

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Willem Klaassen
**Traditsioonilised ja uuenduslikud
rahvastikuproгноosi metoodikad**

Matemaatiline statistika
Bakalaureusetöö (9 EAP)

Juhendajad: PhD Terje Trasberg

MSc Mare Vähi

TARTU 2024

TRADITSIOONILISED JA UUENDUSLIKUD RAHVASTIKUPROGNOOSI METOODIKAD

Bakalaureusetöö

Willem Klaassen

Lühikokkuvõte

Rahvastikuproгноos hindab, millised on rahvastikusündmuste, -koosseisu ja -suuruse muutumine tulevikus. Rahvastikuproгноose koostatakse ülikoolide, rahvusvaheliste organisatsioonide ning riiklike statistikaametite poolt. Käesoleva töö eesmärk on uurida erinevaid traditsioonilisi ja uuenduslikke rahvastikuproгноosi metoodikaid ning rakendada hulka neist Eesti rahvastikul, et võrrelda saadavaid tulemusi ja nende täpsust Statistikaameti 2019. aasta rahvastikuproгноosiga. Uuritakse nelja tüüpi ja rakendatakse kaht tüüpi rahvastikuproгноosi meetodeid.

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

Märksõnad: Demograafia, rahvaloendused, rahvastikustatistika, elutabelid, rahvastik, kirjeldav demograafia, rahvaarv, suremusnäitajad, sündimusnäitajad.

TRADITIONAL AND INNOVATIVE POPULATION PROJECTION METHODS

Bachelor thesis

Willem Klaassen

Abstract

Population forecasting estimates future changes in population events, composition and size. Population forecasts are produced by universities, international organisations and national statistical offices. The aim of this paper is

to examine different traditional and innovative population projection methodologies and apply a number of them to the Estonian population in order to compare the obtained results and their accuracy with the 2019 population projection of Statistics Estonia. Four types of population projection methodologies will be studied and two types will be applied.

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

Key Words: Demography, censuses, population statistics, life tables, population, descriptive demography, population, mortality indexes, fertility measurement.

Sisukord

Sissejuhatus	5
1 Metoodika	7
1.1 Nihke- ehk osismetod	7
1.1.1 R-i paketi demCore osismetodi metoodika	8
1.1.2 Eesti rahvastikuproгноosi metoodika	9
1.2 Tõenäosuslikud rahvastikuproгноosi meetodid	13
1.2.1 R-i paketi bayesPop Bayesiaanlik mudel	14
1.2.2 R-i pakett demography	19
1.3 Mikrosimulatsiooni meetodid	24
1.4 Masinõppe ja närvivõrkude rakendamine rahvastikuproгноosides .	27
1.5 Metoodikate hindamine	30
2 Andmestikud	32
3 Meetodite võrdlus	34
3.1 Osismetod	34
3.2 Tõenäosuslikud meetodid	34
3.3 Mikrosimulatsiooni meetodid	35
3.4 Masinõppe ja närvivõrkude rakendamine rahvastikuproгноosides .	36
3.5 Tulemused aastatel 2020 kuni 2022	36
Kokkuvõte	39
Kasutatud allikad	40

Lisa 1. Proгноос пакетига demCore	44
Lisa 2. Proгноос пакетига demCore	46
Lisa 3. Proгноос пакетига demography	54

Sissejuhatus

Rahvastikuproгноos hindab, millised on rahvastikusündmuste, -koosseisu ja -suuruse muutumine tulevikus. Valitsustele pakub huvi, kuhu ehitada koole, haiglaid ja uusi teid ning kui palju saab olema maksumaksjaid või sõjaväe-ealisi. Ettevõtted aga hoolivad oma klientuuriga toimuvatest muutustest. Intrigeeriv on see ka üldsusele. Rahvastiku koosseis mõjutab ju nii kultuuri kui töö- ja kinnisvaraturgu. Eestis võib tegu olla riigi püsijäämist mõjutava küsimusega. Rahvastikuproгноoside koostamine on rahvastikustatistikute üks kõige tähtsamaid kohustusi (Vähi, Mare *et al.*, 2015).

Rahvastikuproгноose koostatakse ülikoolide, rahvusvaheliste organisatsioonide ning riiklike statistikaametite poolt. Eesti põhiline proгноoside koostaja on Eesti Statistikaamet, mis koostab proгноose iga viie aasta tagant. Viimane neist koostati aastal 2019. Eesti rahvastikuproгноosi eesmärk on hinnata rahvaarvu muutust, sündide ja surmade arvu, välisrännet ja rahvastiku koosseisu ning seda arvutatakse oletatava tulevase sündimuse, suremuse ja välisrände alusel soo ja vanuse gruppide jaoks (“Rahvastikuproгноos | Statistikaamet”, 2024). Eesti rahvastikku proгноositakse ka Eurostati, Ühinenud Rahvaste Organisatsiooni ja Tartu Ülikooli poolt.

Eesti Statistikaameti poolt koostatav rahvastikuproгноos kasutab laialt levinud osismeetodit (*Cohort Component Method*). Osismeetodi aluseks on rahvastiku soovanusjaotus, mida kasutades moodustatakse projektsioon rühmade arvukust uuendades vastavalt sündimuse, suremuse ja rände kohta tehtavatele eeldustele (O’Neill *et al.*, 2001). Eesti rahvastikuproгноosi eeldused luuakse tuginedes antud valdkonna ekspertide arvamusele (Statistikaamet, 2019).

Käesoleva töö eesmärk on uurida erinevaid traditsioonilisi ja uuenduslikke rahvastikuproгноosi metoodikaid ning rakendada hulka neist Eesti rahvastikul, et võrrelda saadavaid tulemusi ja nende täpsust Statistikaameti 2019. aasta rahvastikuproгноosiga.

Töö kirjutati tekstiprotsessoriga $\text{\LaTeX}$ ja erinevaid rahvastikuproгноosi metoodikaid rakendati rakendustarkvaras R.

1 Metoodika

1.1 Nihke- ehk osismetod

Osismetod rahvaarvu prognoosimiseks (*Cohort Component Method of Population Projection*) on olnud kasutuses kõigis pikaajalistes rahvusvahelistes rahvastikuprognoosides (O'Neill *et al.*, 2001). Osismetodi keskmeks on demograafiline bilansivõrrand (*Demographic Balancing Equation*). Bilansivõrrandi kohaselt on rahvaarv ajahetkel T võrdne rahvaarvuga ajahetkel 0, millele on liidetud inimesed, kes ajavahemikus kuni hetkeni T sündisid või sisse rändasid ning millest on lahutatud inimesed, kes surid või rändasid välja. Demograafilise bilansivõrrandi üldkuju on

$$\begin{aligned} N(T) &= N(0) - E[0, T] - D[0, T] + B[0, T] + I[0, T] \\ &= N(0) - D[0, T] + B[0, T] + G[0, T], \\ G[0, T] &= I[0, T] - E[0, T], \end{aligned} \tag{1}$$

kus N on rahvaarv, E on väljaränne, D on surmad, B on sünnid, I on sisseränne ja G on rändesaldo (Katie Paulson *et al.*, 2024).

Osismetodi aluseks on rahvastiku soo-vanusjaotus, mida kasutades moodustatakse projektsioon rühmade arvukust uuendades. Rühmasid uuendatakse vastavalt sündimuse, suremuse ja rände kohta tehtavatele eeldustele. Iga rühm liigub järgmisesse vanuserühma vastavalt antud rühma vanuse- ja soopõhisele suremusele. Rändega arvestatakse vanuse- ja soopõhiseid rändesaldosid vastavatele vanuse ja soo kohortitele liites. Noorima vanuserühma suurust mõjutab sündide arv, mis arvutatakse arvatavat vanusepõhist sündimust kõigile sünnitusealistele naiste rühmadele rakendades. Sünnid jaotatakse sugude vahel arvatavat soojaotust kasutades (O'Neill *et al.*, 2001).

1.1.1 R-i paketi demCore osismetodi metoodika

R-i pakett demCore võimaldab erinevate demograafiliste meetodite rakendamist. Arvutada saab elutabeleid, summaarset sündimuskordajat ja sündimuse vanusekordajat ning osismetodi abil tuleviku rahvaarvu (Katie Paulson *et al.*, 2024). Paketi kasutamiseks on vajalik ka lahendamata küsimuste tõttu nüüdseks arhiveeritud pakett assertive.base (“CRAN - Package assertive.base”, 2024). Antud paketist installleeritigi hilisem arhiveeritud versioon 0.0-9.

Üldine valem ajahetkel $t + n$ vanuses x kuni $x + n$ rahvaarvu leidmiseks ühele soole on:

$${}_nN_x(t + n) = \left[\left({}_nN_{x-n}(t) + \frac{{}_nG_{x-n}[t, t + n]}{2} \right) \cdot \frac{{}_nL_x[t, t + n]}{{}_nL_{x-n}[t, t + n]} + \frac{{}_nG_x[t, t + n]}{2} \right]$$

kus n on vanusegruppide pikkus ning mitu aastat ette prognoositakse ja:

- ${}_nN_x(t)$ - rahvaarv ajahetkel t vanusegrupis x kuni $x + n$,
- ${}_nL[t, t + n]$ - ajavahemikus t kuni $t + n$ elatud inimaastate arv vanusegrupis x kuni $x + n$,
- $\frac{{}_nG[t, t + n]}{2}$ - rändesaldo ajavahemikus t kuni $t + n$ vanusegrupis x kuni $x + n$ (Katie Paulson *et al.*, 2024).

Valem noorima vanusegrupi rahvaarvu leidmiseks ajahetkel $t + n$ ühele soole ilma rändeta on:

$${}_nN_0(t + n) = B[t, t + n] \cdot \frac{{}_nL_0[t, t + n]}{{}_n \cdot l_0[t, t + n]},$$

kus:

$$B[t, t + n] = \sum_{x=0}^A \frac{n}{2} \cdot {}_nF_x[t, t + n] \cdot {}_nL_x[t, t + n] - \frac{{}_nN_x^F(t) + {}_nN_{x-n}^F(t) \cdot \frac{{}_nL_x[t, t + n]}{{}_nL_{x-n}[t, t + n]}}{2},$$

- ${}_nN_x(t)$ - rahvaarv ajahetkel t vanusegrupis x kuni $x + n$,

- ${}_nL_x[t, t+n]$ = ajavahemikus t kuni $t+n$ elatud inimaastate arv vanusegrupis x kuni $x+n$,
- $B[t, t+n]$ = sündide arv ajavahemikus t kuni $t+n$,
- ${}_nF_x[t, t+n]$ = sündimuse vanusekordaja ajavahemikus t kuni $t+n$ naiste vanusegrupis x kuni $x+n$ (Katie Paulson *et al.*, 2024).

Valem kõige vanema vanuserühma rahvaarvu leidmiseks ajahetkel $t+n$ ühele soole ilma rändeta on:

$$\begin{aligned} {}_\infty N_x(t+n) &= \left({}_nN_{x-n}(t) \cdot \frac{{}_nL_x[t, t+n]}{{}_nL_{x-n}[t, t+n]} \right) + \left({}_\infty N_x(t) \cdot \frac{T_{x+n}[t, t+1]}{T_x[t, t+n]} \right) \\ &\approx ({}_nN_{x-n}(t) + {}_\infty N_x(t)) \cdot \frac{T_x[t, t+n]}{T_{x-n}[t, t+n]}. \end{aligned}$$

kus:

- ${}_nN_x(t)$ rahvaarv ajahetkel t vanusegrupis x kuni $x+n$,
- ${}_nL_x[t, t+n]$ elatud inimaastate arv ajavahemikus t kuni $t+n$ vanusegrupis x kuni $x+n$,
- ${}_\infty N_x(t)$ rahvaarv ajahetkel t kõige vanemas vanuserühmas,
- $T_{x+n}[t, t+1]$ elatud inimaastate arv ajavahemikus t kuni $t+1$ vanusest $x+n$ ja vanemates vanuserühmades,
- $T_x[t, t+n]$ on inimeseaaastate arv elatud ajavahemikus t kuni $t+n$ vanusest x ja vanemates vanuserühmades (Katie Paulson *et al.*, 2024).

1.1.2 Eesti rahvastikuproгноosi meetoodika

Eesti Statistikaameti poolt 2019. aastal välja antud kogu Eesti rahvastikuproгноosi mudeli aluseks olid Eesti rahvastiku soo-vanusjaotus hinnatava perioodi alguses

ehk 2019. aasta 1. jaanuaril ning Eesti rahvastiku sündimus, suremus ja välisränne prognoositava perioodi jooksul (Statistikaamet, 2019). Prognoositi vanustele 0 kuni 99 ning lisaks 100 ja rohkem kuni aastani 2080.

