kõikide päevade suhtes (tööaja tüüp) ja miinimum lubatud külastatud päevade arv vastavas ankurpunktis.
7. Multifunktsionaalse ankurpunkti silt „m“ määratakse isikule juhul, kui isikul on tuvastatud ainult kodu või ainult tööaja ankurpunkt, siis vastav esimesel kohal olev ankurpunkt

muudetakse multifunktsionaalseks.

8. Multifunktsionaalse ankurpunkti silt „m“ määratakse esimesel kohal olevatele ankurpunktidele isikutele, kellel pole tööaja ankurpunkti sildiga „w“.
9. Multifunktsionaalse ankurpunkti silt „m“ määratakse isikutele, kellel esimesel kohal olev ankurpunkt on sildiga „h“ või „m“ aga teisel kohal olev ankurpunkt on sildistamata.
10. Kõikidele ülejäänud ankurpunktidele määratakse silt „o“ - teised ankurpunktid.

Kokkuvõttena saab mudeli kohta öelda, et:

1. Kodu ja tööaja sildid määratakse ankurpunktidele, milles on toiminguid tehtud vähemalt neljal erineval päeval kuus ja mis on kasutuselt kahe kõige kasutatavama ankurpunkti hulgas. Mõnel juhul sobib ka kolmas kõige kasutatavam ankurpunkt - vt samm 5 alasamm 6. Vastavalt sellele võib isikul olla maksimaalselt kolm kodu ja tööajaga seotud ankurpunkti.
2. Sildiga „o“ on kõik ankurpunktid, mis ei ole kodu ja tööajaga seotud ankurpunktid ja milles on toiminguid tehtud vähemalt kahel päeval kuus.

(Vent, [2020](#))

### 3 Kasutatav metoodika

Järgnevalt kirjeldatavaid mudeleid kasutatakse töös hindamaks binaarseid tunnust, mis näitab kuiste kodu ankurpunktide kauguse kokkulangemist lähima mobiilsideantenni asukoha kaugusega kodust.

#### 3.1 Logistiline regressioon

Logistilise regressiooni mudel kirjeldab seost sõltuva binaarse tunnuse ja ühe või mitme kirjeldava tunnuse vahel.

Olgu  $Y$  sõltuv tunnus, mille väärtused on 0 või 1 ja mida võib vaadata kui Bernoulli jaotusega  $B(\pi)$  juhuslikku suurust, kus  $\pi = P(Y = 1)$  ja ühtlasi  $\pi = EY$ . Olgu vaatluse all  $k$  sõltumatut tunnust  $X_1, X_2, \dots, X_k$ .

Et siduda uuritav tõenäosus  $\pi$  argumenttunnustega, vaadatakse  $Y$  keskvaartust tinglikuna:  $\pi = E(Y|X_1, X_2, \dots, X_k)$  ja selle põhjal saab öelda, et mudel keskvaartusele hindab sündmuse toimumise tõenäosust.

Logistiline mudel hindab logaritmilist šanssi ja mis on sündmuse esinemise tõenäosuse ja sündmuse mitteesinemise tõenäosuse suhe, mis avaldub kujul  $\frac{\pi}{1-\pi}$ . Mudeli puhul kasutatakse *logit* seosefunktsiooni, et tõenäosuse skaalalt, mis on 0 ja 1 vahel, üle minna kogu reaalteljel muutuvale suurusele:

$$\text{logit}(\pi) = \ln \frac{\pi}{1-\pi}.$$

Seega logistilise regressiooni mudel on defineeritud järgmiselt:

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k,$$

kus kordajad  $\beta_0, \beta_1, \dots, \beta_k$  on tundmatud parameetrid. Defineeritud mudeli korral eeldame, et šansi logaritm avaldub sõltumatute tunnuste kaudu lineaarselt.

Saab avaldada huvipakkuva sündmuse toimumise tõenäosuse  $\pi = P(Y = 1|X_1, X_2, \dots, X_k)$  ehk prognoosi tõenäosuse, mis on antud mudeli puhul

$$\pi = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}.$$

Sellist teisendust võib vaadata kui logistilise jaotuse jaotusfunktsiooni väärtust kohal lineaarkombinatsioon, mis tagab selle, et tõenäosus jääb logistilise regressiooni puhul alati 0 ja 1 vahele st õigetesse piiridesse.

Logistilise regressioonimudeli tulemusi saab interpreteerida šansside suhte ( $OR$ ) kaudu, mis on vaatluste  $i$  ja  $j$  šansside suhe ja avaldub järgmiselt:

$$OR = \frac{\frac{\pi_i}{1-\pi_i}}{\frac{\pi_j}{1-\pi_j}}.$$

Vastav väärtus muutub vahemikus  $[0, \infty)$ .

Kui  $OR$  on positiivne, siis on tegemist samapidise seosega argumendi ja sündmuse esinemise tõenäosuse (uuritava tunnuse) vahel. Negatiivse  $OR$  korral on tegemist vastupidise seosega. Mudeli paremeetrite ja šansside suhte seost saab vaadelda näite põhjal, kus mudelil on üks argumenttunnus  $X$ . Olgu edaspidi  $\ln \frac{\pi}{1-\pi} = \alpha + \beta X$  ja  $\frac{\pi}{1-\pi} = e^{\alpha+\beta X}$ . Logistilise regressiooni mudeli kordaja eksponentväärtus  $e^\beta$  näitab, mitu korda muutub sündmuse toimumise šanss argumendi muutumisel ühe ühiku võrra, st

$$e^{\alpha+\beta(X+1)} = e^\alpha e^{\beta X} e^\beta = e^{\alpha+\beta X} e^\beta = e^\beta \frac{\pi}{1-\pi}.$$

Sellest järeldub, et seos šansside suhte ja regressiooni kordajate vahel avaldub järgmiselt:

$$e^\beta = OR.$$

Saab näidata, et seos  $e^\beta = OR$  kehtib. Kui tunnuse  $X$  väärtus on  $X = x_0$ , siis olgu sündmuse toimumise  $Y = 1$  šanss  $O_{x_0}$  ja kui tunnu  $X$  väärtus on  $X = x_0 + 1$ , siis olgu sündmuse toimumise  $Y = 1$  šanss  $O_{x_0+1}$ . Eelnevast on teada, et

$$O_{x_0} = e^{\alpha+\beta x_0}$$

ja

$$O_{(x_0+1)} = e^{\alpha+\beta(x_0+1)}.$$

Seega

$$OR = \frac{O_{(x_0+1)}}{O_{x_0}} = e^\beta.$$

Mitme argumentiga mudeli puhul peab šansside suhte tõlgendamisel lisama, et teised argumentid ei muutu, st muud tingimused jäävad samaks.

(Hosmer *et al.*, 2013)

## 3.2 Korduvmõõtmistega logistiline regressioon

Töös uuritavat binaarset tunnust saab vaadata korduvalt mõõdetuna, sest mõõtmised on tehtud neljal kuul ning samadele objektidele on kodu ankurpunktid määratud mitme erineva

stsenariumi järgi. Korduvate mõõtmistega arvestamiseks kasutatakse töös üldistatud lineaarset segamudelit. Ühel objektil tehtud mõõtmiste sõltuvusi võetakse arvesse lisades mudelisse indiviidi juhusliku mõju.

Olgu  $N$  sõltumatut objekti vaadeldud korduvalt mingi perioodi vältel. Olgu  $n_i$  vaatluste arv  $i$ . objektil.

Olgu  $Y_{ij}$  binaarne tunnus väärtusega 0 või 1, mis näitab  $i$ . objekti  $j$ . vaatlust. Eeldatakse, et iga  $Y_{ij}$  vaatlus on tehtud ajal  $t_{ij}$ . Sõltuva tunnuse  $i$ . objekti vaatlused saab esitada vektorina järgmiselt:

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix}, \quad i = 1, \dots, N;$$

kus vaatluste vektorid  $\mathbf{Y}_i$  on omavahel sõltumatud.

Tähistame indiviid  $i$  juhusliku mõju  $b_i$ . Töös kasutavate korduvmõõtmistega mudelites on juhuslikku mõju  $b_i$  vaja kahel juhul. Esimesel juhul esitab see indiviidi omapära ajas tehtud mõõtmistes: kodu ankurpunkte vaadatakse kuude lõikes (märts, aprill, mai, juuni). Teisel juhul esitab see indiviidi omapära kodu ankurpunkti määramisel nelja erineva stsenaariumi järgi: stsenaarium 0, stsenaarium 1, stsenaarium 2, stsenaarium 3.

Sarnaselt tavalise logistilise regressiooniga on  $Y_{ij}$  Bernoulli jaotusega  $B(\pi)$  juhuslik suurus, kus  $\pi = P(Y_{ij} = 1|b_i)$ . Uuritav tunnus  $Y_{ij}$  on keskvärtusega  $\pi = E(Y_{ij}|b_i)$  ja dispersiooniga  $D(Y_{ij}|b_i) = E(Y_{ij}|b_i) \cdot [1 - E(Y_{ij}|b_i)]$ .

Olgu vaatluse all  $k$  sõltumatut tunnust  $X_1, X_2, \dots, X_k$ . Sõltumatute tunnuste vaatluse vektori  $i$ . objektil  $j$ . hetkel saab esitada kujul:

$$\mathbf{X}_{ij} = \begin{pmatrix} X_{ij1} \\ X_{ij2} \\ \vdots \\ X_{ijk} \end{pmatrix}, \quad i = 1, \dots, N; j = 1, \dots, n_i.$$

Sõltumatute tunnuste vaatlusi saab grupeerida. Selle tulemusena moodustub vaatluste maatriks:

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{X}'_{i1} \\ \mathbf{X}'_{i2} \\ \vdots \\ \mathbf{X}'_{in_i} \end{pmatrix} = \begin{pmatrix} X_{i11} & X_{i12} & \cdots & X_{i1k} \\ X_{i21} & X_{i22} & \cdots & X_{i2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{in_i1} & X_{in_i2} & \cdots & X_{in_ik} \end{pmatrix}, \quad i = 1, \dots, N,$$

kus veerud näitavad tunnust ja read näitavad tunnuste vaatlusi.

Mudel hinnatakse tinglikule keskväärtusele  $E(Y_{ij}|b_i)$ , mis ühtlasi näitab sündmuse toimumise tõenäosust, ning seotakse fikseeritud ja juhuslike mõjude lineaarkombinatsiooniga. Korduvmõõtmistega logistiline mudel on defineeritud järgmiselt:

$$\ln \left( \frac{P(Y_{ij} = 1|b_i)}{P(Y_{ij} = 0|b_i)} \right) = \ln \frac{\pi}{1 - \pi} = \mathbf{X}'_{ij}\boldsymbol{\beta} + \beta_0 + b_i,$$

kus  $\boldsymbol{\beta}$  on regressiooni parameetrite vektor. Parameetri vektori  $\boldsymbol{\beta}$  mõjud on fikseeritud mõjud.

Saab avaldada huvipakkuva sündmuse toimumise  $\pi = P(Y_{ij} = 1|b_i, X_{ij1}, \dots, X_{ijm}, \dots, X_{ijk})$  ehk prognoosi tõenäosuse, mis on mudeli puhul:

$$\pi = \frac{e^{b_i + \beta_0 + \beta_1 X_{ij1} + \dots + \beta_m X_{ijm} + \dots + \beta_k X_{ijk}}}{1 + e^{b_i + \beta_0 + \beta_1 X_{ij1} + \dots + \beta_m X_{ijm} + \dots + \beta_k X_{ijk}}}.$$

Eeldame, et iga indiviidi korral on juhuslik mõju  $b_i$  ühemõõtmelise normaaljaotusega  $N(0, g_{11})$ .

Korduvmõõtmistega logistilise regressiooni mudeli puhul on parameetri vektori  $\boldsymbol{\beta}$  ühe komponendi  $\beta_m$  tõlgendamiseks kaks varianti. Üks komponent  $\beta_m$  on mingi argumenttunnuse  $X_m$  kordaja mõju.

Esiteks saab kordajat  $\beta_m$  tõlgendada indiviidi siseselt, st juhul, kui juhuslik mõju  $b_i$  ja teised mudeli argumendid  $X_m$  püsivad muutumatuna.

Töös kasutatava mudeli põhjal saab nii tõlgendada kordajat  $\beta_m$  näiteks kuu või stsenaariumi korral: sellisel juhul vaadatakse muutust tunnuses  $X_m$  indiviidisiseselt. Seega nende tunnuste puhul on interpretatsioon sama nagu logistilise regressiooni mudelis.

