

2011. AASTA RAHVA JA ELURUUMIDE LOENDUSE ALAKAETUSE HINNANG

Ene-Margit Tiit
Statistikaamet, Tartu Ülikool

Tänapäeval on kõigis arenenud riikides loenduse alakaetus oluline probleem: osa loendamisele kuuluvaid inimesi ei osale loendusel. Nad ei soovi ennast mitmesugustel põhjustel (soov kaitsta oma privaatsust, tahtmatus riigivõimuga suhelda, sage liikuvus) lasta loendada, mistõttu ei ole nad loendajale kättesaadavad. Enamasti hoiduvad sellised inimesed ka igas vormis iseloendamisest. Loendusmeeskonna ülesannete hulka kuulub seega ka alakaetuse hindamine, millele võib lisanduda alakaetuse kompenseerimine.

Alakaetuse hindamise meetodika

Loenduse alakaetuse hindamiseks on tarvis teada loendamata jäänud isikute (objektide) arvu V , aga ka tegelikult loendamisele kuuluvate isikute arvu ehk üldkogumit N . Loendamise tulemusena on teada loendatud isikute arv K ehk nn loenduskogum. Üldkogumi jaoks on võimalik saada hinnang (nn hinnatav üldkogum (Euroopa ... 2010)), arvutades summa $N = K + V$. Alakaetuse määr on siis $V / (K+V)$, see avaldatakse tavaliselt protsentides ja see näitab, kui suur osa loendamisele kuuluvatest isikutest jäi loendamata. Seega on alakaetuse hindamisel põhiliseks eesmärgiks hinnata loendamata jäänud, kuid loendamisele kuuluvate isikute arvu V .

Alakaetuse hindamiseks on mitmeid meetodeid, neist kõige populaarsem on järeloendus, mille käigus loendatakse spetsiaalselt disainitud valimis isikud uuesti (lühendatud ankeediga) ja leitakse, missugune osa isikuid esines mõlemas loenduses ja missugune puudus põhiloenduses – see teave võimaldab anda kaetusele statistilise hinnangu. Meetodi puuduseks on, et see võimaldab küll avastada juhuslikult loendusest välja jäänud isikud, kuid mitte süstemaatilisi eemalehoidjaid. Eesti jaoks sobib praegu kõige paremini administratiivsete registrite kasutamine, sest seoses kavaga teha järgmine rahva ja eluruumide loendus registripõhiselt, on registrite kvaliteedi testimine ja nende kasutamine rahvastikuanalüüsis igati asjakohane. Registrite kvaliteedi analüüsimisega on rahva ja eluruumide loenduse (REL) meetodid tegelenud juba alates 2008. aastast, kui arutati võimalust teostada rahvaloendus registripõhiselt. Analüüs näitas siis, et Eesti registrid ei olnud veel vajalikul tasemel selleks, et loobuda tavapärasest loendusest (kõik Eurostati väljundis nõutavad tunnused ei ole Eesti registritega kaetud), kuid need pakuvad tänuväärset materjali loenduse korraldamise mitmetel etappidel ja muuhulgas ka selleks, et selgitada välja loendamata jäänud püsielanikke (residente) ja anda sellega usaldusväärne hinnang loenduse alakaetusele.

Registrid ja võimalust kasutada neid loendamata jäänud püsielanike arvu selgitamiseks on üksikasjalikumalt käsitletud artiklis „Rahvaloendajate tegevus küsitluse järel“ (Tiit, Vähi 2012), rakendatavate eeskirjade matemaatiline esitus on artiklites „Rahvaloenduse üldkogumi hindamine“ (Tiit jt 2012) ja „Loendusmeeskond jätkab tööd“ (Tiit 2012), samuti on need tulemused esitatud rahvusvahelisel seminaril (Joint ... 2012).

