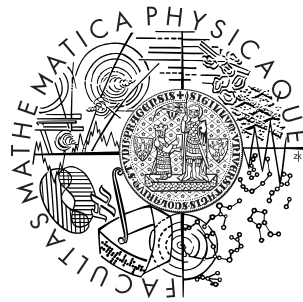


Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## DIPLOMOVÁ PRÁCE



Věra Kuběnová

### Metody projekce úmrtnosti

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Hana Netolická

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

2008

Chtěla bych poděkovat Ing. Haně Netolické za mnoho zajímavých námětů a připomínek při zpracování této práce.

Prohlašuji, že jsem svou diplomovou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 14.4.2008

Věra Kuběnová

# Obsah

Úvod	6
<b>1 Charakteristiky úmrtnosti</b>	<b>7</b>
1.1 Zavedení důležitých pojmů	7
1.2 Údaje v úmrtnostní tabulce	10
1.3 Generační úmrtnostní tabulky	12
1.4 Některé vztahy mezi demografickými ukazateli	13
1.5 Výpočet demografických ukazatelů z dostupných dat	15
1.6 Vyrovnávání dat v úmrtnostních tabulkách	18
<b>2 Kohort efekt</b>	<b>20</b>
2.1 Pojmy a pozorování - kohort efekt	20
2.1.1 Velká Británie	21
2.1.2 Japonsko	21
2.1.3 Kanada	22
2.2 Příčiny vzniku kohort efektu	22
2.3 Česká populace a kohort efekt	23
2.3.1 Dostupnost dat	24
2.3.2 Studie existence kohort efektu v ČR	25
<b>3 Metody projekce úmrtnosti</b>	<b>33</b>
3.1 Přidružené vědní obory a jejich vysvětlení trendů v úmrtnosti	33
3.1.1 Epidemiologie	33
3.1.2 Sociologie a ekonomie	34
3.1.3 Demografie	34
3.2 Klasifikace metod projekcí	35
3.3 Podrobnější popis některých metod, jejich výhody a nevýhody	36
3.3.1 Zákony úmrtnosti	36

3.3.2	Brassova relační metoda . . . . .	39
3.3.3	Lee-Carterův model . . . . .	39
3.3.4	APC modely . . . . .	41
<b>4</b>	<b>Aplikace zvolené metody na data České republiky</b>	<b>42</b>
4.1	Výběr metody aplikované na data ČR . . . . .	42
4.2	Použití vybrané metody . . . . .	43
<b>5</b>	<b>Závěr</b>	<b>52</b>
	<b>Přílohy</b>	<b>54</b>
	<b>Literatura</b>	<b>67</b>

Název práce: Metody projekce úmrtnosti

Autor: Věra Kuběnová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Ing. Hana Netolická

e-mail vedoucího: netolicka@generalippf.eu

Abstrakt: V předložené práci se zabýváme problematikou projekce úmrtnosti. Zavádíme základní pojmy spjaté s danou problematikou, popisujeme úmrtnostní tabulky a speciálně se věnujeme generačním úmrtnostním tabulkám. Zabýváme se velice důležitým jevem ovlivňujícím projekci úmrtnosti, takzvaným kohort efektem. Dále jsou zde diskutovány možné příčiny jeho vzniku. V dostupných datech o populaci České republiky následně zkoumáme možnou existenci kohort efektu v české populaci. Pro porovnání jsou uvedeny i informace o výskytu kohort efektu v jiných vyspělých zemích světa. Zkoumáme různé přístupy k modelování a následné projekci úmrtnosti. Uvádíme rozmanité faktory ovlivňující úmrtnost a způsob jejich analyzování z pohledu různých vědních oborů. Některé z metod projekce úmrtnosti jsou pak popisovány podrobněji. Na konec je vybrána Lee-Carterova metoda projekce úmrtnosti, která je použita pro předpověď vývoje úmrtnosti v České republice pro roky 2007–2060.

Klíčová slova: projekce úmrtnosti, kohort efekt, Lee-Carterova metoda

Title: Mortality projection methods

Author: Věra Kuběnová

Department: Department of Probability and Mathematical Statistics

Supervisor: Ing. Hana Netolická

Supervisor's e-mail address: netolicka@generalippf.eu

Abstract: In this thesis we study the issue of mortality projection. We introduce some basic terms related to mortality, describe life tables and special attention is paid to generation life tables. We are interested in an important phenomenon of so-called cohort effect which influences the mortality projection. There are also discussed possible reasons for creation of the cohort effect. We use available data for population of the Czech Republic for looking into existence of the cohort effect in Czech population. For comparison we mention the existence of the cohort effect in other developed countries. We study different approaches mortality model and project. We set up various factors influencing mortality patterns and the way of their analyzing by some scientific branches. Some of the mortality projection methods are described in more detail. The Lee-Carter method is chosen to forecast mortality for population of the Czech Republic in period 2007–2060.

Keywords: mortality projection, cohort effect, Lee-Carter method

# Úvod

Téma budoucího vývoje úmrtnosti má svoje uplatnění v mnoha oborech. Větší změny v předpokládaném vývoji úmrtnosti či významné lokální výkyvy v tomto vývoji mohou mít velice nepříznivý dopad např. na důchodovou sociální síť státu nebo na podnikání v oblasti životního a penzijního pojištění.

Hlavním cílem této práce je zkoumání různých přístupů k projekci úmrtnosti a aplikace vybrané metody projekce úmrtnosti na populaci České republiky. Důležitým faktorem při předpovědi budoucího vývoje úmrtnosti je vliv tzv. kohort efektu. Ten budeme podrobně analyzovat a zjišťovat jeho případnou existenci v populaci České republiky. Abychom to mohli provést, musíme získat potřebná vstupní data. Vyjdeme tedy z dostupných dat získaných z Českého statistického úřadu. Výsledky porovnáme i s informacemi o kohort efektu v jiných vyspělých zemích a budeme studovat možné příčiny vzniku tohoto jevu.

Tématem souvisejícím s kohort efektem jsou generační úmrtnostní tabulky. Zaměříme se tedy i na jejich podrobnější popis.

Z možných přístupů k projekci úmrtnosti si na základě výsledků z analýzy existence kohort efektu vybereme jeden, který aplikujeme na data z populace České republiky. Na základě dostupných dat vytvoříme projekci úmrtnosti pro populaci ČR.

Abychom mohli úmrtnost studovat, zavedeme nejdříve v první kapitole základní pojmy týkající se úmrtnosti a úmrtnostních tabulek a souvislosti mezi nimi. Zvláště se pak bude věnovat popisu a významu generačních úmrtnostních tabulek. Ukážeme si, jak lze odhadovat demografické ukazatele z dostupných dat a jaké při tom zavádíme předpoklady. Pro další využití vypočtených ukazatelů je třeba jejich hodnoty vyrovnat, čímž uzavřeme první kapitolu.

Ve druhé kapitole se budeme zabývat tzv. kohort efektem a jeho vlivem na předpovídání vývoje úmrtnosti. Vysvětlíme, co vlastně označujeme pojmem kohort efekt, a popíšeme jeho existenci v některých zemích. Rovněž se budeme zabývat možnými příčinami vzniku kohort efektu. Důležitou součástí této diplomové práce bude výzkum existence kohort efektu v české populaci.

Trendy v úmrtnosti z pohledu různých vědních oborů okomentujeme v kapitole třetí. Také zde zavedeme klasifikaci projekcí úmrtnosti a podrobně popíšeme některé vybrané metody.

V poslední části se budeme věnovat aplikaci vybrané metody na data populace České republiky. Vytvoříme předpověď věkově specifických měr úmrtnosti a výsledky budeme ilustrovat mimo jiné různými grafy. Uvedeme zde také srovnání s jinou projekcí úmrtnosti populace ČR, která je součástí populační prognózy dle [5].

# Kapitola 1

## Charakteristiky úmrtnosti

Abychom mohli studovat trendy ve vývoji úmrtnosti a používat při tom získaná data, musíme zavést základní úmrtnostní charakteristiky populace a vztahy mezi nimi. Ukážeme také jak získaná data zpracovat, aby vedla k relevantním výsledkům.

### 1.1 Zavedení důležitých pojmů

Nejdříve je třeba upřesnit, co vlastně pojem úmrtnost znamená. Můžeme ji chápat jako neschopnost jedince dožít se určitého věku nebo naopak jako jakousi „schopnost“ zemřít v určitém věku. Souhrnem těchto individuálních schopností jedinců dostaneme úmrtnost pro celou populaci. Pojem populace vymezíme dle [18] následovně:

***Definice 1.1:***

*Populaci definujeme jako skupinu osob, která má společné znaky. Těmito znaky může být např. území, čas, pohlaví, povolání apod.*

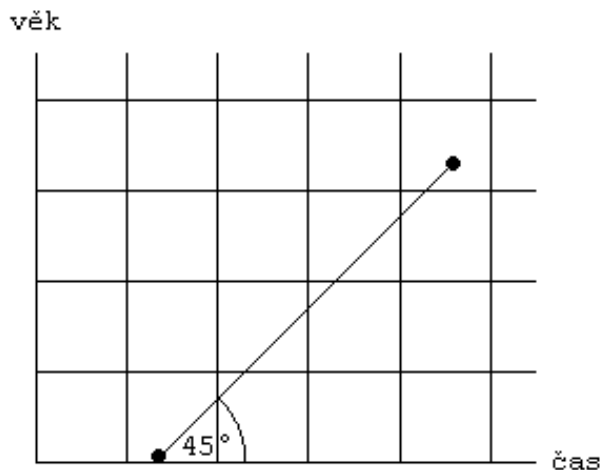
Pro zjednodušení situace se v demografické teorii často používá předpoklad o uzavřenosti populace a předpoklad o spojitém průběhu rození se v populaci. Dle [18] definujeme uzavřenou populaci takto:

***Definice 1.2:***

*Populaci nazýváme uzavřenou populací, jestliže v ní neexistuje migrace.*

V pozorovaných reálných datech ovšem tento předpoklad není splněn. Např. do sledovaného počtu členů populace v daném věku zahrnujeme i členy populace, kteří se přistěhovali.

Pro znázornění populace v závislosti na věku a čase je používaným nástrojem demografická síť, tzv. Lexisův diagram. Život každého jedince je znázorněn úsečkou, tzv. čarou života, která má jeden krajní bod na ose času, tj. okamžik narození jedince, a druhý krajní bod představuje úmrtí jedince (viz obrázek 1.1).



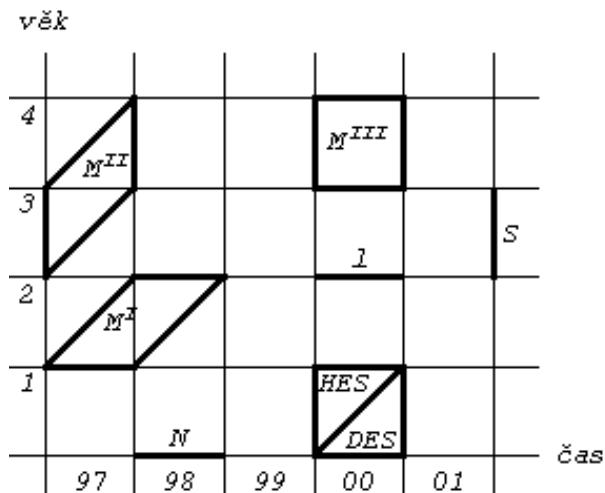
Obrázek 1.1: Čára života jedince v Lexisově diagramu.

Dále z tohoto diagramu můžeme snadno zjistit různé údaje o populaci, které jsou znázorněny na obrázku 1.2 a popsány níže.

Popis vyznačených charakteristik populace, které jsou znázorněny v Lexisově diagramu, je dle [18] následující (viz obrázek 1.2):

- $N$  ...počet narozených v roce 1998 (tj. počet bodů, ve kterých začínají čáry života)
- $l$  ...počet dožívajících se věku 2 roky v roce 2000
- $S$  ...počet žijících dvouletých k 1.1.2002
- $M^I$  ... počet zemřelých ve věku dokončeného roku z narozených v roce 1996, tzv. soubor zemřelých 1.řádu





Obrázek 1.2: Různé charakteristiky populace znázorněné v Lexisově diagramu.

$M^{II}$  ... počet zemřelých v roce 1997 z narozených v roce 1994, tzv. soubor zemřelých 2.řádu

$M^{III}$  ... počet zemřelých v roce 2000 ve věku dokončených 3 let, tzv. soubor zemřelých 3.řádu

$DES$  ... počet zemřelých v roce 2000 nulaletých z narozených v roce 2000, tzv. dolní elementární soubor

$HES$  ... počet zemřelých v roce 2000 nulaletých z narozených v roce 1999, tzv. horní elementární soubor

Data, která o populaci obvykle bývají k dispozici, jsou časově a věkově vymezena, tedy jde o vyznačenou oblast  $M^{III}$ .

Existují i jiné typy demografických sítí, např. podle [8] se můžeme setkat se sítí, kde osa věku svírá s osou času úhel  $135^\circ$  a čáry života jsou pak kolmé na osu času.

Nyní zavedeme dle [18] spojitý model populace s věkovou strukturou.

**Definice 1.3:**

*Za předpokladů, že rození v populaci probíhá spojitě a že populace je uzavřená, definujeme následující spojitě funkce:*

- $b(u)$  jako hustotu počtu zrození,
- $s(x, u)$  jako generační funkci přežití,
- $l(x, u) = b(u) \cdot s(x, u)$  jako hustotu věkového rozložení.

Dalším důležitým pojmem v modelování úmrtnosti je tzv. intenzita úmrtnosti. Tu definujeme pomocí již zavedených funkcí:

**Definice 1.4:**

*Předpokládejme, že rození v populaci probíhá spojitě a že populace je uzavřená. Potom můžeme definovat generační intenzitu úmrtnosti následujícími způsoby:*

$$\begin{aligned} \mu(x, u) &= \lim_{\Delta x \rightarrow 0} \frac{l(x, u) - l(x + \Delta x, u)}{\Delta x \cdot l(x, u)} = -\frac{1}{s(x, u)} \cdot \frac{\partial s(x, u)}{\partial x} = \\ &= -\frac{1}{l(x, u)} \cdot \frac{\partial l(x, u)}{\partial x} = -\frac{\partial \log s(x, u)}{\partial x} = -\frac{\partial \log l(x, u)}{\partial x}, \end{aligned}$$

pro  $\Delta x > 0$ .

Generační intenzitu úmrtnosti tedy můžeme interpretovat jako počet zemřelých ve věku mezi  $x$  a  $x + \Delta x$  z narozených v  $(u, u + du)$  vztažených na  $\Delta x$  a na počet dožívajících se věku  $x$  z narozených v  $(u, u + du)$ . Způsoby definování generační intenzity úmrtnosti jsou ekvivalentní, což je zřejmé z definice 1.3. Často se používá intenzita úmrtnosti, která se od generační intenzity úmrtnosti liší tím, že nezávisí na čase narození jedinců, označujeme ji  $\mu_x$ .

S pomocí definovaných funkcí snadno vypočteme různé charakteristiky populace splňující výše zmíněné předpoklady. Například počet zemřelých ve věku mezi  $x_1$  a  $x_2$  v intervalu  $(t_1, t_2)$ , tj. soubor zemřelých 3.řádu, lze vypočítat jako

$$\int_{x_1}^{x_2} \int_{t_1-x}^{t_2-x} l(x, u) \mu(x, u) du dx = \int_{x_1}^{x_2} \int_{t_1}^{t_2} l(x, t-x) \mu(x, t-x) dt dx.$$

## 1.2 Údaje v úmrtnostní tabulce

Úmrtnostní tabulky jsou buď úplné, tj. obsahují údaje pro roční věkové intervaly, nebo zkrácené, tj. zaznamenávají údaje pro delší věkové intervaly (často pětileté). Lze je interpretovat jako záznamy vymírání hypotetických generací současně narozených jedinců. Základním údajem úmrtnostní tabulky je pravděpodobnost úmrtí  $q_x$ , tedy

pravděpodobnost, že osoba ve věku  $x$  zemře před dosažením věku  $x + 1$ . Tento údaj se odhaduje z dostupných dat a slouží jako základní kámen pro odvození dalších statistik.

