

Tartu Ülikool
Loodus- ja täppisteaduste valdkond
Matemaatika ja statistika instituut

Kristina Muhu

**Eesti täiskasvanud rahvastiku tervisekäitumise uuringu
lisavalimi moodustamine**

Matemaatilise statistika eriala
Bakalaureusetöö (9 EAP)

Juhendajad: Imbi Traat, PhD
Natalja Eigo, MSc

Tartu 2019

Eesti täiskasvanud rahvastiku tervisekäitumise uuringu lisavalimi

moodustamine

Bakalaureusetöö

Kristina Muhu

Lühikokkuvõte.

Bakalaureusetöö eesmärk on lisavalimimahu leidmine Eesti täiskasvanud rahvastiku tervisekäitumise uuringu jaoks, mis on regulaarne uuring, kuid mille valimimahust ei piisa esinduslike tulemuste saamiseks väikestes gruppides. Töö käigus leitakse lisavalimimaht nii, et etteantud täpsusega tulemused saadakse kaheksa maakonna rühma kaupa, mis on jagatud kahte vanuserühma 16 – 34 ja 35 – 64. Töös antakse ülevaade Eesti täiskasvanud rahvastiku tervisekäitumise uuringust ning ühisuuringust Finbalt Health Monitor. Töö teises osas tuletatakse valimimahu arvutamise valemid etteantud täpsuse korral ning leitakse vastavad valimimahud 16 kihis. Kolmandas osas täpsustatakse lisavalimimahud arvestades 2016. aasta Eesti täiskasvanud rahvastiku tervisekäitumise uuringu vastamismäärasid.

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

Märksõnad: juhuvalik, kihtvalik, hinnangu täpsus, valimimaht

Compilation of additional sample for the study of the Health Behavior among Estonian Adult Population

Bachelor's thesis

Kristina Muhu

Abstract.

The purpose of the current bachelor thesis is to find an additional sample size for the study of the Health Behavior among Estonian Adult Population. The latter is a regular survey, but its sample size is not sufficient for producing results with good precision in small groups. In the current work an additional sample size is found so that the results with given precision are obtained by eight county groups divided into two age groups 16 – 34 and 35 – 64. The work gives an overview of the related studies – the Health Behavior among Estonian Adult Population and

the Finbalt Health Monitor. In the second part of the thesis, the formulas for calculating sample size are derived with given accuracy and the sample sizes are found in 16 groups. In the third part, additional sample sizes are adjusted for the response rates of the 2016 study of the Health Behavior among Estonian Adult Population.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematic

Keywords: random sampling, stratified sampling, precision, sample size

Sisukord

Sissejuhatus	5
1 Eesti täiskasvanud rahvastiku tervisekäitumise uuring	7
1.1 Ülevaade uuringu metoodikast	7
1.2 Finbalt Health Monitor ühisuuring	8
1.3 Laiendatud valimiga uuringu vajadus	9
2 Valimimahu määramine	10
2.1 Valikudisain ja hinnangu täpsus	10
2.2 Valimimahu tuletamine etteantud täpsuse korral	11
2.3 Näited valimimahu arvutustest	13
2.4 Hinnangute täpsusest toimunud uuringus	14
2.5 Reeglid valimimahu arvutamiseks	18
3 Hindamine kihtvaliku korral ja lisavalimi koostamine	20
3.1 Hindamine kihtvaliku korral	20
3.2 Vastamismäärad toimunud uuringus	21
3.3 Lisavalimi koostamine	22
Kokkuvõte	25
Viited	26
Lisa	26
Tabelid	27
Litsents	32

Sissejuhatus

Eesti rahvastiku tervis on maakonniti väga erinev. Tervise Arengu Instituut koostöös Sotsiaalministeeriumi ja Statistikaametiga on koostanud 2018. aastal ülevaated maakondade tervise ja heaolu kohta. Näiteks on Lääne-Virumaal suitsetajaid alates 16-aastaste ja vanemate hulgas 30,1%, Tartumaal on seevastu suitsetajate osakaal 19,1%. Ülekaalulisi ja rasvunuid vanuses 16–74 aastat on kõige rohkem Põlvamaal, 61,9% ning kõige vähem Harjumaal, 41,3% [1].

Selleks, et maakondades saaks tervishoiusüsteemi ja -teenuseid õigesti planeerida, on vaja regulaarselt rahvastiku tervist maakondades seirata. Tervise Arengu Instituut on aastate jooksul läbi viinud mitmeid tervisekäitumise uuringuid: Eesti Terviseuuring, Eesti täiskasvanud rahvastiku tervisekäitumise uuring, Paikkonna tervisemõjurite uuring jne. Eesti täiskasvanud rahvastiku tervisekäitumise uuring on pikaajaline uuring, mis katab suuremas osas kõik tervisekäitumisega ja tervisega seotud valdkonnad, küll aga ei ole uuringu tulemused esinduslikud üksikute maakondade kaupa vaid piirkonniti (Kesk-, Kirde-, Lõuna-, Lääne- ja Põhja-Eesti).

Käesoleva töö eesmärgiks on lisavalimimahu leidmine Eesti täiskasvanud rahvastiku tervisekäitumise uuringu jaoks. Töö käigus leitakse lisavalimimaht nii, et esinduslikud tulemused saadakse maakonna gruppide kaupa, mis on jagatud kahte vanuserühma.

Töö esimeses osas antakse lühike ülevaade Eesti täiskasvanud rahvastiku tervisekäitumise uuringust ning selle meetodikast. Samuti kirjeldatakse lühidalt Baltiriikide ja Soome ühisuuringut Finbalt Health Monitor, mille raames Eesti tervisekäitumise uuringut aastatel 1990–2010 läbi viidi. Teises osas keskendutakse valimimahu määramisele: tuletatakse valimimahu arvutamise valemid etteantud täpsuse korral, sealjuures tuuakse selgitavaid näiteid ning Eesti täiskasvanud rahvastiku tervisekäitumise 2016. aasta uuringu andmeid kasutades antakse lühike ülevaade fikseeritud tunnuse korral hinnangute täpsustest. Viimaks fikseeritakse sobiv reegel valimimahu arvutamiseks.

Bakalaureusetöö kolmandas osas kirjeldatakse lühidalt hindamist kihtvaliku korral ning Eesti täiskasvanud rahvastiku tervisekäitumise 2016. aasta uuringu vastamismäärasid. Vastamismää-

rade abil täpsustatakse nõutavaid valimimahtusid selleks, et teha lõplikud ettepanekud lisavalimi koostamise jaoks.

Töö on kirjutatud ja tabelid koostatud kasutades tekstitöötlusprogrammi \LaTeX . Kõik arvutused on tehtud kasutades rakendustarkvara R . Töös kasutatavad andmed pärinevad 2016. aasta Eesti rahvastiku tervisekäitumise uuringust.

Autor tänab juhendajaid Imbi Traati ja Natalja Eigt täpsustavate kommentaaride ja täiendavate nõuannete eest.

1. Eesti täiskasvanud rahvastiku tervisekäitumise uuring

Eesti täiskasvanud rahvastiku tervisekäitumise uuringut (edasipidi TKU) on läbi viidud alates aastast 1990 paarisaastatel Finbalt Health Monitor ühisuuringu raames, kusjuures alates 2012. aastast korraldavad ühisuuringus osalenud riigid (Leedu, Läti, Soome ja Eesti) ise vastavat uuringut. Küll aga lähtutakse siiani ühisuuringu metoodikast ja kokkulepetest selleks, et võrreldavus varasemate tulemustega ei kaoks. Peamised uuritavad valdkonnad on toitumine, suitsetamine, alkoholi tarbimine, kehaline aktiivsus, liikluskäitumine, lisaks veel terviseseisund koos tervise enesehinnanguga, arstiabi ning ravimite kasutamine [2].

