VILNIAUS UNIVERSITETAS

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MAGISTRANTŪROS STUDIJŲ PROGRAMA
FINANSŲ IR DRAUDIMO MATEMATIKA

Skirtingos aktuarinės mirtingumo prielaidos ir jų įtaka mirties rodiklių įverčiams

Different Actuarial Mortality Assumptions and Their Impact on Mortality Estimates

Magistro baigiamasis darbas

Autorius: Laurynas Lukoševičius
VU el. paštas: laurynas.lukosevicius@mif.stud.vu.lt,
Darbo vadovas: Doc. Dr. Andrius Grigutis

Vilnius
2024

MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATINĖS ANALIZĖS KATEDRA

Darbo vadovas Doc. dr. Andrius Grigutis ______________________________
Darbo recenzentas ____

Darbas apgintas 2024 sausio mėn. __ d. Darbas įvertintas ________________

Registravimo NR. ____
Įrašoma atidavimo į katedra data ______________________________________

Skirtingos aktuarinės mirtingumo prielaidos ir jų įtaka mirties rodiklių įverčiams

Santrauka

Kiekvienas verslas yra asocijuojamas su ilgojo laikotarpio nežinomybe ir rizika patirti netikėtus nuostolius. Gyvybės draudimo industrija nėra išimtis. Čia svarbiausia rizika kyla iš mirtingumo rodiklių kintamumo laike, nes draudimo įmonių rezervų skaičiavimai ir kainodara grindžiami klientų mirtingumo duomenimis. Todėl periodiška mirtingumo analizė aktuarams yra būtina. Pagrindiniai mirtingumo analizės žingsniai apima duomenų laikotarpio pasirinkimą, atsitiktinumo ignoravimą duomenyse, tinkamos mirtingumo taisyklės pasirinkimą ir gradavimą, standartinės mirtingumo lentelės interpoliavimo metodą ir saugių pasikliovimų intervalų nustatymą mirtingumo normoms. Šie žingsniai yra esminiai, siekiant kritiškai įvertinti mirtingumo normas ir užtikrintų patikimą bei pelningą įmonės veiklą gyvybės draudimo sektoriuje. Visų šių problemų sprendimui tobula metodika neegzistuoja, todėl darbe yra nagrinėjamas laisvai pasirinktas algoritmas, skirtas saugiai įvertinti klientų mirties rodiklius pagal jų amžių.

Raktiniai žodžiai: Gyvybės draudimas, mirtingumo norma, mirtingumo lentelės interpoliavimas, gradavimas, pasikliovimo intervalai, atsitiktinumas.

Different Actuarial Mortality Assumptions and Their Impact on Mortality Estimates

Abstract

Every business is associated with long-term uncertainty and the risk of incurring unacceptable losses. The life insurance industry is no exception. Here, the primary risk arises from the variability of mortality rates over time, as calculations of insurance company reserves and pricing are based on customers mortality data. Therefore, periodic mortality analysis is crucial for actuaries. Key steps in mortality analysis include choosing the data time frame, accounting for randomness in the data, selecting and graduating the data with an appropriate mortality law, using the standard mortality table interpolation method, and establishing secure confidence intervals for mortality norms. These steps are essential for a critical evaluation of mortality norms, ensuring a reliable and profitable operation in the life insurance sector. Given the absence of a universally perfect methodology to address these challenges, this work explores a specifically chosen algorithm designed to reliably assess clients' mortality indicators based on their age.

Key words: Life Insurance, mortality rate, mortality table interpolation, graduation, confidence intervals, randomness.

Turinys

Įvadas

Gyvybės draudimas – sąvoka, su kuria žmonės XXI amžiuje susiduria vis dažniau. Tai vienas iš populiariausių metodų, skirtas mažinti riziką individams. Pagal Lietuvos Banko viešai skelbiamus duomenis, 2019-ųjų metų pradžioje sumokėtos gyvybės draudimo įmokos Lietuvoje siekė beveik 21 mln eurų, kai 2023-iųjų metų pradžioje šis skaičius viršijo 28 mln eurų. Skaičiai iliustruoja populiarėjančią gyvybės draudimo sąvoką, sėkmingai augančią gyvybės draudimo rinką, bei žmonių pasitikėjimą šias paslaugas teikiančiomis įmonėmis. Šiai dienai gyvybės draudimo kompanijos siūlo ne vieną produktą. Prekiaujama tiek kaupiamojo, tiek investicinio gyvybės draudimo ar net sveikatos draudimo produktais. Paprasčiausia ir geriausiai pažįstama gyvybės draudimo forma - paprastasis rizikinis (klasikinis) gyvybės draudimas su nekintančia draudimo įmoka. Klientas, sudarydamas draudimo sutartį, įsipareigoja draudimo įmonei mokėti sutartyje numatytą draudimo įmoką. Sutartis numato draudiko įsipareigojimą išmokėti iš anksto sutartą gyvybės draudimo išmoką kontrakto naudos gavėjams tik tuo atveju, jei apdraustasis žmogus mirė ir įvykis yra pripažintas draudžiamuoju. Remiantis 2023-iųjų metų pradžios skaičiais, vidutinė rizikinio draudimo įmoka buvo 14,29 eurų, o vidutinė išmoka mirties atveju - 23 000 eurų. Akivaizdu, jog tokio tipo sutartys draudikams kelia nemažą riziką patirti nuostolius. Tuo tarpu draudėjų įsipareigojimai būna kur kas finansiškai menkesni.

Natūralu, jog paslaugos tiekėjas, vykdydamas savo veiklą, siekia tiek maksimalaus pelno, tiek finansinio stabilumo, bei geros reputacijos savo klientams. Būtent gyvybės draudimo atveju, įmonės pelnas yra stipriai priklausomas nuo klientų mirtingumo ir surenkamo įmokų dydžio. Sėkmingai veiklai užtikrinti, draudikas pasikliauja matematikos profesionalų - aktuarų skaičiavimais. Aktuarai padeda savo darbdaviams priimti objektyvius ir skaičiais pagrįstus sprendimus. Klasikinio gyvybės draudimo atveju yra svarbu įvertinti saugų įmokos dydį keliems dešimtmečiams į priekį, mat toks verslo planas yra ilgalaikis ir kontrakto laikotarpis gali siekti net 40 ar net 50 metų. Aktuarams yra svarbu išmanyti, kaip savo skaičiavimuose pritaikyti matematinius ir statisnius metodus bei kokie yra konstruojamo matematinio modelio privalumai ir trūkumai konkrečiai draudimo rūšiai. Tikimybių teorija gyvybės draudimo praktikoje yra neišvengiama. Dažniausiai, įmokų nustatymui ir rezervo, kylančių iš ateityje įvyksiančių žalų (angl. Liability Remaining Coverage, LRC) įvertinimui naudojamas populiarus sąlyginio vidurkio metodas - aktuarinė lygybė. Modelio tikslas sulyginti ateities pinigų srautus - būsimų įmokų dabartinė vertė turi būti lygi ateities išmokų (iš ateityje įvyksiančių žalų) dabartinei vertei. Tokio tipo skaičiavimuose dalyvauja sąlyginių tikimybių sandaugos. Akivaizdu, jog gebėjimas sukonstruoti aktuarinę lygybę draudikams yra be galo svarbus, tačiau tam atlikti yra reikalaujamas pirminis žingsnis - sąlyginių mirtingumo tikimybių kiekvienam žmogaus amžiui įvertinimas, dar kitaip vadinamas mirtingumo lentelių konstravimu.

Sukonstruota mirtingumo lentelė atspindi matematiškai įvertintą tikimybę žmogui mirti tam tikrame amžiaus intervale, su sąlyga, jog žmogus sėkmingai sulaukė šio intervalo pradinio amžiaus. Standartinis amžiaus intervalo ilgis klasikinėje mirtingumo lentelėje - vieneri metai. Be abejonės, šiai procedūrai atlikti yra būtina įmonės portfelio mirčių statistika. Natūralu, jog laikui bėgant, mirties rodikliai keičiasi. To priežastys gali būti įvairios - sparčiai besivystanti šalies medicina, kokybiškesnis kliento rizikos vertinimo procesas (angl. insurance underwriting) ar net šalies ekonominio pakilimo periodas. Dėl visų šių priežasčių yra labai svarbu pasirinkti tinkamą laikotarpį, kurio statistiniai mirčių rodikliai galėtų atspindėti mirtingumą dešimtims metų į priekį, kas reiškia, jog per ilga laiko eilutė gali nuvertinti/pervertinti portfelio mirtingumą. Iš čia kyla praktinė problema - duomenų nepakankamumas ir vyraujantis atsitiktinumas. Empiriniai mirčių duomenys kartais parodo didesnę tikimybę mirti jaunesniam žmogui nei vyresniam. Pavyzdžiui, 49-erių metų žmogui priskirta tikimybė mirti vienerių metų amžiaus intervale yra didesnė, nei penkiasdešimtmečiui. Standartinės (amžius yra sveikasis skaičius) mirtingumo lentelės konstravimas, pašalinant atsitiktinumą, vadinamas mirtingumo gradavimu (angl. mortality graduation). Šis procesas yra duomenų interpoliacija. Jo atlikimo metu svarbu pasirinkti tinkamą mirtingumo dėsnį atitinkančią funkciją. Populiariausi dėsnių pavyzdžiai - Makeham ir Gompertz. Po gradavimo proceso, aktuarai privalo atlikti papildomą mirtingumo interpoliaciją tarp gretimų amžių mirties įverčių, kadangi žmogaus amžius nebūtinai yra sveikasis skaičius. Populiariausios klasikinės mirtingumo lentelės interpoliavimo prielaidos: tolygusis, pastovios mirtingumo galios ir Balducci. Paprastai tariant, mirtingumo lentelės konstravimas, tuo pačiu ir draudimo sutarčių įmokų, draudiko pakankamo kapitalo bei įmonės rezervų skaičiavimai yra tiesiogiai priklausomi nuo mirčių rodikliams įvertinti naudotų aktuarinių prielaidų.

Darbo tikslas yra iliustruoti gyvybės draudimo verslo kompleksiškumą, aktuarinių prielaidų jautrumą mirties rodikliams, pritaikyti laisvai pasirinktą mirtingumo analizės algoritmą, bei jo pagalba atrasti geriausius mirtingumo įverčius Lietuvos populiacijos duomenims. Šio darbo struktūra yra tokia: pirmame skyriuje pristatomi pagrindiniai išgyvenamumo dydžiai. Antrasis skyrius skirtas matematiniui mirtingumo lentelės konstravimo ir analizės algoritmo aprašymui. Trečiame skyriuje pereinama prie praktinės dalies - algoritmo taikymo pasirinktam duomenų masyvui, bei skirtingų simuliacijų rezultatų palyginimo. Ketvirtame skyriuje - tyrimo rezultatai ir išvados.

1. Pagrindiniai aktuariniai išgyvenamumo dydžiai

Tiek mirtingumo gradavimo procesas, tiek mirtingumo lentelės interpoliavimo ir įmokų skaičiavimo metodai reikalauja pagrindinių aktuarinių išgyvenamumo dydžių apibrėžimų ir jų matematinių savybių taikymų. Būtent šiame skyriuje yra aprašomi dažniausiai sutinkami mirtingumo rodikliai ir jų matematinės savybės.

Žmogaus gyvenimo trukmė yra neneigiamas atsitiktinis dydis, kurį žymėsime X. Matematiškai, atsitiktinis dydis yra išmatuojama funkcija, atvaizduojanti galimus įvykius tikimybinėje erdvėje (Ω, $\mathcal{A}$, $\mathbb{P}$) į realių skaičių aibę $\mathbb{R}$. Pagrindinių X charakteristikų nagrinėjimui yra daroma prielaida, jog neneigiamas atsitiktinis dydis X yra absoliučiai tolydus, kas reiškia, jog egzistuoja neneigiama integruojama funkcija $f(x)$, kuriai galioja

$$\mathbb{P}(X < x) = \int\limits_0^x f(u)\,du,\; x \in [0, \infty).$$

1.1 Išgyvenamumo, tikimybinio tankio funkcijos ir sąlyginės tikimybės

Pažymėkime absoliučiai tolydaus atsitiktinio dydžio X (laikas nuo gimimo iki mirties) **pasiskirstymo funkciją** $F(x) = P(X \leq x)$, kuri žymės tikimybę, jog naujagimis neišgyvens iki laiko momento x. Tuomet pasiskirstymo funkcija $F(X)$ tenkina šias savybes:

1. $F(0) = 0$,
2. $\lim_{x\to\infty} F(x) = 1$.

Tikimybė, kad tam tikros populiacijos naujagimis išgyvens iki laiko momento x: $S(x) = P(X > x)$. Čia X - mirties momentą aprašantis tolydus atsitiktinis dydis (naujagimio būsima gyvenimo trukmė). Funkciją $S(x)$ vadinsime **išgyvenamumo funkcija**, kuri tenkina šias savybes:

1. $S(x)$ nedidėja, kai $x \geqslant 0$,
2. $S(0) = 1$,
3. $S(+\infty) = 0$,
4. $S(x)$ yra tolydi apibrėžimo srityje.

Taip pat:

$$S(x) = \int_x^\infty f(y)\,dy, \tag{1.1}$$

čia $f(x)$ - jau prieš tai minėta atsitiktinio dydžio X **tikimybinio tankio funkcija**.

Išgyvenamumo (mirtingumo) modeliavimuose dažnai sutinkamos **sąlyginės tikimybės**:

$${}_tp_x = \frac{S(x+t)}{S(x)},$$

tikimybė, kad amžiaus x individas sulauks $x + t$ gimtadienio, su sąlyga, jog buvo gyvas amžiuje x.

Vienas iš pagrindinių mirtingumo rodiklių aktuarinėje matematikoje:

$${}_tq_x = 1 - {}_tp_x,$$

tikimybė, kad amžiaus x individas mirs amžiaus intervale $[x, x + t]$, su sąlyga, jog buvo gyvas amžiaus x.

Pažymėkime l_x kohortos x amžiaus žmonių skaičių, d_x - tos pačios kohortos x amžiaus žmonių mirčių skaičių. Jei $t = 1$, tuomet pastarosios tikimybės žymimos paprasčiau ir yra išreiškiamos šiomis trupmenomis:

$$p_x = \frac{l_{x+1}}{l_x} = \frac{S(x+1)}{S(x)},$$

$$q_x = \frac{d_x}{l_x} = 1 - p_x.$$

Dažniausiai, aktuarų galutinis tikslas ir yra kuo tiksliau įvertinti pastarąsias tikimybes ir sukonstruoti mirtingumo lentelę.

1.2 Mirtingumo galia

Dažnai naudojamas išgyvenamumo modelių dydis – **mirtingumo galia,** μ_x. Tolydžiu atveju, mirtingumo galią galime apibrėžti kaip sąlyginę tikimybę, kad x amžiaus sulaukęs asmuo mirs per labai trumpą laiko intervalą $[x, x + \Delta x]$:

$$\mu_x = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x(1 - F(x))}.$$

Iš to išplaukia, jog

$$\mu_x = \frac{-\frac{d}{dx}S(x)}{S(x)} = \frac{f(x)}{S(x)}. \tag{1.2}$$

Mirtingumo galia aktuariniuose modeliuose turi stiprų sąryšį su kitais svarbiais aktuarinės matematikos dydžiais:

Išgyvenamumo funkcija:

$$S(x) = e^{-\int_0^x \mu_y \, dy}, \tag{1.3}$$

Vidutinė pilnoji (lauktina) gyvenimo trukmė:

$$\dot{e_x} = \int_0^\infty e^{-\int_0^x \mu_y \, dy} \, dx,$$

Tikimybė žmogui mirti intervale $[x, x + 1]$**, su sąlyga, jog žmogus sulaukė amžiaus** x**:**

$$q_x = 1 - \frac{S(x+1)}{S(X)} = 1 - \frac{e^{-\int_0^{x+1} \mu_y \, dy}}{e^{-\int_0^x \mu_y \, dy}}. \tag{1.4}$$

1.3 Centrinis mirtingumo dažnis

Dar vienas mirtingumą aprašantis matas, kuris siejamas su mirtingumo galia – **centrinis mirtingumo dažnis,** m_x (angl. central death rate). Centrinį mirtingumo dažnį galime apibrėžti kaip mirtingumo galios svertinį vidurkį, svoriais imant tikimybes išgyventi amžiaus intervale $[x, x + 1]$:

$$m_x = \frac{\int_x^{x+1} S(y)\mu_y \, dy}{\int_x^{x+1} S(y) \, dy},$$

Jei μ_{x+s} yra mirtingumo galia, kai žmogaus amžius nėra sveikasis skaičius, t.y jei $0 < s < 1$, tai centrinis mirtingumo dažnis išreiškiamas:

$${}_nm_x = \frac{\int_x^{x+1} {}_np_x\mu_{x+s} \, dy}{\int_x^{x+1} S(y) \, dy}. \tag{1.5}$$

1.4 Nugyventų metų skaičius

Pažymėkime visų asmenų bendrą **nugyventų metų skaičių** (angl. exposure), esant būtent amžiaus x – E_x. Tuomet, tų žmonių, kurie būdami amžiaus x, gyvi sulauks $x + 1$-ojo gimtadienio, indėlis dydyje E_x yra po vienetą (pilnus metus). Natūralu, jog stebėdami pakankamai didelį kiekį x amžiaus žmonių, turėsime tokių atvejų, kai žmogus miršta nesulaukęs $x + 1$-ojo gimtadienio. Šių asmenų įtaka bus dalinė, t.y. skaičiuojamas tik laikas nuo paskutinio gimtadienio iki jų mirties.

Tarkime, jog turime gyvų x amžiaus asmenų skaičių iš pradinės uždaros naujagimių kohortos ir šį dydį pažymėsime - l_x. Jeigu padarysime papildomas prielaidas ir sujungsime du gretimus taškus l_x ir l_{x+1} (kohortos x ir x+1 amžių asmenų skaičiai) tam tikra funkcijos forma, tuomet sandaugą ${}_sp_x\mu_{x+s}$ interpretuosime kaip mirtingumo amžiuje x+s tankį, su sąlyga, kad asmuo buvo gyvas būdamas amžiaus x. Padauginę šį dydį iš l_x, turime $l_{x+s}\mu_{x+s}$, ką interpretuosime kaip mirčių greitį kohortoje, jeigu asmenų amžius yra x + s (0 < s < 1).

Taip pat:

$$l_{x+s}\mu_{x+s}ds$$

galime interpretuoti kaip x-mečių asmenų mirčių skaičių esant jau x+s amžiaus (intervale $[s, s+ds]$).

Tada:

$$sl_{x+s}\mu_{x+s}ds$$

yra asmenų, kurie mirė amžiaus intervale $[x; x+1]$ bendras nugyventų metų skaičius.

Galiausiai, turėdami prielaidas apie l_{x+s} funkcijos formą:

$$\int_0^1 sl_{x+s}\mu_{x+s}ds$$

gauname bendrus asmenų, kurie sulaukė amžiaus x, bet mirė iki $x+1$-ojo gimtadienio, nugyventų metų skaičius.

