UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

BAKALÁŘSKÁ PRÁCE

Konstrukce úmrtnostních tabulek a jejich využití v pojišťovnictví

Vedoucí bakalářské práce:
Mgr. Ondřej Vencálek, Ph.D.
Rok odevzdání: 2013

Vypracovala:
Michaela Polášková
MATPOJ, III. ročník

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci zpracovala samostatně pod vedením pana Mgr. Ondřeje Vencálka, Ph.D. a všechny použité zdroje jsem uvedla v seznamu použité literatury.

V Olomouci dne 18. 4. 2013 ……………………………

Michaela Polášková

Poděkování

Na tomto místě bych ráda poděkovala svému vedoucímu bakalářské práce panu Mgr. Ondřeji Vencálkovi, Ph.D. za jeho odborné rady, cenné připomínky a především za čas, který mi věnoval při konzultacích. Touto cestou děkuji i paní Mgr. Evě Bohanesová, Ph.D. za poskytnutou literaturu.

OBSAH

1 ÚVOD

Při výpočtech pojistného životního a důchodového pojištění se pojišťovny snaží odhadnout, jaká je pravděpodobnost, že se pojištěný dožije či nedožije sjednaného konce pojištění a zda tak může dojít k pojistné události. Tuto pravděpodobnost a další jiné charakteristiky potřebné k určení sazeb pojistného nám udávají tzv. úmrtnostní tabulky. Hlavní charakteristikou, která slouží pro výpočet všech ostatních ukazatelů úmrtnostních tabulek, je tzv. pravděpodobnost úmrtí. Jaké další důležité ukazatele se v úmrtnostních tabulkách nachází, jak se vypočítají a co se děje s úmrtností ve vysokých věcích života, se dozvíte v následujících kapitolách.

Cílem této práce bude seznámit se se způsobem, jakým se úmrtnostní tabulky konstruují. Hlavním cílem práce bude seznámit se s Gompertz-Makehamovým zákonem a následnou aplikací tohoto zákona na reálná data české populace.

Úkolem druhé kapitoly je bližší seznámení čtenáře s úmrtnostními tabulkami a jednotlivými druhy úmrtnostních tabulek. Kapitola tedy obsahuje informace o hlavních ukazatelích úmrtnostních tabulek a vztahy mezi nimi. Dále se čtenář seznámí s nově zavedenou novelou zákona a to úmrtnostními tabulkami nezávislými na pohlaví. Posledním bodem druhé kapitoly budou nejaktuálnější údaje o počtu zemřelých podle příčin v České republice.

Kapitola třetí bude věnována Gompertz-Makehamově zákonu o úmrtnosti. Cílem této kapitoly bude odvodit vztahy pro odhady parametrů tohoto modelu a následné určení optimálních odhadů parametrů Gompertz-Makehamovy funce a porovnání parametrů v letech 2009-2011 a to u žen i mužů české populace.

Kapitola čtvrtá bude věnována fenoménu „Late-life mortality deceleration“, tedy jakémusi zpomalení růstu úmrnosti. V této kapitole zmíním, jaké jsou příčiny „zpomalení“ růstu úmrtnosti a jak se v praxi řeší. Dále bych zde chtěla přiblížit modifikaci Gompertz-Makehamova zákona, která zohledňuje právě dané „zpomalení“ růstu úmrtnosti. V poslední řadě budu aplikovat tento modifikovaný Gompertz-Makehamův zákon na konkrétní data mužské české populace a následně porovnávat s původním modelem.

2 ÚMRTNOSTNÍ TABULKY

Obsahem této kapitoly bude úvodní seznámení s úmrtnostními tabulkami a proč se úmrtností tabulky sestavují. Dále zde budou popsány základní charakteristiky, se kterými se v úmrtnostních tabulkách setkáváme, konstrukce úmrtnostních tabulek, základní druhy úmrtnostních tabulek a v neposlední řadě se čtenář seznámí s úmrtnostními tabulkami nezávislými na pohlaví. Při zpracování této kapitoly jsem vycházela z publikací [2], [3], [7], [8], a internetových zdrojů [9], [10], [11], [12], [16].

2.1 Motivace tvorby úmrtnostních tabulek

Úmrtnostní tabulky a jejich tvorba jsou úzce spjaty s životním a důchodovým pojištěním. Životní pojištění všichni většinou známe jako životní pojištění pro případ smrti. V tomto případě při úmrtí pojištěné osoby pojišťovna vyplatí pojistné plnění ve většině případů někomu z pozůstalých. Další, méně známé, je pojištění pro případ dožití, kdy pojišťovna vyplácí pojistné plnění v případě dožití se určitého věku. V pojištění pro případ smrti se pojistné plnění vyplácí přímo pojištěnému. Za pojištění pro případ dožití můžeme považovat důchodové pojištění. Dále existuje tzv. smíšené pojištění, které je kombinací obou předchozích typů pojištění. V tomto případě pojišťovna vyplácí pojistně plnění buď pozůstalým, pokud pojištěný v průběhu pojištění zemře, nebo je pojistné plnění vypláceno pojištěnému, pokud se pojištěný dožije konce pojištění.

Při uzavření výše uvedených druhů pojištění pojišťovna inkasuje pojistné. Pro stanovení výše pojistného musíme vědět, s jakou pravděpodobností a jak vysoké pojistné plnění bude pojišťovna v budoucnosti vyplácet. Při stanovení výše pojistného vychází pojišťovna především z pravděpodobnostních zákonitostí úmrtnosti v dané populaci. Tyto pravděpodobnostní zákonitosti jsou obsaženy právě v úmrtnostních tabulkách.

2.2 Úmrtnostní tabulky

Úmrtnostní tabulky jsou specifickou metodou užívanou k popsání řádu vymírání určité populace. Tyto tabulky vycházejí z ukazatele pravděpodobnosti úmrtí v jednotlivých věkových kategoriích, kde počet zemřelých vztahujeme k počátečnímu počtu osob vystavených riziku úmrtí, nikoliv ke střednímu stavu obyvatel.

Na základě ukazatele pravděpodobnosti úmrtí lze přejít od reálné populace k tabulkové (fiktivní) populaci, která vychází ze zaokrouhlého počtu narozených osob, obvykle 100 000 osob – toto číslo nazýváme kořen tabulky. Dále aplikací těchto reálných pravděpodobností úmrtí na danou tabulkovou populaci (100 000 obyvatel) dostáváme prostřednictvím specifických výpočtů tabulkové počty jak zemřelých, tak žijících a také získáváme hlavní výstup úmrtnostní tabulky a to střední délku života, která je definovaná jako průměrný počet let, které zbývá osobě ve věku x ještě do konce života prožít.

2.3 Ukazatele v úmrtnostních tabulkách, konstrukce úmrtnostních tabulek

Při konstrukci úmrtnostních tabulek podle metodických poznámek Českého statistického úřadu [11] se vychází z pravděpodobnosti úmrtí q_x, která je odvozena ze specifických měr úmrtnosti m_x. Vstupními daty pro výpočet specifických měr úmrtosti je pro nás počet skutečně zemřelých obyvatel D_x ve věku x z dané populace a daného pohlaví a střední stav populace P_x ve věku x. Uvažujeme populaci homogenní, tedy takovou populaci, která se liší pouze věkem a pohlavím a dále uvažujeme, že populace nepřijímá nové jedince a ztrácí jedince pouze úmrtím.

Střední stav populace je jednou ze základních charakteristik, kterou demografická statistika sleduje. Jedná se o počet x-letých obyvatel v daném kalendářním roce, kteří se dožijí středu sledovaného období. Střed kalendářního roku byl pro tyto účely stanoven Českým statistickým úřadem na půlnoc z 30. 6. na 1. 7. daného kalendářního roku. Střední stav populace v úmrtnostních tabulkách budu značit jako P_x. Počet skutečně zemřelých osob ve věku x budu v tabulkách značit jako D_x.

Východiskem pro výpočet jednotlivých ukazatatelů uváděných v úmrtnostích tabulkách a pro samotnou konstrukci úmrtnostních tabulek jsou pravděpodobnost úmrtí ve věku x a pravděpodobnost dožití ve věku x.

Pravděpodobnost úmrtí x-letého jedince, kterou budu značit jako q_x udává, s jakou pravděpodobností se osoba dožívající se přesného věku x let v daném období nedožije věku $x + 1$ let, tedy že zemře před dosažením věku $x + 1$ let. Tato pravděpodobnost se počítá jako:

$$\boldsymbol{q_x = 1 - e^{-m_x}} \,,$$

kde $\boldsymbol{m_x}$ značí poměr celkového počtu zemřelých z jednotlivých generací D_x ke střednímu stavu populace P_x. Tento poměr se nazývá specifická míra úmrtnosti:

$$\boldsymbol{m_x = \frac{D_x}{P_x}}.$$

S využitím Taylorova rozvoje platí:

$$q_x = 1 - e^{-m_x} \sim 1 - (1 - m_x) = \; m_x.$$

Můžu tedy říct, že $q_x \sim m_x$. Vynásobím-li specifickou míru úmrtnosti 1000 krát, dostanu počet zemřelých na 1000 obyvatel.

Pravděpodobnost úmrtí ve věku 0 (tzv. kojenecká úmrtnost) je rovna podílu zemřelých ve věku 0 a živě narozených:

$$\boldsymbol{q_0 = \frac{D_0}{N^v}} \,,$$

kde N^v značí počet živě narozených osob v daném období a v dané populaci.

Doplňkem pravděpodobnosti úmrtí x-letého jedince q_x je pravděpodobnost dožití ve věku x. Tuto pravděpodobnost budu značit jako p_x a udává, s jakou pravděpodobností se osoba dožívající se přesného věku x let v daném období dožije věku $x + 1$ let. Pravděpodobnost dožití ve věku x se počítá jako:

$$\boldsymbol{p_x = 1 - q_x.}$$

Musí proto platit následující rovnost:

$$\boldsymbol{p_x + \; q_x = 1} \,.$$

Dalším ukazatelem úmrtnostních tabulek je tabulkový počet dožívajících, budu značit l_x. Tabulkový počet dožívajících je hypotetický počet osob, které se dožijí věku x let ze 100 000 živě narozených. Tento tabulkový počet dožívajích se vypočítá jako:

$$\boldsymbol{l_{x+1} = p_x \cdot l_x},$$

kde kořen tohoto rekurentního vzorce je roven:

$$\boldsymbol{l_0 = 100\,000}.$$

Tabulkový počet zemřelých, který budu značit d_x, vyjadřuje hypotetický počet zemřelých osob v dokončeném věku x let, počítá se jako rozdíl dvou po sobě jdoucích tabulkových počtů dožívajících l_x a l_{x+1} :

$$\boldsymbol{d_x = l_x - l_{x+1}}.$$

Tabulkový počet žijících, značí se L_x , je hypotetický průměrný počet žijících v dokončeném věku x let, počítá se jako průměr ze dvou po sobě jdoucích tabulkových počtů dožívajících (kromě věku 0). Jedná se o tzv.celkový počet “člověkoroků”, které osoby ve věku x let do věku $x + 1$ let prožijí:

$$\boldsymbol{L_x = \frac{l_x + l_{x+1}}{2}}.$$

Tabulkový počet žijících osob ve věku 0, tzv.nulaletých jedinců, je odvozen z rozložení zemřelých kojenců v daném období podle ročníku jejich narození. Tabulkový počet “nulaletých” žijícícíh se počítá jako:

$$\boldsymbol{L_0 = l_0 - \alpha \cdot d_0},$$

kde α je tzv. korekční koeficient, který udává, jaká část zemřelých v prvním roce zemřela do 6 měsíců života. Korekční koeficient je definovaný jako:

$$\boldsymbol{\alpha = \frac{d_{0'}}{d_0}},$$

kde $d_{0'}$ je počet zemřelých ve věku 0 do 6 měsíců života a d_0 je celkový počet zemřelých ve věku 0.

Ukazatel T_x, tzv. pomocný ukazatel, udává, kolik let života má tabulková generace v daném věku x ještě před sebou.

$$\boldsymbol{T_x = T_{x+1} + L_x} \, .$$

Tento ukazatel se spočítá jako kumulace počtu žijících osob L_x od nejvyššího věku v tabulce až po námi zjišťovaný věk x:

$$\boldsymbol{T_x = L_x + L_{x+1} + L_{x+2} + \cdots + L_\omega} \, ,$$

nebo také:
$$\boldsymbol{T_x = \sum_{i=\omega-1}^{x} L_i} \, ,$$

kde ω značí nejvyšší věk v úmrtnostní tabulce (105 let) a sčítací index tedy nabývá hodnot od 104 let až po námi počítaný věk x.

Poslední z ukazatelů je tzv. naděje dožití e_x neboli střední délka života, ta udává průměrný počet let, který má osoba právě x-letá naději dožít se při zachování řádu úmrtnosti za sledované období. Střední délka života se spočítá jako:

$$\boldsymbol{e_x = \frac{T_x}{l_x}} \, .$$

Pro odstranění náhodných výkyvů pravděpodobností úmrtí q_x, jsou tyto pravděpodobnosti od věku 4 let vyrovnávány pomocí vztahu:

$$q_x^{vyrovn} = \frac{[105 \cdot q_x + 90 \cdot (q_{x-1} + q_{x+1}) + 45 \cdot (q_{x-2} + q_{x+2}) - 30 \cdot (q_{x-3} + q_{x+3})]}{315}$$

Pro vyšší věky se pravděpodobnost úmrtí vyhlazuje pomocí Gompertz-Makehamova zákona. Touto metodou vyhlazení křivky pravděpodobnosti úmrtí a aplikací na reálná data se budu podrobně zabývat v kapitole 3 a 4.

