

Rahvastikustatistikas kasutusel olevad matemaatilised valemid

Koostanud Ethel Maasing ja Ene-Margit Tiit

Käesolev dokument on rahvastikustatistika meetodika lisamaterjal. Siin on esitatud rahvastikustatistikas kasutatavate põhiliste kordajate ja näitajate arvutusvalemid ja selgitused.

Demograafiline tööturusurveindeks (DTSI) arvutatakse järgmiselt:

$$DTSI = \frac{\sum_5^{14} R}{\sum_{55}^{64} R},$$

kus $\sum_i^n R$ on inimeste arv vanuserühmas i -st n -ni (mõlemad otspunktid kaasa arvatud), i ja n on vanused täisaastates.

Demograafiline tööturusurve indeks, mis arvutatakse 5–14-aastaste ja 55–64-aastaste vanuserühmade arvukuse suhtena, näitab lähikümnendil tööturule saabujate arvu suhet tööturul vanuse tõttu lahkujate arvusse. Kui indeks on ühest suurem, siseneb järgmisel kümnendil tööturule rohkem inimesi, kui seal vanaduse tõttu potentsiaalselt välja langeb. Kui indeks on ühest väiksem, on järgmisel kümnendil väljalangejaid rohkem kui sisenejaid.

Ülalpeetavate määr (ÜM) arvutatakse järgmiselt:

$$\text{ÜM} = \frac{\sum_0^{14} R + \sum_{65}^{120} R}{\sum_{15}^{64} R} * 100,$$

kus $\sum_i^n R$ on inimeste arv vanuserühmas i -st n -ni (mõlemad otspunktid kaasa arvatud), kus i ja n on vanused täisaastates.

Ülalpeetavate määr iseloomustab ülalpeetavate (0–14-aastaste ja vähemalt 65-aastaste) põlvkondade summaarse arvukuse suhet tööealiste (15–64-aastaste) arvukusse (protsentides). Mida väiksem see suhe on, seda väiksem koormus on töötajatel.

Naise viljakas ehk fertiilne iga. Kokkuleppeliselt on naise viljakas iga vanuses 15–49 aastat, st arvestatakse viieaastaseid vanuserühmi [15; 20), ..., [45; 50). Kõiki sündimusnäitajaid arvutatakse viljakas eas naisi arvestades.

Vanuserühmad. Kokkuleppeliselt arvestatakse vanust täisaastates, st viimase sünnipäeva vanusena. See tähendab, et näiteks viieaastane vanuserühm 15–19 hõlmab 15–19-aastaseid.

Sündimuse vanuskordaja (age-specific fertility rate) ${}_a f_x$ arvutatakse vanuserühma $[x, x+a)$ jaoks järgmiselt:

$${}_a f_x = \frac{{}_a B_x}{{}_a \bar{P}_x},$$

kus a on arvutustel kasutatava vanusevahemiku pikkus (tavaliselt üks või viis aastat), ${}_a B_x$ on aasta jooksul naistel vanuses $[x; x+a)$ elusana sündinud laste arv ja ${}_a \bar{P}_x$ on aastakeskmise naiste arv vanuses $[x; x+a)$.

Vahel kasutatakse sündimuse vanuskordajana ka näitajat $1000 {}_a f_x$, mis näitab vaatlusaasta jooksul tuhande vanuses $[x; x+a)$ naise sünnitatud laste arvu.

Sündimuse vanuskordaja näitab sündide intensiivsust vanuserühmiti, st keskmist sünnituste arvu vaatlusaastal iga vanuserühma naise kohta.

Sündimuse erikordaja (*general fertility rate (GFR)*) on tuhandekordne sündimuse vanuskordaja, kus $x = 15$ ja $a = 35$. Näitab elussündide arvu vaatlusaastal tuhande viljakas eas (vanuses 15–49) naise kohta.

Ena keskmine vanus lapse sünnil (viieaastaste vanusevahemike korral) arvutatakse järgmiselt:

$${}_5\bar{x}_f = \frac{\sum(x + 2,5) * {}_5f_x}{\sum {}_5f_x},$$

kus x tähistab vahemike alguspunkte ja $x + 2,5$ keskpunkte ning ${}_5f_x$ on sündimuse vanuskordaja; $x = 15, 20, \dots, 45$ ja summeerimine toimub üle kõigi vanuserühmade (nii lugejas kui ka nimetajas).