Kadangi dydis E_x matuoja bendrą x-mečių nugyventų metų skaičių amžiaus intervale $[x, x+1]$, tuomet:

$$E_x = l_{x+1} + \int_0^1 sl_{x+s}\mu_{x+s}ds = \int_0^1 l_{x+s}ds. \tag{1.6}$$

Prisiminkime centrinio mirtingumo dažnio (1.4) išraišką:

$$m_x = \frac{\int_0^1 {}_sp_x\mu_{x+s}\,ds}{\int_0^1 {}_sp_x\,ds} = \frac{\int_0^1 {}_sp_x l_x\mu_{x+s}\,ds}{\int_0^1 {}_sp_x l_x\,ds} = \frac{\int_0^1 l_{x+s}\mu_{x+s}\,ds}{\int_0^1 l_{x+s}\,ds}$$

Pažymėkime $E_x = \int_0^1 l_{x+s}\,ds$ bei pastebėkime, kad:

$$\int_0^1 l_{x+s}\mu_{x+s}ds = -\int_0^1 dl_{x+s} = -(l_{x+1} - l_x) = d_x,$$

čia d_x – mirčių skaičius amžiaus intervale $[x, x+1]$. Tuomet centrinį mirtingumo dažnio įvertį galime užrašyti išraiška:

$$\hat{m}_x = \frac{d_x}{E_x}.$$

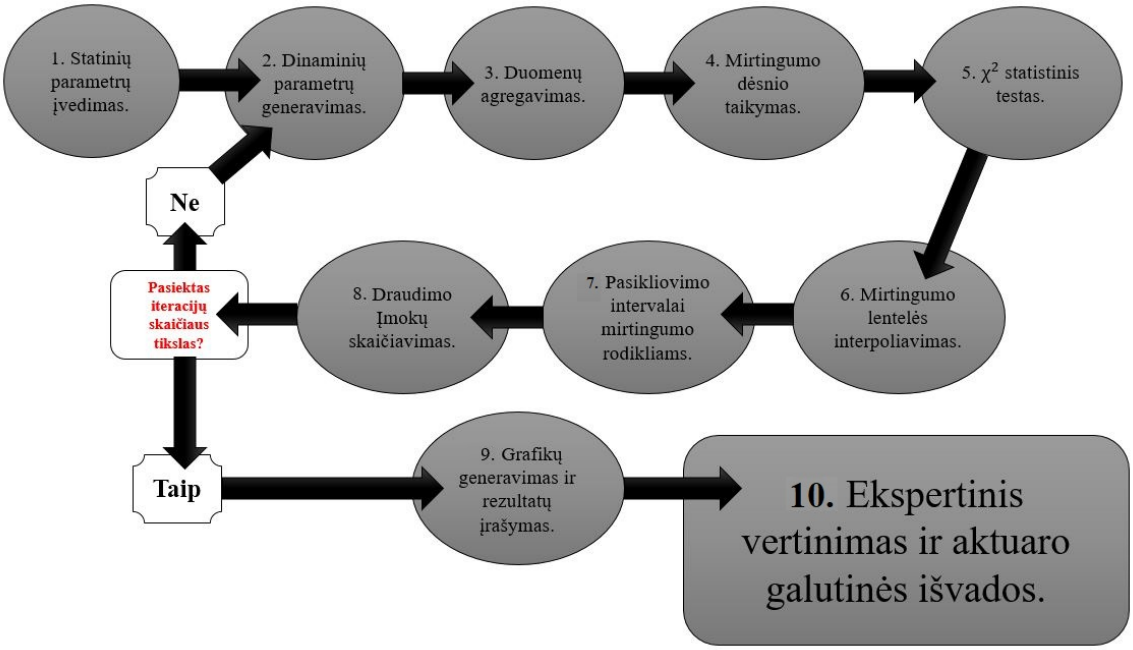
Praktikoje, skirtingomis kalendorinių metų dienomis, dydis l_x yra kintantis. Būtent dėl šios priežasties, praktikoje, yra priimtina skaičiuoti centrinio mirtingumo dažnio įvertį ir taikyti aproksimaciją:

$$\hat{q}_x \approx \hat{m}_x. \tag{1.7}$$

2. Mirtingumo įverčių modeliavimo teorija

Šiame skyriuje aprašoma laisvai pasirinkto mirtingumo įverčių skaičiavimo algoritmo matematinė teorija. 1976-iais metais, britų statistikas George Box leidinyje *"Journal of the American Statistical Association"* įamžino populiarią frazę *"visi modeliai yra klaidingi, tačiau kai kurie gali būti naudojami"*. Laisvas algoritmo pasirinkimas, kaip ir statistiko George Box posakis, reiškia tą patį, jog nėra garantijos, kad taikomi metodai yra patys geriausi ir neegzistuoja joks kitas alternatyvus (ar net geresnis) algoritmas. Šiame darbe aprašoma praktika gali būti taikoma tiek šalies populiacijos, tiek gyvybės draudimo įmonių portfelio duomenims.

Mirtingumo rodiklių įverčiai naudojami gyvybės draudimo produktų kainodaros, įmonės rezervų formavimų pagal *Tarptautinės Finansinės Atskaitomybės Standartus* (angl. International Financial Reporting Standard, IFRS) ar net *Mokumas II* (angl. Solvency II) kapitalo pakankamumo skaičiavimų procesuose. Galutiniams mirties tikimybių įverčiams nemažą įtaką turi aktuarų daromos prielaidos. Būtent dėl šios priežasties, aprašomo algoritmo tikslas yra įvykdyti tam tikrą skaičių iteracijų ir padėti aktuarui atsirinkti geriausiai tinkančias prielaidas bei matematinius/statistinius metodus analizuojamame portfelyje. Kiekvienos simuliacijos metu, modelis atsitiktinai parenka **dinaminius** mirtingumo analizės parametrus, skirtinguose algoritmo žingsniuose atlieka vieną prielaidą ir pritaiko atitinkamą matematinį metodą. Nepaisant to, kai kurie modelyje naudojami parametrai yra **statiniai** ir dėl gyvybės draudimo verslo kompleksiškumo bei duomenų nepakankamumo, turi būti nustatyti pačių aktuarų, pasitelkiant sukaupta patirtimi ir ekspertiniu vertinimu. Galutinis modelio rezultatas - sugeneruotas duomenų sąrašas su kiekvienoje simuliacijoje atrinktų parametrų, jų tinkamumo vertinimo informacija. Taip pat, kompiuteryje išsaugomi ir kiekvienai iteracijai sugeneruoti grafikai. Toliau seka aktuarų ekspertinis vertinimas geriausiems atrinktiems scenarijams ir galutinių įverčių pasirinkimas.



1 pav.: Modelio algoritmas.

Parametro Pavadinimas	Aprašymas	Parametro Rūšis
Iteracijų skaičius.	Vartotojas pasirenka norimą iteracijų skaičių. Didesnis iteracijų skaičius prailgina modelio veikimo laiką ir padidina tikimybę atlikti kokybiškesnę mirtingumo analizę.	Statinis.
Maksimalus galimas laiko eilutės ilgis.	Vartotojas pasirenka didžiausią toleruojamą laiko eilutės ilgį. Per daug ilga/trumpa laiko eilutė gali iškreipti analizės rezultatus.	Statinis.
Mirtingumo analizės periodo pabaiga.	Vartotojas pasirenka analizės periodo pabaigą (kalendorinius metus).	Statinis.
Mirtingumo analizės periodo pradžia.	Modelis atsitiktinai parenka tokią periodo pabaigą, kad analizės periodo ilgis neviršytų maksimalaus galimo laiko eilutės ilgio	Dinaminis.
Išskirtiniai metai, neįtraukti į analizę.	Vartotojo pasirinkimas. Išskiriami metai, kaip COVID-19 pandemija.	Statinis.
Analizuojamo žmonių amžiaus intervalas.	Vartotojo pasirinkimas.	Statinis.
Mirtingumo taisyklių skaičius.	Maksimalus modelio mirtingumo taisyklių skaičius yra 2.	Dinaminis.
Amžiaus taškas, kuriame keičiasi mirtingumo taisyklė.	Jeigu modelio pasirinkimas yra dvi taisyklės, šis amžiaus taškas nurodo, iki kurio amžiaus turi būti taikomas pirmasis dėsnis.	Dinaminis.
Amžiaus grupavimo vektorius.	Mažuose portfeliuose pasireiškia didelis mirtingumo atsitiktinumas. Mirtingumo rodikliai pagal amžių svyruoja ir nėra lengva nustatyti mirtingumo dėsnį bei atlikti gradavimą. Tam praktikoje yra taikomas duomenų grupavimas pagal amžių.	Dinaminis.
Žmonių grupių skaičius.	Parametras priklauso nuo laiko eilutės ilgio ir kalendorinių metų. Kaip pavyzdys, žmonių grupės gali būti trys - jaunimas, suaugę ir pensininkai.	Dinaminis.

1 lentelė: Modelio parametrai.

Pagrindiniai modeliavimo žingsniai, formuojantys subjektyvų mirtingumo vertinimą:

1. Laiko eilutės ilgio pasirinkimas,
2. Skirtingas duomenų grupavimas pagal amžių,
3. Skirtingi mirtingumą aprašantys matematiniai dėsniai (Makeham, Gompertz),
4. Skirtingi mirtingumo lentelės interpoliavimo metodai (tolygusis, Balducci),
5. Mirčių skaičiaus tam tikrame amžiuje pasiskirstymas.

2.1 Atsitiktinis dinaminių parametrų generavimas

Mirtingumo analizės periodo pradžia:

Pažymėkime maksimalų laiko eilutės ilgį, kurį apibrėžia modelio vartotojas - $r \in \mathbb{N}$, statinę analizės periodo pabaigą - $y_{max} \in \mathbb{N}$, tenkinančią savybę $y_{max} \geq r - 1$. Tuomet, diskretus atsitiktinis dydis $Y_i \in \mathbb{N}$, $i \in \mathbb{N}$ (i žymi iteracijos skaičių), kurį vadinsime analizės periodo pradžia, tenkina šią savybę:

$$y_{max} - r + 1 \leq Y_i \leq y_{max}$$

Pavyzdys: jei $r = 5, y_{max} = 2019$*, tada* $Y_i \in [2015, 2019]$.

Mirtingumo taisyklių (dėsnių) skaičius:

Tokie eksponentinio tipo mirtingumo dėsniai, kaip Makeham ar Gompertz, yra paremti logika, jog mirtingumas su amžiumi griežtai didėja. Vėliau pristatyti dėsniai, kaip Perks arba Beard, yra logistinio tipo, reiškiantys, jog mirtingumas griežtai didėja, bet mirtingumo intensyvumas - nebūtinai (daugiau 2.3 skyrelyje). Tačiau draudimo praktikoje pasitaiko atvejų, kai mirtingumo kreivė nėra griežtai auganti. Pavyzdžiui, paauglių atvejis *(angl. accidental hump)*, kuomet jaunuoliai neturėdami pakankamos vairavimo patirties patenka į avarijas ir žūsta. Šiam tikslui, modelyje yra įdiegta dviejų mirtingumo taisyklių kombinacija atliekant gradavimo procesą. Antroji mirtingumo taisyklė yra 3-ojo laipsnio polinominė aproksimacija. Dėl šios priežasties, atrinktas mirtingumo taisyklių skaičius (i-oje iteracijoje) yra atsitiktinis dydis $N_i \in \{1, 2\}$, $i \in \mathbb{N}$.

Amžiaus taškas, kuriame keičiasi mirtingumo taisyklė:

Pažymėkime statinius parametrus, įvedamus vartotojo, analizuojamo žmonių amžiaus apatinį rėžį - $x_{min} \in \mathbb{N}_0$, bei analizuojamo žmonių amžiaus viršutinį rėžį - $x_{max} \in \mathbb{N}$. Paprastumo dėlei, laikysime, kad $x_{max} > x_{min} + 10$ (svarbu tolimesniuose skyreliuose). Diskretų atsitiktinį dydį, amžiaus tašką, kuriame keičiasi mirtingumo dėsnis, žymėsime $X_{mid} \in \mathbb{N}$. Atsitiktinis dydis yra generuojamas tik specifiniu atveju, kai simuliacijos atrinktas mirtingumo dėsnių skaičius yra 2, t.y:

$$\text{Jei } N_i = 2, i \in \mathbb{N}, \quad X_{\text{mid}_i} \in \left[\max\left(\frac{x_{\text{min}} + x_{\text{max}}}{4}, x_{\text{min}} + 10 \right), \frac{x_{\text{min}} + x_{\text{max}}}{2} \right]$$

Pavyzdys: jei $x_{min} = 18$, $x_{max} = 65$ *,tada* $X_{mid} \in [28, 42]$. *Apribojimas yra reikalingas norint išvengti pilnos polinominės aproksimacijos. Apie tai plačiau skyriuose 2.3 ir 2.4.*

Amžiaus grupavimo vektorius:

Dažna, praktiškai sutinkama problema gyvybės draudimo matematikoje - duomenų nepakankamumas kiekviename amžiaus taške. Nepakankamas kiekis duomenų gali privesti prie mirčių rodiklių pervertinimo arba nuvertinimo tam tikroms amžiaus grupėms (ypač polinominė aproksimacija). Aktuarai, siekdami išvengti šios problemos, taiko mirtingumo duomenų grupavimą pagal amžių. Čia ir kyla klausimas, kokio ilgio amžiaus intervalus pasirinkti tinkamam duomenų grupavimui? Ar pasirinktas intervalo ilgis tam tikrai amžiaus grupei nėra per trumpas empirinio atsitiktinumo pašalinimui? Galbūt skirtingoms amžiaus grupėms reikalingi nevienodo ilgio intervalai?
Nagrinėkime modelyje analizuojamą amžiaus intervalą $[x_{min}, x_{max}]$. Sukonstruotą ir kiekvienoje iteracijoje esantį skirtingą amžiaus grupavimo vektorių žymėsime Z_i, o jo elementus $z_{i,j}$, $i \in \mathbb{N}$, $j \in \mathbb{N}$. Indeksas i žymi iteraciją, o j - vektoriaus elemento numerį. Kadangi vienas iš pagrindinių modelio tikslų yra aptikti patį geriausią amžiaus grupavimo vektorių, natūralu, jog skirtingose iteracijose, vektorių ilgiai yra nebūtinai tokio paties ilgio.

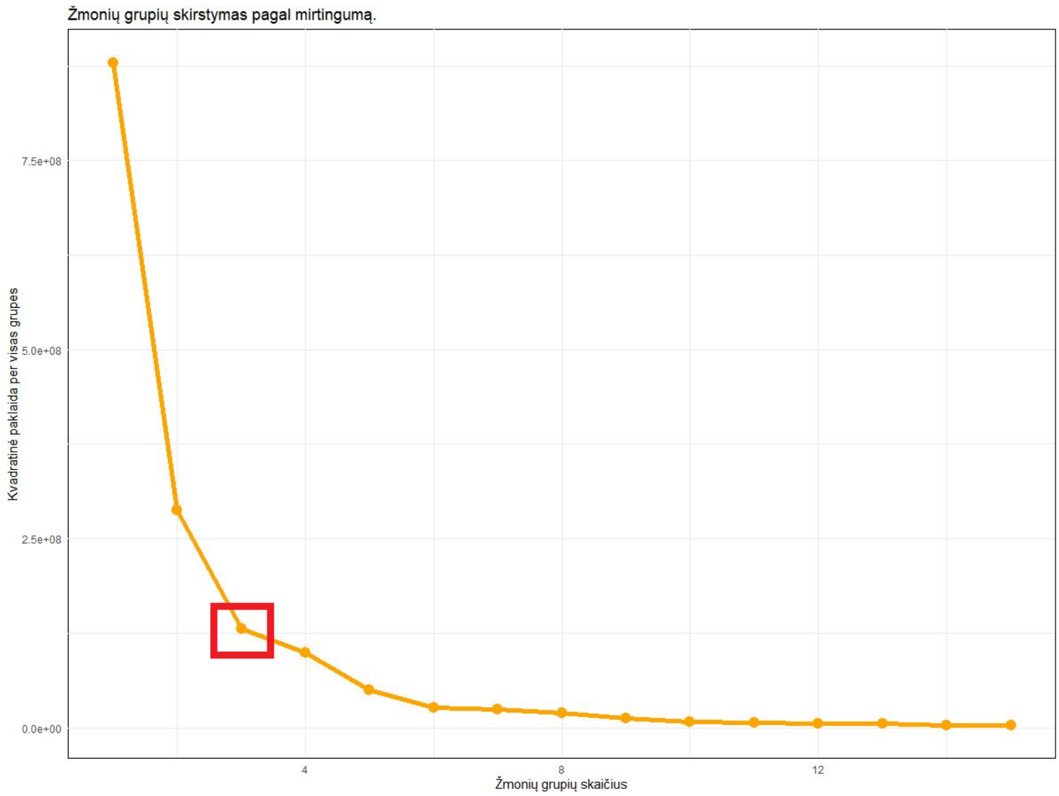
Vektoriaus elementai $z_{i,j}$ tenkina šias savybes:

1. $z_{i,1} = x_{min}$, $z_{i,1} \in \mathbb{N}_0$,
2. $z_{i,n} = x_{max}$, $z_{i,n} \in \mathbb{N}$, - čia n žymi paskutinio vektoriaus elemento indeksą.
3. $z_{i,j+1} = z_{i,j} + 1$, $z_{i,j+1} \in \mathbb{N}$ ir $j = 2n$, kai $n \in \mathbb{N}$,
4. $z_{i,j+2} \geq z_{i,j} + 2$, $z_{i,j+2} \in \mathbb{N}$,

Paskutinis, ketvirtasis apribojimas padaro minimalų grupės ilgį lygų 2. Paprastai tariant, vektoriaus Z_i gretimi nelyginio ir lyginio indeksų elementai reiškia mirtingumo grupavimą "nuo $z_{i,j}$ amžiaus iki amžiaus $z_{i,j+1}$". Čia indeksas $j = 2n - 1$, kai $n \in \mathbb{N}$.

Teisingai sukonstruoto amžiaus grupavimo vektoriaus pavyzdys:

$$Z_1 = (18, 24, 25, 30, 31, 33, 34, 41, 42, 45, 46, 48, 49, 53, 54, 55, 56, 60, 61, 65)$$



2 pav.: "K-vidurkių klasifikavimo" metodo taikymas. "Alkūnės" lūžio taškas (raudoname kvadrate) išskiria tris skirtingas žmonių grupes pagal mirtingumą.

Verta paminėti, jog amžiaus grupavimo vektoriaus generavimui, atsižvelgiama į "aptiktas" žmonių grupes - jaunimas, suaugę žmonės, senjorai ir panašiai. Tokia praktika padeda surasti tinkamą grupavimą mažesniame iteracijų skaičiuje. Grupių aptikimui yra naudojamas **k-vidurkių klasifikavimo** *(angl. K-Means Clustering)* algoritmas. Modelio praktika - nustatyti reikšmingą grupių skaičių duomenyse pagal didėjančius mirtingumo rodiklius bei nurodyti, kuriame amžiuje keičiasi žmonių grupė. Nustatyti amžiaus taškai, kurie reprezentuoja žmonių grupės tame taške pasikeitimą yra naudojami amžiaus grupavimo vektoriuje, be išimties. Tiesa, praktiškai, toks grupavimas yra per didelis, todėl, tarp gretimų, amžiaus grupes reprezentuojančių taškų, generuojami atsitiktinio ilgio vidinių grupių *(angl. sub-group)* intervalai. Trumpai tariant, pirminis modelio tikslas yra nustatyti žmonių grupes ir jų reikšmingą skaičių (jaunimas, suaugę, senjorai ir panašiai). Antrinis tikslas yra parinkti atitinkamo ilgio amžiaus vidinius intervalus kiekvienai žmonių grupei. Praktinis idėjos paaiškinimas būtų toks, jog draudimo kompanijose, ypatingai jauniems žmonėms, duomenų yra ganėtinai mažai (lyginant su 30-40 metų amžiaus žmonių intervalu), todėl, norėdami labiau išvengti atsitiktinumo, jaunimo duomenims turėtume taikyti grupavimą pagal ilgesnius amžiaus vidinius intervalus.

2.2 Mirtingumo duomenų agregavimas

Natūralu, jog vienerių kalendorinių metų mirtingumas gali būti atsitiktinai per mažas/didelis. To pasekoje, aktuarai, norėdami įvertinti realistišką mirtingumą, stebi ilgesnį laikotarpį - trejus, penkerius ar net septynerius metus. Apibrėžkime mirtingumo analizės periodo kalendorinių metų vektorių:

$$Y = (y_1, y_2, y_3, ..., y_n), n \in \mathbb{N}$$

čia $y_1 = y_{max} - r + 1$ ir $y_n = y_{max}$, o r - maksimalus laiko eilutės ilgis.

