

Tartu Ülikool
Loodus- ja täpisteaduste valdkond
Matemaatika ja statistika instituut

Kaidi Kapp

**Modifitseeritud
regressioonhinnang Eesti
tööjõu-uuringu näitel**

Matemaatiline statistika
Magistritöö 30 EAP

Juhendaja:
Imbi Traat (PhD)

Tartu, 2019

Modifitseeritud regressioon hinnang Eesti tööjõu-uuringu näitel

Lühikokkuvõte. Eesti tööjõu-uuring on pidevuuring, kus valimisse sattunud leibkondi küsitatakse korduvalt kindla ajavahemiku järel. Praegu kasutatakse hinnangute parandamiseks kolme kuu libiseva keskmise hinnangut ja kaalude kalibreerimist uuritava ajahetke andmeid kasutades. Modifitseeritud regressioon hinnang (*regressioon composite estimation*) kasutab lisaks uuritava ajahetke andmetele ka eelmistel perioodidel kogutud informatsiooni, et parandada hinnangute täpsust. Käesolevas töös tutvustatakse modifitseeritud regressioon hinnangu meetodikat ning rakendatakse seda Eesti tööjõu-uuringu andmetele kuiste hinnangute arvutamiseks.

Märksõnad: tööjõu-uuringud, longituudanalüüs, rotatsioon, kaalufunktsioonid, kalibreerimine

CERCS teaduseriala: P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Regression composite estimator with application to the Estonian labour force survey

Abstract. Estonian labour force survey is a monthly survey, where sampled households are repeatedly questioned at fixed intervals. Presently, three month moving average and calibration of the weights using data from the time of the study is used to improve estimates. In addition to current data, regression composite estimator uses previously collected data as well in order to improve the efficiency of the estimates. In this paper the methodology of the regression composite estimation is presented and applied to the data of the Estonian labour force survey in order to calculate monthly estimates.

Keywords: labour force surveys, longitudinal analysis, rotation, weighting functions, calibration

CERCS research specialisation: P160 Statistics, operation research, programming, actuarial mathematics

Sisukord

Sissejuhatus	5
1 Eesti tööjõu-uuring	7
1.1 Mõisted ja määratlused tööjõu-uuringus	7
1.2 Andmete kogumine	8
1.2.1 Valimi moodustamine	8
1.2.2 Isikute küsitlemine	9
1.3 Kaalumine ja imputeerimine	10
1.3.1 Disainikaalude arvutamine	11
1.3.2 Kao kompenseerimine homogeensete vastamisgruppide meetodil	11
1.3.3 Kao kompenseerimine logistilise regressiooniga	12
1.3.4 Kalibreerimine	12
1.3.5 Imputeerimine	14
1.4 Parameetrite hindamine ja avaldamine	16
1.4.1 Hinnatavad parameetrid	16
1.4.2 Standardhälbe hinnang	18
2 Modifitseeritud regressioon	20
2.1 Liitkirjeldavate tunnuste defineerimine	20
2.2 Puuduvate väärtuste imputeerimine	21
2.3 Kaalude arvutamine	22
2.4 Parameetrite hindamine	23
3 Rakendus Eesti tööjõu-uuringu andmetele	24
3.1 Andmestiku kirjeldus ja kontroll	24
3.2 Kaalude arvutamine	27
3.3 Kuisid hinnangud	29
Kokkuvõte	35
Kasutatud kirjandus	37

Lisad	39
Lisa 1 Kanada tööjõu-uuringus kasutatavad liitkirjeldavad tunnused .	39
Lisa 2 Töös kasutatavad liitkirjeldavad tunnused	40
Lisa 3 R'i kood	41
Lisa 3.1 Andmete sisselugemine	41
Lisa 3.2 Andmete kontroll	41
Lisa 3.3 Disainikaalude arvutamine	44
Lisa 3.5 Mittevastamisele korrigeeritud kaalude arvutamine . . .	44
Lisa 3.5 Lõplike kaalude ja hinnangute arvutamine	45
Lisa 3.6 Jooniste tegemine	66

Sissejuhatus

Eesti tööjõu-uuring (ETU) on Statistikaameti (SA) poolt läbiviidav pidev-uuring, mille eesmärgiks on saada ülevaade Eesti elanike tööhõivest, töötusest ning tööoludest. Nende andmete põhjal jälgitakse ja analüüsitakse Eesti majanduse ning tööturu olukorda. Samuti kasutatakse kogutud andmeid riigielarve ja muude riiklike prognooside tegemisel ning planeerimisel. [1]

SA andmebaasis avaldatakse kvartali- ja aastaandmed vaadeldava perioodi keskmistena ning kuuandmed kolme kuu libiseva keskmisena [2]. Kolme kuu libiseva keskmise hinnanguid ei ole korrektne nimetada kuuhinnanguteks. Samas kuistelt andmetelt leitud hinnangutel on suur standardhälve. Seetõttu on Statistikaamet huvitatud üle minema uuele metoodikale, mis kasutaks ühe kuu andmeid ja vähendaks hinnangu standardhälvet ilma valimimahtu suurendamata. Selleks on välja pakutud Kanada tööjõu-uuringus kasutatav modifitseeritud regressioonhinnang. Käesoleva töö eesmärk on tutvustada modifitseeritud regressiooni teooriat ning rakendada seda ETU 2012.-2016. aasta andmetele.

Töö on jagatud kolmeks osaks. Esimeses osas kirjeldatakse Eesti tööjõu-uuringu metoodikat. Välja on toodud valimi moodustamise etapid alates valikudisainist kuni andmete kogumise ja esmase kontrollini. Seejärel kirjeldatakse andmete töötlemise samme. See koosneb neljast osast: disainikaalude arvutamine, kao kompenseerimine, kalibreerimine ja imputeerimine. Kasutatav metoodika on täiendatud teoreetiliste valemitega. Lõpuks tuuakse välja tööjõu-uuringus hinnatavad parameetrid ja nende hindamise valemid.

Teises osas antakse ülevaade modifitseeritud regressioonhinnangu teooriast. See on üldistatud regressioonhinnangu edasiarendus, kus osade kirjeldavate tunnuste kontrollsummad ei ole registritest kättesaadavad ning on seetõttu eelmise perioodi andmetelt hinnatud. Selliseid kontrollsummasid nimetatakse liitkontrollsummadeks. Metoodika eeldab valimi rotatsiooni. See tähendab, et valimisse sattunud objektid jäävad valimisse pikemaks ajaks ning neid küsitatakse korduvalt kindla perioodi järel. Modifitseeritud regressioonhinnangu arvutamiseks kasutatavate liitkirjeldavate tunnuste defineerimiseks jagatakse vaadeldava perioodi valim kaheks: uued valimi objektid ning eelnevalt vastanud objektid (roteeruv valim). Kuna uutele valimi objektidele ei ole eelmise perioodi andmed teada, määratakse neile liitkirjeldavate tunnuste väärtused eelmise perioodi hinnanguid ning uuritava kuu vastuseid kasutades. Roteeruvale valimile määratakse liitkirjeldavad tunnused nende eelmise ja uuritava perioodi vastuse kombinatsioonina. Seejärel kaalud kalibreeritakse nii kirjeldavaid kui ka liitkirjeldavaid tunnuseid kasutades. [3]

Kolmas osa on praktiline töö, mis põhineb Eesti tööjõu-uuringu 2012.-2016. aasta andmetel. Praktilises osas kirjeldatakse kasutatavaid tunnuseid, teostatakse andmetele loogiline kontroll ning kontrollitakse töös kirjeldatud meetodite Statistikaametist saadud andmetele rakendamise võimalikkust. Arvutatakse disainikaalud ning mittevastamisele korrigeeritud kaalud nii kolme järjestikuse kuu kui ka ühe kuu andmeid kasutades. Arvutatakse kalibreeritud kaalud ja modifitseeritud regressiooni kaalud ning leitakse hinnangud töötavate, töötute ning tööturul mitteaktiivsete inimeste arvudele ning hinnangute standardhälvetele. Leitud hinnangud esitatakse graafikutena, mille põhjal neid võrreldakse ja kommenteeritakse.

Magistritöö on vormistatud tekstitöötlusprogrammiga $\text{L}^{\text{T}}\text{E}^{\text{X}}$. Allikatele on viidatud tekstis nurksulgude abil. Andmete analüüsiks ja jooniste tegemiseks on kasutatud statistikapaketti R. Tulemi failide varundamiseks ja mugandamiseks on kasutatud tabelarvutusprogrammi Microsoft Excel.

Autor tänab Statistikaameti meeskonda teema väljapakkumise, Eesti tööjõu-uuringu andmete kasutamise loa ning rohkete täpsustuste eest nii meetoodika kirjeldamise kui ka rakendamise etappides.

1 Eesti tööjõu-uuring

Ettevõtete uuring ei hõlma mitteametlikult töötavaid elanikke ning Töötukassal on informatsioon ainult seal registreeritud töötute kohta. Kogu elanikkonda hõlmav Rahvaloendus toimub aga iga 10 aasta tagant. Seetõttu on vaja lisauuringut, mis võimaldaks tööjõudu hinnata ka rahvaloenduste vahele jääval ajal. Selleks viidigi 1995. aastal läbi esimene Eesti tööjõu-uuring. Kuna uuringu läbiviimine on ajas muutunud, kirjeldatakse töös edaspidi 2012.-2016. aastal kasutusel olnud meetodikaid. [4]

1.1 Mõisted ja määratlused tööjõu-uuringus

Käesolev alapeatükk toob välja tööjõudu puudutavad mõisted, mis on esitatud Tööjõu-uuringu metaandmete kirjelduses [2]. Kuna tegemist on definitsioonidega, siis on sõnastused muutmata kujul.

Tööealine rahvastik – rahvastiku majandusliku aktiivsuse uurimisel aluseks võetavas ehk tööjõu-uuringu objektiks olevas vanusevahemikus rahvastik (15–74-aastased).

Majanduslikult aktiivne rahvastik ehk tööjõud – isikud, kes soovivad töötada ja on võimelised töötama (hõivatute ja töötute summa).

Majanduslikult passiivne ehk mitteaktiivne rahvastik – isikud, kes ei soovi töötada või ei ole selleks võimelised.

Leibkond – ühises põhieluruumis (ühisel aadressil) elavate isikute rühm, kes kasutab ühiseid raha- ja/või toiduressursse ja kelle liikmed ka ise tunnistavad, et on ühes leibkonnas. Leibkond võib olla ka üksikisik.

(Tööga) hõivatu – isik, kes uuritava perioodil

- töötas ja sai selle eest tasu kas palgatöötajana, ettevõtjana või vabakutseliseksena;
- töötas otsese tasuta pereettevõttes või oma talus;
- ajutiselt ei töötanud.

Töötu – isik, kelle puhul on korraga täidetud kolm tingimust:

- on ilma tööta (ei tööta mitte kusagil ega puudu ajutiselt töölt);

- on töö leidmisel valmis kohe (kahe nädala jooksul) tööd alustama;
- otsib aktiivselt tööd.

Vaeghõivatu – osaaajatöötaja, kes soovib rohkem töötada ja on valmis lisatööd kohe (kahe nädala jooksul) vastu võtma.

Heitunud isik – mittetöötav isik, kes sooviks töötada ja oleks valmis töö olemasolu korral ka kohe tööle asumata, kuid ei otsi aktiivselt tööd, sest on kaotanud lootuse seda leida.

Tööhõive määr – hõivatute osatähtsus tööealises rahvastikus.

Tööjõus osalemise määr (aktiivsuse määr) – tööjõu osatähtsus tööealises rahvastikus.

Töötuse määr ehk tööpuuduse määr – töötute osatähtsus tööjõus.

1.2 Andmete kogumine

Käesolev alapeatükk kirjeldab Eesti tööjõu-uuringu andmete kogumise etappe ning põhineb ETU meetodika kirjeldusel [4]. Alapeatükk on jagatud kaheks osaks: valimi moodustamine ja isikute küsitlemine.

1.2.1 Valimi moodustamine

ETU üldkogumiks $U = \{1, 2, \dots, N\}$ on kõik alaliselt Eestis elavad tööealised isikud ehk inimesed, kes olid uuringu ajal 15-74-aastased. Üldkogumi maht on seega N . Uuringu freimiks on Rahva ja eluruumide loenduse ja rahvastiku-registri alusel loodud 15-74-aastaste Eesti alaliste elanike loend. [2]

Valim s moodustatakse lihtsa juhusliku kihtvalikuga. Kihid h moodustatakse elukoha piirkonna järgi [2]:

- Tallinn;
- neli suuremat maakonda (Harjumaa(v.a Tallinn), Ida-Virumaa, Pärnumaa, Tartumaa);

- kümme väiksemat maakonda (Jõgevamaa, Järvamaa, Läänemaa, Lääne-Virumaa, Põlvamaa, Raplamaa, Saaremaa, Valgamaa, Viljandimaa, Võrumaa);
- Hiiumaa.

Iga valimisse sattunud isik kaasab valimisse kõik oma tööelised leibkonnaliikmed. Lisaks küsitlusandmetele kasutatakse sotsiaalkindlustusameti sotsiaalkaitse infosüsteemist saadud andmeid isikute töövõime kaotuse protsendi ja puude raskusastme kohta ning Eesti Hariduse Infosüsteemi (EHIS) registri andmeid isikute praeguste õpingute kohta, kui isik õpib Eestis. [2]

Valimisse sattunud leibkondi küsitletakse 4 erineval kuul. Selleks kasutatakse 2-2-2-rotatsiooniplaani, ehk valimisse sattunud leibkondi küsitletakse 2 järjestikusel kvartalil, siis 2 kvartalil ei küsitleta ning seejärel 2 kvartalil küsitletakse. Näiteks kui leibkond kaasatakse esimest korda valimisse 2012. aasta veebruaris, siis küsitletakse teda teist korda 2012. aasta mais, kolmandat korda 2013. aasta veebruaris ning neljandat korda 2013. aasta mais.

Uued valimi objektid valitakse enne küsitlusaasta algust ning jaotatakse ühtlaselt 52 uuringunädala peale - see tähendab, et igal nädalal küsitletakse võrdne arv leibkondi. Enne 2002. aastat arvestati küsitletava vanuse määramisel tema vanust uuringu aasta 1. jaanuari seisuga, kuid alates 2002. aastast arvestatakse ka sünnikuupäeva. Samuti arvestatakse sünnikuupäeva pensioniea määramiseks. Aasta jooksul valimit korrigeeritakse (näiteks infoga surmade kohta). Pärast korrigeerimist saadud valimit nimetatakse loendiveata valimiks. [5]

1.2.2 Isikute küsitlemine

Võimalusel kontakteerutakse valimisse sattunud isikutega enne uuringunädala algust, et leppida kokku sobivaim intervjuu läbiviimise aeg. Kui küsitlaja respondentiga esimesel korral kontakti ei saa, tuleb püüda uuesti mõnel teisel nädalapäeval ja kellaajal, sest valimiisiku ja tema leibkonna elurütm ja kodus viibimise aeg ei ole teada. Parim aeg küsitluse läbiviimiseks on uuringunädalale järgnev nädal.

Küsimustiku põhiosa koosneb isiku- ja leibkonnaküsitlusest, mis on aja jooksul väga vähe muutunud. Põhiankeedile lisatakse erinevaid mooduleid Euroopa Liidu ning Eesti andmekasutajate soovil. Lisatud moodulid jäävad ankeeti kas terveks aastaks või mõneks kvartaliks olenevalt vajalikkust vastatud ankeetide arvust. Eelmise aasta isiku- ja leibkonnaküsitlusega on võimalik tutvuda

Statistikaameti kodulehel [1].

Küsitluse läbiviimiseks kasutatakse sülearvutit. Erandolukorras (turvalisuse kaalutlustel, tehnilise rikke tõttu vms põhjusel) on küsitlejatel lubatud kasutada ka paberankeeti. Tööjõu-uuringu leibkonnaküsitlusele vastab üks leibkonna 15–74-aastastest liikmetest, kelle valib leibkond ise. Isikuküsitlusele vastab iga 15–74-aastane leibkonnaliige enda kohta ise. Kui mõni leibkonnaliige ei ole kättesaadav, siis võib tema eest ja tema kohta vastata mõni teine leibkonnaliige (*proxy*-intervjuu). Küsimustikule saab vastata nii eesti, vene kui ka inglise keeles.

Andmete kontroll algab juba küsitluse ajal, sest elektroonilisse küsitluskeskonda on lisatud loogilised kontrollid. Tekkinud vastuolude korral vead parandatakse või lisatakse seletav kommentaar. Statistikaametisse laekunud andmed kontrollitakse taas ning leitud vead kas parandatakse kohe või pöördutakse küsitleja (vajadusel ka küsitletava) poole täpsustamiseks.

1.3 Kaalumine ja imputeerimine

Alapeatükk põhineb ETU metoodika kirjeldusel [4] ning on täiendatud Valik-uuringute teooria ainekursuse konspekti [6] ja andmetöötlusprogrammi SAS protseduuri CALMAR kirjeldusega [9].

Tööjõu-uuringus on valimi kihid erineva suurusega. Samuti ei valita igas kihis võrdne arv ega ka võrdne protsent isikuid. Lisaks sellele on tööjõu-uuringule vastamine vabatahtlik ning osa valimisse sattunud isikutest jätab küsitlusele vastamata. Nende isikute hulka nimetatakse valimi kaoks. Kõige selle tõttu ei saa valimit kohe üldkogumile üldistada, sest valim ei peegelda tegelikke üldkogumi demograafilisi näitajaid. Selle probleemi lahendamiseks leitakse igale vastanud isikule kaal, mis näitab, mitut üldkogumi objekti isik esindab.

Lõplike kaalude arvutamiseks on kolm etappi:

- disainikaalude arvutamine,
- kao kompenseerimine,
- kalibreerimine.

1.3.1 Disainikaalude arvutamine

Lihtsa juhusliku kihtvaliku (LJKV) korral arvutatakse kaasamistõenäosus iga kihi jaoks eraldi ning see avaldub valemiga

$$f_{h,LJKV} = \frac{n_h}{N_h}, h = 1, \dots, H, \quad (1)$$

kus n_h ja N_h on vastavalt valimi- ja üldkogumimaht kihis h .

Kuna ETUs kaasab iga valimiisik ka kõik tema tööelised leibkonnaliikmed, on isiku kaasamistõenäosus seda suurem, mida rohkem on tema leibkonnas 15-74-aastaseid isikuid. Seega tuleb uuringus kasutatavate kaasamistõenäosuste leidmiseks korrutada LJKV kaasamistõenäosusi leibkonna l tööeliste leibkonnaliikmete arvuga k_l . Saame valemi

$$f_{h_l} = \frac{n_h \cdot k_l}{N_h}.$$

Isiku i kaasamistõenäosus on võrdne tema leibkonna l kaasamistõenäosusega.

Disainikaal näitab, mitut üldkogumi objekti valimisse kuuluv isik esindab ning on pöördvõrdeline kaasamistõenäosusega

$$d_{h_l} = \frac{N_h}{n_h \cdot k_l}. \quad (2)$$

ETUs fikseeritakse kihtide elanike arv N_h iga aasta 1. jaanuari seisuga.

1.3.2 Kao kompenseerimine homogeensete vastamisgruppide meetodil

Aastatel 2012-2014 kasutati kao kompenseerimiseks homogeensete vastamisgruppide meetodit. Selle meetodika kohaselt jagatakse valimiisikud elukoha maakonna (15 maakonda + Tallinn) ja linnastumise astme (linn või maa- piirkond) järgi 31 rühma. Igale rühmale arvutatakse kaoprotsent (mittevas- tanute arv jagatakse rühma suurusega) ning järjestatakse kaoprotsendi järgi. Seejärel moodustatakse rühmi liites seitse võimalikult sarnase suurusega kao- rühma h^* . Igale rühmale leitakse korrigeerimisfaktor valemiga

$$r_{h^*} = \frac{v_{h^*}}{n_{h^*}},$$

kus n_{h^*} on kao kompenseerimise kihi valimimahu suurus ja v_{h^*} on vastanute arv samas kaokihis.