Teisel juhul pole kordaja  $\beta_m$  tõlgendus nii selge ega ühene - siin vaadatakse muutust tunnuses indiviidide üleselt, näiteks tunnuse mobiilioperaator puhul. Sellisel juhul peab eeldama, et eksisteerib kaks indiviidi  $i_1$  ja  $i_2$ , kellel mõlemal on juhuslik mõju  $b_i$  sama, st  $b_{i_1} = b_{i_2}$ . Probleemseks kohaks on selliste indiviidide olemasolu. Sellisel tõlgendamisel peab lisaks ära märkima: eeldusel, et mudeli komponent  $b_i$  ja ülejäänud  $X$  tunnused on fikseeritud.

Korduvmõõtmistega logistilise mudeli interpreteerimiseks vaadatakse samuti šansside suhet.

Näiteks kui  $X_{ijm}$  saab mingi väärtuse  $m$  ja muud väärtused on fikseeritud, siis sündmuse toimumise šanss avaldub järgmiselt:

$$\ln \left( \frac{P(Y_{ij} = 1|b_i, X_{ij1}, \dots, X_{ijm} = m, \dots, X_{ijk})}{P(Y_{ij} = 0|b_i, X_{ij1}, \dots, X_{ijm} = m, \dots, X_{ijk})} \right) = b_i + \beta_0 + \beta_1 X_{ij1} + \dots + \beta_m m + \dots + \beta_k X_{ijk}.$$

Ühe argumentiga  $X_{ij}$  mudeli korral, kus  $X_{ij} = x$ , kehtib:

$$\ln \left( \frac{P(Y_{ij} = 1 | b_i, x)}{P(Y_{ij} = 0 | b_i, x)} \right) = b_i + \beta_0 + \beta x,$$

ja šanss  $O_x$  avaldub järgmiselt:

$$O_x = e^{b_i + \beta_0 + \beta x}$$

Edasi, kui  $X_{ij} = x + 1$ , siis

$$\ln \left( \frac{P(Y_{ij} = 1 | b_i, x + 1)}{P(Y_{ij} = 0 | b_i, x + 1)} \right) = b_i + \beta_0 + \beta(x + 1),$$

ja selle šanss  $O_{x+1}$  on:

$$O_{x+1} = e^{b_i + \beta_0 + \beta(x+1)}.$$

Seega šansside suhe on

$$OR = \frac{O_{x+1}}{O_x} = e^\beta.$$

(Fitzmaurice *et al.*, 2012)

Korduvmõõtmistega logistilise regressiooni mudelite hindamiseks on kasutatud andmetöötlusprogrammi R lisapaketti *lme4* (Bates *et al.*, 2015).



kasutatavate isikute tegelike elukohtade arv on suurim Tallinna linnas, 130 isikut, sellele järgneb Tartu linn, 35 isikut.

### **Lähima mobiilsideantenni asukohta andmestik**

Andmestikus on kõikide isikute kohta välja arvatud koduaadressile geograafiliselt lähima mobiilsideantenni asukohta (edaspidi koduantenni asukohta) kaugus kuude lõikes. Kodule geograafiliselt lähima mobiilsideantenni kaugus isikutel kuude lõikes ei muutu. Koduantenni asukohta on lähim võimalik asukohta, kuhu saab kodu ankurpunkti määrata.

MNO 1 puhul asub kõige kaugem koduantenn 5430 meetri kaugusel ja lähim 1,4 meetri kaugusel. Keskmiselt asuvad koduantennid 631 meetri kaugusel kodudest. MNO 2 puhul asub kõige kaugem koduantenn 6902 meetri kaugusel ja lähim 12 meetri kaugusel. Keskmiselt asuvad koduantennid 522 meetri kaugusel kodudest. Üldine koduantennide asukohtade kaugust kirjeldav statistika on Lisas 1 Tabelis 4.

Koduantenni asukohta kauguse teadmine võimaldab paremini hinnata isikule leitud ankurpunktide täpsust. Näiteks võib maapiirkonnas saadud isiku kodu ankurpunkt asuda tegelikust kodust 2 kilomeetri kaugusel, mis võib tunduda ebatäpse tulemusena. Kui võrrelda saadud ankurpunkti kaugust koduantenni asukohta kaugusega, mis samuti asub 2 kilomeetri kaugusel, siis võimaldab see mõista, et kodu ankurpunkt ei saagi lähemal asuda ning seega võib öelda, et ankurpunkti mudel on tuvastanud kodu asukohta nii täpselt kui antud andmete puhul võimalik on.

### **Mobiilsideantennide andmestik**

Mõlemal mobiilioperaatoril on oma mobiilsideantennide andmestik. Mobiilioperaatorid hoiustavad oma süsteemis mobiilsideantennide informatsiooni, sh asukohta informatsiooni. See on vajalik võrgu planeerimiseks ja hooldustöödeks. Lisaks kõnetoimingute ning andmesidega seonduvatele andmetele edastavad mobiilioperaatorid ka informatsiooni mobiilsideantennide paiknemise kohta - see on vajalik passiivse mobiilpositsioneerimise uuringu läbiviimiseks.

Käesolevas töös kasutatud andmestikus on kaheksa veergu. Analüüsi läbiviimiseks kasutatakse mobiilsideantennide kohta järgmisi tunnuseid:

- mobiilsideantenni ID;
- mobiilsideantenni geograafilised koordinaadid.

MNO 1 puhul on andmestikus 22 209 unikaalset mobiilsideantenni, MNO 2 puhul on neid vähem - 13 991. Samas on unikaalsete mastide arv, mille küljes mobiilsideantennid asuvad, mobiilioperaatorite lõikes sarnane: MNO 1 puhul 1249 unikaalset asukohta ja MNO 2 puhul 1280.

## Ankurpunktide mudeli andmestik

Positiumi ankurpunktide mudeli tulemuseks on andmestik, kus on 16 veergu. Iga määratud ankurpunkti kohta on andmestikus üks rida. Analüüsiks kasutatakse andmestikust järgmisi tunnuseid:

- kuu;
- unikaalne identifitseerimisnumber;
- mobiilsideantenni identifitseerimisnumber;
- toimingute arv;
- päevade arv;
- keskmine kõnede alustamise aeg;
- keskmine kõnede alustamise aja standardhälve;
- nädalavahetuse päevade hulk;
- ankurpunkti tüüp.

Käesolevas töös loetakse kodu ankurpunktideks ka multifunktsionaalsed ankurpunktid, st analüüsitakse punkte, mis on saanud sildid „h“ ja „m“.

## 4.1 Stsenaariumid

Analüüsimaiks Positiumi ankurpunktide määramise mudeli tundlikkust muudetakse mudeli parameetrites keskmist kõnede alustamise kellaja väärtust. Esimesena määratakse ankurpunktid algse Positiumi mudeliga, kus kõne alustamise kaja keskvaartuse lävendiks, mis määrab kodu ankurpunkte, on 17:00. See nimetatakse stsenaariumiks 0 (sts00). Järgmisena muudetakse keskmine kõnede alustamise aeg kella 19:00 - see on stsenaarium 1 (sts01). Stsenaariumi 2 (sts02) puhul on parameetri väärtus kell 21:00 ja stsenaariumi 3 (sts03) puhul on see 13:00.

Stsenaariumi 1 juures oli kellaja valiku puhul ajendiks see, et inimesed, kes töötavad n-ö 9-st 5-ni ei jõua kohe peale tööd koju (liiklusummikud, poes käimine jne). Valitud kellajaks võiksid nad aga kodus olla. Stsenaariumi 2 puhul oli kellaja valiku ajendiks Vanhoof (2018) uuring, milles tuli kõige paremaks kellajaks 21:00 (Vanhoof *et al.*, 2018a). Stsenaarium 3 loodi vastukaaluks õhtustele kellaegadele, et saaks üldistavalt tulemusi interpreteerida.

## 5 Tulemused

### 5.1 Kirjeldav analüüs

#### 5.1.1 Mobiilsidetoimingute ülevaade

Mõlema mobiilioperaatori peale kokku on keskmiselt tehtud 189 toimingut inimese kohta kuus ning keskmiselt on isikud teinud toiminguid 7 päeval kuus määratud ankurpunktides. (Lisa 1 Tabel 3 Read 5 ja 6)

MNO 1 puhul on nelja kuu peale tehtud toimingute arv 122 632. Keskmise tehtud toimingute arv vaadeldud nelja kuu jooksul inimese kohta kuus on 328. Keskmiselt on isikud teinud vaadeldava aja jooksul määratud ankurpunktides toiminguid 6 päeval kuus. (Lisa 1 Tabel 3 Read 1, 5 ja 6)

MNO 2 puhul on nelja kuu peale tehtud toimingute arv 60 971. Keskmise tehtud toimingute arv vaadeldud nelja kuu jooksul inimese kohta kuus on 102. Keskmiselt on isikud teinud vaadeldava aja jooksul määratud ankurpunktides toiminguid 7 päeval kuus. (Lisa 1 Tabel 3 Read 1, 5 ja 6)

Võrgus MNO 2 on toiminguid tehtud poole vähem, kui MNO 1 puhul. See tuleneb asjaolust, et MNO 2 salvestab isikutel ainult CDR andmeid, samal ajal kui MNO 1 salvestab nii CDR kui ka DDR andmeid.

Toimingute alustamise aeg on mobiilioperaatorite puhul keskmiselt 14:49. See näitab, et toiminguid tehakse nii hommikul kui ka õhtul. MNO 1 puhul on keskmine kõnede alustamise aeg varasem (14:19) kui MNO 2 (15:07). See jällegi sõltub asjaolust, et MNO 1 puhul on CDR ja DDR andmed. DDR andmete puhul toimuvad öisel ajal automaatsed andmeside uuendused, mis tõstavad keskmist varasemaks. (Lisa 1 Tabel 3 Rida 7)

#### 5.1.2 Määratud ankurpunktide ülevaade

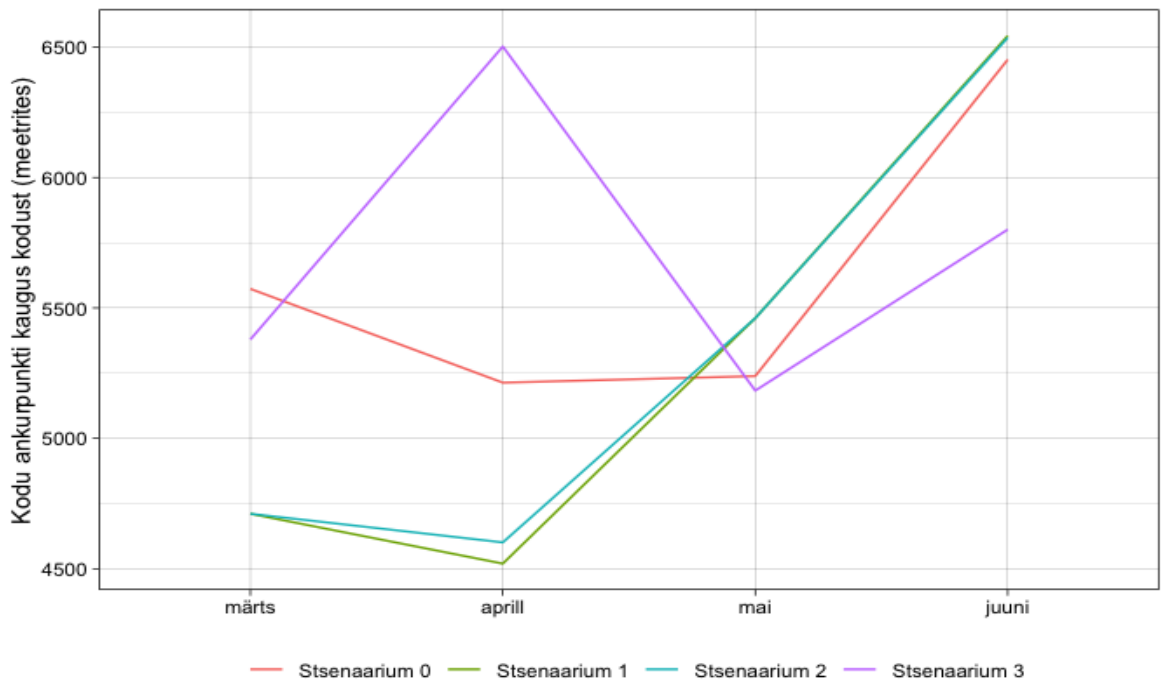
Esmalt tuuakse välja üldine statistika määratud ankurpunktide kohta. Nelja kuu peale on 256 isikule määratud 8251 ankurpunkti. See sisaldab nii kodu, tööaja ja multifunktsionaalseid ankurpunkte kui ka sekundaarseid ankurpunkte. (Lisa 2 Tabel 5 Rida 1)