Pažymėkime amžiaus grupių vektorių:

$$Z = (z_1, z_2, z_3, ..., z_n), n \in \mathbb{N}$$

Mirčių skaičių tam tikrame amžiuje x ir tam tikrais kalendoriniais metais t žymėsime $D_{x,t}$, nugyventų metų skaičių - $E_{x,t}$. Tuomet mirčių skaičių n-oje grupėje žymėsime D_{g_n}, kuris apskaičiuojamas:

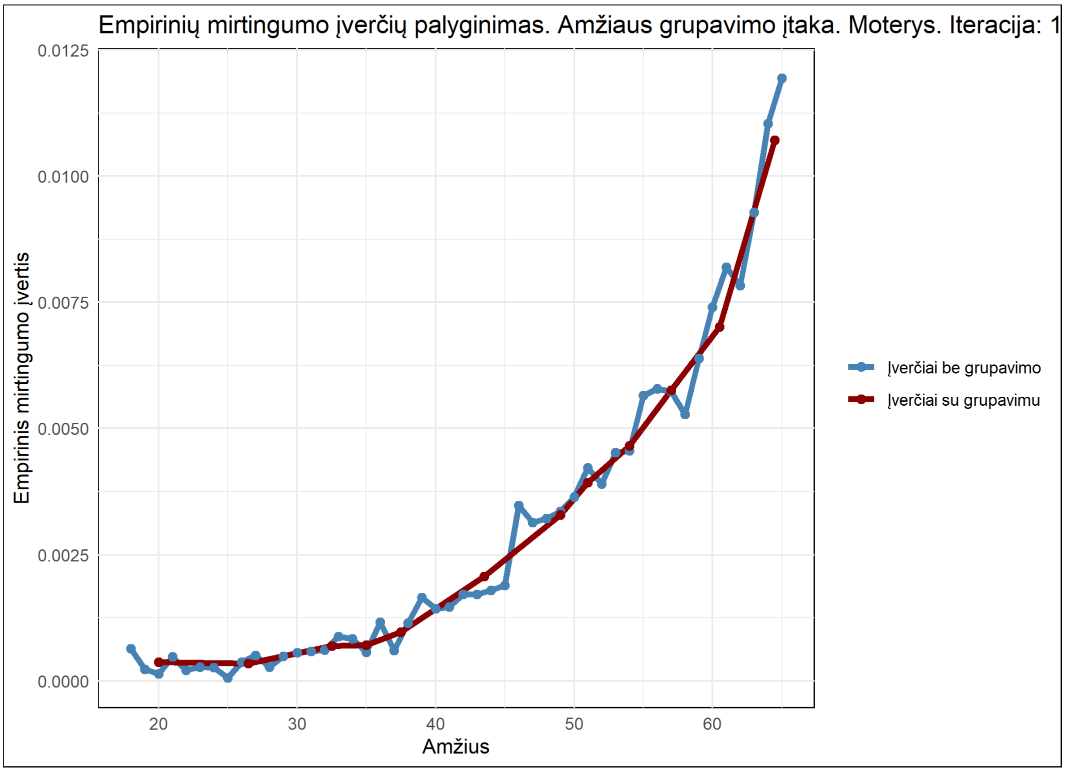
$$D_{g_n} = \sum_{t=y_1}^{y_n} \sum_{x=z_{2n-1}}^{z_{2n}} D_{x,t}.$$

Analogiškai, n-tosios grupės bendras nugyventų metų skaičius:

$$E_{g_n} = \sum_{t=y_1}^{y_n} \sum_{x=z_{2n-1}}^{z_{2n}} E_{x,t}.$$

Galiausiai, n-tosios grupės empirinis mirtingumo įvertis:

$$\hat{q}_{g_n} = \frac{D_{g_n}}{E_{g_n}}.$$



3 pav.: Amžiaus grupavimo taikymo įtaka. 2019-ųjų metų Lietuvos populiacijos duomenys.

2.3 Mirtingumo dėsniai

Kaip ir minėjome (2.1) skyrelyje, amžiaus grupavimas į ilgesnius intervalus ir duomenų agregavimas pagal amžiaus grupes, padeda dalinai išvengti atsitiktinių rodiklių svyravimų. Galutiniam atsitiktinumo pašalinimui yra naudojami parametriniai mirtingumo dėsniai. **Mirtingumo dėsnis (taisyklė)** - tam tikros formos tolydi matematinė funkcija, priskirta atitinkamam amžiaus intervalui, aprašanti mirtingumo elgseną kiekviename amžiaus taške. Kaip ir minėta 2.1 skyrelyje, išskirtiniais atvejais, modeliavimas gali būti atliekamas taikant dvi ar net daugiau mirtingumo taisyklių skirtinguose amžiaus intervaluose.

Šiame skyrelyje aprašysime pagrindinius dėsnius, sutinkamus praktiniame žmogaus gyvenimo trukmės X modeliavime. Dažniausiu atveju, matematikai renkasi mirtingumo modeliavimą pritaikant mirtingumo galios μ_x apibrėžimą. Tokio metodo privalumas yra, jog q_x galima tiesiogiai išisreikšti per μ_x matematinę formulę, bet ne atvirkščiai. Be to, q_x modeliavimui privalo būti taikomas apribojimas, jog tikimybė negali viršyti vieneto. Todėl mirtingumo modeliavimas, naudojant μ_x, tampa gerokai lengvesnis. Žemiau pateikiamos visos, modelyje naudojamos, mirtingumo taisyklės.

1. *Gompertz*[1] *1825 metais pasiūlytas dėsnis.* Žmogaus gyvenimo trukmės X mirtingumo galia yra

$$\mu_x = Be^{\alpha x}\ ,\ x \geqslant 0,$$

kai $B > 0$ ir $\alpha > 0$ yra parametrai, kurie priklauso nuo žmogaus gyvenamosios vietos, lyties. Didesnį portfelį turinčių gyvybės draudimo kompanijų atveju, šie parametrai gali priklausyti ir nuo žmogaus darbo specifikos, ligų istorijos ir dar daug kitų panašių aplinkybių. Pastarasis mirtingumo dėsnis yra standartinis, eksponentinio tipo, reiškiantis, jog mirtingumo intensyvumas su amžiumi **griežtai** didėja.

Apskaičiuokime svarbiausius mirtingumo rodiklius Gompertz atveju. Pritaikę (1.3) išgyvenamumo funkcijos $S(x)$ išraišką, ją apskaičiuosime Gomperz pasiūlytam mirtingumo dėsniui. Išgyvenamumo funkcijos $S(x)$ ir mirtingumo galios μ_x sąryšis gali būti aprašomas diferencialine lygtimi:

$$\frac{dS(x)}{dx} = -\mu_x S(x).$$

Tuomet:

$$\frac{dS(x)}{dx} = -Be^{\alpha x} S(x).$$

Belieka atskirti kintamuosius ir integruoti:

$$\int \frac{1}{S(x)} dS(x) = -B \int e^{\alpha x} dx.$$

Integruojame abi puses:

$$\ln |S(x)| = -\frac{B}{\alpha} e^{\alpha x} + C.$$

Tada:

$$S(x) = e^C e^{-\frac{B}{\alpha} e^{\alpha x}}.$$

Konstantos C nustatymui naudosime pradinę išgyvenamumo funkcijos savybę $S(0) = 1$:

$$e^C e^{-\frac{B}{\alpha}} = 1$$

Iš to išplaukia, kad $e^C = e^{\frac{B}{\alpha}}$, todėl $C = \frac{B}{\alpha}$.
Tuomet $S(x)$ sprendinys Gompertz mirtingumo dėsnio atveju:

$$S(x) = e^{-\frac{B}{\alpha}(e^{\alpha x} - 1)}, \quad x \geqslant 0.$$

Atitinkamai, kitos Gompertz atvejo išraiškos:

$$F(x) = 1 - e^{-\frac{B}{\alpha}(e^{\alpha x} - 1)},\ x \geqslant 0,$$

[1] Benjamin Gompertz (1779 — 1865) – britų aktuaras.

$$f(x) = \frac{dF}{dx} = B\alpha e^{\alpha x} e^{-\frac{B}{\alpha}(e^{\alpha x}-1)}, \quad x \geqslant 0.$$

Ir galiausiai:

$$q_x = 1 - \frac{S(x+1)}{S(x)} = 1 - e^{\frac{B}{\alpha}e^{\alpha x}} e^{\frac{B}{\alpha}e^{\alpha x}(e^{\alpha}-1)}, \quad x \geqslant 0. \tag{2.1}$$

Praktikoje, q_x formulė yra svarbiausia, o toliau jau atliekamas tiesioginis jos taikymas realiems duomenims ir parametrų nustatymas, kurio metodologija bus aprašoma kituose skyreliuose.

2. *Makeham*[2] *1860 metais pasiūlytas mirtingumo dėsnis.* Dėsnis, kaip ir Gompertz atveju, yra standartinis, eksponentinio tipo. Žmogaus gyvenimo trukmės mirtingumo galia yra aprašoma:

$$\mu_x = A + Be^{\alpha x} \ , \ x \geqslant 0,$$

kai parametrai $A \geqslant -B, \alpha > 0, B > 0$ yra randami iš duomenų, apibūdinančių žmogaus gyvenimą. Narys A laikomas nuo amžiaus nepriklausomo, mirtingumą dėl nenatūralių priežasčių lemiančiu parametru. Makeham pasiūlytos versijos atveju pagrindiniai mirtingumo rodikliai yra apskaičiuojami analogiškai:

$$\begin{aligned} S(x) &= e^{-Ax-\frac{B}{\alpha}(e^{\alpha x}-1)}, \, x \geqslant 0, \\ F(x) &= 1 - e^{-Ax-\frac{B}{\alpha}(e^{\alpha x}-1)}, \, x \geqslant 0, \\ f(x) &= (A + Be^{\alpha x})\, e^{-Ax-\frac{B}{\alpha}(e^{\alpha x}-1)}, \, x \geqslant 0, \end{aligned}$$

Galiausiai:

$$q_x = 1 - e^{-A+B(-1+e^{x\alpha})-B(-1+e^{\alpha+x\alpha})}, x \geqslant 0. \tag{2.2}$$

3. *Weibull*[3] *1939 metais pasiūlyta versija.* Žmogaus gyvenimo trukmės mirtingumo galia yra standartiškai griežtai didėjanti su augančiu amžiumi ir aprašoma:

$$\mu_x = cx^n \ , \ x \geqslant 0,$$

kai parametrai $c > 0$ ir $n > 0$ apibūdina žmogaus gyvenimą.

Weibull versijos atveju:

$$\begin{aligned} s(x) &= e^{-\frac{c}{n+1}x^{n+1}}, \, x \geqslant 0, \\ F(x) &= 1 - e^{-\frac{c}{n+1}x^{n+1}}, \, x \geqslant 0, \\ f(x) &= cx^n e^{-\frac{c}{n+1}x^{n+1}}, \, x \geqslant 0. \end{aligned}$$

x-mečio žmogaus tikimybė mirti iki $x+1$-ojo gimtadienio:

$$q_x = 1 - e^{-\frac{c}{n+1}(x+1)^{n+1} + \frac{c}{n+1}x^{n+1}}, x \geqslant 0. \tag{2.3}$$

4. *Perks*[4] *1939 metais pasiūlyta versija.* Žmogaus gyvenimo trukmės mirtingumo galia yra aprašoma ne standartiniu (eksponentinio didėjimo) dėsniu, tačiau logistiniu tipu. Tai reiškia, jog išskirtiniais atvejais, mirtingumo intensyvumas gali sumažėti tam tikrame, senesnių žmonių, amžiaus intervale, o vėliau vėl didėti. Pavyzdys gali būti COVID-19 pandemija, kuomet pensininkai laikosi šalyje įvestos karantino padėties, o vidutinio amžiaus žmonėms neišvengiamai tenka keliauti į darbus. To pasekoje, gali pasireikšti komplikuoti COVID-19 simptomai ir ištikti mirties atvejis.

$$\mu_x = \frac{e^{a+bx}}{1+e^{a+bx}} \ , \ x \geqslant 0,$$

[2] William Matthew Makeham (1826 – 1891) – britų aktuaras.
[3] Ernst Hjalmar Waloddi Weibull (1887 — 1979) – švedų inžinierius ir matematikas.
[4] William Perks – britų aktuaras.

kai parametrai $\alpha > 0, b > 0$ apibūdina žmogaus gyvenimą, turintys identišką prasmę kaip jau prieš tai minėtų mirtingumo dėsnių.

Perks versijos atveju, x-mečio žmogaus tikimybė mirti iki $x+1$-ojo gimtadienio:

$$q_x = 1 - \frac{1}{b} \cdot \ln\left(\frac{1+e^{\alpha+b\cdot(x+1)}}{1+e^{\alpha+b\cdot x}}\right), x \geqslant 0. \tag{2.4}$$

5. *Beard*[5] *1956 metais pasiūlytas dėsnis.* Žmogaus gyvenimo trukmės mirtingumo galia yra aprašoma logistiniu tipu:

$$\mu_x = \frac{\alpha e^{bx}}{\gamma + e^{bx} + 1}\ ,\ x \geqslant 0,$$

   čia $\alpha > 0, b > 0, \gamma > 0$.

   Tada:

$$q_x = 1 - \frac{e^{-\alpha}}{b} \cdot \ln\left(\frac{1+e^{\gamma+\alpha+b\cdot(x+1)}}{1+e^{\gamma+\alpha+b\cdot x}}\right), x \geqslant 0. \tag{2.5}$$

6. *Heligman-Pollard 1980 metais pristatyta mirtingumo taisyklė.* Žmogaus gyvenimo trukmės dėsnis aprašomas mirties ir išgyvenimo tikimybių santykiu:

$$\frac{q_x}{p_x} = A^{(x+B)^C} + D \cdot e^{-E\cdot\left(\ln\left(\frac{x}{F}\right)\right)^2} + GH^2, \quad x \geqslant 0, \tag{2.6}$$

   čia $A, B, C, D, E, F, G, H > 0$, ir $q_x, p_x < 1$.

   Tarkime, kad $\frac{q_x}{p_x} = P$, kai $P \in \mathbb{R}^+$.

   Kadangi $p_x = 1 - q_x$, todėl:

$$q_x = P(1-q_x) = P - Pq_x.$$

   Galiausiai:

$$q_x = \frac{P}{1+P}, \quad x \geqslant 0, \quad P \in \mathbb{R}^+, \quad 0 < q_x < 1.$$

7. *5-ojo laipsnio polinomas.* Išskirtiniais atvejais, mirtingumo intensyvumas gali nebūtinai griežtai didėti. Tiesa, sukonstruoti tokią mirtingumo galios funkciją, kuri būtų universali (tiek eksponentinio, tiek logistinio, tiek mažėjančio mirtingumo intensyvumo atvejais), yra ganėtinai sudėtinga. Mirtingumo rodiklių mažėjimo atvejai, dažniausiai, yra sutinkami jaunesnio amžiaus žmonių atveju *(angl. accidental hump)*. Įprastai, jaunesnio amžiaus žmogaus tikimybė mirti yra pakankamai maža, todėl aproksimuojant duomenis 5-ojo laipsnio polinomu, sunkiai tikėtina viršyti 1. Penktojo laipsnio polinomo funkcija gali prisitaikyti prie atsitiktinio mirtingumo "svyravimo" pakankamai neblogai, tačiau svarbu, jog duomenims funkcija nebūtų per daug pritaikyta *(angl. overfitted)*, todėl modelyje apsiribojama tik ties penktuoju polinomo laipsniu:

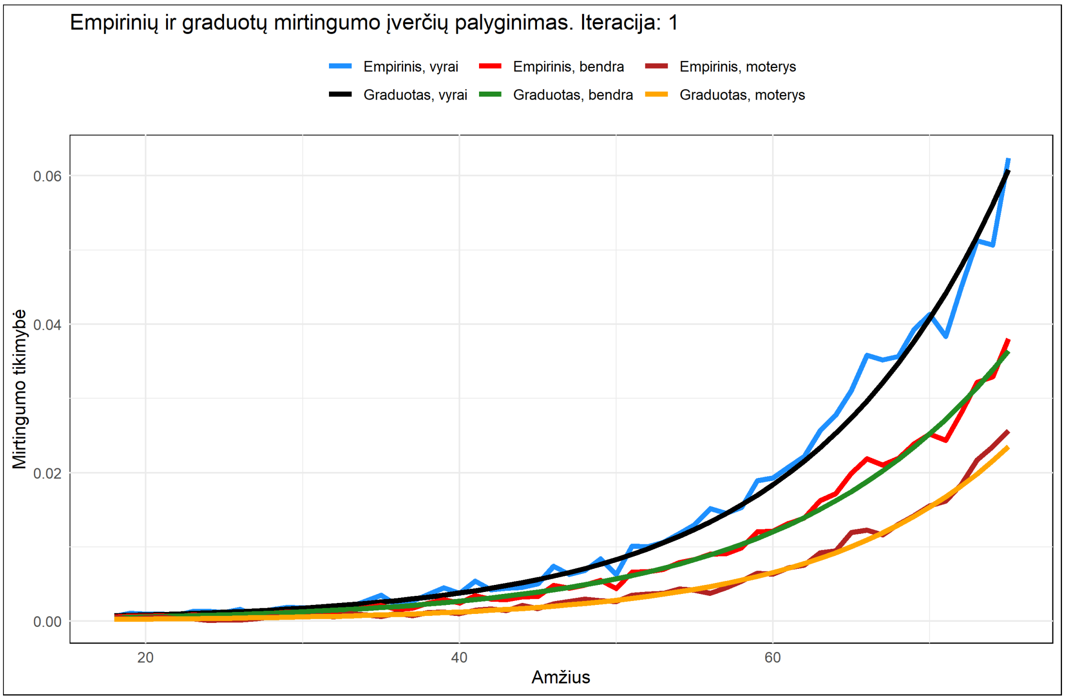
$$q_x \approx a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3 + a_4 \cdot x^4 + a_5 \cdot x^5, \tag{2.7}$$

   čia $x \geqslant 0,\ a_0, ..., a_5 \in \mathbb{R}$

Svarbu paminėti, jog atliekant modelio simuliavimą, visos aukščiau minėtos mirtingumo taisyklės, kiekvienos iteracijos metu, turi vienodą tikimybę būti atrinktos, išskyrus polinominę aproksimaciją. Interpoliavimas polinomu yra naudojamas visais atvejais, kai taikomų mirtingumo dėsnių skaičius yra 2, t.y polinomo taikymo tikimybė iteracijos metu lygi $\frac{1}{2}$.

Kiekviena mirtingumo taisyklė yra unikali savo matematinės funkcijos forma, parametrų skaičiumi ir jų interpretacija. Pasirinktai matematinei funkcijai reikalinga nustatyti tokius parametrus, kurie maksimizuotų taisyklės tikslumą, analizuojamam duomenų masyvui. Kadangi minėtos funkcijos turi daugiau nei vieną parametrą, jų nustatymui taikomas optimizavimo **BFGS gradiento nuolydžio metodas**, aprašomas 2.4 skyrelyje. Išimtis yra 5-ojo laipsnio polinomas, kurio atveju, parametrų nustatymui taikysime **mažiausių kvadratų metodą**, aprašomą 2.5 skiltyje.

[5] James Beard – JAV aktuaras.



4 pav.: Makeham (juoda ir žalia linijos) bei Perks (geltona linija) dėsnių taikymas 2019-ųjų metų Lietuvos populiacijos duomenims.

2.4 BFGS gradiento nuolydžio optimizavimo metodas

BFGS (Broyden-Fletcher-Goldfarb-Shanno) metodas yra iteracinis optimizavimo algoritmas, naudojamas netiesiniam optimizavimo uždaviniui spręsti. Jis priklauso kvazi-Neutono[6] metodų šeimai ir yra sukurtas ieškoti optimalių parametrų, kurie minimizuotų tikslo funkcijos reikšmę. Verta paminėti, jog mirtingumo analizės modelis, kiekvienos išorinės (paties modelio) iteracijos metu atlieka ir 1000 vidinių (BFGS metode) iteracijų ieškant optimaliausio parametrų rinkinio sprendinio.