Vastanute disainikaalu korrigeeritakse valemiga

$$w_l^{NA} = \frac{d_{h_l}}{r_{h^*}}.$$

1.3.3 Kao kompenseerimine logistilise regressiooniga

Kuigi seda ametlikes metoodika aruannetes välja ei ole toodud, toimus 2015. aastal metoodika muudatus ning kao kompenseerimiseks hakati kasutama logistilist regressiooni. Metoodika kirjeldus põhineb artiklil [7] ja Andmeanalüüs II ainekursuse konspektil [8].

Ka selle metoodika puhul otsitakse korrigeerimisfaktorit r_i , mis on isiku i vastamistõenäosus. Korrigeerimisfaktori arvutamiseks kasutatakse logistilist regressioonimudelit, mis kasutab *Logit* seosefunktsiooni $\eta = \text{logit}(\pi) = \ln \frac{\pi}{1-\pi}$, kus π on sündmuse esinemise tõenäosus ehk antud juhul korrigeerimisfaktor r_i ja $\frac{\pi}{1-\pi}$ on vastamise šanss. Logistiline mudel avaldub kujul

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j,$$

kus β_0, \dots, β_j on mudeli tundmatud parameetrid ning x_1, \dots, x_j seletavad tunnused. Eesti tööjõu-uuringus kasutatakse seletavateks tunnusteks elukohta (maakond, linnastumise aste), vanusgruppi (viie aasta kaupa), isiku uuringus osalemise kordsust ja leibkonna liikmete arvu.

Isiku i vastamistõenäosus avaldub kujul

$$r_i = \pi = \frac{e^\eta}{1 + e^\eta}$$

ning tema kaalu korrigeeritakse valemiga

$$w_i^{NA} = \frac{d_{h_l}}{r_i}. \quad (3)$$

1.3.4 Kalibreerimine

Kalibreerimisel kasutatakse kaalude korrigeerimiseks registrist pärit abiinformatsiooni, et uute kaaludega leitud üldkogumijaotused peegeldaksid võimalikult täpselt tegelikke demograafilisi näitajaid. Praktikas on levinud kolm juhtu:

- Info U : registrist on kättesaadav abiinformatsioon üldkogumi kohta. Saadav abiinformatsioon jaguneb kaheks:

- a objekti tasemel informatsioon – iga üldkogumi objekti ($i = 1, 2, \dots, N$) jaoks on teada J abitunnuse väärtused. Olgu objekti i abiinformatsioon $\mathbf{x}_i = (x_{i1}, \dots, x_{iJ})^T$. Pöörame tähelepanu, et \mathbf{x}_i on vektor. Ka edaspidi on vektorid tähistatud paksus kirjas.
- b agregeeritud kujul informatsioon – registrist on kättesaadavad vaid summad $\mathbf{t}_x = \sum_{i \in U} \mathbf{x}_i$. Vastanutelt küsitakse samade tunnuste \mathbf{x}_i väärtusi.
- Info S : abiinformatsioon on uuringu läbiviijatel teada ainult valimi s kohta kas üksikute väärtustena või agregeeritud kujul.
 - Info US : mõned tunnused on teada valimi ja teised üldkogumi tasemel.

Mida suurem on korrelatsioon uuritavate tunnuste ja abitunnuste vahel, seda täpsema hinnangu annavad kalibreeritud kaalud.

Näiteks võib abitunnus \mathbf{x}_i olla defineeritud järgmiselt:

$$\mathbf{x}_i = (\gamma_{1i}, \gamma_{2i}, \dots, \gamma_{ji})^T,$$

kus

$$\gamma_{ji} = \begin{cases} 1, & \text{kui } i \text{ kuulub gruppi } j \\ 0, & \text{muidu.} \end{cases}$$

Seega, on vektori \mathbf{x}_i ühe elemendi väärtuseks 1, kui objekt i vastavasse gruppi kuulub, ning ülejäänud on nullid. Sellist klassifitseerivat abitunnust on võimalik moodustada kõikide tunnuste puhul. Kui on tegemist arvtunnusega (näiteks vanus), siis jagatakse see tunnus kattumatuteks intervallideks ning γ_{ji} näitab kuuluvust konkreetsesse intervalli.

Tööjõu-uuringus kasutatakse demograafilistest andmetest teadaolevat Eesti elanike jaotust soo ja vanuserühma (viie aasta kaupa), isiku elukoha maakonna, linnastumise astme (linn või maa) ning rahvuse (eestlane või mitte) järgi uuringuaasta 1. jaanuari seisuga [4].

Kalibreerimisel otsitakse uusi kaale w_i^c , mis oleksid võimalikult sarnased varasematele kaaludele (antud juhul mittevastamisele korrigeeritud kaaludele w_i^{NA}) ning rahuldaksid kalibreerimisvõrrandit

$$\sum_{i \in s_v} w_i^c \mathbf{x}_i = \mathbf{t}_x.$$

Selleks sobivad kaalud

$$w_i^c = w_i^{NA} \cdot v_i, \tag{4}$$

kus v_i nimetatakse g-kaaludeks. G-kaalud arvutatakse valemiga

$$v_i = F(\mathbf{x}_i^T \cdot \boldsymbol{\lambda}),$$

kus F on kalibreerimisel kasutatav kaugusfunktsioon ja $\boldsymbol{\lambda}$ on Lagrange'i kordajate vektor, mis määratakse võrrandist

$$\sum_{i \in s_v} w_i^{NA} \cdot F(\mathbf{x}_i^T \cdot \boldsymbol{\lambda}) \cdot \mathbf{x}_i = \mathbf{t}_x.$$

Kalibreerimisel kasutatakse järgmiseid meetodeid:

- **Lineaarne meetod** – kasutab kaugusfunktsiooni

$$\sum_{i \in s_v} \frac{(w_i^c - w_i^{NA})^2}{w_i^{NA}} \quad (5)$$

minimiseerimist. Meetod on kiire, kuid kalibreeritud kaalud võivad osutuda negatiivseteks. Samuti ei ole kalibreeritud kaalud tõkestatud ning nende jaotus võib olla ebasümmeetriline.

- **Eksponentsiaalne meetod** – kasutab kaugusfunktsiooni

$$w_i^c \cdot \log\left(\frac{w_i^c}{w_i^{NA}}\right) - w_i^c + w_i^{NA}$$

minimiseerimist. Selle meetodiga saadakse alati positiivsed kaalud. Küll aga ei ole kaalud ülevalt tõkestatud ning võivad seetõttu omandada väga suuri väärtuseid.

- **Logit ja tõkestatud lineaarne meetod** – mõlema meetodi puhul määratakse intervallina piirid g-kaaludele v_i . G-kaalude piirid tuleb määrata vanade kaaludega arvutatud hinnanguid ning kontrollsummasid arvesse võttes. Praktikas määratakse need katse ja eksituse meetodiga, alustades väiksemast intervallist ning seejärel seda suurendades kuni lahendus on leitav.

1.3.5 Imputeerimine

Kui vastaja ei soovi või ei oska küsimusele vastata, peab küsitleja talle küsimust põhjalikumalt selgitama ja aitama küsitletavat lisaküsimustega. Mõnikord võib siiski lõplikuks vastuseks jääda „ei tea“, „ei oska öelda“ või „keeldun vastamast“. Kõige enam jäetakse vastamata kuutöötasu (bruto- ja netotasu) ja leibkonna kuusissetuleku küsimustele. Teistele küsimustele jäetakse vastamata harva. Vastamata jäetakse rohkem *proxy*-intervjuudes, kui isiku eest vastab teine leibkonnaliige. Andmestiku lünkade eemaldamiseks, kasutatakse omistusmeetodeid.

Järgnevalt on välja toodud mõned omistusmeetodid:

- **üldise keskmise omistus** – omistatakse vastanute hulga keskmine. See meetod ei sobi nominaaltunnuste jaoks;
- **klassi keskmise omistus** – valim jagatakse teadaolevate tunnuste väärtuste alusel omistusklassidesse. Puuduvale väärtusele omistatakse klassi keskmine;
- **Hot-Deck** – sama andmestiku doonori väärtuse omistamine. Rakendatakse nii üldise juhusliku omistuse kui ka klassi juhusliku omistuse abil;
- **Regressioonijärgne omistus** – andmestiku puuduvad väärtused prognoositakse regressioonvõrrandiga, mis on moodustatud vastanud objektide andmeid kasutades. Abitunnused peavad olema teada kõikidele valimi objektidele.

Kui vastaja ei soovi oma leibkonna sissetulekut avaldada, siis palutakse tal valida oma leibkonna sissetuleku vahemik ette antud vahemike seast. Sellisel juhul imputeeritakse sissetulek kõikide samas vahemikus leibkonna sissetulekuga vastanute seast *Hot-Deck* meetodiga. Neile, kes ei vastanud ka vahemiku küsimusele, määratakse peretüüp leibkonna liikmete arvu ja vanuse ning laste arvu järgi ning omistatakse selle peretüübi keskmine sissetulek. Erisutatavad peretüübid on:

- üksik alla 65-aastane;
- üksik 65-aastane või vanem;
- paar, kellest vähemalt üks on alla 65-aastane;
- üle 65-aastane paar;
- ühe või enama lapsega üksikvanem;
- ühe lapsega paar;
- kahe lapsega paar;
- kolme lapsega paar;
- üle kolme lapsega paar;
- muu.

Ühtlasi imputeeritakse ka isiku palga puuduvaid väärtusi. Kui vastaja ei nõustu oma täpset palka ütleva, küsitakse talt selle vahemikku ning imputeeritakse täpne palk sellest vahemikust *Hot-Deck* meetodiga. Isikutele, kes keeldusid ka vahemikku avaldamast, palka ei imputeerita.

Alates 2015. aastast kasutatakse imputeerimist küsitlusele mittevastanud isikutele leibkonnaliikmete arvu määramiseks. Kui mittevastanud isiku mõni teine leibkonnaliige on küsitlusele vastanud, määratakse mittevastanud isikule leibkonnaliikmete arv vastanu vastuse põhjal. Neile, kelle leibkonnast ei ole keegi vastanud, määratakse leibkonnaliikmete arv järgmiselt:

1. võetakse isiku leibkonnaliikmete arv valimi võtmise aadressibaasist;
2. kui eelmises punktis leitud liikmete arv on suurem kui 8, imputeeritakse talle leibkonnaliikmete arv kõikide vastanute seast juhusliku regressioonimudeliga.

Siiski tuleb meeles pidada, et imputeerimine ei likvideeri nihet hinnangutes, sest omistatavate väärtuste leidmiseks kasutatakse vastanute andmeid. Samuti alahinnatakse hinnangute dispersiooni. Kõrgemal ametipositsioonil olevad inimesed jätaavad märksa sagedamini palgaküsimustele vastamata ning kuna nende palgad on tavaliselt suuremad, siis on palga hinnang alahinnatud. Seega ei ole mõistlik seda hinnangut järelduste tegemisel kasutada.

1.4 Parameetrite hindamine ja avaldamine

Järgnev alapeatükk põhineb Kanada tööjõu-uuringu meetoodika kirjeldusele [3].

Tööjõu-uuringus kogutud andmed on sisendiks paljudele erinevatele statistikatöödele ning ETU tulemuste põhjal ilmub kord kvartalis pressiteade „Tööhõive ja töötus“ [4]. Kõige esimesena avaldatakse Tööjõu-uuringu tulemused Statistikaameti andmebaasi valdkonnas „Sotsiaalelu / Tööturg“ [2].

1.4.1 Hinnatavad parameetrid

Tööjõu-uuringu üks põhilisi eesmärgi on uuritava tunnuse y kogusumma $t_y = \sum_{i \in P} y_i$ hindamine, kus P on uuritava kuu huvi pakkuv populatsioon. Näiteks võib y_i olla binaarne tunnus, mis näitab, kas isik töötab ($y_i = 1$) või ei tööta ($y_i = 0$). Sellisel juhul näitab t_y uuritava populatsiooni töötajate arvu. Uuritavaks populatsiooniks võib olla nii kogu Eesti tööealised elanikud, maakonna

tööealine rahvastik, kitsendatud vanusegrupp, vms. Kasutades korrigeeritud kaale (4), saame kogusumma hinnanguks

$$\hat{t}_y = \sum_{i \in s_v} w_i^c y_i, \quad (6)$$

kus s_v on vastanute hulk.

Teine huvipakkuv suurus on rahvastiku määra hinnang

$$r_{y_1, y_2} = \frac{\sum_{i \in s_v} w_i^c y_{1i}}{\sum_{i \in s_v} w_i^c y_{2i}}, \quad (7)$$

kus y_1 ja y_2 on kaks huvipakkuvat tunnust. Näiteks töötuse määra hindamiseks on y_{1i} binaarne tunnus, mis näitab, kas uuritav isik on töötu ($y_{1i} = 1$ kui on töötu ja $y_{1i} = 0$ kui isik töötab või on heitunud) ning y_{2i} näitab, kas ta on töøjõus.

Kolmandaks pakub huvi hinnatavate parameetrite keskmise hindamine pikema aja jooksul kui üks kuu. Sellist parameetrit θ nimetatakse T-kuu libisevaks keskmiseks ning see avaldub valemiga

$$\theta = \sum_{t=1}^T \frac{\theta_t}{T},$$

kus θ_t on uuritav parameeter, t on kuu indeks ja T arvutuseks kasutatud kuude arv. T-kuu libiseva keskmise hinnang leitakse valemiga

$$\hat{\theta} = \sum_{t=1}^T \frac{\hat{\theta}_t}{T}, \quad (8)$$

kus $\hat{\theta}_t$ on θ_t hinnang kasutades lõplikke kaale (4). Kolme kuu libisevat keskmist nimetatakse ka libiseva kvartali hinnanguks.

ETU hinnatavad parameetrid avaldatakse järgnevalt[4]:

- kuuandmed kolme kuu libiseva keskmise hinnanguna 60 päeva pärast aruandekuu lõppu;
- kvartaliandmed kvartali kuude keskmisena 45 päeva pärast aruandekvartali lõppu;
- aastakeskmised 12 kuu keskmisena 45 päeva pärast aruandeaasta lõppu.

Kuigi andmebaasides omistatakse libiseva kvartali hinnangud kvartali keskmisele kuule (näiteks aprilli, mai ja juuni andmetel leitud hinnang on andmebaasides kui maikuu hinnang), ei tohiks neid siiski käsitleda kuiste hinnangutena. Hinnangud, mis on arvatatud vähem kui 20 objekti andmetelt, ei avalikustata ning need väljad on andmebaasides tühjad [5].

1.4.2 Standardhälbe hinnang

Järgnev alapeatükk põhineb SASi kasutajatoe koduleheküljel [10].

Eesti tööjõu-uuringus kasutatakse hinnangute arvutamiseks programmi SAS protseduuri *SURVEYMEANS*, mis kasutab vaikimisi dispersiooni hindamiseks Taylori meetodit.

Kogusumma (6) dispersiooni hinnang avaldub kujul

$$\hat{V}(\hat{t}_y) = \sum_{h=1}^H \hat{V}_h(\hat{t}_y),$$

kus H on kihtvaliku kihtide arv, $\hat{V}_h(\hat{t}_y)$ dispersiooni hinnang kihis h ning avaldub valemiga

$$\hat{V}_h(\hat{t}_y) = \frac{n_h(1 - f_h)}{n_h - 1} \sum_{l=1}^{n_h} (y_{hl\cdot} - \bar{y}_{h\cdot\cdot})^2,$$

kus n_h on leibkondade arv kihis h , f_h kaasamistõenäosus (1),

$$y_{hl\cdot} = \sum_{i=1}^{k_l} w_i^c y_i,$$

kus k_l on liikmetearv leibkonnas l , ning i isik leibkonnas l ,

$$\bar{y}_{h\cdot\cdot} = \left(\sum_{l=1}^{n_h} y_{hl\cdot} \right) / n_h.$$

Rahvastiku määra (7) dispersiooni hindamiseks vaatame esmalt näiteks töötusemäära hindamist. Selleks defineeritakse alampopulatsioonis tööjõud indikaator tunnuse y , mis näitab, kas vaadeldav isik on töötu või mitte. Järelikult on töötusemäär sisuliselt defineeritud indikaatoritunnuse keskmine vaadeldud alampopulatsioonis. Analoogse tuletuskäigu saab läbi teha ka teiste rahvastiku määra valemiga hinnatavatele parameetritele. Tähistame keskmise hinnangu \hat{Y} . Kuna \hat{Y} nimetaja (uuritava populatsiooni suurus) on ka juhuslik suurus,

kasutatakse dispersiooni leidmisel kõigepealt Taylori lineariseerimist ja dispersioon avaldub jääkide $y_i - \hat{Y}$ kaudu. Seega on rahvastiku määra dispersiooni hinnang arvutatav kujul:

$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \hat{V}_h(\hat{Y}),$$

kus

$$\hat{V}_h(\hat{Y}) = \frac{n_h(1 - f_h)}{n_h - 1} \sum_{l=1}^{n_h} (e_{hl} - \bar{e}_{h..})^2,$$

$$e_{hl} = \left(\sum_{i=1}^{k_l} w_i^c (y_i - \hat{Y}) \right) / \sum_{i=1}^p w_i^c,$$

kus p on isikute arv uuritavasse alampopulatsiooni (antud näites töøjõudu) kuuluvas valimis,

$$\bar{e}_{h..} = \left(\sum_{l=1}^{n_h} e_{hl} \right) / n_h.$$

Standardhälve on ruutjuur dispersioonist

$$Std(\hat{\theta}) = \sqrt{\hat{V}(\hat{\theta})}, \quad (9)$$

kus $\hat{\theta}$ on huvipakkuva parameetri hinnang.

2 Modifitseeritud regressioon

Käesolev peatükk põhineb Kanada tööjõu-uuringu metoodika kirjeldusel [3] ja metoodikat kirjeldaval artiklil [13].

Kuni 2000. aasta jaanuarini kasutati Kanada tööjõu-uuringus hinnangute parandamiseks regressioonhinnangut ehk kalibreerimist lineaarse meetodiga. See meetod kasutab abiinformatsioonina ainult uuritava kuu andmeid. Kuigi oli teada, et roteeruva valikudisaini puhul aitab varasemate kuude informatsiooni kasutamine hinnangute täpsust parandada, ei oldud seni leitud teistes uuringutes kasutusel olevate meetodite seast ühtegi, mis oleks parandanud hinnangute täpsust ja sobitunud senise kaalude korrigeerimise süsteemiga. Seetõttu loodi uus eelnevate perioodide informatsiooni kasutav meetod, mis ühtlasi:

- muudaks võimalikult vähe juba kasutusel olevat hindamissüsteemi;
- looks kõikidele valimi objektidele unikaalse lõpliku kaalu;
- arvestaks soo- ja vanusrühmade ja geograafiliste kontrollsummadega;
- looks järjekindlad hinnangud (näiteks et hõivatud + töötud = tööjõud ja töötud + mitteaktiivsed = elanike arv).

Välja töötatud meetod nimetati modifitseeritud regressiooniks (*regression composite estimation*). See on sisuliselt ikka kalibreerimine, kuid mõned kontrollsummad on eelmise kuu andmetel põhinevad hinnangud. Samuti ei ole kõik nende hinnangutega seotud kirjeldavad tunnused teada kõikide valimi objektide jaoks ning on seetõttu neile imputeeritud. Neid kontrollsummasid ja kirjeldavaid tunnuseid nimetatakse liitkontrollsummadeks (*composite control totals*) ja liitkirjeldavateks tunnusteks (*composite auxiliary variables*). Kanada uuringus on 25 liitkirjeldavat tunnust (Lisa 1), mis on kõik määratud eelmise kuu kohta.

2.1 Liitkirjeldavate tunnuste defineerimine

Olgu vaadeldav kuu t ning eelmine kuu $t - 1$. Isiku i liitkirjeldavate tunnuste vektor on $\mathbf{z}_{t-1,i}$ ning liitkontrollsummade vektor $\hat{\mathbf{t}}_z$. Kuna $\mathbf{z}_{t-1,i}$ on defineeritud eelmise kuu kohta, tuleb ka liitkontrollsummad $\hat{\mathbf{t}}_z$ leida eelmise kuu andmete põhjal valemiga (6), kus kalibreeritud kaalude w_i^c asemel kasutatakse modifitseeritud kaale w_i^{cc} . Paraku ei ole \mathbf{z}_{t-1} vaadeldud valimi uute objektide jaoks, sest neid ei küsitletud eelmisel korral. Selle probleemi lahendamiseks kasutatakse kahe imputeerimismeetodi kombinatsiooni.