2.4 Druhy úmrtnostních tabulek

Klasifikace úmrtnostních tabulek:

- *Okamžikové úmrtnostní tabulky*

  Okamžiková úmrtnostní tabulka je založena na hypotetickém sledování současně narozených osob. Na tuto hypotetickou populaci se aplikují pravděpodobnosti úmrtí podle věku dané populace. S tímto druhem úmrtnostních tabulek se dnes setkáváme nejčastěji a také ve 3. kapitole budu vycházet právě z okamžikových úmrtnostních tabulek.

- *Generační úmrtnostní tabulky*

  Generační úmrtnostní tabulka představuje záznam skutečného průběhu života konkrétní populace současně narozených jedinců od narození až do smrti posledního z nich. Konstrukce takovéto tabulky je velice obtížná, předpokládá sledování populace v průběhu dlouhé doby. S generačními úmrtnostními tabulkami se dnes již nesetkáváme, protože jak bylo zmíněno výše, jejich konstrukce je z časového hlediska velice náročná.

Dále můžeme úmrtnostní tabulky dělit podle sledovaného věkového intervalu na:

- *Úplné úmrtnostní tabulky*

  V úplných úmrtnostních tabulkách pracujeme s věkovými intervaly o délce jednoho roku.

- *Zkrácené úmrtnostní tabulky*

  Ve zkrácených úmrtnostních tabulkách se vyskytují věkové intervaly delší než jeden rok. Zpravidla volíme věkové intervaly o délce 5 let. Např.: 0 – 4 roky, 5 – 9 let, 10 – 14 let,…atd.

Úplné úmrtnostní tabulky nám poskytují mnohem přesnější informaci o závislosti úmrtnosti na věku, proto budu ve svých výpočtech vycházet z úplných úmrtnostních tabulek.

2.5 Úmrtnostní tabulky nezávislé na pohlaví

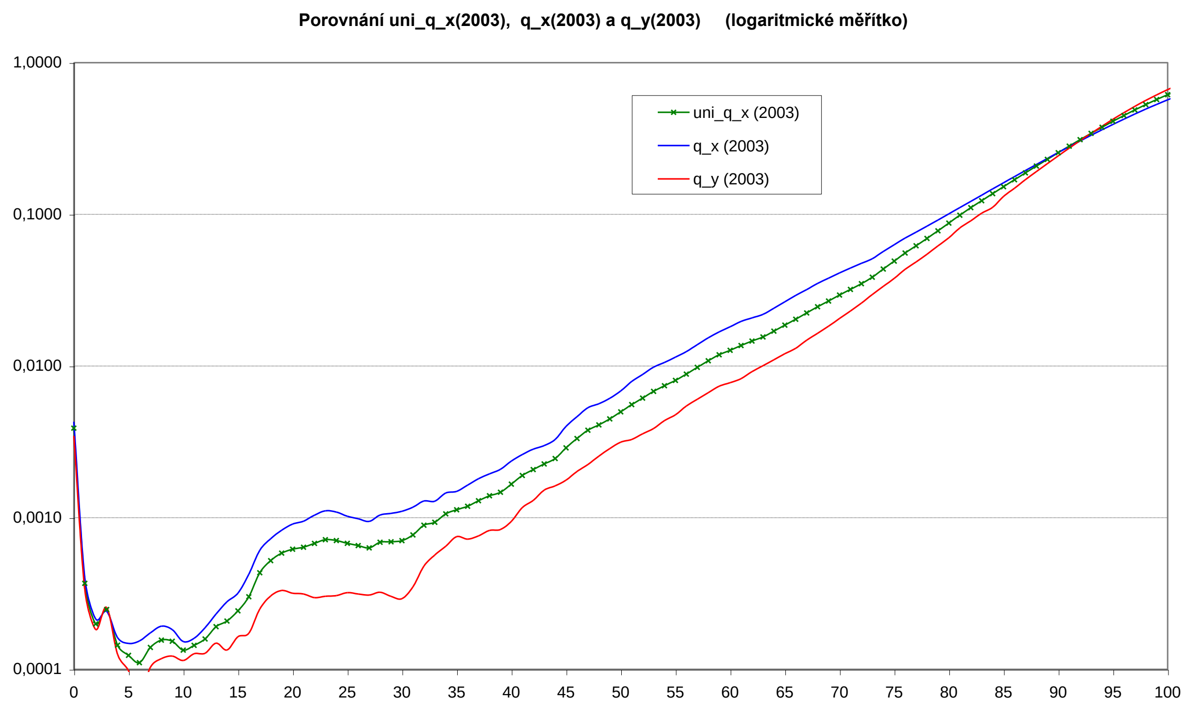
V Evropské unii již od roku 2003 probíhala diskuze týkající se diskriminace žen v oblasti přístupu ke zboží a službám. V oblasti služeb se tato problematika týká hlavně životního pojištění a penzijního připojištění, respektive ostatních finančních produktů sloužících k finančnímu zabezpečení obyvatel ve stáří. Od 21. prosince 2012 byla přijata novela o rovném přístupu, která zakazuje rozlišování pohlaví při kalkulaci pojistného a to zejména u nově uzavřených smluv týkajících se životního pojištění či penzijního připojištění.

Je obecně známo, že průměrný muž se dožívá zhruba o 7 let nižšího věku než průměrná žena, která byla ještě v loňském roce pro pojišťovnu méně riziková, proto platila za stejné životní pojištění mnohem nižší částku než muž.

Ráda bych ukázala, jak se tato novela projevila v kalkulaci pojistného životního pojištění pro případ smrti. Vezmu si proto třicetiletou ženu a třicetiletého muže, kteří chtějí uzavřít běžně placené životní pojištění pro případ smrti s pojistnou částkou 1 000 000 Kč. Muž i žena si toto pojištění sjedná do 65 let věku, u ČSOB pojišťovny [16] by před 21. 12. 2012 žena zaplatila 300 Kč, muž by zaplatil 574 Kč, avšak při nových podmínkách po 21. 12. 2012 zaplatí jak muž, tak žena 537 Kč. Znamená to, že pro ženy se od loňského prosince životní pojištění zdražilo v některých případech i o více než sto procent a to právě z toho důvodu, že pojišťovny přešly od užívání úmrtnostních tabulek pro jednotlivá pohlaví zvlášť k novým tabulkám, které jsou nezávislé na pohlaví.

Zdrojovými daty pro sestavení tabulky nezávislé na pohlaví jsou počty žijících obyvatel v České republice k 1. 7. daného roku v jednotlivých věcích za obě pohlaví, tedy P_x pro muže a P_y pro ženy, kde y značí věk ženy. Dále počty zemřelých D_x a D_y v jednotlivých věcích x, y dané populace. Získaná data za muže a ženy se sečtou a následně se spočítají jednotlivé charakteristiky podle stejných vztahů jako v kapitole 2.3. Rozdíl je pouze v metodách vyrovnání jednotlivých pravděpodobností úmrtí, ale tím se nebudu v této práci více zabývat.

Z *Obrázku 2.1* [9] můžete vidět, jak se vyvíjela úmrtnost v roce 2003 v České republice a to zvlášť u mužů – značeno q_x, u žen – q_y a potom za předpokladu nezávislosti na věku – uni_q_x.



Obrázek 2.1: Porovnání pravděpodobností úmrtí za muže, ženy a unisex.

2.6 Struktura a počty zemřelých v České republice v roce 2011 podle příčin smrti

Jak je uvedeno v [12], v roce 2011 zemřelo 106 848 tisíc osob, což byl skoro stejný počet jako v roce 2010, přičemž počet zemřelých kojenců byl nižší, a to 298 oproti 313 zemřelým kojencům v roce 2010. Naděje dožití při narození oproti roku 2010 vzrostla, u mužů o 0,3 roku na 74,7 roku a u žen vzrostla o 0,2 roku na 80,7 roku.

V níže uvedené *Tabulce 2.1* můžete vidět počty zemřelých podle jednotlivých skupin: "Novotvary", "Nemoci oběhové soustavy", "Nemoci dýchací soustavy", "Nemoci trávicí soustavy", "Vnější příčiny nemocnosti a úmrtnosti" a "Ostatní". Některé skupiny jsou dále rozčleněny do podskupin, ovšem neuváděla jsem všechny skupiny ani podskupiny, vybrala jsem pouze ty nejvíce zastoupené. Z celkových 106 848 zemřelých v roce 2011 zemřelo

54 141 mužů a 52 707 žen. Jak můžete vidět, tak nejčastější příčinou úmrtí u českých obyvatel byly jak u mužů, tak u žen nemoci oběhové soustavy, na jejichž následky zemřelo více jak polovina mužů i žen z celkového počtu zemřelých. V této skupině byly nejzastoupenější podskupinou ischemické nemoci srdeční. Dále potom v loňském roce nejvíce lidí zemřelo na novotvary, nejvíce mužů zemřelo na zhoubný novotvar průdušek či plic a nejvíce žen z této skupiny zemřelo na následky zhoubného novotvaru prsu.

Základní příčina smrti	*Novotvary*			*Nemoci oběhové soustavy*				*Nemoci dýchací soustavy*	*Nemoci trávicí soustavy*	*Vnější příčiny nemocnosti a úmrtnosti*		*Ostatní*
	Zhoubný novotvar tlustého střeva	*Zhoubný novotvar průdušky a plíce*	*Zhoubný novotvar prostaty / prsu*	*Ischemické nemoci srdeční*	*Selhání srdce*	*Cévní nemoci mozku*	*Ateroskler-óza*			*Dopravní nehody*	*Sebevraždy*	
Muži	1 154	3 907	1 314	12 844	2 003	4 306	1 432	3 253	2 562	651	1 337	4 883
Celkem	**15 180**			**24 121**				**3 253**	**2 562**	**4 142**		**4 883**
Ženy	958	1 675	1 725	13 895	2 211	6 497	2 010	2 437	1 969	220	252	5 507
Celkem	**12 359**			**28 604**				**2 437**	**1 969**	**1 831**		**5 507**

Tabulka 2.1: Nejčastější příčiny úmrtí obyvatel České republiky v roce 2011.

Pro zajímavost bych ještě v *Tabulce 2.2* uvedla počty zemřelých dětí mladších jednoho roku v České republice v roce 2011. Počty zemřelých dětí do jednoho roku jsou v České republice již dlouhodobě velice nízké, to stejné můžeme říct i o kojenecké úmrtnosti, která je u nás jedna z nejnižších ve světě. V roce 2011 zemřelo 298 kojenců, což byl historicky nejnižší počet.

	0 dnů	***0 - 6 dnů***	***7 - 27 dnů***	***0- 27 dnů***	***28-364 dnů***	***Celkem***
Počet zemřelých kojenců do 1 roku	*41*	*120*	*66*	*186*	*112*	*298*

Tabulka 2.2: Počty zemřelých kojenců v České republice v roce 2011.

Příčinou úmrtí kojenců do 1 roku života jsou nejčastěji některé stavy vzniklé v perinatálním odbodbí, což je období od 26. týdne vývoje do 4. týdne po porodu. Na tuto příčinu umírá více než polovina kojenců, v roce 2011 asi 57%, což je 170 případů z celkových

298 zemřelých kojenců v České republice v roce 2011. U 20,8%, tedy u 62 zemřelých v kojeneckém věku v České republice v roce 2011, byla příčinou vrozená vada, deformace a chromozomální abnormalita.

3 GOMPERTZ-MAKEHAMŮV ZÁKON ÚMRTNOSTI

Ve třetí kapitole se čtenář seznámí se vznikem Gompertz-Makehamova zákona úmrtnosti. Odvodím zde vztahy pro výpočet počátečních odhadů Gompertz-Makehamovy funkce a hlavním úkolem této kapitoly bude aplikovat Gompertz-Makehamovu funkci na reálná česká data. Při zpracování třet kapitoly jsem čerpala zejména z publikací [3], [4], [5], [6], [7], z internetových zdrojů [13], [15] a k výpočtům jsem použila data dostupná z [17], [18].

3.1 Vznik Gompertz-Makehamova zákona úmrtnosti

V roce 1825 vyslovil B. Gompertz hypotézu, že síla úmrtnosti je závislá na dvou činitelích. První činitel nezávisí na věku a ogranismu a druhý činitel je přímou funkcí věku. Tento zákon omezil na věkové období 10 - 60 let, kdy síla úmrtnosti roste geometrickou progresí v závislosti na věku. Podle Gompertzova zákona úmrtnost exponenciálně narůstá s věkem a míra úmrtnosti je vyjádřena jako:

$$\boldsymbol{\mu_x = B \cdot C^x},$$

kde pro Gompertzovy parametry B a C platí: $B > 0$; $C > 1$; μ_x značí intenzitu úmrtnosti ve věku x.

W. M. Makeham v roce 1860 doplnil tento Gompertzův vzorec o konstantu, Makehamův parameter A, který by měl vystihovat vliv faktorů nezávislých na věku (např. epidemie).

