

UNIVERZITA KARLOVA V PRAZE

Přírodovědecká fakulta

Studijní program: Demografie

Studijní obor: Demografie – sociologie



Srovnání vybraných modelů vyrovnávání a extrapolace křivky úmrtnosti

A comparison of selected models for smoothing and
extrapolation of the death curve

Bakalářská práce

Helena Vykoukalová

Vedoucí práce: RNDr. Boris Burcin, Ph.D.

Praha 2014

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně pod vedením školitele RNDr. Borise Burcina, Ph.D. a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Liberci, 14. srpna 2014

.....

Tímto bych ráda poděkovala vedoucímu mé bakalářské práce, RNDr. Borisi Burcinovi, Ph.D., za odborné vedení této práce, konstruktivní připomínky, cenné rady a především za čas věnovaný této práci. Velký dík patří také Mgr. Aleně Černíkové, Ph.D. za přínosné konzultace v oblasti statistiky. Do třetice chci poděkovat svým blízkým za veškerou podporu a pomoc, které se mi v průběhu vytváření této práce dostalo.

Srovnání vybraných modelů vyrovnávání a extrapolace křivky úmrtnosti

Abstrakt

Důsledkem demografických změn ve 20. století rychle rostla naděje dožití a zvyšoval se podíl věkové skupiny 65 a více let. Cílem této práce je představit šest modelů extrapolace a vyrovnávání křivky úmrtnosti a pomocí empirických dat ověřit jejich přesnost.

Teoretická část práce obsahuje přehled použité literatury a informace o empirických datech. V druhé části jsou představeny vybrané modely vyrovnávání a extrapolace a nastíněny možnosti hodnocení modelů mezi sebou.

Analytická část se zaměřuje na porovnání 36 vyspělých zemí ze 4 kontinentů. Země jsou rozřazeny podle podobnosti do několika menších celků a analyzovány ve vybraných obdobích a podle úrovně naděje dožití při narození. Každý celek je analyzován jak odděleně, tak i společně s ostatními a je doporučen model, který je pro danou skupinu nejvhodnější. V závěru praktické části jsou interpretovány výsledky a vybrány nejlépe hodnocené modely. Jako nejvhodnější byly v analýze podle období i v analýze podle úrovně naděje dožití při narození hodnoceny logistické modely, zejména Kannistö. Naopak nejméně vhodnými se jevíly modely Coale-Kisker a Gompertz-Makeham.

Klíčová slova: úmrtnost, extrapolace, vyrovnávání křivky úmrtnosti, informační kritérium, Coale-Kiskerův model, Gompertzův model, Gompertz-Makehamův model, Heligman-Pollardův model, Kannistův model, Thatcherův model

A comparison of selected models for smoothing and extrapolation of the death curve

Abstract

Due to demographic consequences of the 20th century, a life expectancy has rapidly increased and the proportion of the oldest-age group has grown. The aim of this thesis is to present six models used for smoothing and extrapolation of the data and verify their accuracy using empirical data.

The theoretical part of the thesis contains a summary of bibliography and information about used data sets. The second part presents chosen models of smoothing and extrapolating of the data and it suggests the possibilities of comparison among models.

The analytical part is aimed to compare the data from 36 developed countries over 4 continents. Countries are divided by similar characteristics into a few smaller groups and the groups are analyzed by period and by life expectancy at birth. Every country is analyzed separately as well as with the others and the model that is the best for the group is recommended. At the end of the analytical part, results are interpreted and the best models are chosen. The logistic models – especially Kannistö model – were the most suitable in the analysis by period and by life expectancy at birth. The least suitable models were Coale-Kisker and Gompertz-Makeham model.

Keywords: mortality, extrapolating, smoothing, information criterion, Coale-Kisker model, Gompertz model, Gompertz-Makeham model, Heligman-Pollard model, Kannistö model, Thatcher model

Obsah

Seznam zkratk	7
Seznam obrázků	8
Seznam tabulek	9
Úvod	11
1.1 Teoretická východiska práce	12
2 Prameny dat	14
3 Použité metody	20
3.1 Modely vyrovnávání	20
3.1.1 Gompertz (g)	20
3.1.2 Gompertz-Makeham (m)	21
3.1.3 Kannistö (k)	21
3.1.4 Thatcher (t)	22
3.1.5 Heligman-Pollard (h)	23
3.1.6 Coale-Kisker (c)	23
3.2 Hodnocení modelů	25
3.2.1 Akaikeho informační kritérium	26
4 Analýza vybraných dat	28
4.1 Zpracování vstupních dat	28
4.2 Analýza vybraných států podle období	30
4.2.1 1920–1929	32
4.2.2 1950–1959	33
4.2.3 1990–1999	36
4.2.4 2000–2009	41
4.2.5 Shrnutí	44
4.3 Analýza vybraných států podle úrovně naděje dožití při narození	46
4.3.1 Analýza skupin mužů	47
4.3.2 Analýza skupin žen	52
4.3.3 Shrnutí	55

Závěr	60
Seznam literatury	63
Seznam použitých datových zdrojů	65

Seznam zkratek

AIC	Akaikeho informační kritérium
BIC	Bayesovo informační kritérium
CSV	Comma Separated Values
ČSÚ	Český statistický úřad
HMD	The Human Mortality Database
PNG	Portable Network Graphics
PRB	Population Reference Bureau
SZÚ	Státní zdravotní ústav
WPP	World Population Prospects
c	Coale-Kiskerův model
g	Gompertzův model
h	Heligman-Pollardův model
k	Kannistův model
m	Gompertz-Makehamův model
t	Thatcherův model

Seznam obrázků

2.1	Ukázka vstupního souboru softwaru DeRaS	15
2.2	Ukázka nastavení parametrů v aplikaci DeRaS	16
2.3	Ukázka výstupu aplikace DeRaS, časové řady, ČR, ženy.	17
3.1	Ukázka aplikace vybraných modelů, ČR, muži, 1950.	25
3.2	Ukázka aplikace vybraných modelů, ČR, muži, 2011.	26
4.1	Ukázka oscilace Akaikeho informačního kritéria, Švédsko, muži.	29
4.2	Rozdíl mezi maximální a minimální průměrnou hodnotou Akaikeho informač- ního kritéria podle pohlaví.	31
4.3	Porovnání rozptylu mezi muži a ženami ve skupinách.	57

Seznam tabulek

2.1	Období, ve kterém jsou pro jednotlivé státy poskytována data.	19
3.1	Shrnutí základních informací o použitých modelech.	24
4.1	Skupiny států rozdělené podle období, za které jsou poskytována data.	30
4.2	Pořadí analyzovaných modelů v období 1920–1929, muži.	32
4.3	Pořadí analyzovaných modelů v období 1920–1929, ženy.	33
4.4	Pořadí analyzovaných modelů v období 1950–1959, skupina A, muži.	34
4.5	Pořadí analyzovaných modelů v období 1950–1959, skupina B, muži.	34
4.6	Pořadí analyzovaných modelů v období 1950–1959, skupina A, ženy.	35
4.7	Pořadí analyzovaných modelů v období 1950–1959, skupina B, ženy.	36
4.8	Pořadí analyzovaných modelů v období 1990–1999, skupina A, muži.	37
4.9	Pořadí analyzovaných modelů v období 1990–1999, skupina B, muži.	37
4.10	Pořadí analyzovaných modelů v období 1990–1999, skupina C, muži.	38
4.11	Pořadí analyzovaných modelů v období 1990–1999, skupina A, ženy.	39
4.12	Pořadí analyzovaných modelů v období 1990–1999, skupina B, ženy.	39
4.13	Pořadí analyzovaných modelů v období 1990–1999, skupina C, ženy.	40
4.14	Pořadí analyzovaných modelů v období 2000–2009, skupina A, muži.	41
4.15	Pořadí analyzovaných modelů v období 2000–2009, skupina B, muži.	42
4.16	Pořadí analyzovaných modelů v období 2000–2009, skupina C, muži.	43
4.17	Pořadí analyzovaných modelů v období 2000–2009, skupina A, ženy.	43
4.18	Pořadí analyzovaných modelů v období 2000–2009, skupina B, ženy.	44