Sündimuse väljendamiseks kasutati summaarset sündimuskordajat, mida arvutati järgmiselt:

$$TFR = a \sum a f_x,$$

kus a on vanusevahemiku pikkus, $a f_x$ on sündimuse vanuskordaja ja summeerimine toimub üle kogu viljaka ea (Maasing & Tiit, 2021).

Vanusegrupi $[x, x + a)$ sündimuse vanusekordajat arvutati järgmiselt:

$$a f_x = \frac{a B_x}{a P_x},$$

kus a on vanusevahemiku pikkus, $a B_x$ on aasta jooksul vaadeldavas vanusegrupis elusalt sündinud laste arv ja $a P_x$ on aastakeskmise naiste arv vaadeldavas vanusegrupis (Maasing & Tiit, 2021). Aastakeskmise naiste arv vanusegrupis $[x, x + a)$ on vanusegrupi aasta alguse ja aasta lõpu naiste arvude aritmeetiline keskmine.

Suremuse väljendamiseks kasutati oodatavat eluiga, mida arvutati järgmiselt:

$$e_0 = \frac{S_0}{l_0} \quad \text{ja} \quad e_x = \frac{S_x}{l_x},$$

kus e_0 on keskmine oodatav eluiga sünnil, mis on siiski seotud aastaga x ehk iseloomustab aastal x sündinuid, ja e_x on aastal x sündinute keskmine oodatav eluiga. Tehakse eeldus, et suremus on tulevikus igas eas sama, mis aastal x (Maasing & Tiit, 2021).

S_0 on iga aasta möödudes ellujäänute summa kuni määratud maksimaalse elueani ehk järelejäänud eluaastate arv populatsioonis:

$$S_0 = \sum_{t=0}^w l_t \quad \text{ja} \quad S_x = \sum_{x'=x}^w l_{x'},$$

kus $l_t = l_{t-1} \cdot (1 - q_{t-1}^x)$ (Maasing & Tiit, 2021).

Vanuses t suremise tõenäosust q_t^x arvutatakse järgmiselt:

$$q_t^x = \frac{D_t^x}{P_t^x + D_t'^x},$$

kus D_t^x on vanuses t aastal x surnud inimeste arv, P_t^x on t aasta vanuste inimeste arv aastal x ning $D_t'^x$ on inimeste arv, kes surid vaatlusaastal pärast t aastaseks saamist (Maasing & Tiit, 2021).

Välisrände väljendamiseks kasutati välisrändesaldot, mis on sisse- ja väljarändajate koguarvu ja väljarändajate koguarvu vahe. Mudeli sisendina kasutati rändesaldot 1000 inimese kohta.

Mudeli üldkuju oli

$$N_{s,t,x} = N_{s,t,x-1} \cdot q_t^x + N_{s,t,x-1} \cdot \frac{r_t^x}{1000},$$

kus $N_{s,t,x}$ on soost s vanuses t inimeste arv aastal x ja r_t^x on rändesaldo 1000 elaniku kohta vanuses t ning aastal x .

Esimeses vanuses oli mudel kujul:

$$N_{s,0,x} = \sum_{t=15}^{t=49} \frac{N_{s,t,x-1} + N_{s,t,x}}{2} \cdot {}_1f_t^x,$$

kus $\frac{N_{s,t,x-1} + N_{s,t,x}}{2}$ on aastakeskmise naiste arv ja ${}_1f_t^x$ on hinnatud sündimuse vanusekordaja aastal x .

Viimases vanuses oli mudel kujul:

$$N_{s,t,x} = N_{s,t,x-1} \cdot q_t^x + N_{s,t,x} \cdot q_t^{x+1} + N_{s,t,x-1} \cdot \frac{r_t^x}{1000},$$

kus $N_{s,t,x} \cdot q_t^{x+1}$ on sama vanusegrupi ellujäänute arv, millega arvestatakse, kuna antud vanusegrupist pole võimalik edasi liikuda.

Mudeli eeldusteks olid ekspertarvamuste alusel koostatud hinnangud Eesti rahvas-
tiku sündimuse, suremuse ja välisrände kohta aastateks 2045 ja 2080. Loodi neli
stsenaariumi (Statistikaamet, [2019](#)):

- **Põhistsenaarium:**

- Summarne sündimuskordaja tõuseb 2045 aastaks 1,81 lapseni naise koh-
ta ja 2080 aastaks 1,86 lapseni;
- eluiga pikeneb 2045 aastaks naistel 86,5 ja meestel 79,6 eluaastani ning
2080 aastaks naistel kuni 89 ja meestel kuni 83,5 aastani;
- 1500 inimesega positiivne rändesaldo.

- **Kõrgem sündimus ja rändesaldo:**

- Summaarne sündimuskordaja tõuseb 2045. aastaks 1,90-ni ja 2080. aas-
taks 1,95-ni;
- eluiga pikeneb 2045. aastaks naistel ja meestel 86,5 ja 79,6 aastani ning
2080. aastaks 89 ja 83,5 aastani;
- suurem positiivne rändesaldo, alguses 3000 ja lõpus 2500 inimest.

- **Madalam sündimus, madalam suremus, ränne tasakaalus:**

- Summaarne sündimuskordaja tõuseb 2045. aastaks 1,72-ni ja 2080. aas-
taks 1,77-ni;
- oodatav eluiga sünnil jääb samaks naistel 82,4 ja meestel 73,7 aastat;
- välisränne on tasakaalus.

- **Sündimus tõuseb:**

- Summaarne sündimuskordaja tõuseb oluliselt, aastaks 2045 2,03-ni ja
aastaks 2080 2,08-ni;

- eluiga pikeneb 2045. aastaks naistel ja meestel 86,5 ja 79,6 aastani ning 2080. aastaks 89 ja 83,5 aastani;
- 1500 inimesega positiivne rändesaldo.

Kolmanda stsenaariumi korral tähendab madalam suremus oodatava eluea samaks jäämist.

Stsenaariumites välja toodud ennustatavad summaarsed sündimuskordajad

Summaarne sündimuskordaja ennustati lähteaastast lineaarselt sihtaasta ekspert-hinnaguni muutuma. Seejärel jaotati summaarne sündimuskordaja viljakale eale vastavateks sündimuse vanusekordajateks. Seda tehti arvestades keskmist sündide jaotust üle viljaka ea varasematel aastatel. Samamoodi jaotati oodatavast east saadud suremuskordaja suremuse vanusekordajaks ning rändesaldo kõigi vanuste vahel.

1.2 Tõenäosuslikud rahvastikuproгноosi meetodid

Tõenäosuslike rahvastikuproгноoside koostamise meetodid hõlmavad retrospektiivset analüüsi, aegridade analüüsi ja eksperdipõhiseid lähenemisi. Retrospektiivne analüüs põhineb varasemate prognooside vigade analüüsil. Aegridade analüüs kasutab prognoosi sisendite, nagu sündimus ja suremus, mineviku andmeid, et hinnata statistiline aegrea mudel, mida seejärel kasutatakse suure hulga juhuslike tulevikutrajektooride simuleerimiseks. Simuleeritud trajektooridest koostatakse osismetodi abil hinnatavaid jaotusi prognoosi väljundile. Eksperdipõhises meetodis palutakse ekspertidel koostada iga prognoosi sisendi jaotuseid, mida seejärel kasutatakse prognoosi väljundite hinnatavate jaotuste koostamiseks sarnaselt aegrea meetodile. (Raftery & Ševčíková, 2023)

1.2.1 R-i paketi bayesPop Bayesiaanlik mudel

R-i paketiga bayesPop saab koostada tõenäosuslikke rahvastikuprognose sama metoodikaga mida kasutatakse ÜRO rahvastikuprognosides. Mudel kasutab nende rahvastikuprognoside loomiseks summaarse sündimuskordaja ja oodatava eluea tõenäosuslikke hinnanguid. Summaarse sündimuskordaja ja oodatava eluea hinnanguite loomiseks kasutatakse hierarhilisi Bayesi mudeleid. Luuakse võrdne arv sündimuse, suremuse ja rände trajektoore. Saadud summaarse sündimuskordaja ja oodatava eluea hinnangud teisendatakse vanusejärgseteks ja rahvaarvu prognoositakse osismeetodiga igal simuleeritud sündimuse ja suremuse trajektoorigil. Osismeetodit aluseks on demograafiline bilansivõrrand (1). (Ševčíková & Raftery, 2016)

Sündimus

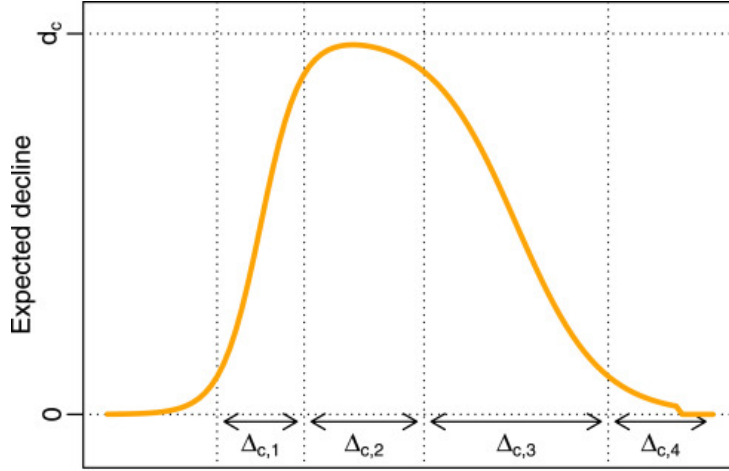
Summaarset sündimuskordajat prognoosides eristatakse kolme järku vastavalt sellele, kui kaugel on prognoosimise hetkel prognoosile vastav populatsioon sündimuskäitumise muutumise protsessis. I järk on aeg enne revolutsiooni kui sündimus on kõrge. II järk on periood, millel jooksul summaarne sündimuskordaja langeb alla taastetaseme 2,1. III järk on madala sündimusega periood. I järk on kõigis maailma riikides möödunud, seega sellega mudelis ei arvestata. (Sevcikova *et al.*, 2023)

II järgus on oodatavat langust summarses sündimuskordajas hinnatud kahekordse logistilise funktsiooniga mille kuju on järgnev:

$$g(f; \theta) = \frac{d}{1 + \exp\left(-\frac{2\log(9)}{\Delta_1} \left(f - \sum_{i=2}^4 \Delta_i + 0.5\Delta_1\right)\right)} - \frac{d}{1 + \exp\left(-\frac{2\log(9)}{\Delta_3} (f - \Delta_4 - 0.5\Delta_3)\right)}, \quad (2)$$

kus $\theta = (d, \Delta_1, \Delta_2, \Delta_3, \Delta_4)$ ja:

- d on oodatava summaarse sündimuskordaja languse ülemine piir, $d > 0$,



Joonis 1: Summaarse sündimuse oodatav langus II järgus vastavalt kahekordsele logistilisele funktsioonile. (Raftery & Ševčíková, 2023)

- esimese logistilise funktsiooni keskpunkt on $0.5\Delta_3 + \Delta_4$, kus Δ_3 on 80% esimese logistilise funktsiooni ulatusest,
- teise logistilise funktsiooni keskpunkt on $0.5\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4$, kus Δ_1 on teise logistilise funktsiooni 80%-lise ulatuse määrang (Raftery & Ševčíková, 2023).

Kuna kõik riigid pole sündimuskäitumise muutumise protsessi II järgus piisavalt kaugelt jõudnud, et neile oodatava summaarse sündimuskordaja languse funktsiooni soovitava täpsusega sobitada, siis kasutatakse teiste riikide andmeid. Kogu maailma võimalikke sündimuse kahanemise mustreid täpsustatakse prognoositava riigi seniste andmetega, et sobitada sündimuse languse kiirust. Eeldatakse, et iga riigi summaarne sündimuskordaja eksleb juhuslikult, nii et ekslemine on kahekordne logistiline funktsioon hetke sündimusest. (Raftery & Ševčíková, 2023)

Mudelil on II järgus kolm taset. Esimene on riigi senisest sündimusest tuletatav sündimus:

$$f_{c,t+1} = f_{c,t} - g(f_{c,t}; \theta_c) + \varepsilon_{c,t},$$

kus $f_{c,t}$ on riigi c summaarne sündimuskordaja perioodil t , $g(f_{c,t}; \theta_c)$ on oodatav langus nagu see on antud valemis 2 ja $\varepsilon_{c,t}$ vea komponent, mille dispersioon sõltub hetke summaarsest sündimuskordajast nii, et $\varepsilon_{c,t} \stackrel{\text{ind}}{\sim} N(0, \sigma(f_{c,t})^2)$. Tase 2 annab sündimuse jaotuse maailmas $\theta_c \sim h(\cdot | \phi)$, kus $\theta_c = (\Delta_{c,1}, \Delta_{c,2}, \Delta_{c,3}, \Delta_{c,4}, d)$ on riigi c oodatava languse funktsiooni parameetrid ja tase 3 täpsustab maailma parameetrite jaotuse $\phi \sim \pi(\cdot)$ (Raftery & Ševčíková, 2023).