Esimene meetod on keskmisega asendamine, kus defineeritakse modifitseeritud vektor

$$\mathbf{z}_{\cdot i}^{(1)} = \begin{cases} \mathbf{z}_{t-1,i} & , \text{ kui } i \in (s_v - s_v^b); \\ \hat{\mathbf{t}}_z/N & , \text{ kui } i \in s_v^b, \end{cases}$$

kus s_v käesoleval kuul küsitlusele vastanud inimeste hulk, s_v^b sellel kuul valimisse lisandunud ja küsitlustele vastanud inimeste hulk, N on 15-74-aastaste elanike arv. See meetod on osutunud efektiivseks kuiste hinnangute arvutamiseks.

Teises meetodis kasutatakse modifitseeritud vektorit

$$\mathbf{z}_{\cdot i}^{(2)} = \begin{cases} \mathbf{z}_{t-1,i} + (\delta_i^{-1} - 1)(\mathbf{z}_{t-1,i} - \mathbf{z}_{t,i}) & , \text{ kui } i \in (s_v - s_v^b); \\ \mathbf{z}_{t,i} & , \text{ kui } i \in s_v^b, \end{cases}$$

kus $\mathbf{z}_{t,i}$ on vektor $\mathbf{z}_{t-1,i}$, kuid selle kuu t andmetega, ja δ_i on tõenäosus, et ($i \in (s_v - s_v^b)$)|($i \in s_v$). Selle hinnang on $\hat{\delta}_i = \sum_{i \in (s_v - s_v^b)} w_i^{NA} / \sum_{i \in s_v} w_i^{NA}$, kus w_i^{NA} on mittevastamisele kohandatud objekti i kaal (3). See imputeerimise meetod sobib erinevate perioodide vahelise muutusena defineeritud populatsiooni parameetrite hindamiseks.

Peamine idee on eelmise perioodi puuduvad väärtused imputeerida kasutades selle perioodi väärtuseid, et tekkida vektor \mathbf{z}_{t-1} uutele valimi objektidele. See on "lubatud", sest liitkirjeldavate tunnuste järjestikuste kuude väärtuste vahel on suur korrelatsioon. Roteeruvale valimi osale (mitmendat korda uuringule vastanud valimi objektidele) täidetakse \mathbf{z}_{t-1} puuduvad väljad võimalikku väärtuse muutust arvesse võttes, et säiliks hinnangute asümptootiline nihketus.

Kuna uuringus on tähtsal kohal nii kuised hinnangud kui ka muutusena defineeritud hinnangud, oleks üheks võimaluseks kasutada mõlemaid eelnevalt defineeritud liitkirjeldavaid tunnuseid. Kui kirjeldavaid tunnuseid on liiga palju, võivad kalibreerimisel tekkida ekstreemsed kaalud (sh negatiivsed). Selle vältimiseks pakuti lõplikuks lahenduseks välja nende kahe imputeerimismeetodi lineaarne kombinatsioon

$$\mathbf{z}_{\cdot i} = (1 - \alpha)\mathbf{z}_{\cdot i}^{(1)} + \alpha\mathbf{z}_{\cdot i}^{(2)},$$

kus α on korrigeerimiskonstant, mille väärtuseks on Kanadas valitud 2/3. Artiklis [11] on kirjeldatud sobiva α väärtuse leidmise meetoodika.

2.2 Puuduvate väärtuste imputeerimine

Valimisse sattunud isikud võivad mõnel kuul uuringule mitte vastata. Vastanute hulgast puudumise põhjuse võib jagada kaheks: vastamata jätmine ja

uuringusse mitte kuulumine.

Kui isik ei vastanud uuringule üldse, jäetakse ta uuringust välja ning tema puudumist arvestavad mittevastamisele korrigeeritud kaalud. Kui isik kuulus mõlemal vaadeldaval kuul valimisse, kuid vastas küsitlusele neist vaid esimesel ($t - 1$), siis moodustatakse omistusklassid ning sellel kuul puuduv andmerida imputeeritakse *Hot-Deck* meetodiga käesoleval kuul vastanute andmete hulgast. Analoogselt imputeeritakse eelmise perioodi andmed isikutele, kes vastasid vaid käesoleval kuul, kuid jätsid eelmisel korral vastamata.

Uuringusse mitte kuulumine võib olla tingitud uuringu populatsiooni mitte kuulumisest näiteks vanuse tõttu: isik oli eelmisel kuul 14-aastane ning ei kuulunud tööjõudu. Selleks kuuks sai ta 15-aastaseks ning seega kaasati ta uuringusse koos ülejäänud leibkonnaga. Sellisel juhul märgitakse tema liitkirjeldavate tunnuste väärtuseks $z_i = 0$, sest ta ei kuulunud eelmisel kuul uuritavasse populatsiooni. Kui inimene vastas eelmisel kuul, aga ei kuulu sellel kuul uuritavasse populatsiooni (näiteks sai 75-aastaseks), jäetakse ta uuringust välja. Idee seisneb z_i määramisel nii, et $\sum_{i \in s_v} w_i^{NA} z_i$ ja \hat{t}_z jäävad eelmise kuu kontrollsummade t_z hinnanguteks.

Teine uuringusse mitte kuulumise põhjus võib olla valimisse sattunud leibkonda mittekuulemine. Seda võib esineda mõlemat pidi: isik oli eelmisel kuul valimisse sattunud leibkonnas, aga on selleks kuuks vaadeldavast leibkonnast lahkunud, või on nüüd liitunud valimisse sattunud leibkonnaga. Mõlemal juhul, eeldusel, et ta kuulus puuduval kuul uuritavasse populatsiooni, imputeeritakse talle puuduva kuu väärtused *Hot-Deck* meetodiga.

2.3 Kaalude arvutamine

Modifitseeritud regressioonhinnangu kaalud w_i^{CC} leitakse lineaarse meetodi kaugusfunktsiooni (5) minimiseerimisel tingimusel

$$\sum_{i \in s_v} w_i^{CC} \begin{pmatrix} \mathbf{x}_i \\ z_i \end{pmatrix} = \begin{pmatrix} t_x \\ \hat{t}_z \end{pmatrix},$$

kus \mathbf{x} on kalibreerimisel kasutatavad kirjeldavad tunnused.

Paneme tähele, et modifitseeritud regressioonhinnangu kaalud w_i^{CC} on ikka kalibreeritud tavapäraseid kontrollsummasid t_x kasutades. Need avalduvad valemiga

$$w_i^{CC} = w_i^{NA} v_i^{CC}, \quad (10)$$

kus

$$v_i^{CC} = (\mathbf{x}_i^T, \mathbf{z}_{.i}^T) \left(\sum_{i \in s_v} w_i^{NA} (\mathbf{x}_i^T, \mathbf{z}_{.i}^T)^T (\mathbf{x}_i^T, \mathbf{z}_{.i}^T) \right)^{-1} (\mathbf{t}_x^T, \hat{\mathbf{t}}^T)$$

on modifitseeritud regressioon hinnangu korrigeerimistegur, kasutades lineaarset meetodit.

Mõnikord võivad lõplikud kaalud tulla negatiivsed. Sellisel juhul teostatakse modifitseeritud regressioon uuesti, kuid mittevastamisele kohandatud kaalud omandavad väärtused:

$$w_i^{NA} = \begin{cases} w_i^{cc} & , \quad \text{kui } w_i^{cc} \geq 0; \\ w_i^{NA} & , \quad \text{muidu.} \end{cases} \quad (11)$$

Kui ka pärast teist vooru esineb jätkuvalt negatiivseid kaale, määratakse negatiivsete kaaludega isikute uuteks kaaludeks 1 ning lepatakse, et modifitseeritud regressiooni tingimused ei ole rahuldatud. Seda esineb praktikas väga harva.

2.4 Parameetrite hindamine

Kanada tööjõu-uuringus hinnatakse samu parameetreid kui Eesti tööjõu-uuringus. Kasutatavad valemid on välja toodud peatükis 1.4.1, kuid kalibreeritud kaalude w_i^c asemel kasutatakse modifitseeritud regressiooniga leitud kaale (10). Kanada valikudisain on üpriski keeruline ning seetõttu kasutatakse standardhälbe hindamiseks *Jackknife* meetodit.

Modifitseeritud regressioon hinnangu kasutamise eesmärgiks on parandada hinnangute efektiivsust (vähendada dispersiooni). Liitkirjeldava tunnuse \mathbf{z}_t hinnangu efektiivsus suureneb märkimisväärselt, kui \mathbf{z}_t ja \mathbf{z}_{t-1} vahel on suur korrelatsioon. Sellele aitab kaasa valimi suur rotatsioon. Efektiivsuse paranemisele aitab kaasa valimist lahkunud objektide eelmise kuu informatsiooni kasutamine, kuid seda vähendab uute valimi objektide eelmise perioodi vastuste puudumine ning mittevastanud isikutele puuduvate väärtuste imputeerimine.

Artiklis [13] tuuakse välja, et regressioon hinnang (kalibreerimine lineaarse meetodiga) ja modifitseeritud regressioon hinnang peaksid eeldatavalt olema mõlemad nihketa hinnangud kuistele hinnangutele ning seetõttu peaksid nende hinnangud ajas lõikuma (ühel kuul on üks hinnang suurem, teisel kuul teine). Praktikas on osutunud modifitseeritud regressioon hinnangud tööjõus hõivatute arvule väiksemateks kui regressioon hinnangud. Selle põhjuseks peetakse valikuviga, sest roteeruva valimi vastamismäär on olnud suurem kui uutel valimi objektidel. Samuti jätavad uute valimi objektide seas suurema tõenäosusega vastamata väiksemad leibkonnand ning tööjõus hõivatud inimesed.

3 Rakendus Eesti tööjõu-uuringu andmetele

Käesoleva peatükki eesmärgiks on rakendada eelnevalt välja toodud meetodid ETU andmetele ning võrrelda saadud hinnanguid ja hinnangute täpsust. Võrdluses tuuakse välja ühe kuu andmetelt arvatud, libiseva kvartali ja modifitseeritud regressiooni hinnangud ning nende standardhälbed. Analüüs on teostatud statistika paketi R. Kasutatud kood on lisa 3.

3.1 Andmestiku kirjeldus ja kontroll

Töö aluseks on 2012.-2016. aasta Eesti Tööjõu-uuringu andmed, mida konfidentsiaalsuskohustus lepingu tõttu ei ole võimalik tööle lisada. Andmestik koosneb 348958 andmerekast ja 31 tunnusest. Käesolevas töös on nende seast kasutatud järgmisi tunnuseid:

- LIIKMENR - on esitatud kujul $leibkonnainumber \times 10000 + liikmenumber$;
- KORDSUS - mitmendat korda on leibkond uuringusse kaasatud;
- MONTH - uuringu kuu number (1-12);
- PERIOOD - periood, mille kohta hinnangut arvutatakse kujul *aasta_kuu*;
- SUGU - isiku sugu: 1-mees, 2-naine;
- VANUSGR - 5-aastane vanusgrupp: 15-19;20-24;...;65-69;70-74. Kokku 12 gruppi;
- LEIB - igas leibkonnas on määratud üks isik vastutavaks: 1-vastutav, NA-ei ole vastutav;
- LIIKMETE - tööealiste leibkonnaliikmete arv;
- KADU - kaotunnus: 1-vastas, 2-kadu;
- LA_MAAK - leibkonna elukoha maakond (kodeeritud);
- LA_MLINN - leibkonna elukoha linnasus: 1-linn, 2-maa;
- IA_MAAK - isiku elukoha maakond (kodeeritud);
- IA_MLINN - isiku elukoha linnasus: 1-linn, 2-maa;
- RAHVUS - isiku rahvus: 1-eestlane, 2-ei ole eestlane;
- GRUPP1 - valimikiht: 1-Tallinn, 2-suured maakonnad, 3-väikesed maakonnad, 4-Hiiumaa;

- RAHVAARV - rahvaarv valimikihis vastava uuringuaasta 1. jaanuaril;
- DKAAL - leibkonna disainikaal;
- MVKAAL - mittevastamisega korrigeeritud disainikaal;
- KAAL - kalibreeritud kaal;
- Staatus - isiku sotsiaalmajanduslik staatus: 1-hõivatud, 2-töötu, 3-mitteaktiivne;
- EMTAK - ettevõtte/asutuse üksuse tegevusala kood: A...U,X;
- SEKTOR - ettevõtte/asutuse üksuse majandussektor: 1-primaarsektor, 2-sekundaarsektor, 3-tertsiaarsektor;
- OMAN_LIIK - ettevõtte/asutuse omanik: 1-avalik (riik või kohalik omavalitsus), 2-eraisik.

Kuna kaalude arvutamise meetoodika on vaadeldava perioodi jooksul muutunud, otsustati arvutada kaalud uuesti uuema, 2015. aastast rakendatava meetoodikaga. Selleks oli vaja teada kõikide valimiobjektide vanuserühmi. Andmete kontrolli käigus selgus, et varasematel aastatel küsitlusele mittevastanud isikutele vanuserühma ei määratud ning seetõttu ei ole uut meetoodikat võimalik 2012.-2014. aasta andmetele rakendada ning need andmed jäeti edaspidisest analüüsist välja.

Libiseva kvartali hinnangu arvutamiseks kasutatakse kolme järjestikuse kuu andmeid ning seetõttu on sama andmerida andmestikus esindatud kuni kolmel korral. Näiteks on kõik 2016. aasta jaanuari kirjed esimest korda perioodiga '2015_12' (detsembri libiseva kvartali hinnangu arvutamiseks), teist korda perioodiga '2016_01' ning kolmandat korda '2016_02'. Erandiks on 2015. aasta jaanuari ning 2016. aasta detsembri andmed, mis esinevad 2 korda, ning 2014. aasta detsembri ja 2017. aasta jaanuari andmed, mis esinevad andmestikus ühel korral.

Andmete kontrollimisel selgus, et 2015. aasta märtsis on jäänud andmestikust välja kaks veebruaris valimisse sattunud, kuid vastamata jätnud isikut. Valimimahud aastatel 2015-2016 on välja toodud tabelis 1. Sarnast ebakõla esines ka 2014. aasta andmetes, kuid valimimahtude erinevused olid märksa suuremad. Ka see oli mõjuvaks argumendiks vanemate andmete analüüsist välja jätmisel.

Eesti tööjõu-uuringus logistilise regressiooniga kao kompenseerimisel kasutatakse ühe tunnusega vastamise kordsust. Aastatel 2015-2016 on 897 korral märgitud isiku kahele järjestikusele vastamise korrale sama vastamise kordsus.

See segadus puudutab oktoober 2015 ja jaanuar 2016 ning jaanuar 2016 ja aprill 2016 kirjeid.

Kõikidest leitud vigadest on Statistikaametit teavitatud. Nende koodis ühtegi märget kahe isiku analüüsist välja jätmise põhjendamiseks ei olnud. Kuna kuu- de löikes on disainikaalude summad võrdsed, ei peetud nende andmeridade välja jäämist probleemiks. Vastamise kordsuse kordumine osutus linkimise veaks. Seetõttu parandati töös kasutatavas andmestikus kõnealustel ridadel tunnuse KORDSUS väärtused.

Tabel 1: Loendiveata valimimaht

	Leibkondade arv		Isikute arv	
	1 kuu	3 kuud	1 kuu	3 kuud
2015				
Jaanuar	1236	3164	2276	5701
Veebruar	982	3180	1743	5737
Märts	962	3211	1718	5728
Aprill	1267	3220	2269	5700
Mai	991	3253	1713	5738
Juuni	995	3241	1756	5651
Juuli	1255	3216	2182	5645
August	966	3210	1707	5635
September	989	3113	1746	5504
Oktoober	1158	3049	2051	5415
November	902	3196	1618	5635
Detsember	1136	3084	1966	5458
2016				
Jaanuar	1046	3195	1874	5644
Veebruar	1013	3316	1804	5821
Märts	1257	3281	2143	5753
Aprill	1011	3284	1806	5781
Mai	1016	3311	1832	5828
Juuni	1284	3314	2190	5783
Juuli	1014	3296	1761	5782
August	998	3270	1831	5797
September	1258	3288	2205	5847
Oktoober	1032	3304	1811	5823
November	1014	3288	1807	5831
Detsember	1242	3428	2213	6015

3.2 Kaalude arvutamine

ETU meetodikast parema ülevaate saamiseks saadeti töö autorile ETU kaalude arvutamise koodi lõik, kust käesoleva töö teostamiseks vajalikud täpsustused üle kontrollida.

Libiseva kvartali hinnangu terooria kohaselt arvutatakse iga kuu andmete kaalud ning leitakse uuritavate parameetrite hinnangud. Seejärel leitakse kolme järjestikuse kuu hinnangute keskmine. Praktikas lahendatakse Eesti tööjõu-uuringus see etapp teisit. Hinnangu leidmiseks võetakse kohe kolme järjestikuse kuu andmed ning arvutatakse neile koos kaalud. Valimimahuks loetakse kolme kuu loendivigadeta valimimahtude summat (tabel 1). Kuiste hinnangute ja modifitseeritud regressioonhinnangute jaoks arvutati käesolevas töös uued di-sainikaalud ühe kuu andmete põhjal valemiga (2).

Mittevastamisele korrigeeritud kaalud arvutati valemiga (3). Korrigeerimis-faktorite arvutamiseks kasutati R'i paketi *survey* protseduuri *svyglm*. Kahjuks ei tulnud töö autori poolt hinnatud mittevastamisele korrigeeritud kaalud samad kui andmestikus. Põhjalikumal andmete kontrollimisel selgus, et kui leibkonnas ei vastanud ükski leibkonnaliige, on andmestikus ainult ühe leibkon-naliikme andmerida, olgugi, et leibkonnaliikmete arv on suurem. Seetõttu on logistilise regressiooniga hinnatud vastamistöenäosus ülehinnatud. Kuna täiendavaid andmeid mittevastanud leibkondadele ei olnud võimalik saada, tuleb selle veaga leppida. Mittevastamisele korrigeeritud kaalud arvutati uuesti nii kolme järjestikuse kuu kui ka ainult vaadeldava kuu andmeid kasutades.

Kalibreerimine viiakse ETUs läbi leibkonna tasandil, seega on kõigil leibkon-naliikmetel võrdne lõplik kaal. Leibkondadest koosneva andmestiku saamiseks luuakse indikaatortunnused soo- ja vanusgruppidele, maakondadele, linnastu-mise astmele ja rahvusele. Seejärel summeeritakse indikaatortunnused leib-konna siseselt ning lisatakse leibkonna keskmine mittevastamisele korrigeeritud kaal. Sellises andmestikus näitab indikaatortunnus, mitu leibkonna lii-get konkreetsesse gruppi kuulub. Kalibreerimisel kasutatavad kontrollsummad arvutati andmestikus ETUs kasutatavaid kalibreeritud kaale summeerides.

Eesti tööjõu-uuringus viiakse kalibreerimine läbi statistikapaketi SAS protse-duuriga *CALMAR* kasutades logit meetodit. Käesolevas töös kasutati tarkvara R paketi *icarus* protseduuri *calibration*. G-kaalude vahemik määrati iga kuu jaoks eraldi katse ja ekstituse meetodil, et see oleks võimalikult kitsas. Kalib-reerides kolme kuu andmeid, oli suurimaks g-kaalude vahemikuks (0, 8; 1, 8). Meetodit korrati ka ühe kuu andmeid kasutades nii logit kui ka lineaarse mee-todiga. Logit meetodi puhul oli suurimaks g-kaalude vahemikuks (0, 7; 2, 1). Lineaarse meetodi puhul esines ka negatiivseid kaale. Sellisel juhul korrati ka-

libreerimise protsessi, kus mittevastamisele korrigeeritud kaalude asemel kasutati kaalude kombinatsiooni (11). Teist korda kalibreerides enam negatiivseid kaale ei tekkinud.