Mirtingumo analizėje, modelis BFGS metodą pritaiko minimizuojant mirties rodiklių paklaidų (kiekviename amžiaus taške) sumą, kurios pagrindas yra **Chi-Kvadrato** *(angl. Chi-Squared)* funkcijos reikšmės. Tarkime, jog turime mirtingumo duomenis, kuriuos atstoja mirčių skaičiaus $\mathbf{D} = (\mathbf{d}_{\mathbf{x}_{\mathbf{min}}}, ..., \mathbf{d}_{\mathbf{x}_{\mathbf{max}}})$ ir nugyventų metų skaičiaus $\mathbf{E} = (\mathbf{e}_{\mathbf{x}_{\mathbf{min}}}, ..., \mathbf{e}_{\mathbf{x}_{\mathbf{max}}})$ vektoriai, bei mirtingumo dėsnio funkciją $\mathbf{q}_{\mathbf{x}}(\mathbf{m})$, kur $\mathbf{m}$ yra parametrų vektorius.

Tikslas yra minimizuoti Chi-Kvadrato funkcijos reikšmių svertinę (pagal kiekviename amžiaus taške nugyventų metų skaičių) sumą, apibrėžtą kaip:

$$\chi^2(\mathbf{m}) = \sum_{x=x_{min}}^{x_{max}} \frac{(D_x - E_x q_x(m))^2}{E_x q_x(m)(1 - q_x(m))},$$

kur:
- x_{min} yra minimalus analizuojamas žmogaus amžius,
- x_{max} yra maksimalus analizuojamas žmogaus amžius,
- D_x yra x-mečių mirčių skaičius,
- $q_x(m)$ yra įvertintas (BFGS metodo iteracijos metu) mirtumo rodiklis, taikant m parametrų rinkinį,
- E_x x-mečių nugyventų metų skaičius.

BFGS metodas iteratyviai atnaujina parametrų vektorių $\mathbf{m}$, siekdamas minimizuoti šią tikslo funkciją. Parametrų vektoriaus atnaujinimo taisyklė $k+1$-ajai iteracijai yra tokia:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{H}_k \nabla \chi^2(\mathbf{x}_k)$$

kur:
- α_k yra parenkama naudojant linijinio posūkio *(angl. linear search)* metodą,

[6]Isaac Newton (1643 – 1727) – anglų fizikas, matematikas ir filosofas.

- $\nabla\chi^2(\mathbf{x}_k)$ yra tikslo funkcijos gradientas pagal $\mathbf{x}$,
- H_k[7] yra atvirkštinė Heso matricos *(angl. Hessian Matrix)* aproksimacija k-tame žingsnyje.

Tikslo funkcijos gradientą $\nabla\chi^2(\mathbf{x})$ pagal parametrą m_i galima skaičiuoti kaip:

$$\frac{\partial\chi^2(\mathbf{m})}{\partial m_i} = -2\sum_{x=x_{\min}}^{x_{\max}} \frac{(D_x - E_x q_x(\mathbf{m}))\cdot\frac{\partial}{\partial m_i}(E_x q_x(\mathbf{m}))}{E_x q_x(\mathbf{m})(1-q_x(\mathbf{m}))}$$

Kaip jau minėta anksčiau, funkciją $q_x(m)$ galima pasirinkti priklausomai nuo konkretaus mirtingumo dėsnio, aprašyto 2.3 skyrelyje. Bendrai, BFGS metodui svarbu parinkti pradinį vektorių rinkinį $\mathbf{m}$, įvertinti tikslo funkcijos $\chi^2(\mathbf{x}_k)$ reikšmę ir apskaičiuoti gradientą $\nabla\chi^2(\mathbf{x}_k)$, kuris kitame x_{k+1} žingsnyje (iteracijoje) labiausiai sumažintų χ^2 reikšmę. Algoritmas kartojamas tol, kol bus aptinktas χ^2 funkcijos konvergavimas arba įvykdytas apibrėžtas maksimalus vidinių iteracijų skaičius, šiuo atveju - 1000.

Pastaba. Grupuotiems duomenims, pagal amžiaus intervalus, pirmiausiai, yra apskaičiuojami viduriniai amžiaus intervalo taškai. Šiems amžiaus taškams yra priskiriami to intervalo empiriniai mirtingumo rodikliai ir tik tuomet yra atliekamas gradavimo procesas. Pavyzdžiui, jeigu turime amžiaus grupių vektorių $Z = (0, 26, 27, 60, 61, 105)$*, tuomet vidurniai intervalų taškai yra atitinkamai 13, 44.5 ir 88. Galiausiai, šiems amžių taškams bus pritaikyta atitinkamo dėsnio kreivė ir atliekamas parametrų kalibravimas BFGS metodu.*

2.5 Mažiausių kvadratų metodas

Mažiausių kvardatų metodas, MKM *(angl. Least Squares Method)* yra taikomas tik mirtingumo rodiklių aproksimavimu 5-ojo laipsnio polinomu. Metodas yra skirtas būtent išskirtiniams atvejams, kai pasireiškia mirtingumo suintensyvėjimas jaunesnio amžiaus žmonių grupėje dėl nenatūralių priežasčių (avarijos ir kita, padidintos rizikos veikla).
Bendru atveju, pažymėkime 5-ojo laipsnio polinomas regresinį modelį:

$$Q = a_0 + a_1X + a_2X^2 + a_3X^3 + a_4X^4 + a_5X^5 + \varepsilon,$$

kur:

- Q - mirčių rodiklių q_x stulpelis,
- X - amžiaus kintamojo matrica,
- $\mathrm{a}_0, a_1, a_2, a_3, a_4, a_5$ - nežinomi modelio koeficientai,
- ε - modelio paklaidos.

Tikslas yra minimizuoti kvadratinių skirtumų sumą tarp empirinių (q_x) ir modeliuojamų reikšmių ($\hat{q}_x$):

$$S = \sum_{x=x_min}^{x_max} (q_x - \hat{q}_x)^2$$

Tuomet, mažiausių kvadratų metodo sprendinys, minimizuojantis S:

$$\alpha = (X^TX)^{-1}X^TQ,$$

$$\begin{aligned}
\alpha &= [a_0, a_1, a_2, a_3, a_4, a_5]^T, \\
X &= \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & x_m^3 & x_m^4 & x_m^5 \end{bmatrix}, \\
Q &= [q_1, q_2, \ldots, q_m]^T.
\end{aligned}$$

Galutinis rezultatas yra α parametrai, vėliau naudojami (2.7) formulėje. Iš tiesų, polinomo parametrų nustatymui gali būti naudojamas anksčiau aprašytas BFGS metodas, atsižvelgiantis į nugyventų metų skaičių kiekviename amžiaus taške. Tačiau, dėl MKM paprastumo ir mažiau reikalaujamų skaičiavimo resursų, būtent šis metodas yra naudojamas jaunesnio amžiaus žmonių modeliavimui.

[7] Heso matrica - kvadratinė matrica, sudaryta iš funkcijos antrosios eilės dalinių išvestinių.

2.6 χ^2 statistinis hipotezių testas

Atlikus gradavimo procesą, atsiranda subjektyvumas, kada įvertinti mirtingumo rodikliai yra pakankamai tikslūs ir atitinka duomenis? Vienas iš dažniausiai naudojamų metodų objektyviam gradavimo proceso vertinimui - statistinis χ^2 hipotezių testas. Pagrindinis tikslas - patikrinti, ar nustatytas mirčių rodiklis kiekviename amžiuje atitinka naudojamą gradavimo formulę. Vėl gi, "statistinis testas" reiškia, jog didžiausia įtaka, kuri hipotezė bus patvirtinta, priklauso nuo amžiaus, kuriame portfelis yra labiausiai koncentruotas. Formali testo prielaida - mirčių skaičius kiekvienoje amžiaus grupėje yra nepriklausomas. Taip pat, galutinis testo rezultatas priklauso nuo pasirinkto pasikliovimo lygmens.

Pirmame testavimo žingsnyje yra apibrėžiamos hipotezės:

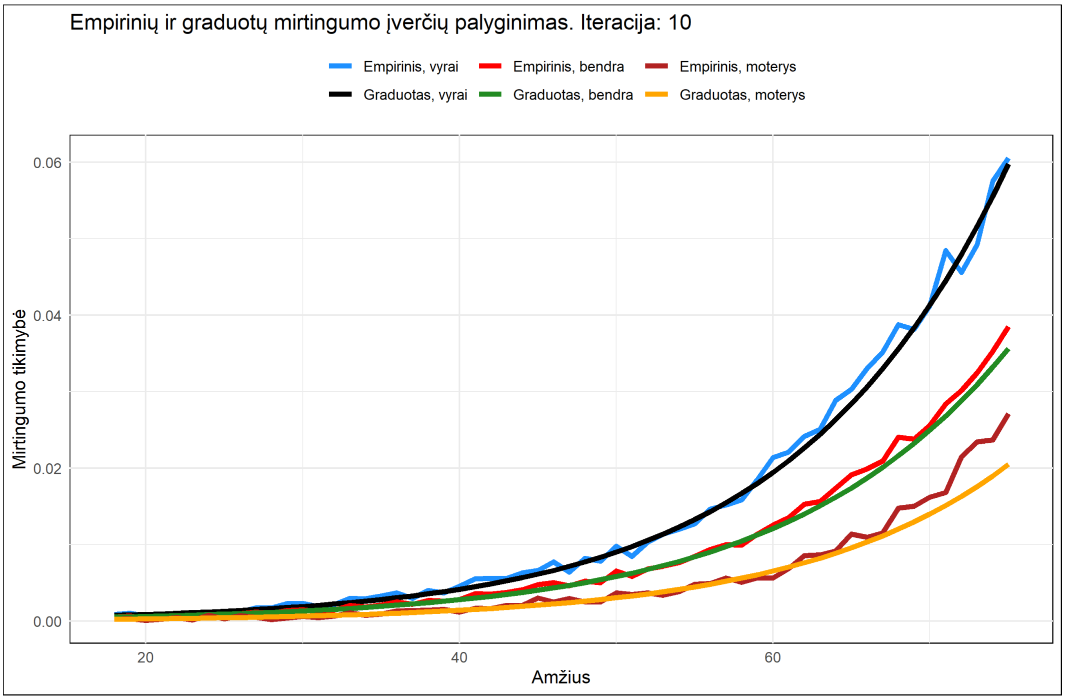
1. ***Nulinė Hipotezė*** H_0***:*** *stebimi mirtingumo rodikliai atitinka tikėtinus mirtingumo rodiklius.*
2. ***Alternatyvioji Hipotezė*** H_1***:*** *stebimi mirtingumo rodikliai neatitinka tikėtinų mirtingumo rodiklių.*

Antrame žingsnyje skaičiuojama Chi-kvadrato testo statistika (χ^2) pagal formulę:

$$\chi^2 = \sum_{x=x_{min}}^{x_{max}} \frac{(D_x - E_x\hat{q}_x)^2}{E_x\hat{q}_x(1-\hat{q}_x)}, \tag{2.8}$$

čia: D_x atitinka realų x-mečių mirčių skaičių duomenyse, $E_x\hat{q}_x$ - tikėtiną x-mečių mirčių skaičių, $\hat{q}_x$ - mirties tikimybės įvertį x amžiaus žmogui.
Trečiame žingsnyje yra atliekamas sprendimas, kurią hipotezę χ^2 tenkina:

1. *Pasirenkamas reikšmingumo lygmuo (angl. significance level):* $\alpha = 0.05$, $0 < \alpha < 1$
2. *Apibrėžiamas laisvės laipsnių (angl. degrees of freedom):* $df = k - 1$. *Čia* k *atitinka amžiaus grupių skaičių.*
3. *Lyginamos (2.8) formulėje gauta* χ^2 *ir Chi-Kvadrato skirstinio kritinė* $\chi_{0.95}$ *reikšmės: jei* $\chi^2 > \chi_{0.95}$ *- nulinė hipotezė atmetama. Priešingu atveju - testo rezultatas statistiškai patenkinamas.*



5 pav.: Mirtingumo gradavimo Lietuvos populiacijos duomenims pavyzdys. Vyrų atveju, χ^2 testo hipotezė H_0 - priimta. Moterų atveju - H_0 atmesta.

2.7 Trečiosios eilės mirties tikimybių skirtumai ir kreivės glotnumo testas

Mirtingumo kreivėms būdinga glotnumo savybė, reiškianti, jog funkcijai nėra priimtina turėti lūžio "kampuotus" taškus. Praktikoje, pastarasis atvejis galimas kombinuojant polinomo ir kito mirtingumo dėsnio funkcijų reikšmes tam tikro amžiaus taško aplinkoje.

Pažymėkime mirties tikimybės q_x pirmosios, antrosios ir trečiosios eilių skirtumus:

1. **Pirmosios eilės skirtumas (Δq_x):**
$$\Delta q_x = q_{x+1} - q_x$$

2. **Antrosios eilės skirtumas ($\Delta^2 q_x$):**
$$\Delta^2 q_x = \Delta(\Delta q_x) = \Delta(q_{x+1} - q_x) = \Delta q_{x+1} - \Delta q_x$$

3. **Trečiosios eilės skirtumas ($\Delta^3 q_x$):**
$$\Delta^3 q_x = \Delta(\Delta^2 q_x) = \Delta(\Delta q_{x+1} - \Delta q_x) = \Delta^2 q_{x+1} - \Delta^2 q_x$$

Tuomet yra skaičiuojamas T statistikos dydis:

$$T = \sum_i |\Delta^3 q(x_i)| \tag{2.9}$$

Pavyzdys:

Amžius	Mirties tikimybė
30	0.0020
31	0.0022
32	0.0027
33	0.0037
34	0.0048
35	0.0065
36	0.0091

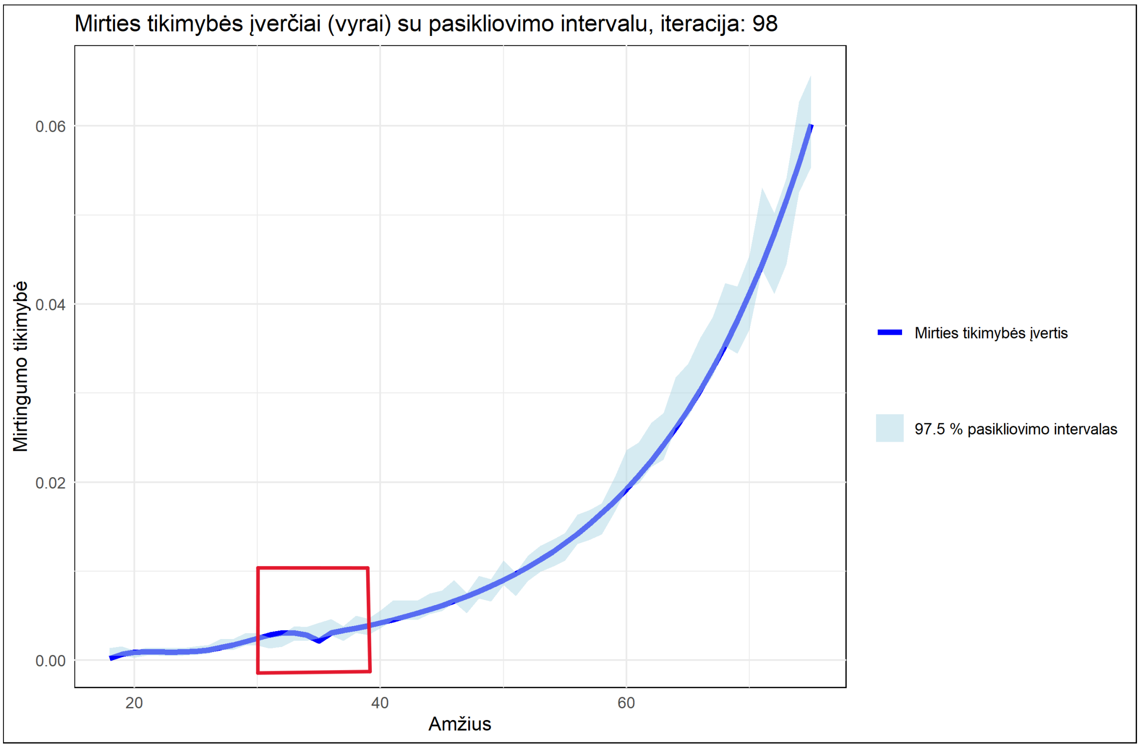
2 lentelė: pavyzdinė mirties tikimybių lentelė.

Apskaičiuoti skirtumai:

Amžius	Δq_x	$\Delta^2 q_x$	$\Delta^3 q_x$
30	0.0002	0.0003	0.0002
31	0.0005	0.0005	-0.0004
32	0.0010	0.0001	0.0005
33	0.0011	0.0006	0.0003
34	0.0017	0.0009	-
35	0.0026	-	-

3 lentelė: mirties tikimybių trečiosios eilės skirtumai.

Iš tiesų, modelyje, trečiosios eilės skirtumai naudojami tik kaip antrinis testas. Nekorektiškai sukombinuotų polinomo ir kito dėsnio kreivių reikšmės, greičiausiai, jog nepatenkintų ir χ^2 hipotezių testo. Tai reiškia, jog gauta T reikšmė gali turėti įtakos tik galutiniame aktuaro ekspertiniame vertinime. Svarbu paminėti, jog algoritme, T statistika yra skaičiuojama dviejų mirtingumo taisyklių taško aplinkoje $[x-5, x+5]$. Todėl, kuo modelyje gauta T reikšmė mažesnė, tuo aktuarams rezultatas yra palankesnis.



6 pav.: Pavyzdys. Polinomo ir Gompertz dėsnio funkcijų kombinavimas. Kreivės glotnumo stoka.

2.8 Mirtingumo lentelės interpoliavimas žmogaus trupmeniniame amžiuje

Žmogaus gyvenimo trukmė - nebūtinai sveikasis skaičius. Statistiniai duomenys (ℓ_x, q_x, d_x, e_x) mirtingumo lentelėse pateikiami tik sveikosioms x reikšmėms. Tačiau, gyvybės draudimo versle, svarbu rasti minėtųjų funkcijų tarpines reikšmes, kadangi draudimo įmokų mokėjimai ir rezervų apskaita įvyksta mėnesiniu periodiškumu. Tam tikslui - reikalingas papildomas mirtingumo lentelės interpoliavimas. Tiek praktikoje, tiek A. Grigučio ir J. Šiaulio mokslinėje publikacijoje "aktuarinė matematika", išskiriami trys pagrindiniai interpoliavimo būdai, kurie remiasi tam tikromis prielaidomis apie $S(x)$ ir $l(x)$ (kohortoje esančių x amžiaus žmonių skaičius) pavidalus, kai x nėra sveikasis skaičius. Dažniausiai sutinkamos yra trys žemiau išvardintos prielaidos:

1. Tolygusis mirtingumo pasiskirstymas. Dažniausiai praktikoje sutinkama prielaida. Šiuo atveju, daroma prielada, jog žmonės vidutiniškai miršta metų viduryje. Tarkime, kad išgyvenimo funkcija $S(x)$ tiesinė bet kuriame intervale $[x, x+1]$, kai $x = 0, 1, \dots,$ t. y.,

$$S(x+t) = (1-t)\,S(x) + t\,S(x+1), \;\; x \in \mathbb{N}_0, \, t \in [0,\, 1].$$

Kitos, tolygaus mirtingumo prielaidos charakteristikos:

$$\begin{aligned} {}_tq_x &= t\,q_x, \\ {}_tp_x &= 1 - t\,q_x, \\ {}_yq_{x+t} &= \frac{y\,q_x}{1 - t\,q_x}, \\ \mu_{x+t} &= \frac{q_x}{1 - t\,q_x}, \\ f_x(t) &= {}_tp_x\,\mu_{x+t} = q_x, \end{aligned}$$

visiems $x \in \mathbb{N}_0, \;\; t \in [0,\, 1], \;\; y \in [0,\, 1], \;\; y + t \leqslant 1.$

2. Mirtingumo galia pastovi. Rečiausiai praktikoje sutinkama prielaida. Tarkime, kad mirtingumo galia $\mu_x = \lambda$, kai λ yra teigiama konstanta, o $x \geqslant 0$. Tuomet

$$\mu_x = -\frac{s'(x)}{s(x)} = \lambda$$

ir, kai $t \geqslant 0$,

$$-\int_x^{x+t} \frac{s'(u)}{s(u)}\,du = \int_x^{x+\lambda} \lambda\,du \quad \Rightarrow \quad \ln s(u)\Big|_x^{x+t} = -\lambda t \quad \Rightarrow \quad \frac{s(x+t)}{s(x)} = e^{-\lambda t}. \tag{2.10}$$

Iš lygybės (2.10) seka, jog

$$ {}_tp_x = \frac{s(x+t)}{s(x)} = e^{-\lambda t}, $$

t. y. tikimybė, jog $x \geqslant 0$ amžiaus asmuo gyvens dar bent t metų, nepriklauso nuo asmens amžiaus x.