II järgus leitakse mudeli jaoks sobivad jaotusparameetrid Bayesi hindamisega kasutades Markovi ahelate Monte Carlo valikut aposterioorsest jaotusest. Parameetreid uuendatakse ühekaupa kasutades Gibbsi valiku algoritmi, Metropolise algoritmi või ka genereerimist tinglikust jaotusest. (Raftery & Ševčíková, 2023)

III järgu alguseks defineeritakse kaht viieaastast perioodi, mille jooksul summaarne sündimuskordaja oli alla kahe, kuid siiski tõusis (Raftery & Ševčíková, 2023). Ka Eesti on jõudnud III järku. Seega Eesti rahvastikku prognoosides rakendatakse just sellele järgule vastavat algoritmi.

III järgu üheks eelduseks on riigi c sündimuse kõikumine ümber lõpliku sündimuse taseme μ_c . Seda kõikumist modelleeritakse esimest järku autoregressioonimudeliga:

$$\begin{aligned} f_{c+1} &= \mu_c + \rho_c(f_{c,t} - \mu_c) + \epsilon_{c,t}, \\ \epsilon_{c,t} &\stackrel{\text{iid}}{\sim} N(0, \sigma^2), \end{aligned}$$

kus $0 \leq \rho_c \leq 1$. (Raftery & Ševčíková, 2023)

Ka III järgus kasutatakse summaarse sündimuse hindamiseks hierarhilisi Bayesi mudeleid. Mudeli teisel tasemel modelleeritakse parameetreid μ_c ja ρ_c :

$$\begin{aligned} \mu_c &\sim TN_{[0,\infty)}(\bar{\mu}, \sigma_\mu^2) \\ \rho_c &\sim TN_{[0,1]}(\bar{\rho}, \sigma_\rho^2), \end{aligned}$$

kus $\bar{\mu}$ on maailma riigipõhiste lähendite keskmine, σ_μ^2 on nende dispersioon ja $TN_{[0,\infty)}$ tähistab kärbitud normaaljaotust. Maailma keskmise jaoks kasutatakse

kindlat eeljaotust:

$$\bar{\mu} \sim U[0, 2.1].$$

See tähendab, et maailma keskmise maksimaalne väärtus on eeldatav taastetase-
mel sündimus. (Raftery & Ševčíková, 2023) Samas σ_μ , $\bar{\rho}$ ja σ_ρ eeljaotused pole
ekspertarvamuse põhjal määratud. Nende kolme parameetri eeljaotustel on vaid
ekspertarvamuse kohaselt mõistlike ulatustega (Raftery & Ševčíková, 2023).

Oodatav eluiga

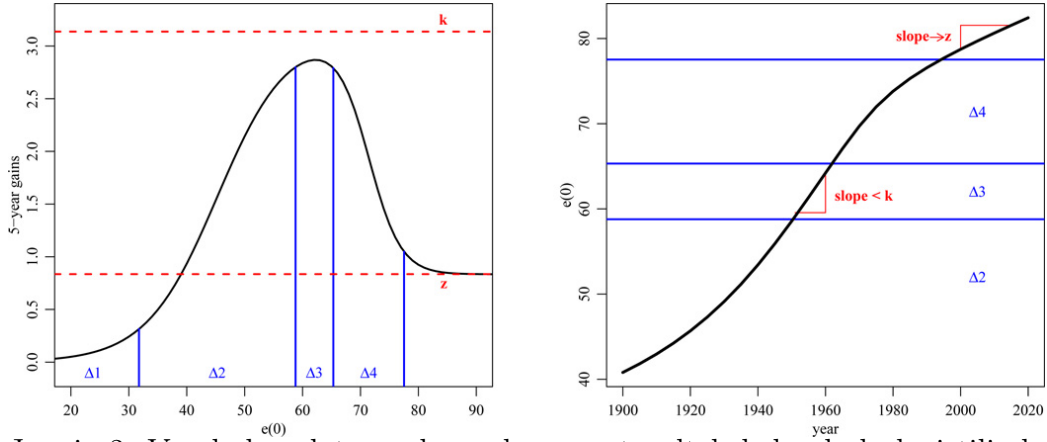
Oodatava eluea kasvu esitamiseks kasutatakse kahekordset logistilist funktsiooni het-
ke oodatavast elueast

$$\begin{aligned} g(l_{c,t}|\theta^c) = & \frac{k^c}{1 + \exp\left(-\frac{A_1}{\Delta_2^c}(l_{c,t} - \Delta_1^c - A_2\Delta_2^c)\right)} \\ & + \frac{z^c - k^c}{1 + \exp\left(-\frac{A_1}{\Delta_4^c}\left(l_{c,t} - \sum_{i=1}^3 \Delta_i^c - A_2\Delta_4^c\right)\right)}, \end{aligned} \quad (3)$$

kus $\theta^c = (\Delta_1^c, \Delta_2^c, \Delta_3^c, \Delta_4^c, k^c, z^c)$ on on kuuest parameetrist koosnev parameetri-
te vektor riigi c kahekordse logistilise funktsiooni jaoks. Parameetrite tähendust
on illustreeritud joonisel 2. Parameetrite vektor θ^c riigi c jaoks on valitud ÜRO
analüütiku poolt viiest võimalusest, mis vastavad ÜRO rahvastikuproгноosi stse-
naariumitele. Konstandid $A_1 = 4.4$, $A_2 = 0.5$ on valitud nii, et parameetrid Δ_i^c
($i = 1, 2, 3, 4$) on interpreteeritaval skaalal. Neid konstante saab muuta ilma tule-
musi muutmata, eeldusel, et nende korrutis, A_1A_2 , jääb muutumatuks. (Raftery
et al., 2013)

Hierarhiline mudel oodatava eluea modelleerimiseks on

$$l_{c,t+1} = l_{c,t} + g(l_{c,t}|\theta^c) + \varepsilon_{c,t+1},$$



Joonis 2: Vasakul oodatava elueea kasv vastavalt kahekordsele logistilisele funktsioonile. Paremal oodatav eluiga, kus kasvu on modelleeritud kahekordse logistilise funktsiooniga. Jaapani näide. (Raftery *et al.*, 2013)

kus

$g(l_{c,t}, \theta^c) =$ kahekordne logistiline funktsioon parameetritega θ^c ,

$$\theta^c = (\Delta_1^c, \Delta_2^c, \Delta_3^c, \Delta_4^c, k^c, z^c),$$

$$\Delta_i^c \mid \sigma_{\Delta_i} \stackrel{\text{iid}}{\sim} \text{Normal}_{[0,100]}(\Delta_i, \sigma_{\Delta_i}^2), \quad i = 1, \dots, 4,$$

$$k^c \mid \sigma_k \stackrel{\text{iid}}{\sim} \text{Normal}_{[0,10]}(k, \sigma_k^2),$$

$$z^c \mid \sigma_z \stackrel{\text{iid}}{\sim} \text{Normal}_{[0,1.15]}(z, \sigma_z^2),$$

kus $\text{Normal}_{[a,b]}(\mu, \sigma^2)$ tähistab normaaljaotust keskmisega μ ja standardhällbega σ , mis on kärbitud väärtuste a ja b vahele. (Raftery *et al.*, 2013)

Tänu eeldusele, et ühe riigi parameetrid on juhuslikult valitud ühisest normaaljaotusest, kasutab mudel ära teiste riikide oodatava eluea kasvu andmeid. Normaaljaotused on kärbitud, et parameetrid oleksid positiivsed. z^c on oodatava eluea kasvu asümptootiline keskmine. (Raftery *et al.*, 2013)

Juhusliku vea jaotuseks on (Raftery *et al.*, 2013):

$$\varepsilon_{c,t} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, (\omega \times f(l_{c,t-1}))^2).$$

Parameetrite eeljaotused on määratud ÜRO mudelite ja stsenaariumite järgi. Spetsiifiline eeljaotuste valik ei mõjuta tulemusi. Aposterioorsed jaotused hinnati Markovi ahelate Monte Carlo valikut kasutades. Parameetreid uuendati Gibbsi valiku algoritmi, Metropolis-Hastingsi algoritmi või Hastingsi valiku abil või ka tinglikust jaotusest genereerides.

Välisränne

Välisrände hinnanguna kasutatakse ÜRO rahvastikuproгноosi rändesaldo hinnanguid (Sevcikova *et al.*, 2023). Rändesaldo eeldused on seatud mineviku rände hinnangute põhjal. Eeldatud tase on seejärel konstantsena tulevikku püsima jääma prognoositud (United Nations, 2022). Eesti puhul on aastaks 2022 ja edasi aastani 2100 eeldatud aastaseks rändesaldoks -1000 ehk tuhat inimest lahkuks rohkem kui Eestisse sisse rändaks.

Kuna ÜRO ei avalda rände projektsioone soo ja vanuse järgi, siis arvutatakse mudelis esiteks rahvastikuproгноos ilma rändeta. Saadavad jäägid võrreldes ÜRO avaldatud rahvastikuproгноosiga võetakse rändesaldo hinnanguks. Saadud vanuse ja soo rändesaldod skaleeritakse seejärel nii, et summa üle sugude ja vanuste vastab terviklikule ÜRO poolt hinnatud rändesaldole. (Sevcikova *et al.*, 2023)

1.2.2 R-i pakett demography

Pakett demography hindab suremust, sündimust ja rännet funktsionaalseid andmemudeleid (*functional data model*) kasutades. Suremuse, sündimuse ja rände hinnanguid kasutatakse osismeetodis tõenäosuslike rahvastikuproгноoside loomiseks soo ja vanuse järgi. Tõenäosuslik rahvastikuproгноos saadakse Monte Carlo simuleerimise abil. Eeldatakse, et mineviku trendid jätkuvad tulevikus. (Hyndman & Booth, 2008)

Funktsionaalsed andmemudelid

Modelleeritav väärtus, $y_t^*(x)$, on kas suremuse vanusekordaja, sündimuse vanusekordaja või rändesaldo vanusele x aastal t . Leitakse $y_t^*(x)$ Boxi-Coxi teisendus, et see normaalkaotusele sarnanevaks transformeerida:

$$y_t(x) = \begin{cases} \frac{1}{\lambda} ([y_t^*(x)]^\lambda - 1) & \text{kui } 0 < \lambda < 1; \\ \ln(y_t^*(x)) & \text{kui } \lambda = 0. \end{cases}$$

(Hyndman & Booth, 2008)

Boxi-Coxi teisendusega väärtusele eeldatakse mudel:

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{\lambda,x}$$
$$s_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k}\phi_k(x) + e_t(x),$$

kus $s_t(x)$ on silutud funktsiooni x -st. Mudeli vead $\varepsilon_{t,x} \sim IID(0, 1)$ on sõltumatu ja sama jaotustega. (Hyndman & Booth, 2008) See eeldus on oluline, et tagada mudeli hinnangute mittekorrigeeritus. $\sigma_\lambda(x)$ võimaldab dispersioonil muutuda koos vanuse ja ajaga, mis on vajalik, et arvestada erinevate vanusegruppide dispersioonide erisusi.

$\mu(x)$ esindab keskmist funktsionaalset sõltuvust vanusest x . Ortogonaalsed baasfunktsioonid $\{\phi_k(x)\}$ on arvutatud peakomponentide meetodiga (PCA) ja valitud selleks, et väljendada mudeli funktsiooni kuju ja dünaamikat. Need võimaldavad esitada keerukaid sõltuvusi vanusest, mis andmetes esinevad. Aegrea kordajad $\{\beta_{t,k}\}$ lisavad mudelile võime kohanduda ajas toimuvate muutustega. Eeldatakse, et need on üksteisest sõltumatud. Juhuslik weakomponent $e_t(x)$ lisab mudelile juhuslikkuse. (Hyndman & Booth, 2008)

Modelleerimise etapid on (Hyndman & Booth, 2008):

1. Siledade funktsioonide $s_\lambda(x)$ hindamine mitteparameetrilise regressiooniga,

mida rakendatakse suurusele $y_t(x)$ iga aasta t kohta.

2. $\mu(x)$ hindamine $s_t(x)$ keskmisena üle aastate.
3. $\beta_{t,k}$ ja $\phi_k(x)$, $k = 1, \dots, K$ hindamine kasutades $[y_\lambda(x) - \mu(x)]$ peakomponente.
4. β_k , $k = 1, \dots, K$ aegridade mudelite hindamine.