Modifitseeritud regressiooni puhul kasutatakse kalibreerimisel lisaks kirjeldavatele tunnustele ka liitkirjeldavaid tunnuseid, mis on defineeritud nii selle kuu kui ka varasemate vastamiskordade vastuseid kasutades. Kanadas kasutatakse 25 liitkirjeldavat tunnust eelmisel kuul töötamise kohta (Lisa 1). Kuna käesoleva töö eesmärgiks on uue meetodika tutvustamine, ei uuritud eraldi missugused liitkirjeldavad tunnused annaksid parima hinnangu ning töös kasutatavad tunnused valiti Kanadas kasutatavatele tunnustele võimalikult sarnased. Sellest lähtuvalt defineeriti liitkirjeldavate tunnustena sotsiaalmajanduslik staatus ja töötamine soo- ja vanusgruppide kaupa. Tegevusalade defineerimiseks kaaluti Eesti majanduse tegevusalade klassifikaatorit 2008. Kuna osades valdkondades töötajaid oli kuiste andmete valimis väga vähe ning mõne puhul ei olnud ühtegi, ei peetud käesolevas töös mõistlikuks nende pealt leitud hinnanguid abiinformatsioonina kasutada. Selle tunnuse asemel kasutati töötavate isikute töökoha ettevõtte/asutuse üksuse majandussektorit (tunnus SEKTOR). Avalikus- või erasektoris töötamisel ei olnud võimalik eristada erasektori töötajaid ja ettevõtjaid ning seetõttu kasutati andmestiku tunnust OMAN_LIIK. Kasutatud tunnused on täpsemalt lahti kirjutatud Lisas 2.

Kuna Eestis kasutatakse 2-2-2 rotatsiooni plaani, oli inimese eelmine vastamise kord kas kolm kuud tagasi, kui inimene vastab teist või neljandat korda, või 9 kuud tagasi, kui vastab kolmandat korda. Seetõttu defineeriti ETU andmeid kasutades liitkirjeldavad tunnused 3 kuud tagasi töötamise kohta ning ka liitkontrollsummad on vastava kuu andmetelt leitud hinnangud. Selleks, et kolme kuu tagused hinnangud oleksid teada, sai modifitseeritud regressiooni rakendada alates aprill 2015 andmetele. Esimesel kolmel kuul hinnati liitkontrollsummad valemiga (6), kus kaalud leiti kalibreerides lineaarse meetodiga. Hilisematel andmetel kasutati liitkontrollsummadena modifitseeritud regressioonihinnanguid. Korrigeerimiskonstandiks valiti $\alpha = \frac{2}{3}$.

Kuna küsitlusele mittevastanud isikute kohta ei olnud töö autoril võimalik kontrollida vastamata jätmise põhjust, otsustati mittevastanud isikutele puuduvaid väärtuseid *Hot-Deck* meetodil mitteimputeerita. Kõik uuritavaid kuul vastamata jätnud isikud loeti valimi kaoks. Isikuid, kes kolm kuud tagasi ei vastanud, aga sellel kuul vastasid, koheldi uute valimi objektidena. Mõnel kuul esines ka inimesi, kes vastasid kolmandat korda mitte kolm kuud, vaid neli kuud enne neljandat korda vastamist. Ka selliseid isikuid loeti käesolevas töös uuteks valimi objektideks.

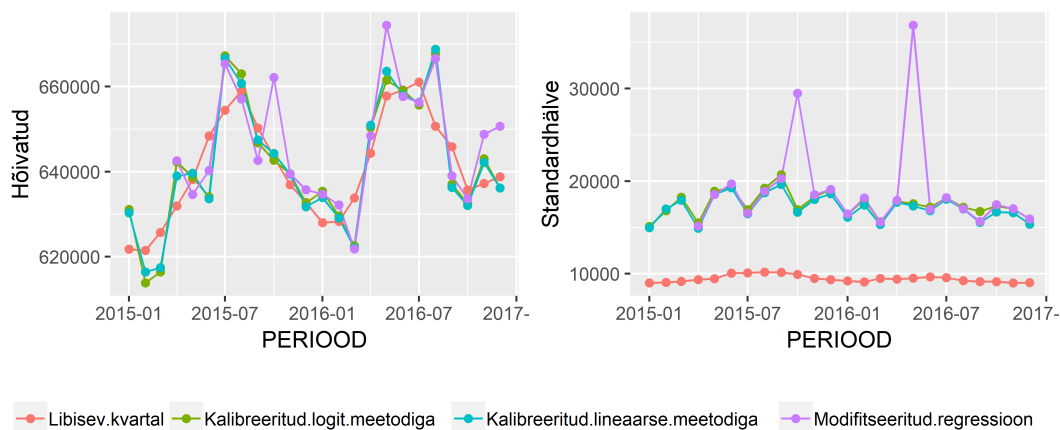
Pärast vajalike tunnuste defineerimist moodustati analoogselt kalibreerimisele

leibkonna põhine andmestik. Selleks liideti leibkonna liikmete kirjeldavate ja liitkirjeldavate indikaatoritunnuste väärtused ning lisati leibkonna keskmine mittevastamisele korrigeeritud kaal. Seejärel arvutati modifitseeritud regressioon hinnangu kaalud (10). Isegi pärast mittevastamisele korrigeeritud kaalude kohandamist valemiga (11) esines enamuse kuudel negatiivseid kaale ehk modifitseeritud regressiooni tingimused ei olnud rahuldavad. Artiklis [13] on välja toodud, et ekstreemsed kaalud (seal hulgas negatiivsed kaalud) võivad tekkida, kui kirjeldavaid ja liitkirjeldavaid tunnuseid on liiga palju.

Seetõttu vähendati liitkirjeldavate tunnuste arvu ning kasutati vaid sotsiaalmajanduslikku staatus ja töötamist soo- ja vanusgruppide kaupa. Ka siis esines esimese kalibreerimise järel osadel kuudel negatiivseid kaale. Nendel kuudel muudeti mittevastamisele korrigeeritud kaale valemiga (11) ning arvutati kaalud uuesti. Sellegi poolest esines oktoober 2015 ning mai 2016 negatiivseid lõplikke kaale. Sellisel juhul asendati negatiivsed kaalud ühtedega ning lepiti, et modifitseeritud regressiooni tingimused ei ole täidetud.

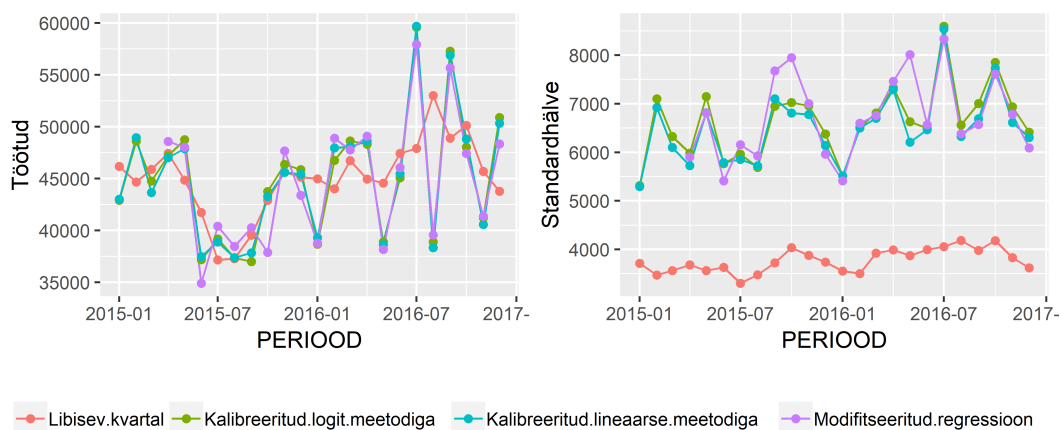
3.3 Kuised hinnangud

Viimaks leiti kuised hinnangud tööjõus hõivatute, töötute ning mitteaktiivsete tööealiste Eesti alaliste elanike arvudele valemiga (6) kasutades leitud kaale. Hinnangute standardhälve hinnati valemiga (9). Tööjõus hõivatud inimeste arvu libiseva kvartali, ühe kuu andmetelt logit meetodiga kalibreeritud kaaludega, ühe kuu andmetelt lineaarse meetodiga kalibreeritud kaaludega ning modifitseeritud regressiooni kaaludega arvutatud hinnangud ja hinnangute standardhälvete hinnangud on välja toodud joonisel 1.



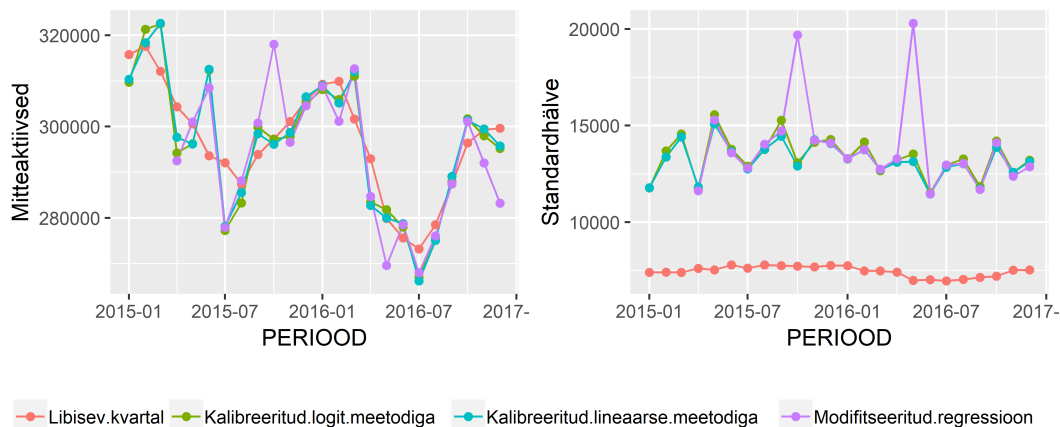
Joonis 1. Tööjõus hõivatud inimeste arvu hinnangud

Jooniselt 1 on näha, et nendel kuudel, mil modifitseeritud regressiooni eeldused ei olnud täidetud (ehk esines negatiivseid kaale), on standardhälbe hinnangud märgatavalt suuremad kui ülejäänud kuudel. Ka neil kuudel, mil eeldused olid täidetud, ei parandanud töös kasutatud liitkirjeldavate tunnuste lisamine hõivatute arvu hinnangu täpsust.



Joonis 2. Töötute inimeste arvu hinnangud

Joonisel 2 on välja toodud töötute arvu hinnangud ning hinnangute standardhälbed. Jooniselt on näha, et töötute arvu hinnangutes oli suurim standardhälve hoopis kuul, mil modifitseeritud regressiooni eeldused olid täidetud (juuli 2016). Sellel kuul olid ka teiste ühe kuu andmeid kasutavate meetodite hinnangute standardhälbed ligikaudu sama suured.



Joonis 3. Töötajates mitteaktiivsete inimeste arvu hinnangud

Töötajates mitteaktiivsete inimeste arvu hinnangud ning standardhälvete hinnangud on esitatud joonisel 3. Sarnaselt hõivatute arvu hinnanguga eristuvad

selgelt kuudel, mil modifitseeritud regressiooni eeldused täidetud ei olnud. Ülejäänud kuudel olid kõik ühe kuu andmete põhjal arvatatud standardhälvete hinnangud ligikaudu võrdsed. Huvitav on täheldada, et kuigi 2016. aasta detsembri ühe kuu andmetelt arvatatud hõivatute ning töötute mitteaktiivsete inimeste hinnangute standardhälvete hinnangud olid sarnased, erinesid modifitseeritud regressioonhinnangu väärtus märgatavalt teistest hinnangutest.

Graafikutel joonistub välja ka hinnangute täpsuse sõltuvus valimi mahust. Kuudel, mil valimi maht on suurem, on hinnangute standardhälvete hinnangud väiksemad. Eesti töötute-uuringus kasutatavate valimimahtude erinevused kuude lõikes on tingitud aastase valimi jagamisest võrdselt 52 nädala peale. Kuine valimi maht sõltub uuringu nädalate jaotamisest - kas vaadeldaval kuul arvestatakse 4 või 5 nädala andmeid. Libiseva kvartali hinnangu standardhälvete jaotus on ühtlasem, sest kolme järjestikuse kuu valimimahtude summad varieeruvad vähem kui kuised valimimahud.

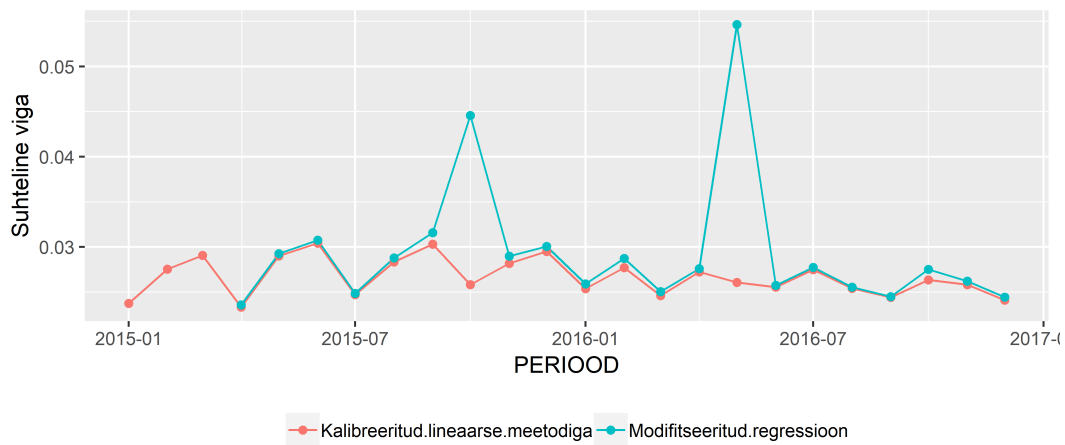
Töös kasutatud andmestik ei joonistu välja süstemaatilist erinevust regressioonhinnangu ja modifitseeritud regressioonhinnangu vahel. See võib tuleneda kasutatud mittevastamisele korrigeeritud kaalude arvutamise meetodist, kus arvestati vastamise kordsust ning leibkonna suurust.

Uuringute üheks kvaliteedinäitajaks on hinnangute suhteline viga. See avaldub valemiga

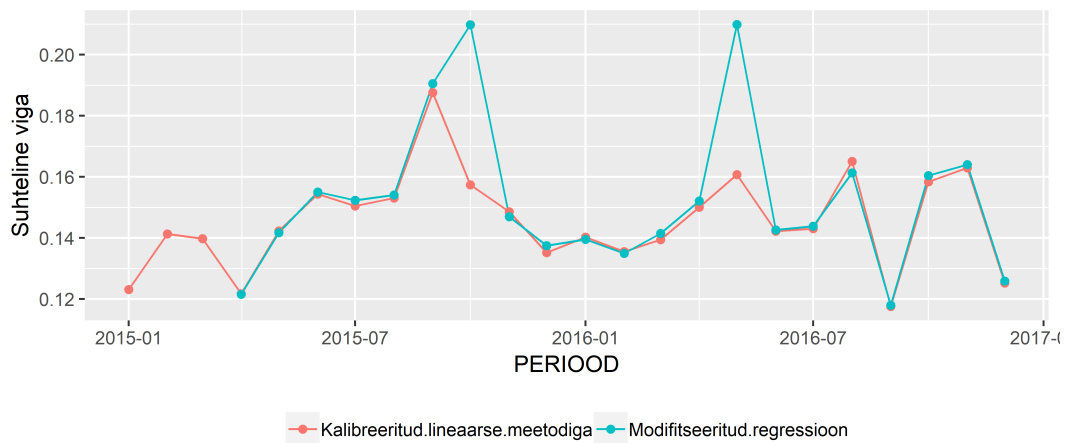
$$CV(\hat{\theta}) = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}},$$

kus $\hat{\theta}$ on uuritava parameetri hinnang ja $\sqrt{\hat{V}(\hat{\theta})}$ on uuritava parameetri hinnangu standardhälve. [6] Käesolevas töös on uuritavaks parameetriks kogusumma hinnang (6). Joonistel 4, 5 ja 6 on välja toodud regressioonhinnangu ja modifitseeritud regressioonhinnangu hõivatute, töötute ning mitteaktiivsete inimeste arvu hinnangute suhtelised vead.

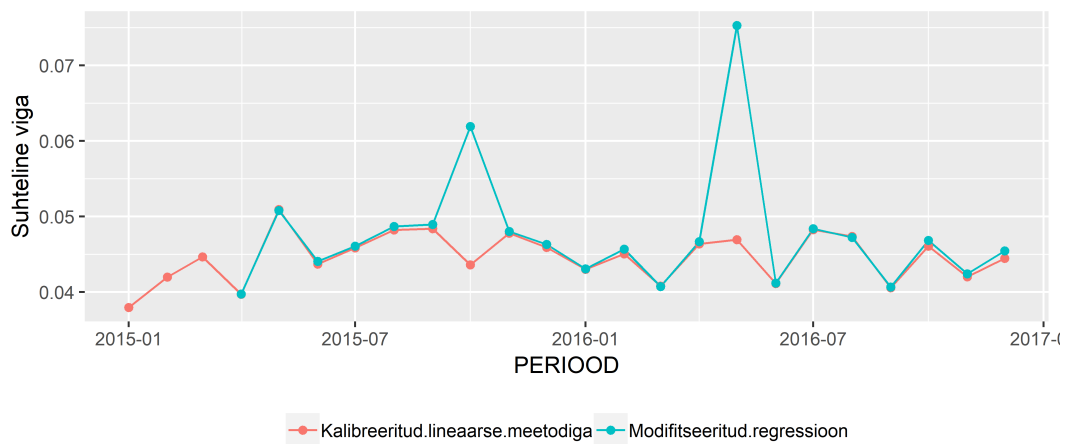
Enamus kuudel on regressioonhinnangute suhtelised vead pisut väiksemad kui modifitseeritud regressioonhinnangute suhtelised vead. See võib olla tingitud kasutatud α suurusest $\frac{2}{3}$, mis kasutab suurema kaaluga muutusena defineeritud suuruste hindamiseks sobivat imputeerimismeetodit.



Joonis 4. Töäjõus hõivatud inimeste arvu hinnangute suhtelised vead



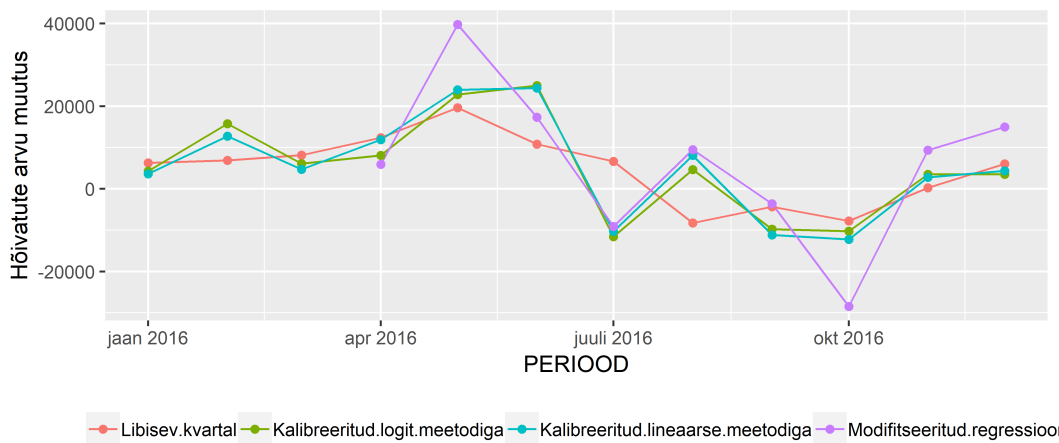
Joonis 5. Töötute arvu hinnangute suhtelised vead



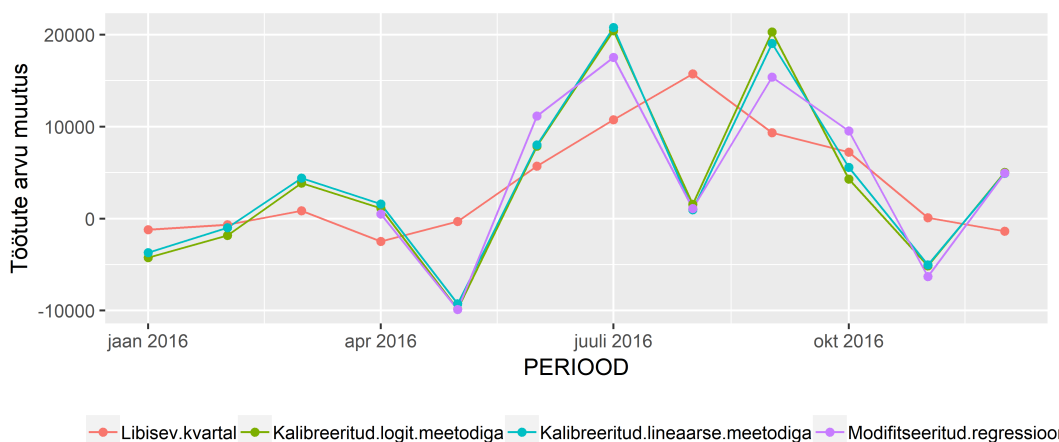
Joonis 6. Töäjõus mitteaktiivsete inimeste arvu hinnangute suhtelised vead
 Töäjõu-uuringutes hinnatakse ka aastast töäjõus hõivatud inimeste arvu muutust. Selleks lahutatakse uuritava kuu hinnangust eelmise aasta sama kuu

hinnangu väärtus. [13] Analoogselt saab vaadelda ka töötute ning tööjõus mitteaktiivsete inimeste arvu muutust. Hõivatute, töötute ja mitteaktiivsete inimeste arvu muutuste hinnangud on esitatud vastavalt joonistel 7, 8 ja 9.