Doplňkem pro pravděpodobnost úmrtí do jedné je pravděpodobnost dožití  $p_x$ , to znamená pravděpodobnost, že osoba ve věku  $x$  se dožije věku  $x + 1$ . Tabulkový počet dožívajících se věku  $x$  označujeme  $l_x$ , což je počet osob, které se dožijí věku  $x$  z  $l_0$  současně narozených, a platí vztah

$$l_{x+1} = p_x \cdot l_x. \quad (1.1)$$

Rozdílem  $l_x - l_{x+1}$  vypočteme tabulkový počet zemřelých  $d_x$ , tj. počet zemřelých ve věku mezi  $x$  a  $x + 1$  z  $l_0$  současně narozených,

$$d_x = l_x - l_{x+1}. \quad (1.2)$$

Další statistikou z úmrtnostních tabulek je tabulkový počet žijících  $L_x$ , který představuje celkovou dobu prožitou tabulkovou generací ve věku  $x$ , tedy v intervalu  $[x, x + 1)$ . Pro tabulkový počet žijících platí vztah

$$L_x = l_{x+1} + a_x d_x, \quad (1.3)$$

kde  $a_x$  označuje průměrnou délku části věkového intervalu  $[x, x + 1)$  prožitou jedincem, který ve věku  $x$  zemře. Pokud bychom položili  $a_x = \frac{1}{2}$  a dosadili za  $d_x$  vyjádření ze vzorce 1.2, pak dostáváme tvar

$$L_x = l_{x+1} + \frac{1}{2} (l_x - l_{x+1}) = \frac{l_x + l_{x+1}}{2}, \quad (1.4)$$

tedy tabulkový počet žijících vypočítaný jako průměr z tabulkového počtu dožívajících se počátku a konce věkového intervalu  $[x, x + 1)$ .

Délka lidského života není nekonečná, proto ji lze shora omezit.

**Definice 1.5:**

Nejnižší věk, kterého se již nikdo nedožije, označujeme symbolem  $\omega$ . Definujeme ho jako  $\omega = \min_{x>0} \{x : \int_{t_1}^{t_2} l(x, t - x) dt = 0; \text{ pro } \forall t_1, t_2 \in (0, \infty), \text{ kde } t_2 > t_1\}$ .

Celkovou dobu prožitou tabulkovou generací ve věku  $\geq x$  značíme  $T_x$ . Platí pro ni vztah

$$T_x = L_x + L_{x+1} + \dots + L_{\omega-1} = L_x + T_{x+1}. \quad (1.5)$$

Podílem  $\frac{T_x}{l_x}$  pak lze spočítat střední délku života jedince ve věku  $x$ , označovanou obvykle  $\overset{\circ}{e}_x$ , tzn.

$$\overset{\circ}{e}_x = \frac{T_x}{l_x}. \quad (1.6)$$

Střední délku života bychom dostali po dosazení 0 do věku jedince tedy jako  $\overset{\circ}{e}_0$ .

Posledním zde popsaným údajem z úmrtnostní tabulky je tabulková míra úmrtnosti. Označuje se  $m_x$  a je definována vztahem

$$m_x = \frac{d_x}{L_x}. \quad (1.7)$$

Výše popsaná úmrtnostní tabulka se nazývá průřezová, je konstruována na základě pozorování z krátkého období (většinou jednoho roku) a popisuje tedy úmrtnost během tohoto období. Jiným druhem jsou např. generační úmrtnostní tabulky, které představují záznam průběhu života konkrétní populace současně narozených jedinců pořizovaný počínaje okamžikem narození všech jedinců této populace a konče smrtí posledního z nich. Další modifikací úmrtnostních tabulek jsou např. selekční tabulky, kde se sleduje další znak jedince, který vede k hodnotě pravděpodobnosti úmrtí odlišné od hodnoty zjištěné pro celou populaci, či tabulky s více dekrementy, kde se sledují různé příčiny odchodu z populace.

### 1.3 Generační úmrtnostní tabulky

V některých oblastech, jako například v oblastech penzijního připojištění nebo důchodového pojištění, není použití průřezových tabulek příliš vhodné. Představme si příklad, že budeme kalkulovat výši pojistného pro produkt penzijního připojištění z hodnot aktuální průřezové úmrtnostní tabulky pro osobu, která je nyní ve věku 40 let, tj. použijeme údaj  $q_{40}$ . Za přibližně 30 let, kdy dojde k začátku vyplácení plnění, bude aktuální hodnota pravděpodobnosti úmrtí ve věku 70 s velmi vysokou pravděpodobností nižší než  $q_{70}$  z průřezové tabulky aktuální v době výpočtu, dojde tedy k nadhodnocení úmrtnosti. Tento rozdíl se navíc obvykle zvyšuje s rostoucí dobou trvání pojištění. Může to tedy pro pojišťovnu znamenat poměrně značné ztráty z výplaty penzí po celou dobu, co jsou vypláceny.

Řešením, které se nabízí a je často používáno, je počítat s nižší úmrtností než je aktuální tabulková úmrtnost, to znamená použít tzv. implicitní přírážku. Tato metoda je ovšem dosti nepřesná.

Vhodnějším řešením je konstrukce generačních úmrtnostních tabulek např. pro daný pojistný kmen, kde pravděpodobnosti úmrtí závisejí na věku a roku narození. Pomocí matematických metod pak modelujeme vývoj úmrtnosti.

Obecně vytvoření vlastních úmrtnostních tabulek pro svůj pojistný kmen je vhodné zvláště proto, že v pojistném kmeni často pozorujeme nižší úmrtnost než v celé populaci (viz [4]).

Příkladem konstrukce generačních úmrtnostních tabulek je hledání trendu ve vývoji úmrtnosti (viz [14]). Obecně hledáme funkci  $\varphi$ , která bude vyhovovat vztahu

$$q_{x,t+h} \approx q_{x,t} \varphi(h).$$

Konkrétní hodnoty parametrů trendové funkce  $\varphi$  se získají statistickými metodami z dostupných dat. Pro odvození generačních úmrtnostních tabulek stačí určit počáteční hodnoty v nějakém čase  $t_0$ , tzv. bazické pravděpodobnosti úmrtí,  $q_{x,t_0}^B$  a hledané pravděpodobnosti úmrtí pak dostaneme z rovnice

$$q_{x,t} = q_{x,t_0}^B \varphi(t - t_0).$$

## 1.4 Některé vztahy mezi demografickými ukazateli

Zavedeme náhodnou veličinu  $X$  se spojitým rozdělením představující délku života nově narozeného jedince. Dále zavedeme s ohledem na [18] označení pro pravděpodobnost, že právě narozený jedinec se dožije věku  $x$ , a pro doplňkovou pravděpodobnost, že zemře do věku  $x$  včetně:

$$P(X > x) = {}_x p_0, \tag{1.8}$$

$$P(X \leq x) = {}_x q_0. \tag{1.9}$$

Označme  $T(x)$  zbývající dobu života jedince, který je naživu ve věku  $x$ . Budeme předpokládat, že rozdělení  $T(x)$  je shodné s podmíněným rozdělením  $X - x$  za podmínky  $X > x$ . Pak platí

$$P(T(x) > t) = P(X - x > t \mid X > x) = \frac{P(X > x + t)}{P(X > x)} = \frac{{}_{x+t} p_0}{{}_x p_0}.$$

Pravděpodobnost, že osoba ve věku  $x$  bude žít dalších  $t$  let, označme  ${}_t p_x$  a pravděpodobnost, že osoba ve věku  $x$  zemře během následujících  $t$  let, označme  ${}_t q_x$ . To

znamená, že

$$P(T(x) > t) = {}_t p_x, \quad (1.10)$$

$$P(T(x) \leq t) = {}_t q_x. \quad (1.11)$$

S použitím náhodné veličiny  $X$  a rovnosti 1.8 můžeme intenzitu úmrtnosti vyjádřit takto:

$$\begin{aligned} \mu_x &= \lim_{h \rightarrow 0^+} \frac{P(x < X \leq x + h \mid X > x)}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{P(X > x) - P(X > x + h)}{h P(X > x)} \\ &= \lim_{h \rightarrow 0^+} \frac{{}_x p_0 - {}_{x+h} p_0}{h {}_x p_0} = -\frac{\frac{\partial}{\partial x} {}_x p_0}{{}_x p_0}. \end{aligned} \quad (1.12)$$

Nyní můžeme vyjádřit intenzitu úmrtnosti v neceločíselném věku. Předpokládejme rovnoměrné rozložení úmrtí na věkovém intervalu  $(x, x + 1)$ . Nechť  $t \in (0, 1)$ .

Z rovnoměrnosti úmrtí na  $(x, x + 1)$  vyplývá, že  $t d_x$  vyjadřuje počet úmrtí ve věku mezi  $x$  a  $x + t$ . Pomocí modifikace 1.2 můžeme psát

$$l_{x+t} = l_x - t d_x = l_x - t(l_x - l_{x+1}) = (1 - t) l_x + t l_{x+1},$$

což je počet žijících v neceločíselném věku  $x + t$  vypočtený pomocí lineární interpolace.

Dále pomocí dosazení do 1.1 za  ${}_t p_x = 1 - {}_t q_x$  dostaneme

$${}_t q_x = \frac{l_x - l_{x+t}}{l_x} = \frac{t d_x}{l_x} = t q_x. \quad (1.13)$$

Tedy  ${}_t p_x = 1 - t q_x$  a ze vzorce 1.12 plyne

$${}_t p_x \mu_{x+t} = \left( -\frac{\partial {}_t p_x}{\partial t} \right) = q_x \Rightarrow \mu_{x+t} = \frac{q_x}{{}_t p_x} = \frac{q_x}{1 - t q_x}. \quad (1.14)$$

Speciálně pak platí pro  $t = \frac{1}{2}$

$$\mu_{x+\frac{1}{2}} = \frac{q_x}{1 - \frac{1}{2} q_x} = \frac{d_x}{l_x - \frac{1}{2} d_x} = \frac{d_x}{L_x} = m_x. \quad (1.15)$$

Předpoklad, který použijeme při vyjádření vztahu mezi pravděpodobnostmi úmrtí a mírou úmrtnosti, je předpoklad konstantní intenzity úmrtnosti během jednoho roku

v daném věku. Předpokládejme tedy, že  $\mu_{x+t} = \mu_x$  pro  $\forall x$  a  $t \in (0, 1)$ . Úpravou 1.12 dostaneme

$$\ln {}_x p_0 = - \int_0^x \mu_y dy,$$

tedy i

$$\ln {}_t p_x = - \int_x^{x+t} \mu_y dy \quad (1.16)$$

a s využitím předpokladu konstantní intenzity úmrtnosti a pro  $t \in (0, 1)$  platí

$${}_t p_x = e^{-\int_x^{x+t} \mu_y dy} = e^{-\mu_x t}. \quad (1.17)$$

Pokud přidáme předpoklad o rovnoměrném rozdělení úmrtí na  $(x, x+1)$  a použijeme-li rovnost  ${}_t p_x = 1 - {}_t q_x$  a vztah 1.15 získáme vzorec

$$q_x = 1 - e^{-m_x}. \quad (1.18)$$

## 1.5 Výpočet demografických ukazatelů z dostupných dat

Data, která máme k dispozici, ovšem nesplňují předpoklady spojitosti. Pozorujeme hodnoty pouze v diskrétních časových okamžicích. Zavedeme úmluvu platnou i v dalším textu, že termín „ve věku  $x$ “ bude vyjadřovat nějakou charakteristiku ve věku mezi  $x$  a  $x+1$ .

Soustředíme se na míru úmrtnosti  $m_x$ , kterou použijeme pro popis úmrtnosti v populaci. Obdobně bychom mohli použít např. pravděpodobnost úmrtí  $q_x$ . Tabulková míra úmrtnosti byla již dříve definována podílem  $\frac{d_x}{L_x}$ . Nyní v reálné populaci za  $d_x$  použijeme  $D_x$ , což tedy představuje počet úmrtí ve věku  $x$  během sledovaného období. Ve jmenovateli by měla být doba, kterou během sledovaného období prožili členové populace ve věku  $x$ . Tuto dobu nyní označíme  $E_x^C$  a nazveme ji centrální expozicí.

Při jejím výpočtu, respektive odhadu, vycházíme z dat o populaci, které jsou k dispozici. Ty popíšeme dle [18] následujícím způsobem:

- $b_x$  nechť je počet osob ve věku  $x$  na začátku sledovaného období,
- $e_x$  pak počet osob ve věku  $x$  na konci sledovaného období,

- $n_x$  představuje počet osob, které se během sledovaného období přistěhovaly ve věku  $x$ ,
- $w_x$  je počet osob, které se během sledovaného období vystěhovaly ve věku  $x$ ,
- $l_x^\circ$  nechť je počet osob, které se ve sledovaném období dožily věku  $x$ .

Pro další postup zavedeme následující, situaci zjednodušující předpoklady:

1. sledované období je jeden kalendářní rok,
2. data narození jedinců z  $b_x$  a  $e_x$  jsou rovnoměrně rozložena v rámci roku,
3. přesný věk jedinců z  $n_x$  a  $w_x$  v okamžiku přistěhování, vystěhování nebo úmrtí je rovnoměrně rozložen na intervalu  $(x, x + 1)$ ,
4. k přistěhování a vystěhování dochází rovnoměrně během roku,
5. přesný věk jedinců z  $D_x$  v okamžiku úmrtí je rovnoměrně rozložen na intervalu  $(x, x + 1)$ .

Z těchto předpokladů vyplývá, že průměrný přesný věk jedince z  $b_x$  na počátku sledovaného období je  $x + \frac{1}{2}$ . Průměrný přesný věk jedince z  $e_x$  na konci sledovaného období je rovněž  $x + \frac{1}{2}$ . A stejně tak je průměrný přesný věk roven  $x + \frac{1}{2}$  pro jedince z  $n_x$  v okamžiku přistěhování, pro jedince z  $w_x$  v okamžiku vystěhování a pro jedince z  $D_x$  v okamžiku úmrtí. Z toho vyplývá:

$$E_x^C = l_x^\circ + \frac{1}{2} b_x - \frac{1}{2} e_x + \frac{1}{2} n_x - \frac{1}{2} w_x - \frac{1}{2} D_x. \quad (1.19)$$

Nyní upravujeme dále. Počet osob ve věku  $x$  na konci sledovaného období, tedy na konci sledovaného roku, lze vyjádřit jako počet osob, které se ve sledovaném roce dožily věku  $x$ , dále přičteme polovinu z počtu osob, které se během tohoto roku přistěhovaly ve věku  $x$ , a od toho odečteme polovinu z počtu osob, které se vystěhovaly během roku ve věku  $x$ , a polovinu z počtu osob, které zemřely ve věku  $x$  ve sledovaném roce. Symbolicky vyjádřeno:

$$e_x = l_x^\circ + \frac{1}{2} n_x - \frac{1}{2} w_x - \frac{1}{2} D_x. \quad (1.20)$$



Po dosazení za  $\dot{l}_x$  do 1.19 dostáváme:

$$E_x^C = \frac{1}{2} (b_x + e_x), \quad (1.21)$$

to znamená střední stav populace ve věku  $x$ . Vidíme tedy, že  $E_x^C$  lze odhadovat buď jako průměr z počátečního a konečného stavu populace ve věku  $x$ , nebo jako stav populace ve věku  $x$  v polovině roku.

### **Definice 1.6:**

Nechť  $D_x$  označuje počet úmrtí v populaci ve věku  $x$  během sledovaného období. Nechť dále  $E_x^C$  označuje centrální expozici populace ve věku  $x$ . Potom podíl  $\frac{D_x}{E_x^C}$  nazýváme věkově specifickou mírou úmrtnosti a značíme ji  $\mathring{m}_x$ .

Pro různá  $x$  dostáváme posloupnost věkově specifických měr úmrtností  $\{\mathring{m}_x\}_1^{\omega-1}$ . Pro  $x = 0$  se tato míra neuzívá, z důvodu neplatnosti předpokladu o rovnoměrném rozložení úmrtí na intervalu  $(0, 1)$  pro nově narozené jedince.

Pro porovnání úmrtnosti v různých populacích se používá jeden ukazatel získaný z řady hodnot  $\{\mathring{m}_x\}_1^{\omega-1}$ , a to obecná míra úmrtnosti  $\mathring{m}$ . Spočítáme ji jako podíl  $\frac{D}{E^C}$ , kde  $D$  je celkový počet úmrtí ve sledovaném období a  $E^C$  je celková expozice. Tato míra je ale závislá na věkové struktuře populace, a proto pro porovnání úmrtnosti v populacích ještě tuto míru standardizujeme, např. pomocí nějaké standardní věkové struktury nebo pomocí standardních věkově specifických měr úmrtností.

## **1.6 Vyrovnávání dat v úmrtnostních tabulkách**

Vyrovnávání odhadnutých dat v úmrtnostních tabulkách, tj. věkově specifických měr úmrtností, pravděpodobností úmrtí či jiných úmrtnostních charakteristik, je nutné ke snížení nesystematických výkyvů způsobených tím, že teoretické hladké hodnoty odhadujeme statistickými postupy z reálných dat. Vyrovnávání úmrtnostních tabulek je potřebné především v případě, že provádíme jejich konstrukci sami na základě menšího množství dat. Úmrtnostní tabulky konstruované Českým statistickým úřadem či jinými centrálními statistickými institucemi jsou obvykle už vyrovnány systémem vyhlazovacích filtrů.

Pro vyrovnávání úmrtnostních tabulek existuje řada metod. Metody rozdělíme dle [6] na

- mechanické,
- analytické.