Meetod ei ole K suuruse suhtes tundlik, kui K on piisavalt suur. Suure K valimisega ei kaasne peale arvutusaja kasvu muid probleeme. Liiga väikese K valimine võib langetada hinnangu täpsust. Seetõttu valitakse selles analüüsis $K = 6$ kõigi komponentide jaoks. $K = 6$ on suurem, kui ükskõik milline komponent vajab. (Hyndman & Booth, 2008)

Dispersioon $\sigma_\lambda^2(x)$ sõltub andmetest. Suremuse korral hinnatakse seda $y_t^*(x) = m_t(x)$, eeldusega, et surmad järgivad Poissoni jaotust parameetriga $m_t(x)E_t(x)$. Seega on $y_t^*(x)$ ligikaudselt dispersiooniga $E_t^{-1}(x)m_t(x)$ ning $y_t(x)$ dispersioon on Taylори reaks arendades ligikaudselt

$$\sigma_t^2(x) \approx [m_t(x)]^{2\lambda-1} E_t^{-1}(x).$$

Sündimuse puhul eeldatakse Poissoni jaotust parameetriga $f_t(x)E_t^F(x)$, mis annab

$$\sigma_t^2(x) \approx [f_t(x)]^{2\lambda-1} E_t^{-1}(x).$$

Rände puhul ei eeldata ühtki jaotust ja $\sigma_\lambda^2(x)$ hinnatakse mitteparameetrilise regressiooniga $[y_t(x) - s_t(x)]^2$ -st x suhtes. (Hyndman & Booth, 2008)

Juhul kui on olemas andmed ajani $t = n$ siis hinnatakse tulevasi väärtusi $y_t(x)$ kui $t = n + 1, \dots, n + h$ kõigi vanuste x korral. $\hat{\beta}_{n,k,h}$ tähistab h -sammu ette prognoosi $\beta_{n+h,k}$, $\hat{y}_{n,h}(x)$ aga $y_{n+h}(x)$ h -sammu ette prognoosi ning $\hat{s}_{n,h}(x)$ tähistab $s_{n+h}(x)$

h -sammu ette prognoosi. Siis hinnang

$$\hat{y}_{n,h}(x) = \hat{s}_{n,h}(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}_{n,k,h} \hat{\phi}_k(x).$$

$y_t^*(x)$ leitakse Boxi-Coxi pöördteisendus abil.

Kohordi tõenäosuslik simulatsioon funktsionaalse andmemudeliga

Rahvastikku simuleeritakse osismeetodiga. Baastasemeks valitakse aasta $n + 1$ 1. jaanuar. Suremuse, sündimuse ja rändesaldo jaoks simuleeritakse suur hulk tuleviku trajektoore, mida kasutatakse soo- ja vanusepõhiste tuleviku trajektooride loomiseks. Iga soo- ja vanusepõhise trajektoori jaoks genereeritakse aegridade mudelitega tinglikult $\beta_{1,k}, \dots, \beta_{n,k}$ suhtes $\beta_{t,k}$ iga $t = n + 1, \dots, n + h$ korral. Geneereeritakse ka juhuslikud suurused $e_t(x)$. (Hyndman & Booth, 2008)

Migrantide surmade ja sündidega arvestamiseks kohandatakse populatsiooni esmalt lisades pool aastasest rändesaldost aasta alguses ja pool aastat lõpus (Hyndman & Booth, 2008). Eeldatakse, et sündmused toimuvad migrantidega keskmiselt poole aasta jooksul (Preston, Heuveline & Guillot, 2001). Rändesaldo esimese poolega kohandatud populatsiooni tähistame R -ga, seega $R_t(x) = P_t(x) + G_t(x, x + 1)/2$, kus $G_t(x, x + 1)$ on vanuse x rändesaldo aasta $n + h$ alguses. Imikute puhul aastat t lisatakse pool $G_t(B, 0)$ sündidele aastal t , st $R_t(B) = B_t + G_t(B, 0)/2$, ja pool lisatakse populatsioonile vanuses 0 aasta t lõpus. Kahe vanima vanuse puhul eeldatakse, et $G_t(p - 1^+, p+)$ jaguneb võrdselt $p - 1$ aasta ja $p+$ aasta vanuste migrantide vahel ning see liidetakse aasta t alguses. (Hyndman & Booth, 2008)

Teisena lisatakse mõlemale soo gruppide suremus. Sugude jaoks võetakse üks simuleeritud trajektoor $\{m_t(x)\}$, kus $t = n + 1, \dots, n + h$. Aastal t vanuses x toimuvate surmade juhusliku arvu saamiseks on vaja aasta keskmist populatsiooni $E_t(x)$, mis omakorda sõltub surmade arvust. Tekib ring, surmade arv sõltub rahvaarvust, mis jälle sõltub surmade arvust. See lahendatakse järgmiselt: kasutades

simuleeritud suremusi $\{m_t(x)\}$ saame elutabelist vanuse suremuskordajad. Raken-
dades saadud suremuskordajaid $R_t(x)$ -le saadakse oodatav kohordi surmade arv
 $\bar{D}_t(x, x+1)$. Need surmad lahutatakse $R_t(x)$ -st, et saada $\bar{R}_{t+1}(x+1)$, mis on hin-
nang $R_{t+1}(x+1)$ -le. Võetakse $R_t(x)$ ja $\bar{R}_{t+1}(x+1)$ keskmine, et saada hinnang
aastakeskmisele rahvaarvule $\bar{E}_t(x)$. Surmade arvu vanuses x aastal t saame Poissoni
jaotusest keskmisega $m_t(x)\bar{E}_t(x)$ juhuslikku valimit võttes. Üle iga vanuserühmade
naabrite paari võetakse aasta t keskmine, et saada kohordi surmad $D_t(x, x+1)$.
 $D_t(x, x+1)$ lahutatakse $R_t(x)$ -st, et saada $R_{t+1}(x+1)$. Kahe vanima vanuserühma
puhul on kohordi surmade arv $D_t(p-1, p+)$ summa $D_t(p-1)/2 + D_t(p+)$. Kohordi
rändesaldo teise poole lisamine annab rahvaarvu vanuses $x+1$ järgmise aasta 1.
jaanuaril $P_{t+1}(x+1)$. (Hyndman & Booth, 2008)

Sünnituste arv saadakse kasutades simuleeritud sündimuse vanusekordaja $f_t(x)$
ja poole rändesaldoga rahvaarvu $R_t(x)$. Eeldatakse, et sündimus vanuses x järgib
Poissoni jaotust keskmisega $f_t(x)[R_t(x) + R_t(x+1)]\frac{1}{2}$, $x = 15, \dots, 49$. Juhuslik
valim sellest jaotusest määrab sünnituste arvu $B_t(x)$ aastal t naistele vanuses x .
Sündide arv aastas, B_t , arvutatakse valemiga

$$B_t = \sum_{x=15}^{49} B_t(x).$$

Sünnid jaotatakse binoomjaotuse järgi. Meessoost sünni tõenäosus on $\frac{\rho}{(\rho+1)}$, kus ρ
on mees- ja naissoost sündimuse suhe. (Hyndman & Booth, 2008)

Nii meeste kui ka naiste puhul rakendatakse suremuse hinnangut $\bar{D}_t(B, 0)$ rahvas-
tikule $R_t(B)$, et saada esimese vanusegrupi suurus $R_{t+1}(0)$. Suremuse hindamiseks
kasutatakse varem tuletatud elutabelist suremuse vanusekordajat sünnist vanuseni
0. Suremuse hinnang lahutatakse $R_t(B)$ -st, et saada hinnang $\bar{R}_{t+1}(0)$. Siis võetak-
se $R_t(0)$ ja $\bar{R}_{t+1}(0)$ keskmine, et hinnata aastakeskmist rahvaarvu $\bar{E}_t(0)$. Aasta t
surmad vanuses 0 saadakse parameetriga $D_t(0) = m_t(0)\bar{E}_t(0)$ Poissoni jaotusest
juhuslikku valimit võttes. See juhuslikult tõmmatud surmade arv, $D_t(0)$, jagatakse
seejärel surmadeks aastal t sündinute seas, $D_t(B, 0)$, ja surmadeks vanuses 0 aastal

$t - 1$ sündinute seas. (Hyndman & Booth, 2008)

Kasutades sündimuse, suremuse ja rände simulatsioone saadakse järgmise ja sellele järgnevate aastate rahvaarv. Kuna seda korratakse kõigil loodud trajektooridel, siis saadakse hulk tuleviku rahvaarvu hinnanguid (Hyndman & Booth, 2008). Saadud hulgast saadakse ka prognoosi juhuslikkuse hinnang.

1.3 Mikrosimulatsiooni meetodid

Rahvastiku mikrosimulatsiooni (*demographic microsimulation*) kasutatakse rahvastiku protsesside analüüsimiseks ja rahvastikuprognooside loomiseks. Modelleeritakse indiviidide elukulgu, eeldades, et nende indiviididega juhtuvad sündmused toimuvad määratud tõenäosuslike reeglite tagajärjel. Erinevate olekute vahel liikumist simuleeritakse Monte Carlo meetodite abil. (Zagheni, 2015)

Olek on indiviidi demograafiliste tunnuste vektor (Tensing, 2012).

Iga indiviidi kohta on teada teatud demograafilised tunnused nagu vanus, sugu, haridustase ja perekonnaseis. Indiviidi tunnused mõjutavad tema olekute vahel liikumise tõenäosusi. Olekute vahel liikumise sündmused võivad olla näiteks suremine, abiellumine, lapse sünd ja ka sisse või välja rändamine. Igal indiviidil on mingi aja jooksul tõenäosus, et juhtub mingisugune sündmus, kuid kõik sündmused ei pea kõigile indiviididele alati võimalikud olema. Näiteks ei saa lesk abielu lahutada. Aeg sündmuse toimumiseni leitakse tõenäosuslikult. Aeg järgmise sündmuseni tekitatakse juhuslikult ja indiviidiga järgmisena juhtuv sündmus on see, milleni tal on vähim aega. (Zagheni, 2015)

Mikrosimulatsioon võib toimuda pideva või diskreetse ajaga. Pideva ajaga mikrosimulatsioonides luuakse igale indiviidile aeg iga sündmuseni, mille juhtumise tõenäosus vastaval indiviidil on. Iga sündmuse risk sõltub indiviidi demograafilistest tunnustest. Diskreetse ajaga simulatsioonides modelleeritakse ajavahemikke, mitte aega sündmuste toimumiseni. Kindel sündmus tohib ühel ajavahemikul toimuda

vaid korra, ühel ajavahemikul tohib toimuda mitu erinevat sündmust. (Zagheni, 2015)

Neid mikrosimulatsioone, kus indiviide saab lisanduda vaid sündmise kaudu, nimetatakse kinnisteks mudeliteks. Vastupidiselt, simulatsioone, kus indiviide saab ka muul viisil lisanduda, kutsutakse avatuteks. Peamine põhjus, miks kinniseid ja avatud mudeleid eristatakse on indiviidile abikaasa leidmise meetod. Kinnistes mudelites saab abikaasa olla vaid teine valimis olev indiviid ning kahe indiviidi sobitamine vajab keerukaid algoritme. Avatud mudelites, aga luuakse vajadusel uusi indiviide. Kinnistes mudelites modelleeritakse rännet luues ülejäänud maailma esindav populatsioon, millest või millesse indiviidid või indiviidide grupid võivad liikuda. (Zagheni, 2015)

Algne populatsioon tuleb tavapäraselt kas kogu rahvastiku rislõikelisest valimist või kogu rahvastiku tunnuste põhjal sünteesitud populatsioonist (Zagheni, 2015).

Mikrosimulatsioonimudelid simuleerivad rahvastikusündmusi rahvastiku makroandmete põhjal. Agendipõhised mudelid simuleerivad teatud käitumismustritega indiviide, kes mõjutavad üksteist vastastikku. Mikrosimulatsioonimudelid keskenduvad makro-mikro interaktsioonide modelleerimisele ning neid kasutatakse poliitiliste otsuste, pikaajaliste rahvastikutrendide ja demograafiliste šokkide mõjude hindamiseks erinevatele huvipakkuvatele muutujatele. Agendipõhised mudelid keskenduvad mikro-makro suunale ning neid kasutatakse sageli teooriate testimiseks ning emergentsete omaduste hindamiseks. (Zagheni, 2015)

Mikrosimulatsioone on rahvastikuproгноosis rakendatud näiteks projekti QuantMig raames. QuantMig-Mic on mikrosimulatsioonimudel, mis on loodud Euroopa Liidu riikide (sh Ühendkuningriik, Island, Norra ja Šveits) rahvastiku prognoosimiseks, keskendudes sellele, kuidas sünnikoht mõjutab erinevaid demograafilisi ja majanduslikke käitumismustreid. Mudel kasutab Monte Carlo meetodeid individuaalsete elukäikude simuleerimiseks pideva ajaga. QuantMig-Mic on CEPAM-Mic mudeli edasi arendus. Mudel võimaldab integreerida ja rakendada erinevaid mig-

ratsioonistsenaariume, uurides nende mõju rahvastiku struktuurile ja tööjõu pakumisele. Mudeli kood ja stsenaariumid on avalikult kättesaadavad Zenodo andmebaasi. (Marois, Potančoková & González-Leonardo, [2023](#))

Mikrosimulatsioonide tegemiseks on olemas ka prognoosiprogramm MicMac ja R-i pakett MicSim.