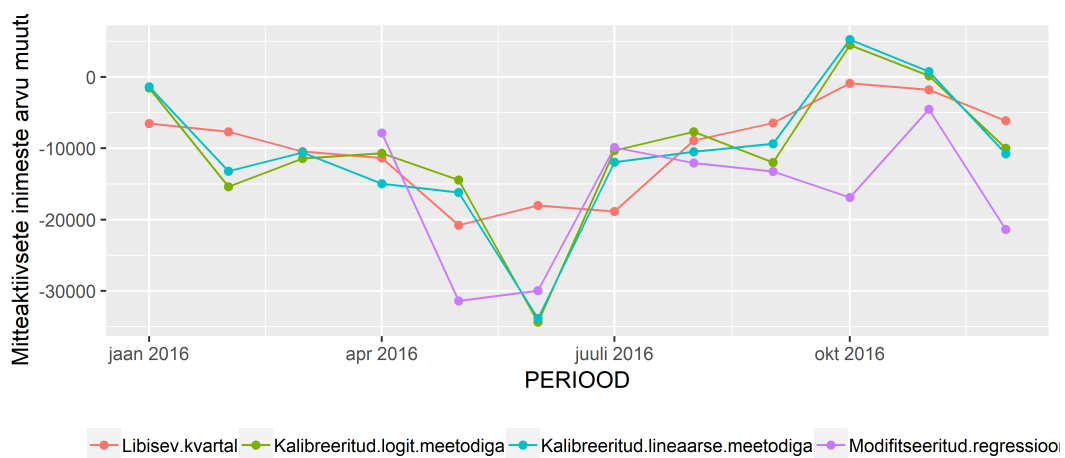
Kuna modifitseeritud regressioonhinnangut sai arvutada alates aprillist 2015 ning töös on kasutatud kahe aasta andmeid, on muutusena defineeritud hinnangud leitavad vaid aprill 2016 kuni detsember 2016 jaoks. Oktoobris 2015 ja mais 2016 ei olnud modifitseeritud regressiooni tingimused rahuldatud ning see mõjutab 2016. aasta mai ja oktoobri aastaseid hinnangute muutuseid. Ülejäänud seitsme kuu hinnangute põhjal on raske öelda, et peale libiseva kvartali hinnangu oleks ühe või teiste meetodi hinnangute jaotus ühtlasem kui teised.



Joonis 7. 2016. aasta tööjõus hõivatud inimeste arvu hinnangute muutus võrreldes 2015. aasta sama kuuga



Joonis 8. 2016. aasta töötute arvu hinnangute muutus võrreldes 2015. aasta sama kuuga



Joonis 9. 2016. aasta töäjõus mitteaktiivsete inimeste arvu hinnangute muutus võrreldes 2015. aasta sama kuuga

Kokkuvõte

Käesoleva töö eesmärgiks oli tutvustada modifitseeritud regressioonhinnangu teooriat ning rakendada meetodit Eesti tööjõu-uuringu 2012.-2016. aasta andmetele. Modifitseeritud regressioonhinnang on regressioonhinnangu edasiarendus, mis kasutab lisaks vaadeldava kuu andmetele ka eelmiste kuude andmeid hinnangute täpsuse parandamiseks. Eelmiste perioodide andmeid kasetades defineeritakse liitkirjeldavad tunnused, mis parandavad nii kuiseid hinnanguid kui ka perioodide vahelise muutusena defineeritud hinnanguid.

Töö esimeses osas kirjeldati Eesti tööjõu-uuringu meetodikat. Tutvustati kasutatavat valikudisaini ning andmete kogumise etappi. Toodi välja uuringus kasutatavate disainikaalude arvutamise, kao kompenseerimise, kalibreerimise ja imputeerimise meetodikate kirjeldused koos teoreetiliste valemitega. Lõpuks toodi välja tööjõu-uuringus hinnatavad parameetrid ja nende hindamise valemid.

Teises osas toodi ülevaade modifitseeritud regressioonhinnangu teooriast. Toodi välja meetodika loomise põhjused ning eesmärgid. Esitati valemid liitkirjeldavate tunnuste defineerimiseks ja hinnangute arvutamiseks ning kirjeldati Kanada tööjõu-uuringus kasutatavaid puuduvate andmete imputeerimise meetodeid.

Kolmandas osas kirjeldati ning kontrolliti Eesti tööjõu-uuringu 2012.-2016. aasta andmeid. Selgus, et 2015.-2016. aastal Eesti tööjõu-uuringus kasutatavat logistilise regressiooniga mittevastamise kompenseerimise meetodit ei olnud võimalik 2012.-2014. aasta andmetele rakendada ning seetõttu kasutati ainult 2015.-2016. aasta andmeid. Toodi välja andmetes esinenud vead ning parandati linkimisel tekkinud vead.

Töös otsustati esitada nelja kaalude korrigeerimise meetodi võrdlus:

- libiseva kvartali hinnang, kus kaalud on korrigeeritud kalibreerides logit meetodiga;
- ühe kuu andmetelt arvatatud hinnang, kus kaalud on korrigeeritud kalibreerides logit meetodiga;
- ühe kuu andmetelt arvatatud hinnang, kus kaalud on korrigeeritud kalibreerides lineaarse meetodiga;
- ühe kuu andmetelt arvatatud hinnang, kus kaalud on korrigeeritud modifitseeritud regressiooni meetodiga.

Modifitseeritud regressioonhinnangu kaalude arvutamisel esines probleeme seoses negatiivsete lõplike kaalude esinemisega. See on kitsaskoht, mille esinemise võimalikkusega tuleb selle meetodi rakendamisel arvestada.

Kuiste hinnangute näidetena arvutati hinnangud tööjõus hõivatute, töötute ning mitteaktiivsete inimeste arvudele ja hinnangute standardhälvetele ning leiti hinnangute suhtelised vead. Perioodide vahelise muutusena defineeritud parameetritest vaadeldi aastast tööjõus hõivatute, töötute ning mitteaktiivsete inimeste arvu muutust. Leitud hinnangute võrdlused esitati graafikutena. Kõige täpsemad hinnangud ning ühtlasemad hinnangute jaotused andis kolme järjestikuse kuu andmeid kasutav libiseva kvartali hinnang. Töös kasutatud tunnuste puhul kalibreerimisel liitkirjeldavate tunnuste lisamine hinnangute täpsust ei parandanud.

Kasutatud kirjandus

- [1] Tööjõu-uuring. [WWW] <https://www.stat.ee/toojou-uuring> (26.10.2016)

- [2] ESMS metaandmed. [WWW] <https://www.stat.ee/esms-metaandmed?code=40701> (26.10.2016)

- [3] *Methodology of the Canadian Labour Force Survey* [WWW] <http://www5.statcan.gc.ca/olc-cel/olc.action?objId=71-526-X&objType=2&lang=en&limit=0> (21.01.2017)

- [4] Ü. Pettai, E. Lelumees. Eesti tööjõu-uuring, metoodika. Tallinn: Statistikaamet, 2012. [WWW] <https://www.stat.ee/metoodika> (26.10.2016)

- [5] Mõisted. [WWW] http://pub.stat.ee/px-web.2001/database/sotsiaalelu/15Tooturg/02heivatud/04luhiajastatistika/tt_047.htm (25.05.2017)

- [6] N. Lepik, I. Traat. Tõenäosuslik valikuuring I. Tartu: Tartu Ülikool, 2016. [WWW] https://courses.ms.ut.ee/MTMS.01.003/2019_spring/uploads/Main/loengud2016.pdf (30.04.2019)

- [7] E. Grau, F. Potter, S. Williams, N. Diaz-Tena. *Nonresponse Adjustment Using Logistic Regression: To Weight or Not To Weight*. [WWW] https://www.researchgate.net/publication/255588889_Nonresponse_Adjustment_Using_Logistic_Regression_To_Weight_or_Not_To_Weight/download (15.04.2019)

- [8] E. Käärrik. Andmeanalüüs II (MTMS.01.007), loengukonspekt. Tartu: Tartu Ülikool, 2017. [WWW] https://courses.ms.ut.ee/MTMS.01.007/2019_spring/uploads/Main/AA2_Loengukonspekt_2017.pdf (30.04.2019)

- [9] *The calibration software CALMAR - What is it?*. <http://vesselinov.com/CalmarEngDoc.pdf> (30.04.2019)

- [10] SAS'i kasutajatugu. https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#surveymeans_toc.htm (15.04.2019)
- [11] E.J. Chen, T.P. Liu. *Choices of Alpha Value in Regression Composite Estimation for the Canadian Labour Force Survey: Impacts and Evaluation*. [WWW] <http://publications.gc.ca/site/eng/9.841687/publication.html> (06.02.2017)
- [12] Statistika andmebaas. <http://pub.stat.ee/px-web.2001/dialog/statfile2.asp>(10.04.2019)
- [13] J. Gambino, B.Kennedy, M.P. Singh. *Regression composite estimation for the Canadian Labour Force Survey: Evaluation and implementation*. [WWW] <https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X20010015855> (26.10.2016)

Lisad

Lisa 1 Kanada tööjõu-uuringus kasutatavad liitkirjeldavad tunnused

Järgnevalt on välja toodud liitkirjeldavad tunnused, mida kasutatakse Kanada tööjõu-uuringus modifitseeritud regressiooni rakendamisel. Kaldkiri tähistab tunnust, mille väärtust saab teiste tunnuste põhjal tuletada. [3]

Töötamine eelmisel kuul

- Töötas
- Oli töötu
- *Oli mitteaktiivne*

Töötamine eelmisel kuul soo-vanusgruppide kaupa

- Töötav mees, 15-24
- Töötu mees, 15-24
- *Mitteaktiivne mees, 15-24*
- Töötav mees, 25+
- Töötu mees, 25+
- *Mitteaktiivne mees, 25+*
- Töötav naine, 15-24
- Töötu naine, 15-24
- *Mitteaktiivne naine, 15-24*
- *Töötav naine, 25+*
- *Töötu naine, 25+*
- *Mitteaktiivne naine, 25+*

Eelmise kuu töökoha tegevusala

- Põllumajandus
- Ehitus
- Info, kultuur ja vabaaeg
- Kommunaalteenused
- Tootmine

- Maavarad
- Transport ja laondus
- Finants-, kindlustus- ja kinnisvaraalaane tegevus
- Kutse-, teadus- ja tehnikaalane tegevus
- Haldus- ja abitegevused
- Haridus
- Tervishoid ja sotsiaaltoolekanne
- Majutus ja toitlustus
- Avalik haldus
- Kaubandus
- *Muud valdkonnad*

Töötamine eelmisel kuul avalikus- või erasektoris

- Avalikus sektoris töötaja
- Erasektoris töötaja
- *Ettevõtja erasektoris*

Lisa 2 Töös kasutatavad liitkirjeldavad tunnused

Järgnevalt on välja toodud käesolevas töös kasutatavad liitkirjeldavad tunnused. Kaldkiri tähistab tunnust, mille väärtust saab teiste tunnuste põhjal tuletada.

Töötamine kolm kuud tagasi

- Töötas
- Oli töötu
- *Oli mitteaktiivne*

Töötamine kolm kuud tagasi soo-vanusgruppide kaupa

- Töötav mees, 15-24
- Töötu mees, 15-24
- *Mitteaktiivne mees, 15-24*
- Töötav mees, 25+
- Töötu mees, 25+

- *Mitteaktiivne mees, 25+*
- Töötav naine, 15-24
- Töötetu naine, 15-24
- *Mitteaktiivne naine, 15-24*
- *Töötav naine, 25+*
- *Töötetu naine, 25+*
- *Mitteaktiivne naine, 25+*

Kolm kuud tagasi olnud töökoha ettevõtte või asutuse üksuse majandussektor

- Primaarsektor
- Sekundaarsektor
- *Tertsiaarsektor*

Töötamine kolm kuud tagasi avalikus- või erasektoris

- Töötas avalikus sektoris
- *Töötas erasektoris*

Lisa 3 R'i kood

Lisa 3.1 Andmete sisselugemine

```
library(survey)
library(icarus)
orig_andmed=read.table("X:/Projektid/TY_ETU_metoodika/Kaidi_
Jogi/etu_2012_2016.csv",header=TRUE, sep=";")
orig_andmed$LEIBKONNANR=round(orig_andmed$LIIKMENR/10000,
digits=0)
orig_andmed$vanus20=paste(orig_andmed$VANUSGR,
orig_andmed$SUGU)
```

Lisa 3.2 Andmete kontroll

```
#Mis andmetega on tegemist?
summary(orig_andmed)
#Mis perioodidel oli isikutele vanusgrupp puudu
abi=orig_andmed[orig_andmed$VANUSGR==0,]
table(abi$PERIOOD)
abi=abi[abi$PERIOOD=='2014_03',]
rm(abi)
#Vaatan ainult sellel kuul vastanud isikuid
```

```

orig_andmed$abi=ifelse(as.numeric(substr(orig_andmed$PERIOD,
6,7))==orig_andmed$MONTH,1,0)
andmed2=orig_andmed[orig_andmed$abi==1,]
table(orig_andmed$PERIOD)
table(andmed2$PERIOD)
#Jatan alles vaid need andmed, mil maarati vanusgrupp
#koikidele valimiobjektidele
orig_andmed$AASTA=as.numeric(substring(
orig_andmed$PERIOD,1,4))
andmed=orig_andmed[orig_andmed$AASTA>2014,]
summary(andmed)
#Vaatan ainult sellel kuul vastanud isikuid
andmed2=andmed[andmed$abi==1,]
#Mitme isiku andmeid kuude loikes arvestatakse
table(andmed$PERIOD)
table(andmed2$PERIOD)
#Missugused vaartused marts arvestuses puudusid?
marts=andmed[andmed$PERIOD=='2015_03',]
marts2=andmed2[andmed2$PERIOD %in% c('2015_02','2015_03',
'2015_04'),]
abi=marts2[!marts2$LIIKMENR %in% marts$LIIKMENR,]
abi=orig_andmed[orig_andmed$LIIKMENR %in% abi$LIIKMENR,]
rm(abi,marts,marts2)
#Mitme leibkonna andmeid kasutati
table(andmed$PERIOD, andmed$LEIB)
table(andmed2$PERIOD, andmed2$LEIB)
#Kui palju on korduvalt vastavaid isikuid
table(andmed$PERIOD, andmed$KORDSUS)
table(andmed2$PERIOD, andmed2$KORDSUS)
andmed2$abi=paste(andmed2$LIIKMENR, andmed2$KORDSUS)
abi=andmed2[duplicated(andmed2$abi),]
abi2=andmed[andmed$LIIKMENR %in% abi$LIIKMENR,]
table(abi$KORDSUS)
#Muudan dubleeritud kirje kordsuse 1 vorra suuremaks
table(andmed$KORDSUS)
table(andmed2$KORDSUS)
andmed=andmed[order(andmed$PERIOD),]
andmed2=andmed2[order(andmed2$PERIOD),]
andmed$KORDSUS=ifelse(paste(andmed$LIIKMENR, andmed$PERIOD)
%in% paste(abi$LIIKMENR, abi$PERIOD), andmed$KORDSUS+1,
andmed$KORDSUS)
andmed2$KORDSUS=ifelse(paste(andmed2$LIIKMENR,
andmed2$PERIOD) %in% paste(abi$LIIKMENR, abi$PERIOD),
andmed2$KORDSUS+1, andmed2$KORDSUS)
rm(abi, abi2)
andmed2$abi=NULL
table(andmed$KORDSUS)

```

```

table(andmed2$KORDSUS)
#Soo jaotus
table(andmed$PERIOD, andmed$SUGU)
table(andmed2$PERIOD, andmed2$SUGU)
#Vanusjaotus
table(andmed$PERIOD, andmed$VANUSGR)
table(andmed2$PERIOD, andmed2$VANUSGR)
#Leibkonnaliikmete arv
leibkonnad=andmed[!is.na(andmed$LEIB),]
table(leibkonnad$PERIOD)
table(leibkonnad$PERIOD, leibkonnad$LIIKMETE)
leibkonnad2=andmed2[!is.na(andmed2$LEIB),]
table(leibkonnad2$PERIOD)
table(leibkonnad2$UN, leibkonnad2$AASTA)
table(leibkonnad2$PERIOD, leibkonnad2$LIIKMETE)
table(andmed$LIIKMETE, andmed$KADU)
#Kadu
table(andmed$PERIOD, andmed$KADU)
table(andmed2$PERIOD, andmed2$KADU)
#Linn ja maakond
table(andmed$LA_MAAK, andmed$LA_MLINN, andmed$AASTA)
table(andmed2$LA_MAAK, andmed2$LA_MLINN)
table(andmed$IA_MAAK, andmed$IA_MLINN, andmed$AASTA)
table(andmed2$IA_MAAK, andmed2$IA_MLINN)
#Kas isiku elukoht ja leibkonna elukoht on samad
andmed$abi=ifelse(is.na(andmed$IA_MAAK), 0,
ifelse(andmed$IA_MAAK==andmed$LA_MAAK, 0, 1))
sum(andmed$abi)
andmed$abi=NULL
#Rahvus
table(andmed$PERIOD, andmed$RAHVUS)
table(andmed2$PERIOD, andmed2$RAHVUS)
#valimijaotus
table(andmed$PERIOD, andmed$GRUPP1)
table(andmed2$PERIOD, andmed2$GRUPP1)
table(leibkonnad$PERIOD, leibkonnad$GRUPP1)
table(leibkonnad2$PERIOD, leibkonnad2$GRUPP1)
#Kontrollin kaalude summasid kuude loikes
aggregate(andmed$DKAAL, by=list(Category=andmed$PERIOD),
FUN=sum)
vastanud=andmed[andmed$KADU==1,]
aggregate(vastanud$MVKAAL, by=list(Category=
vastanud$PERIOD), FUN=sum)
aggregate(vastanud$KAAL, by=list(Category=vastanud$PERIOD),
FUN=sum)
rm(vastanud)
#Leian aastate kaupa kihtide rahvaarvud

```

```

abi=andmed[!duplicated(paste(andmed$AASTA , andmed$GRUPP1)),]
rm(abi)
#Mitme isiku andmereal on andmestikus mittevastanud
#leibkondadel
abi=andmed2[andmed2$KADU==2,]
abi=andmed2[andmed2$LEIBKONNANR %in% abi$LEIBKONNANR,]
abi$loend=rep(1,13445)
a=aggregate(abi$loend ,by=list(Category=paste(abi$PERIOOD ,
abi$LEIBKONNANR)),FUN=sum)
abi$Category=paste(abi$PERIOOD ,abi$LEIBKONNANR)
abi=merge(abi ,a)
abi=abi[abi$x>1,]
rm(abi ,a)