Ena keskmine vanus esimese lapse sünnil (viieaastaste vanusevahemike korral) arvutatakse järgmiselt:

$${}_5\bar{x}_f = \frac{\sum(x + 2,5) * \frac{1}{5}f_x}{\sum \frac{1}{5}f_x},$$

kus x tähistab vahemike alguspunkte ja $x + 2,5$ keskpunkte ning $\frac{1}{5}f_x$ on vaadeldavasse vanuserühma kuuluvate aasta jooksul esimest korda sünnitanud naiste arvu ja sellesse vanuserühma kuuluvate naiste aastakeskmise arvu suhe.

Summaarne sündimuskordaja (*total fertility rate (TFR)*) arvutatakse järgmiselt:

$$TFR = a \sum {}_af_x,$$

kus a on vanusevahemiku pikkus (tavaliselt üks või viis aastat), ${}_af_x$ on sündimuse vanuskordaja ja summeerimine toimub üle kogu viljaka ea.

Summaarne sündimuskordaja näitab keskmist naise kogu elu jooksul sünnitatavate laste arvu eeldusel, et sündimuskäitumine püsib kogu naise viljaka ea jooksul vaatlusaasta tasemel.

Sündimuse üldkordaja (*crude birth rate (CBR)*) arvutatakse järgmiselt:

$$CBR = \frac{B}{\bar{R}} * 1000,$$

kus B on elussündide arv vaadeldaval ajavahemikul ning \bar{R} on aastakeskmise rahvaarv.

Sündimuse üldkordaja näitab aasta elussündide arvu 1000 elaniku kohta.

Taaste brutokordaja (*gross reproduction rate (GRR)*) arvutatakse järgmiselt:

$$GRR = a \sum \frac{{}_aG_x}{{}_a\bar{P}_x},$$

kus a on vanusevahemiku pikkus (tavaliselt üks või viis aastat), ${}_aG_x$ on naistel vanuses $[x; x + a)$ elusana sündinud tütarde arv ja ${}_a\bar{P}_x$ on aastakeskmise naiste arv vanuses $[x; x + a)$ ning summeerimine toimub üle kogu viljaka ea.

Taaste brutokordaja näitab keskmist naise kogu elu jooksul sünnitatud tütarde arvu eeldusel, et sündimuskäitumine püsib kogu naise viljaka ea jooksul vaatlusaasta tasemel.

Taaste netokordaja (*net reproduction rate (NRR)*) arvutatakse elutabeli andmete põhjal järgmiselt:

$$NRR = a \sum \left(\frac{{}_aG_x}{{}_a\bar{P}_x} * l_x \right) / 100000,$$

kus a on vanusevahemiku pikkus (tavaliselt üks või viis aastat), ${}_aG_x$ on naistel vanuses $[x; x + a)$ elusana sündinud tütarde arv, ${}_a\bar{P}_x$ on aastakeskmise naiste arv vanuses $[x; x + a)$ ja l_x ellujäänud naiste arv põlvkonnas vanuses x (vt elutabel). Summeerimine toimub üle kogu viljaka ea.

Arenenud ühiskondades erinevad taaste brutokordaja ja taaste netokordaja väga vähe, sest noorte naiste suremus on väike.

Taastetase on sündide arv, mis tagab loomuliku iibe mittenegatiivsuse. Sündimuse üldkordaja väärtus on 2,06 (varem kasutusel ka 2,1), taaste neto- või brutokordaja väärtus 1.

Suremuse üldkordaja (*crude death rate (CDR)*) arvutatakse järgmiselt:

$$CDR = \frac{D}{\bar{R}} * 1000,$$

kus D on aasta jooksul surnud isikute arv ning \bar{R} on aastakeskmise rahvaarv.

Suremuse üldkordaja näitab aasta surmade arvu 1000 aastakeskmise elaniku kohta.