Kitos, pastovios mirtingumo galios prielaidos charakteristikos:

$$ \begin{aligned} {}_tq_x &= 1 - e^{-\mu_x t}, \\ {}_tp_x &= e^{-\mu_x t}, \\ {}_yq_{x+t} &= 1 - e^{-\mu_x y}, \\ \mu_{x+t} &= \mu_x, \\ f_x(t) &= {}_tp_x\, \mu_{x+t} = e^{-\mu_x t}\, \mu_x, \end{aligned} $$

čia $\mu_x = \lambda > 0$, o $x, \geqslant 0.$, bei $t \in [0,\,1],\ y \in [0,\,1],\ y + t \leqslant 1.$

3. Balducci[8] mirtingumo prielaida. Sakykime, kad išgyvenimo funkcija $S(x)$ tenkina savybę:

$$ \frac{1}{S(x+t)} = \frac{1-t}{S(x)} + \frac{t}{S(x+1)}, $$

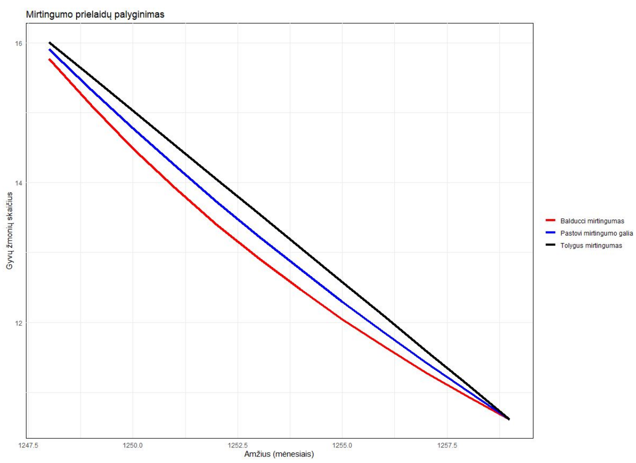
kai $t \in [0,\,1]$ ir $x \in \mathbb{N}_0$.

Pagrindinės Balducci mirtingumo prielaidos charakteristikos:

$$ \begin{aligned} {}_tq_x &= \frac{tq_x}{1-(1-t)q_x}, \\ {}_tp_x &= \frac{p_x}{1-(1-t)q_x}, \\ {}_yq_{x+t} &= \frac{yq_x}{1-(1-y-t)q_x}, \\ \mu_{x+t} &= \frac{q_x}{1-(1-t)q_x}, \\ f_x(t) = {}_tp_x\, \mu_{x+t} &= \frac{p_x\, q_x}{(1-(1-t)q_x)^2}, \end{aligned} $$

visiems $x \in \mathbb{N}_0$, $t \in [0,\,1]$, $y \in [0,\,1]$ ir $y + t \leqslant 1$.

Iš tiesų, Balducci prielaida, praktikoje gali būti pritaikoma vis dažniau. Pavyzdys - COVID-19 pandemija. Tarkime, jog aktuaras nori patikrinti, kaip skiriasi mirtingumas skirtingoms portfelio grupėms. Kadangi COVID-19 pandemijos metas yra žiema, specialistas Balducci prielaidą taiko tokiai asmenų grupei, kurių gimtadienis yra ankstyvą pavasarį. Balducci prielaidos esmė - staigiau padidėjęs mirtingumo intensyvumas (lyginant su tolygaus pasiskirstymo prielaida) prieš ateinantį žmogaus gimtadienį. Šiuo atveju, vėlyvo pavasario metu gimę žmonės, gali susidurti su staigiu mirtingumo paintensyvėjimu prieš artėjantį savo gimtadienį, nes COVID-19 mirtingumas išauga būtent žiemos metu, kai iki sekančio gimtadienio lieka nedaug laiko.



7 pav.: Pavyzdys. Skirtingi mirtingumo lentelės interpoliavimo metodai Lietuvos populiacijos duomenims. Juoda linija - tolygus pasiskirstymas, mėlyna - pastovios mirtingumo galios, o raudona - Balducci.

[8] Gaetano Balducci (XIX–XX a.) – italų aktuaras.

2.9 Mirčių skaičiaus pasiskirstymas ir kvantiliai

Gyvybės draudime, mirčių skirstinio nustatymas yra žingsnis, svarbus kainodaros procese. Skirstinio nustatymas suteikia galimybę apskaičiuoti draudimo įmokų atitinkamus kvantilius. Simboliu D_x pažymėkime x amžiaus žmonių mirčių skaičių tam tikrame periode. Jei mirčių skaičius D_x yra nepriklausomas atsitiktinis dydis, darysime prielaidą, jog mirtys yra pasiskirsčiusios pagal Puasono[9] dėsnį:

$$D_x \sim Poisson(E_x \hat{q}_x),$$

čia E_x - bendras nugyventų metų skaičius tokių asmenų, kuriems būtent x metų, $\hat{q}_x$ - empirinis mirties tikimybės įvertis amžiuje x.

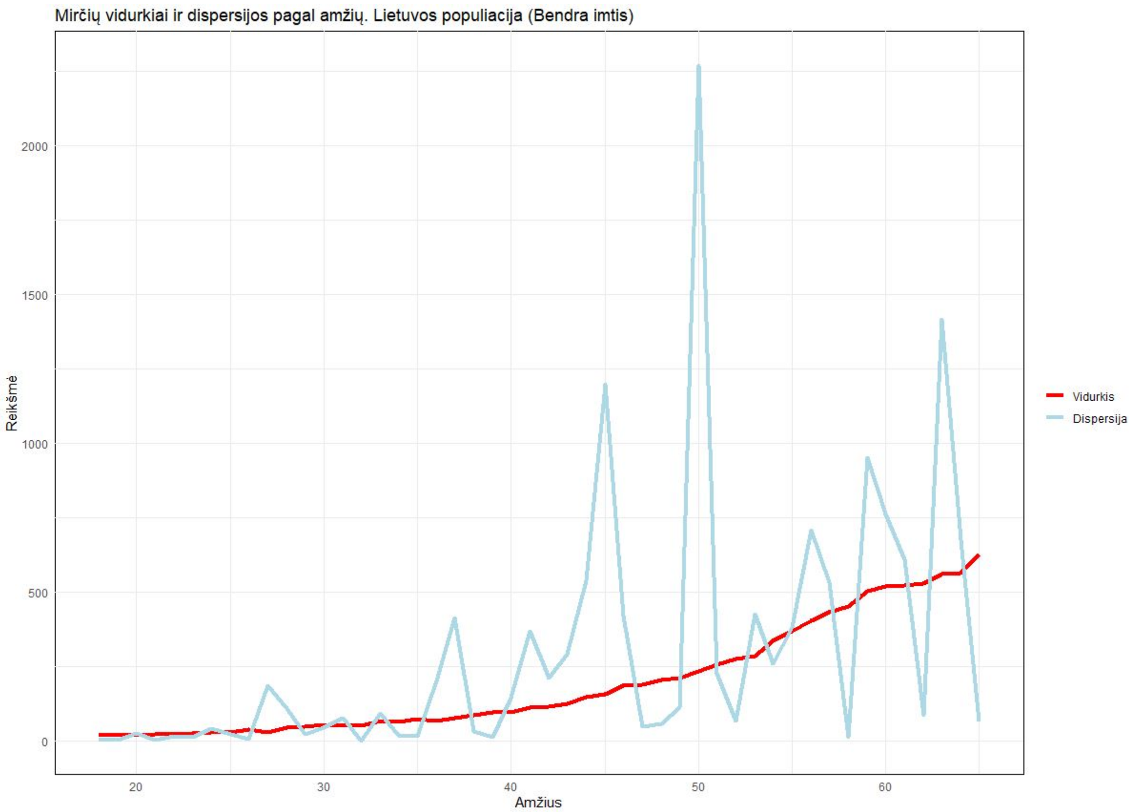
Tuomet, mirčių skaičiaus, amžiuje x, tikimybinio tankio funkcija:

$$P(D_x = k) = e^{-E_x \hat{q}_x} \frac{(E_x \hat{q}_x)^k}{k!},$$

Puasono mirčių skaičiaus, amžiuje x, vidurkis ir dispersija:

$$E[D_x] = E_x \hat{q}_x = Var[D_x] \tag{2.11}$$

Būtent (2.11) Puasono skirstinio vidurkio ir dispersijos savybė yra svarbiausia duomenų tikrinimui. Jei duomenys tenkina šią savybę, tuomet galima taikyti Puasono pasiskirstymo prielaidą mirčių skaičiaus kvantiliams nustatyti. Žemiau, paveikslelyje, pateikiamas pavyzdys, kai Lietuvos populiacijos duomenyse, dispersija svyruoja šalia vidurkio. Tokiu atveju, Puasono pasiskirstymo prielaida yra vizualiai priimtina.



8 pav.: Lietuvos populiacijos (2017-2019) duomenys. Mirčių skaičiaus vidurkis ir dispersija.

Pritaikęs Puasono skirstinio prielaidą mirčių skaičiui, modelis apskaičiuoja 2.5% ir 97.5% draudimo įmokos dydžių (kiekviename analizuojamo amžiaus x taške) kvantilius. Tokie įmokų kvantilių skaičiavimai yra palyginamieji, jog aktuarui būtų lengviau pajausti prielaidų įtaką verslui ir atlikti skaitinį palyginimą tarp statistiškai geriausiai modeliuojamų įverčių.

[9]Simeon Denis Poisson (1781 – 1840) - prancūzų matematikas ir fizikas.

2.10 Klasikinio Gyvybės Draudimo įmokų skaičiavimai

Mirtingumo rodiklių vertinimas yra tiesiogiai susijęs su gyvybės draudimo produktų kainodara. Įmokų skaičiavimo prasmė modelyje - atspindėti aktuaro daromų prielaidų jautrumą įmokų dydžiui. Sakykime, jog gyvybės draudimo kompanija prekiauja klasikiniu gyvybės draudimu. Amžiaus x_{min} klientas moka mėnesinę $P_{x_{min}}$ eurų dydžio įmoką, įsipareigodamas išmokėti S eurų dydžio išmoką tik tuo atveju, jei mirė apdraustasis asmuo ir įvykis yra pripažintas draudžiamuoju. Įvertinus metinius mirtingumo įverčius, padarius prielaidą apie žmogaus gyvenimo trukmę trupmeniniame amžiuje, galima apskaičiuoti draudimo įmokas. Skaičiavimams taikomas būsimų pinigų srautų dabartinės vertės skaičiavimo principas - aktuarinė lygybė.

Apibrėžkime aktuarinę lygybę:

$$P_{x_{min}} + P_{x_{min}} \cdot \left({}_{\frac{1}{12}}p_x \cdot v^{1/12} \right) + P_{x_{min}} \cdot \left({}_{\frac{2}{12}}p_x \cdot v^{2/12} \right) + \ldots + P_{x_{min}} \cdot \left({}_{x_{max}-x-\frac{1}{12}}p_x \cdot v^{x_{max}-x-\frac{1}{12}} \right) =$$
$$= S \cdot \left({}_{\frac{1}{12}}q_x \cdot v^{1/12} \right) + S \cdot \left({}_{\frac{1}{12}}q_{x+\frac{1}{12}} \cdot {}_{\frac{1}{12}}\, p_x \cdot v^{2/12} \right) + \ldots + S \cdot \left({}_{\frac{1}{12}}q_{x_{max}-\frac{1}{12}} \cdot {}_{x_{max}-x-\frac{1}{12}}\, p_x \cdot v^{x_{max}-x-\frac{1}{12}} \right),$$

čia $v = \frac{1}{1+i}$, $i-$ nerizikinga palūkanų norma, $n \in \mathbb{N}$, ${}_{\frac{12n+11}{12}}p_x-$ tikimybė, jog x amžiaus žmogus bus taip pat gyvas po $\frac{12n+11}{12}$ metų, o ${}_{\frac{1}{12}}q_{x+n+\frac{11}{12}}-$ tikimybė, jog $x + n + \frac{11}{12}$ amžiaus žmogus nesulauks savo $x + n + 1$ gimtadienio.

Bendru atveju, aktuarinė lygybė:

$$P_{x_{min}} \left(\sum_{x=x_{\min}}^{x_{\max}-1} \sum_{s=1}^{12} {}_{\frac{12(x-x_{\min})+s}{12}}p_x \cdot v^{\frac{12(x-x_{\min})+s}{12}} \right) =$$
$$= S \left(\sum_{x=x_{\min}}^{x_{\max}-1} \sum_{s=1}^{12} {}_{\frac{1}{12}}q_{x+\frac{12(x-x_{\min})+s}{12}} \cdot {}_{\frac{12(x-x_{\min})+s}{12}}p_x \cdot v^{\frac{12(x-x_{\min})+s}{12}} \right).$$

Galiausiai, įmokos dydis $P_{x_{min}}$:

$$P_{x_{min}} = \frac{S \left(\sum_{x=x_{\min}}^{x_{\max}-1} \sum_{s=1}^{12} {}_{\frac{1}{12}}q_{x+\frac{12(x-x_{\min})+s}{12}} \cdot {}_{\frac{12(x-x_{\min})+s}{12}}p_x \cdot v^{\frac{12(x-x_{\min})+s}{12}} \right)}{\left(\sum_{x=x_{\min}}^{x_{\max}-1} \sum_{s=1}^{12} {}_{\frac{12(x-x_{\min})+s}{12}}p_x \cdot v^{\frac{12(x-x_{\min})+s}{12}} \right)}. \tag{2.12}$$

3. Praktinė dalis

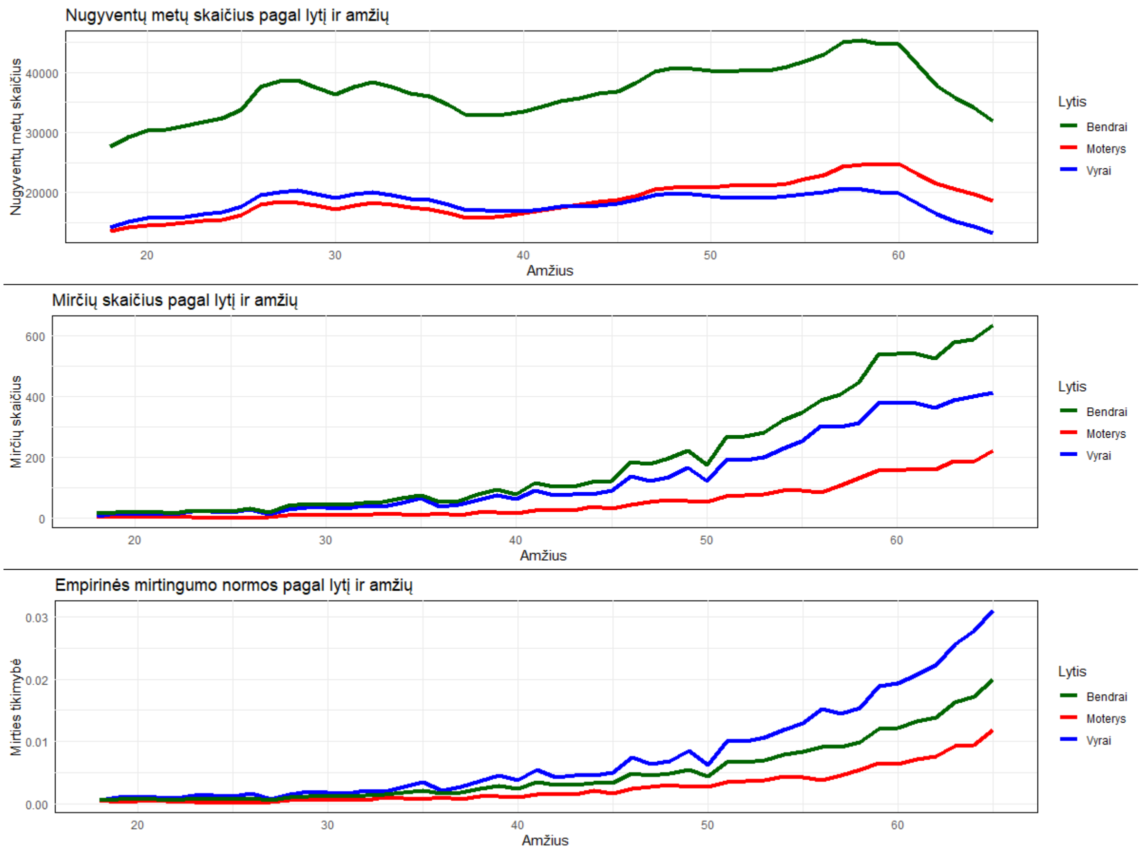
Kadangi realybę atitinkantys ir pagrindines praktines problemas atspindintys, gyvybės draudimo kompnaijos duomenys nėra viešai prieinami, praktinėje dalyje pritaikysime anksčiau aprašytą mirtingumo analizės algoritmą realiems Lietuvos mirtingumo duomenims. Atliksime po 1000 iteracijų vyrų, moterų bei bendros Lietuvos populiacijos duomenims. Šio tyrimo tikslas yra nustatyti tiksliausius mirčių rodiklių įverčius, bei iliustruoti teorinėje dalyje minėtų, aktuarinių prielaidų svarbą ir jautrumą gyvybės draudimo versle, ypatingai - klasikinio gyvybės draudimo atveju, kai mokamos įmokos dydis yra pastovus. Pagrindinė tyrimo prielaida - teisingai sugrupuoti, pagal amžiaus grupes, mirtingumo duomenys yra labiau patikimi nei negrupuoti (pagal amžiaus grupes) duomenys. Be to, duomenų skaidymas į ilgesnių (nei vieneri metai) intervalų amžiaus grupes reikalauja trumpesnės laiko eilutės, ko pasekoje mirtingumo įverčiai atspindi naujausias tendencijas. Galutinis rezultatas - mirtingumo rodiklių, įmokų dydžių ir lauktinosios gyvenimo trukmės (kiekviename amžiaus taške) palyginimas geriausiųjų ir blogiausiųjų, modelio atrinktų, scenarijų atvejais. Papildomai, palyginsime skirtingų mirtingumo dėsnių įvertintas kreives bendrai Lietuvos populiacijai. Tuo pačiu, pateiksime asmenines įžvalgas, kodėl, praktiškai, tokia mirtingumo kreivė yra geriausias įvertis atitinkamam atvejui.