4.19	Pořadí analyzovaných modelů v období 2000–2009, skupina C, ženy.	45
4.20	Skupiny států podle naděje dožití.	46
4.21	Pořadí analyzovaných modelů, skupina I, muži.	47
4.22	Pořadí analyzovaných modelů po vyřazení Švédska a Izraele, skupina I, muži. .	48
4.23	Pořadí analyzovaných modelů, skupina II, muži.	49
4.24	Pořadí analyzovaných modelů, skupina III, muži.	49
4.25	Pořadí analyzovaných modelů, skupina IVA, muži.	50
4.26	Pořadí analyzovaných modelů, skupina IVB, muži.	51
4.27	Pořadí analyzovaných modelů, skupina IV, muži.	51
4.28	Pořadí analyzovaných modelů, skupina I, ženy.	52
4.29	Pořadí analyzovaných modelů, skupina II, ženy.	53
4.30	Pořadí analyzovaných modelů, skupina III, ženy.	54
4.31	Pořadí analyzovaných modelů, skupina IV, ženy.	55
4.32	Shrnutí pořadí ve všech skupinách, muži.	55
4.33	Shrnutí pořadí ve všech skupinách, ženy.	56

Úvod

Systematičnost úmrtí budila od pradávna velký zájem člověka. Není tedy divu, že se naši předci snažili v tomto nevyhnutelném jevu najít zákonitosti, na základě kterých by bylo možné předpovídat vymírání jednotlivých generací. V 17. století se začala vyvíjet věda zvaná demografie – věda o procesu reprodukce lidské populace. Prvotní impulz k jejímu vzniku dal John Graunt, později ji rozvíjel například Halley, Malthus nebo Quételet. Za 350 let rozvoje demografie se zásadním způsobem měnila metodika, teoretické přístupy a techniky sběru dat. Rozdílná byla také témata. Snad největší změny ve zpracování dat přinesl pokrok v oblasti výpočetní techniky.

Evropa dnes směřuje svou pozornost k nejstarším skupinám obyvatel. Zatímco v roce 1950 bylo procentuální zastoupení obyvatel ve věku 60 a více let ve vyspělých státech 11,5 %, v roce 2012 to bylo dvakrát tolik – 23 % (World Population Prospects, 2012). Právě rostoucí naděje dožití a růst procentuálního zastoupení obyvatel ve věku 65 a více let vyvolává ve vyspělých zemích poptávku státních a vládních institucí po zpřesnění odhadů vývoje úmrtnosti. Poznatky o změně věkové skladby obyvatelstva se promítají do různých sfér lidského života, jako je například hospodářství nebo zdravotnictví (penzijní a zdravotní reformy). Prognózování vývoje počtu a zastoupení nejvyšších věkových skupin naráží na problém spolehlivosti empirických dat, která se s rostoucím věkem snižuje. Vlivem malého počtu zemřelých osob ve vysokém věku dochází ke značné fluktuaci pravděpodobností úmrtí. Tato fluktuace může být do značné míry eliminována použitím některého z demografických modelů vyrovnávání křivky úmrtnosti.

Předkládaná práce si klade za cíl představit šest vybraných modelů vyrovnávání a extrapolace křivky úmrtnosti a – pomocí vybraných statistických metod – modely na základě hodnot Akaikeho informačního kritéria vzájemně porovnat a vyhodnotit. Modely budou porovnávány z hlediska času (ve vybraných dekadách) a také na základě úrovně naděje dožití při narození dosažené v roce 2011. Vybrané státy budou rozděleny do několika skupin. Při analýze podle sledovaného období byly vytvořeny tři skupiny podle doby, od které jsou na webových stránkách The Human Mortality Database k dispozici data. Podle druhého kritéria

– úrovně naděje dožití při narození – byly státy rozděleny do čtyř skupin. V závěru analytické části budou interpretovány výsledky a doporučeny modely, které by bylo vhodné užít pro daná data.

1.1 Teoretická východiska práce

Od samotného vzniku demografie se ruku v ruce rozvíjela i demografická metodologie jakožto nástroj, který byl pro etabloující se vědu zcela nezbytný. Demografická metodologie kombinuje prvky statistiky, matematiky, informačních technologií a mnoha dalších věd. V posledních letech se tento obor dostává do popředí zejména kvůli zlepšující se kvalitě, dostupnosti a možnosti zpracování dat (nové technologie). V současnosti se metodologii věnuje mnoho českých (Burcin, Kučera, Rychtaříková, Šídlo, Tesárková) i zahraničních autorů (Bravo, Coelho, Kannistö, Magalhães, Thatcher, Vaupel). V této práci je řešeno metodologické téma – modelování křivky úmrtnosti.

Značnou pozornost modelům vyrovnávání a extrapolace křivky úmrtnosti věnovali čeští autoři Burcin, Tesárková, Šídlo (2010) v článku *Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci*. Článek krátce představuje vývoj modelování úmrtnosti počínaje Gompertzovým zákonem a popisuje nejsilnější tendence ve vývoji úmrtnosti. Druhá část článku je věnována srovnání 13 modelů vyrovnávání (Gompertz-Makeham, modifikovaný Gompertz-Makeham, Himes-Preston-Condarn, Heligman-Pollard, Thatcher, Kannistö, Beard, kubický model, Coale-Kisker, Denuit-Goderniaux, Weibull, polynomická funkce 2. až 3. stupně), mezi nimi je i 5 modelů, které jsou předmětem analýzy předkládané práce. Autoři zde představují principy konstrukce modelů a jejich silné a slabé stránky. Druhou část uzavírá přehledná tabulka shrnující základní údaje každého modelu (název, rok navržení, typ, počet parametrů a výpočetní vzorec).

Třetí část se věnuje aplikaci modelů na českou populaci. K analýze byla použita data z databáze ČSÚ. Pro odhad parametrů jednotlivých modelů autoři zvolili věkovou skupinu 65 až 95 let. Empirické hodnoty q_x byly porovnávány s teoretickými hodnotami q_x pro každý model. Na základě dosažených výsledků bylo možné rozdělit uvažované modely do tří skupin. Do první skupiny patří modely, jejichž hodnoty pravděpodobnosti úmrtí se v předem známém dosažitelném věku blíží 1. Obecně jsou hodnoty vyšší než 0,99 již v relativně nízkých věcích (průměrně 105 až 115 let). Druhá skupina modelů svou konstrukcí také vede k hodnotám pravděpodobnosti úmrtí blízké jedné, tato hodnota je však dosahována spíše limitně (pravděpodobnost úmrtí není jednotková ani pro věky okolo 120 let). Modely spadající do třetí skupiny vedou k hodnotám pravděpodobnosti úmrtí jen okolo 0,6 i pro věky nad 120 let, což

je dáno předpokladem, že hodnoty měr úmrtnosti nerostou neomezeně, ale limitně se blíží k nějaké konečné hodnotě.

V závěru třetí části na příkladu naděje dožití ve vybraném věku (0, 60, 65, 80, 85, 90, 95, 100) autoři demonstrují, jakých rozdílů v hodnotách naděje dožití je možné dosáhnout volbou modelu. Zaměřují se na konstrukci úmrtnostních tabulek ČSÚ a hodnotí Gompertz-Makehamův model, který ČSÚ k vyrovnávání úmrtnostních tabulek využívá, jako poměrně pesimistický v rámci všech analyzovaných modelů.