Uuringut teostab alates 2004. aastast Tervise Arengu Instituudi epidemioloogia ja biostatistika osakond, enne seda viis uuringut läbi Eesti Tervisekasvatuse Keskus. Uuringut on rahastatud Südame- ja veresoonkonnahaiguste ennetamise riikliku strateegia raames ning Rahvastiku tervise arengukava 2009–2020 ning täiendavalt IUT5-1 raames, erinevatel aastatel ka Sotsiaalministeerium [2].

TKU põhiväärtus on selle pikaajalisus. Nii saame jälgida tervisekäitumise muutumist läbi aastate. Vaadates Tervisestatistika ja terviseuuringute andmebaasist andmeid „TKU23: Alkoholi tarvitamise sagedus viimase 12 kuu jooksul soo ja piirkonna järgi“ selgub, et 2004. aastal oli vähemalt mõned korrad nädalas alkoholi tarbijaid Kirde-Eestis 19,1% (Eesti keskmine 23,4%), aga 2014. aastal 12,3% (Eesti keskmine 22,0%) [3].

1.1. Ülevaade uuringu metoodikast

Eesti täiskasvanud rahvastiku tervisekäitumise uuringu metoodikat kirjeldatakse põhjalikult igas vastava aasta aruandes. Aastatega on toimunud uuringu metoodikas mõned muudatused, näiteks oli küsimustiku täitmine internetis 2016. aasta TKU jaoks esmakordne. Järgnev metoodika kirjeldus põhinebki 2016. aasta aruandel. Alates 2012. aastast tellitakse rahvastikuregistrist kihitatud juhuvalim, varem (2004–2010) telliti lihtne juhuvalim. Valimi suurus on olnud kõikidel aastatel 5000 isikut. Eestis elavad 16.–64. aastased isikud (uuringus täiskasvanud) on registris sorteeritud soo, vanuse, rahvuse ja maakonna põhjal, valimisse on võetud kindla sam-

muga soovitud kogus isikuid.¹ Lisaks kõigele eelnevale küsitakse rahvastikuregistrist postiaadress, meiliaadress ja riigiaadress (eesti.ee meiliaadress). Kõikidele aadressidele, mis valimisse sattunud isikul olemas on, saadetakse uuringukutse koos küsimustiku internetiaadressiga, postiaadressile saadetakse postiküsitlus [2].

2016. aasta TKU küsimustik (nagu ka eelmistel aastatel) oli koostatud nii eesti kui vene keeles. Venekeelne küsimustik saadeti neile, kelle emakeeleks oli rahvastikuregistris vene keel või muu keel või kui isiku ees- ja perenimi oli venepärane. Küsimustikus oli 91 küsimust, mis jagunesid järgmisteks teemadeks: toitumine, kehaline aktiivsus ja kaaluindeks, alkoholi ja narkootiliste ainete tarbimine, suitsetamine, liiklus- ja seksuaalkäitumine, tervisehäireid ennetavad meetmed, tervise seisund, arstiabi ja ravimite kasutamine. Küsimustikule vastas 2751 isikut (vastamismäär 55%), kusjuures eristatakse neid, kes ei vastanud (2121 isikut) ja neid, kes ei saanud vastata (128 isikut), näiteks vale aadressi tõttu [2].

1.2. Finbalt Health Monitor ühisuuring

Eesti täiskasvanud rahvastiku tervisekäitumise uuring oli kuni 2010. aastani osa ühisuuringust Finbalt Health Monitor ehk Finbalt (osalesid Eesti, Läti, Leedu ja Soome). Ühisuuringu eesmärk oli jälgida tervisega seotud käitumist, tavasid ja elustiili Soomes ja Baltimaades. Kõigis neljas riigis koguti andmed ristlõikelise uuringu raames [4]. Ühisuuring otsustati lõpetada, sest Euroopa Liit valmistas sel ajal ette uut tervise jälgimissüsteemi, mis oleks hakanud olema liikmesriikidele kohustuslik. Kõik uuringus osalenud riigid jätkasid iseseisvalt tervisekäitumisuuringutega [5].

Ühisuuringu eelkäijaks võib pidada Soomes alates 1978. aastast läbi viidavat Soome täiskasvanud elanikkonna tervisekäitumise uuringut (AVTK). Selles uuritakse suitsetamist, toitumisharjumusi, muid tervisega seotud faktoreid ning muuhulgas ka sotsiaaldemograafilisi ja uuringuga seotud psühhosotsiaalsed tegureid. Soomes läbi viidava uuringuga sarnase uuringuga alustati Eestis 1990. aastal, Leedu ühines 1994. aastal ja Läti alles 1998. aastal. Kogutud andmete põhjal on võimalik võrrelda projektis osalenud riikide tervisekäitumist ja sellega seotud tegureid [6].

¹E-kiri Reilega, 21.12.2018

1.3. Laiendatud valimiga uuringu vajadus

Selles alapeatükis kirjeldatu põhineb meilivahetusel Tiia Perteliga, kes on Tervise Arengu Instituudi Tervise ja heaolu edendamise keskuse juht.

2011. aastal viidi läbi paikkonna tervisemõjurite uuring (PTMU), mille eesmärgiks oli rohkem teada saada Eesti maakondade elanike tervisekäitumisest ja seda mõjutavatest teguritest. Uuringu tulemusi saab kasutada kohalike elanike vajadustele vastavate teenuste planeerimiseks, tulevikus esile kerkida võivate riskide ja vajaduste prognoosimiseks ning kohaliku poliitika kujundamiseks. Uuringu andmed koguti postiküsitluse meetodil, kusjuures ankeedi koostamisel lähtuti juba siis TKU ankeedist; küsitlus saadeti 9668 elanikule vanuses 15–70 eluaastat 77 kohalikust omavalitsusest. Vastajaid oli 4239, vastamismäär 43,8%.

Kuna TKU tulemused ei ole maakonna tasemel usaldusväärsed, siis soovitakse TKU valimit suurendada. Seega lisavalimiga TKU eesmärgiks on maakondade (võimalusel suuremate kohalike omavalitsuste) või osaliselt maakondade gruppide põhiselt andmete analüüsi ja esitamise võimalus. Maavalitsuste kaotamise tõttu on rahvatervise ja turvalisusega seotud ülesanded kohalikele omavalitsustele ühiselt täitmiseks. 2018. aastal tuli kohalikel omavalitustel ühiselt koostada maakonna tervise- ja heoluprofiil, millega tuleb arvestada maakonna arengustrateegia koostamisel. Kuna profiile tuleb mõne aasta pärast uuendada, siis on hea, kui kohalikud omavalitsused saavad juba siis maakonnapõhisemaid tervisekäitumise andmeid kasutada.

Maakonnapõhise uuringu läbiviijaid huvitavad TKU-ga valdavalt samad teemad. Küll aga leidub üksikuid lisaküsimusi, mis võiksid maakondade jaoks TKU-s kajastamist leida (näiteks täpsustada valikute ja võimaluste kättesaadavust või kaasatust otsustusprotsessidesse). Arvatakse, et lisavalimiga uuringut (suure valimimahuga) on mõistlik 6–10 aasta tagant läbi viia. Laiendatud valimiga uuringut rahastatakse programmi „Local Development and Poverty Reduction“ Euroopa Majanduspiirkonna ja Norra toetused 2014–2021 raames.

2. Valimimahu määramine

2.1. Valikudisain ja hinnangu täpsus

Üldkogumi parameetri hinnangu täpsust mõõdetakse hinnangu dispersiooniga või sellest moodustatud teiste suurustega (standardhälve, variatsioonikoeffitsient, usaldusvahemik). Hinnangu dispersioon sõltub aga kasutatud valikudisainist, sealhulgas valimimahust. Punktis 1.1 kirjeldatud viisil valimit võttes imiteeritakse võrdelise paigutusega lihtsat juhuslikku kihtvalikut. Erinevate valimi võtmise meetodite (näiteks lihtne juhuvalik, süstemaatiline valik, kihtvalik, klastervalik jne) kirjeldused on leitavad I. Traadi ja J. Inno õpikust "Tõenäosuslik valikuuring"[7].