Keskmiselt on MNO 1 võrku kasutavatele inimestele määratud rohkem ankurpunkte (49), kui MNO 2 võrku kasutavatele (21), kuigi MNO 2 puhul on valimis olevaid isikuid rohkem kui MNO 1 puhul. See tuleneb asjaolust, et MNO 1 puhul on andmete tihedus suurem (CDR ja DDR andmed) ja mitmekordselt külastatavaid asukohti võib tulla rohkem. Näiteks on ka

MNO 1 puhul maksimaalne määratud ankurpunktide arv isiku kohta kuus suurem (52), kui MNO 2 puhul - 27. (Lisa 2 Tabel 5 Read 2 ja 7)

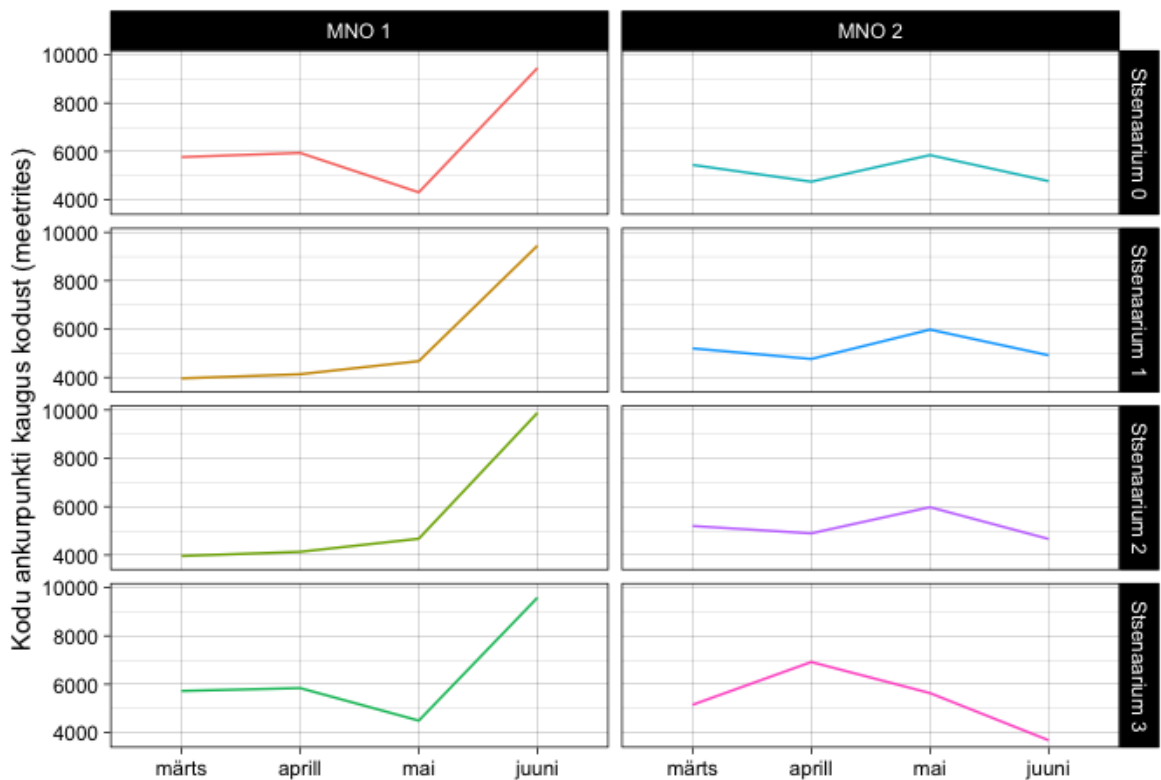
Ankurpunktiks määrati 3748 unikaalset mobiilsideantenni. MNO 1 puhul oli neid rohkem (2272) kui MNO 2 puhul (1476). Unikaalseid ankurpunkti asukohti oli MNO 1 puhul 719 ja MNO 2 puhul 633. Erinevuse põhjuseks võib ka siinjuhul olla andmete tihedus. Keskmiselt määrati isiku kohta kuus ankurpunktiks MNO 1 puhul 13 erinevat mobiilsideantenni ja MNO 2 puhul 5 erinevat mobiilsideantenni. Sama oli ka keskmine määratud ankurpunktide erinevate asukohtade arv. See ühtib ka keskmise määratud ankurpunktide arvuga, seega ühte mobiilsideantenni kasutati ankurpunkti määramisel vaid ühe korra. (Lisa 2 Tabel 5 Read 8, 9, 12, 13 ja 14)

Edaspidi analüüsitakse kodu ankurpunkte, sh multifunktsionaalseid ankurpunkte. Joonisel 2 on välja toodud määratud kodu ankurpunktide keskmine kaugus kodust stsenaariumite lõikes kuude kaupa. Stsenaariumi 1 ja 2 puhul väheneb määratud ankurpunkti täpsus (kaugus kodust suureneb) märgatavalt mai- ja juunikuus. Samuti on stsenaariumi 0 puhul määratud ankurpunkti kaugus kodust kõige suurem juunis. Määratud ankurpunktide täpsuse vähenemine mais ja juunis võib tingitud olla soojadest ilmadest: inimesed veedavad rohkem aega maakodudes ning seda just õhtusel ajal ja nädalavahetustel. See võib viidata sesoonsetele rütmidele (Silm *et al.*, 2010). Stsenaariumi 3 puhul on määratud ankurpunkti täpsus kõige väiksem aprillis. Terve ajaperioodi peale on määratud kodu ankurpunktide keskmine kaugus kodust stsenaariumi 0 puhul 5504 meetrit, stsenaariumi 1 puhul 5211 meetrit, stsenaariumi 2 puhul 5229 meetrit ja stsenaariumi 3 puhul 5606 meetrit.



**Joonis 2.** Määratud kodu ankurpunktide keskmine kaugus kodu asukohast kuude ja stsenaariumide lõikes

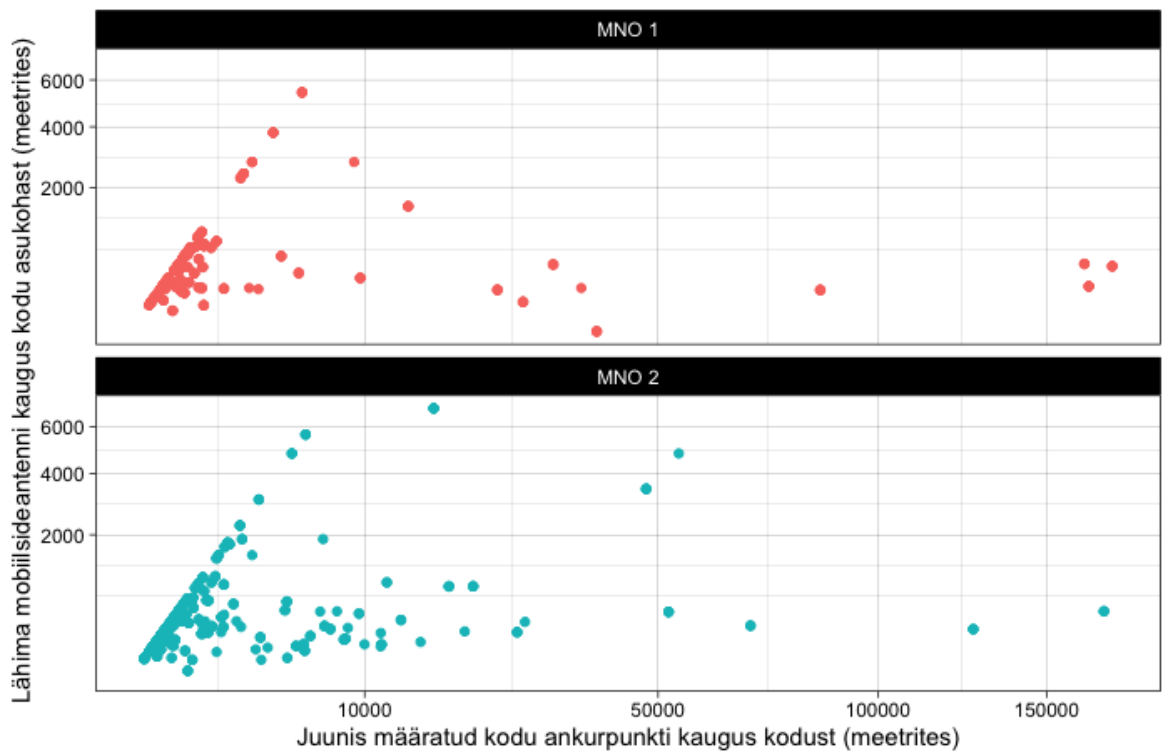
Järgmisena vaadeldakse määratud ankurpunktide keskmist kaugust kodust stsenaariumite ja mobiiloperaatorite lõikes kuude kaupa (Joonis 3). On näha, et MNO 1 puhul on stsenaariumid 0 ja 3 ning stsenaariumid 1 ja 2 omavahel sarnased. Kõikide stsenaariumite puhul asuvad juunikuus määratud ankurpunktid keskmiselt kodust kaugemal kui teistel kuudel. MNO 2 puhul on esimesed kolm stsenaariumit sarnased, stsenaarium 3 on teistest erinev. Võrreldes mobiiloperaatoreid saab öelda, et MNO 2 puhul on määratud ankurpunktide kaugused kodust kuude lõikes ühtlasemad. Samuti on MNO 2 puhul juunikuus määratud ankurpunktid tegelikule kodule lähemal kui MNO 1 puhul.



**Joonis 3.** Määratud kodu ankurpunktide keskmine kaugus kodu asukohast kuude, stsenaariumide ja mobiilioperaatorite lõikes

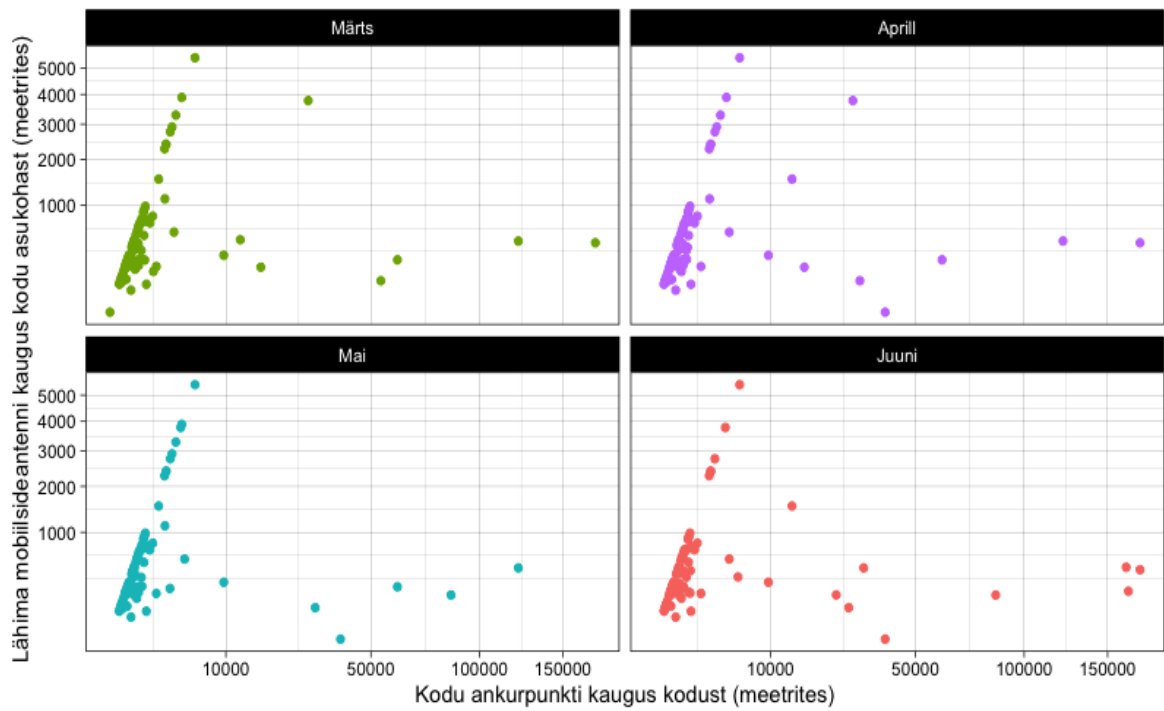
Joonisel 4 on välja toodud koduantenni asukoha kauguse tegelikust kodust ja juunikuus määratud ankurpunktide kauguse vaheline seos. Selle joonise põhjal saab öelda, et MNO 1 puhul on rohkem neid punkte, mis paiknevad tegelikult kodust väga kaugel. Selle tõttu väheneb ka juunikuiste ankurpunktide määramise täpsus MNO 1 puhul (Joonis 3). MNO 2 puhul on punktid küll nähtavalt hajusamalt, aga enamik punkte on määratud lähemal kui 10 kilomeetrit ning seetõttu on ka ankurpunktide täpsus suurem.