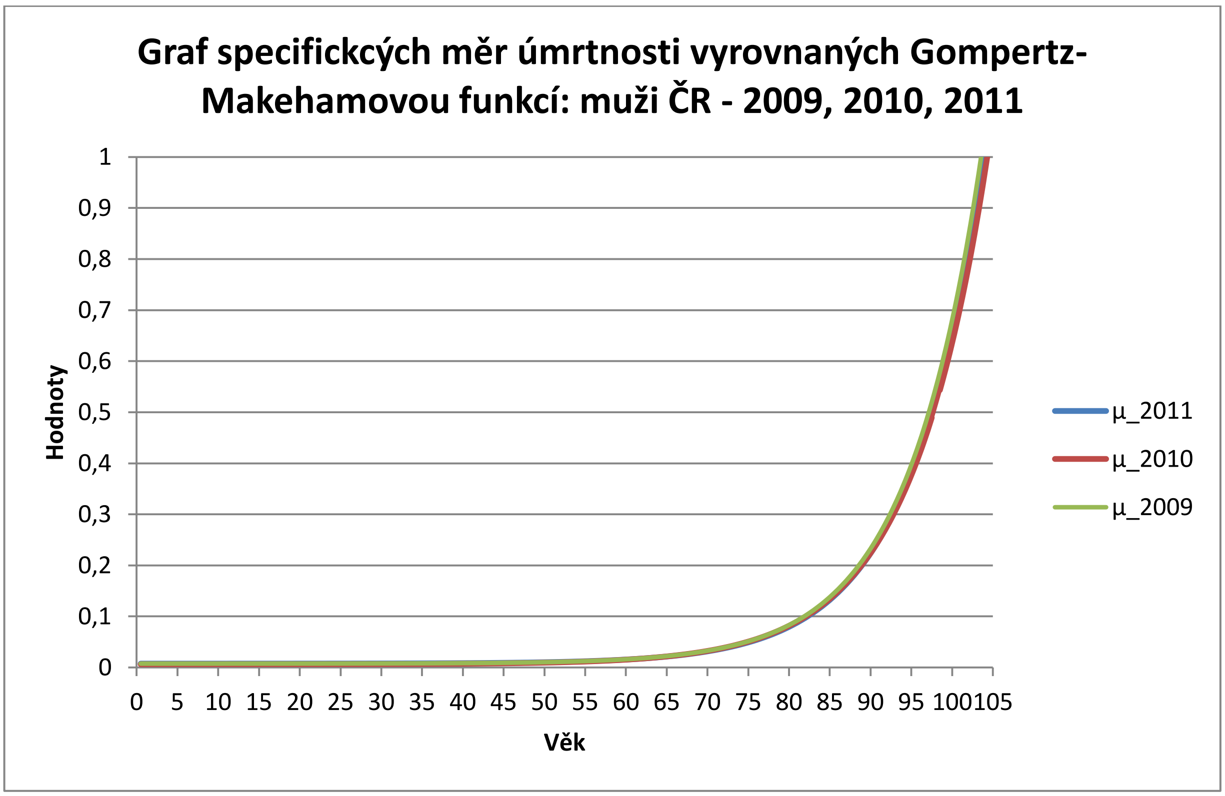
Na základě Gompertzova zákona doplněného o Makehamův parameter byl zformulován vztah, který je označován jako Gompertz- Makehamův zákon. Tento zákon říká, že míra úmrtnosti je součtem věkově nezávislé složky (Makehamův zákon) a složky, která exponenciálně roste s věkem (Gompertzův zákon). Gompertz-Makehamův zákon lze vyjádřit jako:

$$\boldsymbol{\mu_x = A + B \cdot C^x},$$

kde $B > 0$; $C > 1$.

Podmínka pro parametr A chybí. Jak je uvedeno v [5], tak A by mělo být kladné, ovšem odhady z empirických dat vycházející zhruba od poloviny 60. let pro české muže v záporných hodnotách parametru A. Tato skutečnost by se dala vysvětlit jako důsledek pokroku medicíny, která dokáže zachránit život těm, kteří by dříve zemřeli, tedy jako opak náhodných úmrtí, jako náhodná zachránění života oproti letem před 60. rokem 20. století. O tom se budeme moci přesvědčit v kapitole 3.2.5.

Z *Grafu 3.1* můžete vidět, jak by vypadaly hodnoty μ_x specifických měr úmrtnosti vyrovnané Gompertz-Makehamovou funkcí v české populaci mužů v letech 2009-2011.



Graf 3.1: Specifické míry úmrtnosti vyrovnané Gompertz-Makehamovou funkcí v české populaci mužů v letech 2009-2011.

Tento Gompertz-Makehamův zákon patří mezi parametrické modely úmrtnosti. Představuje užitečný nástroj v demografických a aktuárských prognózách úmrtnosti. Model se snaží co nejlépe popsat vztah mezi úmrtností a věkem a to především od 60. roku života.

Jak uvádí [15], ve zvláštních případech může být věkově nezávislá složka (parametr A) zanedbána. Například v současné době v průmyslově vyspělých státech nebo v dobrých laboratorních podmínkách, kdy roste celková intenzita úmrtnosti exponenciálně s věkem, tedy pouze podle Gompertzova zákona.

3. 2 Vyrovnání specifických měr úmrtnosti pomocí Gompertz-Makehamova zákona na reálných datech české mužské populace z roku 2011

V této kapitole budu teoretické poznatky o Gompertz-Makehamově zákoně aplikovat na reálnou českou populaci, konkrétně na hodnoty D_x a P_x z úmrtnostních tabulek z roku 2011, 2010, 2009, zvlášť na mužskou a ženskou populaci. Za pomoci MS Excel 2010 se potom budu snažit najít optimální parametry tohoto modelu.

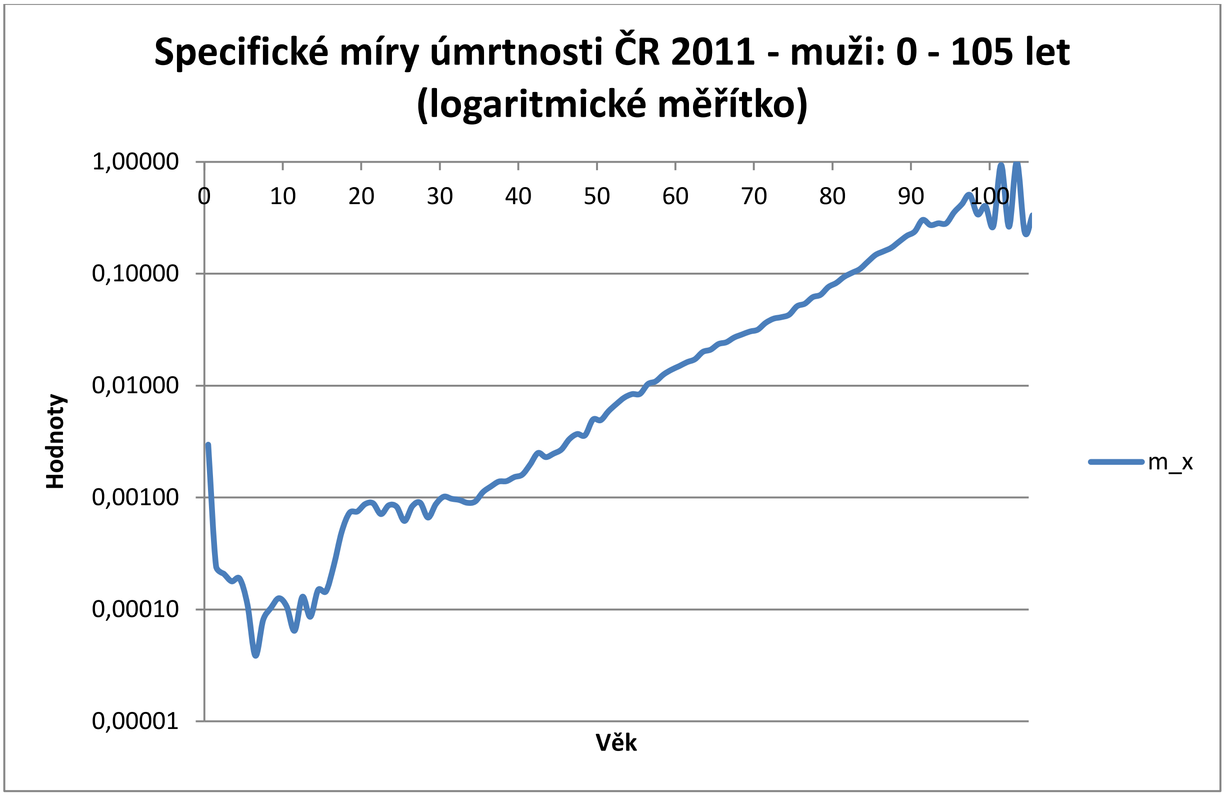
3.2.1 Odvození parametrů Gompertz-Makehamovy funkce

Z webových stránek Českého statistického úřadu [17] jsem si stáhla mužské úmrtnostní tabulky za rok 2011. Při výpočtu jednotlivých specifických měr úmrtnosti budu vycházet z dat pro reálně zemřelé x-leté muže v České republice v roce 2011 a ze středního stavu x-letých můžu v České republice v roce 2011. Jak jsem již uvedla v kapitole 2.3, specifické míry úmrtnosti m_x jsem počítala podle vztahu:

$$\boldsymbol{m_x = \frac{D_x}{P_x}}.$$

Vypočtené specifické míry úmrtnosti můžete vidět v přiloženém excelovském souboru "*muzi_2011.xlsx"* v listu "*Data_muzi_2011*" ve sloupci *L*.

Když se podíváme na průběh úmrtnosti z *Grafu 3.2* specifických měr úmrtnosti m_x (pro lepší přehlednost jsem nechala specifické míry úmrtnosti m_x vykreslit v logaritmickém měřítku), můžeme vidět, že se skládá z několika etap vývoje organismu. Z počátku můžeme vidět, že dochází k vysoké úmrnosti. Příčinou mohou být, jak uvádí [13], dětské nemoci, a jak jsem již uvedla v kapitole 2.6, tak u kojenců do 1 roku života bývají příčinou některé stavy vzniklé v perinatálním období. Poté následuje období poměrně nízké úmrnosti a úmrtnost se začíná opět zvyšovat až před začátkem dospívání. Po období dospívání následuje období konstantní úmrtnosti. Asi od 35. roku života až zhruba do 92. roku života úmrtnost exponenciálně vzrůstá. Jak můžeme vidět z *Grafu 3.2*, tak zhruba od věku 92 let se začínají hodnoty odchylovat od exponenciálního růstu předchozích hodnot. To může být způsobeno tím, že počty žijících v těchto vysokých věcích jsou už poměrně malé a tak můžou být specifické míry úmrtnosti zatíženy velkou náhodnou chybou. Podrobněji se budu touto problematikou zabývat v kapitole 4.



Graf 3.2: Specifické míry úmrtnosti mužů české populace za rok 2011 v logaritmickém měřítku.

Abych eliminovala tyto náhodné chyby, budu se snažit vyrovnat hodnoty specifických měr. Nejvhodnější funkcí pro vyrovnání hodnot specifických měr úmrtnosti je právě Gompertz-Makehamova funkce, kterou jsem již podrobněji rozebrala v kapitole 3.1. Jak později uvidíte z grafů, tak vyrovnání touto funkcí není vhodné pro celé věkové rozmezí, ale věrně kopíruje specifické míry úmrtnosti až od věku 60 let zhruba do věku 85 let. Gompertz-Makehamovu funkci lze vyjádřit jako:

$$\boldsymbol{\mu_x = A + B \cdot C^x},$$

kde $B > 0$, $C > 1$.

Je třeba provést odhad neznámých parametrů $A, B, C.$ Kdyby Gompertz-Makehamova funkce neobsahovala konstantu A, tedy kdyby se jednalo o původní Gompertzovu funkci, mohla bych ji logaritmováním transformovat na lineární funkci a odhad parametrů provést pomocí lineární regrese, v tomto případě však funkce vhodně transformovat nejde, takže budu muset postupovat při odhadu parametrů jiným způsobem.

Nejprve si s pomocí tří stejně dlouhých, na sebe navazujících, věkových intervalů o délce k stanovím počáteční odhady parametrů. Jednotlivé intervaly budou vypadat následovně:

$$\boldsymbol{x_0 - (x_0 + k - 1)},$$

$$\boldsymbol{(x_0 + k) - (x_0 + 2k - 1)},$$

$$\boldsymbol{(x_0 + 2k) - (x_0 + 3k - 1)},$$

kde x_0 značí počátek prvního z intervalu a k je délka intervalu. Nyní si vytvořím soustavu tří rovnic a z ní si vyjádřím parametry $A,\ B,\ C$. Soustavu tří rovnic získám tak, že si do Gompertz-Makehamovy funkce postupně dosadím všechny hodnoty věku z prvního věkového intervalu a dostanu k rovnic, které sečtu. První rovnice po dosazení věkového intervalu $x_0 - (x_0 + k - 1)$ bude vypadat následovně:

$$\boldsymbol{\sum_{x=\,x_0}^{x_0+k-1} m_x = \sum_{x=\,x_0}^{x_0+k-1} (A + B \cdot C^x)}.$$

To stejné udělám i pro druhý a třetí věkový interval:

$$\boldsymbol{\sum_{x=\,x_0+k}^{x_0+2k-1} m_x = \sum_{x=\,x_0+k}^{x_0+2k-1} (A + B \cdot C^x)};$$

$$\sum_{x=x_0+2k}^{x_0+3k-1} \boldsymbol{m_x} = \sum_{x=x_0+2k}^{x_0+3k-1} (\boldsymbol{A} + \boldsymbol{B} \cdot \boldsymbol{C}^{\boldsymbol{x}}).$$

Součty specifických měr úmrtnosti na levých stranách rovnice označím písmeny R_1, R_2, R_3. Pravé strany rovnic upravím a dostanu:

$$\boldsymbol{R_1 = k \cdot A + B \cdot C^{x_0} \cdot (1 + C + C^2 + \ldots + C^{k-1});}$$

$$\boldsymbol{R_2 = k \cdot A + B \cdot C^{x_0+k} \cdot (1 + C + C^2 + \ldots + C^{k-1});}$$

$$\boldsymbol{R_3 = k \cdot A + B \cdot C^{x_0+2k} \cdot (1 + C + C^2 + \ldots + C^{k-1}).}$$