Prognoosiprogramm MicMac

Prognoosiprogramm MicMac on Euroopa Liidu Ühise Teadusuuringute Keskuse poolt arendatud rahvastikuprognoosimise tarkvara, mille eesmärk on ühendada mikro- ja makrotasandi analüüsid demograafilistes uuringutes. See loodi vajaduse tõttu süsteemselt käsitleda individuaalset käitumist ja selle mõju rahvastiku üldisele dünaamikale. MicMac võimaldab analüüsida, kuidas sotsiaal-majanduslikud tegurid, nagu haridustase ja tööhõive, mõjutavad isikute eluteid ning kuidas need individuaalsed trajektoorid omakorda mõjutavad rahvastikutrende.

MicMac on terviklik agendipõhine rahvastikumudel, mis ühendab mikro- ja makrotasandi demograafilised protsessid, et analüüsida rahvastiku dünaamikat. Mudel hindab rahvastiku arengut läbi individuaalsete eluteede ja kohortide analüüsi, võttes arvesse vanuse, soo ja muude demograafiliste tunnustega seotud sündmuste tõenäosusi. MicMac on R-i põhine, kuid vajab töötamiseks ka Java Runtime Environmenti. (Tensing, [2012](#))

MicMac vajab sisendina mudeli võimalikke olekuid, rahvastiku vanusevahemikke, simulatsiooni algus- ja lõppaastat, üleminekute tõenäosusi, mis sõltuvad teistest olekutest ja esialgset populatsiooni. Kaasata saab ka rände. Need sisendid peab mudelile ette andma kindlas formaadis. (Tensing, [2012](#))

R-i pakett MicSim

MicSim on demograafiliste uuringute jaoks loodud R-i pakett, mis võimaldab teha pideva ajaga mikrosimulatsioone. Pakett on mõeldud kasutajatele, kellel ei ole varasemat sügavat kogemust R-i või mikrosimulatsioonidega, pakkudes lihtsat sisendit ja väljundit, mis on kohandatud demograafide vajadustele. (Zinn, 2014)

Pakett võimaldab simuleerida individuaalseid elukäike, mis koosnevad erinevatest olekutest ja olekute vaheliste intervallide kestustest. Elukäikude dünaamika põhineb Markovi ahelatel, mis modelleerivad juhuslike sündmuste mõju indiviidide käitumisele, arvestades nii vanust, kuupäeva kui ka olekutes viibitud aega. See mitmekülgsus võimaldab MicSimil pakkuda detailset vaadet rahvastiku dünaamikale. (Zinn, 2014)

MicSimi uuendatud versioonid on parandanud andmete käitlemise kiirust ja võimaldavad laiendatud funktsionaalsust, nagu näiteks uute sündmuste lisamine simulatsioonide ajal, mis teeb sellest tõhusa tööriista rahvastiku muutuste prognoosimiseks ja analüüsimiseks. (Bosco *et al.*, 2023)

1.4 Masinõppe ja närvivõrkude rakendamine rahvastikuproгноosis

Huvi võib pakkuda ka masinõppe ja närvivõrkude kasutamine rahvastikuproгноosis. Autor ei ole leidnud ühtki näidet, kus seda teeks rahvusvaheline organisatsioon või riiklik statistikaamet. Siiski on kirjutatud selle kohaseid uurimusi. (Şahinarslan, Tekin & Çebi, 2021)

Uuringus *Application of Machine Learning Algorithms for Population Forecasting* analüüsitakse erinevate masinõppemeetodite tõhusust rahvastikuproгноoside koostamisel. Enim keskendutakse XGBoostile (*Extreme gradient boosting*) ja CatBoost algoritmile, kuna need osutusid kõige edukamaks. (Şahinarslan, Tekin & Çebi, 2021)

Uuringus kasutati masinõppe mudeleid nagu XGBoost, CatBoost, lineaarne regressioon, ridge-regressioon ja ARIMA mudelid, et prognoosida Türgi kogu rahvastiku suurus. Mudelite treenimiseks kasutati Maailmapanga andmestikku, mis hõlmas 1595 demograafilist näitajat 262 erinevast riigist aastast 1960 kuni aastani 2017. Masinõppe mudelite võimekust demonstreeriti võrdluses traditsioonilise osismetodiga. Leiti, et osismetod oli üheks aastaks prognoosides nõrgem. (Şahinarslan, Tekin & Çebi, 2021)

Allpool on esitatud tabel 2 artiklist (Şahinarslan, Tekin & Çebi, 2021), mis näitab erinevate masinõppe mudelite rahvastiku hinnanguid võrreldes Türgi 2017. aasta rahvastikuprognoosiga, mida kasutatakse baasmodelina mudelite tõhususe demonstreerimiseks. Prognoose võrreldi Türgi 2017. aasta rahvaarvu vastu, mis oli 80 745 020.

Tabel 1: Türgi 2017. aasta rahvastikuprognoosid erinevate mudelitega. (Şahinarslan, Tekin & Çebi, 2021)

Mudel	MAPE
Lineaar regressioon	0.0493%
Ridge-regressioon	0.0074%
XGBoost	0.0011%
CatBoost	0.0012%
Holt-Winters	0.0104%
ARIMA	0.0080%
Prophet	1.0100%
Osismetod	0.5297%
Türgi Statistikaamet	1.2125%

Tulemusi on hinnatud keskmise absoluutse protsentveaga (6). Tulemused näitavad, et XGBoost ja CatBoost algoritmid saavutasid kõige täpsemaid tulemusi, mis annab mõista, et masinõppel põhinevatel meetoditel on potentsiaali demograafiliste muutuste prognoosimisel.

Uuringus *Comparing Artificial Neural Network and Cohort-Component Models for Population Forecasts* võrreldakse pika lühiajalise mälu (LSTM) võrkude ja osis-

meetodi tõhusust Alabama rahvastikuproгноosi koostamisel. Kaasatud on kõik 67 Alabama maakonda. Tulemused näitavad, et kuigi osismetod on üldiselt täpsem, on teatud juhtudel LSTM närvivõrgud võimelised andma võrreldavaid tulemusi, eriti kui kasutatakse majandusandmeid või kui igale maakonnale on treenitud eraldi mudel vaid selle maakonna andmetega. (Riiman *et al.*, 2019)

Uuringus kasutatakse Ameerika Ühendriikide Rahvaloenduse Büroo andmeid aastakeskmiste rahvaarvude kohta alates aastast 1969. Samuti üle kümne aasta toimuva rahvaloenduse andmeid iga kümnendi kohta vahemikus 1910 kuni 2010. Andmeid täiendatakse majandusandmetega nagu tööhõive ning tegelik sissetulek inimese kohta, mida saadakse Ameerika Ühendriikide Majandusanalüüsi Büroost. (Riiman *et al.*, 2019)

Mitme maakonna mudel on üks kõigi maakondade andmete pealt treenitud mudel, mida rakendati seejärel kõigile maakondadele. Ühe maakonna mudel on aga iga maakonna kohta antud maakonna andmetel eraldi treenitud mudelite kogu. (Riiman *et al.*, 2019)

Uuringus hinnati 2010 aasta Alabama maakondade aastakeskmist või üle 10 aasta tehtava rahvaloenduse rahvaarvu (Riiman *et al.*, 2019).

Tabel 2: Mudelite vigade hinnangute võrdlus. (Riiman *et al.*, 2019)

Meetod	RMSE	MAE	MAPE
Osismetod	5251	3216	6,5%
ANN-LSTM, mitme maakonna mudel			
aastakeskmise rahvaarv	17523	11004	16,7%
rahvaloenduse rahvaarv	5442	3346	6,3%
ANN-LSTM, ühe maakonna mudel			
aastakeskmise rahvaarv	7160	3663	6,3%
rahvaloenduse rahvaarv	4529	2742	5,0%

Tulemusi on hinnatud ruutkeskmise vea (4), keskmise absoluutvea (5) ja keskmise absoluutse protsentveaga (6). Tabelist on näha, et osismetod on üldiselt täpsuse poolest mitme maakonna andmetel treenitud ANN-LSTM mudelist üle, kuid

ühe maakonna andmetel treenitud mudelite tulemused on võrreldava või parema täpsusega.

Kahes uuringus on prognoositud erinevate riikide rahvastikku väga erinevate mudelite ja andmestikega. Seetõttu on nende tulemuste võrdlemine suhteliselt keeruline. Seega, kuigi masinõppemudelite MAPE on oluliselt väiksem närvivõrkude omast, pole neid nende andmete põhjal siiski võimalik eristada.

1.5 Metoodikate hindamine

Rakendatud rahvastikuproгноosi meetodite võrdlemiseks kasutati ruutkeskmist viga (RMSE), keskmist absoluutviga (MAE) ja keskmist absoluutset protsentviga (MAPE). Vigade arvutamiseks kasutati R-i paketti Metrics (“GitHub - mfrasco/-Metrics: An R package for common supervised machine learning metrics.” [2024](#)).

- Keskmise ruutviga määratletakse kui

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2,$$

kus x_i on tegelik väärtus, y_i on hinnatud väärtus ja n on vaatluste arv.

- Ruutkeskmise viga on keskmise ruutvea ruutjuur,

$$\text{RMSE} = \sqrt{\text{MSE}}, \tag{4}$$

mis annab vea samades ühikutes, milles on algandmed.

- Keskmise absoluutviga defineeritakse kui keskmise absoluutne erinevus tegelike ja hinnatud väärtuste vahel,

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|. \tag{5}$$

- Keskmise absoluutse protsentviga väljendab prognooside ja tegelike väärtuste vahelisi erinevusi protsendina,

$$\text{MAPE} = 100 \cdot \frac{1}{n} \sum_{i=1}^n \frac{|x_i - y_i|}{x_i}. \quad (6)$$

2 Andmestikud

Eesti Statistikaamet

Töös kasutati Eesti Statistikaameti andmetabeleid ja statistikatöid aastatest 2016 kuni 2018 ning ka 2019. aasta 1. jaanuari rahvaarvu:

- RV106: ELUSSÜNDINUD SOO JÄRGI
- RV021: RAHVASTIK SOO JA VANUSERÜHMA JÄRGI, 1. JAANUAR
- RV0212: RAHVASTIK AASTA ALGUSES JA AASTAKESKMINE RAHVAARV SOO JA VANUSE JÄRGI
- RV088: RAHVASTIKUPROGNOOS 2020-2080: RAHVASTIKU NÄITAJAD MAAKONNA JA SOO JÄRGI
- Statistikaameti rände andmebaas
- Statistikaameti sündide andmebaas

Eesti Statistikaameti andmeid, peale rahvastikuproгноosi, kasutati paketi demCore osismeetodi rakendamisel vastavalt peatükis [1.1.1](#) kirjeldatud metoodikale.

ÜRO 2022

R-i paketi BayesPop tõenäosusliku rahvastikuproгноosi sisendina kasutati ÜRO wpp2022 (*World Population Prospects 2022*) andmestikku (United Nations, [2022](#)). Samad andmed on olemas ka Statistikaameti andmebaasis, kuid R-i pakett wpp2022 on BayesPopi sisse ehitatud, mistõttu on sealseid andmeid mugav kasutada.

Kasutati järgmiseid tunnuseid (Ševčíková & Raftery, [2016](#)):

- rahvaarv soo ja vanuse järgi,
- ajaloolised suremuse vanusekordajate hinnangud soo järgi,

- ajaloolised sündimuse vanusekordaja hinnangud soo järgi,
- tuleviku sugude osakaalu sünnil hinnangud,
- tuleviku rändesaldo hinnangud.

HMD ja HFD

HMD (*Human Mortality Database*) on maailma riikide suremuskordajate, elutabelite ja rahvaarvude andmebaas ning HFD (*Human Fertility Database*) on sündimuse ja fertiilsuse andmebaas. Mõlemad andmebaasid kasutavad Eesti algandmetena Eesti Statistikaameti väljastatud andmeid (“Human Mortality Database”, [2024](#); “Human Fertility Database”, [2024](#)). HMD ja HFD andmeid kasutatakse R-i paketi Demography sisenditena, kuna need on sobival kujul paketispetsiifilise demogdata objekti loomiseks.