Enamik inimesi võtab puhkuse just suvekuudel (Järv, 2018) ja käesolevas töös on ainuke suvekuu juunikuu. Puhkuse ajal minnakse kodust kaugemale - näiteks maale ja suvilatesse ja selle tõttu väheneb kodu ankurpunktide määramise täpsus. MNO 1 puhul võib juunikuiste ankurpunktide määramise täpsust mõjutada asjaolu, et kodust eemal olles puudub paljudes kohtades WiFi kasutamise võimalus ning selle asemel kasutatakse mobiilset andmesidet. Samuti on juunikuus jaanipäev, mille puhul minnakse näiteks maakohtadesse ning seetõttu võib samuti suurenda andmeside kasutus.

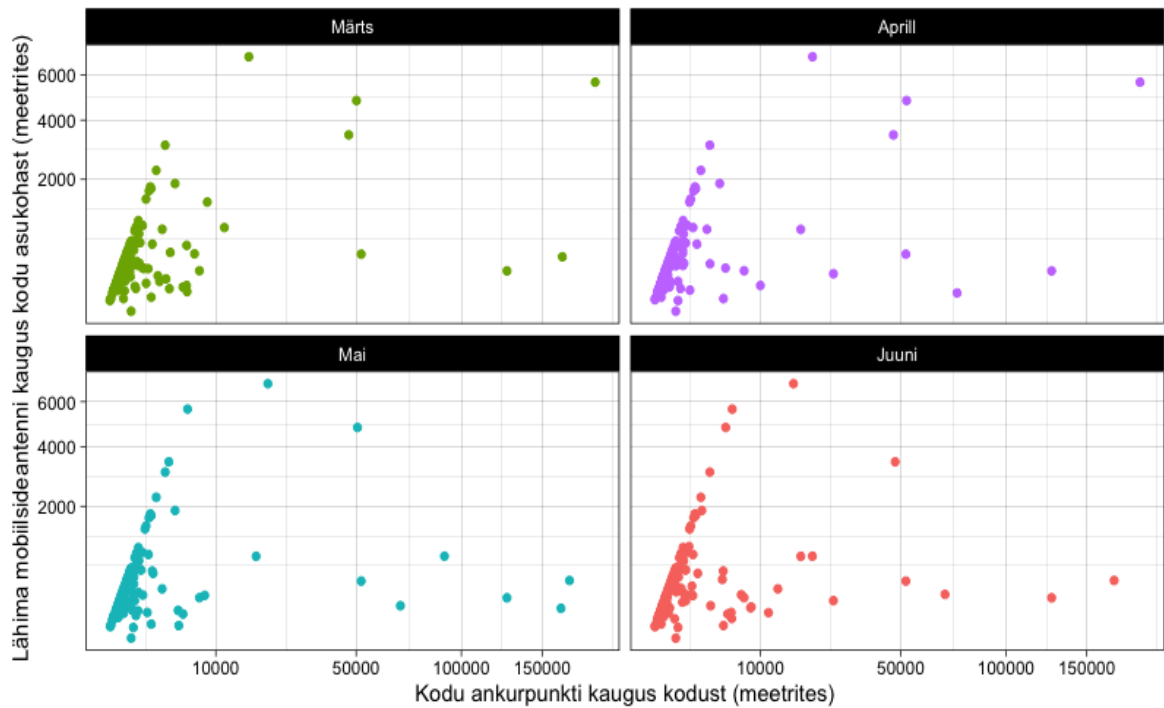


**Joonis 4.** Juunikuus määratud kodu ankurpunktide keskmise kauguse ja koduantenni asukoha kauguse hajuvusdiagramm mobiilioperaatorite lõikes

Joonistel 5 ja 6 on välja toodud lähima mobiilsideantenni kauguse tegelikust kodust ja ankurpunkti kauguse tegelikust kodust vaheline seos stsenaariumi 0 puhul. Siinkohal vaadeldakse ainult stsenaariumit 0, et paremini mõista Positiumi ankurpunktide mudeli tööd ja mudeli määratud kodu ankurpunktide jagunemist. Mõlema mobiilioperaatori puhul paiknevad enamus punktid selgelt nähtavas tõusvas joones, mis näitab, et ühe tunnuse suurenedes suureneb ka teine. Samuti on mõlema mobiilioperaatori puhul näha ka selgelt eristuvaid punkte, erisus väljendub järgmiselt: koduantenn asub kodule lähedal, aga tuvastatud ankurpunkt asub kodust ekstreemselt kaugel.



**Joonis 5.** Koduantenni asukoha kauguse ja kodu ankurpunkti kauguse hajuvusdiagramm MNO 1 puhul (sts00)



**Joonis 6.** Koduantenni asukoha kauguse ja kodu ankurpunkti kauguse hajuvusdiagramm MNO 2 puhul (sts00)

Määratud ankurpunktide hulgas esineb selliseid punkte, mis asuvad kodust ekstreemselt kaugel - kaugemal kui 20 kilomeetrit. Seda on näha ka joonistelt 5 ja 6.

Järgnevalt on uuritud eraldi ja täpsemalt neid ankurpunkte, mis on määratud kodust kaugemalt kui 20 kilomeetrit. Tabelis 1 on välja toodud järgnevate tunnuste arvulised näitajad mõlema mobiilsideoperaatori puhul kõigi stsenaariumite korral:

1. Isikute arv, kellel on vähemalt üks kodu ankurpunkt kuus kaugemal kui 20 kilomeetrit.
2. Isikute arv, kelle kõik ankurpunktid on määratud kaugemal kui 20 kilomeetrit.
3. Isikute kodu ankurpunktide keskmiste kauguste keskmine (meetrites), kui vähemalt üks ankurpunkt asub kaugemal kui 20 kilomeetrit.
4. Isikute kodu ankurpunktide kauguste standardhälvete keskmine (meetrites), kui vähemalt üks ankurpunkt asub kaugemal kui 20 kilomeetrit.

**Tabel 1.** Kaugemal kui 20 kilomeetrit asuvad kodu ankurpunktid

| Tunnus | Stsenaarium 0 |        | Stsenaarium 1 |        | Stsenaarium 2 |        | Stsenaarium 3 |        |
|--------|---------------|--------|---------------|--------|---------------|--------|---------------|--------|
|        | MNO 1         | MNO 2  | MNO 1         | MNO 2  | MNO 1         | MNO 2  | MNO 1         | MNO 2  |
| 1.     | 10            | 14     | 11            | 14     | 11            | 16     | 10            | 13     |
| 2.     | 3             | 2      | 3             | 2      | 3             | 2      | 3             | 5      |
| 3.     | 50 818        | 45 502 | 39 259        | 44 739 | 40 079        | 38 772 | 50 974        | 43471  |
| 4.     | 40 603        | 41 098 | 38 550        | 41 390 | 38 804        | 41 396 | 40 290        | 38 267 |

Võrreldes mobiilioperaatoreid, siis on MNO 1 puhul osakaaluliselt rohkem neid, kellel asub vähemalt üks määratud kodu ankurpunkt kaugemal kui 20 kilomeetrit (MNO 1 puhul 0,1 – 0,11 ja MNO 2 puhul 0,08 – 0,1). See võib tuleneda sellest, et MNO 1 puhul on toimingute hulk suurem ja toiminguid võib olla tehtud rohkem kohtades, mis tegelikult pole kodud, aga parameetrite poolest sobituvad kodu ankurpunktideks. Stsenaariumi 3 korral MNO 2 puhul on 5 isikut, kellel kõik kodu ankurpunktid määratud ekstreemselt kaugemale (Tabel 1), mis on võrreldes teiste stsenaariumitega peaaegu poole rohkem. Selle võib põhjustada Positiumi mudelis määratud keskmine kõnede alustusaja varane kellaaeg, mille tõttu võivad seguneda kodu ja tööaja tunnustega ankurpunktid. Tööaja ankurpunkt võib saada kodu ankurpunkti sildi ja sellepärast võivad määratud kodu ankurpunktide kaugused olla suuremad.

Lisaks sellele, et kõik määratud kodu ankurpunktid asuvad kaugemal kui 20 kilomeetrit võib ankurpunktide puhul olla veel variante:

- Kolmel kuul asuvad määratud kodu ankurpunktid kaugemal kui 20 kilomeetrit ja üks kodu ankurpunkt asub sama kaugel kui koduantenn või enam-vähem samal kaugusel.

- Kahel kuul asuvad määratud kodu ankurpunktid kaugemal kui 20 kilomeetrit ja kaks kodu ankurpunkti asuvad sama kaugel kui koduantenn või enamvähem samal kaugusel.
- Ainult üks määratud kodu ankurpunkt asub kaugemal kui 20 kilomeetrit.

MNO 1 puhul stsenaariumi 0 korral on nendel isikutel, kellel ainult üks kuine määratud ankurpunkt kaugemal kui 20 kilomeetrit, kõigil selleks kuuks juuni. See võib viidata asjaolule, et juunikuus veedeti vähem aega kodus ning kodu ankurpunkti tunnused sai mõni teine tihedasti külastatav asukoht.

Keskmete kauguste erinevused on märgatavad kui võrrelda kõiki määratud ankurpunkte ekstreemselt kaugel asuvate ankurpunktidega. Keskmiselt asuvad MNO 1 puhul määratud kodu ankurpunktid kodust 6 kilomeetri kaugusel ja MNO 2 puhul 5,1 kilomeetri kaugusel. Ekstreemselt kaugel olevad kodu ankurpunktid asuvad aga 40 kuni 50 kilomeetri kaugusel kodust (Tabel 1).

Võrreldes määratud ankurpunkte stsenaariumite kaupa, saab välja tuua nii erinevusi kui ka sarnasusi. Kõige rohkem kodu ankurpunkte (891) tuvastati stsenaariumi 3 puhul, samuti oli selles stsenaariumis kõige vähem neid inimesi, kellele oli määratud kaks kodu ankurpunkti ühes kuus (47). Võrreldes stsenaariume 0, 1 ja 2, siis tuvastatud kodu ankurpunktide arv väheneb seoses kellaaaja hilisemaks muutumisega, vastavalt 877, 874 ja 867. Samuti tuvastati stsenaariumi 3 puhul kõige rohkem multifunktsionaalseid ankurpunkte (653), kusjuures teistel stsenaariumitel oli neid tuvastatud peaaegu poole vähem. Kuna stsenaariumi 3 kellaaeg on päevane - 13:00, siis võivad tööaja ja kodu tunnustega ankurpunktid seguneda, selle tõttu on ka multifunktsionaalseid ankurpunkte rohkem. (Lisa 2 Tabelid 6 ja 7)

Kodu ankurpunktide määramise juures on oluliseks parameetrik keskmine kõnede alustamise aeg, kuna esmajärjekorras määratakse selle põhjal ankurpunktile silt „h“ (kodu ankurpunkti silt) (Positiumi mudeli samm 5 alasamm 2). Vastav parameeter on MNO 1 puhul varasem (14:19) kui MNO 2-1 (15:07), millest võib järeldada, et MNO 1 kasutajad on teinud rohkem toiminguid varasemal ajal (Lisa 1 Tabel 3 Rida 7). Stsenaariumite 0, 1, 2 puhul on keskmine kõnede alustamise aja keskmine väärtus enne kellaaega, mis on mudelis määratud - see jääb vahemikku 13:00-16:00 (Lisa 3 Tabelid 6 ja 7 Read 9-10). Välja toodud stsenaariumite puhul peaks see olema vastavalt pärast kella 17:00, 19:00 või 21:00. Keskmise kõnede alustamise aja varasuse põhjuseks võib olla asjaolu, et kodu silt „h“ määratakse ka esimesel kahel kohal olevatele ankurpunktidele, kui vastava kellaaaja põhjal ja muude parameetrite põhjal (Positiumi mudeli samm 5 alasamm 3) ei suudetud neid märgistada sildiga „h“ või „w“ (Positiumi mudeli samm 5 alasamm 4) ehk tegelikult on kellaaeg nõrk klassifitseerija kodu ankurpunktide puhul. Stsenaariumi 3 puhul on mudelis määratud parameetri väärtus kell 13:00, analüüsist tuleb välja, et tuvastatud kodu ankurpunktide keskmine toimingute alustamise aeg on samuti peale 13:00 (Lisa 2 Tabel 7 Read 9-10).