Nyní se budu chtít zbavit parametru A, proto od rovnice R_3 odečtu rovnici R_2 a od rovnice R_2 odečtu rovnici R_1 :

$$\boldsymbol{R_3 - R_2 = B \cdot C^{x_0+k} \cdot \left(C^k - 1\right) \cdot \left(1 + C + C^2 + \cdots + C^{k-1}\right);}$$

$$\boldsymbol{R_2 - R_1 = B \cdot C^{x_0} \cdot \left(C^k - 1\right) \cdot \left(1 + C + C^2 + \cdots + C^{k-1}\right).}$$

Dále se budu chtít zbavit i parametru B, abych si nejprve mohla vyjádřit parametr C, proto dám zbylé dvě rovnice do poměru a dostanu:

$$\boldsymbol{\frac{R_3 - R_2}{R_2 - R_1} = \frac{B \cdot C^{x_0+k} \cdot \left(C^k - 1\right) \cdot \left(1 + C + C^2 + \cdots + C^{k-1}\right)}{B \cdot C^{x_0} \cdot (C^k - 1) \cdot (1 + C + C^2 + \cdots + C^{k-1})},}$$

a po vykrácení dostavám:

$$\boldsymbol{\frac{R_3 - R_2}{R_2 - R_1} = C^k.}$$

Parametr C vypočítám jako k-tou odmocninu z výše uvedeného výrazu:

$$\boldsymbol{C = \sqrt[k]{\frac{R_3 - R_2}{R_2 - R_1}}.}$$

Nyní si z rovnice $R_2 - R_1 = B \cdot C^{x_0} \cdot (C^k - 1) \cdot (1 + C + C^2 + \cdots + C^{k-1})$ vyjádřím parametr B:

$$\boldsymbol{B} = \frac{\boldsymbol{R_2} - \boldsymbol{R_1}}{\boldsymbol{C^{x_0}} \cdot (\boldsymbol{C^k} - \mathbf{1}) \cdot (\mathbf{1} + \boldsymbol{C} + \boldsymbol{C^2} + \cdots + \boldsymbol{C^{k-1}})}.$$

Tento vzorec ještě dále zjednoduším, jelikož součet v čitateli $(1 + C + C^2 + \cdots + C^{k-1})$ je součtem prvních k členů geometrické posloupnosti, kde první člen $a_1 = 1$ a kvocient $q = C$. V podmínkách Gompertz-Makehamovy funkce jsem výše uvedla, že pro parameter C musí platit: $C > 1$, proto budu počítat součet prvních k členů posloupnosti pro $q \neq 1$ podle vztahu:

$$\boldsymbol{s_n} = \boldsymbol{a_1} \cdot \frac{\boldsymbol{q^n} - \mathbf{1}}{\boldsymbol{q} - \mathbf{1}}.$$

Parametr B potom vyjádřím podle následujícího vztahu:

$$\boldsymbol{B} = \frac{\boldsymbol{R_2} - \boldsymbol{R_1}}{\boldsymbol{C^{x_0}} \cdot (\boldsymbol{C^k} - \mathbf{1}) \cdot \frac{\boldsymbol{C^{k}}-\mathbf{1}}{\boldsymbol{C}-\mathbf{1}}}.$$

Dále si z rovnice $R_1 = k \cdot A + B \cdot C^{x_0} \cdot (1 + C + C^2 + \ldots + C^{k-1})$ vyjádřím parameter $\boldsymbol{A}$:

$$\boldsymbol{A} = \frac{\boldsymbol{R_1} - \boldsymbol{B} \cdot \boldsymbol{C^{x_0}} \cdot (\mathbf{1} + \boldsymbol{C} + \boldsymbol{C^2} + \ldots + \boldsymbol{C^{k-1}})}{\boldsymbol{k}}.$$

Provedu stejnou úpravu součtu prvních k členů geometrické posloupnosti jako u vztahu pro parametr B, potom bude vztah pro vyjádření parametru A :

$$\boldsymbol{A} = \frac{\boldsymbol{R_1} - \boldsymbol{B} \cdot \boldsymbol{C^{x_0}} \cdot \frac{\boldsymbol{C^{k}}-\mathbf{1}}{\boldsymbol{C}-\mathbf{1}}}{\boldsymbol{k}}.$$

Charakteristiky úmrtnosti jsou počítány k dokončeným věkům, jak uvádí [3], tím se rozumí jednoletý věkový interval přesného věku a jelikož jsou charakteristiky úmrtnosti charakteristikami intervalovými, tak pro správné zobrazení v grafech XY bodových budu zadávat hodnotu nezávisle proměnné X (v mém případě se bude jednat o věk) nikoliv jako počátek, ale jako střed daného intervalu, proto za hodnoty věku budu dosazovat hodnoty o 0,5 vyšší než je příslušný dokončený věk, např.: místo věku 60, budu brát v potaz věk 60,5 let. Gompertz-Makehamova funkce bude mít v tomto případě tvar:

$$\boldsymbol{\mu_x} = \boldsymbol{A} + \boldsymbol{B} \cdot \boldsymbol{C^{x+0,5}}.$$

Výše odvozené vztahy pro odhady parametrů A, B a C upravím tak, že si v jednotlivých věkových intervalech k věkům x_0 přičtu hodnotu 0,5. Vztahy pro odhady parametrů budou tedy vypadat následovně:

$$\boldsymbol{C = \sqrt[k]{\frac{R_3 - R_2}{R_2 - R_1}}};$$

$$\boldsymbol{B = \frac{R_2 - R_1}{C^{x_0+0,5} \cdot (C^k - 1) \cdot \frac{C^k-1}{C-1}}} \;;$$

$$\boldsymbol{A = \frac{R_1 - B \cdot C^{x_0+0,5} \cdot \frac{C^k-1}{C-1}}{k}}.$$

Nyní se s pomocí výše odvozených vztahů budu snažit určit počáteční odhady parametrů Gompertz-Makehamovy funkce.

3.2.2 Počáteční odhad parametrů Gompertz-Makehamovy funkce

Za x_0, tedy za počátek prvního z věkových intervalů, si zvolím hodnotu 60 let, jak uvidíte později z grafů, tak Gompertz-Makehamova funkce věrně aproximuje průběh empirických specifických měr úmrtnosti až od věku 60 let. Za k, tedy za délku věkového intervalu jsem si pro začátek stanovila hodnotu $k = 10$. Dostala jsem intervaly:

$60 - 69$ let,

$70 - 79$ let,

$80 - 89$ let.

Za délku věkového intervalu k se později budu snažit dosadit takovou hodnotu, která bude minimalizovat součet vážených čtverců odchylek, protože právě co nejmenší součet vážených čtverců odchylek je vhodný k co nejvěrnější aproximaci.

První si spočítám hodnoty rovnic R_1, R_2a R_3 , kde:

$$\boldsymbol{R_1} = \sum_{x=x_0}^{x_0+k-1} \boldsymbol{m_x}\,; \quad \boldsymbol{R_2} = \sum_{x=x_0+k}^{x_0+2k-1} \boldsymbol{m_x}\,; \quad \boldsymbol{R_3} = \sum_{x=x_0+2k}^{x_0+3k-1} \boldsymbol{m_x}\,.$$

Vypočtené hodnoty součtů empirických specifických měr úmrtnosti v jednotlivých intervalech můžete vidět v přiloženém excelovském souboru v buňkách *D123 – F123*.

x_0	k	R_1	R_2	R_3
60	10	0,22364	0,49999	1,40519

Nyní si podle výše uvedených vztahů odhadnu počáteční odhady Gompertz-Makehamovy funkce. V buňkách *I123 – K123* můžete vidět vypočtené hodnoty:

Počáteční odhady parametrů		
A	B	C
0,0102203	0,0000051	1,1259767

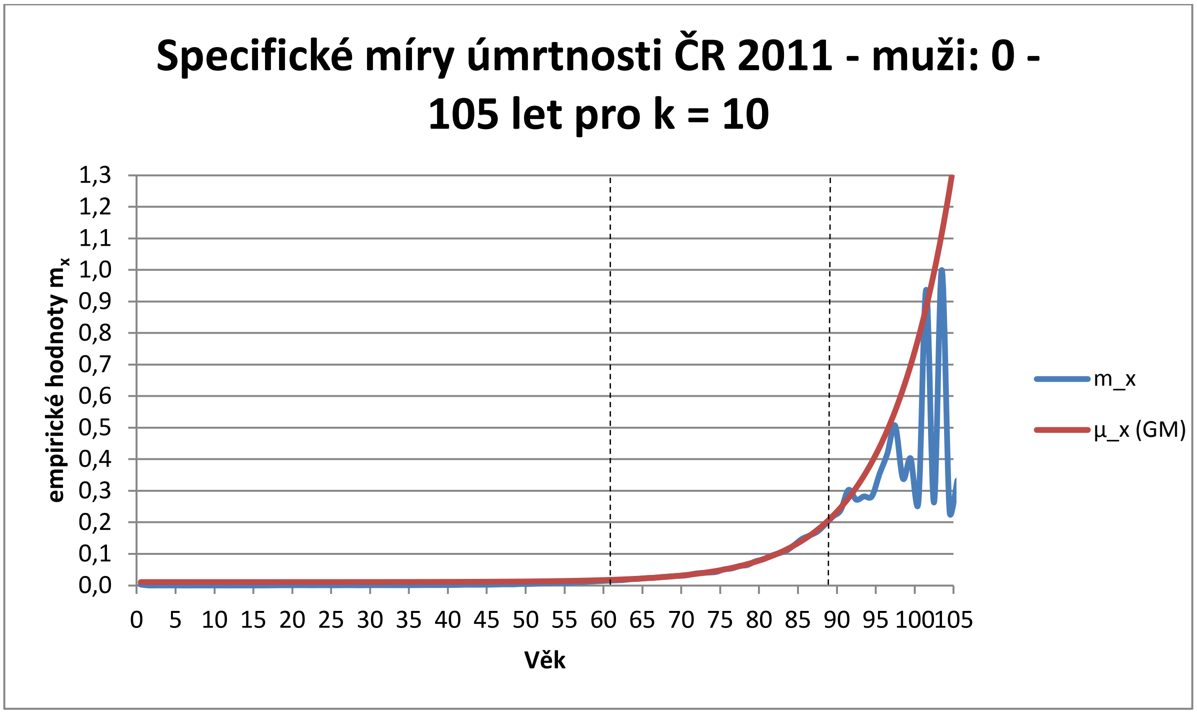
Tyto vypočtené odhady parametrů jsou pouze počátečním odhadem. Dále v kapitole 3.2.3 se budu snažit nalézt takovou délku věkového intervalu k, která bude dávat minimální součet vážených čtverců odchylek.

Výpočet vyrovnaných specifických měr úmrtnosti pomocí Gompertz-Makehamovy funkce budu provádět podle vztahu:

$$\boldsymbol{\mu_x = A + B \cdot C^{x+0,5}}\,.$$

Vypočtené hodnoty můžete vidět v přiloženém excelovském souboru ve sloupci *M* a v následujícím grafu.

Z *Grafu 3.3* můžeme vidět, že Gompertz-Makehamova funkce s odhadnutými parametry při počáteční délce věkového intervalu $k = 10$ dobře modeluje úmrtnost zejména ve věku 60 - 89 let. Z grafu je patrné, že pro vyšší věky je již úmrtnost nadhodnocována.



Graf 3.3: Specifické míry úmrtnosti mužů z ČR za rok 2011, porovnání vyrovnaných a původních hodnot specifických měr úrmtnosti pro počáteční délku intervalu $k = 10$.

3.2.3 Odhad nejvhodnější délky věkového intervalu k

V této kapitole se budu snažit určit délku věkového intervalu k tak, aby byl součet čtverců odchylek co možná nejnižší. Nejprve si zavedu určité předpoklady nutné k určení vztahu pro součet vážených čtverců odchylek.

Předpokládejme, že úmrtí lidí v populaci jsou nezávislá a stejně rozdělená, potom můžu celkový počet zemřelých D_x ve věku x považovat za náhodnou veličinu s binomickým rozdělením s parametry P_x, μ_x.

Binomické rozdělení, jak je uvedeno v [6], s parametry P_x, μ_x má náhodná veličina D_x, která nabývá hodnot $j = 0, 1, 2, \dots, P_x$ s pravděpodobnostmi:

$$\boldsymbol{p_{xj} = P(D_x = j) = \binom{P_x}{j} \cdot \mu_x^j \cdot (1 - \mu_x)^{P_x - j}},$$

kde $j = 0, 1, 2, \dots, P_x$.

$\mu_x \in (0,1), P_x \in \mathrm{N}$ jsou tzv. parametry tohoto rozdělení. Zkráceně budu označovat:

$$\boldsymbol{D_x \sim Bi(P_x\,;\, \mu_x)},$$

kde P_x je střední stav populace ve věku x a μ_x jsou hodnoty specifických měr úmrtnosti vyrovnané pomocí Gompertz-Makehamovy funkce.

Střední hodnota náhodné veličiny $\boldsymbol{D_x}$ je rovna:

$$\boldsymbol{E(D_x)} = \sum_{j=0}^{P_x} j \cdot \binom{P_x}{j} \cdot \mu_x^j \cdot (1-\mu_x)^{P_x-j} = \sum_{j=1}^{P_x} \frac{P_x!}{(j-1)! \cdot (P_x-j)!} \cdot \mu_x^j \cdot (1-\mu_x)^{P_x-j} =$$

$$= P_x \cdot \mu_x \sum_{j=1}^{P_x} \binom{P_x-1}{j-1} \cdot \mu_x^{j-1} \cdot (1-\mu_x)^{P_x-j} =$$

$$= P_x \cdot \mu_x \sum_{i=0}^{P_x-1} \binom{P_x-1}{i} \cdot \mu_x^{i} \cdot (1-\mu_x)^{P_x-1-i} =$$

$$= P_x \cdot \mu_x \cdot [\mu_x + (1-\mu_x)]^{P_x-1} = \boldsymbol{P_x \cdot \mu_x}.$$

Ve výše uvedeném výpočtu jsem zvolila substituci $\left| \begin{smallmatrix} i=j-1 \Rightarrow j=i+1 \\ P_x-j=P_x-(i+1)=P_x-i-1 \end{smallmatrix} \right|$.