Kuigi TKU kihtides toimub süstemaatiline valik, lähendatakse hinnangu dispersiooni lihtsa juhusliku valiku dispersiooniga ja kasutatakse sellele vastavat dispersioonihinnangut. Süstemaatilise valiku korral puudub nihketa dispersioonihinnang. Lihtsa juhusliku valiku dispersioonihinnang on siin üldlevinud lähend [7, lk 112]. Veelgi enam lihtsustades võime punktis 1.1 kirjeldatud viisil valimit võttes lugeda selle lihtsaks juhuslikuks valimiks (st loobume kihtidest), kuna kaasamistõenäosused on kõigil objektidel samad.

Antud uuringus hinnatakse põhiliselt osakaale, enamasti väljendatuna protsentides. Olgu osakaalu p hinnang \hat{p} . Osakaalu hinnangu dispersioon avaldub lihtsa juhusliku tagasispanekuta valiku korral kujul:

$$D\hat{p} = \frac{p(1-p)}{n} \cdot \left(1 - \frac{n}{N}\right), \quad (1)$$

kus n on valimi suurus, N on üldkogumi suurus ja $1 - \frac{n}{N}$ on lõpliku üldkogumi parandus, mis praktikas on sageli ligikaudu 1, sest valimimaht n moodustab väikese osa üldkogumi mahust N [7, lk 95].

Sageli kasutatakse hinnangu täpsuse iseloomustamiseks variatsioonikoeffitsienti,

$$cv(\hat{p}) = \frac{\sqrt{D\hat{p}}}{p},$$

mis on ühikuvaba suurus ja sobib seetõttu paremini erinevate hinnangute võrdlemiseks. Variatsioonikoeffitsienti nimetatakse ka suhteliseks veaks, mida arvutatakse sageli ka kui poole usaldusvahemiku osakaalu suurusest p [9, lk 97-98].

2.2. Valimimahu tuletamine etteantud täpsuse korral

Valimimahu fikseerimiseks on vaja kindlaks määrata lubatav viga. Kuna igas uuringus on nii suuri kui väikeseid hinnanguid \hat{p} , on õigem fikseerida suhteline viga, mis näitab, kui suur võib vea osakaal hinnangust olla.

Järgnev lõik tugineb S. L. Lohr'i raamatule „*Sampling: Design and Analysis*“ [8, lk 39-41]. Tähistagu U üldkogumit ja s valimit. Olgu y uuritav tunnus ja $\bar{y}_U = \sum_{k \in U} \frac{y_k}{N}$ üldkogumi keskmine. Selle nihketa hinnanguks on lihtsa juhusliku valiku korral valimi keskmine $\bar{y} = \sum_{k \in s} \frac{y_k}{n}$. Soovitud täpsust ehk hinnangu \bar{y} erinevust suurusest \bar{y}_U kirjeldatakse tõenäosuslikult,

$$P(|\bar{y} - \bar{y}_U| \leq e) = 1 - \alpha,$$

kus e on hinnangu viga. Usaldusvahemike keeles on suurus e pool usaldusvahemiku laiust ja $1 - \alpha$ on usaldusnivoo. Suure valimimahu korral on \bar{y} normaaljaotusega ja siis avaldub e kujul:

$$e = z_{\alpha/2} \sqrt{D\hat{p}} = z_{\alpha/2} \sqrt{\left(1 - \frac{n}{N}\right) \frac{S}{\sqrt{n}}}, \quad (2)$$

kus $z_{\alpha/2}$ on normaaljaotuse $N(0, 1)$ $\alpha/2$ -täiendkvantiil ning

$$S^2 = \frac{1}{N} \sum_{k \in U} (y_k - \bar{y}_U)^2.$$

Kui y_k on binaarne, siis $\bar{y}_U = p$, $y_k^2 = y_k$ ja S^2 avaldub kujul:

$$\begin{aligned} S^2 &= \frac{1}{N} \sum_{k \in U} (y_k - p)^2 = \frac{1}{N} \sum_{k \in U} (y_k^2 - 2y_k p + p^2) = \\ &= \frac{1}{N} \left(\sum_{k \in U} y_k - Np^2 \right) = \frac{1}{N} (Np - Np^2), \end{aligned}$$

millest

$$S^2 = p(1 - p). \quad (3)$$

Seega osakaalu hinnates avaldub viga samuti valemiga (2), kus S tuleb valemist (3). Ilma populatsiooni paranduseta on viga e kujul:

$$e = z_{\alpha/2} \sqrt{D\hat{p}} = z_{\alpha/2} \frac{S}{\sqrt{n}}. \quad (4)$$

Avaldades võrdusest (2) tundmatu n saame [8, lk 39-41]:

$$n = \frac{z_{\alpha/2}^2 S^2}{e^2 + \frac{z_{\alpha/2}^2 S^2}{N}}. \quad (5)$$

Analoogiliselt avaldame võrdusest (4) tundmatu n :

$$\sqrt{n} = \frac{z_{\alpha/2} S}{e},$$

millest

$$n = \frac{z_{\alpha/2}^2 S^2}{e^2}. \quad (6)$$

Valemitest (5) ja (6) saab leida vastavalt lubatava vea suurusele vajaliku valimimahu n . Sageli on vaja n leida vastavalt etteantud suhtelisele veale c . Nõudes, et $e = c \cdot p$, kus $0 \leq c \leq 1$, soovime, et viga ei oleks suurem kui $c \cdot 100\%$ osakaalust p . Vastavad n avaldised leiame seosest (5),

$$n = \frac{z_{\alpha/2}^2 p(1-p)}{(cp)^2 + \frac{z_{\alpha/2}^2 p(1-p)}{N}} = \frac{z_{\alpha/2}^2 (1-p)}{c^2 p + \frac{z_{\alpha/2}^2 (1-p)}{N}} \quad (7)$$

ja seosest (6),

$$n = \frac{z_{\alpha/2}^2 p(1-p)}{(cp)^2} = \frac{z_{\alpha/2}^2 (1-p)}{c^2 p}. \quad (8)$$

Kahe osakaalu p_1 ja p_2 võrdlemisel uuritakse nende usaldusvahemikke. Suuruse p usaldusvahemik ligikaudsel usaldusnivool $1 - \alpha$ avaldub seosega

$$UI(p) = \hat{p} \pm \hat{e}, \quad (9)$$

kus \hat{e} on antud valemitega (2) või (4), milles S on asendatud hinnanguga $\hat{S} = \hat{p}(1 - \hat{p})$. Kui usaldusvahemikud ei kattu, on osakaalud erinevad. Alternatiivselt saab erinevuse tuvastamiseks usaldusnivool $1 - \alpha$ leida usaldusvahemiku kahe osakaalu vahele $p_1 - p_2$. Olgu nüüd kahe osakaalu p_1 ja p_2 hinnangud sõltumatud \hat{p}_1 ja \hat{p}_2 . Vahe $p_1 - p_2$ hinnanguks olgu $\hat{p}_1 - \hat{p}_2$. Dispersiooni hinnanguks on siis [9, lk 105]:

$$\hat{D}(\hat{p}_1 - \hat{p}_2) = \frac{\hat{p}_1(1 - \hat{p}_2)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_1)}{n_2}.$$

Suurte n_1 ja n_2 korral on $\hat{p}_1 - \hat{p}_2$ ligikaudu normaaljaotusega:

$$\hat{p}_1 - \hat{p}_2 \sim N \left(p_1 - p_2, \frac{\hat{p}_1(1 - \hat{p}_2)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_1)}{n_2} \right).$$

Usaldusvahemik usaldusnivooga ligikaudu $1 - \alpha$ avaldub kujul:

$$UI(p_1 - p_2) = \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}. \quad (10)$$

Kui intervall ei sisalda 0, siis on (tegelikud) p_1 ja p_2 erinevad [9, lk 105].