Mechanické vyrovnávání úmrtnostních tabulek provádíme např. pomocí klouzavých průměrů, jejichž obecný tvar je

$$\hat{q}_x = \sum_{k=-r}^s a_k q_{x+k}, \quad (1.22)$$

kde  $\{a_k\}$  je vhodný systém vah,  $q_x$  jsou vyrovnávané odhadnuté pravděpodobnosti úmrtí a  $\hat{q}_x$  jsou výsledné vyrovnané pravděpodobnosti úmrtí.

Příkladem takovéto metody je Schärtlinova devítibodová metoda (viz [6]). Ta je popsána vzorcem

$$\hat{q}_x = \frac{1}{27} (-q_{x-4} + 2q_{x-2} + 8q_{x-1} + 9q_x + 8q_{x+1} + 2q_{x+2} - q_{x+4}). \quad (1.23)$$

Jedná se tedy o centrované vážené klouzavé průměry délky 9.

Dalším příkladem (viz [6]) je Wittsteinova devítibodová metoda tvaru

$$\hat{q}_x = \frac{1}{25} (q_{x-4} + 2q_{x-3} + 3q_{x-2} + 4q_{x-1} + 5q_x + 4q_{x+1} + 3q_{x+2} + 2q_{x+3} + q_{x+4}).$$

Metodika používaná ČSÚ při vyrovnávání pravděpodobností úmrtí pro věky 4–80 let v úmrtnostních tabulkách má následující tvar:

$$\begin{aligned} \hat{q}_x = \frac{1}{315} & (-30q_{x-3} + 45q_{x-2} + 90q_{x-1} + 105q_x + \\ & + 90q_{x+1} + 45q_{x+2} - 30q_{x+3}). \end{aligned} \quad (1.24)$$

Analytické vyrovnávání je vyrovnávání matematickou křivkou (viz také kapitola 3.3.1 Zákony úmrtnosti). Pro odhad parametrů prokládané křivky lze použít různé statistické metody jako např. metodu maximální věrohodnosti, metodu nejmenších čtverců atd. Vyrovnávání lze také provést pomocí splinů.

Dle [6] je v praxi osvědčeným způsobem vyrovnávání  $\ln q_x$  lomenou čarou. Řeší se optimalizační úloha

$$\min_{a, b, c > 0} \left\{ \sum_{x=x_0}^{x_1} L_x |\ln q_x - a| + \sum_{x=x_1+1}^{x_2} L_x |\ln q_x - (b + cx)| \right\}, \quad (1.25)$$

s vhodně zvoleným bodem zlomu  $x_1$  a s váhami  $L_x$  (počet  $x$ -letých v dané populaci). Jestliže  $\hat{a}$ ,  $\hat{b}$ ,  $\hat{c}$  jsou minimalizující hodnoty 1.25, pak je obvyklé použít pro výpočet vyrovnaných pravděpodobností úmrtnosti vzorec

$$\hat{q}_x = \begin{cases} \exp(\hat{a} + \hat{\sigma}^2/2) & \text{pro } x = x_0, x_0 + 1, \dots, x_1, \\ \exp(\hat{b} + \hat{c}x + \hat{\sigma}^2/2) & \text{pro } x = x_1 + 1, x_1 + 2, \dots, x_2, \end{cases}$$

kde  $\hat{\sigma}_2$  je vhodně odhadnutý reziduální rozptyl v 1.25.

Uvedené metody jsou poměrně jednoduché a v praxi vykazují dobré výsledky.

# Kapitola 2

## Kohort efekt

### 2.1 Pojmy a pozorování - kohort efekt

Soustředme se nejdříve na vymezení pojmu kohorta. Někdy se totiž stává, že pojmy kohorta a generace bývají zaměňovány. Za generaci je v demografii označována skupina osob se stejným kalendářním rokem narození (např. generace narozených v roce 1984). Někdy je takto označován i soubor osob narozených v intervalu několika let. Z hlediska sociologie znamená generace skupinu osob, která je spojená dobově podmíněným stylem myšlení a jednání, žijící ve shodných historických a kulturních podmínkách.

Oproti tomu pojmem kohorta označujeme skupinu osob určenou nějakými konkrétními demografickými událostmi ve stejném časovém intervalu, např. kohorta dětí narozených v roce 1997, kohorta žen, které uzavřely manželství v roce 2004 apod. Kohorta je základem pro tzv. kohortní analýzu. Její podstatou je sledování, co se děje s konkrétní kohortou během chronologického vývoje. Např. u sňatkové kohorty 1995 se sleduje, jak se její příslušníci rozvádějí, umírají, kdy a kolik se jim rodí děti atd. (viz [24]).

Kohorta můžeme mít různou šířku, to znamená, že do kohorty zahrnujeme jedince, určené danou demografickou vlastností, narozené v časovém intervalu dané délky, resp. šířky.

Kohort efektem obecně rozumíme konkrétní ráz skupiny osob ohraničené časem a danou demografickou vlastností, kterou je pro účely této práce rok narození.

Pozorování vývoje úmrtnosti v některých populacích (vymezených územními celky) přineslo zajímavý výsledek. Obecně platným pravidlem je klesající trend v úmrtnosti, a to především díky zlepšujícím se životním podmínkám a kvalitnější lékařské péči. Bylo zjištěno, že existují generace, ve kterých je míra zlepšení v úmrt-

nosti větší než v generacích předchozích i v generacích následujících. Tento pozorovaný jev byl nazván „cohort effect“. V textu je používán počestěný výraz kohort efekt a je tím tedy míněna tato konkrétní situace bez ohledu na širší obecný význam kohort efektu.

### 2.1.1 Velká Británie

Zmíněná pozorování proběhla např. ve Velké Británii. Pozorování tamní populace dokazují, že žijící členové generace narozených v pozdních dvacátých letech a třicátých letech 20. století vykazují rychlejší zlepšování v úmrtnosti během osmdesátých a devadesátých let 20. století než ostatní předchozí i následující generace (viz [15]). Vzhledem k tomu, že tito lidé jsou ve věku odchodu do důchodu, má tento efekt podstatný význam pro podnikání v oblasti životního a penzijního pojištění. Při používání dřívějších úmrtnostních tabulek ve výpočtech bez zohlednění pozorované vlastnosti této generace by tedy nepřesné předpovědi budoucího vývoje úmrtnosti mohly mít nepříznivý dopad na oblasti sociálního a důchodového pojištění. Za střed kohorty se považuje generace narozená v roce 1931. Obdobný trend byl zjištěn rovněž u osob pobírajících starobní důchod a u osob se sjednaným životním pojištěním, ale středem této kohorty je generace narozená v roce 1926 (dle [20]).

Díky tomuto pozorování byla ve Velké Británii zahájena rozsáhlá diskuse a následně i výzkum kohort efektu a jeho vlivu na projekce úmrtnosti. Britské seskupení The Actuarial Profession se úmrtností zabývá ve speciálním projektu The Continuous Mortality Investigation, jehož výsledky jsou publikovány a předávány odborné veřejnosti k posouzení a k případnému připomínkování. Důležitou součástí tohoto projektu je rovněž snaha o ustálení termínů v projekci úmrtnosti (viz [17]). V projektu jsou vytvářeny projekce úmrtnosti za platnosti různých předpokladů o možném rozsahu, do kterého bude pozorovaný kohort efekt pokračovat a ovlivní tak budoucí úmrtnost.

### 2.1.2 Japonsko

Velká Británie není ovšem jedinou zemí, kde byl kohort efekt pozorován. Zaměřme se nyní na sledování trendů v úmrtnosti pro Japonsko. Pozorování rychlosti zlepšování úmrtnosti prokázalo, že kohorta žen narozených okolo roku 1915 zažívá větší míru zlepšování úmrtnosti než předchozí i následující generace (viz [15]). Příklad kohort efektu v japonské populaci je obzvláště zajímavý, protože kohort efekt zde přetrvává i do vysokého věku jedince (nad 80 let) z této kohorty. U mužské populace lze pozorovat dva kohort efekty, první pro narozené v období 1910–1920 a druhý

pro narozené v období 1935–1945. Neznamená to, že vlastnost pokračování kohort efektu ve vysokém věku jedinců musí v budoucnu vykazovat i kohort efekt ve Velké Británii nebo kdekoliv jinde, ale dokazuje to existenci této možnosti. Data pro tento výzkum pocházejí z Japonského statistického úřadu a Ministerstva zdraví a sociální péče v Japonsku.

Na druhou stranu samozřejmě existují i státy, kde kohort efekt nebyl pozorován nebo byl jen velmi slabý.

### 2.1.3 Kanada

Hlavním sociálním faktorem ovlivňujícím kanadskou populaci ve 20. století je migrace. Dle [12] byl podíl obyvatel narozených mimo Kanadu na počátku 20. století 22% z celkové populace a 16% z celkové populace na konci 20. století. Navíc těžiště zemí, ze kterých přicházeli imigranti do Kanady, se během druhé poloviny 20. století přesunulo z Anglie, USA a severní Evropy do Asie, což má také dopad na úmrtnostní schéma populace. V celkovém důsledku míry zlepšení úmrtnosti nevykazují znaky kohort efektu, neboť se zde více odráží přísun nových jedinců z různých prostředí než změny v životních podmínkách stávajících obyvatel.

## 2.2 Příčiny vzniku kohort efektu

Položme si otázku, jaká je příčina vzniku kohort efektu. Budeme se věnovat především podnětům vzniku tohoto jevu ve Velké Británii. Kohort efekt byl pravděpodobně způsoben značně odlišnými životními zkušenostmi generací vzdálených od sebe pouze několik let.

Nejvýznamnějším faktorem v odlišných životních zkušenostech těchto generací je vypuknutí druhé světové války. Jedním z možných důvodů pro rychlejší zlepšení úmrtnosti generací narozených v letech 1925 až 1944 jsou nepříznivé životní podmínky, kterým byly vystaveny předchozí generace, jenž válku zažily v aktivním věku. Příslušníci generací narozených po roce 1926 se nejspíš neúčastnili žádného z hlavních konfliktů války.

Dále strava v poválečné Británii měla zdravotní přínos a pozitivní dopad na vývoj dětí v tomto období. Sice zde probíhalo dávkové přidělování potravin, ale spotřeba čerstvé zeleniny či ryb byla vyšší než např. během 90. let (dle [19]). Naopak spotřeba masa a sýru byla v tomto poválečném období nižší. Různé výzkumy uvádějí, že na vzdory své střídmosti byla poválečná strava v mnoho ohledech zdraví prospěšná.

Dalším faktorem přispívajícím ke vzniku kohort efektu byly významné sociální změny ve společnosti. Bylo například zavedeno právo na bezplatné středoškolské vzdělání nebo založen the National Health Service, tj. veřejně financovaný systém zdravotní péče. Ve srovnání s předchozími lety se tedy sociální podmínky velmi zlepšily.

Mezi faktory zapříčiňující vznik kohort efektu také neodmyslitelně patří kouření. Dle [15] průměrná spotřeba cigaret mužů ve Velké Británii rovnoměrně rostla v období 1900 až 1940, zůstala konstantní v letech 1940–1960 a následně klesala v 60. až 80. letech 20. století. Od té doby spotřeba cigaret vykazuje spíše stabilní trend. Ženská populace vykazuje celkově nižší spotřebu než mužská, ale s obdobnými časovými trendy. Během druhé světové války byly cigarety zdarma distribuovány mezi vojáky a až po válce začal probíhat výzkum vyšetřující dopad kouření cigaret na zdraví. Jakmile začal negativní dopad kouření na zdraví vycházet najevo, spotřeba klesala. Generace narozené okolo roku 1920 mohly začít kouřit zhruba ve 30. letech, dostávaly cigarety za války a „kouřily“ přibližně 20 let, než se dostalo pozornosti zdravotním následkům kouření. Na druhou stranu narození okolo roku 1940 dosáhli dospělosti v čase, kdy už byly neblahé následky kouření zkoumány a diskutovány.

Sledování spotřeby cigaret má svůj podstatný význam při sledování vývoje úmrtnosti, neboť kouření je obecně rizikovým faktorem pro mnohé z hlavních příčin úmrtí. Existuje mnoho lékařských studií, které se dopady kouření na vznik různých onemocnění zabývají. Poskytují tak důkaz, že např. rakovina plic, nemoc úzce spojená s kouřením cigaret, je jednou z příčin existence kohortních trendů.

Dalším možným faktorem ovlivňujícím vznik kohort efektu je dle [15] počet narozených tzn. počáteční počet členů generace. Počet narozených ve 20. století má ve Velké Británii obecně klesající trend, ale existují 3 období, kdy proběhl tzv. baby boom (po první a druhé světové válce a v polovině 60. let). Za povšimnutí stojí, že léta narození generací vykazujících kohort efekt spadají do období mezi dvěma takovými výkyvy v porodnosti.

Jak vyplýne z úvodní části textu třetí kapitoly, různé vědní obory mají pro kohort efekt různé vysvětlení. Faktorem zůstává, že kohort efekt je zjevně způsoben kombinací několika faktorů.

## 2.3 Česká populace a kohort efekt

Nyní se můžeme zabývat problematikou existence kohort efektu v české populaci. Budeme vycházet z dat získaných o české populaci, a to konkrétně z počtu zemřelých dle dosaženého věku a z věkového složení obyvatelstva za roky 1980 až 2006. Tato

detailní data se podařilo získat z Českého statistického úřadu (dále jen ČSÚ) a ze zdroje [22].

### 2.3.1 Dostupnost dat

Omezenost dostupných detailních dat je způsobena především historickými změnami v územních celcích, na kterých byly informace zjišťovány. Např. v letech 1938–1940 se zmíněné demografické údaje vztahují na území tzv. protektorátu Čechy a Morava a rovněž údaje za období 1941–1944 nezahrnují pohraniční oblasti současné České republiky. Bylo tedy nutné tato data přepočítat na území České republiky.

Data za dřívější období, tzn. 1900 až 1913 z dob Rakouska-Uherska, jsou ve členění podle zemí neúplné a podrobně vedené jen v letech 1904, 1907, 1910 a 1913. Údaje za ostatní roky dle [23] většinou není možné seriózně rekonstruovat. Rovněž i rekonstrukce nevytvořených či nedochovaných ukazatelů za období první světové války je nereálná.

Existující data z období druhé světové války byla deformována chybějícími údaji o zemřelých mimo protektorát (tzn. i o zemřelých v koncentračních táborech), o zemřelých v nacistických věznicích atd. Rekonstrukce těchto dat sice proběhla, ale nemohla napravit všechny nepřesnosti v již publikovaných datech za protektorát. Výsledky rekonstrukce dat z období 1938 až 1944 byly akceptovány a používají se dále v časových řadách. Nicméně tato data postrádají detaily jako členění zemřelých dle věku či počty živě narozených dle věku matek apod. Detaily se zabývali různí demografové, kteří své výsledky publikovali v odborných časopisech. Tyto výsledky nebyly nikdy vydány v demografických příručkách, ale jsou v časových řadách obecně respektovány.

Nezajímavé z hlediska dostupnosti dat jsou rovněž roky 1945 a 1946, kdy byla oddělována data za české a německé obyvatelstvo. Odvozené ukazatele byly uváděny pouze za české obyvatelstvo. S časovým odstupem není již pro toto dělení důvod, proto v použitých časových řadách jsou uváděna data za všechny obyvatele bez ohledu na národnost.

Dle [23] jsou již publikovaná data získaná rekonstrukcí obecně respektována i přes občas zjevné nepřesnosti např. v datech o věkové struktuře obyvatelstva v padesátých letech, která jsou citována z publikace Věkové složení obyvatelstva v letech 1920–1937 a 1945–1979 (vydané ČSÚ v roce 1981).

Český statistický úřad nepovažuje rekonstruované údaje v časových řadách za období 1900–1913 a 1938–1945 za definitivní a předpokládá jejich upřesnění za spolupráce demografů a historiků.

Vzhledem k výše uvedeným faktům a vzhledem k tomu, že starší údaje jsou



dostupné pouze pro pětileté věkové intervaly, se zabýváme především údaji z let 1980 až 2006. Zde jsou podrobné a přesné údaje k dispozici. Na nich lze tudíž provádět podrobné analýzy úmrtnosti.

### 2.3.2 Studie existence kohort efektu v ČR

Pokusíme se z dostupných dat zjistit, zda lze kohort efekt sledovat i v České republice. Budeme analyzovat data zvláště pro mužskou a ženskou populaci. V textu uvedeme grafy pro mužskou populaci České republiky, grafy pro ženskou populaci České republiky jsou zobrazeny v přílohách.