```

Lisa 3.3 Disainikaalude arvutamine

```

#Leian loendivigadeta valimimahu
a=as.data.frame(table(leibkonnad2$PERIOOD ,
leibkonnad2$GRUPP1))
colnames(a)=c("PERIOOD" ,"GRUPP1" ,"NI_LOENDIVEATA")
andmed2=merge(andmed2 ,a)
rm(a ,leibkonnad ,leibkonnad2)
#Arvutan uue disainikaalu uhe kuu andmeid kasutades
andmed2$DKAAL=andmed2$RAHVAAARV (andmed2$NI_LOENDIVEATA*
andmed2$LIIKMETE)
#Kontrollin , et disainikaalude summa oleks uldkogumimaht
abi=andmed2[!is.na(andmed2$LEIB),]
aggregate(abi$DKAAL*abi$LIIKMETE ,by=list(Category=
abi$PERIOOD),FUN=sum)
abi=andmed[!is.na(andmed$LEIB),]
aggregate(abi$DKAAL*abi$LIIKMETE ,by=list(Category=
abi$PERIOOD),FUN=sum)
rm(abi)

```

Lisa 3.4 Mittevastamisele korrigeeritud kaalude arvutamine

```

#Muudan kodeeritud tunnused faktoriteks
andmed$LA_MAAK=factor(andmed$LA_MAAK)
andmed$LA_MLINN=factor(andmed$LA_MLINN)
andmed$KORDSUS=factor(andmed$KORDSUS)
#Arvutan iga kuu kohta mittevastamisele korrigeeritud kaalu
#logistilise regressiooniga 3 kuu andmeid kasutades
period=andmed$PERIOOD
period=period[!duplicated(period)]
andmed=andmed[order(andmed$PERIOOD),]
mvkaalud=NA
for (i in 1:24) {

```

```

kuu=period[i]
kuu_andmed=andmed[andmed$PERIOD==kuu,]
disain=svydesign(ids=~0,weights=~DKAAL,
data=kuu_andmed)
minuglm=svyglm(I(KADU==1)~LA_MAAK+LA_MLINN+vanus20+KORDSUS
+LIIKMETE,data=kuu_andmed,design=disain,
family="quasibinomial")
kuu_andmed$tn=predict(minuglm,newdata=kuu_andmed,
type="response")
kuu_andmed$MVKAAL2=ifelse(kuu_andmed$KADU==1,
kuu_andmed$DKAAL/kuu_andmed$tn,NA)
mvkaalud=c(mvkaalud,kuu_andmed$MVKAAL2)}
andmed$MVKAAL2=mvkaalud[2:length(mvkaalud)]
rm(disain,minuglm,kuu_andmed,i,kuu,mvkaalud,period)
#Kordan sama uhe kuu andmetel
andmed2$LA_MAAK=factor(andmed2$LA_MAAK)
andmed2$LA_MLINN=factor(andmed2$LA_MLINN)
andmed2$KORDSUS=factor(andmed2$KORDSUS)
period=andmed2$PERIOD
period=period[!duplicated(period)]
mvkaalud=NA
for(i in 1:24){
kuu=period[i]
kuu_andmed=andmed2[andmed2$PERIOD==kuu,]
disain=svydesign(ids=~0,weights=~DKAAL,
data=kuu_andmed)
minuglm=svyglm(I(KADU==1)~LA_MAAK+LA_MLINN+vanus20+KORDSUS
+LIIKMETE,data=kuu_andmed,design=disain,
family="quasibinomial")
kuu_andmed$tn=predict(minuglm,newdata=kuu_andmed,
type="response")
kuu_andmed$MVKAAL2=ifelse(kuu_andmed$KADU==1,
kuu_andmed$DKAAL/kuu_andmed$tn,NA)
mvkaalud=c(mvkaalud,kuu_andmed$MVKAAL2)}
andmed2$MVKAAL2=mvkaalud[2:length(mvkaalud)]
rm(disain,minuglm,kuu_andmed,i,kuu,mvkaalud)

```

Lisa 3.5 Lõplike kaalude ja hinnangute arvutamise

```

#Tekitan kalibreerimiseks vajalikud indikaatoritunnused
fun_ind=function(and){
and$vgr1m=ifelse(and$vanus20=='1_1',1,0)
and$vgr2m=ifelse(and$vanus20=='2_1',1,0)
and$vgr3m=ifelse(and$vanus20=='3_1',1,0)
and$vgr4m=ifelse(and$vanus20=='4_1',1,0)
and$vgr5m=ifelse(and$vanus20=='5_1',1,0)
and$vgr6m=ifelse(and$vanus20=='6_1',1,0)

```

```

and$vgr7m=ifelse (and$vanus20=='7_1',1,0)
and$vgr8m=ifelse (and$vanus20=='8_1',1,0)
and$vgr9m=ifelse (and$vanus20=='9_1',1,0)
and$vgr10m=ifelse (and$vanus20=='10_1',1,0)
and$vgr11m=ifelse (and$vanus20=='11_1',1,0)
and$vgr12m=ifelse (and$vanus20=='12_1',1,0)
and$vgr1n=ifelse (and$vanus20=='1_2',1,0)
and$vgr2n=ifelse (and$vanus20=='2_2',1,0)
and$vgr3n=ifelse (and$vanus20=='3_2',1,0)
and$vgr4n=ifelse (and$vanus20=='4_2',1,0)
and$vgr5n=ifelse (and$vanus20=='5_2',1,0)
and$vgr6n=ifelse (and$vanus20=='6_2',1,0)
and$vgr7n=ifelse (and$vanus20=='7_2',1,0)
and$vgr8n=ifelse (and$vanus20=='8_2',1,0)
and$vgr9n=ifelse (and$vanus20=='9_2',1,0)
and$vgr10n=ifelse (and$vanus20=='10_2',1,0)
and$vgr11n=ifelse (and$vanus20=='11_2',1,0)
and$vgr12n=ifelse (and$vanus20=='12_2',1,0)
and$mk1=ifelse (and$IA_MAAK==1,1,0)
and$mk37=ifelse (and$IA_MAAK==37,1,0)
and$mk39=ifelse (and$IA_MAAK==39,1,0)
and$mk44=ifelse (and$IA_MAAK==44,1,0)
and$mk49=ifelse (and$IA_MAAK==49,1,0)
and$mk51=ifelse (and$IA_MAAK==51,1,0)
and$mk57=ifelse (and$IA_MAAK==57,1,0)
and$mk59=ifelse (and$IA_MAAK==59,1,0)
and$mk65=ifelse (and$IA_MAAK==65,1,0)
and$mk67=ifelse (and$IA_MAAK==67,1,0)
and$mk70=ifelse (and$IA_MAAK==70,1,0)
and$mk74=ifelse (and$IA_MAAK==74,1,0)
and$mk78=ifelse (and$IA_MAAK==78,1,0)
and$mk82=ifelse (and$IA_MAAK==82,1,0)
and$mk84=ifelse (and$IA_MAAK==84,1,0)
and$mk86=ifelse (and$IA_MAAK==86,1,0)
and$m1inn1=ifelse (and$IA_MLINN==1,1,0)
and$m1inn2=ifelse (and$IA_MLINN==2,1,0)
and$rahvus1=ifelse (and$RAHVUS==1,1,0)
and$rahvus2=ifelse (and$RAHVUS==2,1,0)
return (and)}
andmed=fun_ind (andmed)
andmed2=fun_ind (andmed2)
rm (fun_ind)
#Leian kontrollsummad kasutades esialgses andmestikus olnud
#loplike kaale
abi=andmed [andmed$KADU==1,]
#Summeerin kaalud kategooriate kaupa
kontroll1=aggregate (abi$KAAL, by=list (Category=paste(

```

```

abi$PERIOD ,abi$vanus20)),FUN=sum)
#Eraldan aasta arvu
kontroll1$AASTA=substring(kontroll1$Category,1,4)
#Soo-ja vanuse kategooria
kontroll1$Category=substring(kontroll1$Category,9,12)
#kuna aasta igal kuul on vordsed kontrollsummad, jattan
#alles iga aasta kohta uhe kirje
kontroll1=kontroll1[!duplicated(paste(kontroll1$Category,
kontroll1$AASTA)),]
kontroll2=aggregate(abi$KAAL,by=list(Category=paste(
abi$PERIOD,abi$IA_MAAK)),FUN=sum)
kontroll2$AASTA=substring(kontroll2$Category,1,4)
kontroll2$Category=substring(kontroll2$Category,9,10)
kontroll2=kontroll2[!duplicated(paste(kontroll2$Category,
kontroll2$AASTA)),]
kontroll3=aggregate(abi$KAAL,by=list(Category=paste(
abi$PERIOD,abi$RAHVUS)),FUN=sum)
kontroll3$AASTA=substring(kontroll3$Category,1,4)
kontroll3$Category=substring(kontroll3$Category,9,9)
kontroll3=kontroll3[!duplicated(paste(kontroll3$Category,
kontroll3$AASTA)),]
kontroll4=aggregate(abi$KAAL,by=list(Category=paste(
abi$PERIOD,abi$IA_MLINN)),FUN=sum)
kontroll4$AASTA=substring(kontroll4$Category,1,4)
kontroll4$Category=substring(kontroll4$Category,9,9)
kontroll4=kontroll4[!duplicated(paste(kontroll4$Category,
kontroll4$AASTA)),]
rm(abi)
#Loon funktsiooni, mis tekitab leibkonnabaasil andmestiku
fun_leib=function(i,and){
  #Jattan ainult uuritava kuu andmed indikaatortunnuste,
  #MVKAALu ja leibkonnanimbriga
  kuu=period[i]
  kuu_andmed=and[and$PERIOD==kuu,c(35,32,36:79)]
  #Moodustan andmestiku, kus igal leibkonnal on 1 rida,
  #indikaatortunnuste summa ja keskmine mittevastamisele
  #korrigeeritud kaal
  kuu_andmed2=aggregate(.~LEIBKONNANR,kuu_andmed,sum)
  abi=aggregate(MVKAAL2~LEIBKONNANR,kuu_andmed,mean)
  kuu_andmed2$MVKAAL2=abi$MVKAAL2
  kuu_andmed2$leibkuu=paste(kuu_andmed2$LEIBKONNANR,kuu)
  return(kuu_andmed2)}
#Tostan andmed2 veerud umber nii, et need oleks samas
#jarjekorras kui andmestikul andmed
andmed2=andmed2[,c(3:6,1,7:18,2,19:34,36:80)]
andmed$leibkuu=paste(andmed$LEIBKONNANR,andmed$PERIOD)
andmed2$leibkuu=paste(andmed2$LEIBKONNANR,andmed2$PERIOD)

```

```

#Leian kalibreeritud kaalud 3 kuu andmetelt logit meetodiga
aasta=2015
#Tekitan 2015. aasta kontrollsummade maatriksi
margins=c('vgr1m','vgr1n','vgr10m','vgr10n','vgr11m',
'vgr11n','vgr12m','vgr12n','vgr2m','vgr2n','vgr3m','vgr3n',
'vgr4m','vgr4n','vgr5m','vgr5n','vgr6m','vgr6n','vgr7m',
'vgr7n','vgr8m','vgr8n','vgr9m','vgr9n','mk1','mk37','mk39',
'mk44','mk49','mk51','mk57','mk59','mk65','mk67','mk70',
'mk74','mk78','mk82','mk84','mk86','rahvus1','rahvus2',
'mlinn1','mlinn2')
margins=as.data.frame(margins)
margins$a=rep(0,44)
margins$total=c(kontroll1[kontroll1$AASTA==aasta,2],
kontroll2[kontroll2$AASTA==aasta,2],
kontroll3[kontroll3$AASTA==aasta,2],
kontroll4[kontroll4$AASTA==aasta,2])
margins=as.matrix(margins)
#Kalibreerime iga kuu eraldi, sest piirid vajavad igal kuul
#eraldi tapsustamist
leibkond=fun_leib(1, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8,1.5), description=FALSE)
ued_andmed=merge(andmed, leibkond[,c(48,47)])
leibkond=fun_leib(2, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8,1.5), description=FALSE)
ued_andmed2=merge(andmed, leibkond[,c(48,47)])
ued_andmed=rbind(ued_andmed, ued_andmed2)
leibkond=fun_leib(3, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8,1.5), description=FALSE)
ued_andmed2=merge(andmed, leibkond[,c(48,47)])
ued_andmed=rbind(ued_andmed, ued_andmed2)
leibkond=fun_leib(4, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8,1.5), description=FALSE)
ued_andmed2=merge(andmed, leibkond[,c(48,47)])
ued_andmed=rbind(ued_andmed, ued_andmed2)
leibkond=fun_leib(5, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8,1.5), description=FALSE)
ued_andmed2=merge(andmed, leibkond[,c(48,47)])

```

```

uued_andmed=rbind(uued_andmed,uued_andmed2)
leibkond=fun_leib(6,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.6),description=FALSE)
uued_andmed2=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed2)
leibkond=fun_leib(7,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.5),description=FALSE)
uued_andmed2=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed2)
leibkond=fun_leib(8,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.6),description=FALSE)
uued_andmed2=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed2)
leibkond=fun_leib(9,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed2=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed2)
leibkond=fun_leib(10,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.8),description=FALSE)
uued_andmed2=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed2)
leibkond=fun_leib(11,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.7),description=FALSE)
uued_andmed2=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed2)
leibkond=fun_leib(12,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.6),description=FALSE)
uued_andmed2=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed2)
#Kordan eelnevat ka 2016 aasta andmetel
aasta=2016
margins=c('vgr1m','vgr1n','vgr10m','vgr10n','vgr11m',
'vgr11n','vgr12m','vgr12n','vgr2m','vgr2n','vgr3m','vgr3n',

```

```

'vgr4m', 'vgr4n', 'vgr5m', 'vgr5n', 'vgr6m', 'vgr6n', 'vgr7m',
'vgr7n', 'vgr8m', 'vgr8n', 'vgr9m', 'vgr9n', 'mk1', 'mk37', 'mk39',
'mk44', 'mk49', 'mk51', 'mk57', 'mk59', 'mk65', 'mk67', 'mk70',
'mk74', 'mk78', 'mk82', 'mk84', 'mk86', 'rahvus1', 'rahvus2',
'mlinn1', 'mlinn2')
margins=as.data.frame(margins)
margins$a=rep(0,44)
margins$total=c(kontroll1[kontroll1$AASTA==aasta, 2],
kontroll2[kontroll2$AASTA==aasta, 2],
kontroll3[kontroll3$AASTA==aasta, 2],
kontroll4[kontroll4$AASTA==aasta, 2])
margins=as.matrix(margins)
leibkond=fun_leib(13, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8, 1.5), description=FALSE)
ued_andmed3=merge(andmed, leibkond[, c(48, 47)])
ued_andmed=rbind(ued_andmed, ued_andmed3)
leibkond=fun_leib(14, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8, 1.5), description=FALSE)
ued_andmed3=merge(andmed, leibkond[, c(48, 47)])
ued_andmed=rbind(ued_andmed, ued_andmed3)
leibkond=fun_leib(15, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.9, 1.5), description=FALSE)
ued_andmed3=merge(andmed, leibkond[, c(48, 47)])
ued_andmed=rbind(ued_andmed, ued_andmed3)
leibkond=fun_leib(16, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.9, 1.5), description=FALSE)
ued_andmed3=merge(andmed, leibkond[, c(48, 47)])
ued_andmed=rbind(ued_andmed, ued_andmed3)
leibkond=fun_leib(17, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8, 1.6), description=FALSE)
ued_andmed3=merge(andmed, leibkond[, c(48, 47)])
ued_andmed=rbind(ued_andmed, ued_andmed3)
leibkond=fun_leib(18, andmed)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8, 1.7), description=FALSE)
ued_andmed3=merge(andmed, leibkond[, c(48, 47)])

```

```

uued_andmed=rbind(uued_andmed,uued_andmed3)
leibkond=fun_leib(19,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed3=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed3)
leibkond=fun_leib(20,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed3=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed3)
leibkond=fun_leib(21,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed3=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed3)
leibkond=fun_leib(22,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.6),description=FALSE)
uued_andmed3=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed3)
leibkond=fun_leib(23,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.6),description=FALSE)
uued_andmed3=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed3)
leibkond=fun_leib(24,andmed)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.5),description=FALSE)
uued_andmed3=merge(andmed,leibkond[,c(48,47)])
uued_andmed=rbind(uued_andmed,uued_andmed3)
#Leian hinnangud
#Loon hinnangute andmestiku 2015_01 andmetega
kuu=period[1]
abi=uued_andmed[uued_andmed$PERIOOD==kuu,]
abi=abi[abi$KADU==1,]
disain=svydesign(data=abi,ids=~LEIBKONNANR,
strata=~GRUPP1,weights=~uus_kaal,fpc=~RAHVAARV)
hinnangud_libisev=as.data.frame(svytotal(~factor(Status),
disain,data=kuu_andmed,na.rm=TRUE))
hinnangud_libisev$PERIOOD=rep(kuu,3)

```

```

#Kordan sama ulejaanud kuudel ja lisan tulemused eelmisele
#andmestikule
for (i in 2:24) {
  kuu=period[i]
  abi=uued_andmed[uued_andmed$PERIOOD==kuu,]
  abi=abi[abi$KADU==1,]
  disain=svydesign(data=abi,ids=~LEIBKONNANR,
  strata=~GRUPP1,weights=~uus_kaal,fpc=~RAHVAAARV)
  a=as.data.frame(svytotal(~factor(Status),disain,
  data=kuu_andmed,na.rm=TRUE))
  a$PERIOOD=rep(kuu,3)
  hinnangud_libisev=rbind(hinnangud_libisev,a)}
#Ekspordin tulemuse
write.table(hinnangud_libisev,"X:/Projektid/TY_ETU_
metoodika/Kaidi_Jogi/hinnangud_libisev.txt",dec=",")
#Leian kalibreeritud kaalud 1 kuu andmetelt logit meetodiga
aasta=2015
margins=c('vgr1m','vgr1n','vgr10m','vgr10n','vgr11m',
'vgr11n','vgr12m','vgr12n','vgr2m','vgr2n','vgr3m','vgr3n',
'vgr4m','vgr4n','vgr5m','vgr5n','vgr6m','vgr6n','vgr7m',
'vgr7n','vgr8m','vgr8n','vgr9m','vgr9n','mk1','mk37','mk39',
'mk44','mk49','mk51','mk57','mk59','mk65','mk67','mk70',
'mk74','mk78','mk82','mk84','mk86','rahvus1','rahvus2',
'mlinn1','mlinn2')
margins=as.data.frame(margins)
margins$a=rep(0,44)
margins$total=c(kontroll1[kontroll1$AASTA==aasta,2],
kontroll2[kontroll2$AASTA==aasta,2],
kontroll3[kontroll3$AASTA==aasta,2],
kontroll4[kontroll4$AASTA==aasta,2])
margins=as.matrix(margins)
#Kalibreerime iga kuu eraldi, sest piirid vajavad igal kuul
#eraldi tapsustamist
leibkond=fun_leib(1,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed2=merge(andmed2,leibkond[,c(48,47)])
leibkond=fun_leib(2,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(3,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",

```

```

bounds=c(0.7,1.6),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(4,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.5),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(5,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(6,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.9),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(7,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.7),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(8,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.7,2),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(9,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.8),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(10,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.7),description = FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(11,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=

```

```

margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.7, 2.1), description=FALSE)
uued_andmed3=merge(andmed2, leibkond[, c(48, 47)])
uued_andmed2=rbind(uued_andmed2, uued_andmed3)
leibkond=fun_leib(12, andmed2)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.7, 1.9), description=FALSE)
uued_andmed3=merge(andmed2, leibkond[, c(48, 47)])
uued_andmed2=rbind(uued_andmed2, uued_andmed3)
aasta=2016
margins=c('vgr1m', 'vgr1n', 'vgr10m', 'vgr10n', 'vgr11m',
'vgr11n', 'vgr12m', 'vgr12n', 'vgr2m', 'vgr2n', 'vgr3m', 'vgr3n',
'vgr4m', 'vgr4n', 'vgr5m', 'vgr5n', 'vgr6m', 'vgr6n', 'vgr7m',
'vgr7n', 'vgr8m', 'vgr8n', 'vgr9m', 'vgr9n', 'mk1', 'mk37', 'mk39',
'mk44', 'mk49', 'mk51', 'mk57', 'mk59', 'mk65', 'mk67', 'mk70',
'mk74', 'mk78', 'mk82', 'mk84', 'mk86', 'rahvus1', 'rahvus2',
'mlinn1', 'mlinn2')
margins=as.data.frame(margins)
margins$a=rep(0, 44)
margins$total=c(kontroll1[kontroll1$AASTA==aasta, 2],
kontroll2[kontroll2$AASTA==aasta, 2],
kontroll3[kontroll3$AASTA==aasta, 2],
kontroll4[kontroll4$AASTA==aasta, 2])
margins=as.matrix(margins)
leibkond=fun_leib(13, andmed2)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8, 1.5), description=FALSE)
uued_andmed3=merge(andmed2, leibkond[, c(48, 47)])
uued_andmed2=rbind(uued_andmed2, uued_andmed3)
leibkond=fun_leib(14, andmed2)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8, 1.6), description=FALSE)
uued_andmed3=merge(andmed2, leibkond[, c(48, 47)])
uued_andmed2=rbind(uued_andmed2, uued_andmed3)
leibkond=fun_leib(15, andmed2)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.8, 1.7), description=FALSE)
uued_andmed3=merge(andmed2, leibkond[, c(48, 47)])
uued_andmed2=rbind(uued_andmed2, uued_andmed3)
leibkond=fun_leib(16, andmed2)
leibkond$uus_kaal=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="logit",
bounds=c(0.7, 1.8), description=FALSE)