Kasutatavatest registritest olulisim on rahvastikuregister, mis eeldatavasti sisaldab kõiki Eestis elavaid inimesi, kuid mille Eesti elanike osa on mõnevõrra ülekaetud, sest kõik väljarändajad ei ole enda lahkumist rahvastikuregistris registreerinud. Registreerimata illegaalsete immigrantide arv on Eestis hinnanguliselt võrdlemisi väike, sest riigi immigratsioonipoliitika on olnud üsna karm ja peale selle ei ole Eesti eriti ahvatlev sihtriik. Ülejäänud registritest on olulisemad **ravi-kindlustuse andmekogu**, mis sisaldab üle 90% Eesti inimestest, **maksukohustuslaste register**, mis katab olulise osa tööealisest elanikkonnast, **Eesti Hariduse Infosüsteem (EHIS)**, mis sisaldab peaaegu kõikselt õppijate andmeid. Täiendavat infot pakuvad veel **riiklik pensionikindlustuse register** (sisaldab nii pensionide, peretoetuste kui ka hüvitiste saajaid),

sotsiaalteenuste ja -toetuste andmeregister (STAR), samuti **liiklusregister**. Kokku kasutati loendamata jäänud residentide arvu määramiseks enam kui tosinat registrit ja alamregistrit, kusjuures otsustamise aluseks oli analüüsitava isikute aktiivsus analüüsitava registrite andmestikes 2011. aastal.

Potentsiaalsete residentidena käsitleti ja analüüsiti kõiki neid rahvastikuregistris paikneva isikuid, kelle püsielukohaks oli 1. jaanuaril 2012 Eesti ja kes ei olnud REL-i käigus loendatud ei püsielanike, lahkunute ega ajutiste elanikena. Isikuid, kellel oli küll Eesti isikukood, kuid kes loendusmomendil rahvastikuregistri andmetel ei elanud püsivalt Eestis, potentsiaalsete residentide hulka ei arvatud ja neid ei analüüsitud. Niisuguste isikute lisamine analüüsitava leubanuks potentsiaalsete residentide arvu suurendada vähem kui ühe protsendi võrra tegelikust rahvaarvust, kuid oleks suurendanud kaasamisvea tõenäosust. Analüüsimiseks kasutati viit eeskirja (mudelit), mida on üksikasjalikult kirjeldatud artiklites „Rahvaloendajate tegevus küsitluse järel (Tiit, Vähi 2012) ja „Rahvaloenduse üldkogumi hindamine“ (Tiit jt 2012). Need eeskirjad on:

- lineaarne diskriminantanalüüs võrdsete teoreetiliste vigade eeldusel;
- lineaarne diskriminantanalüüs eeldusel, et esimest liiki viga on väiksem kui 0,05;
- logistiline diskriminantanalüüs võrdsete teoreetiliste vigade eeldusel;
- logistiline diskriminantanalüüs eeldusel, et esimest liiki viga on väiksem kui 0,05;
- eksperthinnang.

Diskriminantanalüüsi puhul võeti otsustamisreegli tuletamisel aluseks „kindlasti residentide“ kogum – isikud, kes olid püsielanikena loendatud ja kuulusid ka rahvastikuregistri andmetel Eesti alaliste elanike hulka – ning „kindlasti mitteresidentide“ kogum – inimesed, kes olid lahkunutena loendatud ja ei olnud ka rahvastikuregistri andmetel Eesti elanikud.

Vastavalt REL-i teadusnõukogu ekspertide soovitusel püüti hindamise käigus minimeerida isikute ekslikult residentide hulka arvamise tõenäosust. Selleks kasutati kahel juhul ebasümmeetrilist otsustusreeglit: piirati isiku ekslikult residendiks lugemise tõenäosus (1. liiki viga) etteantud tõenäosusega, mis enamiku soo- ja vanuserühmade puhul oli 0,02 ja kokkuvõttes väiksem kui 0,05. Loomulikult suurenes selle arvel teist liiki vea, isiku ekslikult mitteresidendiks lugemise, tõenäosus.