Loomuliku iibe kordaja (*crude rate of natural increase (CRNI)*) arvutatakse sündimuse ja suremuse üldkordaja vahena:

$$CRNI = CBR - CDR.$$

Esmasabiellumise vanuskordaja naiste/meeste viieaastasest vanuserühmas arvutatakse järgmiselt:

$${}_5m_x = \frac{{}_5M_x}{{}_5\bar{P}_x},$$

kus ${}_5M_x$ on esmasabiellumise arv naistel/meestel vanuses $[x; x + 5)$ ja ${}_5\bar{P}_x$ on aastakeskmise naiste/meeste arv vanuses $[x; x + 5)$.

Arvutatakse vanuserühmade 15–19, 20–24, ..., 45–49 jaoks.

Kasutatakse ka kujul $1000 * {}_5m_x$ näitamaks esmasabiellumise sagedust 1000 vastavasse vanuserühma kuuluva mehe/naise kohta.

Keskmine vanus esmasabiellumisel naistel/meestel arvutatakse järgmiselt:

$${}_5\bar{x}_m = \frac{\sum (x + 2,5) * {}_5m_x}{\sum {}_5m_x},$$

kus ${}_5m_x$ on esmasabiellumise vanuskordaja viieaastasest vanuserühmas naistel/meestel (arvestatavad vanuserühmad on 15–49).

Aastail 1970–1977 on võetud ${}_5m_{40}$ ja ${}_5m_{45}$ väärtusteks tinglikult ${}_{10}m_{40}$.

Summaarne esmasabiellumuskordaja naistel/meestel (*total first marriage rate (TFMR)*) arvutatakse järgmiselt:

$$TFMR = 5 \sum {}_5m_x,$$

kus ${}_5m_x$ on esmasabiellumise vanuskordaja viieaastasest vanuserühmas naistel/meestel (arvestatavad vanuserühmad on 15–49). Summaarne esmasabiellumuskordaja näitab enne 50. sünnipäeva esimest korda abielluvate meeste/naiste osatähtsust rahvastikus eeldusel, et abiellumuskäitumine püsib vaatlusperioodi jooksul muutumatuna.

Oodatav eluiga ja elutabel

Oodatava eluea arvutamisel on abiks elutabel. **Elutabel** sisaldab iga-aastaseid andmeid rahvastiku suremuse kohta, kusjuures andmete võrreldavuse eesmärgil on tabel koostatud 100 000-isikulise mudelpopulatsiooni jaoks.

Tabeli aluseks on määratud vanuses t suremise tõenäosus q_t^x vaatlusaastal x , $t = 1, 2, \dots$

Selle näitaja hinnang arvutatakse vanuses t aastal x surnud inimeste arvu D_t^x ja selles vanuses inimeste koguarvu P_t^x jagatisena. Pildi teeb veidi segasemaks tõsiasi, et aastal x vanuses t surnute seas leidub niihästi neid, kes said t -aastaseks aastal x , kui ka neid, kelle viimane sünnipäev, kui nad t -aastaseks said, jääb hoopiski eelmisesse aastasse ehk aastasse $x-1$. Kui kogurahvaarvuna kasutatakse aastakeskmist rahvaarvu, siis ei sisalda see arv sel aastal pärast t -aastast sünnipäeva lahkunud ning nihke vältimiseks tuleb suhte arvutamisel rahvaarvu selle võrra parandada. Seega on arvutuskäik järgmine:

$$q_t^x = \frac{D_t^x}{P_t^x + D_t^{x-1}}$$

kus D_t^{x-1} tähistab nende isikute arvu, kes surid vaatlusaastal pärast oma t -ndat sünnipäeva.

Elutabelis lähtutakse eeldusest, et aastal x sünnib 100 000 inimest, st 0-aastaste põlvkonna suurus on $l_0 = 100\,000$. Esimese aasta lõpuks on nende arv surmade tõttu vähenenud $100\,000 * q_0$ võrra ja ellujääjaid on $l_1 = 100\,000 * (1 - q_0)$. Järgmise aasta **ellujäänute arvu** leidmiseks lahutatakse viimase aasta ellujäänute arvust aasta jooksul surnute arv ja nii väheneb ellujäänute põlvkond iga aastaga. Inimese maksimaalne võimalik eluiga tähistatakse tähega w , mis muutub vastavalt populatsioonile ja ajale. Traditsiooniliselt on selleks kasutatud arvu 100, ja kuigi praegu reaalne maksimaalne eluiga maailmas ületab 100 aasta piiri, põhjustab see siiski võrdlemisi väikest viga. Mõnel pool kasutatakse viimasel ajal ka väärtust $w = 120$.