Ze zahraniční literatury byl důležitým zdrojem monografie *The Force of Mortality at Ages 80 to 120* (Thatcher, Kannistö, Vaupel, 1998), kterou vydala univerzita v Odense a která je svým zaměřením podobná výše zmíněnému článku (Burcin, Tesárková, Šídlo, 2010). Autoři vybrali 6 modelů (Gompertz, Weibull, Heligman-Pollard, Coale-Kisker, Kannistö a logistický model) a pracovali s daty z 12 vyspělých zemí (Rakousko, Dánsko, Anglie, Finsko, SRN, Francie, Itálie, Japonsko, Nizozemsko, Norsko, Švédsko a Švýcarsko)¹, které dosahovaly v roce 1994 naděje dožití při narození pro obě pohlaví dohromady 75–79 let. Sledována byla populace ve věku 80 a více let, celkem vzorek obsahoval 32 milionů osob v tomto věku. Výzkum zpracovával data ze tří dekád (1960–1970, 1970–1980 a 1980–1990). Předmětem výzkumu byla také generace 1871–1880. Autoři porovnávali teoretické hodnoty q_x jednotlivých modelů s pozorovanými hodnotami q_x . Pro každý z modelů bylo vytvořeno 8 grafů – 4 pro muže a 4 pro ženy. Modely byly následně hodnoceny podle toho, jak výrazně se teoretické hodnoty odlišovaly od pozorovaných. Podle tohoto kritéria byl nejlépe hodnocen logistický model, který velmi dobře popisoval populaci od 80 do 90 až 95 u mužů a do 95 až 100 let u žen. Za ním se v tomto pořadí umístily modely Kannistö, Weibull, Heligman-Pollard a Coale Kisker. Nejhorší byl hodnocen Gompertzův model, který výrazně nadhodnocoval úmrtnost u nejstarších věkových skupin.

Mimo zdroje uvedené v této kapitole byly poznatky čerpány z řady dalších odborných publikací a textů, které jsou uvedeny v připojeném seznamu literatury.

¹Původně bylo zemí 13, Island byl vyřazen kvůli malému počtu sledovaných událostí u osob ve věku 90 a více let

Kapitola 2

Prameny dat

Zdrojová data pro hlavní demografické ukazatele byla čerpána z online databáze *The Human Mortality Database*. Stránky vznikly ve spolupráci Kalifornské univerzity v Berkeley a institutu Maxe Plancka pro demografický výzkum v Rostocku. *The Human Mortality Database* poskytuje data pro 37 vyspělých států světa. Pro každý stát jsou na webových stránkách k dispozici minimálně tři soubory dat – střední stav populace P_x ¹ tříděný podle věku, kalendářního roku a pohlaví, počet zemřelých D_x tříděný opět podle věku, kalendářního roku a pohlaví a počet narozených N_x tříděný podle roku a pohlaví (HMD).

Tato práce analyzuje všech 37 zemí, které jsou na HMD aktuálně k dispozici. Data o počtu zemřelých a narozených byla HMD poskytována v podobě třetích hlavních souborů. Analyzovány byly údaje o středních stavech populací a počtech zemřelých tříděné do jednoletých intervalů podle kalendářního roku a jednotek věku. Pouze poslední interval byl širší než jeden rok, a to 110 a více let. Dále byla tato data poskytována odděleně pro muže a ženy. Další proměnnou vstupující do analýzy je počet narozených v daném kalendářním roce, opět je rozdělený podle pohlaví.

Období, ve kterém jsou data poskytována pro jednotlivé země, jsou různá, jak ukazuje tabulka 2.1. Nejdelší časovou řadu nalézáme u Švédska (1751–2011), nejkratší potom u Chile – 14 let (1992–2005). U naprosté většiny zemí jsou data bez přerušení poskytována pro celé v tabulce 2.1 uvedené období. Jedinou výjimkou je Belgie, pro kterou jsou data k dispozici od roku 1841 do roku 2009 mimo let 1914–1918, kde chybí hodnoty P_x a D_x . Počet narozených v tomto období evidován je. Kvůli tomuto je Belgie analyzována až od roku 1919 místo od roku 1841.

U většiny států se jednalo o **změnu území**, kdy se měnily hranice dané země. Výjimkou byl Nový Zéland a Austrálie, kde se změna váže k objektu sčítání. V případě Nového Zélandu spočívá změna v tom, že od roku 1901 do roku 1948 byli evidováni pouze obyvatelé nemaorského původu občané. Od roku 1948 do roku 2008 je zaznamenávána celková populace (Maoři i obyvatelé nemaorského původu) a samostatně také údaje o Maorech. Pro potřeby

¹Proměnná x značí věk.

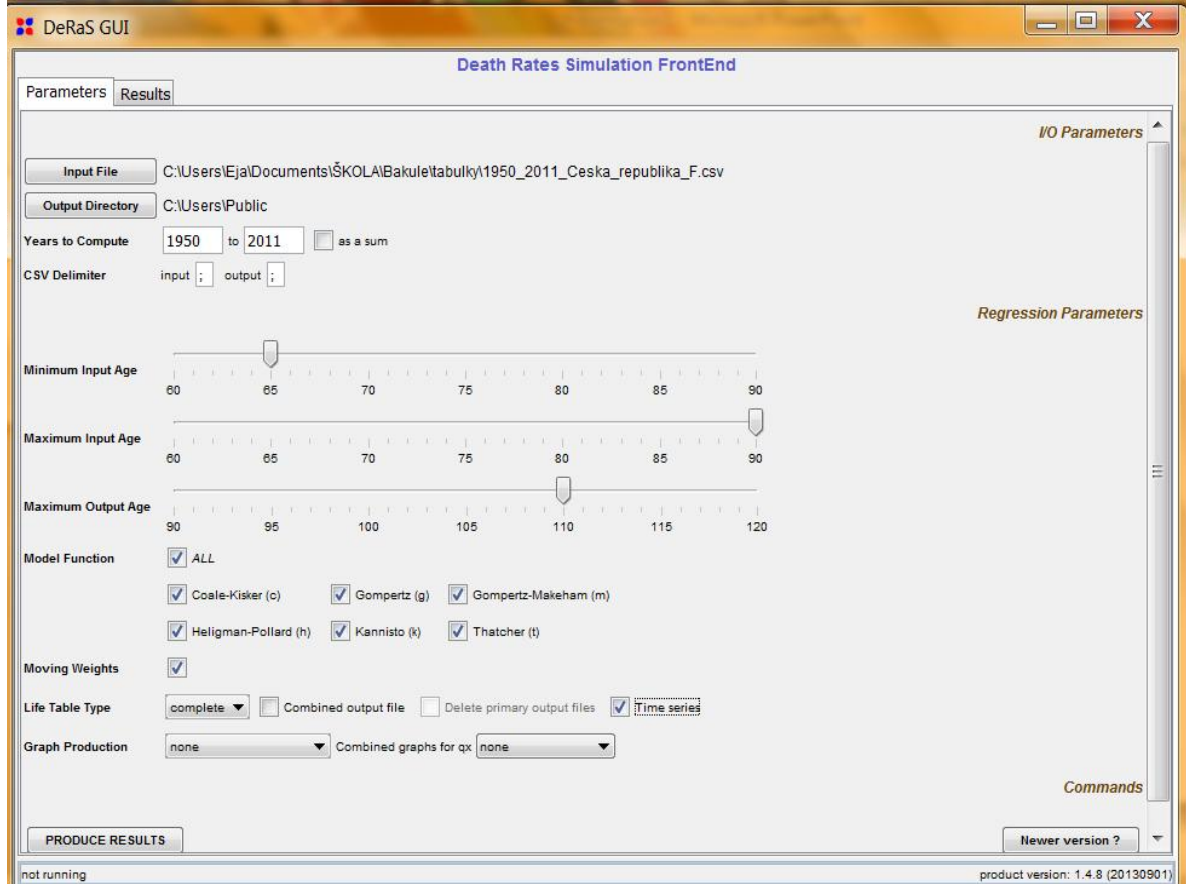
předkládané práce byla zvolena data za obyvatele celkem v období 1948–2008. Analogicky je řešena i legislativní úprava v Austrálii, .