Kõigil juhtudel olid aluseks samad lähteandmed ehk analüüsitava isikute aktiivsus administratiivsetes registrites 2011. aasta jooksul. Diskriminantanalüüsi mudelites kasutatakse lähtetunnuste seast automaatselt valitud optimaalset (parima eristusvõimega ja mittekorreleeritud) tunnustekomplekti. See komplekt osutus lineaarse ja logistilise diskriminantanalüüsi korral mõnevõrra erinevaks. Eksperthinnangu koostamise lähtuti tunnuste valikul registrite sisulisest analüüsist. Kõigi otsustusreeglite tulemused olid võrdlemisi sarnased, kuid nn ebasümmeetriliste kriteeriumite korral oli residentidena määratud isikute hulk mõnevõrra väiksem kui sümmeetriliste kriteeriumite korral.

Integreeritud kriteeriumi saamiseks kasutati loetletud viit eeskirja samaväärselt ja loeti residentideks need isikud, kelle puhul seda valikut toetas neli kriteeriumit viiest loetletust.

Selgus, et niisuguse valiku korral osutus residendiks 46% kõigist analüüsitavatest isikutest. Ülejäänud isikud vaadeldavast isikute hulgast loeti mitteresidentideks. Esimest liiki (kaasamise) vea hinnatav tõenäosus oli sel juhul 0,044. Seega moodustaks oletatav ekslikult residentideks määratud isikute arv kogu elanikkonnast vähem kui 0,1%, ekslikult mitteresidentideks määratud isikute arv jääks alla 0,5% koguelanikkonnast.

Kõigi nende isikute kohta, kes residentideks loeti, on võimalik registritest saada suur osa loendusel küsitavast infost. Seega põhimõtteliselt oleks võimalik nad lisada ka loendatud isikute hulka (nagu seda tehakse näiteks Lätis). Eesti loendusmetoodika, mis nägi ette registre kasutamist loendusel, sh ka imputeerimisel, seda võimaldaks. Tuleb aga tõdeda, et rahvusvaheliselt ei ole alakaetuse parandamiseks ühtset metoodikat välja töötatud ja soovitatud (ilmselt pole see lähiajal võimalikki), seetõttu esitab enamik riike loendustulemused parandamata kujul. Nii otsustati toimida ka Eesti Statistikaametis.

Hindamise tulemused

Esitatud metoodika tulemusena hinnati loendamata residentideks ligikaudu 28 000 isikut.

Loenduskogumi suurus on 1 294 455 isikut. Seega on üldkogumi hinnang üle 1 322 000 isiku. See tulemus on väga hästi kooskõlas jooksva rahvastikuarvestustega (arvestab ka rännet).

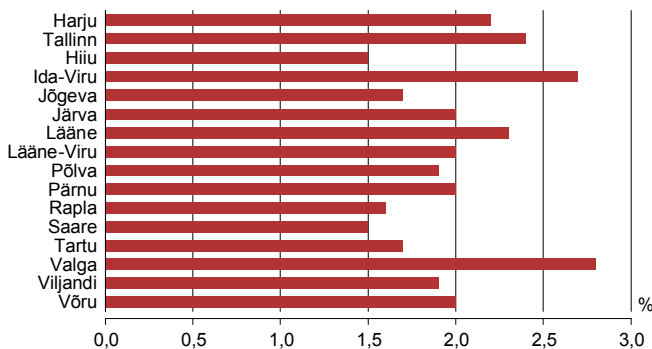
Alakaetuse määr on $28\,000 / 1\,322\,000 = 2,1\%$ ja kaetuse määr vastavalt 97,9%.

Alakaetus maakondade kaupa

Alakaetus ei ole päris ühtlane, vaid varieerub maakonniti peaaegu kaks korda. Selgub, et alakaetus on keskmisest suurem Valgemaal ja Ida-Virumaal, s.o piiriäärsetes maakondades, kus elab suhteliselt rohkem teiste rahvuste esindajaid (joonis 1).