HMD ja HFD andmestikest kasutati järgmisi tunnuseid (Rob J Hyndman, [2014](#)):

- 2019. aasta 1. jaanuari rahvaarv,
- ajaloolised suremuse vanusekordajad,
- ajaloolised sündimuse vanusekordajad.

3 Meetodite võrdlus

Metoodika peatükis kirjeldatud rahvastikuproгноosi meetoditest on töös rakendatud osismeetodi ja tõenäosuslikke meetodeid. Rakendamiseks kasutati vastavalt demCorei, BayesPopi ja Demography R-i pakette.

Rahvastikku on prognoositud aastast 2019 aastani 2080 nagu ka Statistikaameti 2019. aasta prognoosis. Prognoose võrreldakse 2020. kuni 2022. aasta tõelise rahvaarvuga ning omavahel aastani 2080. Tõeliseks rahvaarvuks on võetud Statistikaameti poolt välja antud 2020. kuni 2022. aasta väärtused. Võrdlemiseks on kasutatud metoodikate hindamise peatükkis välja toodud vea hindamise mõõte.

3.1 Osismeetod

Osismeetod on enim kasutatav rahvastikuproгноosi meetod (Preston, Heuveline & Guillot, 2001). Osismeetodiga loodud rahvastikuproгноosidele on seetõttu lihtsam teiste riikide kohta tehtud või teiste asutuste poolt läbi viidud prognoose võrdluseks tuua, kuna on palju teisi sarnaste või võrreldavate metoodikatega prognoose. On suurem kindlus, et prognooside erinevused ei tulene metoodika erinevustest. Osismeetodi progноosi kvaliteet sõltub paljuski sündimuse, suremuse ja rände eeldest.

Osismeetod on deterministlik rahvastikuproгноosi meetod. See võib väikeseid populatsioone progноosides osutada probleemiks, kuna reaalsus on juhuslik ehk stohastiline (Raftery & Ševčíková, 2023).

3.2 Tõenäosuslikud meetodid

Tõenäosuslikud progноosid võimaldavad täpsuse hindamist standardviagde ja usaldusintervallidega, mis on ka paljude teiste valdkondade progноoside puhul oodatud. See aitab kasutajatel mõista progноosi usaldusväärsust ja vähendada otsuste tege-

misel ebakindlust. Tõenäosuslik lähenemine annab ka võimaluse hinnata, kas ajas toimuvad muutused on juhuslikud või olulised. Samuti saab tõenäosuslike meetodite abil hinnata riskistsenaariume ja nende realiseerumise tõenäosusi. Ka ÜRO 2022 aasta rahvastikuproгноos oli tõenäosuslik. (Raftery & Ševčíková, 2023)

Tõenäosuslikud meetodid on arvutuslikult keerukamad näiteks osismetodist (Raftery & Ševčíková, 2023). See võib raskendada mudeli kasutamist kui mudeli parameetrite arv on suur.

3.3 Mikrosimulatsiooni meetodid

Mikrosimulatsioonid kasutavad individuaalset lähenemist, kus analüüsi üksuseks on indiviid. See võimaldab modelleerida keerukaid elukäigu üleminekuid ja demograafilisi protsesse, mida makro tasandi mudelid tehtud lihtsustuste tõttu ei pruugi suuta teha. Mikrosimulatsiooni mudelid on ka mitmekülgsed, võimaldades prognoosi kaasta suurel hulgal erinevaid tunnuseid, millel võib olla palju erinevaid väärtusi. See suurendab mudeli võimalike kasutusviiside hulka. (Bosco *et al.*, 2023)

Mikrosimulatsioonid on arvutuslikult keerukamad nii osismetodist kui ka tõenäosuslikest meetoditest, kuid see ei ole tänu arvutusvõimeksue kasvule enam suur piirang (Bosco *et al.*, 2023). Mikrosimulatsioon nõuab palju ja kavaliteetseid andmeid, mis on oluliselt spetsiifilisemad, kui teiste meetodite puhul ning ei pruugi olla saadaval. Andmete puudus võib simulatsiooni kalibreerimise võimatuks muuta. (Zagheni, 2015)

Antud töös pole mikrosimulatsiooni meetodeid rakendatud. Seda just sisendandmete eeltöötamise keerukuse ja mahukuse tõttu. Autori arvates oleks Eesti kontekstis siiski kasulik mikrosimulatsioonide rakendamist proovida. Põhjenduseks oleks erinevate riiklike registrite paljususe ja detailsus. Eriti huvitav oleks mikrosimulatsiooni algpopulatsioonina kasutada kõiki rahvastikuregistris olevaid Eestis elavaid isikuid. Oleks võimalik simuleerida kogu Eesti rahvastikku.

3.4 Masinõppe ja närvivõrkude rakendamine rahvastikuproгноosides

Närvivõrgud suudavad kohaneda populatsioonide isepärasustega neid prognoositava populatsiooni andmetel treenides. Kuid masinõppe ja närvivõrkude mudelid on sageli keerukad ja nõuavad olulist arvutusressurssi, mis võib olla takistuseks nende rakendamisel piiratud ressurssidega keskkondades. Närvivõrkudega seotud meetodite puhul on oluline arvestada nende potentsiaaliga. Kuid hetkel ei ole need veel rahvastikku prognoosides oluliselt paremad kui traditsioonilised võtted. (Riiman *et al.*, 2019)

Masinõppe mudelid vajavad efektiivseks õppimiseks suuri andmekogumeid. Kui andmed on piiratud või pole piisavalt esinduslikud, võib see viia ebatäpsete prognoosideni (Şahinarslan, Tekin & Çebi, 2021).

3.5 Tulemused aastatel 2020 kuni 2022

Nii Eesti Statistikaameti stsenaariumid kui ka antud töös rakendatud meetodite prognoosid on väga sarnased 2020. kuni 2022. aasta päris rahvaarvudega. Võrreldav periood on lühike ja kohe pärast sisendina kasutatud ajavahemiku lõppu. Tabelis (3) on toodud kõigi töös kasutatud prognooside tulemused aastatel 2020 kuni 2022.

Tabel 3: Rahvastikuproгноosid aastateks 2020 kuni 2022.

Mudel/Aasta	2020	2021	2022
Päris	1 328 889	1 330 068	1 331 796
SA1	1 324 025	1 323 391	1 322 814
SA2	1 325 545	1 326 498	1 327 574
SA3	1 322 526	1 319 753	1 316 488
SA4	1 324 025	1 323 552	1 323 292
demCore	1 327 416	1 329 452	1 330 926
demography	1 333 458	1 335 612	1 336 869
bayesPop	1 330 742	1 332 124	1 329 436

Tabelis (3) tähistavad SA1-SA4 Eesti Statistikaameti 2019. aasta rahvastikuprognosi stsenaariume. Stsenaariumid on vastavalt samad nagu Eesti rahvastikuprognosi metoodika kirjelduses.

Aastatele 2020 kuni 2022 prognoosis kõige täpsemalt paketi demCore osismeetod. Seda üle kõigi veahinnagute (4). demCore'i sisendiks kasutati prognoosiperioodile eelneva kolme aasta keskmisi. Neid keskmisi kasutati siis konstantselt kuni aastani 2080. Mudeli edu võib tuleneda sellest, et kolm aastat on liiga lühike periood suurteks muutusteks rahvastikusündmuste üldistes trendides. Kiiresti muutuv on vaid ränne. Eesti näitel toimus hüppeline tõus sisserändes tänu Ukraina sõjale. See muutus jääb aga võrdluseks valitud perioodist välja.

Paremuselt teiseks osutus sellel perioodil paketi bayesPop tõenäosuslik meetod, mis on teine parim samuti üle kõigi vea mõõtude. Kolmas parim on Eesti Statistikaameti kõrgema sündimuse ja rändesaldo stsenaarium.

Kõige halvemaks prognoosiks perioodile 2020-2022 osutus Statistikaameti madalama sündimuse ja suremuse ning tasakaalus rändesaldoga stsenaarium. Nii sündimus kui suremus on perioodil langenud, kuid rändesalod on pidevalt positiivne olnud. Seega selle stsenaariumie eeldused on täitmata, millest tuleneb ka halb tulemus.

Tabel 4: Prognoositud tulemuste vigade hinnangud.

Mudel	RMSE	MAE	MAPE
SA1	7045	6841	0,5142%
SA2	3731	3712	0,2790%
SA3	11273	10662	0,8013%
SA4	6793	6628	0,4982%
demCore	1050	986	0,0742%
Demography	5077	5062	0,3805%
bayesPop	2100	2090	0,1571%

Lisas 3.5 on toodud ka Eesti Statistikaameti ja töös rakendatud meetodite kogu rahvastiku prognooside joonised (3, 4) aastani 2080. Joonisel (5) on toodud ka tõenäosuslike meetodite prognooside mediaanid koos 95% usaldusintervallidega.

On näha, et paketi bayesPop prognoosi usaldusintervall on oluliselt kitsam kui demogrphy prognoosi oma.

Kokkuvõte

Bakalaureusetöö eesmärk on uurida erinevaid traditsioonilisi ja uuenduslikke rahvastikuproгноosi metoodikaid ning rakendada neid Eesti rahvastikul.

Uuriti osismeetodit, tõenäosuslikke meetodeid, mikrosimulatsiooni meetodeid ja masinõppele ning närvivõrkudele tuginevaid rahvastikuproгноosi meetodeid.

Osismeetod on levinuim rahvastikuproгноosi meetod ja seda rakendavad kõik pikaajaliste rahvastikuproгноose loovad asutused. Osismeetodi tuumaks on demograafiline bilansivõrrand, mis arvutab rahvaarvu sündide, surmade ja rändesaldo alusel. Töös uuriti osismeetodi rakendamist Eesti Statistikaameti poolt ja R-i paketiga demCore.

Tõenäosuslikud meetodid võimaldavad genereerida tõenäosuslikke hinnanguid rahvaarvule ja rahvastiku koosseisule. Tõenäosuslike rahvastikuproгноoside koostamise meetodid hõlmavad retrospektiivset analüüsi, aegridade analüüsi ja eksperdi-põhiseid lähenemisi. Pakett bayesPop võimaldab luua proгноose sama metoodikaga, mida kasutati ÜRO 2022. aasta rahvastikuproгноosis.

Mikrosimulatsiooni meetoditega modelleeritakse indiviidide elukulgu, eeldades, et nende indiviididega juhtuvad sündmused toimuvad määratud tõenäosuslike reeglite tagajärjel. Töös uuriti proгноosiprogrammi MicMac ja R-i paketti MicSim, mis võimaldab rahvastikuproгноosi mikrosimulatsioone teha.

Masinõppe ja närvivõrkude kasutamine rahvastikuproгноosides ei ole proгноose koostavate organisatsioonide seas levinud. Hetkel ei ole need veel rahvastikku proгноosides oluliselt paremad kui traditsioonilised võtted.

Rakendatud meetoditest kõige täpsemaks osutus paketi demCore osismeetod.

Kasutatud allikad

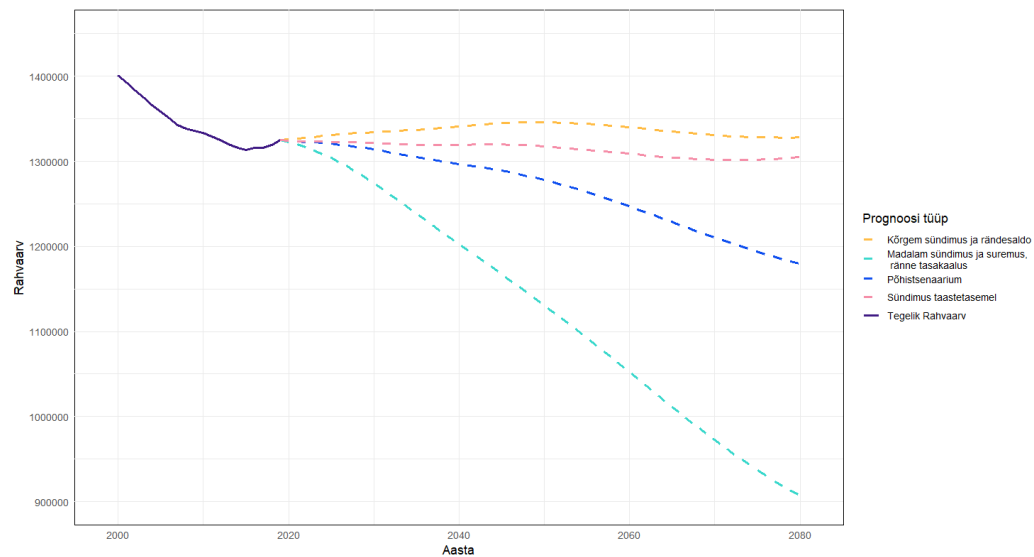
- Bosco, Claudio, Sabine Zinn, Daniela Ghio, Maurizio Teobaldelli ja Stefano Maria Iacus (1. juuni 2023). *MicSim, a microsimulation model for population dynamics*. DOI: [10.21203/rs.3.rs-2986170/v1](https://doi.org/10.21203/rs.3.rs-2986170/v1). URL: <https://www.researchsquare.com/article/rs-2986170/v1> (vaadatud 14.05.2024).
- CRAN - *Package assertive.base* (2024). URL: <https://cran.r-project.org/web/packages/assertive.base/index.html> (vaadatud 25.04.2024).
- GitHub - *mfrasco/Metrics: An R package for common supervised machine learning metrics*. (2024). URL: <https://github.com/mfrasco/Metrics?tab=readme-ov-file> (vaadatud 10.05.2024).
- Human Fertility Database (2024). URL: <https://www.humanfertility.org/> (vaadatud 14.05.2024).
- Human Mortality Database (2024). URL: <https://www.mortality.org/> (vaadatud 14.05.2024).
- Hyndman, Rob J. ja Heather Booth (juuli 2008). “Stochastic population forecasts using functional data models for mortality, fertility and migration”. *International Journal of Forecasting* 24.3, lk. 323–342. ISSN: 01692070. DOI: [10.1016/j.ijforecast.2008.02.009](https://doi.org/10.1016/j.ijforecast.2008.02.009). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0169207008000319> (vaadatud 13.05.2024).
- Katie Paulson, Grant Nguyen, Charlton Callender, Spencer Pease ja Haidong Wang (2024). *demCore: Core Functions for Demography*. URL: <https://ihmeuw-demographics.github.io/demCore/articles/ccmpp.html> (vaadatud 02.04.2024).
- Maasing, Ethel ja Ene-Margit Tiit (2021). *Rahvastikustatistika valemid*. URL: https://www.stat.ee/sites/default/files/2021-09/Rahvastikustatistika_valemid.pdf.