Iš tiesų, šalies populiacijos analizės atveju, tokie statiniai parametrai, kaip "maksimalus galimas laiko eilutės ilgis" ar "išskirtiniai metai, neįtraukti į analizę", neturi daug reikšmės. Įprastai, dėl pakankamai didelio duomenų kiekio, šalies mirtingumo įvertinimui pakanka vienerių metų. Teoriškai, statinių parametrų svarba turėtų būti kur kas svarbesnė analizuojant gyvybės draudimo portfelį, todėl parametrai yra įtraukti, su idėja sukurti kuo universalesnį modelį. darbo tikslas yra įvairiapusiškai iliustruoti gyvybės draudimo kompleksiškumą, todėl pasirinkta laiko eilutė yra ilgesnė, nei vieneri metai ir tuo pačiu patikrinsime, kaip skiriasi draudimo įmokų dydžiai, naudojant skirtingų periodų laiko eilutes.

3.1 Duomenys

Realiai mirčių statistikai naudosime HMD, Human Mortality Database *(www.mortality.org)* viešai prieinamus mirtingumo duomenis. Duomenų bazėje yra prieinama tiek mirčių skaičiaus D_x, tiek bendrų nugyventų metų skaičiaus E_x statistikos. Bendras nugyventų metų skaičius yra nebūtinai sveikasis skaičius, reiškiantis, jog mirusiųjų žmonių dalinė įtaka šiam dydžiui - įvertinta. Naujausi, prieinami duomenys Lietuvai - 2020-ųjų metų. Kadangi duomenys yra populiacijos mąsto, reiškiantys, jog čia atsitiktinumo yra ne tiek daug, kiek gyvybės draudimo kompanijų portfeliuose, todėl apsiribosime tik ties trijų metų laiko eilutės ilgiu. Dėl tyrimo korektiškumo priežasties, 2020-ųjų metų nenagrinėsime, kadangi šis laikotarpis yra COVID-19 pandemijos pradžia, kuri gali iškreipti analizės rezultatus. Todėl tyrime fokusuojamasi į 2017-2019-ųjų metų periodo mirtingumo situaciją. Nagrinėsime tiek vyrų, tiek moterų, bei bendros populiacijos atvejus.

Lietuvos populiacija yra labiausiai koncentruota tiek 30-ies ir 60-ies amžių taškais, t.y bendras nugyventų metų skaičius šiuose taškuose - didžiausias visais atvejais. Tai reiškia, jog šie amžiaus taškai turės didžiausią įtaką χ^2 statistikos ir tam tikro mirtingumo dėsnio parametrų reikšmėms. Kaip įprasta, analizuojamuose duomenyse, vyrų mirtingumas yra didesnis, lyginant su moterimis. Žemiau yra pateikiami, 2017-2019 kalendorinių metų periodą reprezentuojantys, grafikai. Žalia spalva atitinka bendrą Lietuvos populiaciją (neskirtstant pagal lytis), mėlyna - vyrų, o raudona - moterų populiaciją:



9 pav.: Lietuvos populiacijos mirtingumas. 2017-2019-ųjų metų periodo duomenys.

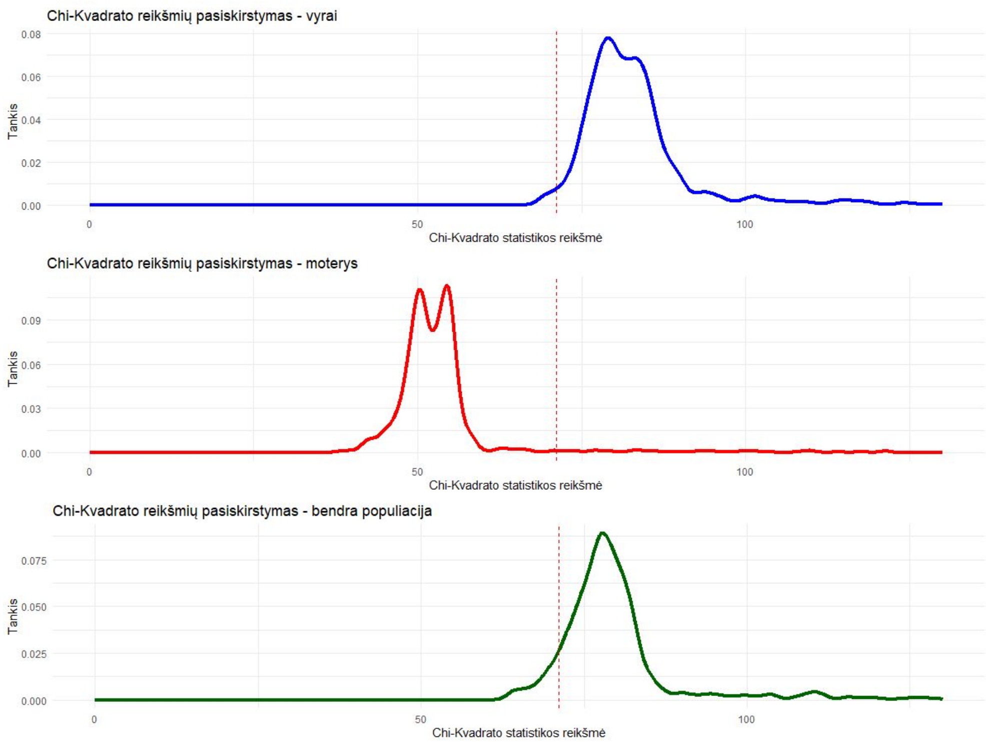
3.2 Bendra modelio simulacijos rezultatų apžvalga

Prieš modelio simuliacijų paleidimą, vartotojas privalo įvesti tokius statinius parametrus, kaip maksimalus laiko eilutės ilgis, iteracijų skaičius, analizės periodo pabaiga ir t.t. Statiniai parametrai, kurie buvo naudojami šiam tyrimui, yra pateikiami lentelėje žemiau:

Parametro pavadinimas	Parametro rūšis	Parametro reikšmė
Iteracijų skaičius.	Statinis.	1000.
Maksimalus galimas laiko eilutės ilgis	Statinis.	3 metai.
Mirtingumo analizės periodo pabaiga	Statinis.	2019-ieji metai.
Išskirtiniai metai, neįtraukti į analizę.	Statinis.	-
Analizuojamo žmonių amžiaus intervalas.	Statinis.	Nuo 18 iki 65 metų.

4 lentelė: Statiniai modelio parametrai, naudojami tyrime.

Mirtingumo įverčių tikslumas yra matuojamas χ^2 statistikos reikšmėmis. Šiuo atveju, laisvės laipsnių skaičius amžiaus intervale $[18, 65]$: $df = 48 + 1 = 49$. Pakankamai didelis įvykdytų iteracijų skaičius sugeneravo χ^2 statistikos reikšmių pasiskirstymus visais trim atvejais.



10 pav.: χ^2 statistikos reikšmių pasiskirstymai.

Empirinių skirstinių grafikai atspindi, jog statistiškai tinkamus mirtingumo parametrus yra sunkiausia aptikti vyrams, lengviausia - moterims. Grafikų interpretavimas yra toks: kuo empirinis skirstinys yra kairiau χ kritinės reikšmės (vertikali, raudona linija), tuo didesnis aptiktų, statistiškai patenkinamų mirtingumo kreivių skaičius. Žemiau lentelėje pateikiamos statistinės charakteristikos. Kadangi 95% pasikliovimo lygmens χ kritinė reikšmė yra 65.17, todėl kiekvienam atvejui, statistiškai gerai pavyko įvertinti mirtingumo normas.

Metrika	χ^2, vyrai	χ^2, moterys	χ^2, bendra populiacija
Vidurkis	98,77	65,42	91,10
Standartinis nuokrypis.	77,51	54,80	62,55
Maksimali reikšmė.	143,88	126,41	132,32
Minimali reikšmė.	64,97	37,91	63,15

5 lentelė: Modelio simuliacijos statistika χ^2 reikšmėms.

Mirtingumo dėsnis	χ^2, vyrai	χ^2, moterys	χ^2, bendra populiacija
Gompertz	65,06	41,35	64,42
Makeham	64,97	37,99	64,55
Weibull	66,84	39,42	63,17
Beard	66,09	47,91	65,29
Perks	65,87	38,04	64,34
Heligman-Pollard	67,13	41,55	64,19
Makeham ir polinomo kombinacija	65,26	41,96	63,15
Heligman-Pollard ir polinomo kombinacija	66,28	37,91	63,74

6 lentelė: Geriausi modelio scenarijai kiekvienam mirtingumo dėsniui.

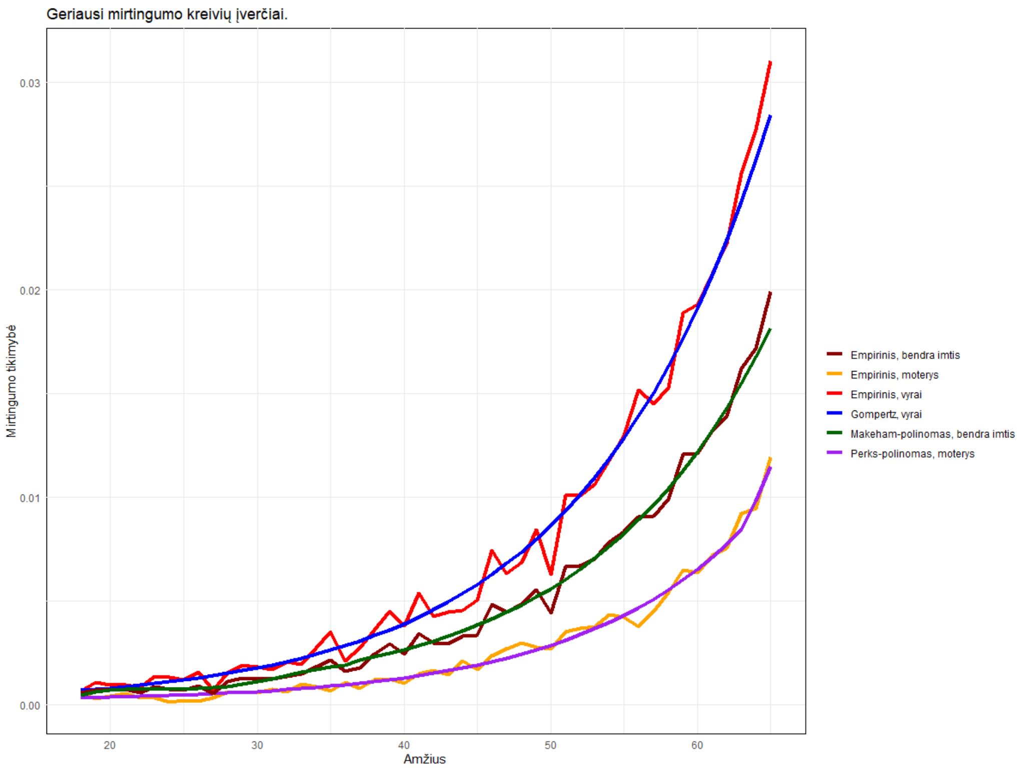
3.3 Statistiškai geriausi mirtingumo įverčiai

Praktiškai, vyrams labiausiai pasiteisino Gompertz mirtingumo dėsnis, moterims Makeham ir trečiojo laipsnio polinomo kombinacija, bendrai populiacijai - Heligman-Pollard ir trečiojo laipsnio polinomo derinys. Šiems scenarijams, sugeneruoti dinaminiai parametrai yra pateikiami lentelėje žemiau:

Parametro pavadinimas	Populiacija	Reikšmė
Mirtingumo taisyklių skaičius.	Vyrai	1
Mirtingumo taisyklių skaičius.	Moterys	2
Mirtingumo taisyklių skaičius.	Bendra	2
Amžiaus taškas, kuriame keičiasi mirtingumo taisyklė.	Vyrai	-
Amžiaus taškas, kuriame keičiasi mirtingumo taisyklė.	Moterys	29
Amžiaus taškas, kuriame keičiasi mirtingumo taisyklė.	Bendra	37
Amžiaus grupavimo vektorius.	Vyrai	$Z_{245} = (18, 23, 24, 30, 31, 33, 34, 44, 45, 51, 52, 61, 63, 65)$
Amžiaus grupavimo vektorius.	Moterys	$Z_{748} = (18, 29, 30, 37, 38, 45, 46, 52, 53, 58, 59, 62, 63, 65)$
Amžiaus grupavimo vektorius.	Bendra	$Z_{270} = (18, 29, 30, 36, 37, 41, 42, 48, 49, 58, 59, 63, 64, 65)$
Žmonių grupių skaičius.	Vyrai	3
Žmonių grupių skaičius.	Moterys	3
Žmonių grupių skaičius.	Bendra	3
Mirtingumo dėsnis.	Vyrai	Gompertz
Mirtingumo dėsnis.	Moterys	Makeham-polinomas
Mirtingumo dėsnis.	Bendra	Heligman-Pollard-polinomas

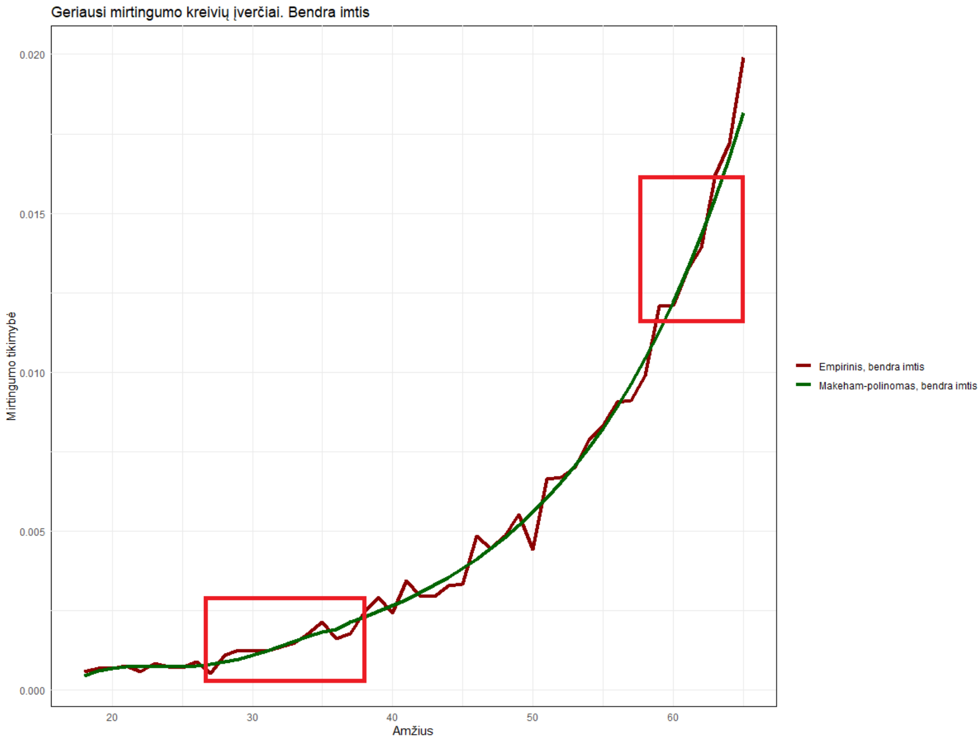
7 lentelė: Geriausių įverčių dinaminiai parametrai.

Taip pat, žemiau yra pateikiamas šių mirtingumo kreivių ir empirinių įverčių palyginimo grafikas:



11 pav.: Geriausių mirtingumo kreivių ir empirinių įverčių palyginimas.

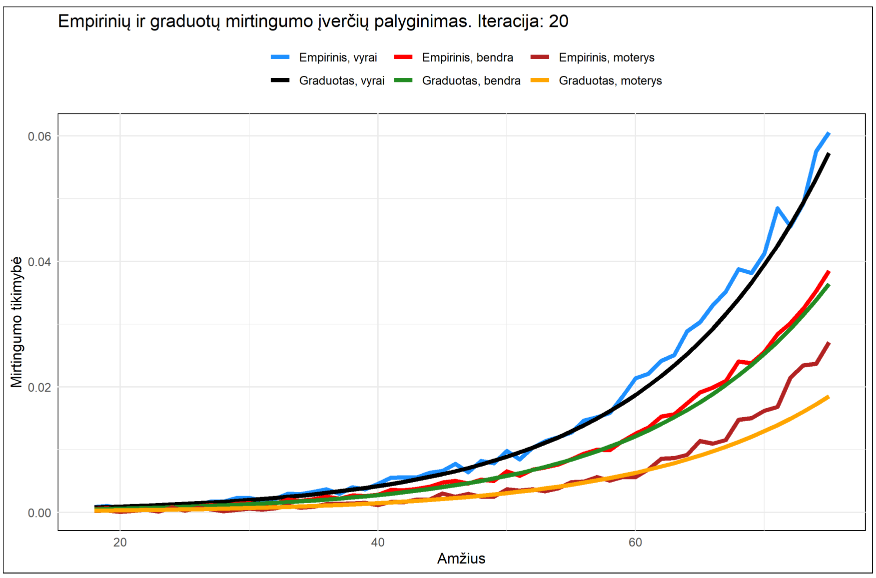
Kaip ir buvo galima tikėtis, geriausių mirtingumo kreivių įverčiai labiausiai yra priartėję prie to amžiaus intervalo, kuriame bendras nugyventų metų skaičius yra didžiausias. Tokie rezultatai yra įtakojami sukonstruotos χ^2 funkcijos, kurios reikšmių suma yra minimizuojama BFGS metodu. Žemiau, paveikslėlyje, pateikiamas pavyzdys bendros populiacijos atveju. Amžiaus intervalai, kuriuose populiacija yra labiausiai koncentruota, pažymėti raudonais stačiakampiais:



12 pav.: Mirtingumo vertinimas trečiojo laipsnio polinomo ir Makeham dėsnio kombinacija.

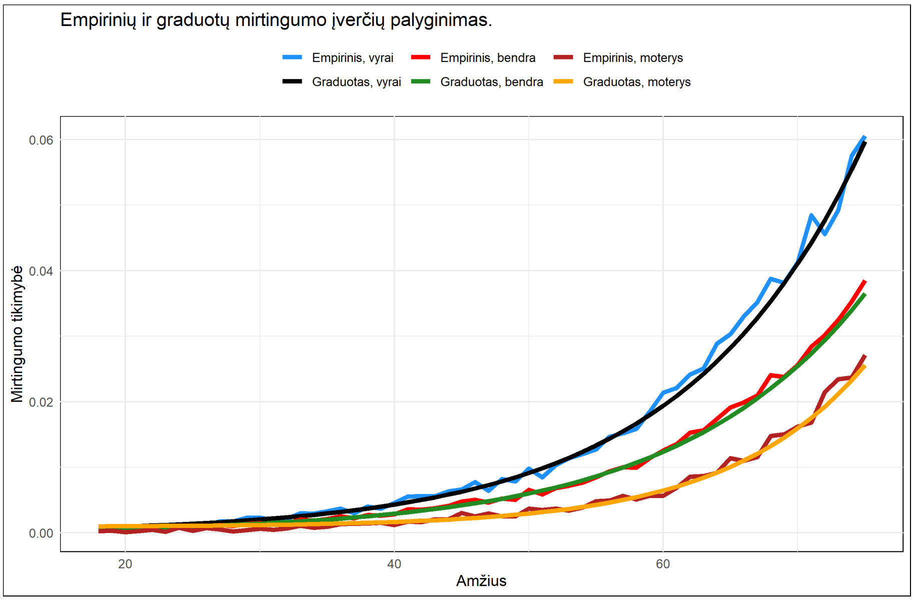
3.4 Statistiškai blogiausi mirtingumo įverčiai

Vienas iš tyrimų tikslų, panagrinėti statistiškai blogiausius mirtingumo įverčius ir pabandyti suprasti, kokia yra pagrindinė priežastis, dėl ko atsiranda mirties rodiklių pervertinimas/nuvertinimas. Panagrinėjus dešimtis variantų, buvo pastebėta, jog duomenų grupavimas pagal amžiaus intervalus labai nesiskyrė nuo palankiausiųjų scenarijų atvejų. Žemiau pateikiamas pavyzdys, kuomet taikytas Perks dėsnis tiek vyrams, tiek moterims, nuvertino mirtingumą. Tačiau pritaikytas Makeham dėsnis bendrai populiacijai, su tokiu pačiu duomenų grupavimu, statistiškai patenkino rodiklių atitikimą empiriniams duomenims.