Joonis 1. Alakaetuse määr maakondades ja Tallinnas, 2011

Figure 1. Rate of under-coverage in counties and in Tallinn city, 2011

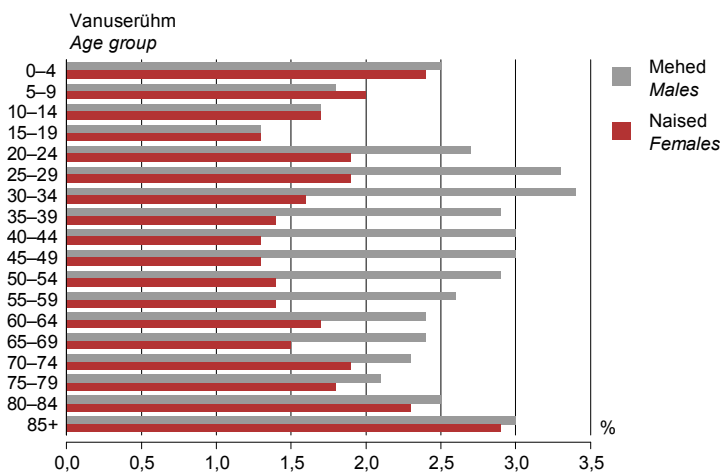


Alakaetus soo- ja vanuserühma kaupa

Selgub, et kõige väiksem on kooliõpilaste alakaetus, suurenedes järsult noorukite puhul (joonis 2). Tööealiste isikute seas on meeste alakaetus märksa suurem kui naiste puhul. See on seletatav meeste üldiselt väiksema motiveeritusega osaleda uuringutes ja võimalik, et osalt ka ühe osa meeste hoolimatusega (sotsiaalsete) normide suhtes. On ka võimalik, et loendamata jättis end osa pendelmigrantidest, kes töötavad väljaspool Eestit, kuid kelle perekond on Eestis ja kes veedavad siin suurema osa oma töövabast ajast.

Joonis 2. Alakaetuse määr soo- ja vanuserühma järgi, 2011

Figure 2. Rate of under-coverage by sex and age group, 2011

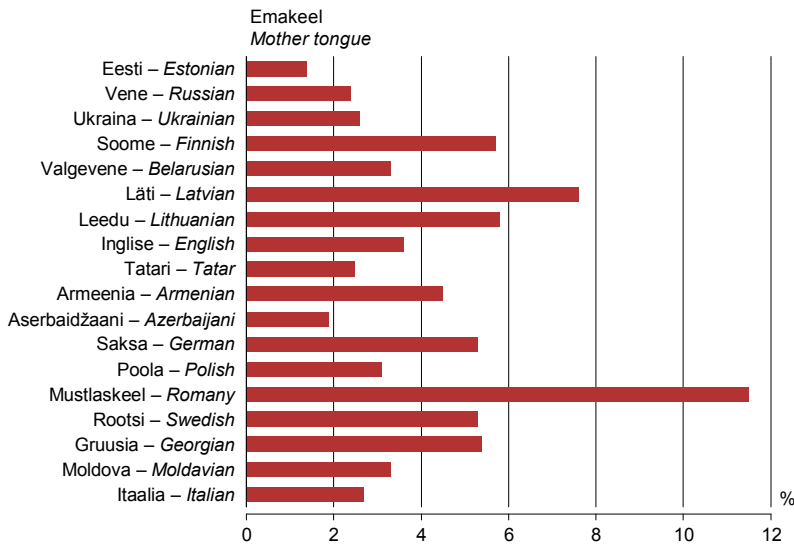


Alakaetus sõltuvalt kodakondsusest, rahvusest ja emakeelest

Selgitamaks, kes on need inimesed, kes loendamata jäid, vaatlesime nende jagunemist rahvuse, kodakondsuse ja emakeele järgi – need kõik on tunnused, mille kohta on praeguseks ametlikud andmed avaldatud ja mis on olemas ka rahvastikuregistris. Kõige kõnekam oli jaotus **emakeele järgi** (joonis 3), kuigi selle tunnuse kohta ei olnud rahvastikuregistris andmeid ligi viiendikul residendiks määratud isikutest. Kõigist loendamata jäänud inimestest, kelle kohta emakeel oli märgitud, oli ligi pooltel (46%) emakeeleks riigikeelest erinev keel – ilmselt ei ulatunud nendeni loenduse teavituskampaania. Siinjuures olulisel osal nendest inimestest erines emakeel ka teisest teavituskaleest – vene keelest.