```

```

uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(17,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.7),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(18,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.7,1.7),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(19,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.7),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(20,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.7,1.5),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(21,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.9),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(22,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.8,1.6),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(23,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",
bounds=c(0.9,1.6),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
leibkond=fun_leib(24,andmed2)
leibkond$uus_kaal=calibration(data=leibkond,marginMatrix=
margins,colWeights="MVKAAL2",method="logit",

```

```

bounds=c(0.8,1.8),description=FALSE)
uued_andmed3=merge(andmed2,leibkond[,c(48,47)])
uued_andmed2=rbind(uued_andmed2,uued_andmed3)
#Leian hinnangud
kuu=period[1]
abi=uued_andmed2[uued_andmed2$PERIOD==kuu,]
abi=abi[abi$KADU==1,]
disain=svydesign(data=abi,ids=~LEIBKONNANR,
strata=~GRUPP1,weights=~uus_kaal,fpc=~RAHVAAARV)
hinnangud_logit_kuu=as.data.frame(svytotal(~factor(Status),
disain,data=kuu_andmed,na.rm=TRUE))
hinnangud_logit_kuu$PERIOD=rep(kuu,3)
for (i in 2:24) {
  kuu=period[i]
  abi=uued_andmed2[uued_andmed2$PERIOD==kuu,]
  abi=abi[abi$KADU==1,]
  disain=svydesign(data=abi,ids=~LEIBKONNANR,
strata=~GRUPP1,weights=~uus_kaal,fpc=~RAHVAAARV)
  a=as.data.frame(svytotal(~factor(Status),disain,
data=kuu_andmed,na.rm=TRUE))
  a$PERIOD=rep(kuu,3)
  hinnangud_logit_kuu=rbind(hinnangud_logit_kuu,a)}
#Ekspordin tulemuse
write.table(hinnangud_logit_kuu,"X:/Projektid/TY_ETU_
metoodika/Kaidi_Jogi/hinnangud_logit_kuu.txt",dec=",")
#Leian kalibreeritud kaalud 1 kuu andmetelt lineaarse
#meetodiga
aasta=2015
margins=c('vgr1m','vgr1n','vgr10m','vgr10n','vgr11m',
'vgr11n','vgr12m','vgr12n','vgr2m','vgr2n','vgr3m','vgr3n',
'vgr4m','vgr4n','vgr5m','vgr5n','vgr6m','vgr6n','vgr7m',
'vgr7n','vgr8m','vgr8n','vgr9m','vgr9n','mk1','mk37','mk39',
'mk44','mk49','mk51','mk57','mk59','mk65','mk67','mk70',
'mk74','mk78','mk82','mk84','mk86','rahvus1','rahvus2',
'mlinn1','mlinn2')
margins=as.data.frame(margins)
margins$a=rep(0,44)
margins$total=c(kontroll1[kontroll1$AASTA==aasta,2],
kontroll2[kontroll2$AASTA==aasta,2],
kontroll3[kontroll3$AASTA==aasta,2],
kontroll4[kontroll4$AASTA==aasta,2])
margins=as.matrix(margins)
#Kalibreerime iga kuu eraldi, kui leidub negatiivseid kaale,
#siis kordame nii, et MVKAALude asemel on kalibreeritud
#kaalud, kus negatiivsed kaalud on asendatud MVKAALudega
leibkond=fun_leib(1,andmed2)
leibkond$uus_kaal2=calibration(data=leibkond,marginMatrix=

```

```

margins, colWeights="MVKAAL2", method="linear",
description=FALSE)
uued_andmed3=merge(andmed2, leibkond[,c(48,47)])
for (i in 2:12){
  leibkond=fun_leib(i, andmed2)
  leibkond$uus_kaal2=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="linear",
description=FALSE)
  if (min(leibkond$uus_kaal2)<0){
    leibkond$MVKAAL2=ifelse(leibkond$uus_kaal2<0,
leibkond$MVKAAL2, leibkond$uus_kaal2)
    leibkond$uus_kaal2=calibration(data=leibkond,
marginMatrix=margins, colWeights="MVKAAL2",
method="linear", description=FALSE)}
  abi=merge(andmed2, leibkond[,c(48,47)])
  uued_andmed3=rbind(uued_andmed3, abi)}
aasta=2016
margins=c('vgr1m', 'vgr1n', 'vgr10m', 'vgr10n', 'vgr11m',
'vgr11n', 'vgr12m', 'vgr12n', 'vgr2m', 'vgr2n', 'vgr3m', 'vgr3n',
'vgr4m', 'vgr4n', 'vgr5m', 'vgr5n', 'vgr6m', 'vgr6n', 'vgr7m',
'vgr7n', 'vgr8m', 'vgr8n', 'vgr9m', 'vgr9n', 'mk1', 'mk37', 'mk39',
'mk44', 'mk49', 'mk51', 'mk57', 'mk59', 'mk65', 'mk67', 'mk70',
'mk74', 'mk78', 'mk82', 'mk84', 'mk86', 'rahvus1', 'rahvus2',
'mlinn1', 'mlinn2')
margins=as.data.frame(margins)
margins$a=rep(0,44)
margins$total=c(kontroll1[kontroll1$AASTA==aasta,2],
kontroll2[kontroll2$AASTA==aasta,2],
kontroll3[kontroll3$AASTA==aasta,2],
kontroll4[kontroll4$AASTA==aasta,2])
margins=as.matrix(margins)
for (i in 13:24){
  leibkond=fun_leib(i, andmed2)
  leibkond$uus_kaal2=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="linear",
description=FALSE)
  if (min(leibkond$uus_kaal2)<0){
    leibkond$MVKAAL2=ifelse(leibkond$uus_kaal2<0,
leibkond$MVKAAL2, leibkond$uus_kaal2)
    leibkond$uus_kaal2=calibration(data=leibkond,
marginMatrix=margins, colWeights="MVKAAL2",
method="linear", description=FALSE)}
  abi=merge(andmed2, leibkond[,c(48,47)])
  uued_andmed3=rbind(uued_andmed3, abi)}
#Kontrollin, et ei oleks jaanud negatiivseid kaale
min(uued_andmed3$uus_kaal2)
#Leian hinnangud

```

```

kuu=period[1]
abi=uued_andmed3[uued_andmed3$PERIOD==kuu,]
abi=abi[abi$KADU==1,]
disain=svydesign(data=abi,ids=~LEIBKONNANR,
strata=~GRUPP1,weights=~uus_kaal2,fpc=~RAHVAARV)
hinnangud_lin_kuu=as.data.frame(svytotal(~factor(Status),
disain,data=kuu_andmed,na.rm=TRUE))
hinnangud_lin_kuu$PERIOD=rep(kuu,3)
for (i in 2:24) {
  kuu=period[i]
  abi=uued_andmed3[uued_andmed3$PERIOD==kuu,]
  abi=abi[abi$KADU==1,]
  disain=svydesign(data=abi,ids=~LEIBKONNANR,
strata=~GRUPP1,weights=~uus_kaal2,fpc=~RAHVAARV)
  a=as.data.frame(svytotal(~factor(Status),disain,
data=kuu_andmed,na.rm=TRUE))
  a$PERIOD=rep(kuu,3)
  hinnangud_lin_kuu=rbind(hinnangud_lin_kuu,a)}
#Ekspordin tulemuse
write.table(hinnangud_lin_kuu,"X:/Projektid/TY_ETU_
metoodika/Kaidi_Jogi/hinnangud_lin_kuu.txt",dec=","")
#Modifitseeritud regressioonhinnang
#Kontrollin, kas koigil vastanud isikutel on Status maaratud
abi=andmed[andmed$KADU==1,]
summary(abi$Status)
#Kas koikidel tootajatel on EMTAK maaratud
abi2=abi[abi$Status==1,]
summary(abi2$SEKTOR)
sum(is.na(abi2$EMTAK))
table(abi2$EMTAK,abi2$PERIOD)
#Kontrollin ka ainult uhe kuu andmestikku
abi=andmed2[andmed2$KADU==1,]
summary(abi$Status)
#Kas koikidel tootajatel on EMTAK maaratud
abi2=abi[abi$Status==1,]
summary(abi2$SEKTOR)
sum(is.na(abi2$EMTAK))
summary(abi2$EMTAK)
table(abi2$EMTAK,abi2$PERIOD)
#Kas teist ja neljandat korda vastanute eelmine vastamise
#kord oli 3 kuud tagasi
abi=andmed2[andmed2$PERIOD=='2015_11',]
abi=abi[abi$KADU==1,]
abi2=andmed2[andmed2$PERIOD=='2015_08',]
abi2=abi2[abi2$KADU==1,]
abi3=andmed[andmed$PERIOD=='2015_08',]
abi3=abi3[abi3$KADU==1,]

```

```

a=abi[abi$LIIKMENR %in% abi2$LIIKMENR,]
aa=abi[abi$LIIKMENR %in% abi3$LIIKMENR,]
aaa=aa[!aa$LIIKMENR %in% a$LIIKMENR,]
aaa=andmed2[andmed2$LIIKMENR %in% aaa$LIIKMENR,]
abi=andmed2[andmed2$PERIOD== '2015_12',]
abi2=andmed2[andmed2$PERIOD== '2015_09',]
abi3=andmed[andmed$PERIOD== '2015_09',]
a=abi[abi$LIIKMENR %in% abi2$LIIKMENR,]
aa=abi[abi$LIIKMENR %in% abi3$LIIKMENR,]
abi=andmed2[andmed2$PERIOD== '2016_02',]
abi2=andmed2[andmed2$PERIOD== '2015_11',]
abi3=andmed[andmed$PERIOD== '2015_11',]
a=abi[abi$LIIKMENR %in% abi2$LIIKMENR,]
aa=abi[abi$LIIKMENR %in% abi3$LIIKMENR,]
rm(abi,abi2,abi3,a,aa,aaa)
#Defineerin uued soo-ja vanusgruppides tootamise tunnused
uued_andmed3$toogrupp=ifelse((uued_andmed3$Status==1)*
(uued_andmed3$SUGU==1)*(uued_andmed3$VANUSGR %in%
c(1,2)), "Stat1M15_24",
ifelse((uued_andmed3$Status==2)*(uued_andmed3$SUGU==1)*
(uued_andmed3$VANUSGR %in% c(1,2)), "Stat2M15_24",
ifelse((uued_andmed3$Status==3)*(uued_andmed3$SUGU==1)*
(uued_andmed3$VANUSGR %in% c(1,2)), "Stat3M15_24",
ifelse((uued_andmed3$Status==1)*(uued_andmed3$SUGU==1)*
(!uued_andmed3$VANUSGR %in% c(1,2)), "Stat1M25",
ifelse((uued_andmed3$Status==2)*(uued_andmed3$SUGU==1)*
(!uued_andmed3$VANUSGR %in% c(1,2)), "Stat2M25",
ifelse((uued_andmed3$Status==3)*(uued_andmed3$SUGU==1)*
(!uued_andmed3$VANUSGR %in% c(1,2)), "Stat3M25",
ifelse((uued_andmed3$Status==1)*(uued_andmed3$SUGU==2)*
(uued_andmed3$VANUSGR %in% c(1,2)), "Stat1N15_24",
ifelse((uued_andmed3$Status==2)*(uued_andmed3$SUGU==2)*
(uued_andmed3$VANUSGR %in% c(1,2)), "Stat2N15_24",
ifelse((uued_andmed3$Status==3)*(uued_andmed3$SUGU==2)*
(uued_andmed3$VANUSGR %in% c(1,2)), "Stat3N15_24",
ifelse((uued_andmed3$Status==1)*(uued_andmed3$SUGU==2)*
(!uued_andmed3$VANUSGR %in% c(1,2)), "Stat1N25",
ifelse((uued_andmed3$Status==2)*(uued_andmed3$SUGU==2)*
(!uued_andmed3$VANUSGR %in% c(1,2)), "Stat2N25",
ifelse((uued_andmed3$Status==3)*(uued_andmed3$SUGU==2)*
(!uued_andmed3$VANUSGR %in% c(1,2)), "Stat3N25", NA)))))))))
#Kui inimene ei toota, on ta tookoha oman_liik ja Sektor NA,
#muudan need nullideks
uued_andmed3$OMAN_LIIK=ifelse(is.na(uued_andmed3$OMAN_LIIK),
0,uued_andmed3$OMAN_LIIK)
uued_andmed3$SEKTOR=ifelse(is.na(uued_andmed3$SEKTOR),
0,uued_andmed3$SEKTOR)

```

```

#Loon funktsiooni, mis tekitab liitkirjeldavad tunnused
liit=function(kuu, alpha){
  #Leian sellel kuul vastanud isikute andmestiku
  see_kuu=periood[kuu]
  see_kuu_vast=uued_andmed3[uued_andmed3$PERIOOD==see_kuu,]
  see_kuu_vast=see_kuu_vast[see_kuu_vast$KADU==1,]
  #Leian sellel kuul vastanud isikute eelmise perioodi
  #andmed
  eelm_kuu=periood[kuu-3]
  eelm_kuu_and=uued_andmed3[uued_andmed3$PERIOOD==eelm_kuu,]
  eelm_kuu_and=eelm_kuu_and[eelm_kuu_and$KADU==1,]
  rot_eelm_kuu=eelm_kuu_and[eelm_kuu_and$LIIKMENR %in%
  see_kuu_vast$LIIKMENR,]
  #Jagan selle kuu andmed kaheks, uued objektid ja roteeruv
  #valim
  see_kuu_uus=see_kuu_vast[!see_kuu_vast$LIIKMENR %in%
  rot_eelm_kuu$LIIKMENR,]
  see_kuu_rot=see_kuu_vast[see_kuu_vast$LIIKMENR %in%
  rot_eelm_kuu$LIIKMENR,]
  #Leian toenaosuse, et inimene on rotatsioonigrupis, kui ta
  #vastas
  tn=sum(see_kuu_rot$MVKAAL2)/sum(see_kuu_vast$MVKAAL2)
  #Leian liitkirjeldavad tunnused uutele valimi objektidele
  #Selleks leian Staatuse hinnangud eelmisel kuul
  disain=svydesign(data=eelm_kuu_and, ids=~LEIBKONNANR,
  strata=~GRUPP1, weights=~uus_kaal2, fpc=~RAHVAARV)
  kontroll5=as.data.frame(svytotal(~factor(Status), disain,
  data=eelm_kuu_and, na.rm=TRUE))
  a=kontroll1[kontroll1$AASTA==substring(see_kuu, 1, 4),]
  RAHVAARV=round(sum(a$x))
  see_kuu_uus$Stat1=(1-alpha)*kontroll5$total[1]/RAHVAARV
  +alpha*(see_kuu_uus$Status==1)
  see_kuu_uus$Stat2=(1-alpha)*kontroll5$total[2]/RAHVAARV
  +alpha*(see_kuu_uus$Status==2)
  #Tootamine vanusgruppides
  kontroll6=as.data.frame(svytotal(~factor(toogrupp),
  disain, data=eelm_kuu_and, na.rm=TRUE))
  see_kuu_uus$Stat1M15_24=(1-alpha)*kontroll6$total[1]/
  RAHVAARV+alpha*(see_kuu_uus$toogrupp=="Stat1M15_24")
  see_kuu_uus$Stat2M15_24=(1-alpha)*kontroll6$total[5]/
  RAHVAARV+alpha*(see_kuu_uus$toogrupp=="Stat2M15_24")
  see_kuu_uus$Stat1M25=(1-alpha)*kontroll6$total[2]/
  RAHVAARV+alpha*(see_kuu_uus$toogrupp=="Stat1M25")
  see_kuu_uus$Stat2M25=(1-alpha)*kontroll6$total[6]/
  RAHVAARV+alpha*(see_kuu_uus$toogrupp=="Stat2M25")
  see_kuu_uus$Stat1N15_24=(1-alpha)*kontroll6$total[3]/
  RAHVAARV+alpha*(see_kuu_uus$toogrupp=="Stat1N15_24")

```

```

see_kuu_uus$Stat2N15_24=(1-alpha)*kontroll6$total[7]/
RAHVAAARV+alpha*(see_kuu_uus$toogrupp=="Stat2N15_24")
#Omandi liik
kontroll7=as.data.frame(svytotal(~factor(OMAN_LIIK),
disain,data=eelm_kuu_and,na.rm=TRUE))
see_kuu_uus$Oman1=(1-alpha)*kontroll7$total[1]/RAHVAAARV
+alpha*(see_kuu_uus$OMAN_LIIK==1)
#Tootamine sektorites
kontroll8=as.data.frame(svytotal(~factor(SEKTOR),disain,
data=eelm_kuu_and,na.rm=TRUE))
see_kuu_uus$Sektor1=(1-alpha)*kontroll8$total[1]/RAHVAAARV
+alpha*(see_kuu_uus$SEKTOR==1)
see_kuu_uus$Sektor2=(1-alpha)*kontroll8$total[2]/RAHVAAARV
+alpha*(see_kuu_uus$SEKTOR==2)
#Liitkirjeldavad tunnused rotatsioonigrupile
see_kuu_rot=see_kuu_rot[order(see_kuu_rot$LIKMEMR),]
rot_eelm_kuu=rot_eelm_kuu[order(rot_eelm_kuu$LIKMEMR),]
#Staatus
see_kuu_rot$Stat1=(1-alpha)*(rot_eelm_kuu$Status==1)
+alpha*((rot_eelm_kuu$Status==1)+(1/tn-1)*
((rot_eelm_kuu$Status==1)-(see_kuu_rot$Status==1)))
see_kuu_rot$Stat2=(1-alpha)*(rot_eelm_kuu$Status==2)
+alpha*((rot_eelm_kuu$Status==2)+(1/tn-1)*
((rot_eelm_kuu$Status==2)-(see_kuu_rot$Status==2)))
#Tootamine vanusgruppides
see_kuu_rot$Stat1M15_24=(1-alpha)*
(rot_eelm_kuu$toogrupp=="Stat1M15_24")+alpha*
((rot_eelm_kuu$toogrupp=="Stat1M15_24")+1/tn-1)*
((rot_eelm_kuu$toogrupp=="Stat1M15_24")-
(see_kuu_rot$toogrupp=="Stat1M15_24"))
see_kuu_rot$Stat2M15_24=(1-alpha)*
(rot_eelm_kuu$toogrupp=="Stat2M15_24")+alpha*
((rot_eelm_kuu$toogrupp=="Stat2M15_24")+1/tn-1)*
((rot_eelm_kuu$toogrupp=="Stat2M15_24")-
(see_kuu_rot$toogrupp=="Stat2M15_24"))
see_kuu_rot$Stat1M25=(1-alpha)*
(rot_eelm_kuu$toogrupp=="Stat1M25")+alpha*
((rot_eelm_kuu$toogrupp=="Stat1M25")+1/tn-1)*
((rot_eelm_kuu$toogrupp=="Stat1M25")-
(see_kuu_rot$toogrupp=="Stat1M25"))
see_kuu_rot$Stat2M25=(1-alpha)*
(rot_eelm_kuu$toogrupp=="Stat2M25")+alpha*
((rot_eelm_kuu$toogrupp=="Stat2M25")+1/tn-1)*
((rot_eelm_kuu$toogrupp=="Stat2M25")-
(see_kuu_rot$toogrupp=="Stat2M25"))
see_kuu_rot$Stat1N15_24=(1-alpha)*
(rot_eelm_kuu$toogrupp=="Stat1N15_24")+alpha*