Seda, et kellaaeg on nõrk klassifitseerija kodu ankurpunktide puhul, võib järeldada ka asjaolust, et stsenaariumite 0, 1, 2 puhul on kellaaaja faktori põhjal määratud kodu ankurpunktide osakaal väike. Kõige suurem on see MNO 2 stsenaarium 0 puhul, kus vastavalt keskmisele kõnede alustusajale on koduks määratud ankurpunktide osakaal 0,33. Stsenaariumi 0 MNO 1 puhul ja stsenaariumite 1, 2 puhul on see osakaal alla 0,1. Stsenaariumi 3 puhul on aga vastavalt kellaaajale määratud kodu ankurpunktide osakaal suur - MNO 1 puhul 0,67 ja MNO 2 puhul 0,97. (Lisa 2 Tabelid 6 ja 7 Rida 15)

Kellaaajast tulenevalt määratud kodu ankurpunktide madala osakaalu põhjal võib järeldada, et keskmine kõnede alustusaja tunnus pole kodu ankurpunktide määramise juures väga oluline. Sarnasele järeldusele jõudis ka Vanhoof oma artiklis (Vanhoof *et al.*, 2018a). Suurem roll on päevade arvul (Vanhoof *et al.*, 2018a) - kodu on tavaliselt üks tihedamini külastatavaid tegevuskohti. Seetõttu võib kodu ankurpunkti vaid eelnevalt nimetatud tunnuste põhjal õigesti määrata. Probleem tekib inimeste puhul, kellel on mitu kodu ja ka näiteks mitu töökohta ning seetõttu on toimingute hulk ja päevade arv ühtlasemalt jaotunud nende tegevuskohtade vahel ning n-ö top 1 ankurpunkti on raskem eristada.

Tööaja ankurpunktide määramisel on oluline parameeter „Kõnede alustusaja standardhälve“. Kui see on väiksem kui või võrdne 0,175, siis saab ankurpunkt tööaja sildi „w“ (Positiumi mudeli samm 5 alasamm 3). Võrreldes kõnede alustusaja standardhälvete keskmisi mobiilioperaatorite vahel, siis võib öelda, et MNO 2 puhul on multifunktsionaalsed ankurpunktid eelmainitud näitaja poolest iseloomulikud tööajale - parameetri väärtus on 0,16. Samas kui MNO 1 puhul on kodu ja multifunktsionaalsed ankurpunktid kõnede alustusaja standardhälbe poolest iseloomulikud kodu ankurpunktidele - nimetatud tunnus on kõikidel juhtudel suurem kui 0,175. (Lisa 2 Tabelid 6 ja 7 Read 12-14).

Järgneva analüüsi jaoks jaotatakse isikud koduantenni asukoha kauguse alusel nelja gruppi: lähedal, keskmisel kaugusel, kaugel ja väga kaugel (Tabel 2). Klasside piirid on valitud koduantenni asukoha kauguse kvartiilide alusel (Lisa 1 Tabel 4). Eeldame, et koduantenni asukoha kauguse põhjal on võimalik eristada isikuid, kes elavad tihedamini asustatud piirkondades ja neid, kes elavad vähem asustatud piirkondades. Gruppi jaotamine annab võimaluse analüüsida määratud ankurpunktide kaugusi erinevate kauguste gruppide kaupa.

**Tabel 2.** Koduantenni asukoha kvartiilide põhjal grupeeritud isikute jaotus (sts00)

|                           | Lähedal | Keskmisel kaugusel | Kaugel    | Väga kaugel |
|---------------------------|---------|--------------------|-----------|-------------|
| <b>Kaugus (meetrites)</b> | 0 – 175 | 175 – 280          | 280 – 534 | 534 – ...   |
| <b>Sagedus</b>            | 64      | 65                 | 63        | 64          |

Stsenaariumi 0 puhul hinnati erinevaid tunnuseid kauguse gruppides (Lisa 2 Tabel 8). Gruppide lõikes on näha väikest erinevust osakaaludes, mille puhul vähemalt üks määratud kodu ankurpunkt langeb kokku koduantenni asukohaga (kauguste mõistes) - gruppides „Keskmisel kaugusel“ ja „Väga kaugel“ on need madalaimad (0,59) ja grupis „Lähedal“ on vastav osakaal suurim (0,67).

Määratud kodu ankurpunktide keskmised kaugused on gruppide lõikes väga erinevad. Välja tuleb vastuolu, et grupis „Väga kaugel“ on keskmine määratud ankurpunktide kaugus kodust väiksem (3050 meetrit) kui grupis „Lähedal“ (4596 meetrit). Selle tingib asjaolu, et grupis „Lähedal“ asuvad n-ö valesi määratud kodu ankurpunktid kaugemal kui grupis „Väga kaugel“. Põhjuseks võib olla inimeste ruumilise käitumise juhuslikkus: puhkustel käimine, mitu töökohta, mitu kodu jne, mis teeb raskemaks ankurpunktide õige määramise. Seega olenemata sellest, kas inimene elab maal või linnas, st lähim mobiilsideantenni asukoht on vastavalt kaugel või lähedal, esineb olukordi, kus kodu ankurpunkt määratakse valesi. Samuti on oluline ära märkida, et koduantenni puhul räägitakse võimalikust parimast asukohast, kuhu saab kodu ankurpunkti määrata, mitte kohast, kuhu peaks ankurpunkti määrama. (Lisa 2 Tabel 8 Rida 1)

Võrreldes keskmisi ankurpunktide kaugusi kodust, tingimusel, et määratud ankurpunktide keskmine kaugus kodust ja koduantenni asukoha kaugus ei ole samad, on näha, et teises grupis („Keskmisel kaugusel“) on vastav keskmine kaugus suurim (9770 meetrit). Samuti on samas grupis ka üldine ankurpunktide keskmine kaugus suurim (8157 meetrit). Seega saab järeldada, et selles grupis on kõige vähem neid ankurpunkte, mille kaugus kodust ühtib koduantenniga. Seda on näha ka tabelis 8 - koduantenniga sama kaugel asuvate ankurpunktide osakaal on kõige väiksem teises grupis (0,48), ülejäänud gruppides on see 0,55 või 0,56. (Lisa 2 Tabel 8 Read 4 ja 9)

## 5.2 Logistilise regressiooni mudeli rakendamine

Logistilise regressiooni mudel hinnatakse erinevate stsenaariumite puhul binaarsele tunnusele  $Y = 0/1$ , mis näitab, kas määratud ankurpunkt ja koduantenni asukoht langevad kokku vähemalt üks kord nelja kuu jooksul (1) või ei (0) (kauguse mõistes).

Kirjeldavad tunnused on koduantenni asukoha kaugus ja mobiilioperaator - tasemetega MNO 1 ja MNO 2. Tunnuse „Koduantenni kaugus“ puhul vaatame läbi kaks varianti: kaugus pideva tunnuseks ja grupeerituna kvartiilide põhjal (nagu kasutati eelnevas kirjeldavas analüüsis).

Samuti vaadatakse koosmõju tunnuste koduantenni kaugus (meetrites) ja mobiilsideoperaator vahel, et teada saada, kas andmete tihedus võiks olenevalt koduantenni asukoha kaugusest erinevalt mõjuda. Mobiilioperaatoritel on erinev andmetihedus, kuna MNO 1 puhul on and-

mestikus nii CDR kui ka DDR andmed, samal ajal kui MNO 2 puhul on ainult CDR andmed.

Hinnatakse 8 logistilise regressiooni mudelit:

- Logistilise regressiooni mudel stsenaariumite 0, 1, 2 ja 3 korral, kus koduantenni kaugus on pideva tunnusena meetrites.
- Logistilise regressiooni mudel stsenaariumite 0, 1, 2 ja 3 korral, kus koduantenni kaugus on grupeeritud kvartiilide põhjal.

Vaadeldud argumenttunnused ei osutunud statistiliselt oluliseks mitte ühegi mudeli puhul, olulisuse nivoo  $\alpha = 0,05$  korral (Lisa 3). Seega ei saa väita, et antud andmetiheduse piirides mõjutab erinev andmetihedus Positiumi ankurpunktide mudeli tööd. Antud andmete põhjal selgus, et andmetiheduselt 1005 toimingut päevas (MNO 1) andmetihedusele 500 toimingut päevas (MNO 2) liikudes tulemustes mõju näha ei olnud. Andmetiheduse hinnanguna kasutatakse keskmist toimingute hulka päevas (Lisa 1 Tabel 3 Rida 3). Sellest saab järeldada, et antud andmete puhul andmetihedusel ankurpunktide mudelile mõju ei olnud, st antud andmete puhul ei määra mudel suurema toimingute hulga puhul ankurpunkte paremini/halvemini kui väiksema toimingute hulga puhul.

Samuti ei mõjuta Positiumi mudeli tööd lähima mobiilsideantenni asukoha kaugus, st mudel ei määra erinevalt kodu ankurpunkte neil, kellel on koduantenni asukoht lähemal võrreldes nendega, kellel asub koduantenn kaugemal.

### 5.3 Korduvmõõtmistega logistilise regressiooni mudeli rakendamise

Korduvmõõtmistega logistilise regressiooni mudel hinnatakse kahele erinevale binaarsele tunnusele.

Esimesel juhul hinnatakse see tunnusele, mis näitab, et vastaval kuul määratud ankurpunkt asub samal kaugusel kodust kui koduantenn. Mudel hinnatakse kõigi stsenaariumite puhul. Mudelis on kirjeldavateks tunnusteks koduantenni kaugus grupeerituna kvartiilide põhjal, mobiilioperaator - tasemetega MNO 1 ja MNO 2 ning kuu nimetus - tasemetega märts, aprill, mai ja juuni.

Mudel hinnatakse iga stsenaariumi jaoks eraldi. Kõigi nelja mudeli puhul (Lisa 4 Mudel 4.1., Mudel 5.1., Mudel 6.1., Mudel 7.1.) on referentstasemeteks mobiilioperaator MNO 1, koduantenni asukoha kauguse grupp „Lähedal“ ja märtsikuu.

Kõikide eelmainitud mudelite puhul osutusid statistiliselt oluliselt erinevaks märtsikuu ja mai-

kuu, olulisuse nivoo  $\alpha = 0,05$  korral (Lisa 4 Tabelid 17, 18, 19, 20). Eeldusel, et mobiilioperaator ja koduantenni asukoha kauguse grupp jäävad hinnangute puhul samaks, siis saab mudeli põhjal öelda, et määratud ankurpunkti kauguse ja koduantenni asukoha kauguse kokkulangemise šanss vastavas kuus erineb:

1. stsenaariumi 0 puhul  $e^{0,7504} = 2,21$  korda, kui võrrelda ankurpunkte, mis on määratud märtsikuus nende ankurpunktidega, mis on määratud maikuus;
2. stsenaariumi 1 puhul  $e^{0,757} = 2,13$  korda, kui võrrelda ankurpunkte, mis on määratud märtsikuus nende ankurpunktidega, mis on määratud maikuus;
3. stsenaariumi 2 puhul  $e^{0,761} = 2,14$  korda, kui võrrelda ankurpunkte, mis on määratud märtsikuus nende ankurpunktidega, mis on määratud maikuus;
4. stsenaariumi 3 puhul  $e^{0,662} = 1,17$  korda, kui võrrelda ankurpunkte, mis on määratud märtsikuus nende ankurpunktidega, mis on määratud maikuus.

Šanss, et määratud ankurpunkt ja koduantenn asuvad kodust samal kaugusel on stsenaariumite 0, 1 ja 2 puhul kaks korda kõrgem mais võrreldes märtsiga. Kõige väiksem šanss kauguste kokkulangemise osas on stsenaariumi 3 korral.

Vaadates joonist 2, siis ei tundu, et erinevus tuleb märtsi ja mai vahel nagu mudeli põhjal selgub. Põhjuseks on see, et joonisel on vaadatud määratud ankurpunkti keskmisi kaugusi meetrites, mille järgi erineb juunikuu teistest kuudest enim. Logistiline regressiooni mudel vaatab aga osakaale. Osakaal, et määratud ankurpunkt ja koduantenn asuvad märtsikuus sama kaugel on 0,488, maikuus on see osakaal 0,543. Märtsikuul aprilli (0,504) ja juuniga (0,480) nii suurt osakaalude erinevust ei tule.

Teisel juhul hinnatakse korduvmõõtmistega logistilise regressiooni mudel tunnusele, mis näitab, et vähemalt üks korda nelja kuu jooksul asub määratud ankurpunkt sama kaugel kui koduantenn. Kirjeldavad tunnused on koduantenni asukoha kaugus grupeerituna kvartiilide põhjal, mobiilioperaator - tasemetega MNO 1 ja MNO 2 ja stsenaarium - tasemetega sts00, sts01, sts02 ja sts03.

Mudeli 8.1. puhul on referentstasemeteks mobiilioperaator MNO 1, koduantenni asukoha kauguse grupp „Lähedal“ ja stsenaarium 0.