13 pav.: Mirtingumo nuvertinimas vyrų ir moterų populiacijoms. Logistinis Perks dėsnis.

Kuomet mirtingumo taisyklė yra pakeičiama iš Perks dėsnio į Makeham, rezultatai tiek vyrams, tiek moterims, keičiasi kardinaliai. Svarbu paminėti, jog kiti parametrai išliko tokie patys. Todėl galima daryti prielaidą, kad šiuo atveju, populiacijų mirtingumo vertinimo rezultatai labiausiai priklauso nuo pasirinktos mirtingumo taisyklės.

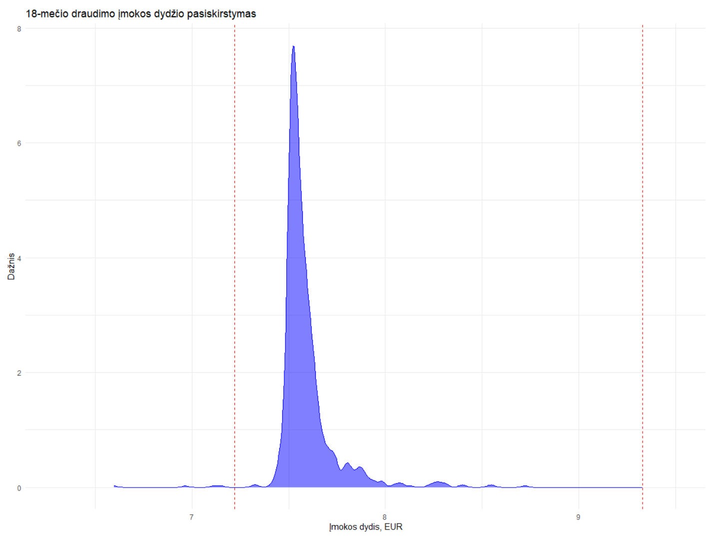


14 pav.: Gradavimas Makeham mirtingumo dėsniais.

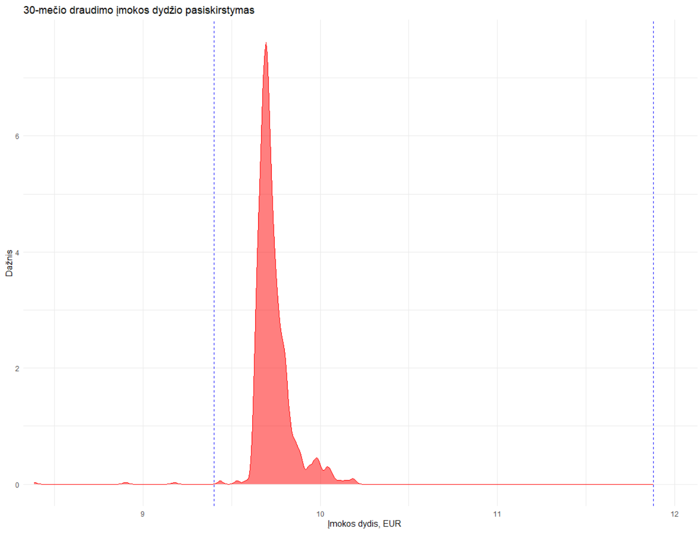
3.5 Draudimo įmokų dydžiai taikant skirtingas aktuarines prielaidas

Mirties tikimybės gali būti gan neintuityvus matavimo vienetas, jeigu mus domina klausimas, kokia yra skirtingų aktuarinių prielaidų įtaka gyvybės draudimo kompanijos verslui. Simuliuojant modelį, kiekvienos iteracijos metu, įmokos buvo skaičiuojamos 18-kos, 35-erių ir 50-ies metų amžiaus žmonėms. Kadangi įmokų skaičiavimuose yra draudžiama išskirti lytis, todėl panagrinėsime bendrąjį Lietuvos populiacijos atvejį.

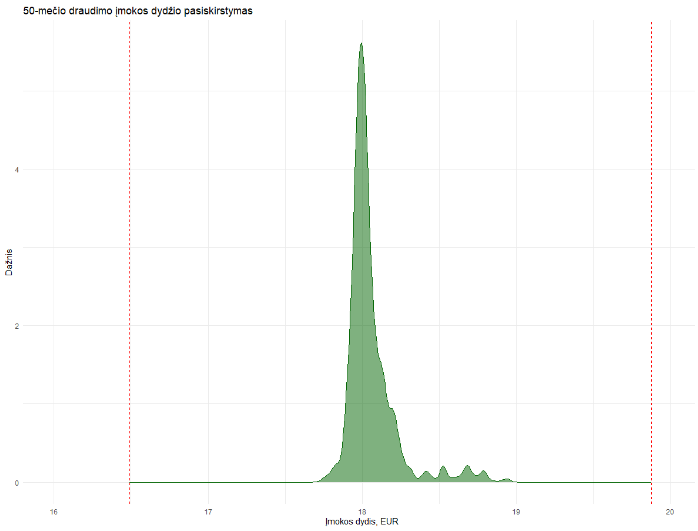
Tarkime, jog gyvybės draudimo kompanija prekiauja klasikinio gyvybės draudimo, produktu. Draudimo sumos dydį laikysime 20 000 EUR. Klientas, nepriklausomai nuo jo amžiaus, įsipareigoja mokėti vienodo dydžio įmoką, nustatytą pagal (2.12) formule. Žemiau pateikiami atitinkamo amžiaus žmonių įmokų pasiskirstymai, gauti iš modelio 1000 iteracijų, dėl naudotų skirtingų prielaidų. Empiriniai skirstiniai sudaryti ir iš tokių atvejų, kai žmogaus trupmeniniame amžiuje, galioja tiek tolygaus, tiek Balducci bei pastovios mirtingumo galios pasiskirstymo mirtingumo dėsniai, aprašyti 2.8 skyriuje. Grafikuose, vertikaliai brūkšninės linijos žymi 2.5% ir 97.5% įmokų dydžių kvantilius, kurie apskaičiuoti darant prielaidą, jog mirčių skaičius yra pasiskirstęs Puasono dėsnį.



15 pav.: 18-mečio empirinis įmokos dydžio pasiskirstymas.



16 pav.: 30-mečio empirinis įmokos dydžio pasiskirstymas.



17 pav.: 50-mečio empirinis įmokos dydžio pasiskirstymas.

Iš tiesų, mirtingumo lentelės skirtingo interpoliavimo efektas yra menkas. Gauti įmokų dyžiai yra labai panašūs tiek tolygaus, tiek Balducci ar pastovios mirtingumo galios prielaidų atvejais. Žemiau, lentelėje, pateikiama aštuoniolikmečio, trisdešimtmečio bei penkiasdešimtmečio asmenų įmokų dydžių statistika.

Metrika	**Prielaida trupmeniniame amžiuje**	P_{18}	P_{30}	P_{50}
Vidurkis	Tolygusis mirtingumo pasiskirstymas.	7,58	9,81	18,12
Standartinis nuokrypis.	Tolygusis mirtingumo pasiskirstymas.	0,1434	0,1472	0,1645
Maksimali reikšmė.	Tolygusis mirtingumo pasiskirstymas.	9,67	11,94	22,43
Minimali reikšmė.	Tolygusis mirtingumo pasiskirstymas.	6,41	8,72	17,04
Vidurkis	Pastovios mirtingumo galios pasiskirstymas.	7,57	9,79	18,08
Standartinis nuokrypis.	Pastovios mirtingumo galios pasiskirstymas.	0,1588	0,1354	0,1517
Maksimali reikšmė.	Pastovios mirtingumo galios pasiskirstymas.	9,66	11,91	22,37
Minimali reikšmė.	Pastovios mirtingumo galios pasiskirstymas.	6,40	8,71	17,01
Vidurkis	Balducci pasiskirstymas.	7,57	9,78	18,06
Standartinis nuokrypis.	Balducci pasiskirstymas.	0,1471	0,1384	0,1506
Maksimali reikšmė.	Balducci pasiskirstymas.	9,64	11,88	22,34
Minimali reikšmė.	Balducci pasiskirstymas.	6,40	8,69	16,99

8 lentelė: Įmokų dydžiai ir jų 1000 iteracijų rezultatų statistika.

Taip pat, įmokų dydžių vidurkiai geriausiųjų mirtingumo kreivių įverčių atveju, pagal skirtingą laiko eilutės ilgį:

Periodas	P_{18}	P_{30}	P_{50}
2017-2019 metai	7,51	9,81	18,16
2018-2019 metai	7,54	9,83	18,18
2019 metai	7,47	9,80	18,13

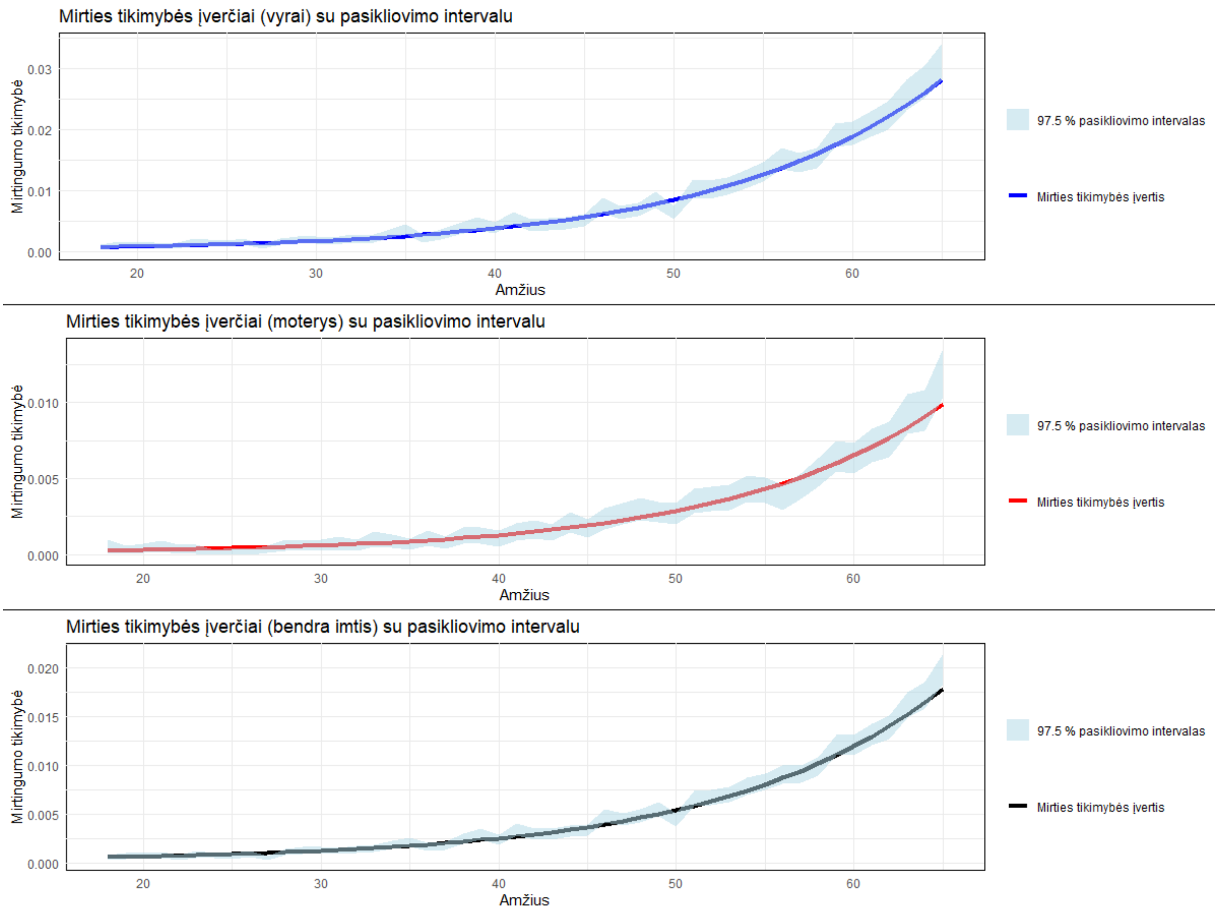
9 lentelė: Įmokų dydžiai ir jų 1000 iteracijų rezultatų statistika.

Galime daryti prielaidą, jog gautiems rezultatams, laiko eilutės ilgio įtaka - minimali.

4. Tyrimo išvados

Nors mirtingumo analizė praktikoje yra sudėtingas procesas, kurio galutinis rezultatas yra priklausomas nuo daugelio naudojamų parametrų, tačiau, Lietuvos populiacijos atveju, didžiausią įtaką rezultatams turėjo pasirinkto mirtingumo dėsnio taikymas. Gompertz dėsnis geriausiai nusakė vyrų mirtingumą, moterų - Makeham ir trečiojo laipsnio polinomo kombinacija, bendros populiacijos - Heligman-Pollard ir trečiojo laipsnio polinomo derinys.

Sukurtas modelis yra universalus, tačiau labiau pritaikytas praktiniame, gyvybės draudimo kompanijų skaičiavimų naudojime. Duomenų grupavimas į ilgesnių amžių intervalų grupes, populiacijų atveju, sukuria minimalų efektą, nes mirčių statistika kiekviename amžiaus taške yra pakankama. Taip pat, χ^2 reikšmių sumos minimizavimo tikslas yra mažiau priimtinas populiacijų mirtingumo atvejui. Manau, jog šalies mirtingumo vertinime, turi būti vienodai atsižvelgiama į kiekvieną žmogaus amžiaus tašką, kai šiame darbe analizuoto modelio atveju, labiausiai orientuojamasi į tuos taškus, kuriuose populiacija yra daugiausiai koncentruota. Nepaisant to, sukonstruoti mirtingumo įverčiai patenkino tiek statistinius testus, tiek loginį vertinimą. Visi mirtingumo kreivės taškai yra tarp 2.5% ir 97.5% kvantilių įverčių, todėl galutinis mirtingumo įverčius laikau patikimais.



18 pav.: Mirtingumo įverčių palyginimas su empirinėmis 2.5% ir 97.5% kvantilių reikšmėmis.

Literatūra

[1] Bowers N. L. Jr., Gerber H. U. Actuarial mathematics. The society of actuaries, 1997.

[2] Čekanavičius V., Murauskas G. Statistika ir jos taikymai I. Vilnius: TEV 2006, p. 192-203.

[3] Grigutis A., Šiaulys J. Aktuarinė matematika. Paskaitų konspektas.

[4] Pitacco M., Denuit M., Haberman S. Modelling Longevity Dynamics for Pensions and Annuity Business. New York: Oxford University 2009.

[5] Greg L. Somers, Richard G. Oderwald. Predicting Mortality with a Weibull Distribution. p. 85-103.