Rozptyl $var\ (D_x)$ náhodné veličiny D_x vypočtu za pomoci vztahu:

$$\boldsymbol{E\left({D_x}^2\right) = E[D_x \cdot (D_x - 1)] + E(D_x)}.$$

$$\boldsymbol{E[D_x \cdot (D_x - 1)]} =$$

$$= \sum_{j=0}^{P_x} j \cdot (j-1) \cdot \binom{P_x}{j} \cdot \mu_x^j \cdot (1-\mu_x)^{P_x-j} = \sum_{j=2}^{P_x} \frac{P_x\,!}{(j-2)! \cdot (P_x-j)!} \cdot \mu_x^j \cdot (1-\mu_x)^{P_x-j} =$$

$$= P_x \cdot (P_x-1) \cdot \mu_x^2 \sum_{j=2}^{P_x} \frac{(P_x-2)!}{(j-2)! \cdot (P_x-j)!} \cdot \mu_x^{j-2} \cdot (1-\mu_x)^{P_x-j} =$$

$$= P_x \cdot (P_x-1) \cdot \mu_x^2 \sum_{i=0}^{P_x-2} \binom{P_x-2}{i} \cdot \mu_x^{i} \cdot (1-\mu_x)^{(P_x-2)-i} =$$

$$= P_x \cdot (P_x - 1) \cdot \mu_x^2 \cdot [\mu_x + (1-\mu_x)]^{P_x - 2} = \boldsymbol{P_x \cdot (P_x - 1) \cdot \mu_x^2}.$$

Ve výše uvedeném výpočtu jsem zvolila substituci $\left| \substack{i = j-2 \Rightarrow j = i+2 \\ P_x - j = P_x - (i+2) = P_x - i - 2 = (P_x - 2) - i} \right|$.

Potom $var\,(D_x) = E(D_x^2) - [E(D_x)]^2 = [P_x \cdot (P_x - 1) \cdot \mu_x^2 + P_x \cdot \mu_x]$- $(P_x \cdot \mu_x)^2$=

$$= \boldsymbol{P_x \cdot \mu_x \cdot (1 - \mu_x)}$$

Při použití centrálních limitních vět můžu říct, že binomická náhodná veličina D_x, tedy počty zemřelých, mají přibližně normální rozdělení s parametry:

$$\boldsymbol{D_x \sim N\left(P_x \mu_x;\ P_x \mu_x (1 - \mu_x)\right)}.$$

Po vydělení předchozího vztahu středním stavem populace P_x dostávám:

$$\boldsymbol{\frac{D_x}{P_x} = m_x \sim \left(\frac{P_x \mu_x}{P_x};\ \frac{\mu_x (1 - \mu_x)}{P_x}\right)},$$

po úpravě potom dostanu:

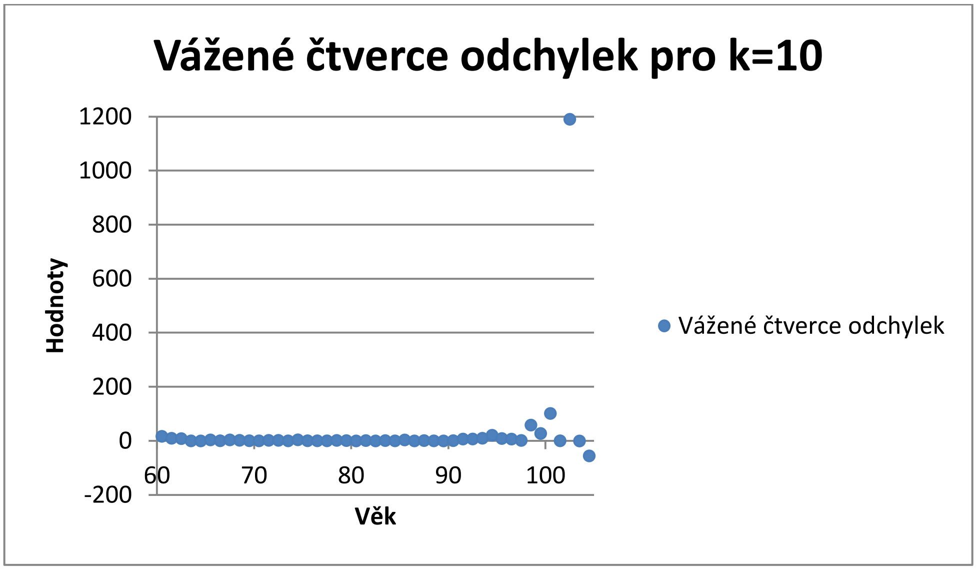
$$\boldsymbol{\frac{D_x}{P_x} = m_x \sim \left(\mu_x;\ \frac{\mu_x (1 - \mu_x)}{P_x}\right)}.$$

Na základě výše uvedených úvah budu součet vážených čtverců odchylek počítat, jak uvádí [3], podle vztahu:

$$\boldsymbol{\sum_{x=60}^{z} \frac{P_x \cdot (m_x - \mu_x)^2}{\mu_x \cdot (1 - \mu_x)}}.$$

Hodnoty vážených čtverců odchylek budu počítat od věku x_0, tedy od věku 60 let až po věk z, který stanovím odhadem podle grafu hodnot vážených čtverců odchylek tak, že se podívám, kde začnou hodnoty vážených čtverců odchylek víc kolísat. Vypočtené hodnoty můžete vidět v přiloženém excelovském souboru ve sloupci *N*.

Jak můžete vidět z *Grafu 3.4*, tak hodnoty vážených čtverců odchylek od zhruba 98 let života začínají až moc kolísat, proto budu součet vážených čtverců počítat pro věkové rozmezí 60 – 97 let.



Graf 3.4: Hodnoty vážených čtverců odchylek pro délku věkového intervalu k=10.

Součet čtverců odchylek pro délku věkového intervalu $k = 10$ vyšel 132,72. Nyní se budu snažit změnit délku věkového intervalu tak, aby mi vyšel co nejmenší součet čtverců odchylek. Postupně jsem zkoušela měnit délky intervalu k a nejmenší součet čtverců odchylek 105,26 vyšel pro délku věkového intervalu $k = 8$.

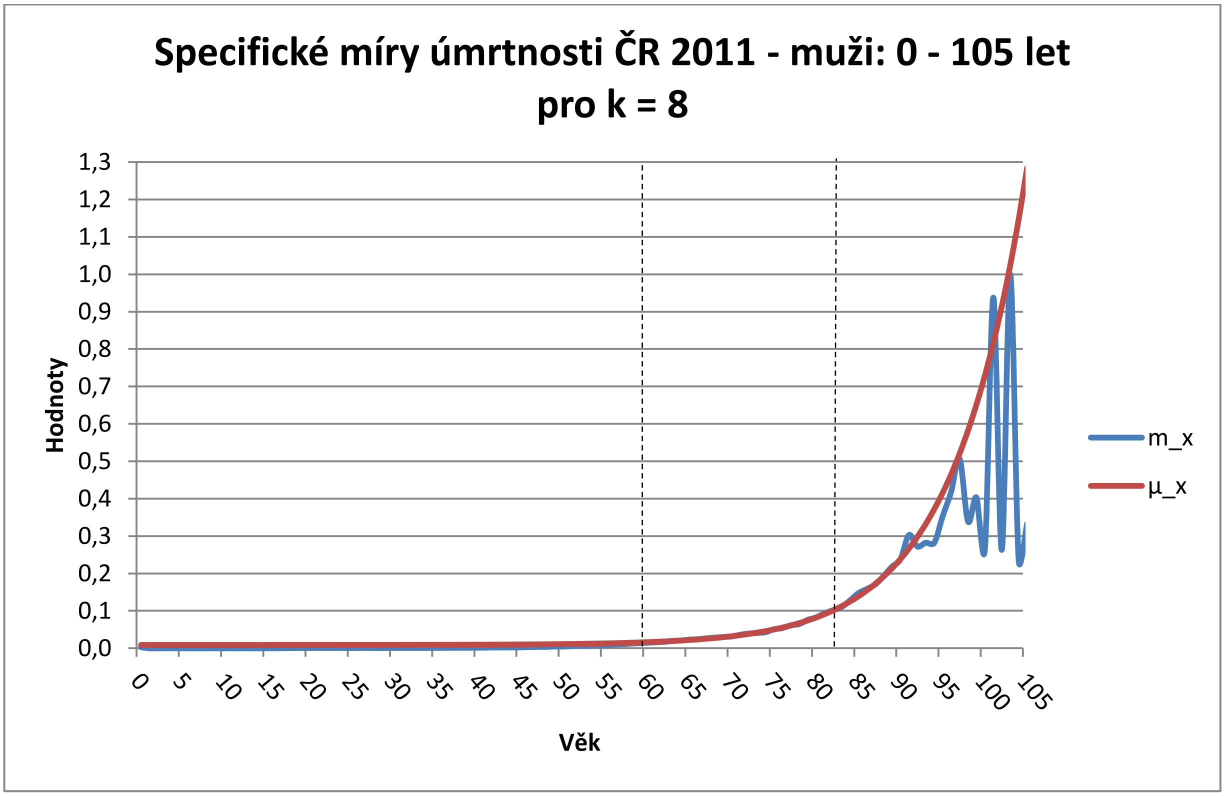
Pro $k = 8$ dostávám věkové intervaly:

$$60 - 67 \text{ let,}$$

$$68 - 75 \text{ let,}$$

$$76 - 83 \text{ let.}$$

Z *Grafu 3.5* můžete vidět porovnání specifických měr úmrtnosti u mužů české populace za rok 2011 získaných empirickým měřením s vyrovnanými hodnotami těchto měr za pomocí Gompertz-Makehamovy funkce pro délku věkového intervalu $k = 8$. Z grafu můžeme vyčíst, že při upravené délce věkového intervalu $k = 8$ vyrovnané hodnoty Gompertz-Makehamovy funkce dobře modelují úmrtnost zejména ve věku 60 – 83 let.



Graf 3.5: Specifické míry úmrtnosti mužů z ČR za rok 2011, porovnání vyrovnaných a původních hodnot specifických měr úrmtnosti pro upravenou délku intervalu k = 8.

Se změnou délky věkového intervalu k se změnily i hodnoty součtů specifických měr úmrtnosti m_x označené jako R_1, R_2 a R_3:

x_0	k	R_1	R_2	R_3
60	8	0,16446	0,30223	0,64581

A hodnoty parametrů A, B a C jsou pro délku věkového intervalu $k = 8$ rovny:

Počáteční odhady parametrů		
A	B	C
0,0090301	0,0000074	1,1210120

3.2.4 Optimalizace odhadů parametrů Gompertz-Makehamovy funkce

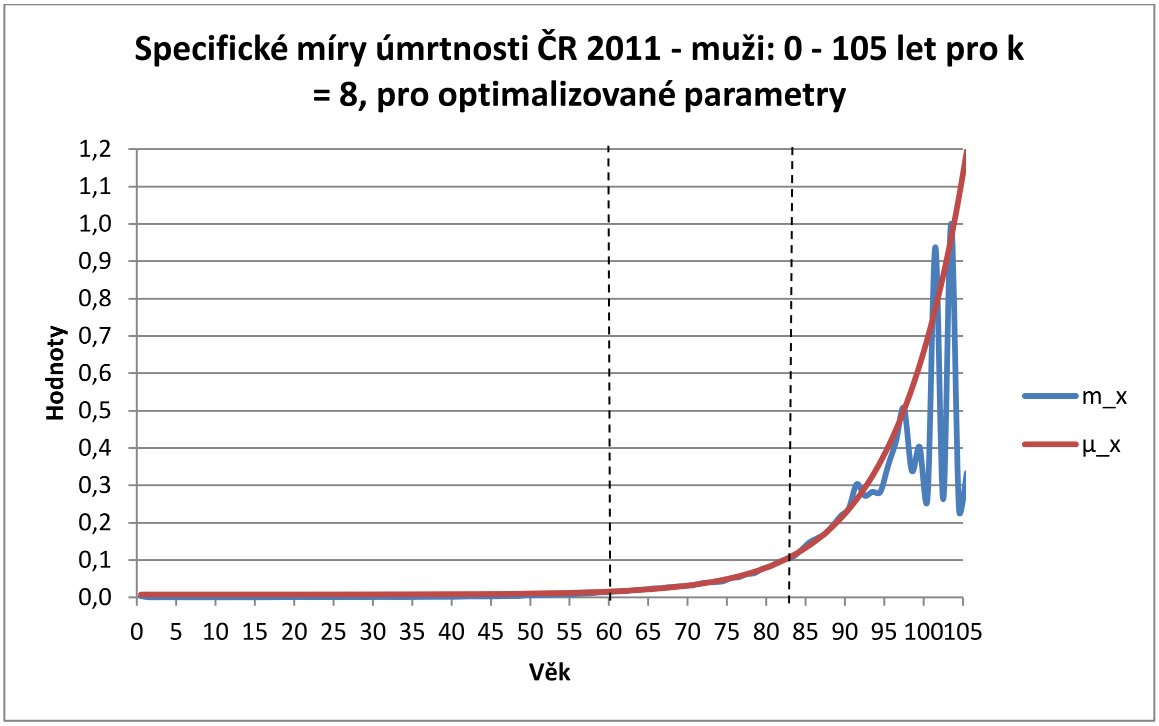
Nyní budu výše uvedené parametry A, B a C optimalizovat za pomoci funkce Řešitel tak, abych součet čtverců 105,26 minimalizovala ještě víc. Optimalizaci provedu do jiného excelovského souboru: *"2011_muzi_optimalizace.xlsx"*, protože po optimalizaci funkcí Řešitel již budou hodnoty parametrů zadány „na pevno", nikoliv jako vzorce a pro lepší přehlednost bude vhodné zachovat i původní hodnoty, kde můžu stle měnit délku věkového intervalu k. Jelikož se hodnoty parametrů liší až o 6 řádů a funkce Řešitel by nemusela správně minimalizovat hodnoty s takovým rozdílem řádů, hodnoty parametrů A a B si převedu na řád jednotek, stejně jako je parametr C. Ve sloupci μ_x, hodnot vyrovnaných Gompertz-Makehamovou funkcí, vydělím parameter A číslem 1000 a parametr B číslem 1 000 000. Nyní jsou všechny tři parametry vyjádřeny v řádu jednotek. Dále si do funkce Řešitel zadám podmínky pro minimalizaci parametrů:

$$B \geq 0;\ C \geq 1 \ \text{ a } \ \mu_{97,5} \leq 1.$$

Funkce Řešitel minimalizovala součet čtverců na hodnotu 94,56 a po přepočtení parametrů zpět do správných řádů vychází parametry Gompertz-Makehamovy funkce:

Optimalizované odhady parametrů		
A	B	C
0,0074686	0,0000109	1,1161854

Všechny hodnoty μ_x po optimalizaci parametrů najdete v excelovském souboru *"2011_muzi_optimalizace.xlsx"* v listu *„Data_muzi_2011"* ve sloupci *M*. Pro lepší přehlednost přehlednost přikládám i graf:



Graf 3.6: Specifické míry úmrtnosti mužů z ČR za rok 2011, porovnání vyrovnaných optimalizovaných a původních hodnot specifických měr úrmtnosti pro $k = 8$.