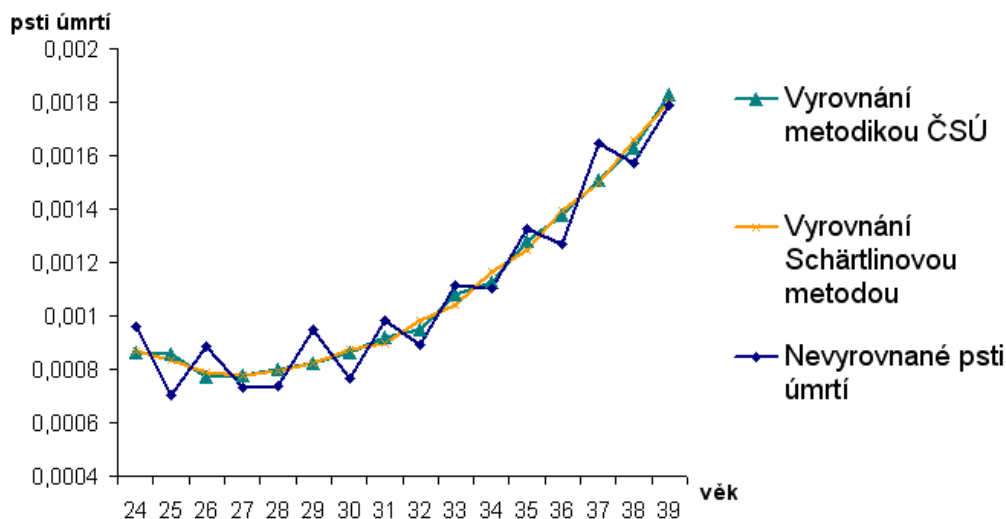
Dostupná data jsou věková složení obyvatel České republiky podle pohlaví a věku za roky 1980 až 2006, a to vždy stav na počátku a na konci roku (viz [25]-[38]). Spočteme tedy střední stav populace pro jednotlivé roky a věky a dle 1.21 tak odhadneme centrální expozice  $E_{x,t}^C$ , kde  $x$  určuje věk a  $t$  určuje rok této centrální expozice a  $x \in \{20, 21, \dots, 85\}$ ,  $t \in \{1980, 1981, \dots, 2006\}$ . Rozsah věkového intervalu je dán poměrně nízkou úmrtností v mladším věku (s výjimkou kojenecké úmrtnosti), kde nezaznamenáváme velkou změnu ve vývoji úmrtnosti, a prudkým nárůstem úmrtnosti ve věku vyšším než 85 let, kde data nejsou zcela vypovídající, neboť jde o hodnoty pozorované na nízkém počtu osob.

Druhým dostupným údajem jsou počty zemřelých podle věku, pohlaví a rodinného stavu z let 1980 až 2006 (viz [39] - [52]), kde rodinný stav není pro naši studii relevantní, a tak ho nadále zanedbáváme. Tyto údaje označujeme  $D_{x,t}$ , což interpretujeme jako počet úmrtí v dosaženém věku  $x$  v roce  $t$ . Podle definice 1.6 nyní můžeme vyjádřit věkově specifické míry úmrtnosti  $\hat{m}_{x,t}$ , kde opět  $t$  určuje rok pozorování této věkově specifické míry úmrtnosti.

Za zjednodušujících předpokladů, že intenzita úmrtnosti je v daném celočíselném věku konstantní a že přesný věk zemřelých jedinců v okamžiku úmrtí je rovnoměrně rozložen na  $(x, x + 1)$ , můžeme vypočítat pravděpodobnosti úmrtí v jednotlivých letech pomocí vztahu 1.18.

Nyní dle kapitoly 1.5 použijeme pro odstranění náhodných výkyvů vyrovnávací metodu používanou Českým statistickým úřadem 1.24. Vyrovnání pravděpodobností úmrtí dle různých metod přináší výsledky se zanedbatelným rozdílem ve vyrovnaných datech, viz obrázek 2.1. Rozpětí 16 let bylo na obrázku 2.1 zvoleno proto, aby na detailu bylo možné pozorovat rozdíly v nevyrovnaných datech a jejich vyrovnání. Pro celé věkové rozpětí by znázorněné křivky splynuly díky malým absolutním rozdílům téměř v jednu křivku.

Další vyrovnání provedeme k zamezení vzniku nežádoucí volatility úmrtnosti v daném věku při přechodu z roku na rok. Při sledování vývoje úmrtnosti po-



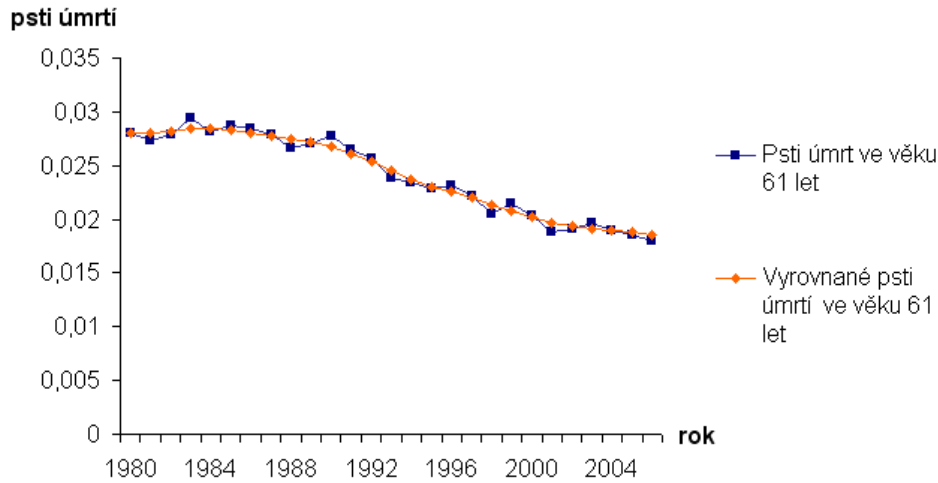
Obrázek 2.1: Pravděpodobnosti úmrtí mužské populace v roce 2006 ve věku 24–39 let.

rovnáváme úmrtnost v jednotlivých letech, proto by náhodné výkyvy v úmrtnosti z roku na rok v daném věku mohly výsledky výrazně zkreslit. Při studiu existence kohort efektu zase porovnáváme vývoj úmrtnosti ve sledovaných generacích, a to pomocí měř zlepšení úmrtnosti, které se dle definice 2.1 počítají na základě dat ze dvou po sobě jdoucích let. Tedy i pro další zkoumání kohort efektu je vyrovnání pravděpodobností úmrtí v daném věku přínosné.

Pro toto vyrovnání použijeme jednoduchou metodu centrovaných vážených klouzavých součtů délky 7 na již vyrovnaná data v jednotlivých letech  $\hat{q}_x$ . Matematický vzorec pro takové vyrovnání je následující:

$$\begin{aligned} \tilde{q}_{x,t} = \frac{1}{16} (\hat{q}_{x,t-3} + 2\hat{q}_{x,t-2} + 3\hat{q}_{x,t-1} + 4\hat{q}_{x,t} + \\ + 3\hat{q}_{x,t+1} + 2\hat{q}_{x,t+2} + \hat{q}_{x,t+3}), \end{aligned} \quad (2.1)$$

kde  $x$  označuje věk a  $t$  rok vyrovnávané či vyrovnané pravděpodobnosti úmrtí. U krajních hodnot, kde nejsou data pro vyrovnání klouzavými součty dostupná, tj. pro roky 1980–1982 a 2004–2006, byly váhy u chybějících hodnot nahrazeny 0 a odpovídajícím způsobem byl upraven i jmenovatel v 2.1. Graf zobrazující věkové



Obrázek 2.2: Pravděpodobnosti úmrtí mužské populace ve věku 61 let v letech 1980–2006.

vyrovnání  $\tilde{q}_{61,t}$  oproti tímto způsobem nevyrovnaným hodnotám  $\hat{q}_{61,t}$  je znázorněn na obrázku 2.2. Věk 61 byl vybrán náhodně pro názornost.

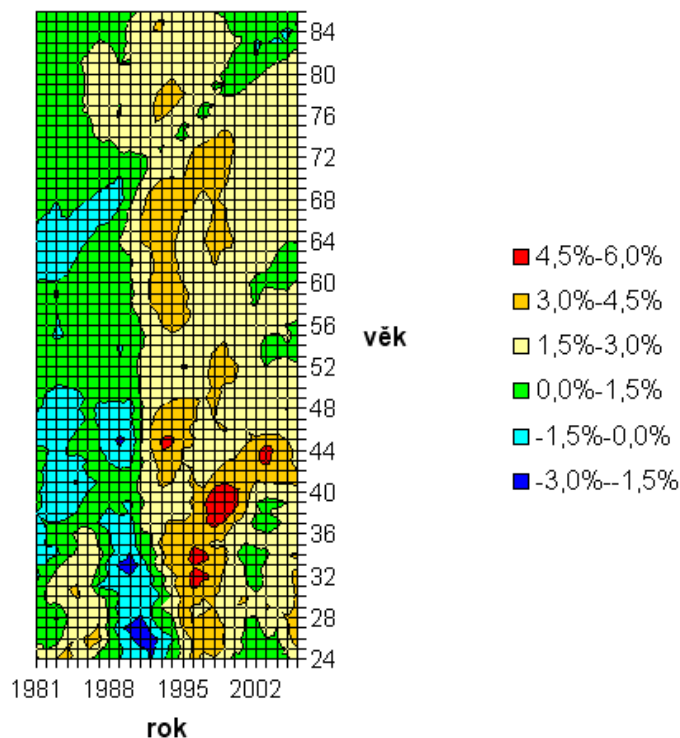
Pro zjištění, zda z takto dostupností omezených dat můžeme rozhodnout o existenci kohort efektu v české populaci, sestrojíme ukazatele, tj. míry zlepšení úmrtnosti, které budou vypovídat o rychlosti změny resp. poklesu vyrovnaných pravděpodobností úmrtí v čase pro jednotlivé věky.

**Definice 2.1:**

Definujeme míru zlepšení úmrtnosti v roce  $t$  pro věk  $x$  jako  $\frac{\tilde{q}_{x,t-1} - \tilde{q}_{x,t}}{\tilde{q}_{x,t-1}}$  a označíme ji  $M_{x,t}$ .

Z vypočtených měř zlepšení úmrtnosti jsme sestrojili graf (viz obrázek 2.3) pro českou mužskou populaci, tzv. heat map.

Heat map je označení pro grafickou reprezentaci dat, kde hodnoty jsou reprezentovány barvami v dvoudimenzionálním grafu. Na obrázku 2.3 je stupnice barev od modré po červenou, kde červená představuje největší zlepšení míry úmrtnosti, modrá nejmenší zlepšení resp. největší zhoršení. Kdyby existoval kohort efekt v zobrazených datech, pak by v grafu musela být významná oblast jedné (např. červené) barvy ve směru osy prvního kvadrantu kartézské soustavy souřadnic. Takovou vlastnost grafu

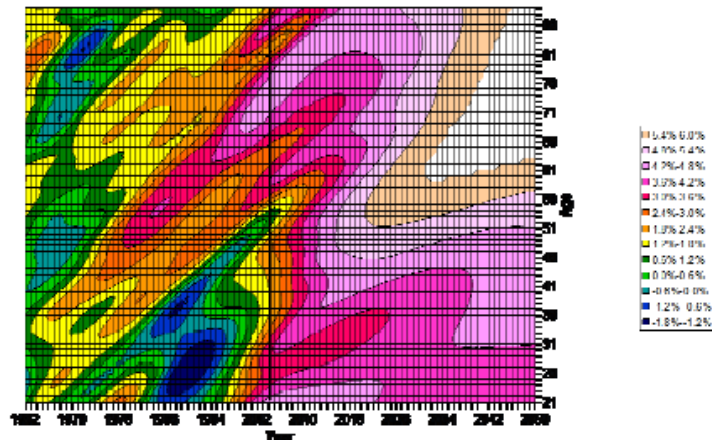


Obrázek 2.3: Heat Map měr zlepšení úmrtnosti pro českou mužskou populaci ve věku 24–86 v letech 1981–2006.

vidíme na obrázku 2.4, který znázorňuje obdobnou heat map pro britskou mužskou populaci, kde byl kohort efekt pozorován.

Pro ženskou část české populace je obdobný graf heat map zobrazen v příloze 1. Z tohoto grafu, stejně jako pro mužskou českou populaci, nevyplývá existence výraznějšího kohort efektu.

Kohort efekt můžeme obecně vysledovat i z jiných zobrazení dat. Jako další příklad jsme zvolili zobrazení měr zlepšení úmrtnosti mužské populace pro 5tileté věkové intervaly, vypočtené jako průměr hodnot za jednotlivé věky spadající do daného věkového intervalu, ve vybraných rocích 1981, 1986, 1991, 1996, 2001 a 2006, viz obrázek 2.5. Z toho obrázku by se dalo usoudit o existenci kohorty narozených v letech 1917–1921, která vykazuje menší zlepšení úmrtnosti než populace, které ji obklopují. Podle podrobnějšího grafu 2.3 se nejedná o příliš výrazný rozdíl v mírách



Obrázek 2.4: Heat Map měr zlepšení úmrtnosti pro britskou mužskou populaci včetně projekce dle [21].

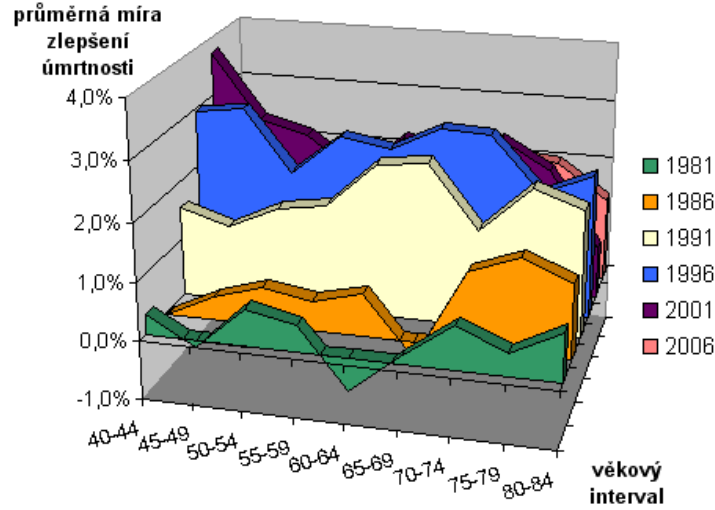
zlepšení a pro projekci úmrtnosti to hraje zanedbatelnou roli, neboť jedinci z této kohorty jsou dnes ve věku 85–90 let.

Pro porovnání přikládáme ještě obrázek 2.6, na kterém je sestaven podobný graf, ale pro britskou mužskou populaci, kde byl kohort efekt pozorován.

Zmíňme ještě, že tento kohort efekt byl pozorován i v ženské populaci Anglie a Walesu.

Z těchto grafů (a obdobných grafů pro ženskou populaci, viz přílohy) lze usoudit, že kohort efekt v české populaci na dostupných datech nelze vysledovat. Jeho existenci však nelze zcela vyloučit, protože data, která jsme analyzovali, pokrývají poměrně krátké časové období. Máme k dispozici i data za delší časový interval, ovšem tato data nejsou tak podrobná. Jedná se o střední stavy obyvatelstva a počty zemřelých pro Českou republiku za období 1950–2005 pro pětileté věkové intervaly (viz [53] a [54]). Vypočítáme tedy příslušné míry úmrtnosti  $\hat{m}_{x,t}$ , kde  $x$  je nyní označení věkového intervalu, pravděpodobnosti úmrtí, které vyrovnáme, a odpovídající míry zlepšení úmrtnosti  $M_{x,t}$ .

Když chceme pozorovat na těchto datech generační trend ve vývoji měr úmrtnosti, musíme brát data vždy po 5 kalendářních letech, aby se pozorovaná generace celá přesunula do další zaznamenané věkové kategorie. Za  $t$  budeme postupně dosazovat roky 1950, 1955, 1960, ..., 2005. Dále musíme použít modifikovanou míru zlepšení



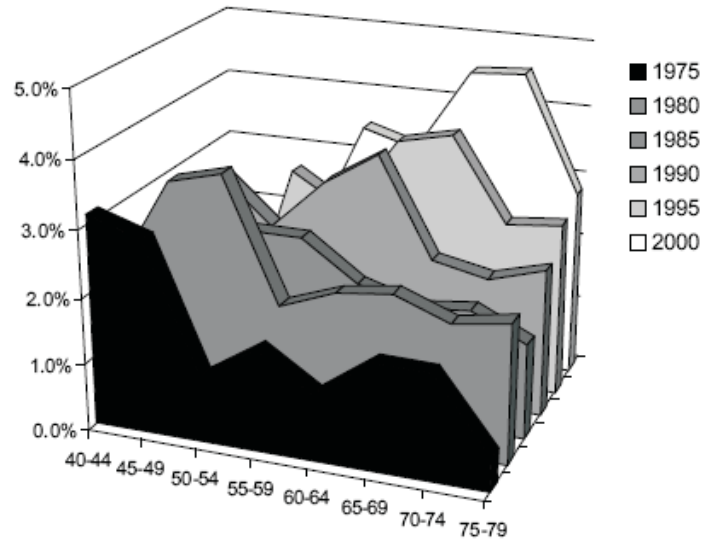
Obrázek 2.5: Průměrné míry zlepšení úmrtnosti dle věkových skupin a roku pro českou mužskou populaci.