Mudelis osutusid statistiliselt oluliselt erinevaks stsenaariumid 0 ja 3, olulisuse nivoo  $\alpha = 0,05$  korral (Lisa 4 Tabel 21). Mudeli põhjal saab öelda, et määratud ankurpunkti kauguse ja koduantenni asukoha kauguse kokkulangemise šanss vähemalt üks kord 4 kuu jooksul erineb  $e^{-4,0602} = 0,23$  korda, kui ankurpunktid on määratud stsenaariumi 3 puhul võrreldes stsenaariumiga 0 ja eeldusel, et mobiilioperaator ja koduantenni asukoha kauguse grupp on hinnangul samad.

## Kokkuvõte

Bakalaureuse töö eesmärk oli uurida Positiumi ankurpunktide mudeli tundlikkust kodu ankurpunkti tuvastamise näitel. Tundlikkuse uurimiseks muudeti Positiumi mudelis ühe parameetri väärtust. Muudetavaks parameetriks oli keskmine kõnede alustamise aeg. Samuti sooviti teada, kui tundlik on ankurpunktide mudel andmetiheduse suhtes.

Selleks kirjeldati esmalt passiivsetel mobiilpositsioneerimise andmetel töötavat Positiumi ankurpunktide mudelit, et mõista, missuguste põhimõtete alusel määratakse kodu ja multifunktsionaalsed ankurpunktid. Samuti anti töös ülevaade mobiilpositsioneerimise ja ankurpunktidega seonduvast teoreetilisest taustast.

Töös toodi näide alusandmete kohta ja anti ülevaade analüüsi jaoks kasutatavatest andmetest. Analüüsiti alusandmeid, mille puhul toodi välja isiku tehtud toimingutega seonduvad näitajad. Määratud kodu ankurpunktide peal viidi läbi põhjalik analüüs: vaadati määratud kodu ankurpunkte erinevate stsenaariumite, mobiilioperaatorite ja kuude lõikes. Peamise tunnusega vaadati määratud kodu ankurpunkti kaugust kodust. Selle näitaja abil analüüsiti ekstreemselt kaugel (kaugemal kui 20 km) asuvaid kodu ankurpunkte ning stsenaariumi 0 puhul analüüsiti neid ka kvartiilide järgi jaotatud gruppides. Samuti võrreldi vastavat näitajat geograafiliselt lähima mobiilsideantenni asukoha kaugusega kodust.

Positiumi ankurpunktide mudeli tundlikkuse hindamiseks viidi läbi logistiline regressioon. Logistilise regressiooni mudel hinnati binaarsele tunnusele, mis näitas määratud kodu ankurpunkti kauguse ja koduantenni kauguse kokkulangemist. Tulemuseks saadi, et Positiumi ankurpunktide mudeli tööd ei mõjuta lähima mobiilsideantenni kaugus ega ka mobiilioperaatorite erinev andmetihedus (MNO 1 puhul CDR ja DDR andmed, MNO 2 puhul ainult CDR andmed).

Lisaks hinnati korduvmõõtmistega logistilise regressiooni mudel samale binaarsele tunnusele. Mudel hinnati erinevate stsenaariumite puhul, et testida kodu ankurpunktide määramise erinevust kuude lõikes. Tulemusena saab väita, et kasutatud andmete põhjal määras Positiumi mudel kodu ankurpunkte erinevalt märtsis ja maikuus - šanss, et määratud ankurpunkt ja koduantenn asuvad kodust samal kaugusel oli stsenaariumite 0, 1 ja 2 puhul kaks korda kõrgem mais võrreldes märtsiga. Samuti hinnati stsenaariumite vahelist erinevust, et mõista keskmise kõnede alustamise aja mõju kodu ankurpunktide määramisele. Mudeli põhjal saab väita, et kodu ankurpunktide määramine, st määratud kodu ankurpunkti kauguse ja koduantenni kauguse kokkulangemine on erinev stsenaariumite 0 ja 3 vahel - šanss, et kodu ankurpunkti kaugus ja koduantenni kaugus langevad kokku on stsenaariumi 0 korral peaaegu 4 korda kõrgem kui stsenaariumi 3 korral.

Analüüside tulemusena saab järeldada, et Positiumi ankurpunktide mudel ei ole tundlik kellaaja kui sisendparameetri muutumise suhtes. Võib väita, et keskmine kõnede alustamise aeg on Positiumi mudeli puhul nõrk klassifitseerija. Samas saab öelda, et päevane kellaeg (13:00) parameetri väärtusena ei sobi: stsenaariumi 3 puhul oli kodu ankurpunkti kauguse ja koduantenni kauguse kokkulangemise šanss erinev võrreldes teiste stsenaariumitega ja kodu ankurpunktid määrati kodust kõige kaugemale. Ankurpunktide mudeli puhul saab kodu ankurpunktide määramisel kõige olulisemaks parameetriks lugeda päevade arvu, millal on toiminguid tehtud, sealjuures ka toimingute ühtlast jaotumist päevade vältel (Vanhoof *et al.*, 2018a). Logistilise regressiooni tulemusena saab öelda, et Positiumi mudel ei ole andmetiheduse suhtes tundlik. Samas saab kahe mobiilioperaatori puhul esinevaid erinevusi selgitada andmetiheduse erinevuse tõttu.

Edaspidistes analüüsides võiks hinnata ka teiste Positiumi ankurpunkti mudeli parameetrite mõju ankurpunktide määramisele. Samuti vaadata erinevate parameetrite koosmõju. Käesolevas töös sesoonsust ei uuritud, kuna andmed ei võimaldanud seda põhjalikult teha. Küll aga tuli välja kohti, kus esines sesoonset mõju andmetele. Seega järgnevate analüüside puhul on soovitatav kasutada andmeid, mis hõlmavad pikemat perioodi.

## Kirjandus

- Ahas, R., Aasa, A., Roosa, A., Mark, Ü., Silm, S. (2008). „Evaluating passive mobile positioning data for tourism surveys: An Estonian case study“. *Tourism Management* 29, lk. 469–486. DOI: [10.1016/j.tourman.2007.05.014](https://doi.org/10.1016/j.tourman.2007.05.014).
- Ahas, R., Aasa, A., Silm, S., Aunap, R., Kalle, H., Mark, Ü. (2007). „Mobile Positioning in Space-Time Behaviour Studies: Social Positioning Method Experiments in Estonia“. *Cartography and Geographic Information Science* 34.4, lk. 259–273. DOI: [10.1559/152304007782382918](https://doi.org/10.1559/152304007782382918).
- Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., Zook, M. (2015). „Everyday space-time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn“. *International Journal of Geographical Information Science* 29.11, lk. 2017–2039. DOI: [10.1080/13658816.2015.1063151](https://doi.org/10.1080/13658816.2015.1063151).
- Ahas, R., Silm, S., Järv, O., Saluveer, E., Tiru, M. (2010). „Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones“. *Journal of Urban Technology* 17.1, lk. 3–27. DOI: [10.1080/10630731003597306](https://doi.org/10.1080/10630731003597306).
- Axhausen, K., Zimmermann, A., Schönfelder, S., Rindsfuser, G., Haupt, T. (2002). „Observing the rhythms of daily life: A six-week travel diary“. *Transportation* 29, lk. 95–124. DOI: [10.1023/A:1014247822322](https://doi.org/10.1023/A:1014247822322).
- Bates, Douglas, Mächler, Martin, Bolker, Ben, Walker, Steve (2015). „Fitting Linear Mixed-Effects Models Using lme4“. *Journal of Statistical Software* 67.1, lk. 1–48. DOI: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Crane, R., Crepeau, R. (1998). „Does neighborhood design influence travel?: A behavioral analysis of travel diary and GIS data“. *Transportation Research Part D: Transport and Environment* 3.4, lk. 225–238. ISSN: 1361-9209. DOI: [https://doi.org/10.1016/S1361-9209\(98\)00001-7](https://doi.org/10.1016/S1361-9209(98)00001-7).
- Farrow, A., Taylor, H., Golding, J. (1997). „Time Spent in the Home by Different Family Members“. *Environmental Technology* 18.6, lk. 605–613. DOI: [10.1080/09593331808616578](https://doi.org/10.1080/09593331808616578).
- Fitzmaurice, G.M., Laird, N.M., Ware, J.H. (2012). *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics. Wiley, lk. 341–408.
- Griffin, A., Bell, S. (2009). „Applications of Signal Detection Theory to Geographic Information Science“. *Cartographica* 44, lk. 145–158. DOI: [10.3138/carto.44.3.145](https://doi.org/10.3138/carto.44.3.145).

- Hosmer, D.W., Lemeshow, S., Sturdivant, R.X. (2013). *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley, lk. 1–56.
- Hu, T., Luo, J., Kautz, H., Sadilek, A. (2015). „Home Location Inference from Sparse and Noisy Data: Models and Applications“. Teoses: *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, lk. 1382–1387.
- Järv, O., Ahas, R., Witlox, F. (2013). „Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records“. *Transportation Research Part C: Emerging Technologies* 38, lk. 122–135. DOI: <https://doi.org/10.1016/j.trc.2013.11.003>.
- Järv, P.-R. (juuli 2018). *Eestimaalaste peamine reisisiht oli mullu Läti*. Vaadatud: 05.05.2020. URL: <https://blog.stat.ee/2018/07/26/eestimaalaste-peamine-reisisiht-oli-mullu-lati/>.
- Krumm, J. (2007). „Inference Attacks on Location Tracks“. Teoses: *Pervasive Computing*. Toim. Anthony LaMarca, Marc Langheinrich, Khai N. Truong. Berlin, Heidelberg: Springer Berlin Heidelberg, lk. 127–143. ISBN: 978-3-540-72037-9.
- Mooses, V., Silm, S., Ahas, R. (2016). „Ethnic segregation during public and national holidays: a study using mobile phone data“. *Geografiska Annaler: Series B, Human Geography* 98.3, lk. 205–219. DOI: [10.1111/geob.12100](https://doi.org/10.1111/geob.12100).
- Paršova, D. (2019). „Mobility patterns in university campuses: an example of the University of Tartu“. Magistritöö. Tartu Ülikool.
- Pontes, T., Magno, G., Vasconcelos, M., Gupta, A., Almeida, J., Kumaraguru, P., Almeida, V. (2012). „Beware of What You Share: Inferring Home Location in Social Networks“. Teoses: lk. 571–578. ISBN: 978-1-4673-5164-5. DOI: [10.1109/ICDMW.2012.106](https://doi.org/10.1109/ICDMW.2012.106).
- Poom, A., Orru, K., Ahas, R. (2017). „The carbon footprint of business travel in the knowledge-intensive service sector“. *Transportation Research Part D: Transport and Environment* 50, lk. 292–304. DOI: <https://doi.org/10.1016/j.trd.2016.11.014>.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Saluveer, E., Raun, J., Tiru, M., Altin, L., Kroon, J., Snitsarenko, T., Aasa, A., Silm, S. (2020). „Methodological framework for producing national tourism statistics from

- mobile positioning data“. *Annals of Tourism Research* 81, lk. 102895. DOI: <https://doi.org/10.1016/j.annals.2020.102895>.
- Silm, S., Ahas, R. (2010). „The Seasonal Variability of Population in Estonian Municipalities“. *Environment and Planning A: Economy and Space* 42.10, lk. 2527–2546. DOI: [10.1068/a43139](https://doi.org/10.1068/a43139).
- Thierry, B., Chaix, B., Kestens, Y. (2013). „Detecting activity locations from raw GPS data: A novel kernel-based algorithm“. *International journal of health geographics* 12, lk. 14. DOI: [10.1186/1476-072X-12-14](https://doi.org/10.1186/1476-072X-12-14).
- Tiru, M. (2014). „Overview of the sources and challenges of mobile positioning data for statistics“. Teoses: *In Proceedings of the international conference on Big Data for official statistics*. Beijing, China, lk. 17–18.
- Tiru, M., Saluveer, E., Ahas, R., Aasa, A. (2010). „The Positium Barometer: A Web-Based Tool for Monitoring the Mobility of Tourists“. *Journal of Urban Technology* 17, lk. 71–89. DOI: [10.1080/10630731003597348](https://doi.org/10.1080/10630731003597348).
- Vanhoof, M., Lee, C., Smoreda, Z. (2018a). „Performance and sensitivities of home detection from mobile phone data“. *CoRR* abs/1809.09911. arXiv: [1809.09911](https://arxiv.org/abs/1809.09911). URL: <http://arxiv.org/abs/1809.09911>.
- Vanhoof, M., Reis, F., Ploetz, T., Smoreda, Z. (2018b). „Assessing the quality of home detection from mobile phone data for official statistics“. *CoRR* abs/1809.07567. arXiv: [1809.07567](https://arxiv.org/abs/1809.07567).
- Vent, K. (2014). „Inimese tegevuskohtade leidmine nutitelefonipõhiste käitumisandmestike alusel“. Magistritöö. Tartu Ülikool.
- (veebruar 2020). *Intervjuu: Positiumi ankurpunktide mudeli meetoodika*. Tartu.