Při porovnání *Grafu 3.6* s *Grafem 3.5* před nalezením optimalizovaných parametrů můžeme vidět, že hodnoty μ_x se změnily minimálně. Rozdíly hodnot μ_x před optimalizací a po optimalizaci se pohybují zhruba v řádech setin až statisícin. Ale i přesto se součet čtverců odchylek po optimalizaci zmenšil o více než 10 jednotek.

Pokud porovnáme hodnoty specifických měr úmrtnosti m_x s vyrovnanými hodnotami μ_x při optimálních parametrech, můžeme vidět, že modelové hodnoty μ_x korespondují s hodnotami m_x zjištěnými empirickým měřením, což je patrné i z *Grafu 3.6.*

3.2.5 Porovnání výsledků obou pohlaví za roky 2009 - 2011

Výše uvedené postupy jsem aplikovala na hodnoty hodnoty D_x a P_x z úmrtnostních tabulek mužů české populace za roky 2010 a 2009 a na data z úmrtnostních tabulek žen české populace za stejné roky, tedy 2011, 2010 a 2009. Přičemž součty vážených čtverců odchylek jsem vždy počítala do věku 97 let. V následujícím odstavci budu interpretovat výsledky, ke kterým jsem došla. Veškeré výpočty a grafy k jednotlivým pohlavím a rokům 2009 – 2011 můžete vidět v přiložených excelovských souborech pojmenovaných podle pohlaví a roku.

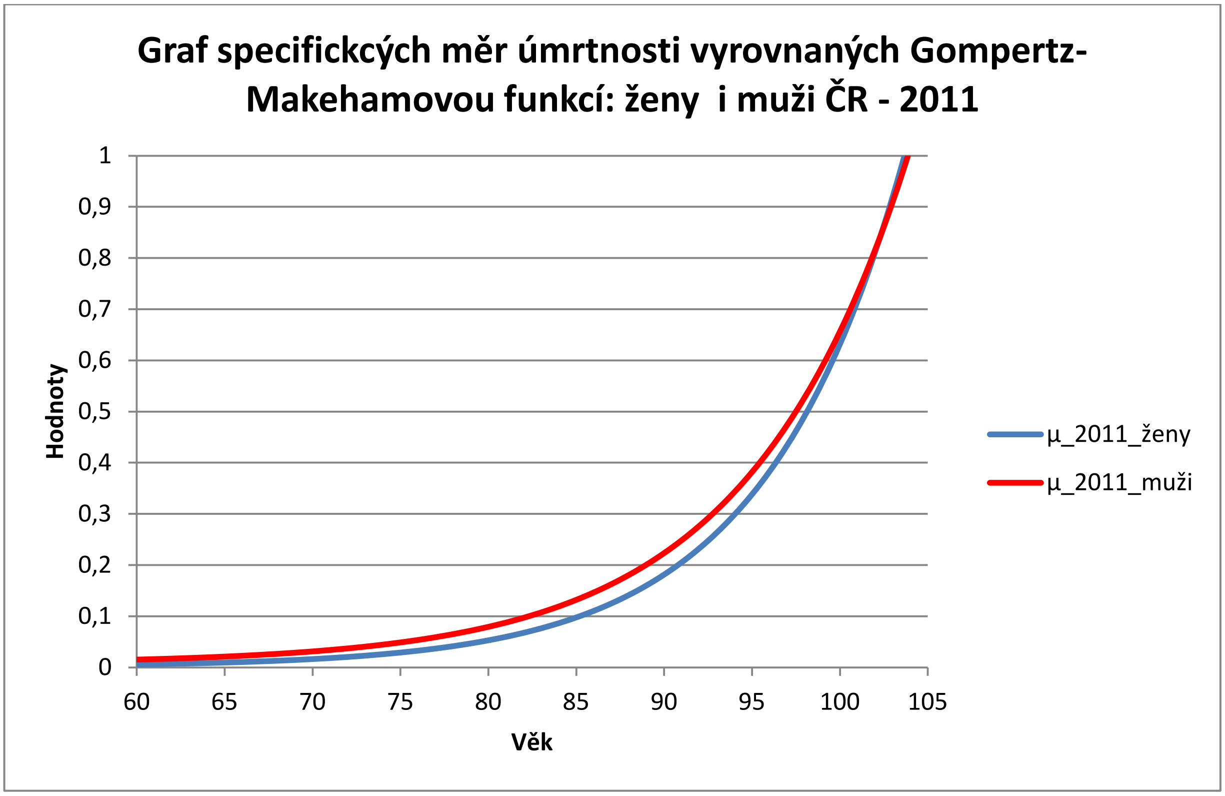
Z *Tabulky 3.1* a *Tabulky 3.2* můžete vyčíst, jak se vyvíjely hodnoty optimalizovaných parametrů Gompertz-Makehamovy funkce v letech 2009-2011 u žen a u mužů české populace. Jak vidíme, tak hodnoty za jednotlivé roky v samotné mužské populaci se liší jen minimálně, to stejné v ženské populaci, jelikož za pouhé tři roky se úmrtnost znatelně nezměnila. Je totiž obecně známo, že lidé se dožívají čím dál tím vyššího věku a lidí v důchodovém věku přibývá. V níže uvedeném *Grafu 3.7* můžeme vidět rozdíl v úmrtnosti u mužů a žen české populace v roce 2011. Ostatní křivky za roky 2010 a 2009 nemělo smysl do grafu vykreslovat, jelikož byly takřka shodné s křivkou za rok 2011. Větší rozdíly v úmrtnostech uvidíme v *Tabulce 3.3*, kde budu srovnávat data mužů za roky 2011, 2010, 2009 a 1990. Pokud ovšem porovnám obě pohlaví mezi sebou, tak už můžeme vidět podstatnější změny v úmrtnostech. Naděje dožití při narození je o 7 let vyšší než u mužů, to potvrzuje i *Graf 3.7*. Trend křivky u mužů i žen je velice podobný, avšak u žen jsou hodnoty při stejném trendu výrazně nižší, jelikož ženy se v průměru dožívají právě o 7 let vyšších věků než muži. Největší rozdíly můžeme vidět zhruba v 80 letech života. Rozdíly v úmrtnosti mužů jsem nechala vykreslit do *Grafu 3.8,* kde můžeme vidět rozdíly v úmrtnostech mužů a žen, konkrétně za rok 2011. Z grafu je patrné, že do 93 let života odchylky v úmrtnosti mužů a žen narůstají exponenciálně, ale od tohoto roku se již začínají odchylky zmenšovat, a jak můžeme vyčíst z *Grafu 3.7,* tak v nejvyšších věcích (cca 100 let a víc) se začíná růst úmrtnosti u mužů více zpomalovat než u žen. To můžete vyčíst i z *Grafu 3.9*, kde jsou porovnány specifické úmrtnosti mužů a žen v roce 2011.

muži	A	B	C
2009	0,007595876	0,000011958	1,115587801
2010	0,006557672	0,000015538	1,111929544
2011	0,007468609	0,000010932	1,116185374

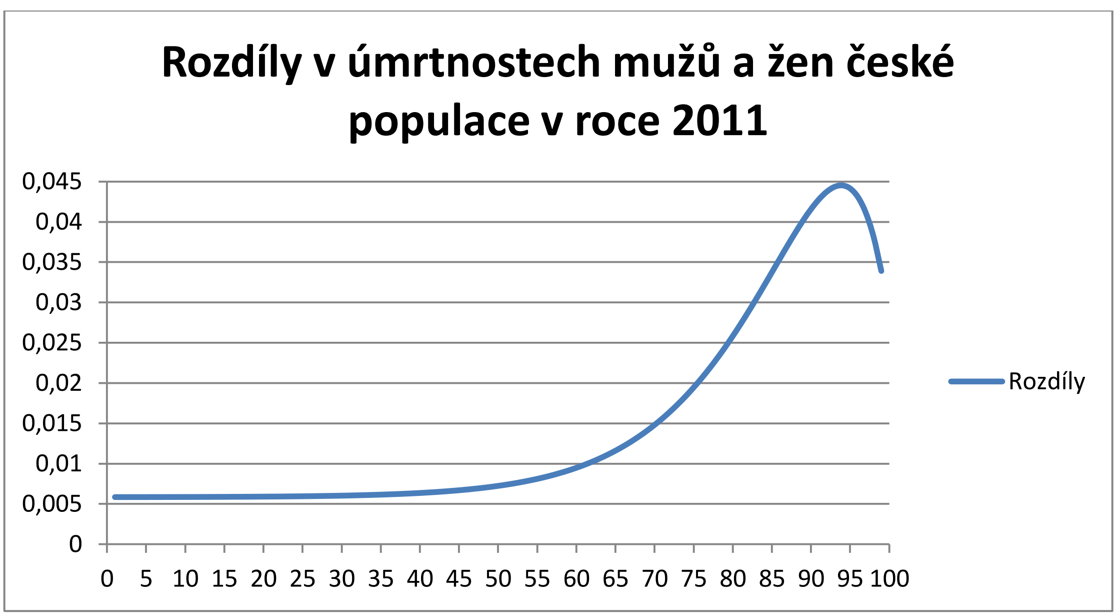
Tabulka 3.1: Optimální hodnoty parametrů Gompertz-Makehamovy funkce za roky 2009-2011 u mužů české populace.

ženy	A	B	C
2009	0,001603823	0,000002373	1,133681213
2010	0,001642811	0,000002281	1,133722411
2011	0,001608575	0,000002255	1,133613739

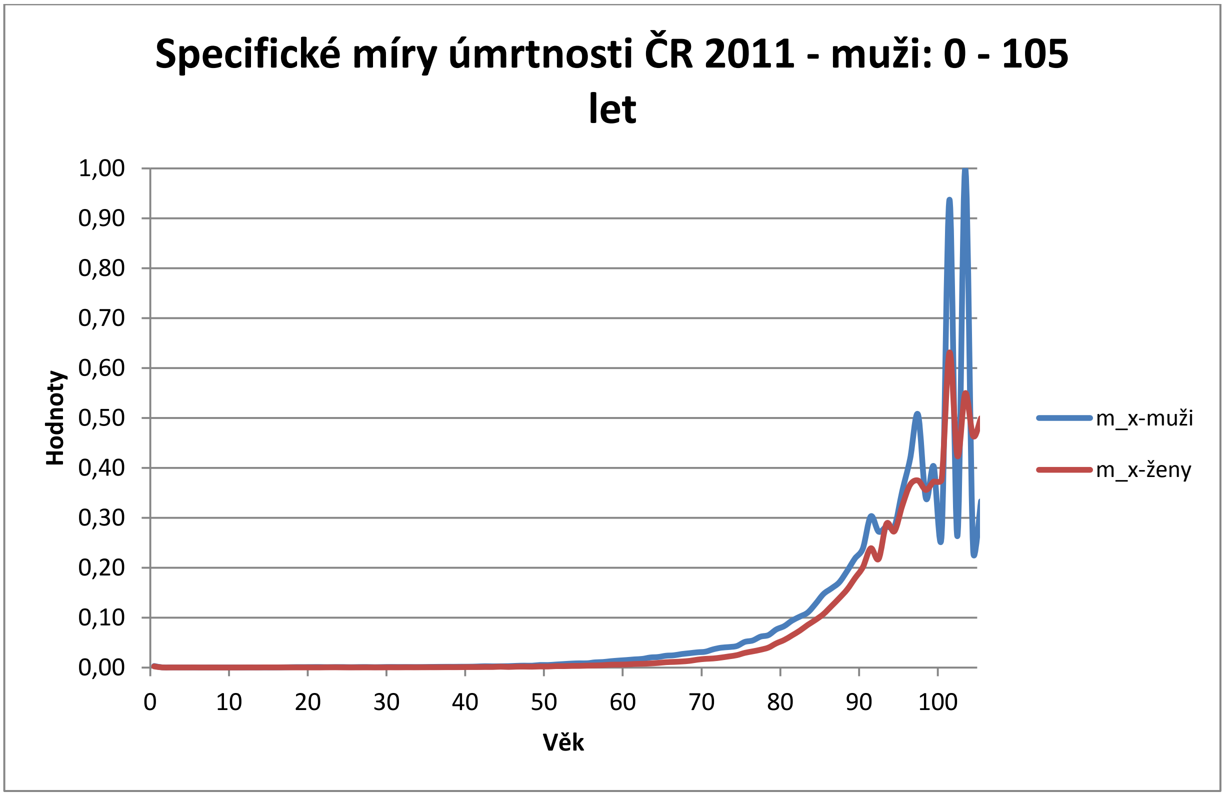
Tabulka 3.2: Optimální hodnoty parametrů Gompertz-Makehamovy funkce za roky 2009-2011 u žen české populace.