Zdrojová data byla upravena pro software DeRaS, který slouží ke konstrukci úmrtnostních tabulek (viz 2.1). Velkým přínosem softwaru DeRaS pro předkládanou práci je možnost vyrovnávání úmrtnostní křivky a extrapolace pravděpodobností úmrtí až do nejvyšších věků pomocí šesti modelů a také možnost zpracovat velké množství dat během krátké doby. Vstupní soubory ve formátu CVS obsahují informaci o středním stavu populace a počtu zemřelých rozřazených podle pohlaví, jednotek věku a kalendářního roku, a počtu živě narozených (B_v nebo N_v) podle pohlaví a kalendářního roku. Písmeno M nebo F v horní části udává, zda se jedná o soubor mužů nebo žen. Dalším identifikátorem ve vstupním souboru je název státu umístěný rovněž u levého horního okraje.

Obrázek 2.1: Ukázka vstupního souboru softwaru DeRaS. Zdroj: HMD, vlastní zpracování.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Version	1.0												
2	Country	Ceska_republika												
3	Region													
4	Sex	F												
5														
6	Dx	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	:
7	0	5198	4572,1	3386,15	2673,06	2143,14	1903,09	1800	1656	1429	1138	1059	1019	:
8	1	404	414	309,02	290	206,02	175	163	173	155	164	90	96	:
9	:	:	:	:	:	:	:	:	:	:	:	:	:	:
10	:	:	:	:	:	:	:	:	:	:	:	:	:	:
11	110+	0	0	0	0	0	0	0	0	0	0	0	0	:
12														
13	Px	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	:
14	0	86643,99	86761,1	85857,74	83318,65	81124,91	79687,82	78445,02	76057,71	71092,88	65008,34	62120,31	62423,44	6344
15	1	85901,42	84695,91	85235,94	84676,01	82437,4	80484,98	79162,94	77883,47	75669,79	70758,3	64978,52	62229,7	6212
16	:	:	:	:	:	:	:	:	:	:	:	:	:	:
17	:	:	:	:	:	:	:	:	:	:	:	:	:	:
18	110+	0	0	0	0	0	0	0	0	0	0	0	0	:
19														
20	Bv	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	:
21		90851	89792	87096	83598	81787	80294	79139	75217	68960	62510	62418	63661	65
22														
23														

Obrázek 2.2: Ukázka nastavení parametrů v aplikaci DeRaS. Zdroj: DeRaS.



Pravděpodobnost úmrtí počítá software DeRaS nepřímou metodou (Burcin, Tesárková, Kománek, 2014):

$$q_x = 1 - e^{-u_x} \quad (2.1)$$

kde u_x značí míry úmrtnosti podle věku

$$u_x = \frac{D_x}{P_x} \quad (2.2)$$

proměnná D_x označuje absolutní hodnoty počtu zemřelých v dokončeném věku x , proměnná P_x pak střední stav populace v dokončeném věku x . Pravděpodobnost úmrtí pro věk 0 je počítána jako

$$q_0 = \frac{D_0}{B_v} \quad (2.3)$$

kde D_0 značí absolutní počet zemřelých v dokončeném věku 0 a B_v značí počet živě narozených v daném roce.

DeRaS využívá pro vyrovnání úmrtnosti metodu klouzavých průměrů pro nízký a střední věk a úmrtnostní modely pro střední a vysoký věk, u kterých se parametry odhadují pomocí

metody nelineárních nejmenších čtverců. Software nabízí šest modelů vyrovnávání a extrapolace křivky úmrtnosti – Coale-Kisker, Gompertz, Gompertz-Makeham, Heligman-Pollard, Kannistö a Thatcher. Principem konstrukce těchto modelů se zabývá kapitola 3.1.

DeRaS umožňuje nastavení několika parametrů, viz obrázek 2.2, jako je období, za které budou úmrtnostní tabulky počítány, věkový interval, ve kterém budou data vstupovat do analýzy – minimální (60 let) a maximální (90 let), a nastavení posledního věkového intervalu, pro který bude spočtena úmrtnostní tabulka. Dále nabízí volbu mezi šesti výše zmíněnými modely, které vyrovnají a extrapolují míry a kvocienty úmrtnosti ve vyšším věku. Software umožňuje i nastavení parametrů výstupních souborů – mimo jiné výběr mezi zkrácenou a úplnou úmrtnostní tabulkou – a grafů.

Obrázek 2.3: Ukázka výstupu aplikace DeRaS, časové řady, ČR, ženy. Zdroj: DeRaS.

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2	Complete Life Table														
3	Country	Ceska_republika													
4	Region														
5	Sex	F													
6	Year / Period	1950 - 2011													
7	Program Version	1.2.9													
8	Copyright	(c)2010-2012 by B. Burcin, K. Hulikova Tesarkova, D. Komanek													
9	qx unadj.	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963
10	0	0,05722	0,05092	0,03888	0,03198	0,0262	0,0237	0,02275	0,02202	0,02072	0,01821	0,01697	0,01601	0,01824	0,01705
11	1	0,00469	0,00488	0,00362	0,00342	0,0025	0,00217	0,00206	0,00222	0,00205	0,00232	0,00138	0,00154	0,00166	0,00136
12	:														
13	110+	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
14	qx	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963
15	0	0,05722	0,05092	0,03888	0,03198	0,0262	0,0237	0,02275	0,02202	0,02072	0,01821	0,01697	0,01601	0,01824	0,01705
16	1	0,00469	0,00488	0,00362	0,00342	0,0025	0,00217	0,00206	0,00222	0,00205	0,00232	0,00138	0,00154	0,00166	0,00136
17	:	0,58284	0,70897	0,69138	0,79044	0,73307	0,6984	0,7837	0,74433	0,71281	0,79509	0,7141	0,76077	0,81255	0,76942
18	110+	1	1	1	1	1	1	1	1	1	1	1	1	1	1
19	lx	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963
20	0	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000
21	1	94279	94908	96112	96802	97380	97630	97726	97798	97928	98179	98303	98399	98176	98295
22	:	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	110+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	dx	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963
25	0	5721	5092	3888	3198	2620	2370	2274	2202	2072	1821	1697	1601	1824	1705
26	1	442	463	348	331	243	212	201	217	200	227	136	152	163	134
27	:	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	110+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	lx	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963
30	0	95543	95929	96748	97259	97712	97914	97992	98049	98154	98365	98474	98557	98368	98473
31	1	94057	94677	95938	96637	97258	97524	97625	97690	97828	98066	98235	98323	98094	98228
32	:	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	110+	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	ex	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963
35	0	66.88	67.72	69.2	69.86	70.55	71.52	71.96	71.65	72.55	72.64	73.32	73.55	72.88	73.56

Pro účely této práce vstupovala do analýzy věková skupina 65 až 90 let. Konečný interval byl stanoven na 110 a více let a bylo využito všech šesti modelů. Jako výstup z aplikace DeRaS jsme zvolili časové řady všech proměnných v úmrtnostní tabulce za celé zvolené období (viz

obrázek 2.3. Grafy byly vygenerovány jen k účelům prezentace rozdílů mezi jednotlivými modely v kapitole 3.1.