# Lisad

## Lisa 1. Algandmete ülevaade

Ülevaade mobiilsidetoimingutest (vastavad kirjeldused käivad Tabeli 3 kohta):

1. Tehtud toimingute arv nelja kuu peale.
2. Keskmise tehtud toimingute arv kuus.
3. Keskmise tehtud toimingute arv päevas.
4. Keskmise tehtud toimingute arv inimese kohta.
5. Keskmise tehtud toimingute arv inimese kohta kuus.
6. Keskmise päevade arv kuus, millal isikud on toiminguid teinud.
7. Keskmise kõnede alustusaeg inimese kohta kuus.
8. Keskmise kõnede alustusaja standardhälve inimese kohta kuus.
9. Keskmise nädalavahetuse päevade hulk inimese kohta kuus.

**Tabel 3.** Mobiilsidetoimingute statistika

| Tunnus | MNO 1   | MNO 2  | Üldine  |
|--------|---------|--------|---------|
| 1.     | 122 632 | 60 971 | 183 603 |
| 2.     | 30 658  | 15 242 | 45 900  |
| 3.     | 1005    | 500    | -       |
| 4.     | 1226    | 390    | 717     |
| 5.     | 328     | 102    | 189     |
| 6.     | 6       | 7      | 7       |
| 7.     | 14:19   | 15:07  | 14:49   |
| 8.     | 0,16    | 0,14   | 0,15    |
| 9.     | 1,6     | 1,5    | 1,6     |

**Tabel 4.** Koduantenni kaugusi kirjeldav statistika

| Miinumum | q <sub>1</sub> | Mediaan | Keskmine | q <sub>3</sub> | Maksimum |
|----------|----------------|---------|----------|----------------|----------|
| 1,43     | 175,0          | 279,5   | 564,8    | 534,3          | 6902,8   |

## Lisa 2. Määratud ankurpunktide ülevaade

Üldine informatsioon määratud ankurpunktide kohta (vastavad kirjeldused käivad Tabeli 5 kohta):

1. Määratud ankurpunktide arv nelja kuu peale.
2. Keskmine määratud ankurpunktide arv unikaalsete ID kohta.
3. Maksimaalne määratud ankurpunktide arv isiku kohta.
4. Minimaalne määratud ankurpunktide arv isiku kohta.
5. Kuude lõikes keskmine määratud ankurpunktide arv.
6. Kuude lõikes määratud ankurpunktide arvu standardhälve.
7. Maksimaalne määratud ankurpunktide arv kuus inimese kohta.
8. Keskmine määratud ankurpunktide arv kuus inimese kohta.
9. Ankurpunktiks määratud unikaalsete mobiilsideantennide arv.
10. Ankurpunktiks määratud keskmine unikaalsete mobiilsideantennide arv isiku kohta.
11. Ankurpunktiks määratud keskmine unikaalsete mobiilsideantennide arv kuude lõikes.
12. Ankurpunktiks määratud keskmine unikaalsete mobiilsideantennide arv kuude lõikes isiku kohta.
13. Ankurpunktiks määratud unikaalsete asukohtade arv.
14. Ankurpunktiks määratud keskmine unikaalsete asukohtade arv kuude lõikes isiku kohta.

**Tabel 5.** Määratud ankurpunktide statistika

| <b>Tunnus</b> | <b>MNO 1</b> | <b>MNO 2</b> | <b>Üldine</b> |
|---------------|--------------|--------------|---------------|
| 1.            | 4937         | 3314         | 8251          |
| 2.            | 49           | 21           | 32            |
| 3.            | 178          | 88           | 178           |
| 4.            | 1            | 1            | 1             |
| 5.            | 1234         | 828          | 2062          |
| 6.            | 117          | 29           | 119           |
| 7.            | 52           | 27           | 52            |
| 8.            | 13           | 5            | 8             |
| 9.            | 2272         | 1476         | 3748          |
| 10.           | 38           | 14           | 24            |
| 11.           | 944          | 657          | 1601          |
| 12.           | 13           | 5            | 8             |
| 13.           | 719          | 633          | -             |
| 14.           | 13           | 5            | -             |

Üldine informatsioon kodu ja multifunktsionaalsete ankurpunktide kohta (vastavad kirjeldused käivad Tabelite 6 ja 7 kohta):

1. Tuvastatud kodu ankurpunktide arv nelja kuu peale.
2. Keskmise tuvastatud kodu ankurpunktide arv kuus.
3. Mitmel inimesel on tuvastatud 1 kodu ankurpunkt kuus?
4. Mitmel inimesel on tuvastatud 2 kodu ankurpunkti kuus?
5. Keskmise tuvastatud kodu ankurpunktide arv inimese kohta.
6. Tuvastatud multifunktsionaalsete ankurpunktide arv.
7. Keskmise tuvastatud multifunktsionaalsete ankurpunktide arv kuus.
8. Keskmise tuvastatud multifunktsionaalsete ankurpunktide arv inimese kohta.
9. Keskmise kõnede alustusaeg kodu ja multifunktsionaalsete ankurpunktide puhul inimese kohta kuus.
10. Keskmise kõnede alustusaeg kodu ankurpunktide puhul inimese kohta kuus.

11. Keskmise kõnede alustusaeg multifunktsionaalsete ankurpunktide puhul inimese kohta kuus.
12. Kõnede alustusaja standardhälvete keskmine kodu ja multifunktsionaalsete ankurpunktide puhul inimese kohta kuus.
13. Kõnede alustusaja standardhälvete keskmine kodu ankurpunktide puhul inimese kohta kuus.
14. Kõnede alustusaja standardhälvete keskmine multifunktsionaalsete ankurpunktide puhul inimese kohta kuus.
15. Mudeli samm 5 alasamm 1 põhjal määratud kodu ankurpunktide osakaal, st nende kodu ankurpunktide osakaal, mis on saanud sildi "h" keskmise kõnede alustuskellaaja põhjal.

**Tabel 6.** Määratud kodu ja multifunktsionaalsete ankurpunktide statistika (sts00 ja sts01)

| Tunnus | Stsenaarium 0 |       |        | Stsenaarium 1 |       |        |
|--------|---------------|-------|--------|---------------|-------|--------|
|        | MNO 1         | MNO 2 | Üldine | MNO 1         | MNO 2 | Üldine |
| 1.     | 397           | 490   | 887    | 396           | 478   | 874    |
| 2.     | 99            | 122   | 221    | 99            | 119   | 218    |
| 3.     | 301           | 454   | 755    | 300           | 438   | 738    |
| 4.     | 48            | 18    | 66     | 48            | 20    | 68     |
| 5.     | 4             | 3,6   | 3,8    | 4             | 3,5   | 3,7    |
| 6.     | 140           | 200   | 340    | 138           | 200   | 338    |
| 7.     | 35            | 50    | 85     | 34            | 50    | 84     |
| 8.     | 2,1           | 2,4   | 2,3    | 2,1           | 2,3   | 2,2    |
| 9.     | 13:35         | 15:50 | 14:53  | 13:34         | 15:44 | 14:48  |
| 10.    | 13:41         | 16:08 | 15:05  | 13:39         | 16:00 | 14:59  |
| 11.    | 13:18         | 15:09 | 14:23  | 13:20         | 15:06 | 14:23  |
| 12.    | 0,24          | 0,18  | 0,20   | 0,24          | 0,18  | 0,20   |
| 13.    | 0,236         | 0,18  | 0,20   | 0,236         | 0,19  | 0,21   |
| 14.    | 0,25          | 0,16  | 0,20   | 0,25          | 0,16  | 0,20   |
| 15.    | 0,08          | 0,33  | -      | 0,008         | 0,06  | -      |

**Tabel 7.** Määratud kodu ja multifunktsionaalsete ankurpunktide statistika (sts02 ja sts03)

| Tunnus | Stsenaarium 2 |       |        | Stsenaarium 3 |       |        |
|--------|---------------|-------|--------|---------------|-------|--------|
|        | MNO 1         | MNO 2 | Üldine | MNO 1         | MNO 2 | Üldine |
| 1.     | 395           | 472   | 867    | 381           | 510   | 891    |
| 2.     | 98            | 118   | 216    | 95            | 127   | 222    |
| 3.     | 301           | 434   | 735    | 321           | 476   | 797    |
| 4.     | 47            | 19    | 66     | 30            | 17    | 47     |
| 5.     | 4             | 3,4   | 3,7    | 4             | 3,7   | 4      |
| 6.     | 138           | 205   | 343    | 227           | 426   | 653    |
| 7.     | 34            | 51    | 85     | 56            | 106   | 163    |
| 8.     | 2,1           | 2,3   | 2,2    | 2,6           | 3     | 2,9    |
| 9.     | 13:34         | 15:41 | 14:47  | 13:42         | 15:24 | 14:45  |
| 10.    | 13:39         | 15:57 | 14:57  | 13:37         | 15:37 | 14:47  |
| 11.    | 13:20         | 15:06 | 14:24  | 13:49         | 15:09 | 14:41  |
| 12.    | 0,24          | 0,18  | 0,21   | 0,225         | 0,16  | 0,187  |
| 13.    | 0,237         | 0,19  | 0,21   | 0,22          | 0,16  | 0,18   |
| 14.    | 0,25          | 0,16  | 0,20   | 0,24          | 0,16  | 0,19   |
| 15.    | 0             | 0,004 | -      | 0,67          | 0,97  | -      |

Vaatlused on jaotatud koduantenni asukoha kauguste kvartiilide põhjal gruppidesse ja gruppides viiakse läbi analüüs tunnuse "Määratud ankurpunkti keskmine kaugus kodust" peal, kus hinnatakse järgmised näitajad (vastavad kirjeldused käivad Tabeli 8 kohta):

1. Määratud kodu ankurpunkti keskmine kaugus kodust (meetrites).
2. Määratud kodu ankurpunkti keskmine kaugus kodust standardhälve (meetrites).
3. Keskmine kaugus, kui määratud kodu ankurpunkti keskmine kaugus kodust ja koduantenni asukoha kaugus on samad (meetrites).
4. Keskmine kaugus, kui määratud kodu ankurpunkti keskmine kaugus kodust ja koduantenni asukoha kaugus ei ole samad (meetrites).
5. Keskmine kaugus, kui vähemalt üks määratud kodu ankurpunkti kaugus kodust ja koduantenni asukoha kaugus on samad (meetrites).
6. Keskmine kaugus, kui vähemalt üks määratud kodu ankurpunkti kaugus kodust ja koduantenni asukoha kaugus ei ole samad (meetrites).

7. Osakaal, et vähemalt üks määratud kodu ankurpunkti kaugus langeb kokku koduantenni kaugusega.
8. Kuude arv, mitmel kuul langeb ühe isiku vähemalt üks määratud kodu ankurpunkti kaugus kokku koduantenni kaugusega.
9. Koduantenniga samal kaugusel asuvate ankurpunktide osakaal.
10. Isikute osakaal, kellel vähemalt üks määratud kodu ankurpunkt asub kodust kaugemal kui 10 kilomeetrit.
11. Isikute osakaal, kellel vähemalt üks määratud kodu ankurpunkt asub kodust kaugemal kui 50 kilomeetrit.