Graf 3.7: Porovnání hodnot vyrovnaných Gompertz-Makehamovou funkcí u žen a mužů české populace v roce 2011.



Graf 3.8: Rozdíly v úmrtnosti mužů a žen české populace v roce 2011.

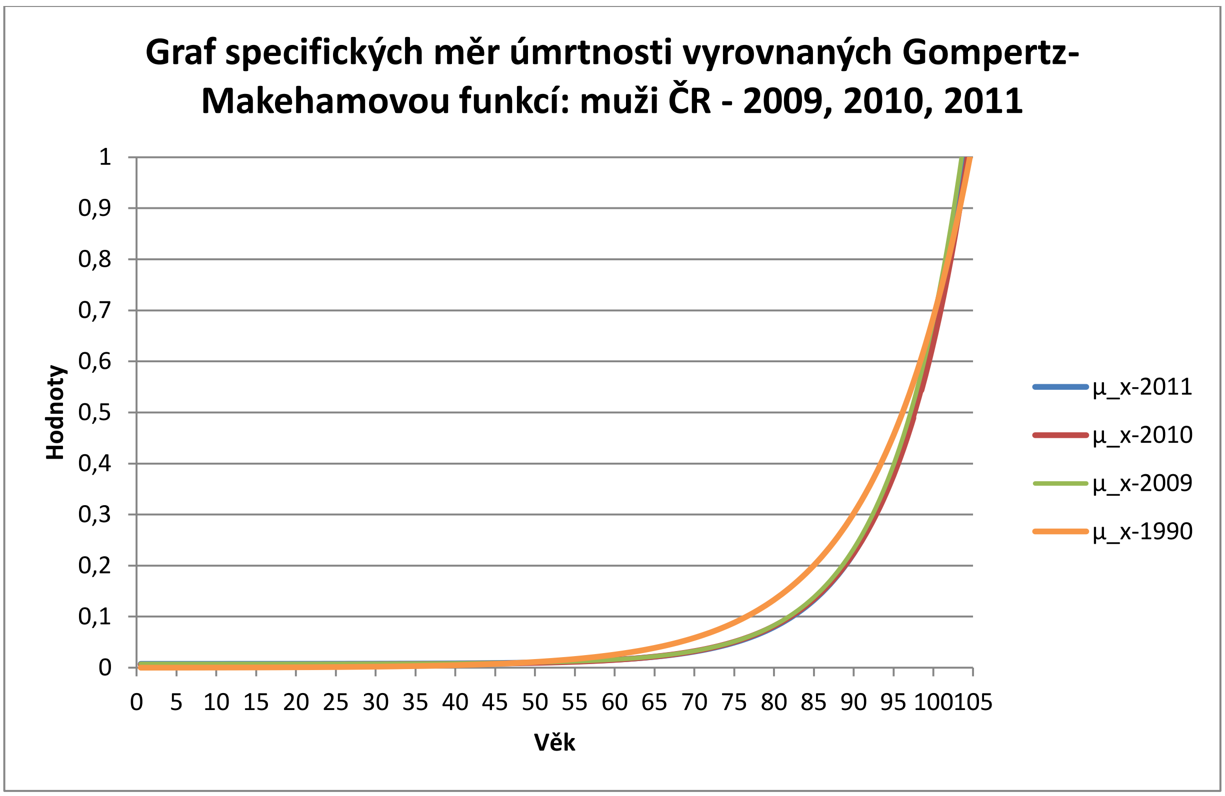


Graf 3.9: Porovnání specifických měr úmrtnosti u mužů a žen české populace v roce 2011.

Jelikož u jednotlivých pohlaví nebyly za porovnávané 3 roky patrné větší rozdíly v úmrtnostech, provedla jsem odhad parametrů Gompertz-Makehamovy funkce na datech mužské české populace z roku 1990 [18], kde by už mohly být patrnější větší rozdíly. V *Tabulce 3.3* vidíte optimalizované odhady parametrů Gompertz-Makehamovy funkce mužské české populace za roky 1990, 2009, 2010 a 2011. Jak jsem zmínila v kapitole 3.1, tak není podmínkou, aby byl parametr $A > 0$ a shodou okolností právě v roce 1990 vyšel parametr A po optimalizaci přesně 0, přičemž před optimalizací vycházel zhruba 0,006. Z *Grafu 3.10* potom můžeme vyčíst, že oproti rokům 2011, 2010 a 2009 se úmrtnost oproti roku 1990 značně liší. Je to logické, jelikož před 20 roky se lidé nedožívali tak vysokých věků jako dnes. To můžeme vidět také na střední délce života při narození, která se u mužů české populace v roce 1990 pohybovala kolem hodnoty 67,6 let, přičemž v roce 2011 to bylo zhruba 74,7.

muži	A	B	C
2009	0,007595876	0,000011958	1,115587801
2010	0,006557672	0,000015538	1,111929544
2011	0,007468609	0,000010932	1,116185374
1990	0,000000000	0,000188470	1,085460666

Tabulka 3.3: Parametry Gompertz-Makehamovy funkce u mužské české populace za roky 2009,2010,2011 a 1990.



Graf 3.10: Porovnání specifických měr úmrnosti vyrovnaných Gompertz-Makehamovou funkcí u mužů české populace v letech 1990, 2009, 2010 a 2011.

4 LATE-LIFE MORTALITY DECELERATION

V této poslední kapitole se čtenář seznámí s fenoménem “Late-life mortality deceleration”, s modifikací Gompertz-Makehamova zákona, která do jisté míry souvisí s tímto fenoménem a následnou aplikací modifikovaného Gompertz-Makehamova zákona na reálnou českou populaci, konkrétně na mužskou populaci z roku 2011. Při zpracování této kapitoly jsem čerpala především z publikací [1], [14].

4.1 Seznámení s fenoménem „Late-life mortality deceleration“

Spolehlivé odhady úmrtnosti ve vyšším věku jsou nezbytné pro určení přesné prognózy úmrtnosti a velikosti populace nejstarších věkových skupin. Úmrtnost ve vyšších věcích má tendenci odchylovat se od Gompertzova zákona úmrtnosti. Late-life mortality deceleration je fenomén, kdy se růst intenzit úmrtnosti v nejvyšších věcích života přestane zvyšovat konstantní rychlostí. Míru úmrtnosti ve vysokých věcích není jednoduché odhadnou, jelikož ve většině zemí se nejvyšších věků dožívá jen velmi malý počet osob, proto jsou takováto data velice vzácná a často můžou být nadhodnocována. Odhad intenzit úmrnosti ve vysokých věcích představuje vážné problémy pro výzkumné pracovníky.

Pozorované zpomalení růstu úmrtnosti v nejvyšších věcích souvisí s heterogenitou v datech. Z důvodu nedostatku dat pro nejvyšší věky často dochází ke sloučení dat za více časových období a právě tento heterogenní efekt může být příčinou zpomalení růstu úmrtnosti v nejvyšších věcích. Jak jsem již zmínila, populace se dožívá čím dál vyšších věků, proto když si sloučíme data pro nejvyšší věky např. za posledních 10 let a budeme chtít vidět intenzitu úmrtnosti v nejvyšším věku, tak může dojít ke zkreslení a v konečném efektu právě ke zmiňovanému „zpomalení“ úmrtnosti, jelikož se tento heterogenní soubor dat skládá z dat za různé generace.

Další z důvodů zpomalení úmrtnosti v nejvyšších věcích je fakt, že v nejvyšších věcích může být míra úmrtnosti podezřele vysoká a potom můžeme považovat odhady úmrtnostních měr za nevhodné. Tento problém můžeme vidět na datech z úmrtnostních tabulek České republiky, konkrétně u mužů za rok 2011, a to ve 103. roce života. Střední stav populace

103-letých osob je 2 a počet skutečně zemřelých osob ve věku 103 je také roven číslu 2, potom odhadovaná míra úmrtnosti m_x vychází 1, což se výrazně odchyluje od trendu předchozích hodnot.

Dalším problémem je, že v nejvyšších věcích dochází k chybnému nahlášení skutečného věku a poté dochází k nadhodnocování věku. Zpomalení úmrtnosti může být způsobeno právě špatnou kvalitou údajů. Pokud by hypotéza o špatné kvalitě nahlášených údajů platila, potom by se zpomalení úmrtnosti v nejvyšších věcích mělo pravděpodobně zmenšovat, a to s rostoucí kvalitou dat. Za účelem ověření této hypotézy byla provedena studie kvality dat pro databázi zemřelých správy sociálního zabezpečení (Social Security Administration Death Master File, dále budu značit zkratkou DMF). Jedním ze způsobů, jak získat dobré odhady úmrtnosti v nejvyšších věcích je porovnávat dostupné mezinárodní záznamy osob dožívajících se tohoto věku a následné zpřísnění kontrol prověření nahlášených údajů.

Byl navržen alternativní přístup, který dává možnost řešit problém s heterogenními soubory dat a to pomocí sestavování údajů pro více homogenních generací. Řešení problému s extrémně vysokými mírami úmrtnosti je založeno na údajích z DMF. Údaje z DMF umožňují zkoumat problematiku metodou vymřelých generací. Dostupnost informací o měsíci narození a měsíci úmrtí sledovaných generací poskytuje jedinečnou příležitost, jak získat odhady měr pro každý měsíc věku a tím eliminovat pochybné extrémně vysoké míry rizika ve vysokých věcích.

4.2 Modifikace Gompertz-Makehamova zákona

Empirická data jsou důkazem toho, že Gompertz-Makehamova funkce je vhodná pouze do věku přibližně 85 let. Gompertz-Makehamův zákon je totiž založen na předpokladu, že rychlost nárůstu úmrtnosti je s věkem konstnantí, přičemž jak jsem uvedla v kapitole 4.1 ve skutečnosti se v nejvyšších věcích nárůst úmrtosti zpomaluje. V předchozí kapitole jsem také uvedla, že v nejvyšších věcích se již výrazně modelové hodnoty Gompertz-Makehamovy funkce odchylují od empiricky naměřených specifických měr úmrtnosti a úmrtnost lze v těchto nejvyšších věcích vyjádřit pomocí funkce rostoucí čím dál tím pomaleji. Na základě tohoto předpokladu byla zformulována modifikace Gompertz-Makehamova zákona, která obsahuje již čtyři parametry, oproti původnímu modelu se třemi parametry. Tento modifikovaný model lze vyjádřit jako:

$$\boldsymbol{\mu_x^{modif} = A + B \cdot C^{x_0 + \frac{1}{\gamma} \cdot ln[\gamma \cdot (x - x_0) + 1]}},$$

kde $x > x_0$; μ_x^{modif} je modifikovaná intenzita úmrtnosti ve věku x; x_0 je věk, od kterého je úmrtnost vyrovnávána pomocí modifikované Gompertz-Makehamovy funkce; A, B a C jsou parametry původní Gompertz-Makehamovy funkce a γ je parametr, který vyjadřuje pokles rychlosti nárůstu úmrtnosti s věkem.

Věk x_0, od kterého probíhá vyrovnávání, se volí zhruba okolo věku 80 – 85 let. Tento model je specificky zaměřen na nejvyšší věky, u kterých dochází k poklesu růstu měr úmrtnosti. Vyrovnané hodnoty potom budou připomínat průběh klasické Gompertz-Makehamovy funkce, ale v nejvyšších věcích budou růst pomaleji. Tento fakt totiž odpovídá nejnovějším poznatkům o vývoji úmrtnosti osob v nejvyšších věcích, jak jsem již uvedla v kapitole 4. 1.

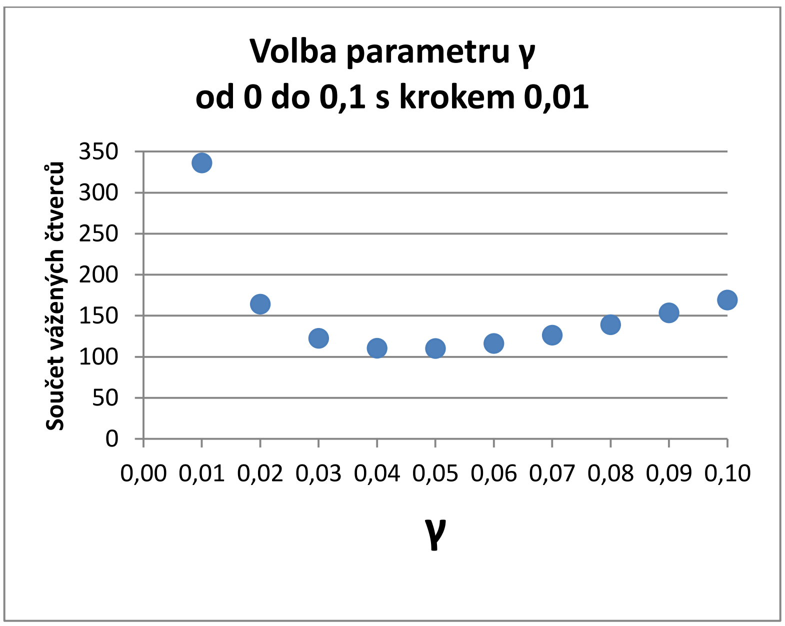
4.3 Aplikace modifikovaného Gompertz-Makehamova zákona na českou populaci

V této poslední části bakalářské práce budu aplikovat modifikaci Gompertz-Makehamova zákona na reálnou českou populaci, konkrétně na mužskou českou populaci v roce 2011. K následujícím výpočtům použiji vyrovnané hodnoty původní Gompertz-Makehamovy funkce pro mužskou populaci z roku 2011, ke kterým jsem došla v kapitole 3.2.4.