2.3. Näited valimimahu arvutustest

Fikseerides osakaalu p ning lubatava vea e , saame leida nõutava valimimahu n , kui $z_{\alpha/2} = 1,96$. Suurustega p ja e on määratud suhteline viga $c = \frac{e}{p}$. Oletame, et N on nii suur, et lõpliku populatsiooni parandus ei oma tähtsust. Valimimahu n leiame kasutades valemit (6).

Kui $p = 0,5$, siis $S^2 = p(1 - p) = 0,25$. Olgu $e = 0,03$, siis valemi (6) korral

$$n = \frac{1,96 \cdot 0,25}{0,03^2} = 1067$$

ja suhteline viga on $c = 0,06$. Teisisõnu, hinnates osakaalu $p = 0,5$ ja võttes $n = 1067$, saame hinnangu $\hat{p} \approx 0,5$ ja usaldusvahemiku laiuse ligikaudu $0,03$. Hinnatud suhteline viga on sellisel juhul ligikaudu 6% , tabel 1.

Tabel 1. Valimimaht n ja suhteline viga c fikseeritud vea e ja osakaalu p korral, valem (6), kus $z_{\alpha/2} = 1,96$.

Osakaal p	Lubatav viga e					
	0,01		0,03		0,05	
	n	c	n	c	n	c
0,01	381	1	43	3	16	5
0,1	3458	0,1	385	0,3	139	0,5
0,5	9604	0,02	1067	0,06	385	0,1

Tabelis 1 on toodud näiteid valimimahu arvutustest. Näeme, et iga fikseeritud p korral lubatava vea e kasvades, valimimaht väheneb. Iga fikseeritud e korral p kasvades valimimaht kasvab.

Paneme tähele, et kindlaks määratud (kuigi väike) fikseeritud viga e moodustab väikesest osakaalust p suure osa. Seetõttu uurime n väärtusi fikseeritud suhtelise vea korral, tabel 2.

Tabelist 2 on näha, et väikese osakaalu $p = 0,01$ korral on väikese suhtelise vea $c = 0,01$ tagamiseks vaja, et valimimaht n oleks väga suur, $n = 3\,803\,184$ (see on kordades suurem, kui Eesti elanike arv). Lubades suhtelisel veal olla suurem, näiteks $c = 0,2$, saame mitu korda väiksema valimimahu $n = 9508$, kuigi ka see on valimimahu mõttes suur. Kui $c = 0,3$ ja $p = 0,01$, siis valimimaht peab olema $n = 4226$. Seega väikeste osakaalude korral tuleb isegi suure valimimahu korral suhteline viga küllaltki suur.

Tabel 2. Valimimaht n , valemist (6), fikseeritud osakaalu p ja suhtelise vea c korral.

Osakaal p	Suhteline viga c			
	0,01	0,1	0,2	0,3
	n	n	n	n
0,01	3 803 184	38 031	9508	4226
0,1	345 744	3458	864	385
0,2	153 664	1537	385	171
0,5	38 416	385	97	43

Kuna praktikas on uuringu ressursid piiratud, tuleb loobuda nõudest hinnata väikesi osakaale suure täpsusega. Valimimahu määramisel tuleb kindlaks määrata osakaalu suurus, mida on vaja hinnata nõutud täpsusega.

2.4. Hinnangute täpsusest toimunud uuringus

Uurime andmestiku TKU16 (2016. aasta TKU) tunnuse *smoind3* realiseerunud osakaale (tabelid 9 ja 10, Lisa) ja vigasid e ja c maakondade kahes vanusegrupis (tabel 3). Tunnus *smoind3* näitab suitsetamise staatust skaalal: 1 – „Igapäevasuitsetaja“, 2 – „Juhusuitsetaja“, 3 – „Endine suitsetaja“, 4 – „Mittesuitsetaja ehk pole kunagi suitsetanud“.

Tabel 3. Igapäevasuutsetajate osakaalu hinnang \hat{p} maakondade vanusegruppides, selle vead e ja c ning usaldusintervall (TKU16).

Igapäevasuutsetajad		TKU16		e	c	$UI(p)$
Maakond	Vanusegrupp	\hat{p}	n			
Harjumaa	16 – 34	0,155	402	0,035	0,226	(0,120; 0,190)
	35 – 64	0,203	763	0,029	0,143	(0,174; 0,232)
Hiiumaa	16 – 34	0,091	11	0,170	1,868	(-0,079; 0,261)
	35 – 64	0,250	14	0,266	0,904	(0,024; 0,476)
Ida-Virumaa	16 – 34	0,301	73	0,105	0,349	(0,196; 0,406)
	35 – 64	0,249	220	0,057	0,229	(0,192; 0,306)
Jõgevamaa	16 – 34	0,333	21	0,201	0,604	(0,132; 0,534)
	35 – 64	0,304	48	0,130	0,428	(0,174; 0,434)
Järvamaa	16 – 34	0,167	18	0,172	1,03	(-0,005; 0,339)
	35 – 64	0,276	60	0,113	0,409	(0,163; 0,389)
Läänemaa	16 – 34	0,182	22	0,161	0,885	(0,021; 0,343)
	35 – 64	0,257	36	0,142	0,553	(0,115; 0,399)
Lääne-Virumaa	16 – 34	0,255	47	0,124	0,486	(0,131; 0,379)
	35 – 64	0,247	82	0,093	0,377	(0,154; 0,340)
Põlvamaa	16 – 34	0,174	23	0,155	0,891	(0,019; 0,329)
	35 – 64	0,179	40	0,119	0,665	(0,060; 0,298)
Pärnumaa	16 – 34	0,326	46	0,135	0,414	(0,191; 0,461)
	35 – 64	0,268	130	0,076	0,284	(0,192; 0,344)
Raplamaa	16 – 34	0,231	27	0,159	0,688	(0,072; 0,390)
	35 – 64	0,273	46	0,129	0,473	(0,144; 0,402)
Saaremaa	16 – 34	0,217	23	0,168	0,774	(0,049; 0,385)
	35 – 64	0,188	51	0,107	0,569	(0,081; 0,295)
Tartumaa	16 – 34	0,178	120	0,068	0,382	(0,110; 0,246)
	35 – 64	0,225	190	0,062	0,243	(0,193; 0,317)

Valgamaa	16 – 34	0,238	21	0,182	0,765	(0,056; 0,420)
	35 – 64	0,162	39	0,115	0,710	(0,047; 0,277)
Viljandimaa	16 – 34	0,259	27	0,165	0,637	(0,094; 0,424)
	35 – 64	0,149	69	0,084	0,564	(0,065; 0,233)
Võrumaa	16 – 34	0,050	21	0,093	1,86	(-0,043; 0,143)
	35 – 64	0,213	61	0,103	0,484	(0,110; 0,316)

Tabelis 3 vaatleme igapäevasuitsetajate osakaalu hinnangut \hat{p} . Uurime võrreldavust vanusegruppides maakondade vahel. Hinnangu viga e on arvatud valemi (2) abil ja usaldusvahemik $UI(p)$ on leitud valemiga (9). Valimimaht n on maakonnas vastavas vanusegrupis vastanute arv tabelist 11 (Lisa) ning valemis vajalik N on tabelist 13 (Lisa). Paneme tähele, et lõpliku populatsiooni parandus on erinevate maakondade ja vanusegruppide korral väike (ligikaudu 1), näiteks Hiiumaa esimese vanusegrupi korral $\sqrt{(1 - \frac{11}{2293})} \approx 1$ või Lääne-Virumaa teise vanusegrupi korral $\sqrt{(1 - \frac{82}{23659})} \approx 1$.