Časová dostupnost	Země	Časová dostupnost	Země
1751–2011	Švédsko	1948–2008	Nový Zéland
1816–2010	Francie	1950–2009	Irsko
1835–2011	Dánsko	1950–2009	Maďarsko
1838–2010	Island	1950–2009	Slovensko
1841–2009	Belgie	1950–2011	Česko
1846–2009	Norsko	1958–2009	Polsko
1850–2009	Nizozemsko	1959–2009	Ukrajina
1872–2009	Itálie	1959–2010	Bělorusko
1876–2011	Švýcarsko	1959–2010	Rusko
1878–2009	Finsko	1959–2011	Estonsko
1908–2009	Španělsko	1959–2011	Litva
1921–2009	Austrálie	1959–2011	Lotyšsko
1921–2009	Kanada	1960–2009	Lucembursko
1922–2011	Anglie a Wales	1970–2010	Taiwan
1933–2010	USA	1983–2009	Izrael
1940–2009	Portugalsko	1983–2009	Slovinsko
1947–2009	Japonsko	1990–2011	Německo
1947–2010	Bulharsko	1992–2005	Chile
1947–2010	Rakousko		

Tabulka 2.1: Období, ve kterém jsou pro jednotlivé státy poskytována data. Zdroj: HMD

Kapitola 3

Použité metody

3.1 Modely vyrovnávání

Modely vyrovnávání a extrapolace dat jsou dnes běžně užívaný nástroj, který eliminuje zkreslení dané malým počtem pozorovaných událostí. Dnes jsou tyto modely používány řadou českých i zahraničních institucí a organizací, jako je například ČSÚ, který používá při tvorbě úmrtnostních tabulek Gompertz-Makehamův model v kombinaci s King-Hardyho metodou (ČSÚ, 2014), HMD vyrovnává data pomocí Kannistova modelu.

K vyhlazení křivky úmrtnosti lze použít celou řadu modelů, jako jsou například Weibullův mocninný model, kubický nebo různě modifikované logistické modely. Výběr optimálního modelu závisí na volbě analyzovaného státu. Ve této práci je analyzováno šest modelů vyrovnávání a extrapolace křivky úmrtnosti, viz tabulka 3.1, které jsou následně hodnoceny pomocí informačních kritérií.

3.1.1 Gompertz (g)

Již v roce 1825 představil Benjamin Gompertz svůj model vyrovnávání intenzit úmrtnosti, který předpokládal, že – mimo nejstarší a nejmladší věky – roste intenzita úmrtnosti **exponenciálně**. Relativní přírůstek je tedy v každém věku konstantní. Model je popsán rovnicí (Burcin, Tesárková, Šídlo, 2010, s. 78):

$$\mu_x = ae^{bx} \quad (3.1)$$

kde a , b jsou parametry modelu, x je věk a μ_x značí intenzitu úmrtnosti ve věku x .

Gompertz definoval svůj zákon na základě studia úmrtnostních tabulek a zvažoval důsledky aplikace zákona na stále starší skupinu obyvatel. Ideu exponenciální funkce vysvětluje tak, že s rostoucím věkem **narůstá** počet poškozených buněk v lidském těle, a proto se ve vyšších věcích rapidně snižuje pravděpodobnost přežití.

V monografii *The Force of Mortality at Ages 80 to 120* se pro muže model jevil jako poměrně spolehlivý až do věku 95 let. U žen spolehlivost tohoto modelu klesala už mnohem dříve – v 80 letech. Nad touto hranicí model nadhodnocuje (Thatcher, Kannisto, Vaupel, 1998).

3.1.2 Gompertz-Makeham (m)

Další velmi dobře známý model je Gompertz-Makehamův, který ČSÚ používá pro vyrovnávání intenzit úmrtnosti od věku 80 let (ČSÚ, 2014). William Makeham analyzoval tabulky úmrtnosti od 20 let až do 80 let. Na základě této analýzy doplnil Gompertzův zákon o tzn. **Makehamovu korekci** – parametr, který zohledňuje úmrtí **nezáviselí** na věku (nehody, úrazy, sebevraždy) (Rabušic, 1995, s. 34). Tento model lze vyjádřit jako (Thatcher, Kannistö, Vaupel, 1998):

$$\mu_x = c + ae^{bx} \quad (3.2)$$

kde a a b jsou parametry modelu a c je konstanta zahrnující úmrtí nezávislá na věku.

Stejně jako Gompertzův i Gompertz-Makehamův model popisuje růst úmrtnosti jako exponenciální. To je důvod, proč jsou při aplikaci na data 13 zemí v článku *The Force of Mortality at Ages 80 to 120* pozorované hodnoty v nejvyšších věcích nižší než očekávané (Thatcher, Kannistö, Vaupel, 1998).

3.1.3 Kannistö (k)

V roce 1992 představil Väinö Kannistö svůj **logistický** model (Thatcher, Kannistö, Vaupel, 1998), který k modelování využívá logitové transformace¹ měr úmrtnosti. Kannistö nevytvořil žádný obecný zákon, ale sestavil výpočet na základě pozorovaných hodnot. Model může být popsán rovnicí (Thatcher, Kannistö, Vaupel, 1998):

$$\mu_x = c + \frac{ae^{bx}}{1 - ae^{bx}} \quad (3.3)$$

kde a , b a c jsou parametry modelu,

¹transformace se používá u normálního rozdělení a umožňuje zhustit hodnoty na méně obsazeném konci škály a zároveň rozlišit hodnoty nahustěné na opačném konci škály. Je popsána vzorcem (Jeřábek, 1993):

$$y = \ln \frac{x}{1 - x}$$

nebo také (Coelho, Magalhães, Bravo, 2007, s.11)

$$\mu_x = \frac{e^{[\theta_0 + \theta_1 * (x-80)]}}{1 - e^{[\theta_0 + \theta_1 * (x-80)]}} \quad (3.4)$$

pro $x \geq 80$, kde θ_0 a θ_1 jsou nezáporné parametry modelu.

Kannistův model je obecně považován za jeden z **nejvhodnějších** modelů pro vyspělé země. Z empirických dat vyplynulo, že pro muže relativně přesně popisoval hodnoty zhruba do věku 90 let, u žen se tento věk pohybal okolo 95 let. Po dosažení hraničního věku 90 let v případě mužů začíná hodnoty mírně nadhodnocovat (Thatcher, Kannistö, Vaupel, 1998). Tento model je mimo jiné využíván projektem The Human Mortality Database k vyrovnávání a extrapolaci úmrtnostních tabulek.

3.1.4 Thatcher (t)

Další **logistický** model popsal v roce 1999 Thatcher. Thatcherův model bývá uváděn ve dvou variantách, jako tří a čtyř parametrický. Čtyř parametrický model je formálně vyjádřen jako (Thatcher, 1999 s. 8):

$$\mu_x = \frac{xz}{1+z} + \gamma \quad (3.5)$$

kde

$$z = \alpha \exp \beta x$$

$$= \exp[\beta(x - \phi)]$$

kde

$$\phi = \frac{-\ln \alpha}{\beta}$$

Avšak po aplikaci vzorce 3.5 na data 13 vyspělých zemí se ukázalo, že x se vždy **blíží k jedné**, tedy je pro výpočet nevýznamné (Thatcher, 1998). Proto začali autoři používat tří parametrickou verzi, která může být vyjádřena takto:

$$\mu_x = \frac{xz}{1+z} + \gamma \quad (3.6)$$

α , β a γ jsou parametry modelu. Thatcherův model dobře vystihuje data až do věku 100 let (Thatcher, 1999, s. 20–26).

3.1.5 Heligman-Pollard (h)

V roce 1980 definovala dvojice autorů Heligman a Pollard nový zákon úmrtnosti, který definoval 8 parametrů a 3 fáze vývoje úmrtnosti závislé na věku. První fáze popisuje pokles úmrtnosti v dětském věku. Prostřední modeluje náhlý nárůst úmrtnosti kolem věku 23 let. Poslední fáze odráží vývoj úmrtnosti v dospělém věku (Carriere, 1992, s. 77–78). Oproti ostatním modelům je Heligman-Pollard vhodný pro celé věkové rozpětí.

Pro potřeby předkládané práce byla použita zjednodušená **dvouparametrická** verze, která se zaměřuje na věky od 50 let a u které první dvě podmínky mohou být vynechány. Pomocí logitové transformace ji lze definovat jako (Thatcher, Kannistö, Vaupel, 1998):

$$\text{logit}(q_x) = \ln(a) + bx \quad (3.7)$$

kde a a b jsou parametry modelu a q_x je pravděpodobnost úmrtí. Nabývá hodnot mezi 0 a 1, q_x může být dále definováno jako:

$$q_x = \frac{ae^{bx}}{1 + ae^{bx}} \quad (3.8)$$

Heligman-Pollardův zákon úmrtnosti spolehlivě modeluje data do 95 let u mužů a 92 let u žen, poté mírně pravděpodobnost úmrtí nadhodnocuje (Thatcher, Kannistö, Vaupel, 1998).