```

```

((rot_eelm_kuu$toogrupp=="Stat1N15_24")+(1/tn-1)*
((rot_eelm_kuu$toogrupp=="Stat1N15_24")-
(see_kuu_rot$toogrupp=="Stat1N15_24")))
see_kuu_rot$Stat2N15_24=(1-alpha)*
(rot_eelm_kuu$toogrupp=="Stat2N15_24")+alpha*
((rot_eelm_kuu$toogrupp=="Stat2N15_24")+(1/tn-1)*
((rot_eelm_kuu$toogrupp=="Stat2N15_24")-
(see_kuu_rot$toogrupp=="Stat2N15_24")))
#Omandi liik
see_kuu_rot$Oman1=(1-alpha)*(rot_eelm_kuu$OMAN_LIIK==1)
+alpha*((rot_eelm_kuu$OMAN_LIIK==1)+(1/tn-1)*
((rot_eelm_kuu$OMAN_LIIK==1)-(see_kuu_rot$OMAN_LIIK==1)))
#Tootamine sektorites
see_kuu_rot$Sektor1=(1-alpha)*(rot_eelm_kuu$SEKTOR==1)
+alpha*((rot_eelm_kuu$SEKTOR==1)+(1/tn-1)*
((rot_eelm_kuu$SEKTOR==1)-(see_kuu_rot$SEKTOR==1)))
see_kuu_rot$Sektor2=(1-alpha)*(rot_eelm_kuu$SEKTOR==2)
+alpha*((rot_eelm_kuu$SEKTOR==2)+(1/tn-1)*
((rot_eelm_kuu$SEKTOR==2)-(see_kuu_rot$SEKTOR==2)))
#Liidan andmestikud
see_kuu_vast=rbind(see_kuu_uus,see_kuu_rot)
return(see_kuu_vast)}
#Loon funktsiooni, mis tekitab leibkonnabaasil andmestiku
fun_leib=function(i,alpha){
  #Jattan ainult uuritava kuu andmed
  kuu=period[i]
  kuu_andmed=liit(i,alpha)
  #Jatan alles kokku liidetavad tunnused
  kuu_andmed=kuu_andmed[,c(33,36,37:80,83:93)]
  #Moodustan andmestiku, kus igal leibkonnal on 1 rida
  kuu_andmed2=aggregate(.~LEIBKONNANR,kuu_andmed,sum)
  abi=aggregate(MVKAAL2~LEIBKONNANR,kuu_andmed,mean)
  kuu_andmed2$MVKAAL2=abi$MVKAAL2
  kuu_andmed2$leibkuu=paste(kuu_andmed2$LEIBKONNANR,kuu)
  return(kuu_andmed2)}
#Teostan modifitseeritud regressiooni
mod=function(kuu_nr){
  alp=2/3
  leibkond=fun_leib(kuu_nr,alp)
  kuu=period[kuu_nr]
  aasta=substring(kuu,1,4)
  #Loon kontrollsummade loendi
  margins=c('vgr1m','vgr1n','vgr10m','vgr10n','vgr11m',
'vgr11n','vgr12m','vgr12n','vgr2m','vgr2n','vgr3m','vgr3n',
'vgr4m','vgr4n','vgr5m','vgr5n','vgr6m','vgr6n','vgr7m',
'vgr7n','vgr8m','vgr8n','vgr9m','vgr9n','mk1','mk37',
'mk39','mk44','mk49','mk51','mk57','mk59','mk65','mk67',

```

```

'mk70', 'mk74', 'mk78', 'mk82', 'mk84', 'mk86', 'rahvus1',
'rahvus2', 'mlinn1', 'mlinn2', 'Stat1', 'Stat2', 'Stat1M15_24',
'Stat2M15_24', 'Stat1M25', 'Stat2M25', 'Stat1N15_24',
'Stat2N15_24', 'Oman1', 'Sektor1', 'Sektor2')
margins=as.data.frame(margins)
margins$a=rep(0,55)
#Leian liitkirjedavate tunnuste kogusumma hinnangud 3 kuud
#tagasi
eelm_kuu=period[kuu_nr-3]
eelm_kuu_and=uued_andmed3[uued_andmed3$PERIOD==eelm_kuu,]
eelm_kuu_and=eelm_kuu_and[eelm_kuu_and$KADU==1,]
disain=svydesign(data=eelm_kuu_and, ids=~LEIBKONNANR,
strata=~GRUPP1, weights=~uus_kaal2, fpc=~RAHVAARV)
kontroll5=as.data.frame(svytotal(~factor(Status), disain,
data=eelm_kuu_and, na.rm=TRUE))
kontroll6=as.data.frame(svytotal(~factor(toogrupp),
disain, data=eelm_kuu_and, na.rm=TRUE))
kontroll7=as.data.frame(svytotal(~factor(OMAN_LIIK),
disain, data=eelm_kuu_and, na.rm=TRUE))
kontroll8=as.data.frame(svytotal(~factor(SEKTOR), disain,
data=eelm_kuu_and, na.rm=TRUE))
margins$total=c(kontroll1[kontroll1$AASTA==aasta,2],
kontroll2[kontroll2$AASTA==aasta,2],
kontroll3[kontroll3$AASTA==aasta,2],
kontroll4[kontroll4$AASTA==aasta,2],
kontroll5$total[1:2], kontroll6$total[c(1,5,2,6,3,7)],
kontroll7$total[2], kontroll8$total[2:3])
margins=as.matrix(margins)
#Kalibreerime iga kuu eraldi, kui leidub negatiivseid
#kaale, siis kordame nii, et MVKAALude asemel on
#kalibreeritud kaalud, kus negatiivsed kaalud on asendatud
#MVKAALudega
leibkond$uus_kaal2=calibration(data=leibkond, marginMatrix=
margins, colWeights="MVKAAL2", method="linear",
description = FALSE)
if(min(leibkond$uus_kaal2)<0){
  leibkond$MVKAAL2=ifelse(leibkond$uus_kaal2<0,
leibkond$MVKAAL2, leibkond$uus_kaal2)
  leibkond$uus_kaal2=calibration(data=leibkond,
marginMatrix=margins, colWeights="MVKAAL2",
method="linear", description=FALSE)}
#Asendan eelmises andmestikus tunnuse uus_kaal2, praegu
#leitud kaaludega
abi=uued_andmed3[uued_andmed3$PERIOD==kuu,]
abi$uus_kaal2=NULL
abi=merge(abi, leibkond[,c(58,59)])
uued_andmed32=uued_andmed3[!uued_andmed3$PERIOD==kuu,]

```

```

    uued_andmed3=rbind(abi,uued_andmed32)
    return(uued_andmed3)}
#Kalibreerin iga kuu eraldi, alates aprill 2015, sest meil
#on vaja 3 kuud tagasi arvutatud andmeid
for(i in 4:24){
  uued_andmed3=mod(i)}
#Leian, kas kuudel on olnud negatiivseid kaale
aggregate(uued_andmed3$uus_kaal2,
by=list(Category=uued_andmed3$PERIOD),FUN=min)
#Eemaldan liitkirjeldavaid tunnuseid ning kalibreerin
#uuesti
mod=function(kuu_nr){
  alp=2/3
  leibkond=fun_leib(kuu_nr,alp)
  kuu=period[kuu_nr]
  aasta=substring(kuu,1,4)
  #Loon kontrollsummade loendi
  margins=c('vgr1m','vgr1n','vgr10m','vgr10n','vgr11m',
  'vgr11n','vgr12m','vgr12n','vgr2m','vgr2n','vgr3m',
  'vgr3n','vgr4m','vgr4n','vgr5m','vgr5n','vgr6m','vgr6n',
  'vgr7m','vgr7n','vgr8m','vgr8n','vgr9m','vgr9n','mk1',
  'mk37','mk39','mk44','mk49','mk51','mk57','mk59','mk65',
  'mk67','mk70','mk74','mk78','mk82','mk84','mk86',
  'rahvus1','rahvus2','mlinn1','mlinn2','Stat1','Stat2',
  'Stat1M15_24','Stat2M15_24','Stat1M25','Stat2M25',
  'Stat1N15_24','Stat2N15_24')
  margins=as.data.frame(margins)
  margins$a=rep(0,52)
  #Leian liitkirjedavate tunnuste kogusumma hinnangud 3
  #kuud tagasi
  eelm_kuu=period[kuu_nr-3]
  eelm_kuu_and=uued_andmed3[uued_andmed3$PERIOD==eelm_kuu,]
  eelm_kuu_and=eelm_kuu_and[eelm_kuu_and$KADU==1,]
  disain=svydesign(data=eelm_kuu_and,ids=~LEIBKONNANR,
  strata=~GRUPP1,weights=~uus_kaal2,fpc=~RAHVAARV)
  kontroll5=as.data.frame(svytotal(~factor(Status),
  disain,data=eelm_kuu_and,na.rm=TRUE))
  kontroll6=as.data.frame(svytotal(~factor(toogrupp),
  disain,data=eelm_kuu_and,na.rm=TRUE))
  margins$total=c(kontroll1[kontroll1$AASTA==aasta,2],
  kontroll2[kontroll2$AASTA==aasta,2],
  kontroll3[kontroll3$AASTA==aasta,2],
  kontroll4[kontroll4$AASTA==aasta,2],
  kontroll5$total[1:2],kontroll6$total[c(1,5,2,6,3,7)])
  margins=as.matrix(margins)
  #Kalibreerime iga kuu eraldi, kui leidub negatiivseid
  #kaale, siis kordame nii, et MVKAALude asemel on

```

```

#kalibreeritud kaalud, kus negatiivsed kaalud on
#asendatud MVKAALudega
leibkond$uus_kaal2=calibration(data=leibkond,
marginMatrix=margins,colWeights="MVKAAL2",
method="linear",description=FALSE)
if (min(leibkond$uus_kaal2)<0){
  leibkond$MVKAAL2=ifelse(leibkond$uus_kaal2<0,
  leibkond$MVKAAL2,leibkond$uus_kaal2)
  leibkond$uus_kaal2=calibration(data=leibkond,
marginMatrix=margins,colWeights="MVKAAL2",
method="linear",description=FALSE)}
#Asendan eelmises andmestikus tunnuse uus_kaal2, praegu
#leitud kaaludega
abi=uued_andmed3[uued_andmed3$PERIOOD==kuu,]
abi$uus_kaal2=NULL
abi=merge(abi,leibkond[,c(58,59)])
uued_andmed32=uued_andmed3[!uued_andmed3$PERIOOD==kuu,]
uued_andmed3=rbind(abi,uued_andmed32)
return(uued_andmed3)}
#Kalibreerin iga kuu eraldi, alates aprill 2015, sest meil
#on vaja 3 kuud tagasi arvutatud andmeid
for(i in 4:24){
  uued_andmed3=mod(i)}
#Leian, kas kuudel on olnud negatiivseid kaale
aggregate(uued_andmed3$uus_kaal2,
by=list(Category=uued_andmed3$PERIOOD),FUN=min)
#Leian hinnangud
kuu=period[4]
abi=uued_andmed3[uued_andmed3$PERIOOD==kuu,]
abi=abi[abi$KADU==1,]
disain=svydesign(data=abi,ids=~LEIBKONNANR,
strata=~GRUPP1,weights=~uus_kaal2,fpc=~RAHVAARV)
hinnangud_mod=as.data.frame(svytotal(~factor(Status),
disain,data=kuu_andmed,na.rm=TRUE))
hinnangud_mod$PERIOOD=rep(kuu,3)
for(i in 5:24) {
  kuu=period[i]
  abi=uued_andmed3[uued_andmed3$PERIOOD==kuu,]
  abi=abi[abi$KADU==1,]
  disain=svydesign(data=abi,ids=~LEIBKONNANR,
strata=~GRUPP1,weights=~uus_kaal2,fpc=~RAHVAARV)
a=as.data.frame(svytotal(~factor(Status),disain,
data=kuu_andmed,na.rm=TRUE))
a$PERIOOD=rep(kuu,3)
hinnangud_mod=rbind(hinnangud_mod,a)}
#Ekspordin tulemuse
write.table(hinnangud_mod,"X:/Projektid/TY_ETU_metoodika/

```

```
Kaidi_Jogi/hinnangud_mod.txt", dec=",")
```

Lisa 3.6 Jooniste tegemine

```
tulem=read.csv("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
tulem.csv", header=TRUE, sep=";", dec=",", na.strings="NA")
#Muudan perioodi kuupaevaks
tulem$PERIOD=paste(tulem$PERIOD, rep(01,72), sep="_")
tulem$PERIOD=as.Date(tulem$PERIOD, format="%Y_%m_%d")
#Jagan andmestiku kolmeks
hoivatud=tulem[tulem$Status=="Hoivatud",]
tootud=tulem[tulem$Status=="Tootud",]
mitteaktiivsed=tulem[tulem$Status=="Mitteaktiivsed",]
tulem_sd=read.csv("C:\\Users\\Kaidi\\Documents\\
Magistritoo\\tulem_sd.csv", header=TRUE, sep=";", dec=",",
na.strings="NA" )
#Muudan perioodi kuupaevaks
tulem_sd$PERIOD=paste(tulem_sd$PERIOD, rep(01,72),
sep="_")
tulem_sd$PERIOD=as.Date(tulem_sd$PERIOD,
format="%Y_%m_%d")
#Jagan andmestiku kolmeks
hoivatud_sd=tulem_sd[tulem_sd$Status=="Hoivatud",]
tootud_sd=tulem_sd[tulem_sd$Status=="Tootud",]
mitteaktiivsed_sd=tulem_sd[tulem_sd$Status=="Mitteaktiivsed",]
#Teen joonised
library("reshape2")
library("ggplot2")
library("gridExtra")
library("lubridate")
hoivatud=melt(hoivatud[,c(1,3,5,7,9)], id="PERIOD")
hoivatud_sd=melt(hoivatud_sd[,c(1,3:6)], id="PERIOD")
p=ggplot(data=hoivatud, aes(x=PERIOD, y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Hoivatud")+
theme(legend.position="bottom")+
theme(legend.title=element_blank())
q=ggplot(data=hoivatud_sd, aes(x=PERIOD, y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Standardhalve")
g_legend=function(a.gplot){
  tmp=ggplot_gtable(ggplot_build(a.gplot))
  leg=which(sapply(tmp$grobs, function(x) x$name)==
"guide-box")
  legend=tmp$grobs[[leg]]
```

```

    return(legend)}
mylegend=g_legend(p)
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
hoivatud.png",grid.arrange(arrangeGrob(p+
theme(legend.position="none"),q+
theme(legend.position="none"),nrow=1),mylegend,nrow=2,
heights=c(10,3)),width=18,height=8,units="cm",
dpi=600)
tootud=melt(tootud[,c(1,3,5,7,9)],id="PERIOD")
tootud_sd=melt(tootud_sd[,c(1,3:6)],id="PERIOD")
p=ggplot(data=tootud,aes(x=PERIOD,y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Tootud")+
theme(legend.position="bottom")+
theme(legend.title=element_blank())
q=ggplot(data=tootud_sd,aes(x=PERIOD,y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Standardhalve")
mylegend=g_legend(p)
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
tootud.png",grid.arrange(arrangeGrob(p+
theme(legend.position="none"),q+
theme(legend.position="none"),nrow=1),mylegend,nrow=2,
heights=c(10,3)),width=18,height=8,units="cm",
dpi=600)
mitteaktiivsed=melt(mitteaktiivsed[,c(1,3,5,7,9)],
id="PERIOD")
mitteaktiivsed_sd=melt(mitteaktiivsed_sd[,c(1,3:6)],
id="PERIOD")
p=ggplot(data=mitteaktiivsed,aes(x=PERIOD,y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Mitteaktiivsed")+
theme(legend.position="bottom")+
theme(legend.title=element_blank())
q=ggplot(data=mitteaktiivsed_sd,aes(x=PERIOD,y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Standardhalve")
mylegend=g_legend(p)
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
mitteaktiivsed.png",grid.arrange(arrangeGrob(p+
theme(legend.position="none"),q+
theme(legend.position="none"),nrow=1),mylegend,nrow=2,
heights=c(10,3)),width=18,height=8,units="cm",

```

```

dpi=600)
hoivatud_suht=hoivatud
hoivatud_suht$value=hoivatud_sd$value/hoivatud_suht$value
hoivatud_suht=hoivatud_suht[!hoivatud_suht$variable %in%
c("Libisev.kvartal","Kalibreeritud.logit.meetodiga"),]
p=ggplot(data=hoivatud_suht, aes(x=PERIOD, y=value,
colour=variable))+geom_line()+geom_point()+
scale_y_continuous(name="Suhteline_viga")+
theme(legend.position="bottom")+
theme(legend.title=element_blank())
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
hoivatud_suht.png", p, width=18, height=8, units="cm",
dpi=600)
tootud_suht=tootud
tootud_suht$value=tootud_sd$value/tootud_suht$value
tootud_suht=tootud_suht[!tootud_suht$variable %in%
c("Libisev.kvartal","Kalibreeritud.logit.meetodiga"),]
p=ggplot(data=tootud_suht, aes(x=PERIOD, y=value,
colour=variable))+geom_line()+geom_point()+
scale_y_continuous(name="Suhteline_viga")+
theme(legend.position="bottom")+
theme(legend.title=element_blank())
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
tootud_suht.png", p, width=18, height=8, units="cm", dpi=600)
mitteaktiivsed_suht=mitteaktiivsed
mitteaktiivsed_suht$value=mitteaktiivsed_sd$value/
mitteaktiivsed_suht$value
mitteaktiivsed_suht=mitteaktiivsed_suht[
!mitteaktiivsed_suht$variable %in% c("Libisev.kvartal",
"Kalibreeritud.logit.meetodiga"),]
p=ggplot(data=mitteaktiivsed_suht, aes(x=PERIOD, y=value,
colour=variable))+geom_line()+geom_point()+
scale_y_continuous(name="Suhteline_viga")+
theme(legend.position="bottom")+
theme(legend.title=element_blank())
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
mitteaktiivsed_suht.png", p, width=18, height=8, units="cm",
dpi=600)
tulem=read.csv("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
tulem_vahe.csv", header=TRUE, sep=";", dec=".",
na.strings="NA")
#Muudan perioodi kuupaevaks
tulem$PERIOD=paste(tulem$PERIOD, rep(01,72), sep="_")
tulem$PERIOD=as.Date(tulem$PERIOD, format="%Y_%m_%d")
tulem=tulem[order(tulem$PERIOD),]
tulem=tulem[year(tulem$PERIOD)>2015,]
#Jagan andmestiku kolmeks

```

```

hoivatud=tulem[tulem$Status=="Hoivatud",]
tootud=tulem[tulem$Status=="Tootud",]
mitteaktiivsed=tulem[tulem$Status=="Mitteaktiivsed",]
hoivatud=melt(hoivatud[,c(1,11,12,13,14)], id="PERIOD")
p=ggplot(data=hoivatud,aes(x=PERIOD,y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Hoivatute_aru_muutus")+
theme(legend.position="bottom")+ theme(legend.title=element_blank())
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
hoivatud_vahe.png",p,width=18,height=8,units="cm",dpi=600)
tootud=melt(tootud[,c(1,11,12,13,14)],id="PERIOD")
p=ggplot(data=tootud,aes(x=PERIOD,y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Tootute_aru_muutus")+
theme(legend.position="bottom")+
theme(legend.title=element_blank())
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
tootud_vahe.png",p,width=18,height=8,units="cm",dpi=600)
mitteaktiivsed=melt(mitteaktiivsed[,c(1,11,12,13,14)],
id="PERIOD")
p=ggplot(data=mitteaktiivsed,aes(x=PERIOD,y=value,
colour=variable))+geom_line()+geom_point()+
theme(legend.position="none")+
scale_y_continuous(name="Mitteaktiivsete_inimeste_aru_muutus")+theme(legend.position="bottom")+
theme(legend.title=element_blank())
ggsave("C:\\Users\\Kaidi\\Documents\\Magistritoo\\
mitteaktiivsed_vahe.png",p,width=18,height=8,units="cm",
dpi=600)

```

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Kaidi Kapp,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Modifitseeritud regressioonhinnang Eesti tööjõu-uuringu näitel“, mille juhendaja on Imbi Traat, 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, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kaidi Kapp
15.05.2019