Tabel 3 osutab, et osakaalu hinnangute korral on enamikel juhtudel suhteline viga liiga kõrge, mis tähendab, et vea osakaal hinnangus on liiga suur, mõnel juhul ületab 100%. Koostame analoogiliselt tabelile 3 uue tabeli, kus maakonnad on ühendatud (tabel 4), milles vajalikud valimimahud on tabelist 12 (Lisa). Nüüd on suhtelised vead vähematel juhtudel üle 40%, siiski on juba 30% vea osakaal hinnangus liiga suur.

Kui soovime nüüd sama vanusegrupi erinevate maakondade osakaale eristada, vaatame usaldusintervalle. Paneme tähele, et selliste tulemuste korral nagu tabelites 3 ja 4, saame kahte tegelikku osakaalu eristada alles väga suure erinevuse korral. Kuigi igapäevasuitsetajate hinnangulised osakaalud on Läänemaa ja Jõgevamaa esimeses vanusegrupis erinevad, vastavalt 0,182 ja 0,333 (tabel 3), ei saa me väita, et tegelikud osakaalud on erinevad, sest usaldusvahemikud (0,021; 0,343) ja (0,132; 0,534) kattuvad.

Leiame eelnevate hinnatud osakaalude korral osakaalude vahele usaldusvahemiku, valem (10), saame: $UI(p_1 - p_2) = -0,151 \pm 0,258 = (-0,409; 0,107)$. Kuna vahe usaldusvahemik sisaldab nulli, siis tegelike igapäevasuitsetajate osakaalude erinevust Läänemaa ja Jõgevamaa

esimeses vanusegrupis väita ei saa.

Võimalik on teha järeleandmisi usaldusnivoos. Kui $1 - \alpha = 0,9$, siis $z_{\alpha/2} = 1,64$ ning usaldusvahemikud muutuvad kitsamaks. Usaldusvahemikud osakaalude 0,182 ja 0,333 korral on sellisel juhul (0,047; 0,317) ja (0,165; 0,501), need on kitsamad, kuid kattuvad. Seega ka $\alpha = 0,1$ korral ei saa öelda, et tegelikud osakaalud on erinevad.

Tabel 4. Igapäevasuutsetajate osakaalu hinnang \hat{p} ühendatud maakondade vanusegruppides, selle vead e ja c ning usaldusintervall (TKU16).

Igapäevasuutsetajad		TKU16		e	c	$UI(p)$
Maakond	Vanusegrupp	\hat{p}	n			
Harjumaa	16 – 34	0,155	402	0,035	0,226	(0,120; 0,190)
	35 – 64	0,203	763	0,029	0,143	(0,174; 0,232)
Tartumaa	16 – 34	0,178	120	0,068	0,382	(0,110; 0,246)
	35 – 64	0,225	190	0,062	0,243	(0,193; 0,317)
Ida-Virumaa	16 – 34	0,301	73	0,105	0,349	(0,196; 0,406)
	35 – 64	0,249	220	0,057	0,229	(0,192; 0,306)
Raplamaa, Läänemaa	16 – 34	0,195	83	0,085	0,436	(0,110; 0,280)
Hiiumaa, Saaremaa	35 – 64	0,237	147	0,069	0,291	(0,168; 0,306)
Põlvamaa, Valgamaa	16 – 34	0,156	65	0,088	0,565	(0,068; 0,244)
Võrumaa	35 – 64	0,190	140	0,065	0,342	(0,125; 0,255)
Lääne-Virumaa	16 – 34	0,231	65	0,102	0,442	(0,129; 0,333)
Järvamaa	35 – 64	0,259	142	0,072	0,278	(0,187; 0,331)
Pärnumaa	16 – 34	0,326	46	0,135	0,414	(0,191; 0,461)
	35 – 64	0,268	130	0,076	0,284	(0,192; 0,344)
Jõgevamaa, Viljandimaa	16 – 34	0,292	48	0,128	0,438	(0,164; 0,420)
	35 – 64	0,212	117	0,074	0,349	(0,138; 0,286)

Märkame, et kui hinnatud osakaalude suhtelised vead on väikesed ja hinnatud osakaalude erinevus suur, siis saame tegelikke osakaale üldkogumis eristada. Näiteks võrdleme Harjumaa ja Ida-Virumaa igapäevasuutsetajate hinnangulisi osakaale esimeses vanusegrupis, osakaalud vas-

tavalt 0,155 ja 0,301 (tabel 3). Usaldusvahemikest (0,120; 0,190) ja (0,196; 0,406) näeme, et tegelikud igapäevasuitsetajate osakaalud on erinevad.

2.5. Reeglid valimimahu arvutamiseks

Usaldusintervalli laius on määratud poole usaldusvahemiku laiusega e , mis on aga määratud suhtelise veaga c , $e = c \cdot p$. Suurus e sõltub lisaks valimimahule n ka osakaalust p . Seega n määramisel tuleb fikseerida ka osakaal p . Tabelist 2 (alapeatükk 2.3) on näha, et osakaalu $p = 0,2$ hindamiseks suhtelise veaga $c = 0,2$ on vaja valimimahtu $n = 385$. Siis $e = c \cdot p = 0,04$.

Vaatleme 30 kihti ehk 15 maakonda kahes vanusegrupis. Kui igas kihis on valimimaht 385 (Hiiumaa ja mõne teise maakonna jaoks kahes vanusegrupis pole selline n mõistlik), siis valimimaht 30 kihi jaoks on $30 \cdot 385 = 11\,550$, mis võib aga liiga suureks osutada, seetõttu vähendame usaldusnivood: $1 - \alpha = 0,9$. Nüüd osakaalu $p = 0,2$ ja suhtelise vea $c = 0,2$ korral on valimimaht ühe kihi jaoks 269 ning 30 kihi jaoks $30 \cdot 269 = 8070$.

Kui kihte on 16 (kaheksa ühendatud maakonna rühma), siis esimesel juhul ($z_{\alpha/2} = 1,96$) on vaja valimimahtu suurusega $16 \cdot 385 = 6160$ ning teisel juhul ($z_{\alpha/2} = 1,64$) $16 \cdot 269 = 4304$.

Uurime nüüd fikseeritud n korral teiste osakaalude viga e . Kuna $e = c \cdot p$, siis valemist (4) saame avaldada:

$$cp = z_{\alpha/2} \frac{S}{\sqrt{n}},$$

millest

$$c = z_{\alpha/2} \sqrt{\frac{1-p}{pn}}. \quad (11)$$

Näeme, et p suurenedes fikseeritud n korral suhteline viga väheneb. Seega olles määranud valimimahu n nõudest, et osakaalu $p = 0,2$ hindamisel on suhteline viga 0,2, saavutame, et kõigi osakaalude $p > 0,2$ hinnangud on täpsemad, nende suhtelised vead on väiksemad kui 0,2. Samas osakaalude $p < 0,2$ hinnangud on ebatäpsemad.

Kontrollime, kas osakaalude hinnangud 0,1 ja 0,2 vihjavad tegelike osakaalude erinevusele. Vastavad usaldusvahemikud on $z_{\alpha/2} = 1,64$ korral (0,070; 0,130) ja (0,160; 0,240) ning see-

tõttu on tegelikud osakaalud erinevad. Analoogselt saame hinnangute 0,3 ja 0,4 korral: usaldusvahemikud (0,254; 0,346) ja (0,351; 0,449) ei kattu, seega on tegelikud osakaalud erinevad.