úmrtnosti tvaru

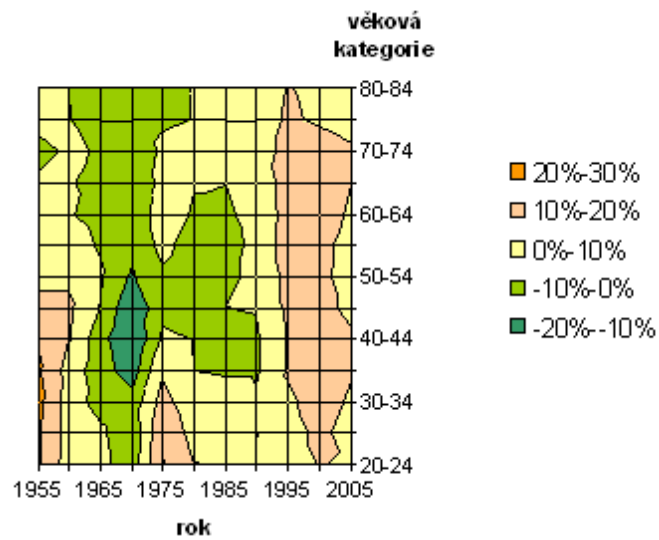
$$M_{x,t}^{mod} = \frac{\dot{m}_{x,t-5} - \dot{m}_{x,t}}{\dot{m}_{x,t-5}}, \quad (2.2)$$

a to proto, abychom zaznamenali vývoj úmrtnosti právě ve sledované generaci. Za  $x$  dosazujeme nikoliv jednotlivé věky, ale věkové kategorie dle zdrojových dat, tzn.  $x = \{20 - 24, 25 - 29, 30 - 34, \dots, 75 - 79, 80 - 84\}$ . Pokud bychom pro tato ne příliš podrobná data sestrojili heat map modifikovaných měř zlepšení úmrtnosti a druhou vybranou reprezentaci těchto modifikovaných měř, dostaneme grafy, ze kterých rovněž existence kohort efektu v české mužské populaci nevyplývá (viz obrázek 2.7 a 2.8).

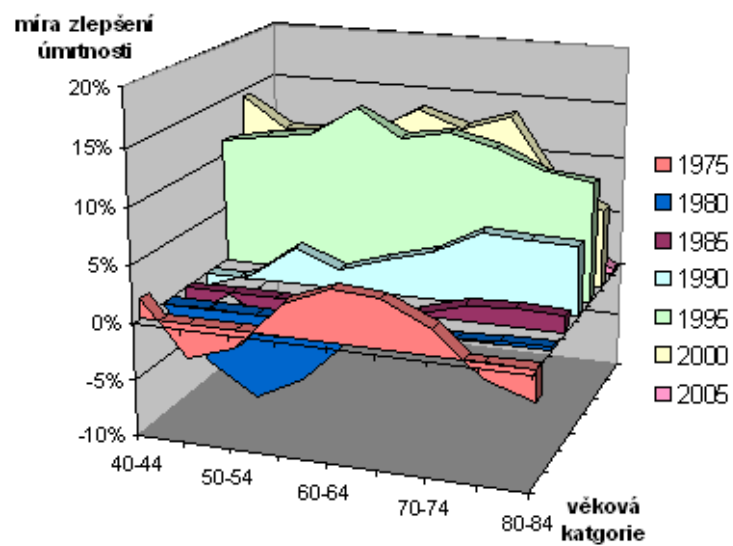
Na základě uvedených závěrů z grafů 2.3, 2.5, 2.7 a 2.8 usuzujeme, že existenci kohort efektu nelze v české mužské populaci jednoznačně prokázat. Stejně závěry plynou z obdobných grafů i pro ženskou část populace obyvatel ČR, viz přílohy. Proto v dalším studiu vývoje úmrtnosti v České republice a při projekcích úmrtnosti populace České republiky nebudeme vliv možného kohort efektu zohledňovat.



Obrázek 2.6: Průměrné míry zlepšení úmrtnosti dle věkových skupin a roku pro mužskou populaci Anglie a Walesu (dle [15]).



Obrázek 2.7: Heat Map modifikovaných měr zlepšení úmrtnosti pro českou mužskou populaci v letech 1955-2005.



Obrázek 2.8: Modifikované míry zlepšení úmrtnosti dle věkových kategorií a roku pro českou mužskou populaci.



# Kapitola 3

## Metody projekce úmrtnosti

### 3.1 Přidružené vědní obory a jejich vysvětlení trendů v úmrtnosti

Úmrtnost není jen záležitostí pozorování nějakého matematického trendu v jejím vývoji. Naopak je ovlivňována množstvím externích faktorů jako jsou ekonomické, sociologické či zdravotní podmínky prostředí, ve kterém se populace nachází. Proto se do posuzování vývoje úmrtnosti zapojují mnohé vědní obory.

#### 3.1.1 Epidemiologie

Epidemiologové se mimo jiné zabývají vlivem průběhu raného dětství na zdraví jedince v dospělosti, což bylo poprvé diskutováno ve 40. letech 20. století. Průběh dětství je ovlivňován různými faktory jako je životní prostředí či výživa dítěte. Z výzkumů byly učiněny závěry, že při predikci rizika úmrtnosti je rok narození důležitější než rok možného úmrtí. Tento přístup byl ale v padesátých letech zamítnut viz [15].

V nedávné době byla diskuse o dopadu podmínek v dětství na zdraví v dospělosti znovu otevřena. Pozornosti se dostává například vlivu porodní váhy na riziko vysokého krevního tlaku v dospělosti. Dle [15] bylo dokázáno, že nízká hmotnost kojence má za následek vyšší riziko koronárních onemocnění v dospělosti, nebo že oblasti v Anglii s vysokou úmrtností byly dříve charakteristické chudými životními podmínkami, což je demonstrováno vyšší kojeneckou úmrtností a malou tělesnou výškou dospělých.

Epidemiologové tedy navrhují zkoumat rozdíly mezi generacemi z hlediska např. porodní váhy, mateřské výživy a podobně. Z toho pak lze odvozovat možné dopady

na vývoj úmrtnosti. Tyto faktory mohou pomoci objasnit trendy v úmrtnosti.

### 3.1.2 Sociologie a ekonomie

V rámci sociologie a ekonomie probíhá zkoumání, jak jsou různé generace ovlivněné odlišnými sociálními a ekonomickými životními podmínkami. Například značný rozdíl v úmrtnostních mírách je mezi kohortou narozených v období 1916–1920 a mezi narozenými v období 1941–1945 (viz [15]). První kohorta zažila v dětství hospodářskou krizi 30. let a druhou světovou válku v rané dospělosti a navíc u podstatné části členů kohorty se před a během války velmi rozmohlo kouření cigaret. Oproti tomu generace narozených v letech 1941–1945 vyrůstala v období poválečné euforie. Bylo to období přidělového systému a regenerace a při vstupu této generace na trh práce byla ekonomika na vzestupu a prosperovala.

Výzkum tohoto typu pomáhá rozvíjet možná vysvětlení pro trendy v úmrtnosti a dává příležitost k pohledu na projekce úmrtnosti z širší perspektivy. Konkrétně kdybychom chtěli předpovědět míry úmrtnosti pro generaci 65tiletých, můžeme zohlednit fakt, že známe přesně ekonomické a sociální podmínky, které tuto generaci doposud provázely.

### 3.1.3 Demografie

Z hlediska demografie vývoj úmrtnosti ovlivňují tři hlavní faktory. Těmito faktory jsou okamžik narození a okamžik úmrtí (více viz [16]) a samozřejmě věk.

Okamžik narození je určující pro výše popsaný kohort efekt. Jde v podstatě o vnitřní (např. biologické) a vnější (např. socioekonomické) vlivy, kterým je jedinec vystaven od okamžiku narození a jejichž účinky se během jeho života akumulují a tím dávají za vznik rozdílným zdravotním a úmrtnostním rizikům pro jedince narozené v odlišných obdobích.

Druhým faktorem označeným jako okamžik úmrtí se rozumí proměnlivost úmrtnosti v čase, která působí na jedince ve všech věkových skupinách. Zahrnuje všechny historické události a životní prostředí ovlivňující faktory, například světové války, ekonomické krize, hladomory či pandemie nakažlivých onemocnění. Výkyvy v úmrtnosti týkající se okamžiku úmrtí mohou být rovněž důsledkem zlepšení situace ve veřejném zdravotnictví nebo průlomovým objevem v lékařství (např. nalezení nového léku či diagnostické techniky). Změna tohoto druhu vede k poklesu úmrtnostních měř ve všech věkových skupinách.

Třetí faktor, věk, ovlivňuje úmrtnost zřejmým způsobem. S rostoucím věkem úmrtnost roste. Jedinou výjimkou v tomto trendu je kojenecká úmrtnost, kde nejvyšší

úmrtnost sledujeme ihned po narození během následující 24 hodin. Z hlediska celého prvního roku života jedince je nejvyšší míra úmrtnosti v prvním měsíci jeho života, pak klesá.

Uvedené faktory ovlivňující úmrtnost nelze zkoumat samostatně, neboť mezi nimi existuje přímá souvislost:

okamžik narození + dosažený věk = možný okamžik úmrtí (tzn. 2. faktor).

Byly sestrojeny mnohé modely úmrtnosti zohledňující různou kombinaci těchto faktorů. V současné době je kladen větší důraz na věk a dobu (okamžik úmrtí) než na okamžik narození. Děje se tak, neboť ve vyspělých zemích úmrtnostní míry s rostoucím věkem rostou, a to až do přibližně sto násobku měř v mladém věku, a protože se zdravotní péče a životní podmínky velice zlepšily. Tím se i zvýšila očekávaná délka života jedince. Na druhou stranu analýza trendů v úmrtnosti z hlediska okamžiku narození dává vysvětlení pro rozdílnou rychlost ve zlepšování úmrtnosti mezi generacemi. To je ovšem poměrně zanedbatelný faktor v porovnání se zlepšováním díky příhodným podmínkám doby a v porovnání s vlivem věku na úmrtnost. Nicméně při modelování budoucí úmrtnosti pro účely ocenění např. důchodového pojištění hrají i relativně malé výkyvy v míře zlepšení úmrtnosti významnou roli.

## 3.2 Klasifikace metod projekcí

Existuje mnoho způsobů, jak přistupovat k předpovědi vývoje úmrtnosti. Tradičně se predikcí úmrtnosti zabývají aktuárové a demografové, nicméně v posledních době byly nové modely pro předpověď úmrtnosti vyvinuty v souvislosti s výzkumem v oblasti stárnutí např. v evoluční biologii, genetice či gerontologii. Při výběru vhodné metody projekce se posuzuje několik skutečností, a to především typ, rozsah a kvalita dat dostupných v okamžiku projekce. Obecně platná klasifikace metod není definována. Přirozeně lze ale metody klasifikovat například následujícím způsobem:

1. extrapoláční metody
  - (a) vztahující se k jednotlivým věkům,
  - (b) vztahující se k celému věkovému rozpětí,
2. metody založené na modifikaci standardního profilu úmrtnosti,
3. metody založené na epidemiologickém a biomedicínském výzkumu,
4. kombinace uvedených metod.

Negativním znakem extrapolačních metod je nedostatek informací o vlivech způsobujících budoucí změny v úmrtnosti, neboť extrapolací modelujeme pouze trend pozorovaný v datech z minulosti. Extrapolační metodou vztahující se pouze k jednomu věku je například aplikace vzorce geometrické posloupnosti na míru úmrtnosti v daném věku. Celého věkového rozpětí se týkají například zákony úmrtnosti.

Metodou založenou na modifikaci standardního profilu úmrtnosti nazýváme nějakou změnu věkově specifických měr úmrtnosti zvolené vzorové hypotetické nebo skutečné populace tak, abychom dostali předpovídané věkově specifické míry úmrtnosti. Takovou metodou je například Brassův dvou parametrický model a jeho modifikace. Do této kategorie by se dala zařadit i speciální třída modelů s nějakým externím faktorem. Ten se do modelu uměle přidává, aby přiblížil úmrtnost vzorové skutečné populace k úmrtnosti zkoumané populace, která vykazuje podobný charakter vývoje úmrtnosti jako populace vzorová.

Více komplexní metody přidávají k věkové struktuře faktor kohorty. Vznikají tak třífaktorové modely zohledňující věk, současnou dobu a kohortu (tzv. APC modely z anglického age-period-cohort models). Metody založené na epidemiologickém a biomedicínském výzkumu především vylepšují předpovědi úmrtnosti vytvořené podle předchozích metod. Základním kamenem těchto metod je projekce úmrtnosti dle příčiny úmrtí.

## 3.3 Podrobnější popis některých metod, jejich výhody a nevýhody

### 3.3.1 Zákony úmrtnosti

Zákony úmrtnosti popisují úmrtnost pomocí matematických funkcí věku. Počet parametrů se obvykle pohybuje od dvou do devíti. Parametrické funkce jsou vyjádřením věkově specifických úmrtnostních měr, pravděpodobností úmrtí a podobně. Parametrizace je aplikována na data, která jsou pozorována v uzavřené populaci, tedy např. na data pro kohortu jedinců narozených v určitém časovém intervalu. Zákony úmrtnosti se uplatňují při vyrovnávání dat a redukci chyb, při konstrukci úmrtnostních tabulek, při doplňování neúplných dat a pro předpověď vývoje úmrtnosti a další.

Příklady prvních zákonů zahrnují např. Moivreův, Gompertzův, Makehamův, Oppermanův, Thieleho či Weibullův model. Modely Heligmena a Pollarda nebo Rogerse a Littlea patří do kategorie novějších zákonů úmrtnosti. Mnoho funkcí zákonů úmrtnosti vzniklých v nedávné době má svůj základ v již dříve zavedených

funkcích. Můžeme tedy rozdělit funkce vyjadřující zákony úmrtnosti na ty, které nějakým způsobem vycházejí z Gompertzova modelu (viz [11]), a na ty, které mají jiné specifikace. Dále můžeme dělit tyto funkce na jednosložkové, které nerozlišují různé příčiny úmrtí, a vícesložkové, které různé příčiny úmrtí, významné pro danou věkovou skupinu, rozlišují. Některé zákony úmrtnosti jsou platné pro celé věkové rozpětí, jiné se vztahují pouze k vybraným věkovým intervalům.

Formulace matematických modelů zákonů úmrtnosti umožňuje jejich snadné a na první pohled zřejmě dělení na:

- polynomické,
- ostatní.

Dle [13] polynomické funkce nejsou přirozeně interpretovatelné, ale jsou oblíbenou interpolační a vyrovnávací technikou. Důvodem pro užití polynomických funkcí při modelování věkového schéma úmrtnosti je možnost aproximace většiny funkcí Taylorovým polynomem do libovolného stupně přesnosti.

Ostatní, tzn. ne polynomické, funkce jsou dle [13] především aditivní vícesložkové modely. Novější ne polynomické modely často předpokládají, že příčiny úmrtí jsou odlišné v dětství, dospělosti a stáří, proto zahrnují tři složky pro tyto životní etapy. Složky jsou definovány tak, aby neměly žádný dopad na úmrtnost ve věku, pro který nejsou určeny. Zároveň součet těchto složek je definován pro celé věkové rozpětí.

Původ zákonů úmrtnosti spadá do první poloviny 18. století. První pochází od francouzského matematika Moivre a má tvar

$$\mu(x) = \frac{1}{\omega - x}. \quad (3.1)$$

Interpretace je zřejmá. S věkem přibližujícím se k hodnotě  $\omega$  definované dle 1.5 klesá rozdíl ve jmenovateli a tedy roste intenzita úmrtnosti. Důležitým pro další vývoj zákonů úmrtnosti byl Gompertzův model z roku 1825. Prosazoval myšlenku, že schopnost jedince odolávat destrukci ubývá s rostoucím věkem úměrně velikosti této schopnosti. Pokud bychom zlomek  $\frac{1}{\mu(x)}$  označili jako tzv. míru odolnosti vůči destrukci, pak lze Gompertzova idea vyjádřit následujícím způsobem:

$$\frac{d\left(\frac{1}{\mu(x)}\right)}{dx} = -h \left(\frac{1}{\mu(x)}\right), \quad h > 0.$$

Nyní budeme výraz dále upravovat:

$$\begin{aligned} \ln\left(\frac{1}{\mu(x)}\right) &= -hx + k, \quad k, h > 0, \\ \mu(x) &= e^{-k} e^{hx}, \quad h, k > 0. \end{aligned}$$

Integrační konstanta  $k$  musí být kladná, protože jinak by  $\mu(x)$  bylo větší než jedna pro všechna  $x$ , což je v rozporu s jeho významem. Označíme-li  $b = e^{-k}$  a  $c = e^h$ , získáme tvar Gompertzovy funkce

$$\mu(x) = bc^x, \quad b \in (0, 1), \quad c > 1. \quad (3.2)$$

V roce 1867 britský aktuár Makeham doplnil Gompertzův model o složku nezávislou na věku

$$\mu(x) = a + bc^x. \quad (3.3)$$

Rozdělil tak příčiny smrti na související a nesouvisející s věkem. Modelováním úmrtnosti se zabýval i Opperman, který zformuloval model tvaru

$$\mu(x) = \frac{a}{\sqrt{x}} + b + c\sqrt{x}.$$

Dánský matematik Thiele navrhl v roce 1872 komplexní model pokrývající celé věkové rozpětí, na který navazuje mnoho autorů. Má tvar

$$\mu(x) = a_1e^{-b_1x} + a_2e^{-\frac{1}{2}b_2(x-c)^2} + a_3e^{b_3x}. \quad (3.4)$$

V 80. letech 20. století publikovali Heligman a Pollard svůj hlavní model

$$\mu(x) = A^{(x+B)^C} + De^{-E(\ln x - \ln F)^2} + GH^x$$

a jeho modifikace. Další modely navazující na Thieleho model se lišily například tím, že druhá komponenta modelu byla nahrazena parabolickou funkcí či dvojitou exponenciální funkcí. Některé modely dosáhly takového počtu parametrů, že už se jejich případná aplikace v praxi neobejde bez vysokých nároků na čas či výpočetní techniku. V roce 1994 Rogers a Little vytvořili jeden multiexponenciální model, který je dostatečný pro zachycení tvaru věkové struktury úmrtnosti a dalších demografických procesů. Má pět složek a třináct parametrů a autoři tvrdí, že jde o obecnou funkci, jejíž redukovanou formou lze jednoduše modelovat jakékoliv křivky úmrtnosti.