3.1.6 Coale-Kisker (c)

V roce 1990 byl autory Coalem a Kiskerem představen **lineární model**, který se zakládá na trochu jiném principu než předchozí modely exponenciální a logaritmické. Autoři definovali model, který porovnává změny úmrtnosti ve věku x a $x + 1$. Předpokládali, že míry úmrtnosti nejsou v nejvyšších věcích konstantní, ale **lineárně klesají**. Na základě rozdílů měr úmrtnosti mezi dvěma po sobě jdoucími věky vymezili autoři proměnnou k_x , která může být definována jako: „logaritmus podílu dvou měr úmrtnosti pro dva po sobě jdoucí věky“ (Burcin, Tesárková, Šídlo, 2010, s. 82). Model je popsán vzorcem (Coelho, Magalhães, Bravo, 2007, s. 8–9):

$$k_x = \ln\left(\frac{m_x}{m_{x-1}}\right) \quad (3.9)$$

kde k_x značí růst měr úmrtnosti m_x od věku x do věku x_{max} . Model je lineární až po dosažení určitého věku (například 80 let), k_x potom může být vyjádřeno jako:

$$k_x = k_{80} + (x - 80) * s; x \geq 80 \quad (3.10)$$

kde k_{80} a s jsou parametry modelu.

K tomuto modelu se vážou dva předpoklady. První, že míry ve věku $x = 80$ musí být poměrně **spolehlivé**, k_x lze odhadnout z empirických hodnot. Druhá podmínka se týká **nejvyššího věku**, který autoři stanovili na 110 let. Hodnoty míry úmrtnosti v tomto věku určili na 1,0 pro muže a 0,8 pro ženy, stanovení těchto hodnot umožní odhad parametru s . Tyto hodnoty byly striktně stanoveny, aby nedocházelo ke křížení hodnot u pohlaví v nejvyšším věku (Coelho, Magalhães, Bravo, 2007, s. 8–9).

Model pak odpovídá exponenciálně-kvadratické funkci, která může být vyjádřena jako (Burcin, Tesárková, Šídlo, 2010, s. 82):

$$m_x = e^{a*x^2+b*x+c} \quad (3.11)$$

kde a , b a c jsou parametry modelu.

Empirická data považovala model za vhodný pro data v intervalu 85–95 let u mužů i u žen. Mimo tento interval jsou hodnoty q_x nadhodnoceny i podhodnoceny. Model je závislý na kvalitě vstupních dat (Thatcher, Kannistö, Vaupel, 1998).

Název	Princip výpočtu	Vzorec	Počet parametrů
Coale-Kisker	lineární	$m_x = e^{a*x^2+b*x+c}$	3
Gompertz	exponenciální	$\mu_x = ae^{bx}$	2
Gompertz-Makeham	exponenciální	$\mu_x = c + ae^{bx}$	3
Heligman-Pollard	logistický	$\logit(q_x) = \ln(a) + bx$	2
Kannistö	logistický	$\mu_x = \frac{e^{[\theta_0 + \theta_1 * (x-80)]}}{1 - e^{[\theta_0 + \theta_1 * (x-80)]}}$	2
Thatcher	logistický	$\mu_x = \frac{xz}{1+z} + \gamma$	3

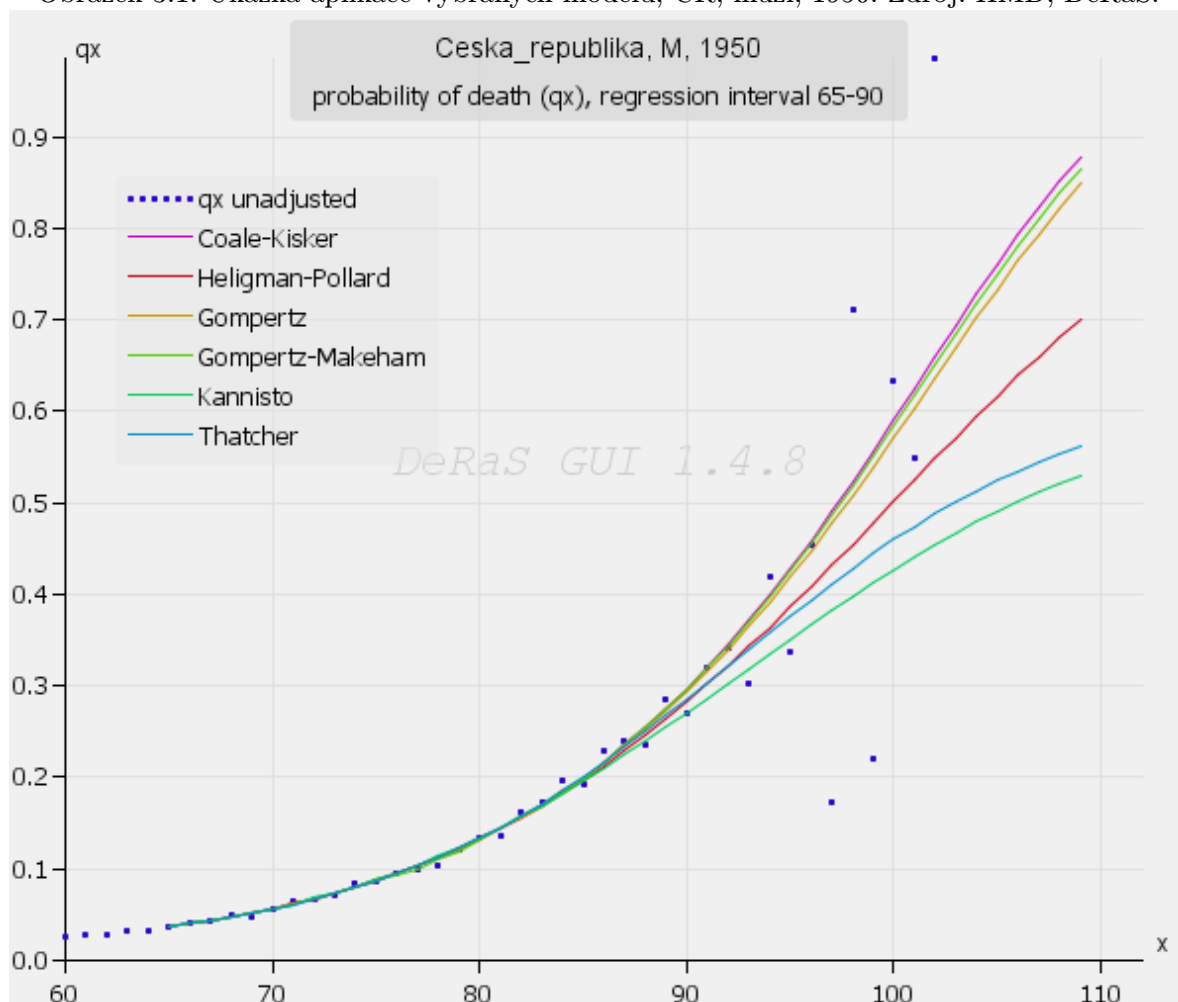
Tabulka 3.1: Shrnutí základních informací o použitých modelech.

V tabulce 3.1 jsou shrnuty všechny vybrané modely. Tři modely jsou logistické, dva exponenciální a jeden lineární. Modely se liší nejen principem výpočtu, ale i počtem parametrů – Coale-Kisker, Gompertz-Makeham a Thatcher jsou tříparametrické, ostatní počítají jen se dvěma parametry.