Järgnevalt võrdleme osakaalude hinnanguid 0,05 ja 0,1. Usaldusvahemikud on $z_{\alpha/2} = 1,64$ korral vastavalt (0,028; 0,072) ja (0,070; 0,130). Kuna usaldusvahemikud kattuvad, siis usaldusnivool 0,9 ei saa väita, et tegelikud osakaalud on erinevad. Samuti näeme, et osakaalu 0,05 suhteline viga on küllaltki suur, moodustades 44% hinnangust, tabel 5.

Tabel 5. Vead c ja e , valemitest (4) ja (11), fikseeritud valimimahu korral.

Osakaal p	$n = 269, z_{\alpha/2} = 1,64$	
	c	e
0,05	0,436	0,022
0,1	0,3	0,03
0,2	0,2	0,04
0,3	0,153	0,046
0,4	0,122	0,049
0,7	0,065	0,046
0,8	0,05	0,04

Eelnevat arutelu arvestades on mõistlik vaadelda 16 kihti, st kaheksa ühendatud maakonda kahes vanusegrupis. Seega antud töös fikseerime valimimahu määramiseks usaldusnivoo 0,9 ($z_{\alpha/2} = 1,64$) ja nõuame, et oskaal $p = 0,2$ oleks hinnatav suhtelise veaga $c = 0,2$ (20%). Saame vajalikuks valimimahuks $n = 269$. Töö kolmandas peatükis korrigeerime leitud valimimahu vastamismääradega.

3. Hindamine kihtvaliku korral ja lisavalimi koostamine

3.1. Hindamine kihtvaliku korral

Olgu üldkogumis H kihti U_h , mis on lõikumatud ja mille korral $U = \cup_{h=1}^H U_h$. Olgu $s_h \subset U_h$ valim kihist U_h . Olgu N_h kihi U_h maht ja n_h valimi s_h maht. Üldkogumi kogusumma avaldub siis kujul:

$$Y = \sum_U Y_h = \sum_{h=1}^H \sum_{k \in U_h} y_k \quad (12)$$

ja keskmine kujul:

$$\bar{Y} = \frac{Y}{N} = \sum_U \frac{y_k}{N}. \quad (13)$$

Kogusumma nihketa hinnang on Horvitz-Thompsoni hinnang, kus üldkogumi summa asemel kasutatakse valimisummat, mille liidetavad on jagatud kaasamistõenäosustega. Kihtvaliku korral on kaasamistõenäosused $\pi_k = P(s \ni k)$ kihi sees konstantsed [7, lk 124-128]:

$$\pi_k = \frac{n_h}{N_h}, \text{ kui } k \in U_h.$$

Valemist (12) saame kogusumma hinnangu:

$$\hat{Y} = \sum_{h=1}^H \sum_{k \in s_h} \frac{y_k}{\pi_k} = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{k \in s_h} y_k. \quad (14)$$

Kao korral on y_k teada vastanute hulgas r_h , kus $r_h \subset s_h$ ja $r = \cup_{h=1}^H r_h$. Eeldame, et vastamistõenäosused valimis s_h on konstantsed $\frac{m_h}{n_h}$, kus m_h on vastanute arv. Kasutades vastamistõenäosusi, saame valemist (14) kogusumma hinnangu [7, lk 124-128]:

$$\hat{Y}_r = \sum_{h=1}^H \frac{N_h}{n_h} \cdot \frac{n_h}{m_h} \sum_{k \in r_h} y_k = \sum_{h=1}^H \frac{N_h}{m_h} \sum_{k \in r_h} y_k = \sum_{h=1}^H N_h \bar{y}_{r_h}, \quad (15)$$

kus $\bar{y}_{r_h} = \frac{1}{m_h} \sum_{k \in r_h} y_k$. Praktikas arvutatakse hinnang kaalude abil:

$$\hat{Y}_r = \sum_{k \in r} w_k y_k,$$

kus valemist (15) $w_k = \frac{N_h}{m_h}$, kui $k \in r_h$. Keskmise $\bar{Y} = \frac{Y}{N}$ hinnanguks on $\frac{\hat{Y}_r}{N}$. Osakogumi keskmise hinnangu saamiseks summeeritakse $w_k y_k$ vastava osakogumi vastanute hulgas ja jagatakse

osakogumi mahuga [7, lk 126]. Näiteks kihi keskmise $\bar{Y}_h = \sum_{U_h} y_k / N_h$ hinnanguks tuleb

$$\hat{Y}_{r_h} = \sum_{k \in r_h} \frac{w_k y_k}{N_h} = \frac{1}{m_h} \sum_{k \in r_h} y_k = \bar{y}_{r_h}.$$

Lihtsustatud dispersioonivalem (arvestab, et m_h on fikseeritud konstant) tuleb valemist (15):

$$D(\hat{Y}_r) = \sum_{h=1}^H N_h^2 D(\bar{y}_{r_h}).$$

Selle töö kontekstis on y_k binaarne ja \bar{y}_{r_h} on osakaalu hinnang \hat{p} vastanute hulgas r_h , mille dispersioon on antud valemiga (1), kus n ja N on asendatud vastavate suurustega m_h ja n_h .

3.2. Vastamismäärad toimunud uuringus

Tabelis 6 on välja toodud 2016. aasta TKU vastanute osakaalud protsentidena ühendatud maakondades. Näeme, et vanusegrupis 35 – 64 ületab vastamismäär kõikides maakondades 50%, kusjuures kõrgeim vastanute osakaal on Jõgeva- ja Viljandimaa rühmas, 74%, madalaim Harjumaa, 55%. Vanusegrupis 16 – 34 on madalaim vastamismäär Pärnumaal, 41%, ning kõrgeim Rapla-, Lääne-, Hiiu- ja Saaremaa rühmas, 57%.

Soovime, et lõplik valimimaht 16 kihis oleks vastamismääruga korrigeeritud. Kadu arvestav valimimaht on Harjumaa esimeses vanusegrupis $n = \frac{269}{0,46} = 585$ ja teises vanusegrupis $n = 489$. Pärnumaal saame aga vastavalt $n = 657$ ja $n = 396$.

Kui tavaliselt vastanud erinevad mittevastanutest, siis sellisel viisil valimimahu suurendamine ei vähenda nihet hinnangutes. Valimimahu suurendamine vähendab aga hinnangu dispersiooni. Nihke vähendamiseks kasutatakse hindamisetapil lisainformatsiooni ja nihet kompenseerivaid hinnanguid. Antud juhul toome hinnangute valemid eeldusel, et vastanud ja mittevastanud on ühesugused.

Tabel 6. Vastamismäärad (%) ühendatud maakondade vanusegruppides.

Maakond	Vastamismäär		
	16 – 34	35 – 64	Kõik
Harjumaa	46	55	52
Tartumaa	55	59	58
Ida-Virumaa	42	56	51
Raplamaa, Läänemaa, Hiiumaa, Saaremaa	57	62	60
Põlvamaa, Valgamaa, Võrumaa	49	67	60
Lääne-Virumaa, Järvamaa	54	66	61
Pärnumaa	41	68	58
Jõgevamaa, Viljandimaa	45	74	73
Kõik	48	59	55

3.3. Lisavalimi koostamine

Valimi koostamisel on põhivalimi suurus 5000, mis võetakse nii, nagu seda on seni tehtud TKU läbiviimisel. Tulemuseks on võrdsete kaasamistõenäosustega valim, mis imiteerib lihtsat juhuvalikut. Lisavalimi puhul rakendame kihtides ehk ühendatud maakondade vanusegruppides lihtsat juhuvalikut, nii et saaksime soovitud valimimahu kihtides, tabel 7.