- Marois, Guillaume, Michaela Potančoková ja Miguel González-Leonardo (2023). “Technical report: QuantMig-Mic microsimulation population projection model”. Publisher: International Institute for Applied Systems Analysis. URL: https://pure.iiasa.ac.at/id/eprint/18973/1/QuantMig_IIASA_Deliverable%20D8.2_forsubmission_21-07-2023_corr.pdf (vaadatud 14.05.2024).
- O'Neill, Brian C., Deborah Balk, Melanie Brickman ja Markos Ezra (13. juuni 2001). “A Guide to Global Population Projections”. *Demographic Research* 4, lk. 203–288. ISSN: 1435-9871. URL: <https://www.demographic-research.org/volumes/vol4/8/4-8.pdf> (vaadatud 12.04.2024).
- Preston, Samuel H., Patrick Heuveline ja Michel Guillot (2001). *Demography: Measuring and Modeling Population Processes*. 9. [pr.] Oxford: Blackwell. 291 lk. kokku. ISBN: 978-1-55786-214-3.
- Raftery, Adrian E., Jennifer L. Chunn, Patrick Gerland ja Hana Ševčíková (1. juuni 2013). “Bayesian Probabilistic Projections of Life Expectancy for All Countries”. *Demography* 50.3, lk. 777–801. ISSN: 0070-3370, 1533-7790. DOI: [10.1007/s13524-012-0193-x](https://doi.org/10.1007/s13524-012-0193-x). URL: <https://read.dukeupress.edu/demography/article/50/3/777/169587/Bayesian-Probabilistic-Projections-of-Life> (vaadatud 13.05.2024).
- Raftery, Adrian E. ja Hana Ševčíková (1. jaanuar 2023). “Probabilistic population forecasting: Short to very long-term”. *International Journal of Forecasting* 39.1, lk. 73–97. ISSN: 0169-2070. DOI: [10.1016/j.ijforecast.2021.09.001](https://doi.org/10.1016/j.ijforecast.2021.09.001). URL: <https://www.sciencedirect.com/science/article/pii/S0169207021001394> (vaadatud 11.05.2024).
- Rahvastikuproгноos / Statistikaamet* (2024). URL: <https://www.stat.ee/et/avasta-statistikat/valdkonnad/rahvastik/rahvastikuproгноos> (vaadatud 09.05.2024).

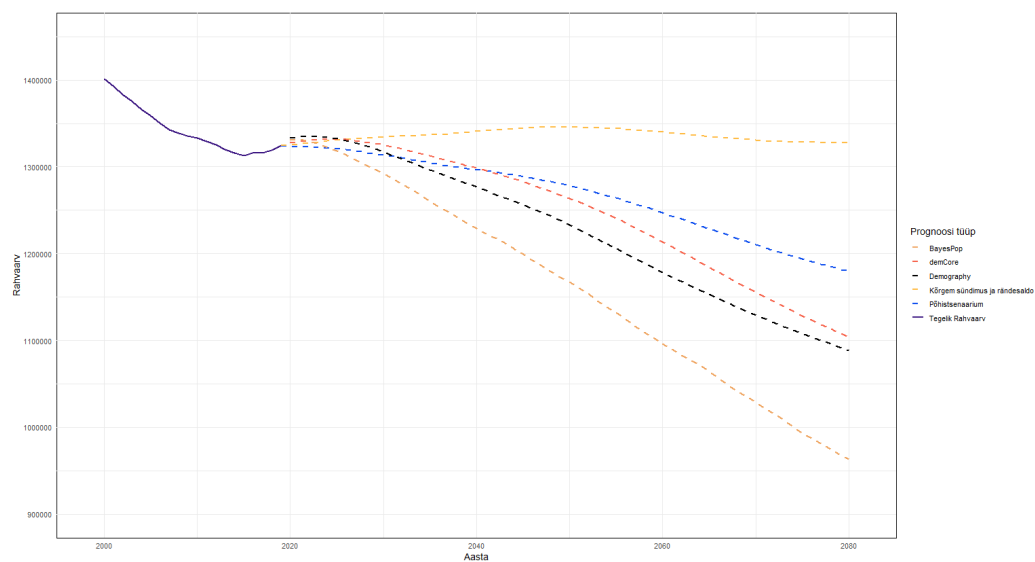
- Riiman, Viktoria, Amalee Wilson, Reed Milewicz ja Peter Pirkelbauer (2019). “Comparing artificial neural network and cohort-component models for population forecasts”. *Population review* 58.2. Publisher: Sociological Demography Press. URL: <https://muse.jhu.edu/pub/251/article/736444/summary> (vaadatud 14.05.2024).
- Rob J Hyndman (24. juuli 2014). *Rob J Hyndman - Coherent population forecasting using R*. Rob John Hyndman. URL: <https://robjhyndman.com/hyndsight/coherent-population-forecasting/index.html> (vaadatud 13.05.2024).
- Şahinarslan, Fatih Veli, Ahmet Tezcan Tekin ja Ferhan Çebi (2021). “Application of machine learning algorithms for population forecasting”. *International Journal of Data Science* 6.4, lk. 257. ISSN: 2053-0811, 2053-082X. DOI: [10.1504/IJDS.2021.122770](https://doi.org/10.1504/IJDS.2021.122770). URL: <http://www.inderscience.com/link.php?id=122770> (vaadatud 14.05.2024).
- Sevcikova, H, N Li, V Kantorova, P Gerland ja A E Raftery (2023). “bayesPop: Probabilistic Population Projection”.
- Statistikaamet, Eesti (2019). *2019. AASTA RAHVASTIKUPROGNOOS AASTANI 2080. METOODIKA*. URL: <https://www.stat.ee/sites/default/files/2024-02/Rahvastikuproгноос.%20Metoodika%20%282019%29.pdf>.
- Ševčíková, Hana ja Adrian E. Raftery (6. detsember 2016). “bayesPop: Probabilistic Population Projections”. *Journal of Statistical Software* 75, lk. 1–29. ISSN: 1548-7660. DOI: [10.18637/jss.v075.i05](https://doi.org/10.18637/jss.v075.i05). URL: <https://doi.org/10.18637/jss.v075.i05> (vaadatud 04.05.2024).
- Zagheni, Emilio (2015). “Microsimulation in demographic research”. *International encyclopedia of social and behavioral sciences* 15, lk. 343–346. URL:

- http://www.zagheni.net/uploads/3/1/7/9/3179747/microsimulation_in_demographic_research.pdf (vaadatud 14.05.2024).
- Zinn, Sabine (1. detsember 2014). “The MicSim Package of R: An Entry-Level Toolkit for Continuous-Time Microsimulation”. *International Journal of Microsimulation* 7, lk. 3–32. DOI: [10.34196/ijm.00105](https://doi.org/10.34196/ijm.00105).
- Tensing, Anu (2012). *Proгноосiprogrammi MicMac juhend*.
- United Nations (23. august 2022). *World Population Prospects 2022: Summary of Results*. Statistical Papers - United Nations (Ser. A), Population and Vital Statistics Report. United Nations. ISBN: 978-92-1-001438-0. DOI: [10.18356/9789210014380](https://doi.org/10.18356/9789210014380). URL: <https://www.un-ilibrary.org/content/books/9789210014380> (vaadatud 13.05.2024).
- Vähi, Mare, Tiit, Ene-Margit, Leetmaa, Kadri, Väiko, Annika, Krusell, Siim, Tammur, Alis ja Tõnurist, Anu (2015). *Rahvastiku areng: = Population trends*. Toim. Taimi Rosenberg. Tallinn: Eesti Statistika. 180 lk. kokku. ISBN: 978-9985-74-578-6.

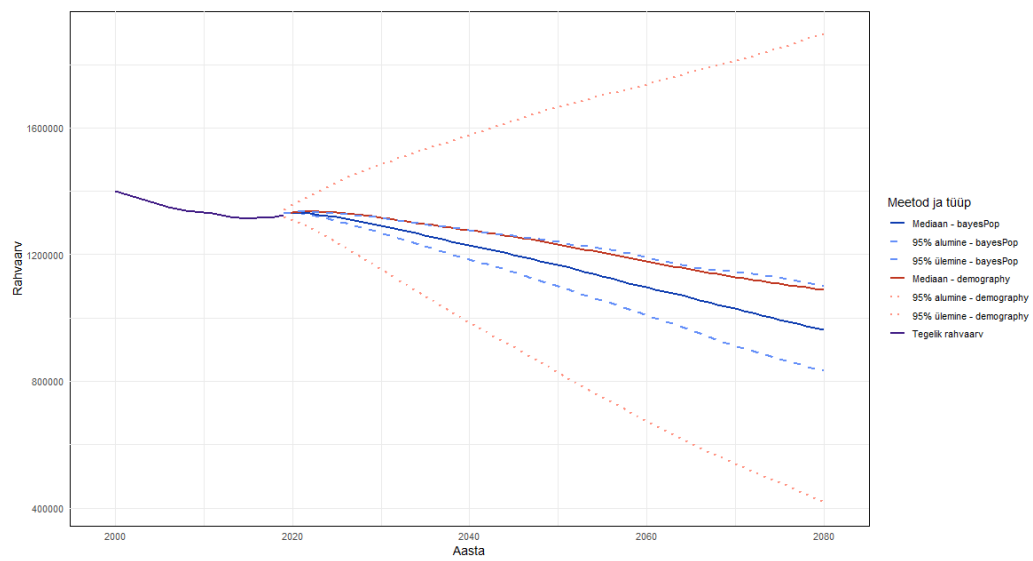
Lisa 1. Prognoos paketiga demCore



Joonis 3: Eesti Statistikaameti 2019. aasta rahvastikuprognosi stsenaariumid aastani 2080.



Joonis 4: Rakendatud meetoditega loodud rahvastikuprognosisid ning Statistikaameti põhi- ja nii kõrgema sündimuse kui rändesaldoga stsenaarium aastani 2080.



Joonis 5: bayesPop ja demography rahvastikuproгноoside mediaanid koos 95% usaldusintervallidega aastani 2080.