5. Priedas

Žemiau yra pateikiamas R-Studio programinis kodas, su apibrėžtomis skaičiavimų funkcijomis ir vykdymo algoritmu, iliustruotu pirmąjame paveikslėlyje (psl 7).

```
# Funckijų apibrėžimų sub-kodas.

chi_sq_testing <- function(graduated_rates, observed_rates, exposures, deaths, conf_level){

  vector_1 <- ((deaths - exposures * graduated_rates)^2)
  vector_1 <- vector_1[vector_1 > 0]

  vector_2 <- (exposures * graduated_rates * (1 - graduated_rates))
  vector_2 <- vector_2[vector_2 > 0]

  chi_sqrt <- sum(vector_1 / vector_2)

  degrees_of_freedom <- length(vector_1)

  chi_critical_value <- qchisq(conf_level, degrees_of_freedom)

  result <- list(
    chi_square_statistic = round(chi_sqrt, 3),
    chi_critical_value = round(chi_critical_value, 3),
    degrees_of_freedom = degrees_of_freedom
  )

  return(result)
}
poly_fit <- function(coef_vector, x_value, order_polynomial){

  value <- 0

  for (order in sort(seq((order_polynomial+1):1),decreasing = TRUE)){

    value <- value + as.double(coef_vector[order])*x_value^(order-1)


  }


  return (as.double(value))

}
wssplot <- function(data, nc=15, seed = 1234){

  wss <- (nrow(data)-1)*sum(apply(data,2,var))

  for (i in 2:nc){
    #set.seed(seed)
    wss[i] <- sum(kmeans(data, centers = i)$withinss)
  }

  plot(1:nc, wss,
  type = "b", xlab = "Lizdų skaičius",
  ylab = "Kvadratinė paklaida per visas grupes",
  col = "orange", lwd = 2,
  main = "Lizdų skaičiaus pasirinkimas")

}
wssplot_w_gg <- function(data, nc = 15, seed = 1234) {

  wss <- (nrow(data) - 1) * sum(apply(data, 2, var))

  for (i in 2:nc) {
    set.seed(seed)
    wss[i] <- sum(kmeans(data, centers = i)$withinss)
  }

  df <- data.frame(K = 1:nc, WSS = wss)

  ggplot(df, aes(x = K, y = WSS)) +
    geom_line(color = "orange", size = 2) +
    geom_point(color = "orange", size = 4) +
    labs(x = "Žmonių grupių skaičius",
    y = "Kvadratinė paklaida per visas grupes",
    title = "Žmonių grupių skirstymas pagal mirtingumą.") +


    # Applying the provided theme
    theme_minimal() +
    theme(
      plot.background = element_rect(fill = "white"),  # Set plot background to white
      panel.background = element_rect(fill = "white")  # Set panel background to white
    )

}
wssanalysis <- function(data, nc=15, seed = 1234){

  wss <- (nrow(data)-1)*sum(apply(data,2,var))

  for (i in 2:nc){
```

```
    #set.seed(seed)
    wss[i] <- sum(kmeans(data, centers = i)$withinss)
  }



  return(wss)
}
wssanalysis_extra <- function(data, nc, seed = 1234) {
  #set.seed(seed)
  kmeans_model <- kmeans(data, centers = nc)

  # Sum of squares within clusters
  wss <- kmeans_model$tot.withinss

  # Silhouette score
  cluster_labels <- as.numeric(as.character(kmeans_model$cluster))
  return(cluster_labels)
}
crude_rate_calc <- function(exposure_dataframe, claims_dataframe, period){

  exposure_dataframe
  <- exposure_dataframe[year <= analysis_data_to & year >= (analysis_data_to - period+1) & exposure_dataframe$country %in% countries, ]
  claims_dataframe
  <- claims_dataframe[year <= analysis_data_to & year >= (analysis_data_to - period+1) & claims_dataframe$country %in% countries, ]

  names(exposure_dataframe)[which(colnames(exposure_dataframe) == "count")] <- "count_exposure"

  exposure_dataframe
  <- exposure_dataframe[, lapply(.SD, sum, na.rm = TRUE),
  by =  c("age"), .SDcols = c("count_exposure")]
  claims_dataframe
  <- claims_dataframe[, lapply(.SD, sum, na.rm = TRUE),
  by =  c("age"), .SDcols = c("count")]

  crude_df <- merge(exposure_dataframe, claims_dataframe,
  by.x = "age", by.y = "age", all.x = TRUE)
  crude_df[is.na(count),"count"] <- 0

  crude_df$rate <- crude_df$count / crude_df$count_exposure

  return(crude_df)
}
group_data <- function(exposure_dataframe, claims_dataframe, groups_vector, variable_deaths, variable_exposure){

  count <- 0

  groupped_claims_data <- data.frame(age_group = character(),
                                     age = double(),
                                     #sex = character(),
                                     deaths = integer(),
                                     exposure = double(),
                                     death_rate_observed = double(),
                                     group_length = integer())

  for (i in seq(1,length(groups_vector)-1, by = 2)){
    count <- count + 1
    age_0 <- groups_vector[i]
    age_1 <- groups_vector[i+1]

    claims <-
    sum(claims_dataframe[claims_dataframe$death_age >= age_0
    & claims_dataframe$death_age <= age_1, variable_deaths ])
    exposure <- as.double(sum(exposure_dataframe[exposure_dataframe$age >= age_0 & exposure_dataframe$age <= age_1, variable_exposure]))

    groupped_claims_data[count, "age_group"] <- paste(age_0,age_1,sep="-")

    #if (length(sex) >1){
    #  groupped_claims_data[count,"sex"] <- "unisex"
    #} else{
     # groupped_claims_data[count,"sex"] <- sex
    #}

    groupped_claims_data[count,"deaths"] <- claims
    groupped_claims_data[count,"exposure"] <- exposure
    groupped_claims_data[count,"death_rate_observed"] <- round(claims/exposure,6)
    groupped_claims_data[count,"age"] <- (age_1-age_0+1)/2 + age_0
    groupped_claims_data[count,"group_length"] <- (age_1-age_0+1)
  }

  return (groupped_claims_data)
}
build_table <- function(mortality_table,cohort){

  built_tables <- data.frame(cohort = character(),
                             age = integer(),
                             death_rate = double())


  count <- 0
  for (age in 0:60){
    count <- count + 1

    built_tables[count,"cohort"] <- cohort
    built_tables[count,"age"] <- age

    death_rate <- as.double(mortality_table[mortality_table$Year == cohort + age
    & mortality_table$Age == age, 3])
    built_tables[count,"death_rate"] <- death_rate
  }
  return (built_tables)

}
cheby <- function(N, X) {
  if (N == 0) {
```

```
    return(1)
  }
  else if (N == 1) {
    return(X)
  }
  else {
    return(2 * X * cheby(N - 1, X) - cheby(N - 2, X))
  }
}
boole <- function(f){
  (7*f(0) + 32*f(0.25) + 12*f(0.5) + 32*f(0.75) + 7*f(1)) / 90

}
gompertz.mu <- function(p, x) {
  exp(p[1] + p[2] * x)
}
gompertz.int <- function(p, x){
  ((exp(p[2]) - 1) / p[2]) * exp(p[1] + p[2] * x)
}
makeham.mu <- function(p, x) {
  exp(p[3]) + exp(p[1] + p[2] * x)
}
makeham.int <- function(p, x){
  exp(p[3]) + ((exp(p[2]) - 1) / p[2]) * exp(p[1] + p[2] * x)
}
weibull.mu <- function(p, x) {
  p[1] * x^(p[2])
}
weibull.int <- function(p, x) {
  if (p[2] == -1) {
    # Avoid singularity when p[2] is -1
    return(exp(p[1] * (log(x + 1) - log(x))))
  } else {
    return(exp(p[1]/(p[2] + 1) * (x + 1)^(p[2] + 1) + p[1]/(p[2] + 1) * x^(p[2] + 1)))
  }
}
#weibull.int <- function(p, x){
 # exp(p[1]/(p[2] + 1) * (x + 1)^(p[2] + 1) + p[1]/(p[2] + 1) * x^(p[2] + 1))
#}
perks.mu <- function(p, x){
  exp(p[1] + p[2] * x) / (1 + exp(p[1] + p[2] * x))
}
perks.int <- function(p, x){
  (1/p[2]) * log((1 + exp(p[1] + p[2] * (x + 1)))
                 / (1 + exp( p[1] + p[2] * x )))
}
beard.mu <- function(p, x) {
  exp(p[1] + p[2] * x) /(1 + exp(p[1] + p[3] + p[2] * x))
}
beard.int <- function(p, x){
  (exp(-p[3])/p[2]) * log((1 + exp(p[1] + p[3] + p[2] * (x + 1)))
                          / (1 + exp(p[1] + p[3] + p[2] * x)))
}
makeham.perks.mu <- function(p, x) {
  (exp(p[3]) + exp(p[1] + p[2] * x)) /
    (1 + exp(p[1] + p[2] * x))
}
makeham.perks.int <- function(p, x){
  exp(p[3]) + ((1 - exp(p[3]))/p[2]) * log((1 + exp(p[1] + p[2] * (x + 1)))
                                           / (1 + exp(p[1] + p[2] * x)))
}
makeham.beard.mu <- function(p, x) {
  (exp(p[3]) + exp(p[1] + p[2] * x)) /
    (1 + exp(p[1] + p[4] + p[2] * x))
}
makeham.beard.int <- function(p, x){
  exp(p[3]) + ((exp(-p[4]) - exp(p[3]))/p[2]) *
    log((1 + exp(p[1] + p[4] + p[2]*(x+1))) /
          (1 + exp(p[1] + p[4] + p[2] * x)))
}
gmrs.mu <- function(p, r, s, x) {
  TX <- (x - 70) / 50
  sum1 <- 0
  if (r != 0) {
    for (i in 1:r) { sum1 <- sum1 + p[i] * cheby(i - 1, TX) }
  }
  sum2 <- 0
  if (s != 0) {
    for (j in 1:s) { sum2 <- sum2 + p[j + r] * cheby(j - 1, TX) }
  }

  return(sum1 + exp(sum2))
}

gmrs.int <- function(p, r, s, x) {
  integrand <- function(tt) { gmrs.mu(p, r, s, x + tt) }
  boole(integrand)
}

gmrs <- function(r, s) {
  list (mu = function(p, x) { gmrs.mu (p, r, s, x) },
        int = function(p, x) { gmrs.int(p, r, s, x) } )
}

law.set <- function(p, law) {
  off <- if(!MODEL.M) 0 else -0.5
  list(mu = function(x) { law$mu (p, x + off) },


       int = function(x) { law$int(p, x + off) })
}

ll <- function(p, f, d) {
  mu <- pmax(MININIMUM.MU, f(p, d$x))
  sum(ifelse(d$e == 0, 0, -d$e * mu + d$d * log(d$e*mu) - lfactorial(d$d)))
```

```
}

fit <- function(p.init , f, d) {

  ll.fit <- function(p) { ll(p, if (!MODEL.M) f$int else f$mu, d) }

  grr <- function(p) { grad(ll.fit , p, method = GRAD.METHOD) }

  model <-optim(par = p.init , gr = grr , fn = ll.fit , method=OPTIM.METHOD ,
                control = list(fnscale=-1,
                               reltol=REL.TOLERANCE ,
                               maxit=MAX.ITERATIONS, trace = 1))

  hess.calc <- hessian(ll.fit , model$par , method = GRAD.METHOD)

  return(list(ll=model$value , pars = model$par , npars = length(model$par),
              hessian = hess.calc , fun=law.set(model$par , f), convergence_info = model))
}

fit.gompertz <- function(d) {
  fit(c(INIT.ALPHA , INIT.BETA), gompertz , d)
}

fit.makeham <- function(d) {
  model.gompertz <- fit.gompertz(d)
  fit(c(model.gompertz$pars , INIT.EPSILON), makeham , d)
}

fit.weibull <- function(d) {
  fit(c(INIT.C , INIT.n), weibull , d)
}
fit.perks <- function(d) {
  fit(c(INIT.ALPHA , INIT.BETA), perks , d)
}

fit.beard <- function(d) {
  model.perks <- fit.perks(d)
  fit(c(model.perks$pars , INIT.RHO), beard , d)
}

fit.makeham.perks <- function(d) {
  model.perks <- fit.perks(d)
  fit(c(model.perks$pars , INIT.EPSILON), makeham.perks , d)
}


fit.makeham.beard <- function(d) {
  model.makeham.perks <- fit.makeham.perks(d)
  fit(c(model.makeham.perks$pars , INIT.RHO), makeham.beard , d)
}


fit.gmrs <- function(d, r, s) {
  p.init <- if (s == 0) NULL
  else if (s == 1) INIT.B1
  else c(INIT.B1, INIT.B2, rep(INIT.BX, s - 2))
  gmrs.fit <- if (s != 0) fit(p.init , gmrs(0, s), d) else NULL

  if (r == 0) {
    return(gmrs.fit)
  }
  else {
    for (i in 1:r) {
      r.pars <- if (i == 1) INIT.AX else c(gmrs.fit$pars[1:(i-1)],
                                           INIT.AX)
      s.pars <- if(s != 0) gmrs.fit$pars[i:(i+s-1)] else NULL
      gmrs.fit <- fit(c(r.pars , s.pars), gmrs(i, s), d)
    }
    return(gmrs.fit)
  }
}

fit.gm <- function(r, s) {
  function(d) {
    fit.gmrs(d, r, s)
  }
}



graduate.poisson <- function(f, d) {

  # Set results
  result <- f(d)

  ll <- result$ll
  pars <- result$pars
  npars <- result$npars
  hessian <- result$hessian

  m <- if(!MODEL.M) function(x) { result$fun$int(x) }
  else function(x) { result$fun$mu(x + 0.5) }

  fun <- list(mu = result$fun$mu,
              int = result$fun$int ,
              qx = function(x) { 1 - exp(-result$fun$int(x)) },
              m = m)
  rates <- m(d$x)

  # Calculate standard errors and p-values for parameter estimates
  std.dev <- sqrt(diag(solve(-hessian)))
  z.stat <- pars / std.dev
  p.values <- 2*pnorm(abs(z.stat), 0, 1, lower.tail = FALSE)

  # Calculate deviance and residuals
```

```
  O <- d$d
  E <- rates * d$e
  dev <- sum(2 * (ifelse(O == 0, 0, O * log(O/E)) - (O - E)))
  res <- sign(O - E) * sqrt(2 * (ifelse(O == 0, 0, O * log(O/E)) - (O - E)))

  # Determine dispersion coefficient
  dis <- dev / (length(d$x) - npars)

  # Chi -squared calculations
  chi <- sum(((O - E)^2) / ifelse(E == 0, 1, E))
  chi.p <- pchisq(chi , df = length(d$x) - npars
  , lower.tail = FALSE)

  # Information criteria
  aic <- -2 * ll + 2 * npars
  bic <- -2 * ll + log(length(d$x)) * npars

  # Signs test
  signs.p <- sum(res > 0)
  signs.n <- sum(res < 0)
  signs.a <- if(signs.p <= signs.n) "less" else "greater"
  signs.test <- binom.test(signs.p, signs.p + signs.n,
                           alternative = signs.a)$p.value

  runs <- length(rle(sign(res))$lengths)
  runs.p <- runs.test(res , "left.sided", 0, "exact", FALSE)$p.value

  lower = log(qpois(0.025, O) / d$e)
  upper = log(qpois(0.975, O) / d$e)

  details <- list(pars = pars , npar = npars , std.dev = std.dev ,
                 z.stat = z.stat , p.values = p.values , ll = ll, aic = aic ,
                 bic = bic , dis = dis , chi = chi , chi.p = chi.p,
                 signs.p = signs.p, signs.n = signs.n,
                 signs.test = signs.test , runs = runs ,
                 runs.p = runs.p, rates = log(rates),
                 q_x_fun = m, lower_b = lower, upper_b = upper)

  lower = log(qpois(0.025, O) / d$e)
  upper = log(qpois(0.975, O) / d$e)
  par(mfrow=c(2, 1))

  obs <- log(d$d / d$e)
  mod <- log(rates)

  if(length(mod) == 1) mod <- rep(mod , length(obs))
  y.min <- min(obs[is.finite(obs)],
              mod[is.finite(mod)],

              lower[is.finite(lower)])
  y.max <- max(obs[is.finite(obs)],
              mod[is.finite(mod)],
              upper[is.finite(lower)])

  #plot (d$x, lower , pch = 150, cex=1.25, col = "blue",
   #     xlab = "Amžius", ylab = " LN(q_x) ",
    #    ylim = c(y.min ,y.max), axes = FALSE)
  #axis(1, at=c(seq(from=floor(min(d$x)/5)*5,to=ceiling(max(d$x)/5)*5,by=5)),
   #    las = 1)
  #axis(2, las = 1)
  #box()
  #points(d$x, upper , pch = 150, cex = 1.25, col = "green")
  #points(d$x, obs , col = "black", cex = 1.2)
  #lines(d$x, mod , col = "red", lwd = 2)
  #legend("bottomright", lwd=c(NA, 2, NA, NA), lty=c(NA, 1, NA, NA),
      #   pch=c(1, NA, 150, 150), col=c("black", "red", "green", "blue"),
     #    legend=c("Empirinis įvertis","Graduotas įvertis","97.5% Percentilis",
    #              "2.5% Percentilis"),
   #      bty="n", cex=0.7)

  #plot(d$x, res , col = "blue",
  #     xlab = "Amžius", ylab = "Paklaidos", axes = FALSE)
  #axis(1, at=c(seq(from=floor(min(d$x)/5)*5,to=ceiling(max(d$x)/5)*5,by=5)),
   #    las = 1)
  #axis(2, las = 1)
  #box()
  #abline(h = 0, col = "red")
  #abline(h = 1.96, lty = 3)
  #abline(h = -1.96, lty = 3)

  return(list(fun = fun , details = details, k = mod))
}



# Initial Makeham, Perks and Gompertz parameters:
INIT.ALPHA <- -10
INIT.BETA <- 0.1
INIT.EPSILON <- -10
INIT.RHO <- 1
INIT.C   <- 0.5
INIT.n   <- 1.08

# Initial GMRS parameters:
INIT.B1 <- -5
INIT.B2 <- 5
INIT.BX <- 2
INIT.AX <- 2



OPTIM.METHOD <- "BFGS"
GRAD.METHOD <- "Richardson"
REL.TOLERANCE <- 1e-20
MAX.ITERATIONS <- 1000
```

```
MININIMUM.MU <- 1e-20


MODEL.M <- TRUE




gompertz <- list(mu = gompertz.mu, int = gompertz.int)
makeham <- list(mu = makeham.mu, int = makeham.int)
weibull <- list(mu = weibull.mu, int = weibull.int)
perks <- list(mu = perks.mu, int = perks.int)
beard <- list(mu = beard.mu, int = beard.int)
makeham.perks <- list(mu = makeham.perks.mu, int = makeham.perks.int)
makeham.beard <- list(mu = makeham.beard.mu, int = makeham.beard.int)



standard_mortality_table <- function(mortality_table, initial_kohort_size) {

  std_mort_table <- data.frame(age_in_months = mortality_table$age_in_months)
  std_mort_table$l_x <- 0

  l_x <- initial_kohort_size

  for (i in 1:nrow(std_mort_table)){

    if (i == 1){
      std_mort_table$l_x <- l_x

    } else {
      decrement <- l_x * as.double(mortality_table[i, "mortality_rate"])
      l_x <- l_x - decrement
      std_mort_table$l_x[i] <- l_x
    }
  }

  return(std_mort_table)
}




# PAGRINDINIS PARAMETRŲ NUSTATYMO IR VALDYMO KODAS.




# 00: nustatome R-Studio paketų instaliavimo lokaciją savo kompiuteryje:
.libPaths(new="C:/Users/laury/OneDrive/Dokumentai/R/win-library/4.1")

# 01: nustatome R-Studio darbinę aplinką šiam kodui:
setwd("C:/Users/laury/Desktop/Magistras")

# 02: instaliuojame (jeigu reikia, kitu ateveju - užkomentuoti) reikalingus R-Studio paketus:
#install.packages("data.table")
#install.packages("numDeriv")
#install.packages("randtests")
#install.packages("ggplot2")

# 03: įkeliame R-Studio paketus, reikalingus darbui atlikti:
library(data.table)
library(numDeriv)
library(randtests)
library(ggplot2)

# 04: nustatome pagrindinius Mirtingumo Analizės parametrus:
source("01_define_functions.R")







# Iteracijų skaičius:
iterations_number                <- 1

# Skirtingų asmenų amžius:
first_person_age                 <- 18
second_person_age                <- 35
third_person_age                 <- 50


# Automatiškai sukuriame naują aplanką rezultatams:
current_timestamp <- format(Sys.time(), "%Y%m%d")
folder_name <- paste("01_R_output/01_image_output/Results_", current_timestamp, sep = "")

if (!dir.exists(folder_name)) {
  dir.create(folder_name)
}


while (iterations_counter < iterations_number){

  tryCatch({

    # Laiko eilutės ilgio parametrai:
    max_time_series_length            <-    1
    analysis_data_to                  <- 2019
    analysis_data_from                <- analysis_data_to - round(runif(1, min = 1, max = max_time_series_length),0) + 1
    years_exclude                     <- c()
```

```
    # Analizuojamo Amžiaus parametrai:
    age_from                          <-    18
    age_to                            <-    65


    # Gradavimo parametrai:
    number_of_mortality_laws          <- round(runif(1, min = 1, max = 2),0)
    apply_mortality_rule_males_from   <- ifelse(number_of_mortality_laws == 1, age_from, round(runif(1, min = age_from + 10, max = round(age_from - (age_from - a
    apply_mortality_rule_males_till   <- age_to
    apply_mortality_rule_total_from   <- apply_mortality_rule_males_from
    apply_mortality_rule_total_till   <- age_to
    apply_mortality_rule_females_from <- apply_mortality_rule_males_from
    apply_mortality_rule_females_till <- age_to
    polynomial_young_order_male       <- round(runif(1,min = 3, max = 5),0)
    polynomial_young_order_female     <- round(runif(1,min = 3, max = 5),0)
    polynomial_young_order_total      <- round(runif(1,min = 3, max = 5),0)
    graduation_old_rule_male          <- sample(list(fit.makeham, fit.makeham.beard, fit.perks, fit.beard, fit.makeham.perks, fit.gompertz),1)[[1]]
    graduation_old_rule_female        <- sample(list(fit.makeham, fit.makeham.beard, fit.perks, fit.beard, fit.makeham.perks, fit.gompertz),1)[[1]]
    graduation_old_rule_total         <- sample(list(fit.makeham, fit.makeham.beard, fit.perks, fit.beard, fit.makeham.perks, fit.gompertz),1)[[1]]
    apply_poly_grad_from_male         <- age_from
    apply_poly_grad_from_female       <- age_from
    apply_poly_grad_from_total        <- age_from
    # Clustering Analizės parametrai žmonių amžiui:
    apply_clustering_analysis         <- TRUE
    min_number_cluster                <-   3
    max_number_cluster                <-   6

    # c(0:105)
    # Žmonių amžiaus grupavimas (rankiniu būdu, jeigu nėra taikoma Clustering Analizė. T.y apply_clustering_analysis = "FALSE")
    age_grouping_vector_males         <- c(0,7,8,12,13,17,18,67,68,72,73,80,81,98,99,99,100,100,101,101,102,102)
    age_grouping_vector_females       <- c(0,7,8,12,13,17,18,62,63,67,68,72,73,80,81,98,99,99,100,100,101,101,102,102)
    age_grouping_vector_total         <- c(0,7,8,12,13,17,18,100,101,101,102,102)
    # Gybybės draudimo kompanijos parametrai:
    interest_rate                     <-  0.01
    portfolio_size                    <-    10
    risk_insurance_amount             <- 20000
    premium_safety_margin             <-  0.15


    # 05: nuskaitome reikalingus mirtingumo duomenis:
    if (iterations_counter == 0){
      source("02_read_data.R")
    }


    # 06: apskaičiuojame taškinius (empirinius) mirtingumo įverčius:
    source("03_crude_rate_calc.R")

    # 07: grupuojame duomenis pagal amžių:
    source("04_group_data_manually.R")

    # 08: graduojame mirtingumo kreivę ir gaminame metinę mirtingumo lentelę:
    source("05_mortality_graduation_process.R")

    # 09: statistinis gradavimo testavimas (goodness-of-fit):
    source("08_goodness_of_fit_testing.R")

    # 10: interpoliuojame mirtingumo lenteles:
    source("06_mortality_table_interpolation.R")

    # 11: skaičiuojame draudimo įmokas:
    source("07_premiums_calc.R")

    # 12: išsaugojame rezultatus:
    source("09_save_calculated_results.R")



    iterations_counter <- iterations_counter + 1

  }, error = function(e) {

    iterations_counter <- iterations_counter + 1

  })

}
```