## Predikce založená na zákonech úmrtnosti

Predikce úmrtnosti z parametrických funkcí je obvykle založena na modelech časových řad. Často používaným nástrojem pro modelování časových trendů pozorovaných v datech demografických procesů je Box-Jenkinsova metodologie. Další metodou pro určení parametrů v zákonech úmrtnosti je nelineární regrese. Je třeba uvědomit si, že pokud má být předpověď co nejpřesnější, měly by být parametry závislé na čase. Potom dostaneme časovou řadu odhadnutých parametrů a předpovíáme její vývoj.

### 3.3.2 Brassova relační metoda

William Brass v roce 1968 navrhl modelovat změny úmrtnosti v čase, nikoliv průběh intenzity úmrtnosti (více viz [3]). Změny úmrtnosti navrhoval popisovat vzhledem k nějakému standardu, proto se metoda nazývá relační. Vychází z průběhu křivky počtu dožívajících se věku  $x$ , tzn.  $l(x)$ , který tvarem připomíná logistickou křivku (viz [9]). Transformací osy  $x$  vhodnou funkcí  $g(x)$  dosáhneme toho, že křivka počtu dožívajících má tvar logistické křivky a její předpis je  $l(x) = \frac{l(0)}{1+e^{g(x)}}$ . Dostáváme vyjádření transformační funkce

$$g(x) = \ln \frac{l(0) - l(x)}{l(x)} = \text{logit } l(x).$$

Předpokládáme, že lze všechny křivky počtu dožívajících převést na modifikaci jedné standardní logistické křivky. Modifikaci lze provést dvěma parametry, označme je  $a$  a  $b$ , vyjadřujícími posunutí a změnu rychlosti růstu. Transformační funkce se změně lineárně, tj.  $g^{mod}(x) = a + bg(x)$ . Model úmrtnosti zapíšeme ve tvaru

$$\text{logit } l(x, t) = a(t) + b(t)\text{logit } l^{standard}(x) + \varepsilon(x, t),$$

kde  $\varepsilon(x, t)$  je náhodná chyba s normálním rozdělením a dále řešíme jako regresní úlohu.

Nevýhodou tohoto modelu je, že nelze zaznamenat např. změny úmrtnosti jen pro určitou věkovou kategorii. Další obtíže mohou nastat při volbě standardní úmrtnosti. Lze například zkonstruovat počáteční odhad (např. průměr z jednotlivých ráďů vymírání, které popisujeme) a ten dále vylepšovat.

### 3.3.3 Lee-Carterův model

V roce 1992 dva američtí profesoři Ronald D. Lee a Lawrence R. Carter (viz [10]) navrhli model pro matici měř úmrtnosti

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \quad \text{tj.} \quad (3.5)$$

$$m_{x,t} = e^{a_x + k_t b_x + \varepsilon_{x,t}}, \quad t = t_1, t_1 + 1, \dots, t_1 + T - 1, \quad (3.6)$$

kde  $a_x$  lze interpretovat jako průměrnou věkově specifickou složku úmrtnosti,  $b_x$  vyjadřuje trend měnící se úmrtnosti vztažený k věku  $x$  a  $k_t$  představuje obecnou úroveň úmrtnosti v čase. Poslední člen modelu  $\varepsilon_{x,t}$  vyjadřuje odchylku modelu od logaritmu pozorovaných měř úmrtnosti a předpokládáme o něm, že odpovídá bílému šumu se střední hodnotou  $E \varepsilon_{x,t} = 0$  a rozptylem  $\text{Var } \varepsilon_{x,t} = \sigma_x^2 I$ , který je relativně malý. Přidáme ještě podmínky  $\sum_t k_t = 0$  a  $\sum_x b_x = 1$ .

K vyjádření  $a_x$  vysčítáme model přes všechna  $t$  a s použitím zavedených podmínek upravíme:

$$\sum_t \ln(m_{x,t}) = T a_x + b_x \sum_t k_t + \sum_t \varepsilon_{x,t}, \quad (3.7)$$

$$a_x = \sum_{t=t_1}^{t_1+T-1} \frac{\ln m_{x,t}}{T}. \quad (3.8)$$

Pro dané  $x$  se tedy  $a_x$  určí jako několikaletý průměr logaritmu měř úmrtnosti.

Pro odhad  $b_x$  a  $k_t$  se používají tři základní metody:

- metoda singulárního rozkladu (Singular Value Decomposition, SVD),
- metoda vážených nejmenších čtverců,
- metoda maximální věrohodnosti.

V [10] je použita metoda singulárního rozkladu matice  $\ln(m_{x,t}) - a_x$ . Pokud singulární rozklad matice zapíšeme jako  $Y\Phi Z^T$ , kde  $Y$  a  $Z$  jsou ortonormální matice a  $\Phi$  je diagonální matice, pak dle [10] vychází singulární rozklad

$$\text{SVD}[\ln(m_{x,t}) - a_x] \approx \varphi_1 y_x^{(1)} z_t^{(1)} \quad (3.9)$$

a odhady můžeme tedy vyjádřit jako

$$\hat{b}_x = y_x^{(1)}, \quad (3.10)$$

$$\hat{k}_t = \varphi_1 z_t^{(1)}, \quad (3.11)$$

neboť první singulární hodnota  $\varphi_1$  je mnohem větší a tedy i významnější než ostatní.

Jediný parametr závislý na čase je  $\hat{k}_t$ , který lze dle [10] projektovat do budoucna pomocí Box-Jenkinsovy metodologie (použitý model je ARIMA(0,1,0)). Předpovídaná věková a časově specifická míra úmrtnosti má pak tvar

$$m_{x,t_0+t} \approx \exp(a_x + b_x \hat{k}_{t_0+t}). \quad (3.12)$$

Kdybychom ze získaných odhadů parametrů zpětně spočítali matici měř úmrtnosti, dle [10] zjistíme, že se modelované a skutečné hodnoty neshodují. Důvodem je použití singulárního rozkladu na logaritmy měř úmrtnosti a ne přímo na míry úmrtnosti. V [10] je tedy navrhováno určit  $\hat{k}_t$  iteračními metodami ze skutečných měř úmrtnosti pomocí

$$D_t = \sum_x \exp(a_x + b_x \hat{k}_t) L_{x,t}, \quad (3.13)$$

kde  $D_t$  je celkový počet zemřelých v čase  $t$ ,  $L_{x,t}$  je počet jedinců v generaci ve věku  $x$  v čase  $t$ ,  $a_x$  a  $b_x$  jsou již dříve odhadnuté parametry.



### 3.3.4 APC modely

APC modely (tzn. age-period-cohort models) byly vyvinuty k rozlišení vlivu současné doby a vlivu kohorty na úmrtnost. Vliv současné doby odráží současné faktory jako je zdraví populace apod., zatímco ve vlivu kohorty se odráží historické faktory. Parametry modelu popisují trajektorii kalendářního roku a kohort efektu, který je daný parametry popisujícími věkovou strukturu v úmrtnosti. Obecně se dle [13] jedná o logaritmicko-lineární model tvaru

$$\ln \lambda_{ijk} = \alpha_i + \beta_j + \gamma_k, \quad (3.14)$$

kde  $\alpha_i$ ,  $i = 1, \dots, I$  vyjadřuje logaritmicko-lineární vliv věku,  $\beta_j$ ,  $j = 1, \dots, J$  vyjadřuje logaritmicko-lineární vliv současné doby a  $\gamma_k$ ,  $k = 1, \dots, K$  vyjadřuje logaritmicko-lineární vliv kohorty. Označení  $\lambda_{ijk}$  je použito pro míru úmrtnosti určenou všemi třemi uvedenými faktory. Omezení pro parametry jsou následující:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0. \quad (3.15)$$

Parametry jsou na sobě lineárně závislé,  $j = i + k$ , a proto není možné získat jedinečnou množinu parametrů  $\alpha_i$ ,  $\beta_j$  a  $\gamma_k$ , která by optimálně modelovala pozorovaná data. Dle [13] lze nějaký ze všech možných modelů určit pomocí metody Poissonovy maximální věrohodnosti nebo metodou vážených nejmenších čtverců.

Jedná se o oblíbený model v epidemiologii při určování úmrtnosti dle příčin smrti.

# Kapitola 4

## Aplikace zvolené metody na data České republiky

### 4.1 Výběr metody aplikované na data ČR

Vzhledem k tomu, že jsme z analýzy pravděpodobností úmrtí v kapitole 2.3 učinili závěr o převaze vlivu doby na vývoj úmrtnosti nad vlivem kohorty, budeme vybírat z modelů zohledňující věk a současnou dobu.

Použijeme Lee-Carterův model, který je relativně jednoduchý. Má několik parametrů, ale pouze jeden parametr je závislý na čase. Jde tedy o odhad parametrů modelu z dostupných dat a následný odhad budoucích hodnot jediného parametru. Pak už můžeme získat odhad budoucích měr úmrtnosti.

Obecně platí, že k projektování budoucí úmrtnosti musí být úmrtnost nejdříve dostatečně přesně modelována. Ovšem ne všechny modely úmrtnosti jsou vhodné pro její předpovídání. Např. modely s mnoha parametry sice vystihují výborně tvar křivky úmrtnosti, ale na druhou stranu je velký počet parametrů limituje v možnosti predikce úmrtnosti.

Z tohoto hlediska je Lee-Carterův model výhodný. Dominantní je v něm časová složka, zatímco věkové složky jsou v čase zafixovány. Časová složka je projektována pomocí metody autoregresních integrovaných klouzavých průměrů.

Lee-Carterova metoda má dle [2] řadu výhod, např. jde o úsporný demografický model, z hlediska počtu parametrů, který je kombinovaný se standardní statistickou metodou pro časové řady. Další předností je, že metoda nevyžaduje žádné subjektivní odhady. Dále je výhodné, že předpověď je založena na stabilních dlouhodobých trendech a že jsou touto metodou rovněž získány intervaly spolehlivosti pro předpovídaný vývoj úmrtnosti.

Hlavní nedostatek Lee-Carterovy metody je dle [2] předpoklad o neměnnosti věkových složek modelu v čase. Z modelování úmrtnosti v praxi vyplývá návrh jednoduchého řešení tohoto problému, což je vytvořit predikci na takových datech, aby předpoklad o neměnnosti věkové složky byl použit pouze na dobu přibližně půl století. Při použití Lee-Carterova modelu v sedmi ekonomicky nejvyspělejších zemích světa na data z období 1950–1994 bylo totiž identifikováno univerzální schéma poklesu úmrtnosti charakterizované téměř lineárním poklesem právě v dominantní časové složce modelu.

Od vzniku Lee-Carterova modelu byla navržena řada jeho modifikací, které mimo jiné právě potlačují nedostatek v časové neměnnosti věkové složky modelu. Tyto změny ale bývají často na úkor jednoduchosti a úspory parametrů v modelu, proto použijeme pro předpověď úmrtnosti v ČR původní Lee-Carterův model.

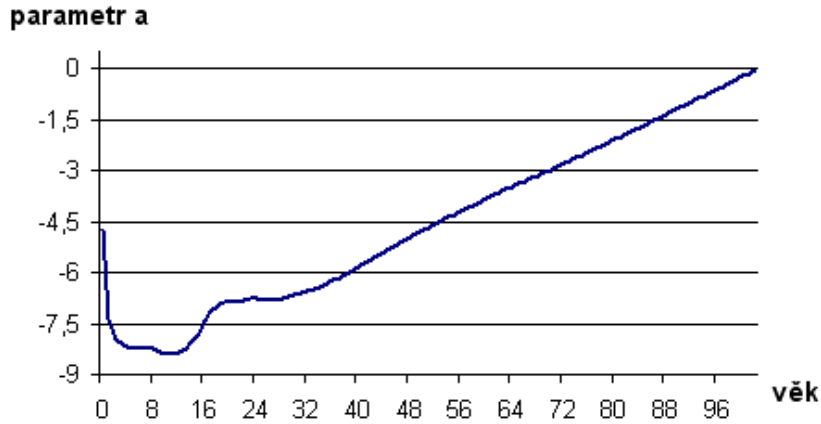
## 4.2 Použití vybrané metody

Lee-Carterovu metodu použijeme zvláště na mužskou a zvláště na ženskou populaci České republiky. Opět zde zaznamenáme výsledné grafy pro mužskou populaci. Grafy pro ženskou část populace jsou uvedené v přílohách. Data, ze kterých budeme úmrtnost projektovat, jsou podrobně popsána a analyzována v kapitole 2.3. Jde o data za období 1980–2006.

Z grafů heat map pro mužskou i ženskou populaci (viz obrázek 2.3 a příloha 1.) můžeme vysledovat změnu ve zlepšování úmrtnosti na přelomu osmdesátých a devadesátých let. Je třeba zvážit, na kolik by tento mírný zlom mohl ovlivnit výsledky projekce. Aby tato změna neovlivnila výsledky projekce, mohli bychom použít pro projekci data až z devadesátých let. Ale vzhledem k tomu, že by pak období, ze kterého v projekci vycházíme, bylo už velmi krátké, použijeme data již od roku 1980.

Pro všechny jednotlivé věky dle vzorce 3.8 spočítáme hodnoty parametru  $a_x$ . Parametr  $a_x$  je tedy vyjádřen jako průměr logaritmů měr úmrtností  $m_{x,t}$ ,  $t \in \{1980, 1981, \dots, 2006\}$ ,  $x \in \{0, \dots, 103\}$ . Výsledné hodnoty parametru  $a_x$  jsou znázorněny na obrázku 4.1. Křivka znázorňující závislost hodnoty parametru  $a_x$  na věku je dle očekávání v nízkých věcích klesající kvůli vyšší kojenecké úmrtnosti a pak už je s rostoucím věkem rostoucí, což je v souladu s interpretací  $a_x$  jako průměrné věkově specifické složky úmrtnosti.

Parametr  $b_x$  odráží míru vlivu časového vývoje úmrtnosti na věk  $x$ . Z obrázku 4.2 můžeme vyčíst, že se úmrtnost snižovala v čase hlavně v nízkých věcích, především u novorozenců se snížila několikanásobně. Nejmenší snížení úmrtnosti v čase bylo zaznamenáno ve věku kolem 20ti let. Pro další věky už nejsou zaznamenané velké



Obrázek 4.1: Parametr  $a_x$  pro mužskou populaci.

výkyvy zlepšení úmrtnosti v čase. Klesající trend křivky ve vysokých věcích znamená, že úmrtnost nejstarších se snižuje mírněji. Parametry  $b_x$  jsou vypočteny pomocí metody singulárního rozkladu matice jako ve vztahu 3.10, tedy pomocí pouze prvního členu rozkladu.