Tabelis 7 on lisavalimi soovitud vastajate arv saadud lahutades TKU16 vastanute arv fikseeritud kõigist vastajate arvust 269 (leitud alapeatükis 2.5). Lisavalimi kogumaht on leitud jagades lisavalimi vastajate arvu eelmises alapeatükis 3.2 leitud vastava rühma vastamistõenäosusega tabelist 6.

Tabelist 7 näeme, et kogu valimi suurus peab olema 9110. Kuna TKU valimi koostamine peab jääma nii, nagu seda seni on tehtud, ei saa me Harjumaa vanusegruppides valimimahte vähendada, kuna aga vastajate valimimahud ületavad kihtides fikseeritud vastajate valimimahu 269, siis Harjumaal lisavalimit kummastki vanusegrupist võtta ei ole vaja.

Tabel 7. Lisavalimi ja TKU16 valimimahud ühendatud maakondade vanusegruppides.

Maakond	Vanusegrupp	TKU16		Lisavalim		Kokku
		Vastanud	Valim	Vastajad	Valim	
Harjumaa	16 – 34	402	871	-	-	871
	35 – 64	763	1388	-	-	1388
Tartumaa	16 – 34	120	219	149	271	490
	35 – 64	190	320	79	134	454
Ida-Virumaa	16 – 34	73	174	196	467	641
	35 – 64	220	396	49	88	484
Raplamaa, Läänemaa	16 – 34	83	146	186	327	473
Hiiumaa, Saaremaa	35 – 64	147	238	122	197	435
Põlvamaa, Valgamaa	16 – 34	65	132	204	416	548
Võrumaa	35 – 64	140	210	129	192	402
Lääne-Virumaa	16 – 34	65	121	204	378	499
Järvamaa	35 – 64	142	216	127	193	409
Pärnumaa	16 – 34	46	111	223	544	655
	35 – 64	130	192	139	205	397
Jõgevamaa, Viljandimaa	16 – 34	48	107	221	492	599
	35 – 64	117	159	152	206	365
Kokku	16 – 64	2751	5000	2180	4110	9110

Tehes edasisi lihtsustusi Rahvastikuregistri tarbeks, kus valim tegelikult võetakse, ümardame lisavalimimahud üles ja võimalikult ühesuguseks. Saame, et lisavalimisse on vaja võtta maakonnagruppide Raplamaa, Läänemaa, Hiiumaa, Saaremaa ja Lääne-Virumaa, Järvamaa nooremast vanusegrupist 400 isikut, analoogiliselt Ida-Virumaa ja Põlvamaa, Valgamaa, Võrumaa ja Jõgevamaa, Viljandimaa korral 500 isikut ning Tartumaa ja Pärnumaa korral vastavalt 300 ja 600 isikut. Lisavalimisse on vaja võtta kõigi maakonnagruppide (va Harjumaa ja Ida-Virumaa) vanemast vanuserühmast 200 isikut, Ida-Virumaa korral 100 isikut. Kokku on lisavalimimahuks 4500, tabel 8.

Kõike eelnevat arvesse võttes on järgmise TKU korral vaja võtta kaks valimit:

- põhivalim 5000 isikut, nii nagu seda on seni tehtud;
- lisavalim 4500 isikut lihtsa juhusliku valiku teel 16 kihis (ühendatud maakondade vanusegrupid) vastavalt valimimahtudele tabelist 8.

Tabel 8. Lisavalimi ümardatud valimimahud ühendatud maakondade vanusegruppides.

Maakond	Vanusegrupp	Lisavalim	
		Esialgne	Ümardatud
Tartumaa	16 – 34	271	300
	35 – 64	134	200
Ida-Virumaa	16 – 34	467	500
	35 – 64	88	100
Raplamaa, Läänemaa Hiiumaa, Saaremaa	16 – 34	327	400
	35 – 64	197	200
Põlvamaa, Valgamaa Võrumaa	16 – 34	416	500
	35 – 64	192	200
Lääne-Virumaa Järvamaa	16 – 34	378	400
	35 – 64	193	200
Pärnumaa	16 – 34	544	600
	35 – 64	205	200
Jõgevamaa, Viljandimaa	16 – 34	492	500
	35 – 64	206	200
Kokku	16 – 64	4110	4500

Kokkuvõte

Bakalaureusetöö eesmärgiks oli lisavalimimahu leidmine Eesti täiskasvanud rahvastiku tervisekäitumise uuringu jaoks, nii et piisava täpsusega tulemused saadakse maakonnagruppide kaupa kahes vanuserühmas 16 – 34 ja 35 – 64. Töö esimeses osas anti lühike ülevaade Eesti täiskasvanud rahvastiku tervisekäitumise uuringust ning selle metoodikast, samuti kirjeldati Finbalt Health Monitor uuringut, mille raames Baltiriigid ja Soome 1990. aastast kuni 2012. aastani ühisuuringut läbi viisid, kuid hiljem iseseivalt jätkasid. Laiendatud valimiga uuringut on vaja selleks, et saada parem ülevaade Eesti täiskasvanud rahvastiku tervisest ja tervisekäitumisest maakondade sees.

Töö teises peatükis tuletati valimimahu arvutamise valemid etteantud täpsuse korral ning vaadeldi valimimahtude muutumist erinevate fikseeritud osakaalude ja hinnangu täpsuste korral. Kõigele lisaks uuriti põgusalt 2016. aasta tervisekäitumise uuringus igapäevasuutsetajate osakaalude hinnangute täpsusi nii 15 maakonna kui ka kaheksa ühendatud maakonna kahes vanuserühmas. Edasi määrati kindlaks reegel valimimahu arvutamiseks: vaadeldi 16 kihti, kaheksa ühendatud maakonda kahes vanuserühmas, kusjuures usaldusnivool 0,9 nõuti, et osakaal $p = 0,2$ oleks hinnatav suhtelise veaga $c = 0,2$. Reeglist lähtuvalt saadi, et valimimaht kihis peab olema 269.

Kolmandas osas kirjeldati parameetrite hindamist kihtvaliku korral. Sealjuures vaadeldi 2016. aasta Eesti täiskasvanud rahvastiku tervisekäitumise uuringu vastamismäärasid kaheksa maakonnagrupi kahes vanuserühmas. Leitud vastamismäärasid arvestades leiti valimimahud 16 kihis. Tulemuseks saadi, et lisavalimiga TKU jaoks on vaja võtta kaks valimit: põhivalim 5000 isikut nii nagu seda on seni tehtud ning lisavalim 4500 isikut lihtsa juhuvaliku teel kihtides.

Viited

- [1] Maakondi võrdlev tabel, 2018. https://www.terviseinfo.ee/images/M_images/Maakondade_v6rdlusgraafik_2018.pdf. (vaadatud 27.04.2019)
- [2] Tekkel, M., Veidema, T. (2017). Eesti täiskasvanud rahvastiku tervisekäitumise uuring, 2016. *Tervise Arengu Instituut*.
<http://rahvatervis.ut.ee/bitstream/1/6665/1/Tervisekaitumine2017.pdf>.
(vaadatud 11.02.2019)
- [3] TKU23: Alkoholi tarvitamise sagedus viimase 12 kuu jooksul soo ja piirkonna järgi.
http://pxweb.tai.ee/PXWeb2015/pxweb/et/05Uuringud/05Uuringud__02TKU__03Alkohol/TKU23.px/table/tableViewLayout2/?rxid=d6dcecaf-759c-49b1-a786-9f147c590a54. (vaadatud 6.03.2019)
- [4] Helasoja, V., Prättälä, R., Dregval, L., Pudule, I., Kasmel, A. (2002) Late response and item nonresponse in the Finbalt Health Monitor Survey. *European Journal of Public Health* 2002, 12: 117-123.
https://www.researchgate.net/publication/11300669_Late_response_and_item_nonresponse_in_the_Finbalt_Health_Monitor_survey. (vaadatud 1.04.2019)
- [5] Tekkel, M., Veideman, T. (2013). Eesti täiskasvanud rahvastiku tervisekäitumise uuring, 2012. *Tervise Arengu Instituut*.
https://intra.tai.ee/images/prints/documents/136479842690_TKU_2012.pdf. (vaadatud 11.02.2019)
- [6] Prättälä, R., Helasoja, V., Finbalt Group, Kasmel, A., Klumbiene, J., Pudule, I. (2003). Finbalt Health Monitor. In: McQueen D.V., Puska P. (eds) *Global Behavioral Risk Factor Surveillance*. Springer, Boston, MA https://link.springer.com/chapter/10.1007%2F978-1-4615-0071-1_7 (vaadatud 1.04.2019)
- [7] Traat, I., Inno, J. (1997). *Tõenäosuslik valikuuring*. Tartu: TÜ kirjastus.
- [8] Lohr, Sharon L. (1999). *Sampling: Design and Analysis*. Brooks/Cole Publishing Company.
- [9] Traat, I. (2006). *Matemaatilise statistika põhikursus*. Tartu: TÜ kirjastus.