**Tabel 8.** Gruppides hinnatud näitajad stsenaariumi 0 puhul

|     | <b>Lähedal</b> | <b>Keskmisel kaugusel</b> | <b>Kaugel</b> | <b>Väga kaugel</b> |
|-----|----------------|---------------------------|---------------|--------------------|
| 1.  | 4596           | 8157                      | 6184          | 3050               |
| 2.  | 13 667         | 22 305                    | 20 887        | 7986               |
| 3.  | 117            | 236                       | 390           | 1759               |
| 4.  | 5738           | 9770                      | 6794          | 3516               |
| 5.  | 5653           | 7354                      | 2696          | 3421               |
| 6.  | 2432           | 9285                      | 12 250        | 2542               |
| 7.  | 0,67           | 0,59                      | 0,64          | 0,59               |
| 8.  | 2              | 1,8                       | 2             | 2                  |
| 9.  | 0,55           | 0,48                      | 0,56          | 0,55               |
| 10. | 0,2            | 0,1                       | 0,08          | 0,1                |
| 11. | 0,09           | 0,06                      | 0,05          | 0,08               |

## Lisa 3. Logistilise regressiooni mudelite väljundid

**Tabel 9.** Mudel 0.1 Logistilise regressiooni mudel (sts00)

| Tunnus                          | Tase  | OR   | 95% usaldusintervall | Olulisuse tõenäosus |
|---------------------------------|-------|------|----------------------|---------------------|
| Mobiilioperaator                | MNO 1 | -    | -                    | -                   |
|                                 | MNO 2 | 1,68 | (0, 90; 3, 16)       | 0,108               |
| Koduantenni kaugus <sup>1</sup> |       | 1,05 | (0, 99; 1, 12)       | 0,134               |
| Koduantenn*MNO 2 <sup>1</sup>   |       | 0,95 | (0, 88; 1, 02)       | 0,201               |

**Tabel 10.** Mudel 0.2 Logistilise regressiooni mudel (sts00)

| Tunnus             | Tase            | OR   | 95% usaldusintervall | Olulisuse tõenäosus |
|--------------------|-----------------|------|----------------------|---------------------|
| Mobiilioperaator   | MNO 1           | -    | -                    | -                   |
|                    | MNO 2           | 1,23 | (0, 73; 2, 08)       | 0,440               |
| Koduantenni kaugus | Lähedal         | -    | -                    | -                   |
|                    | Keskmine kaugus | 0,71 | (0, 34; 1, 47)       | 0,360               |
|                    | Kaugel          | 0,85 | (0, 41; 1, 77)       | 0,663               |
|                    | Väga kaugel     | 0,7  | (0, 33; 1, 44)       | 0,327               |

**Tabel 11.** Mudel 1.1 Logistilise regressiooni mudel (sts01)

| Tunnus                          | Tase  | OR   | 95% usaldusintervall | Olulisuse tõenäosus |
|---------------------------------|-------|------|----------------------|---------------------|
| Mobiilioperaator                | MNO 1 | -    | -                    | -                   |
|                                 | MNO 2 | 1,05 | (0, 90; 3, 16)       | 0,108               |
| Koduantenni kaugus <sup>1</sup> |       | 1,05 | (0, 99; 1, 12)       | 0,134               |
| Koduantenn*MNO 2 <sup>1</sup>   |       | 0,96 | (0, 89; 1, 02)       | 0,201               |

<sup>1</sup>OR on arvatud 100 meetrise muutuse kohta

**Tabel 12.** Mudel 1.2 Logistilise regressiooni mudel (sts01)

| <b>Tunnus</b>         | <b>Tase</b>     | <b>OR</b> | <b>95% usaldusintervall</b> | <b>Olulisuse tõenäosus</b> |
|-----------------------|-----------------|-----------|-----------------------------|----------------------------|
| Mobiilioperaator      | MNO 1           | -         | -                           | -                          |
|                       | MNO 2           | 1,23      | (0, 73; 2, 08)              | 0,440                      |
| Koduantenni<br>kaugus | Lähedal         | -         | -                           | -                          |
|                       | Keskmine kaugus | 0,71      | (0, 34; 1, 47)              | 0,360                      |
|                       | Kaugel          | 0,85      | (0, 41; 1, 77)              | 0,663                      |
|                       | Väga kaugel     | 0,7       | (0, 33; 1, 44)              | 0,327                      |

**Tabel 13.** Mudel 2.1 Logistilise regressiooni mudel (sts02)

| <b>Tunnus</b>                   | <b>Tase</b> | <b>OR</b> | <b>95% usaldusintervall</b> | <b>Olulisuse tõenäosus</b> |
|---------------------------------|-------------|-----------|-----------------------------|----------------------------|
| Mobiilioperaator                | MNO 1       | -         | -                           | -                          |
|                                 | MNO 2       | 1,76      | (0, 94; 3, 32)              | 0,079                      |
| Koduantenni kaugus <sup>1</sup> |             | 1,05      | (0, 99; 1, 12)              | 0,134                      |
| Koduantenn*MNO 2 <sup>1</sup>   |             | 0,94      | (0, 87; 1, 00)              | 0,090                      |

**Tabel 14.** Mudel 2.2 Logistilise regressiooni mudel (sts02)

| <b>Tunnus</b>         | <b>Tase</b>     | <b>OR</b> | <b>95% usaldusintervall</b> | <b>Olulisuse tõenäosus</b> |
|-----------------------|-----------------|-----------|-----------------------------|----------------------------|
| Mobiilioperaator      | MNO 1           | -         | -                           | -                          |
|                       | MNO 2           | 1,19      | (0, 70; 2, 00)              | 0,522                      |
| Koduantenni<br>kaugus | Lähedal         | -         | -                           | -                          |
|                       | Keskmine kaugus | 0,71      | (0, 34; 1, 46)              | 0,351                      |
|                       | Kaugel          | 0,85      | (0, 41; 1, 77)              | 0,663                      |
|                       | Väga kaugel     | 0,65      | (0, 31; 1, 33)              | 0,241                      |

---

<sup>1</sup>OR on arvatud 100 meetrise muutuse kohta

**Tabel 15.** Mudel 3.1 Logistilise regressiooni mudel (sts03)

| <b>Tunnus</b>                   | <b>Tase</b> | <b>OR</b> | <b>95% usaldusintervall</b> | <b>Olulisuse tõenäosus</b> |
|---------------------------------|-------------|-----------|-----------------------------|----------------------------|
| Mobiilioperaator                | MNO 1       | -         | -                           | -                          |
|                                 | MNO 2       | 1,20      | (0,64; 2,24)                | 0,570                      |
| Koduantenni kaugus <sup>1</sup> |             | 1,05      | (0,99; 1,12)                | 0,134                      |
| Koduantenn*MNO 2 <sup>1</sup>   |             | 0,96      | (0,89; 1,02)                | 0,227                      |

**Tabel 16.** Mudel 3.2 Logistilise regressiooni mudel (sts03)

| <b>Tunnus</b>      | <b>Tase</b>     | <b>OR</b> | <b>95% usaldusintervall</b> | <b>Olulisuse tõenäosus</b> |
|--------------------|-----------------|-----------|-----------------------------|----------------------------|
| Mobiilioperaator   | MNO 1           | -         | -                           | -                          |
|                    | MNO 2           | 0,89      | (0,53; 1,50)                | 0,672                      |
| Koduantenni kaugus | Lähedal         | -         | -                           | -                          |
|                    | Keskmine kaugus | 0,74      | (0,36; 1,48)                | 0,392                      |
|                    | Kaugel          | 1,04      | (0,51; 2,12)                | 0,915                      |
|                    | Väga kaugel     | 0,86      | (0,42; 1,75)                | 0,680                      |

---

<sup>1</sup>OR on arvatud 100 meetrise muutuse kohta

## Lisa 4. Korduvmõõtmistega logistilise regressiooni mudelite väljundid

**Tabel 17.** Mudel 4.1 Korduvmõõtmistega logistilise regressiooni mudel (sts00)

| Tunnus             | Tase            | OR   | 95% usaldusintervall | Olulisuse tõeäosus |
|--------------------|-----------------|------|----------------------|--------------------|
| Mobiilioperaator   | MNO 1           | -    | -                    | -                  |
|                    | MNO 2           | 1,08 | (0, 29; 4, 05)       | 0,914              |
| Koduantenni kaugus | Lähedal         | -    | -                    | -                  |
|                    | Keskmine kaugus | 0,50 | (0, 08; 3, 03)       | 0,448              |
|                    | Kaugel          | 0,42 | (0, 07; 2, 51)       | 0,339              |
|                    | Väga kaugel     | 0,48 | (0, 08; 2, 98)       | 0,434              |
| Kuu                | märts           | -    | -                    | -                  |
|                    | aprill          | 1,24 | (0, 76; 2, 03)       | 0,397              |
|                    | mai             | 2,21 | (1, 28; 3, 49)       | <b>0,003</b>       |
|                    | juuni           | 0,90 | (0, 55; 1, 47)       | 0,672              |

**Tabel 18.** Mudel 5.1 Korduvmõõtmistega logistilise regressiooni mudel (sts01)

| Tunnus             | Tase            | OR   | 95% usaldusintervall | Olulisuse tõeäosus |
|--------------------|-----------------|------|----------------------|--------------------|
| Mobiilioperaator   | MNO 1           | -    | -                    | -                  |
|                    | MNO 2           | 0,91 | (0, 26; 3, 20)       | 0,884              |
| Koduantenni kaugus | Lähedal         | -    | -                    | -                  |
|                    | Keskmine kaugus | 0,50 | (0, 09; 2, 78)       | 0,430              |
|                    | Kaugel          | 0,47 | (0, 09; 2, 60)       | 0,389              |
|                    | Väga kaugel     | 0,57 | (0, 10; 3, 21)       | 0,527              |
| Kuu                | märts           | -    | -                    | -                  |
|                    | aprill          | 1,35 | (0, 83; 2, 21)       | 0,229              |
|                    | mai             | 2,13 | (1, 30; 3, 50)       | <b>0,003</b>       |
|                    | juuni           | 0,90 | (0, 55; 1, 48)       | 0,689              |

**Tabel 19.** Mudel 6.1 Korduvmõõtmistega logistilise regressiooni mudel (sts02)

| Tunnus             | Tase            | OR   | 95% usaldusintervall | Olulisuse tõeäosus |
|--------------------|-----------------|------|----------------------|--------------------|
| Mobiilioperaator   | MNO 1           | -    | -                    | -                  |
|                    | MNO 2           | 0,85 | (0, 24; 3, 01)       | 0,802              |
| Koduantenni kaugus | Lähedal         | -    | -                    | -                  |
|                    | Keskmine kaugus | 0,53 | (0, 10; 2, 98)       | 0,475              |
|                    | Kaugel          | 0,51 | (0, 09; 2, 82)       | 0,440              |
|                    | Väga kaugel     | 0,56 | (0, 10; 3, 17)       | 0,515              |
| Kuu                | märts           | -    | -                    | -                  |
|                    | aprill          | 1,35 | (0, 83; 2, 22)       | 0,228              |
|                    | mai             | 2,14 | (1, 30; 3, 52)       | <b>0,003</b>       |
|                    | juuni           | 0,82 | (0, 50; 1, 34)       | 0,423              |

**Tabel 20.** Mudel 7.1 Korduvmõõtmistega logistilise regressiooni mudel (sts03)

| Tunnus             | Tase            | OR   | 95% usaldusintervall | Olulisuse tõeäosus |
|--------------------|-----------------|------|----------------------|--------------------|
| Mobiilioperaator   | MNO 1           | -    | -                    | -                  |
|                    | MNO 2           | 0,36 | (0, 08; 1, 63)       | 0,183              |
| Koduantenni kaugus | Lähedal         | -    | -                    | -                  |
|                    | Keskmine kaugus | 0,28 | (0, 04; 2, 25)       | 0,233              |
|                    | Kaugel          | 0,52 | (0, 07; 3, 85)       | 0,521              |
|                    | Väga kaugel     | 0,72 | (0, 09; 5, 66)       | 0,757              |
| Kuu                | märts           | -    | -                    | -                  |
|                    | aprill          | 1.43 | (0, 76; 2, 69)       | 0,263              |
|                    | mai             | 1,94 | (1, 03; 3, 65)       | <b>0,040</b>       |
|                    | juuni           | 1,17 | (0, 62; 2, 19)       | 0,630              |

**Tabel 21.** Mudel 8.1 Korduvmõõtmistega logistilise regressiooni mudel (stsenaariumite võrdlemine)

| <b>Tunnus</b>      | <b>Tase</b>     | <b>OR</b> | <b>95% usaldusintervall</b> | <b>Olulisuse tõenäosus</b> |
|--------------------|-----------------|-----------|-----------------------------|----------------------------|
| Mobiilioperaator   | MNO 1           | -         | -                           | -                          |
|                    | MNO 2           | 1,01      | (0, 10; 10, 16)             | 0,993                      |
| Koduantenni kaugus | Lähedal         | -         | -                           | -                          |
|                    | Keskmine kaugus | 0,67      | (0, 03; 16, 10)             | 0,804                      |
|                    | Kaugel          | 0,93      | (0, 04; 24, 38)             | 0,964                      |
|                    | Väga kaugel     | 0,71      | (0, 03; 17, 47)             | 0,831                      |
| Stsenaarium        | sts00           | -         | -                           | -                          |
|                    | sts01           | 1,00      | (0, 18; 5, 44)              | 1,000                      |
|                    | sts02           | 0,69      | (0, 13; 3, 75)              | 0,668                      |
|                    | sts03           | 0,02      | (0, 00; 0, 18)              | <b>0,001</b>               |

## Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Merili Konik,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Ankurpunktide mudeli tundlikkus andmetiheduse ja kellaja kui sisendparameetri suhtes", mille juhendajad on Anne Selart ja Kaisa Vent, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, alates 13.05.2025 kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktis 1 ja 2 nimetatud reprodutseerimise õigus jääb alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Merili Konik

15.05.2020