Jak jsem uvedla v předchozí kapitole, tak pro tuto modifikaci Gompertz-Makehamova zákona se za věk x_0 volí hodnoty okolo věku 80 – 85 let. Já jsem zvolila hodnotu x_0 právě 85 let, protože do tohoto věku mi přijdou empirické specifické míry úmrtnosti velice dobře kopírovany vyrovnanými hodnotamy Gompertz-Makehamovu funkcí.

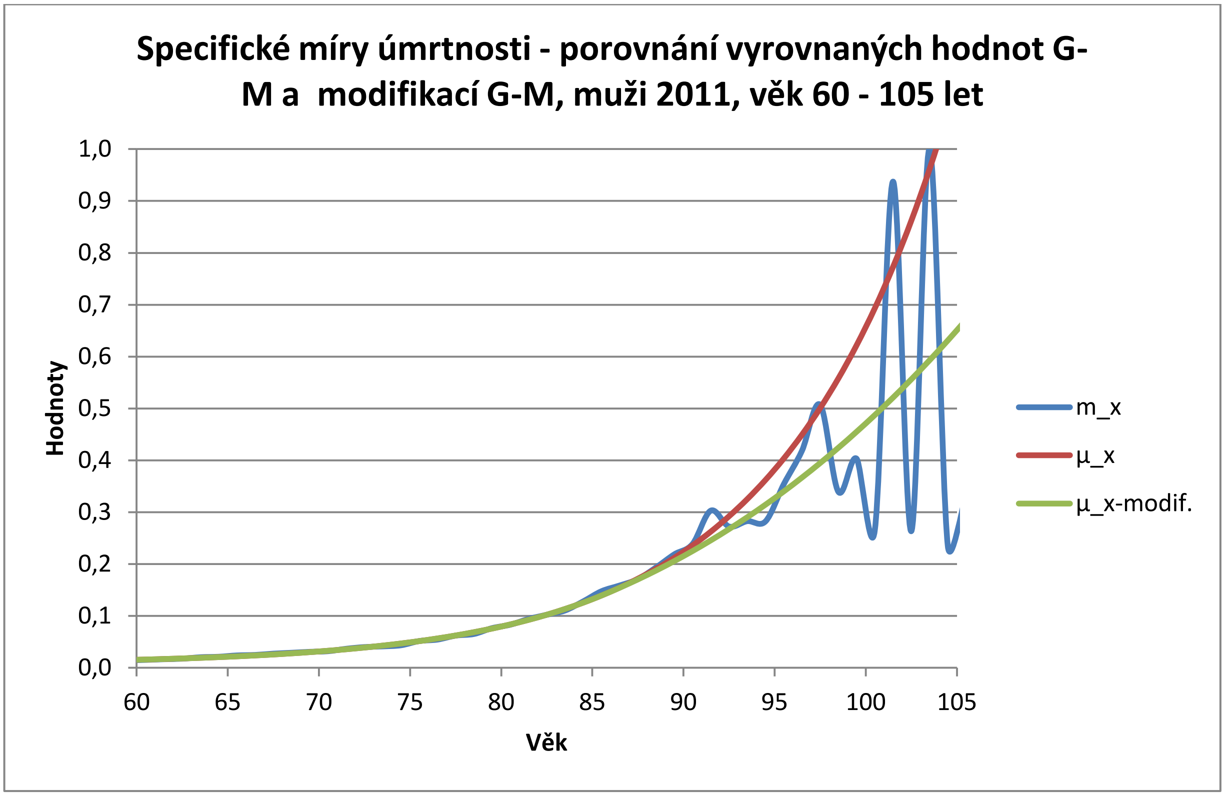
Do věku 85 let tedy použiji hodnoty vyrovnané původní Gompertz-Makehamovou funkcí a od tohoto se již budu snažit modifikací Gompertz-Makehamova zákona zpomalit růst úmrntnosti tak, aby co nejvěrněji kopíroval hodnoty empirických specifických měr úmrnosti.

Hlavním cílem tedy bude zvolit parametr γ, který určuje pokles rychlosti nárůstu úmrtnosti. Tento parametr se budu snaži zvolit optimálně, tedy tak, aby součet vážených čtverců odchylek byl pro věky 85 – 105 pokud možno co nejmenší. Vzala jsem si hodnoty parametru γ od 0 do 0,1 s krokem 0,01 a v následujícím *Grafu 4.1* můžete vidět, že nejmenší součet vážených čtverců odchylek vychází pro hodnotu parametru $\gamma = 0{,}04$.



Graf 4.1: Hodnoty vážených čtverců odchylek při dané volbě parametru γ.

V *Grafu 4.2* potom vidíte, jaký je průběh úmrtnosti podle klasického Gompertz-Makehamova zákona a jak vypadá průběh úmrtnosti podle modifikovaného Gompertz-Makehamova zákona s volbou parametru $\gamma = 0{,}04$. Z *Grafu 4.2* jde pěkně vidět, že modifikované hodnoty Gompertz-Makehamova zákona pro počáteční hodnotu vyrovnání $x_0 = 85$, opravdu věrněji kopírují empirické specifické míry úmrtnosti ve věku 85 – 105 let. Veškeré výpočty můžete vidět v přiloženém excelovském souboru *“muzi_2011_modifikace.xlsx”*.



Graf 4.2: Porovnání vyrovnaných hodnot pomocí Gompertz-Makehamova zákona a pomocí modifikovaného Gompertz-Makehamova zákona.

5 ZÁVĚR

Začátek práce byl věnován seznámení se s úmrtnostními tabulkami, jejich klasifikací základními charakteristikami, s jejichž pomocí se samotné úmrtnostní tabulky konstruují. Pro zajímavost jsem uvedla i počty zemřelých v roce 2011 podle nejčastějších příčin smrti.

Třetí kapitola se již zabývala samotným Gompertz-Makehamovým zákonem. Nejprve jsem čtenáře seznámila se základní myšlenkou a samotným vznikem modelu. V další části kapitoly jsem se již zabývala odvozením základních vztahů pro výpočet odhadů parametrů Gompertz-Makehamova modelu a následně jsem v MS Excel 2010 s pomocí těchto vztahů vypočítala počáteční odhady parametrů na mužské české populaci za rok 2011. Poté jsem hledala vhodnou délku intervalu k, která minimalizovala součet vážených čtverců odchylek, tak jsem získala věrohodnější odhad parametrů. Ale pro co nejvěrnější vyrovnání specifických měr úmrtnosti jsem již minimální odhady vážených čtverců odchylek ještě více zmenšila a to pomocí „excelovské“ funkce Řešitel, která zoptimalizovala parametry A, B a C tak, že součet čtverců vážených odchylek minimalizovala ještě více. Stejný postup jsem aplikovala i na data za roky 2010 a 2009, a to i na data ženská. V závěru kapitoly jsem porovnala výsledky, které se v jednotlivých pohlavích nikterak zvláštně nelišily. Rozdíl byl patrný až při porovnání mužů a žen vzájemně. Křivka úmrtnosti žen, vyhlazená Gompertz-Makehamovou funkcí, se realizuje nižšími hodnotami oproti křivce mužské. Tento výsledek je ovšem očekávaný, protože střední délka života při narození u žen činila v roce 2011 80,7 roku a u mužů 74,7 roku, což je značný rozdíl. Pro objektivnější porovnání jsem výše zmiňovaný postup aplikovala na mužskou populaci za rok 1990, kde už byly rozdíly v rámci daného pohlaví mnohem větší, ovšem tento výsledek také nebyl nikterak překvapující, protože před 20 lety se lidé dožívali nižších věků. Ve všech výše uvedených případech Gompertz-Makehamova funkce velice dobře kopírovala průběh skutečné úmrtnosti zhruba v letech 60 – 85 let, v nejvyšších letech se už značně odchylovala od empiricky naměřených hodnot.

V poslední kapitole jsem potom přiblížila fenomén zvaný „Late-life mortality deceleration“, který nás, jak už samotný název napovídá, seznámil se zpomalením růstu intenzity úmrtnosti v nejvyšších věcích života. S tímto zpomalením souvisí i modifikace Gompertz-Makehamova zákona. Jde o původní Gompertz-Makehamovu funkci doplněnou o čtvrtý parametr, který vyjadřuje pokles rychlosti nárůstu úmrtnosti od určitého roku života. V závěru kapitoly jsem potom optimálně odhadla tento čtvrtý parametr a porovnala modifikované hodnoty s hodnotami původní Gompertz-Makehamovy funkce.

Díky psaní bakalářské práce jsem si prohloubila vědomosti získané během studia. Lépe jsem se naučila orientovat v tabulkovém procesoru MS Excel 2010 a seznámila jsem se s některými jeho novými funkcemi, které jsem doposud neznala. Také jsem se naučila pracovat s cizojazyčnou a odbornou literaturou.

POUŽITÁ LITERATURA A WEBOVÉ ODKAZY

Literatura:

[1] BURCIN, B., TESÁRKOVÁ, K., ŠÍDLO, L., *Nejpoužívanější metody vyrovnání a extrapolace křivky úmrtnosti a jejich aplikace na českou populace,* Demografie revue pro výzkum populačního vývoje 2/2010, 79, dostupné z <http://www.czso.cz/csu/2012edicniplan.nsf/t/1000344DDD/$File/demografie_2_2010.pdf> [citováno 17. 3. 2013]

[2] CIPRA, T., *Pojistná matematika: teorie a praxe*, 2. aktualizované vydání. Praha: Ekopress, 2006.

[3] FIALA, T., *Výpočty aktuárské demografie v tabulkovém procesoru*, Praha: VŠE v Praze Nakladatelství Oeconomica, 2005.

[4] GAVRILOVA, N. S., GAVRILOV, L. A, *Stárnutí a dlouhověkost: Zákony a prognózy úmrtnosti pro stárnoucí populace*, Demografie revue pro výzkum populačního vývoje 2/2011, 109-128, dostupné z <http://notes3.czso.cz/csu/2011edicniplan.nsf/t/2C0035B3D9/$File/180311q2.pdf> [citováno 20. 1. 2013].

[5] KOSCHIN, F., *Aktuárská demografie. Praha*: VŠE v Praze, Nakladatelství Oeconomica, 2002.

[6] KUNDEROVÁ, P., *Základy pravděpodobnosti a matematické statistiky*, 1. vydání, Univerzita Palackého v Olomouci, 2004.

[7] PAVLÍK, Z., RYCHTAŘÍKOVÁ, J., ŠUBRTOVÁ, A., *Základy demografie*. Praha: Academia, 1986.

[8] SMETANA P., CIPRA, T., *Úmrtnostní tabulky nezávislé na pohlaví pro ČR (unisex tabulky) a důsledky pro pojistně-matematické výpočty – 1. Část*, příloha časopisu Pojistný obzor 5/2005

Internetové zdroje:

[9] SMETANA P., *Modifikace úmrtnostních tabulek v pojišťovnictví: generační a unisex tabulky* dostupné z <www.actuaria.cz/upload/051111_VystoupeniSAV2005.ppt> [citováné 13. 1. 2013]

[10] *Definice úmrtnostní tabulky* dostupné z <http://www.demografie.info/?cz_umrtnosttabulky=> [citováno 12. 1. 2013]

[11] *Úmrtnostní tabulky – Metodika* dostupné z <http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_metodika> [citováno 13. 1. 2013]

[12] *Počty zemřelých za rok 2011 podle příčin*, dostupné z <http://www.czso.cz/csu/2012edicniplan.nsf/t/6A002ECDC6/%24File/400712a6.doc> [citováno 20. 2. 2013]

[13] *Úmrtnost lidského organismu v průběhu jeho vývoje* dostupné z <http://apps.faf.cuni.cz/Gerontology/mechanisms/mortality/mortality_human.asp> [citováno 20. 2. 2013]

[14] GAVRILOVA, N. S., GAVRILOV, L. A, *Mortality measurement at advanced ages: A study of the social security administration death master file* dostupné z <http://longevity-science.org/pdf/Mortality-NAAJ-2011.pdf> [citováno 10. 2. 2013]

[15] *Gompertz-Makeham law of mortality*, dostupné z <http://en.wikipedia.org/wiki/Gompertz%E2%80%93Makeham_law_of_mortality> [citováno 20. 1. 2013]

[16] *Porovnání pojistného před 21. 12. 2012 a po 21. 12. 2012*, dostupné z <http://www.csobpoj.cz/cs/o-spolecnosti/pro-media/tiskove-zpravy/Documents/2012_11_12_TZ_Zeny-berou-pojistovny-utokem-muzi-vyckavaji.pdf> [citováno 13. 2. 2013]

Zdroje dat pro výpočty:

[17] *Úmrtnostní tabulky za ČR 1920 - 2011*, dostupné z <http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_za_cr_od_roku_1920/$File/cr_ut_1920_2011.zip> [citováno 2. 11. 2012]

[18] *Počty zemřelých a střední stav obyvatelstva*, dostupné z <http://popin.natur.cuni.cz/html2/index.php?item=3.7> [citováno 30. 3. 2013]

SEZNAM PŘÍLOH

Příloha 1: CD ROM s výpočty a grafy