Lisa

Tabelid

Tabel 9. Tunnuse *smoind3* osakaalud maakonna ja kahe vanuserühma põhjal.

Maakond	Vanuserühm	p_1	p_2	p_3	p_4
Harjumaa	16 – 34	0,155	0,133	0,276	0,436
	35 – 64	0,203	0,066	0,278	0,453
Hiiumaa	16 – 34	0,091	0,091	0,455	0,364
	35 – 64	0,250	0,000	0,167	0,583
Ida-Virumaa	16 – 34	0,301	0,110	0,192	0,397
	35 – 64	0,249	0,051	0,267	0,433
Jõgevamaa	16 – 34	0,333	0,048	0,286	0,333
	35 – 64	0,304	0,065	0,239	0,391
Järvamaa	16 – 34	0,167	0,111	0,389	0,333
	35 – 64	0,276	0,000	0,224	0,500
Läänemaa	16 – 34	0,182	0,182	0,182	0,455
	35 – 64	0,257	0,000	0,400	0,343
Lääne-Virumaa	16 – 34	0,255	0,085	0,191	0,468
	35 – 64	0,247	0,062	0,296	0,395
Põlvamaa	16 – 34	0,174	0,043	0,261	0,522
	35 – 64	0,179	0,051	0,256	0,513
Pärnumaa	16 – 34	0,326	0,109	0,261	0,304
	35 – 64	0,268	0,039	0,362	0,331
Raplamaa	16 – 34	0,231	0,231	0,308	0,231
	35 – 64	0,273	0,068	0,318	0,341
Saaremaa	16 – 34	0,217	0,087	0,217	0,478
	35 – 64	0,188	0,021	0,292	0,500

Tartumaa	16 – 34	0,178	0,136	0,305	0,381
	35 – 64	0,225	0,075	0,235	0,465
Valgamaa	16 – 34	0,238	0,000	0,238	0,524
	35 – 64	0,162	0,027	0,270	0,541
Viljandimaa	16 – 34	0,259	0,111	0,185	0,444
	35 – 64	0,149	0,030	0,299	0,522
Võrumaa	16 – 34	0,050	0,150	0,100	0,700
	35 – 64	0,213	0,098	0,262	0,426

Tabel 10. Tunnuse *smoind3* osakaalud ühendatud maakondade ja kahe vanuserühma põhjal.

Maakond	Vanuserühm	p_1	p_2	p_3	p_4
Harjumaa	16 – 34	0,155	0,133	0,276	0,436
	35 – 64	0,203	0,066	0,278	0,453
Tartumaa	16 – 34	0,178	0,136	0,305	0,381
	35 – 64	0,225	0,075	0,235	0,465
Ida-Virumaa	16 – 34	0,301	0,110	0,192	0,397
	35 – 64	0,249	0,051	0,267	0,433
Raplamaa, Läänemaa Hiiumaa, Saaremaa	16 – 34	0,195	0,159	0,268	0,378
	35 – 64	0,237	0,029	0,317	0,417
Põlvamaa, Valgamaa Võrumaa	16 – 34	0,156	0,062	0,203	0,578
	35 – 64	0,190	0,066	0,263	0,482
Lääne-Virumaa Järvamaa	16 – 34	0,231	0,092	0,246	0,431
	35 – 64	0,259	0,036	0,266	0,439
Pärnumaa	16 – 34	0,326	0,109	0,261	0,304
	35 – 64	0,268	0,039	0,362	0,331
Jõgevamaa, Viljandimaa	16 – 34	0,292	0,083	0,229	0,396
	35 – 64	0,212	0,044	0,274	0,469

Tabel 11. Sagedused küsitlusele vastamise, maakonna ja vanuserühma järgi.

Maakond	Vastas			Ei vastanud			KOKKU		
	16 – 34	35 – 64	Kõik	16 – 34	35 – 64	Kõik	16 – 34	35 – 64	Kõik
Harjumaa	402	763	1165	469	625	1094	871	1388	2259
Hiiumaa	11	14	25	6	8	14	17	22	39
Ida-Virumaa	73	220	293	101	176	277	174	396	570
Jõgevamaa	21	48	69	23	16	39	44	64	108
Järvamaa	18	60	78	24	16	40	42	76	118
Läänemaa	22	36	58	15	22	37	37	58	95
Lääne-Virumaa	47	82	129	32	58	90	79	140	219
Põlvamaa	23	40	63	19	20	39	42	60	102
Pärnumaa	46	130	176	65	62	127	111	192	303
Raplamaa	27	46	73	20	33	53	47	79	126
Saaremaa	23	51	74	22	28	50	45	79	124
Tartumaa	120	190	310	99	130	229	219	320	539
Valgamaa	21	39	60	20	285	48	41	67	108
Viljandimaa	27	69	96	36	26	62	63	95	158
Võrumaa	21	61	82	28	22	50	49	83	132
Kõik	902	1849	2751	979	1270	2249	1881	3119	5000

Tabel 12. Vastanute sagedused maakonna ja vanuserühmade järgi.

Ühendatud maakonnad	Vastas		
	16 – 34	35 – 64	Kõik
Harjumaa	402	763	1165
Tartumaa	120	190	310
Ida-Virumaa	73	220	293
Raplamaa, Läänemaa, Hiiumaa, Saaremaa	83	147	230
Põlvamaa, Valgamaa, Võrumaa	65	140	205
Lääne-Virumaa, Järvamaa	65	142	207
Pärnumaa	46	130	176
Jõgevamaa, Viljandimaa	48	117	165
Kokku	902	1849	2751

Tabel 13. Maakondade mahud kahes vanuserühmas, Statistikaameti 2017. aasta andmed.

Maakonnad	Osakogumite mahud		
	15 – 34	35 – 64	Kõik
Harjumaa	263 169	397 470	660 639
Hiiumaa	2293	4018	6311
Ida-Virumaa	28 708	62 667	91 375
Jõgevamaa	7571	12 104	19 675
Järvamaa	7158	12 094	19 252
Läänemaa	5503	9944	15 447
Lääne-Virumaa	14 045	23 659	37 704
Põlvamaa	6818	11 388	18 206
Pärnumaa	19 569	32278	51 847
Raplamaa	8150	13 797	21 947
Saaremaa	8121	13 534	21 655
Tartumaa	61 825	88 715	15 0540
Valgamaa	7040	12 003	19 043
Viljandimaa	11 550	18 356	29 906
Võrumaa	8225	13 328	21 553
Kokku	459 745	725 355	1 185 100

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Kristina Muhu,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Eesti täiskasvanud rahvastiku tervisekäitumise uuringu lisavalimi moodustamine", mille juhendajad on Imbi Traat ja Natalja Eigo, 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.

Kristina Muhu

08.05.2019