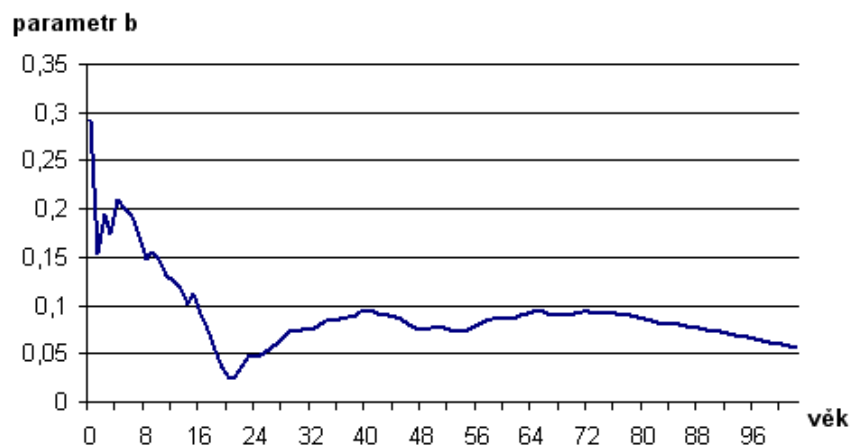
Zda je předpoklad o použití pouze prvního členu singulárního rozkladu matice  $\ln(m_{x,t}) - a_x$  správný, můžeme ověřit testem významnosti jednotlivých členů singulárního rozkladu matice (dle [1])

$$\frac{\varphi_i^2}{\sum_{j=1}^{27} \varphi_j^2}, \quad (4.1)$$

pro  $i \in \{1, 2, \dots, s\}$ , kde  $s$  je takové, že hodnota  $\varphi_s$  je nula nebo velmi blízká nule. Takto lze postupovat, neboť singulární hodnoty  $\varphi$  jsou uspořádány v diagonální matici sestupně a nesou informaci o významnosti sloupců matice  $Y$  a odpovídajících sloupců matice  $Z$ . Výsledek takového testu vidíme na obrázku 4.3. Z něj usuzujeme, že můžeme pro odhady  $b_x$  a  $k_t$  využít pouze první člen singulárního rozkladu dané matice.

Parametr  $k_t$  opět odhadneme z metody singulárního rozkladu matice dle 3.11. Parametr představuje pokles úmrtnosti v čase a tento trend můžeme pozorovat i na obrázku 4.4.

Pro posouzení použitelnosti hodnot parametru  $k_t$ , které jsou vypočteny z logaritmu měr úmrtnosti, jsme sestrojili i upravené parametry  $k_t$  dle vzorce 3.13. Pro



Obrázek 4.2: Parametr  $b_x$  pro mužskou populaci.

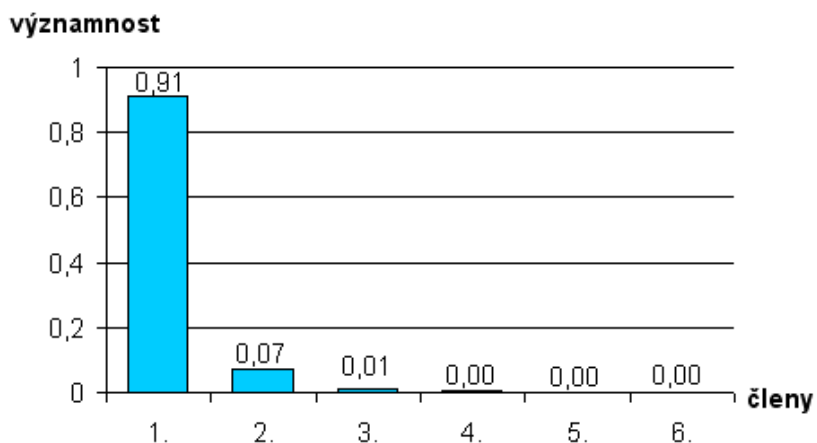
odlišení je budeme označovat  $K_t$ . Tyto vypočtené hodnoty jsou zobrazeny na obrázku 4.5.

Ze zobrazených hodnot parametru  $K_t$  (viz 4.5) usuzujeme, že bude lepší pro výpočet použít parametr  $k_t$  získaný metodou singulárního rozkladu, neboť tyto hodnoty méně kolísají v čase než hodnoty parametru  $K_t$  a skoro lineárně klesají. To nám vzhledem k interpretaci tohoto parametru, že jde o pokles úmrtnosti v čase, vyhovuje více. Pro rostoucí hodnoty úmrtnosti by parametr  $k_t$  samozřejmě rostl.

Na obrázku 4.4 vidíme, že zhruba v devadesátých letech klesala úmrtnost mužské populace rychleji než v okolních obdobích. Jinak jde o téměř lineární pokles úmrtnosti mužské populace, což je ovlivněno mimo jiné poměrně krátkou délkou dat, ze kterých parametr odhadujeme.

Nyní máme hodnoty parametrů určené a můžeme se věnovat projekci. Na čase je závislý jediný parametr  $k_t$ , který dle [10] budeme projektovat pomocí Box-Jenkinsovy metodologie, konkrétně pomocí modelu ARIMA (1,1,0). V [10] je použit model ARIMA(0,1,0), který ale není Lee-Carterovu metodou pevně definován jako jediný možný. Náš použitý model byl identifikován na základě analýzy hodnot autokorelační a parciální autokorelační funkce  $k_t$  (více viz [7]).

Zvoleným modelem použitým pro predikci vývoje parametru  $k_t$  jsme získali předpověď parametru na dalších 54 let, tj. do roku 2060 (viz 4.6). Vzhledem k poměrně krátké historii, ze které vycházíme, je vzdálenost mezi dolním a horním odhadem vývoje parametru  $k_t$  s rostoucím časem poměrně velká. Horní a dolní odhad je konstruován na 95%ní hladině spolehlivosti. Výsledek predikce parametru  $k_t$  včetně



Obrázek 4.3: Významnost členů singulárního rozkladu matice úmrtností pro mužskou populaci.

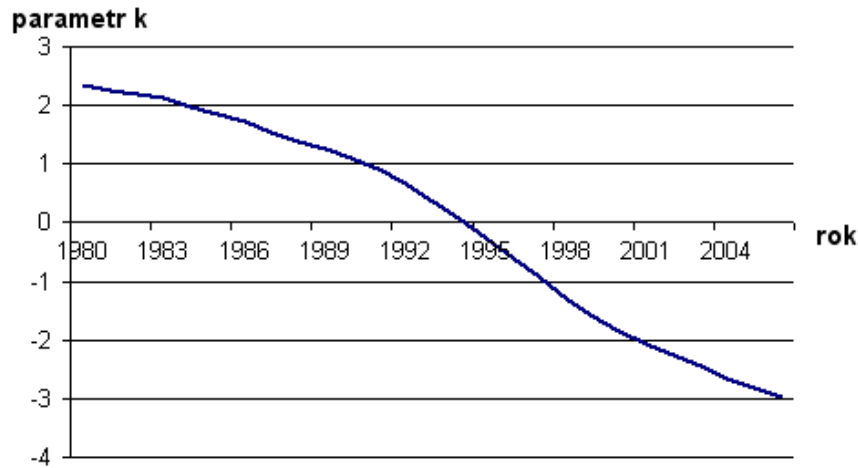
zmíněného horního a dolního odhadu, tj. intervalu spolehlivosti, je zobrazen na obrázku 4.6.

Na základě všech získaných hodnot parametrů  $a_x$ ,  $b_x$  a  $k_t$  nyní dle 3.12 vypočteme predikované věkově specifické míry úmrtnosti  $m_{x,t}$ , pro  $x \in \{0, 1, \dots, 103\}$  a  $t \in \{2007, 2008, \dots, 2060\}$ . Přesné predikované hodnoty věkově specifických měr úmrtnosti jsou uvedeny v příloze.

Pro ilustraci výsledků předpovědi je na obrázku 4.7 vyobrazena míra zlepšení úmrtnosti pro věky 24, ..., 96. Hodnoty míry zlepšení úmrtnosti pro roky 1980–2006 byly pro účely tohoto obrázku rovněž vypočteny dle 3.12. Z projekce vyplývá, že v budoucnu je očekávaná míra zlepšení úmrtnosti nejmenší pro vysoké věky 90 let a více a pro věky 19–27 let. Naopak největší míra zlepšení úmrtnosti se předpokládá v nejnižších věcích, zhruba do věku 12ti let. Z obrázku 4.7 ještě můžeme vyčíst, že z pohledu věkových skupin se očekává větší míra zlepšení úmrtnosti kolem věku 40ti a 65ti let.

Pro lepší názornost výsledků jsme vypočítali předpovídanou střední délku života (viz vzorec 1.6). Vybrané hodnoty střední délky života mužské populace a jejich porovnání jsou uvedeny v tabulce 4.1. Procentuálně největší zlepšení je v tabulce zaznamenáno pro věk 80ti let, nejmenší pak pro střední délku života při narození.

Další možné znázornění výsledků projekce úmrtnosti, tzn. klesající trend ve vývoji úmrtnosti české mužské populace, je zachyceno na obrázku 4.8. Jde o tzv. procento přežívajících a určuje se pomocí pravděpodobností přežití. Počet narozených

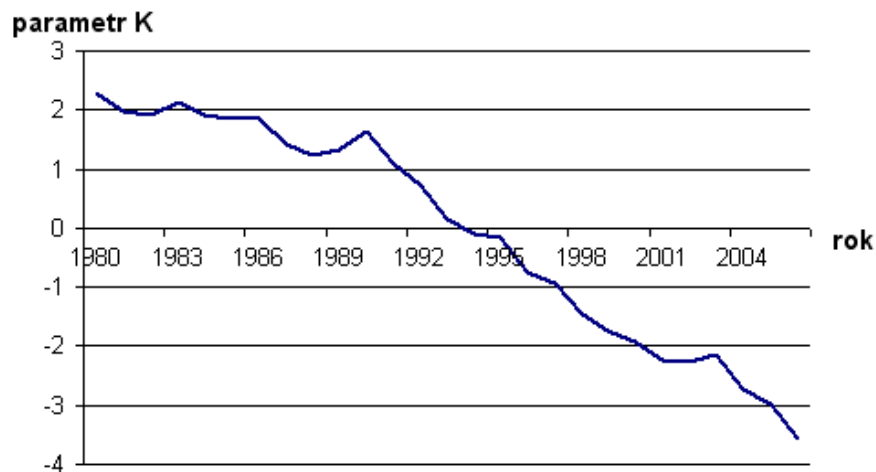


Obrázek 4.4: Parametr  $k_t$  pro mužskou populaci.

představuje 100% a tento počet se vynásobením příslušnou pravděpodobností přežití postupně snižuje (dle 1.1). V čase se pravděpodobnosti přežití v daném věku zvyšují, z čehož vyplývá „zobdelníkující“ se tvar výsledných křivek pro vzestupně řazené roky, viz obrázek 4.8. Znamená to, že z procentuálního hlediska se stále více jedinců populace bude dožít vyšších věků. Např. zatímco z pravděpodobností přežití pro rok 1980 jsme vypočetli, že věku 80ti let se dožije necelých 20% z nově narozených jedinců, z pravděpodobností přežití pro rok 2060 předpovídáme, že věku 80ti let se dožije až 67% z nově narozených jedinců.

Všechny předložené výsledky tedy dle očekávání potvrzují hypotézu celkem výrazného snižování úmrtnosti v populaci České republiky. Projekce vychází z poměrně krátkého časového úseku dvaceti sedmi let, kdy bylo v důsledku významného rozvoje v různých vědních oblastech, jako např. v lékařství, zaznamenáno výrazné zlepšení v úmrtnosti. Otázkou zůstává, zda je možné, aby tento nastavený trend poměrně velkého zlepšování úmrtnosti vydržel ve stejné intenzitě i pro následující roky, v našem případě po dobu 54 let. Na to má vliv mnoho externích faktorů, jako je například výskyt nové choroby. Omezení vlivu těchto externích faktorů lze dosáhnout za podmínky, že budeme mít k dispozici data za delší časové období.

Výsledky projekce jsme porovnali s výsledky z populační prognózy vytvořené B. Burcinem a T. Kučerou a publikované v [5]. Tato studie je zaměřena kromě úmrtnosti rovněž na další složky např. na plodnost, migraci a změny ve věkové struktuře. Prognóza je tvořena třemi variantami budoucího vývoje, a to střední,



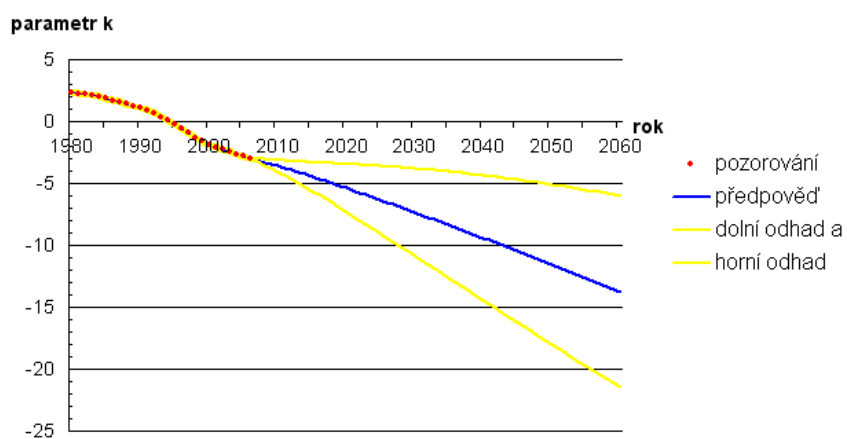
Obrázek 4.5: Upravený parametr  $K_t$  pro mužskou populaci.

v době sestavení prognózy nejpravděpodobnější směr vývoje, a nízkou, resp. vysokou variantou, které dle [5] pomyslně ohraničují pole, mimo které by budoucí změny prakticky neměly sledovanou populaci vyvést. Očekávaný vývoj každé ze složek byl také prognózován ve třech variantách, přičemž jednotlivé varianty souhrnné prognózy vznikly použitím analogicky označených variant očekávaného vývoje složek.

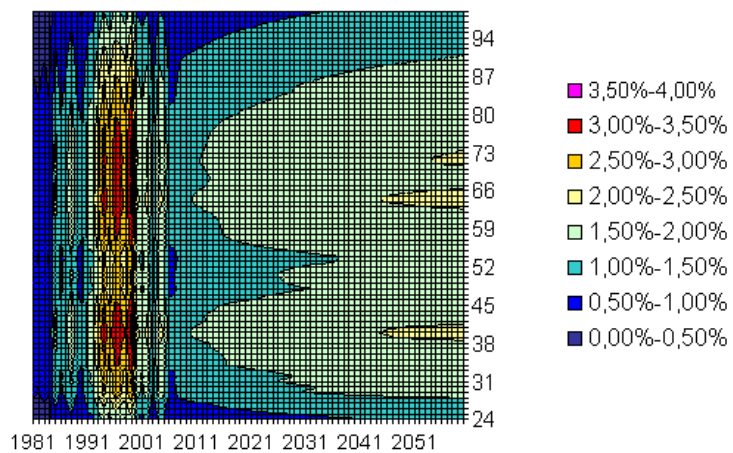
Výsledné hodnoty výzkumu [5] pro složku úmrtnosti jsou uvedeny na obrázku 4.9. Získané hodnoty z naší projekce Lee-Carterovou metodou se pohybují mezi nízkou a střední variantou populační prognózy [5]. Nejdříve jsou blíže nízké prognóze a pro vzdálenější roky se přesunou spíše ke střední prognóze.

Při posuzování výsledků projekce úmrtnosti je tedy třeba vzít v úvahu, že z poměrně krátkého časového období vytváříme dlouhou projekci. Tudíž se podmínky, které ovlivňovaly výchozí hodnoty pro předpověď vývoje úmrtnosti, mohou v budoucím čase, pro který je projekce vytvořena, změnit.





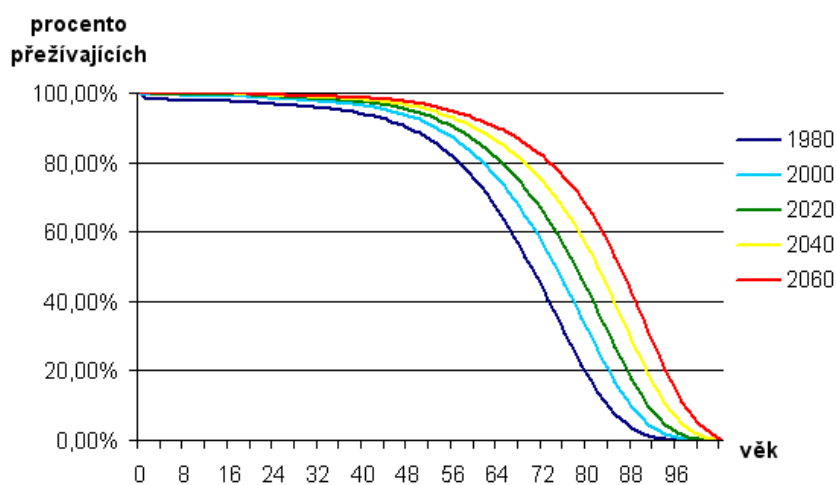
Obrázek 4.6: Predikce parametru  $k_t$  pro mužskou populaci modelem ARIMA(1,1,0).



Obrázek 4.7: Míra zlepšení úmrtnosti předpovídaných hodnot pro mužskou populaci.

věk \ rok	1980	1996	2012	2028	2044	2060	celkové zlepšení v %
0	66,83	70,28	73,77	76,68	79,77	82,85	23,97%
20	48,58	51,23	54,29	57,01	59,99	63,01	29,68%
40	29,88	32,37	35,26	37,86	40,71	43,62	45,98%
60	14,18	16,15	18,50	20,65	23,07	25,57	80,29%
80	4,75	5,63	6,75	7,84	9,15	10,58	123,03%
100	1,10	1,25	1,43	1,59	1,79	1,98	79,71%

Tabulka 4.1: Vybrané hodnoty střední délky života mužské populace.