Lisa 2. Prognosis paketi demCore

```
library(demCore)
library(dplyr)
library(tidyr)
library(readxl)
library(data.table)

options(scipen = 999, encoding = 'utf-8')

setwd(dirname(rstudioapi::getActiveDocumentContext())$path
))

# Sugude oskaal sündide hulgas

synnid_sugu <- read.csv("synnid_soo_jargi.csv", sep = "\t",
  na.strings = "..") %>%
  filter(2015 < Aasta, Aasta < 2019) %>% summarise(value
    = mean(Kõik.rahvused.Poisid / Kõik.rahvused.Tüdrukud
  ))
srb <- data.table(year_start = seq(2019, 2080, 1),
  year_end = seq(2020, 2081, 1),
  value = rep(synnid_sugu$value, 5))

# Sündimuse vanusekordaja
```

```

synd_vanuskordaja <- read.csv("synd_vanuskordaja.csv")
  %>% select(VANUS_EMA_H, Vanuskordaja18)
colnames(synd_vanuskordaja) <- c("age_start", "value")
synd_vanuskordaja <- synd_vanuskordaja %>% mutate(age_end
  = age_start + 1)

year_start <- rep(seq(2019, 2080, 1), each = nrow(synd_
  vanuskordaja))
year_end <- rep(seq(2020, 2081, 1), each = nrow(synd_
  vanuskordaja))
age_start <- rep(synd_vanuskordaja$age_start, length(seq
  (2019, 2080, 1)))
age_end <- rep(synd_vanuskordaja$age_end, length(seq
  (2019, 2080, 1)))
value <- rep(synd_vanuskordaja$value, length(seq(2019,
  2080, 1)))

asfr <- data.frame(year_start, year_end, age_start, age_
  end, value) %>% arrange(year_start, year_end, age_
  start, age_end) %>% data.table()

# asfr %>% group_by(year_start) %>% summarise(sum = sum(
  value))
# näide[["srb"]] %>% group_by(year_start) %>% summarise(
  sum = sum(value*5))

# Rahvaarv 2019. aasta 1. jaanuari seisuga

```

```

rahvaarv_2019 <- read.csv("rahvaarv_aasta_alguses.csv",
  sep = "\t", na.strings = "..")
colnames(rahvaarv_2019) <- c("aasta", "vanus", "mehed", "
  naised")
rahvaarv_2019[rahvaarv_2019$vanus == '100□ja□vanemad', 2]
  <- 100
rahvaarv_2019 <- rahvaarv_2019 %>% filter(vanus != '85□ja
  □vanemad')
rahvaarv_2019$vanus <- as.numeric(rahvaarv_2019$vanus)
rahvaarv_2019 <- rahvaarv_2019 %>% mutate(age_end =
  replace(vanus + 1, vanus == 100, Inf)) %>%
  pivot_longer(cols = c(mehed, naised), names_to = "sugu"
    , values_to = "arv")
baseline <- rahvaarv_2019 %>% select(aasta, sugu, vanus,
  age_end, arv) %>% mutate(sugu = ifelse(sugu == 'naised
  ', 'female', 'male')) %>%
  arrange(sugu, vanus)
colnames(baseline) <- c("year", "sex", "age_start", "age_
  end", "value")
baseline <- data.table(baseline)

```

Vanuse suremuskordaja

```

rahvaarv <- read.csv("aastakeskmine_rahvaarv_soo_jargi.
  csv", sep = "\t", na.strings = "..")
colnames(rahvaarv) <- c("aasta", "vanus", "mehed", "
  naised")

```

```

rahvaarv[rahvaarv$vanus == '100_aja_vanemad', 2] <- 100
rahvaarv <- rahvaarv %>% filter(vanus != '85_aja_vanemad')
rahvaarv$vanus <- as.numeric(rahvaarv$vanus)
kesk_rahvaarv <- rahvaarv %>% group_by(vanus) %>%
  summarise(male = mean(mehed), female = mean(naised))
kesk_rahvaarv <- kesk_rahvaarv %>% pivot_longer(cols = c(
  male, female), names_to = "sex", values_to = "kesk")
  %>%
  arrange(sex, vanus)

surmad <- read.csv("surmad_vanuse_järgi.csv", sep = "\t",
  na.strings = "..")
colnames(surmad) <- c("aasta", "piirkond", "vanus", "
  mehed", "naised")
surmad <- surmad %>% mutate(age_start = as.numeric(ifelse
  (vanus != '100_aja_vanemad', vanus, 100)),
  age_end = age_start + 1)
kesk_surmad <- surmad %>% group_by(age_start) %>%
  summarise(male = mean(mehed), female = mean(naised))
kesk_surmad <- kesk_surmad %>% pivot_longer(cols = c(male
  , female), names_to = "sex", values_to = "kesk") %>%
  arrange(sex, age_start)

mx <- kesk_surmad$kesk / kesk_rahvaarv$kesk
ax <- mx_to_ax(mx = mx, age_length = 1)
ax <- ifelse(is.na(ax), 0.5, ax)
sex <- kesk_surmad$sex
age_start <- baseline$age_start
age_end <- baseline$age_end

```

```

dt <- data.table::data.table(sex, age_start, age_end, mx,
  ax)

lifetable(dt, c("sex", "age_start", "age_end"))
nSx <- nSx_from_lx_nLx_Tx(dt, c("sex", "age_start", "age_
  end"), terminal_age = 100) %>% arrange(sex, age_start)
colnames(nSx) <- c("sex", "age_start", "age_end", "value"
  )
survival <- data.frame(year_start = rep(seq(2019, 2080,
  1), each = nrow(nSx)),
  year_end = rep(seq(2020, 2081, 1),
    each = nrow(nSx)), nSx[rep(seq
    _len(nrow(nSx)), times = 5), ])
  %>%
  arrange(year_start, year_end, sex, age_start, age_end)
  %>% data.table()

# Rändesaldo

net_migration <- read.csv("net_migration.csv")
# net_migration <- net_migration %>% filter(2015 < year_
  start, year_start < 2019) %>% group_by(sex, age_start,
  age_end)

aastakeskmine <- read.csv("rahvaarv2018-23.csv", sep = "\
  t") %>% filter(Vanus == 'Kokku', Aasta == 2018) %>%
  select(Aastakeskmine.rahvaarv.Mehed.ja.naised)

```

```

net_migration <- net_migration %>% filter(2015 < year_
  start, year_start < 2019) %>% group_by(sex, age_start,
  age_end) %>%
  summarise(value = round(mean(value)))
net_migration$value <- net_migration$value / baseline$
  value
net_migration <- data.table(year_start = rep(seq(2019,
  2080, 1), each = nrow(net_migration)),
  year_end = rep(seq(2020,
    2081, 1), each = nrow(net_
    migration)),
  sex = rep(net_migration$sex,
    5),
  age_start = rep(net_migration
    $age_start, 5),
  age_end = rep(net_migration$
    age_end, 5),
  value = rep(net_migration$
    value, 5))

initial_estimates_constant <- list(srb = srb, asfr = asfr
  , baseline = baseline,
  survival = survival,
  net_migration = net
    _migration)

```

```
# Modeli seeded
```

```
settings <- list(  
  years = seq(2019, 2080, 1),  
  sexes = c("female", "male"),  
  ages = seq(0, 100, 1),  
  ages_mortality = seq(0, 100, 1),  
  ages_asfr = seq(15, 49, 1)  
)
```

```
# Ennustus
```

```
estonia_population_constant <- ccmpp(  
  inputs = initial_estimates_constant,  
  settings = settings  
)
```

```
# Salvestamine
```

```
write.csv(estonia_population_constant, "estonia_  
  population_constant.csv")
```

```
# per_year_estonia_pop_constant <- estonia_population_  
  constant %>% group_by(year) %>% summarise(pop = sum(  
    value))
```

Lisa 3. Prognosis paketi demography

```
library(HMDHFDplus)
library(demography)
library(dplyr)

options(scipen = 999, encoding = 'utf-8')

setwd(dirname(rstudioapi::getActiveDocumentContext())$path
))

# Funktsioon demogdata objektide loomiseks
demogdata_2 <- function(mx_or_fert, pop, label = "Estonia
  ", type) {
  mx <- mx_or_fert
  obj <- list(type = type, label = label, lambda = 0)
  obj$year <- sort(unique(mx[, "Year"]))
  n <- length(obj$year)
  m <- length(unique(mx[, "Age"]))
  obj$age <- mx[seq(m), "Age"]
  mnames <- names(mx)[-c(1:2, NCOL(mx))]
  n.mort <- length(mnames)
  obj$rate <- obj$pop <- list()
  for (i in seq(n.mort)) {
    obj$rate[[i]] <- matrix(mx[, i + 2], nrow = m, ncol =
      n)
    obj$rate[[i]][obj$rate[[i]] < 0] <- NA
    obj$pop[[i]] <- matrix(pop[, i + 2], nrow = m, ncol =
```

```

      n)
    obj$pop[[i]][obj$pop[[i]] < 0] <- NA
    dimnames(obj$rate[[i]]) <- dimnames(obj$pop[[i]]) <-
      list(obj$age, obj$year)
  }
  names(obj$pop) <- names(obj$rate) <- tolower(mnames)

  return(structure(obj, class = "demogdata"))
}

```

```

# Andmete laadmine HMDst ja HFDst

```

```

username <- "*****"
password_HMD <- "*****"
password_HFD <- "*****"

```

```

country = "EST"

```

```

mx <- readHMDweb(country, item = "Mx_1x1",
                  username = username,
                  password = password_HMD,
                  fixup = TRUE)

pop <- readHMDweb(country, item = "Exposures_1x1",
                  username = username,
                  password = password_HMD,

```

```

fixup = TRUE)

est.pop <- hmd.pop("EST", username, password_HMD, "
  Estonia")
# est.pop <- extract.ages(est.pop, age=0:100)
est.pop <- extract.ages(est.pop, age=0:100)

asfr <- readHFDweb(country, item = "asfrRR",
  username = username, password =
    password_HFD, fixup = TRUE)

# Eesti suremuse demogdata objekt

combined_data_mx <- merge(mx, pop, by = c("Year", "Age"))

mx_100 <- combined_data_mx %>% filter(Age > 99) %>% group
  _by(Year) %>% summarise(
  mx_Female = sum(Female.x * Female.y, na.rm = TRUE) /
    sum(Female.y, na.rm = TRUE),
  mx_Male = sum(Male.x * Male.y, na.rm = TRUE) / sum(Male
    .y, na.rm = TRUE),
  mx_Total = sum(Total.x * Total.y, na.rm = TRUE) / sum(
    Total.y, na.rm = TRUE),
  .groups = 'drop') %>% mutate(Age = 100)

mx2 <- mx %>% filter(Age <= 100) %>% merge(mx_100, by = c
  ("Year", "Age"), all.x = TRUE) %>%
  mutate(Female = ifelse(Age == 100, mx_Female, Female),

```

```

      Male = ifelse(Age == 100, mx_Male, Male),
      Total = ifelse(Age == 100, mx_Total, Total)) %>%
        select(-c(mx_Female, mx_Male, mx_Total))

pop2 <- pop %>% filter(Age <= 100)

mx3 <- mx %>% filter(Age <= 100)

est.mort = demogdata_2(mx3, pop2, label = "Estonia", type
  = "mortality")
est.mort = extract.years(est.mort, years = 1959:2018)

# Eesti sündimuse demogdata objekt

asfr2 <- asfr %>% filter(15 <= Age, Age <= 49) %>% mutate
  (female = ASFR*1000) %>% select(-ASFR) %>% select(Year
    , Age, female, OpenInterval)

pop3 <- pop %>% filter(15 <= Age, Age <= 49)

est.fert = demogdata_2(asfr2, pop3, label = "Estonia",
  type = "fertility")
est.fert = extract.years(est.fert, years = 1959:2018)

# Eesti rände demogdata objekt

```

```
est.mig <- netmigration(est.mort, est.fert, mfratio =  
  1.079309)
```

```
# Silumine
```

```
estmort.sm <- smooth.demogdata(est.mort, obs.var="  
  theoretical")  
estfert.sm <- smooth.demogdata(est.fert, obs.var="  
  theoretical")  
estmig.sm <- smooth.demogdata(est.mig)
```

```
# Suremus
```

```
mort.fit <- coherentfdm(estmort.sm)
```

```
# Sündimus
```

```
fert.fit <- fdm(estfert.sm)
```

```
# Ränne
```

```
mig.fit <- coherentfdm(estmig.sm)
```

```
# Prognosis
```

```
# Suremus
```

```

mortf <- forecast(mort.fit, h=62, max.d=1)

# Sündimus
fertf <- forecast(fert.fit, h=62, max.d=1)

# Ränne
migf <- forecast(mig.fit, h=62, stationary=TRUE)

# Simulatsioon

est.sim <- pop.sim(mortf, fertf, migf, est.pop, N = 1000,
  mfratio = 1.079309)

# Salvestamine

popm.mean <- apply(est.sim$male, c(1,2), mean, na.rm = TRUE)
popf.mean <- apply(est.sim$female, c(1,2), mean, na.rm =
  TRUE)
colnames(popf.mean) <- colnames(popm.mean)
pop.mean <- popm.mean + popf.mean

write.csv(popm.mean, "demog_mees.csv")
write.csv(popf.mean, "demog_naine.csv")

```

```
write.csv(pop.mean, "demog_kokku.csv")
```

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Willem Klaassen,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Traditsioonilised ja uuenduslikud rahvastikuprognooosi metoodikad", mille juhendajad on Terje Trasberg ja Mare Vähi, 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 litsentsiga CC BY NC ND 4.0, 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.

Willem Klaassen

15.05.2024