Joonis 3. Alakaetuse määr emakeele järgi, 2011^a

Figure 3. Rate of under-coverage by mother tongue, 2011^a

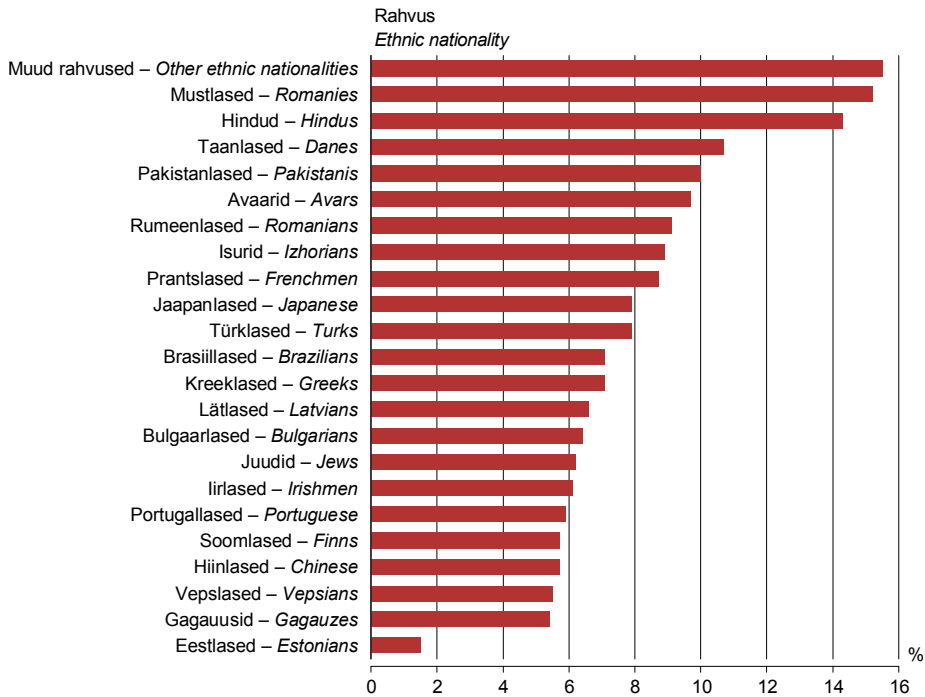


^a Joonisel on kõik vähemalt 200 esindajaga emakeelerühmad järjestatud vastavalt nende arvukusele elanikkonnas.

^a The figure includes all mother tongues with at least 200 representatives, presented in a descending order based on the share of the mother tongue in the population.

Vaadates, missuguste **rahvuste** esindajad on jäänud loendamata (joonis 4), selgub, et nendest 13%-l küll puudub rahvuse määratlus rahvastikuregistris, kuid neist, kellel on rahvus määratud, on eestlasi vaid 55%. Seega on suur osa, tõenäoliselt vähemalt pooled, loendamata jäänud residentidest mitte-eestlased. See seletab ka nende suurema sageduse Ida-Virumaal, kus eestlased on vähemuses. Ilmselt on lätlaste arvestatav osa elanikkonnas ja nende vähene haaratus loendusel põhjuseks, miks Valgamaa kaetus oli ülejäänud maakondadest madalam.

Joonis 4. Alakaetuse määr rahvuse järgi, 2011^a
Figure 4. Rate of under-coverage by ethnic nationality, 2011^a



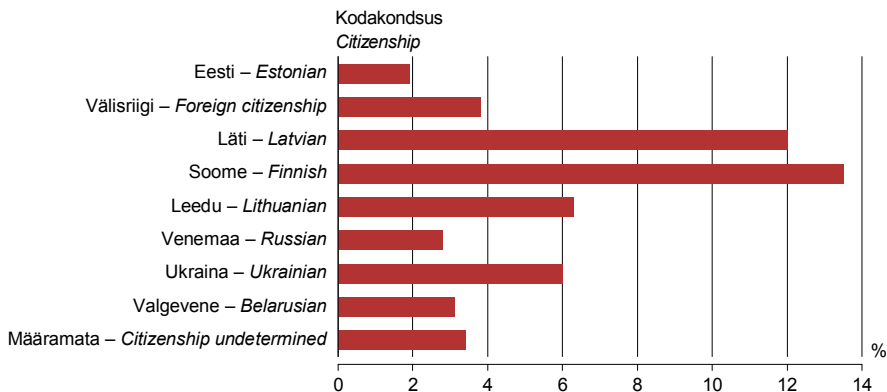
^a Joonisel on kõik vähemalt 25 esindajaga rahvusrühmad, kelle puhul alakaetus ületab 5%.