Obrázek 4.8: Procento přežívajících v mužské populaci ve vybraných rocích.

Rok	Naděje dožití při narození					
	muži			ženy		
	nízká	střední	vysoká	nízká	střední	vysoká
2003	72,0	72,4	72,7	78,5	78,8	79,1
2010	73,4	74,1	74,9	79,7	80,3	81,0
2020	75,4	76,5	77,6	81,3	82,4	83,4
2040	78,4	80,4	81,8	83,6	85,4	86,7
2065	81,3	84,0	85,7	85,8	88,3	89,8

Obrázek 4.9: Očekávaný vývoj celkové úrovně úmrtnosti dle [5] - základní varianty prognózy.

# Kapitola 5

## Závěr

V této práci jsme se zabývali metodami projekce úmrtnosti. Zavedli jsme základní charakteristiky a vztahy mezi nimi a ze získaných dat jsme sami demografické ukazatele odhadovali. Aby byly vypočtené ukazatele použitelné pro předpověď úmrtnosti a pro studii existence kohort efektu, museli jsme je nejdříve vyrovnat. K tomu jsme kvůli konzistenci s hodnotami prezentovanými Českým statistickým úřadem používali stejné metody jako ČSÚ. Rovněž jsme se zabývali kohort efektem a souvisejícími generačními úmrtnostními tabulkami.

Dále jsme zavedli pojem kohort efektu, rozebrali jsme možné příčiny jeho vzniku a popsali zkušenosti s tímto jevem v jiných vyspělých zemích. Po té jsme začali zjišťovat, zda je tento jev možné vysledovat i v české populaci, zvláště pro mužskou a ženskou část populace. Přes krátké období, ze kterého máme data o české populaci k dispozici, jsme usoudili, že kohort efekt, který by mohl nějak ovlivnit predikci vývoje úmrtnosti, pozorovat nelze. Některá data sice mírný kohort efekt naznačují, ten ale není tak významný, aby ho bylo nutné v projekci zohledňovat.

V další části textu jsme přistoupili k popisu vývoje úmrtnosti z hlediska různých vědních oborů. Úmrtnost je totiž ovlivněna mnoha faktory, které jsou zkoumány mnoha vědními obory, ať už z hlediska medicínského, ekonomického či sociologického. Klasifikovali jsme metody projekce úmrtnosti a z jednotlivých tříd jsme vybrali jednu, které jsme se věnovali podrobněji.

Pro data z populace České republiky jsme vybrali Lee-Carterovu metodu projekce úmrtnosti. Jedním z důvodů výběru je i to, že jsme nevysledovali v české populaci kohort efekt a tudíž jsme použili model, kde v předpovědi úmrtnosti dominuje časová složka, tedy vliv doby, nikoliv okamžiku narození.

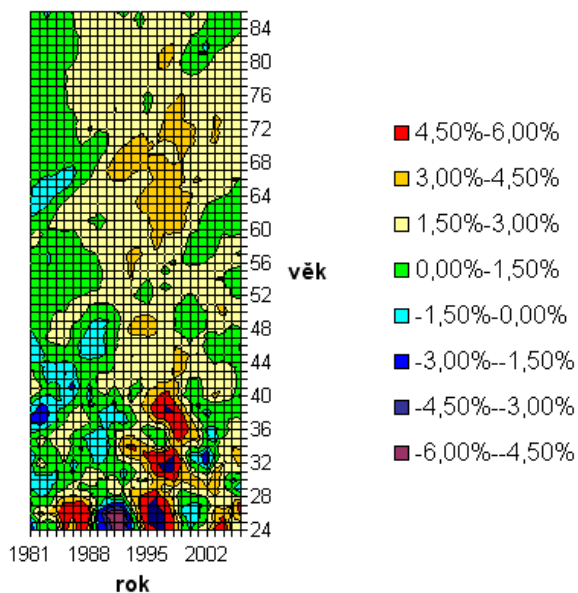
Pomocí softwaru Mathematica 6 a statistického softwaru NCSS 2007 jsme vytvořili předpovídané věkově specifické míry úmrtnosti pro roky 2007–2060. Ovšem i zde

je třeba vzít v úvahu krátké období, ze kterého naše předpovědi vycházejí a specifika tohoto období. Toto omezení může přinášet pro předpovědi na delší časové období jisté zkreslení výsledků předpovědi. Vypočítali jsme i předpovídané střední délky života v tomto období a výsledky jsme porovnali se studií úmrtnosti z populační prognózy B. Burcina a T. Kučery. Výsledky jsme se snažili také co nejlépe ilustrovat pomocí různých grafů.

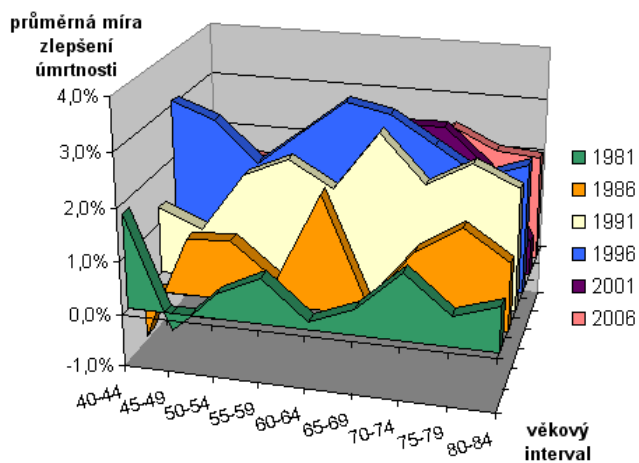
# Přílohy

## Seznam příloh:

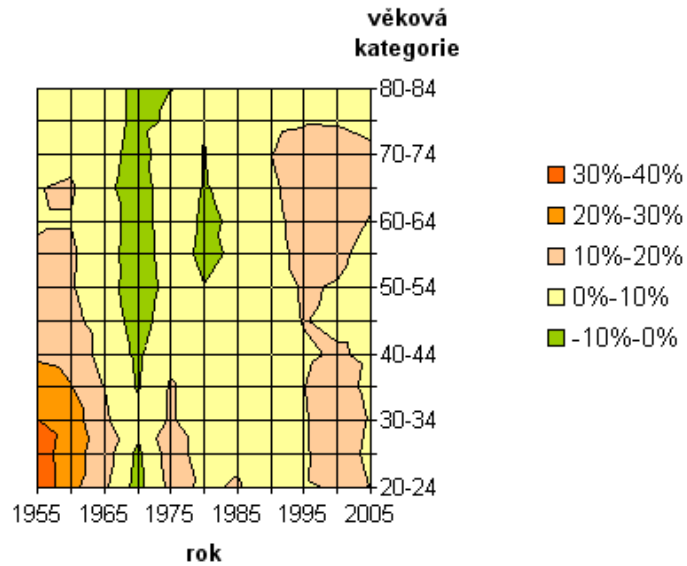
- Příloha 1. : Heat Map měř zlepšení úmrtnosti pro českou ženskou populaci ve věku 24–86 v letech 1981–2006.
- Příloha 2. : Průměrné míry zlepšení úmrtnosti dle věkových skupin a roku pro českou ženskou populaci.
- Příloha 3. : Heat Map modifikovaných měř zlepšení úmrtnosti pro českou ženskou populaci v letech 1955-2005.
- Příloha 4. : Predikce parametru  $k_t$  pro ženskou populaci modelem ARIMA(4,1,0).
- Příloha 5. : Procento přežívajících v ženské populaci ve vybraných rocích.
- Příloha 6. : Předpovídané věkově specifické míry úmrtnosti pro mužskou populaci v letech 2007–2060.
- Příloha 7. : Předpovídané věkově specifické míry úmrtnosti pro ženskou populaci v letech 2007–2060.



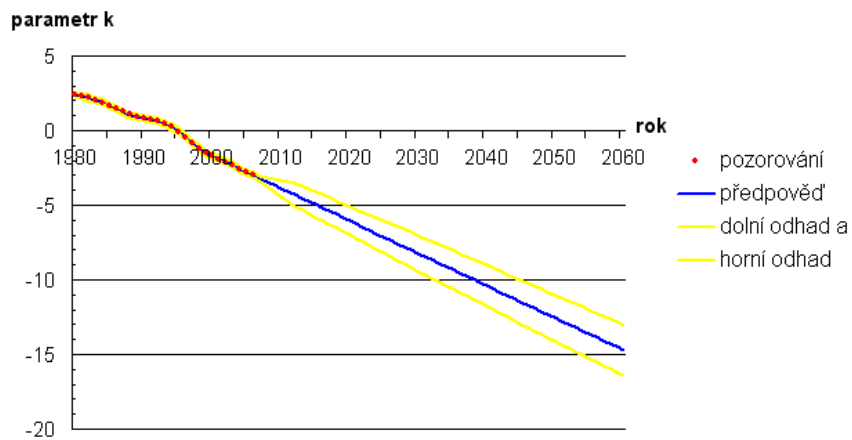
Obrázek 5.1: Příloha 1.: Heat Map měř zlepšení úmrtnosti pro českou ženskou populaci ve věku 24–86 v letech 1981–2006.



Obrázek 5.2: Příloha 2.: Průměrné míry zlepšení úmrtnosti dle věkových skupin a roku pro českou ženskou populaci.

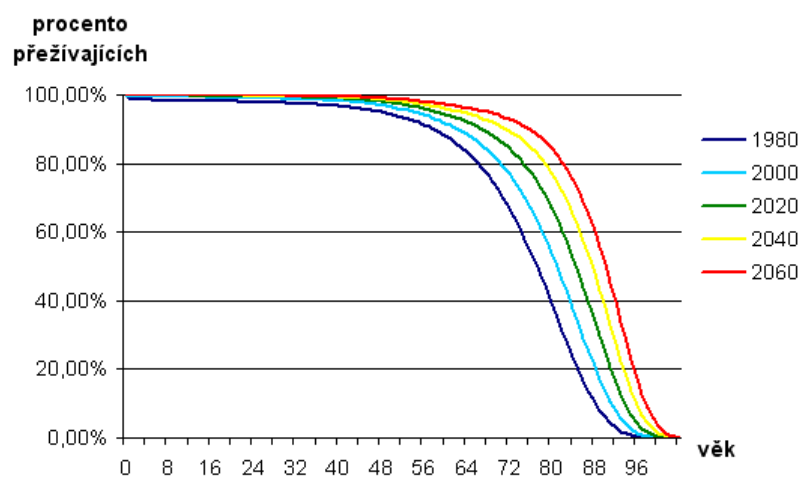


Obrázek 5.3: Příloha 3.: Heat Map modifikovaných měr zlepšení úmrtnosti pro českou ženskou populaci v letech 1955-2005.



Obrázek 5.4: Příloha 4.: Predikce parametru  $k_t$  pro ženskou populaci modelem ARIMA(4,1,0).





Obrázek 5.5: Příloha 5.: Procento přežívajících v ženské populaci ve vybraných rocích.













# Literatura

- [1] Berrar, D.P., Dubitzky, W., Granzow, M.: Singular value decomposition and principal component analysis. *A Practical Approach to Microarray Data Analysis*, 91–109, 2003.
- [2] Booth, H., Maindonald, J., Smith, L.: Applying Lee-Carter under conditions of variable mortality decline. *Population Studies* 56, 325–336, Great Britain, 2002.
- [3] Brass, W.: Mortality models and their uses in demography. *Transactions of the Faculty of Actuaries*, 33, 123–132, 1974.
- [4] Broekhoven van, H.: Market value of liabilities mortality risk: A practical model. *North American Actuarial Journal*, Volume 6 (2), 95–106, 2002.
- [5] Burcin, B., Kučera, T.: Nová kmenová prognóza populačního vývoje České republiky (2003–2065). *Demografie* 46, 100–111, 2004.
- [6] Cipra, T.: *Pojistná matematika - teori a praxe*. EKOPRESS, Praha, 2006.
- [7] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL, Praha, 1986.
- [8] Koschin, F.: *Aktuárská demografie*. Vysoká škola ekonomická, Praha, 2002.
- [9] Koschin, F.: *Vybrané demografické modely.*, Vysoká škola ekonomická, Praha, 1995.
- [10] Lee, R., Carter, L.: Modeling and Forecasting U.S. Mortality. *Journal of American Statistical Association* 419 (87), 659–671, 1992.
- [11] Pitacco, E.: From Halley to "frailty": A review of survival models for actuarial calculations. *Giornale dell'Istituto Italiano degli Attuari*, Terst, 2005.



- [12] Richards, S., Ellam, J., Hubbard, J., Lu, J., Makin, S., Millre, K.: Two-dimensional mortality data: Patterns and projections. Fakulty of Actuaries, Velká Británie, 2007.
- [13] Tabeau, E.: Forecasting mortality in developed countries. Kluwer Academic Publishers, 2001.
- [14] Šacl, F.: Generační úmrtnostní tabulky. Diplomová práce, MFF UK, 2005.
- [15] Willets, R.C.: The Cohort Effect: Insights and Explanations. British Actuarial Journal 10, 833–877, 2004.
- [16] Yang, Y.: The Triumph of Cohort-Effects in the Explanation of Mortality Change: A New Age-Period-Cohort Analysis of Adult Cause-Specific Mortality in the United States. Department of Sociology And Population Research Center&Center on Aging at NORC, The University of Chicago, 2006.
- [17] Continuous Mortality Investigation (CMI), <http://www.actuaries.org.uk>, 2008.
- [18] Poznámky z kurzu Demografie vedeným RNDr. L. Mazurovou, Ph.D., kód NFAP001, MFF UK, Praha, 2007.
- [19] The health of adult Britain 1841-1994. Office of National Statistics, London, 1997.
- [20] Working paper 1: An interim basis for adjusting the '92' series mortality projections for cohort effects. Continuous Mortality Investigation (CMI), London, 2002.
- [21] Working paper 27: The "library" of Mortality Projections. Continuous Mortality Investigation (CMI), London, 2007.
- [22] POPIN Czech Republic - Czech Republic Population Information, <http://popin.natur.cuni.cz/html2/index.php>, 2008.
- [23] Časové řady základních dat demografické statistiky ČR v letech 1900-1999. Český statistický úřad, Praha, 2000.
- [24] Příručka demografické statistiky, obyvatelstvo. Český statistický úřad, <http://www.czso.cz/cz/cisla/0/02/020100/obyvatel.htm>, 2008.
- [25] Věkové složení obyvatelstva České republiky v roce 1993. Český statistický úřad, Praha, 1994.

- [26] Věkové složení obyvatelstva České republiky v roce 1994. Český statistický úřad, Praha, 1995.
- [27] Věkové složení obyvatelstva České republiky v roce 1995. Český statistický úřad, Praha, 1996.
- [28] Věkové složení obyvatelstva České republiky v roce 1996. Český statistický úřad, Praha, 1997.
- [29] Věkové složení obyvatelstva České republiky v roce 1997. Český statistický úřad, Praha, 1998.
- [30] Věkové složení obyvatelstva České republiky v roce 1998. Český statistický úřad, Praha, 1999.
- [31] Věkové složení obyvatelstva České republiky v roce 1999. Český statistický úřad, Praha, 2000.
- [32] Věkové složení obyvatelstva České republiky v roce 2000. Český statistický úřad, Praha, 2001.
- [33] Věkové složení obyvatelstva České republiky v roce 2001. Český statistický úřad, Praha, 2002.
- [34] Věkové složení obyvatelstva České republiky v roce 2002. Český statistický úřad, Praha, 2003.
- [35] Věkové složení obyvatelstva České republiky v roce 2003. Český statistický úřad, Praha, 2004.
- [36] Věkové složení obyvatelstva České republiky v roce 2004. Český statistický úřad, Praha, 2005.
- [37] Věkové složení obyvatelstva České republiky v roce 2005. Český statistický úřad, Praha, 2006.
- [38] Věkové složení obyvatelstva České republiky v roce 2006. Český statistický úřad, Praha, 2007.
- [39] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 1993.

- [40] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 1994.
- [41] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 1995.
- [42] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 1996.
- [43] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 1997.
- [44] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 1998.
- [45] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 1999.
- [46] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 2000.
- [47] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 2001.
- [48] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 2002.
- [49] Zemřelí podle věku, ročníku narození, pohlaví a rodinného stavu. Český statistický úřad, Praha, 2003.
- [50] Zemřelí podle věku, pohlaví a rodinného stavu. Český statistický úřad, Praha, 2004.
- [51] Zemřelí podle věku, pohlaví a rodinného stavu. Český statistický úřad, Praha, 2005.
- [52] Zemřelí podle věku, pohlaví a rodinného stavu. Český statistický úřad, Praha, 2006.
- [53] Obyvatelstvo podle pětiletých věkových skupin v letech 1920 - 2005 (stav k 1. 7.). Český statistický úřad, Praha, 2006.
- [54] Zemřelí podle pohlaví a věku v letech 1930 - 2005. Český statistický úřad, Praha, 2006.