Ellujäänute arv määrab kogu populatsiooni summaarse ees ootavate eluaastate arvu:

$$S_0 = \sum_{t=0}^w l_t.$$

Samamoodi on leitav kõigi x -aastaseks saanud isikute summaarne ees ootavate eluaastate arv:

$$S_x = \sum_{t=x}^w l_t.$$

Keskmise oodatavate eluaastate arvu saamiseks tuleb summaarne eluaastate arv jagada vastaval aastal elus olevate isikute arvuga:

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

Suurust e_0 nimetatakse oodatavaks elueaks sünnil, kuid see on siiski suurus, mis iseloomustab vaatlusaastat x – see näitab aastal x sündinud inimese oodatavat keskmist eluiga tingimusel, et kogu selle isiku eluea jooksul jäävad suremusnäitajad täpselt samaks, mis on aastal x . Tegelikuses see nii ei ole, sest enamasti suremus kahaneb ja oodatav eluiga pikeneb. Seega keskmiselt elab aastal x sündinud isik märgatavalt kauem, kui näitab arv e_0 .

Miks kasutatakse integraaliga valemit?

Kuna põhimõtteliselt muutub vanus pidevalt, mitte täisarvuliste aastatena, siis esitatakse oodatava elada jäänud aastate arvu valem sageli integraali kujul:

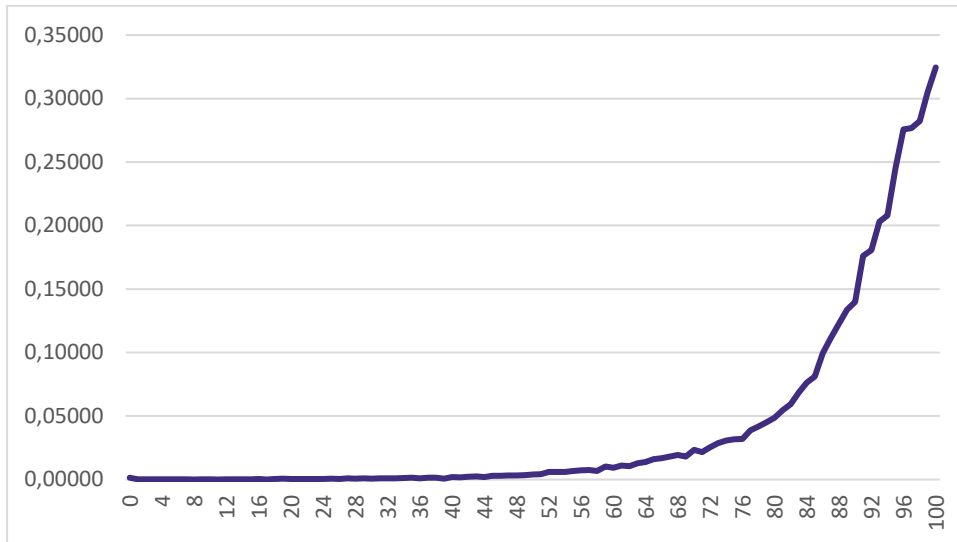
$$e_x^0 = \frac{1}{l_x} \int_x^w l_t dt,$$

kuid praktilises arvutustöös kasutatakse alati integraali asemel summat.