Obrázky 3.1 a 3.2 ilustrují rozdíly mezi jednotlivými modely v roce 1950 a 2011. V tomto případě byly modely aplikovány na data České republiky. V obou případech není mezi jednotlivými modely až zhruba do věku 85 let žádný velký rozdíl. Ve většině vyspělých zemí je právě věk 85 let brán jako kritická hranice spolehlivosti empirických dat (Burcin, Tesárková, Šídlo, 2010, s. 77). Nad touto hranicí se vyrovnané a extrapolované hodnoty pravděpodobnosti úmrtí

začínají výrazněji odlišovat. V obou letech vykazuje minimální hodnoty pravděpodobnosti úmrtí Kannistův model, nejvyšší hodnoty předpokládá Coale-Kiskerův model, jak vyplývá z grafu 3.1 a 3.2, ale také z článku *Nejpoužívanější metody extrapolace a vyrovnávání křivky úmrtnosti a jejich aplikace na českou populaci* (Burcin, Tesárková, Šídlo, 2010, s. 84–87), kterému je věnována část podkapitoly 1.1.

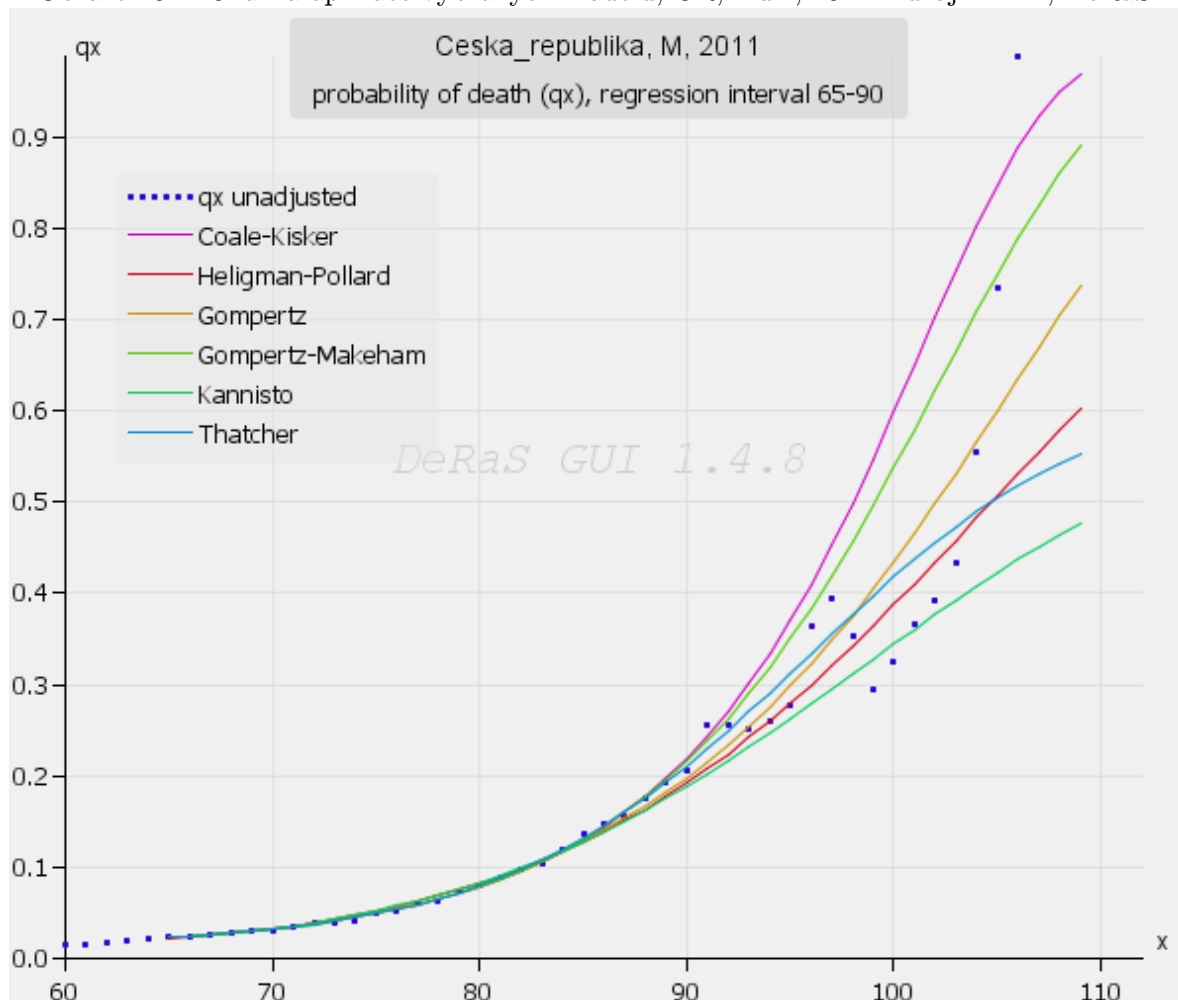
Obrázek 3.1: Ukázka aplikace vybraných modelů, ČR, muži, 1950. Zdroj: HMD, DeRaS.



3.2 Hodnocení modelů

Existuje celá řada postupů, kterými lze hodnotit přesnost modelů. V článku *The Force of Mortality at Ages 80 to 120* (Thatcher, Kannistö, Vaupel, 1998) hodnotí autoři všech šest modelů na základě podílů teoretických a pozorovaných hodnot. Jednoduše lze říct, že čím menší rozdíl, tím model získává lepší hodnocení (Thatcher, Kannistö, Vaupel, 1998). Jinou

Obrázek 3.2: Ukázka aplikace vybraných modelů, ČR, muži, 2011. Zdroj: HMD, DeRaS.



metodou porovnávání bývá **reziduální součet čtverců**, který je druhou mocninou rozdílu pozorovaných a teoretických hodnot. Z reziduálního součtu čtverců vychází i hodnocení, které je použito v této práci.

Dalším možným hodnocením pro výběr **optimálního modelu** jsou informační kritéria. Oproti metodě reziduálních čtverců informační kritéria zohledňují počet parametrů modelu. Existuje jich celá řada – Akaikeho, Bayesovo, Hannan-Quinovo informační kritérium. Všechna tato kritéria se používají k výpočtu **logaritmu věrohodnosti** (viz vzorec 3.12), který odečítají od násobku počtu parametrů.

3.2.1 Akaikeho informační kritérium

Pro účely předkládané práce bylo vybráno Akaikeho kritérium (dále jen **AIC**), které Hirotugu Akaike představil v roce 1973. Toto kritérium je z výše uvedených nejstarší (Bozdogan, 2000,

s. 62–63). Jeho základem je **logaritmus věrohodnosti** $\ln(L)$, který může být definován jako (Bozdogan, 2000, s. 62–63):

$$\ln(L) = \ln\left(\Pi\left(\frac{1}{2\pi * sd(Res)}\right)\right) - \frac{\sum_{i=1}^n \frac{Y_i - X_i}{s_i}}{1} \quad (3.12)$$

kde s_i je odhad směrodatné odchylky pro i -té pozorování, $sd(Res)$ značí odhad směrodatné odchylky reziduí, Y_i jsou pozorované hodnoty proměnné, X_i očekávané hodnoty a n je počet pozorování. Logaritmus věrohodnosti $\ln(L)$ udává, jaká část variability je vysvětlená modelem. Obecně platí, že čím vyšších hodnot logaritmus věrohodnosti dosahuje, tím je model spolehlivější.

U dat v předkládané práci lze očekávat konstantní variabilitu, proto byl použit modifikovaný vzorec, kdy s_i je nahrazeno odhadem směrodatné odchylky reziduí, který může být vyjádřen jako:

$$sd(Res) = \sqrt{\frac{\sum_{i=1}^n (Y_i - X_i)^2}{n - 1}} \quad (3.13)$$

Samotné Akaikeho informační kritérium je dáno vzorcem (Bozdogan, 2000, s. 62–63):

$$AIC = 2 * k - 2 * \ln(L) \quad (3.14)$$

kde k je počet parametrů.