^a The figure includes all ethnic nationalities with at least 25 representatives and with an under-coverage rate above 5%.

Joonisel 4 on võrdluseks esitatud ka eestlased, kelle puhul alakaetuse määr oli madalaim – 1,5%. Siit on näha, et sagedamini kippusid loendusest hoiduma pigem eksotiliste, Eestist ruumiliselt ja kultuuriliselt kaugemate maade esindajad.

Kodakondsuse mõju kaetusele on mõnevõrra väiksem, Eesti **kodakondsuseta** on veerand loendamata inimestest. Joonisel 5 on esitatud nende riikide kodanikud, keda Eestis on vähemalt 1000.

Joonis 5. Alakaetuse määr kodakondsuse järgi, 2011
Figure 5. Rate of under-coverage by citizenship, 2011



Kokkuvõte

2011. aasta rahvaloenduse kaetuse määr on 97,9%, st väga lähedal REL-kogu kehtestatud kontrollarvule 98%, mis tähendaks väga head tulemust.

Loendamata jäänud residentidest on arvestuslikult umbes pooled mitte-eestlased, kelle emakeel pole eesti keel. Nende hulgas on omakorda väga paljude erinevate riikide ja keelte esindajaid. Ilmselt on nende seas niihästi välismaa(de)lt pärinevaid vahetusüliõpilasi kui ka võõrtöölisi, kes loendusreeglite kohaselt on (olid) loendushetkel Eesti elanikud.

Nähtavasti ei olnud mitte-eestlastel, eriti neil, kes ka eesti keelt ei valda, motivatsiooni loendusel osaleda, seda enam, et nende kodumaal võis loendus olla juba toimunud.

Peale selle oli loendamata jäänud hulk tööealisi mehi, kelle seas võis olla niihästi kodust eemal viibinuid, pendelmigrante kui ka inimesi, kes lihtsalt ei soovinud suhelda (just viimaste kohta laekus loendusjärgsel perioodil mitmeid teateid). Väga eakate loendamata jäänute puhul võis mitteloendamise põhjuseks olla nende olematu aktiivsus.

Seega ei leidnud kinnitust hüpotees, et valdavalt jäid loendamata Eestist lahkunud eestlased, kes varasemate sidemete toel kasutavad Eesti riigi teenuseid ja seetõttu esinevad Eesti registrites. Niiuguseid inimesi saab analüüsitava isikute seas olla vaid alla poole.

Allikad Sources

Euroopa komisjoni määrus. (2010). (EL) nr 1151/2010, artikkel 2, lõige (8).

Joint UNECE-UNFPA Training Workshop on Censuses Using Registers.(2012). 21 May. Geneva

Statistikaamet. (2012). [e-andmebaas] <http://pub.stat.ee/px-web.2001/Database/Rahvaloendus/REL2011/REL2011.asp>

Tiit, E-M., Vähi, M. (2012). Rahvaloendajate tegevus küsitluse järel. – Eesti Statistika Kvartalikirj. *Quarterly Bulletin of Statistics Estonia*, nr 2, lk 102–111. Tallinn: Statistikaamet.

Tiit, E-M., Meres, K., Vähi, M. (2012). Rahvaloenduse üldkogumi hindamine. – Eesti Statistika Kvartalikirj. *Quarterly Bulletin of Statistics Estonia*, nr 3, lk 79–108. Tallinn: Statistikaamet.

Tiit, E-M. (2012). Loendusmeeskond jätkab tööd. – Sirp, 23.08.