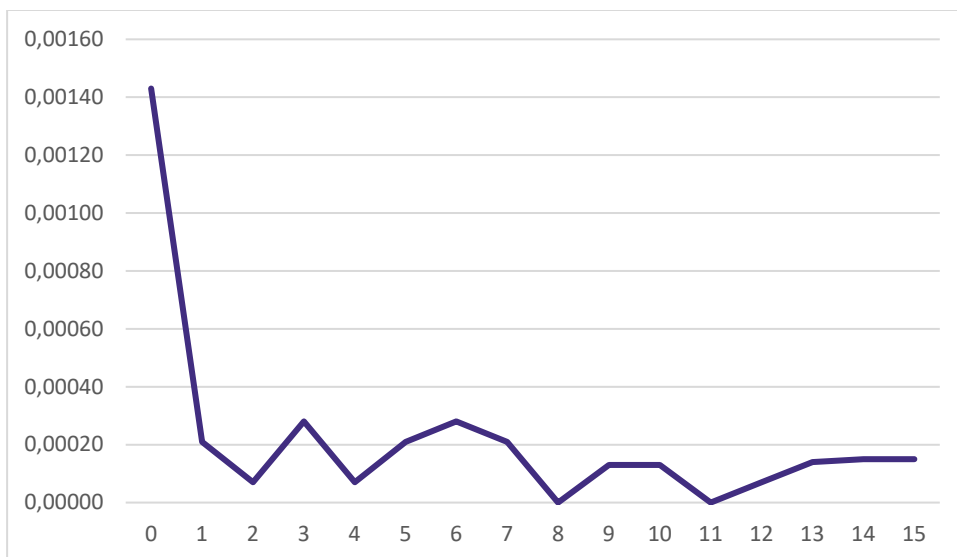
Silumine

Praktilises arvutustöös esineb teinegi mure, eriti kui on tegemist suhteliselt väikese rahvaarvuga. Selgub, et suremusnäitajad varieeruvad aastati võrdlemisi palju (vt alljärgnevat graafikuid).

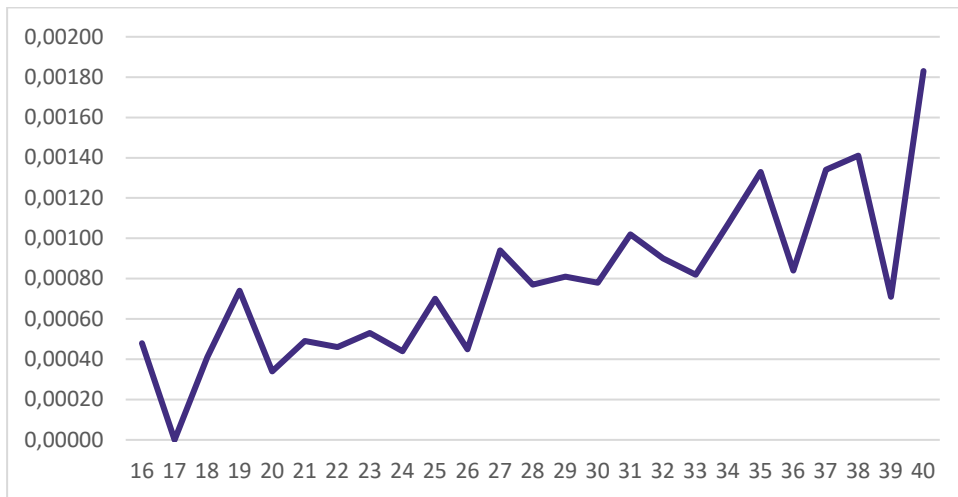
Joonis 1. Suremustõenäosus vanuses 0–100 Eestis, 2019



Joonis 2. Suremustõenäosus vanuses 0–15 Eestis, 2019



Joonis 3. Suremustõenäosus vanuses 16–40 Eestis, 2019



Kuna tugevasti hajuvad suremustõenäosused võivad arvutustulemustes, st oodatavates elukestustes, põhjustada ebaloogilisi ja raskesti seletatavaid efekte, siis suremustõenäosusi silutakse ehk igale aastale t vastav suremuse tõenäosus q_t^x asendatakse selle aasta ja lähiaastate suremistõenäosuste lineaarkombinatsiooniga, mis arvutatakse järgmiselt:

$$q_t^x = \sum_{i=t-2}^{t+2} q_i^x a_i,$$

kus kaalud a_i on mittenegatiivsed ja nende summa on 1. Tuntakse ja kasutatakse eri kaalumismeetodeid, mida on otstarbekas rakendada eri vanusepiirkonnas suremuskordajatele või ka teistele näitajatele erinevalt.

Viieaastase vanusevahemiku kasutamine

Suremuskordajaid aitab stabiilsemaks muuta ka viieaastaste vanuserühmade kasutamine üheaastaste vanuserühmade asemel. Kogu arvutustöö on põhimõtteliselt samasugune nagu üheaastaste vanuserühmade korralgi.