Kapitola 4

Analýza vybraných dat

Cílem předkládané práce je najít model, který bude nejlépe vyrovnávat křivku úmrtnosti dané země nebo skupiny zemí. Země byly hodnoceny na základě dvou kritérií. Prvním z nich bylo sledované období, druhým úroveň intenzity úmrtnosti, kterou v tomto případě určovala naděje dožití při narození v roce 2011.

4.1 Zpracování vstupních dat

Samotné zpracování dat bylo poměrně časově náročnou etapou bakalářské práce. Data poskytovaná HMD jsou dostupná v textovém formátu (.txt), proto bylo nutné je dále upravovat. Data jsme nejprve vložili do tabulkového procesoru, kde jsme vytvořili vstupní soubory pro software DeRaS, které obsahovaly data o **zemřelých, středním stavu populace a narozených**, jak ukazuje obrázek 2.1.

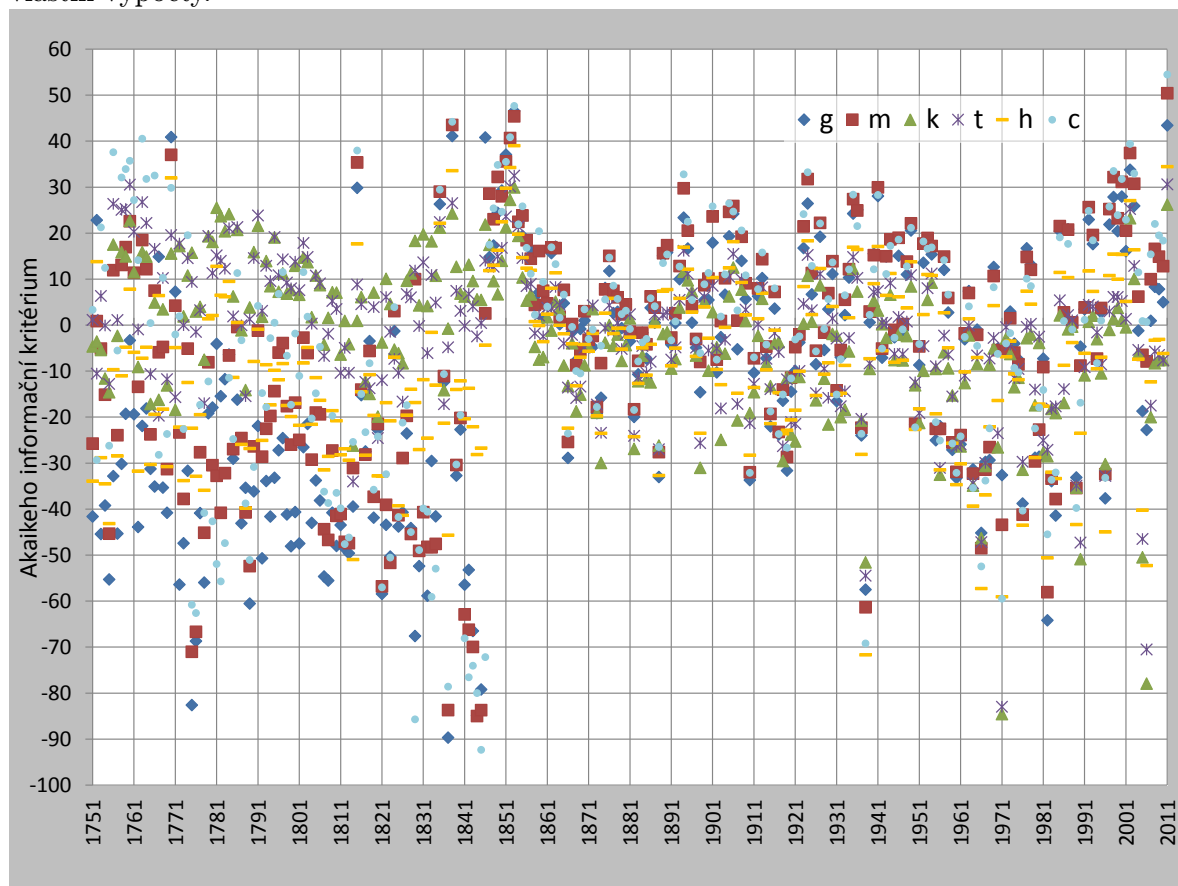
Celkem tak vzniklo 74 vstupních souborů, které byly převedeny do formátu CSV a následně zpracovány softwarem DeRaS. Díky funkci „Time series“¹ (viz obrázek 2.2) jsme získali proměnné q_x pozorované, které vycházelo z naměřených hodnot, a q_x odhadované, které bylo vypočítáno na základě odhadů modelů vyrovnávání, l_x , d_x , L_x a e_x pro každý model za každý rok podle věku.

Konečný otevřený interval úmrtnostní tabulky byl stanoven na **110 a více let**. K dalším výpočtům byly použity pouze pravděpodobnosti úmrtí q_x , jak pozorované tak odhadované, ze kterých byly dále vypočítány hodnoty Akaikeho informačního kritéria. Alternativou hodnocení by mohlo být i vypočtení Akaikeho informačního kritéria z ukazatele naděje dožití e_x . Z důvodu většího množství výpočtů, kvůli kterému se zvyšuje riziko chyby dané zaokrouhlováním a lidským faktorem, byly pro analýzu vybrány hodnoty pravděpodobnosti úmrtí q_x .

Z výstupů DeRaSu bylo vypočítáno Akaikeho informační kritérium pomocí výše uvedeného vzorce 3.14. Vypočteny byly hodnoty jen pro věkovou skupinu **65 až 109 let**. Pod spodní

¹Pomocí funkce byla pro každou proměnnou vytvořena časová řada, viz obrázek 2.3.

Obrázek 4.1: Ukázka oscilace Akaikeho informačního kritéria, Švédsko, muži. Zdroj: HMD, vlastní výpočty.



hranicí tohoto intervalu se odhadované pravděpodobnosti úmrtí všech vybraných modelů lišily zanedbatelně. Horní hranice 109 let byla stanovena tak, aby se teoretické pravděpodobnosti úmrtí jednotlivých modelů dostatečně různily. Pravděpodobnost úmrtí v konečném intervalu 110 a více let se vždy rovná 1. Na základě součtu rozdílů mezi pozorovanými a očekávanými hodnotami pravděpodobnosti úmrtí a další jeho úpravou byl vypočítán logaritmus věrohodnosti, který následně zohlednil různý počet parametrů (viz vzorec 3.14). Počty parametrů jsou uvedeny výše v tabulce 3.1. Tím byla získána hodnota Akaikeho informačního kritéria pro každou zemi za každý rok, kdy byla k dispozici data.

Získané hodnoty Akaikeho informačního kritéria mohou být charakterizovány následovně:

- Mezi hodnotami Akaikeho informačního kritéria a časem není souvislost, jak ukazuje obrázek 4.1. Na příkladu Švédska je vidět, že hodnoty v celém období **oscilovaly**. Nejméně oscilují uprostřed sledovaného období, asi od roku 1860 do roku 1931. Velmi podobné je to i v případě dalších zemí.

- Čím nižší hodnoty Akaikeho informační kritérium dosahuje, tím je model pro danou zemi vhodnější. Hodnoty mohou být i minusové.

Hodnoty Akaikeho informačního kritéria se velmi různí podle daného státu. Zatímco se průměrná hodnota za celé období u všech modelů u Španělska pohybovala mezi 4,4–19,8, u Izraele se tyto hodnoty pohybovaly v intervalu od –33,2 – –21,6. Aby mohly být modely porovnávány mezi sebou v rámci skupin, rozhodli jsme se jim přiřadit **známky**. Hodnota 1 byla přiřazena modelu, který byl pro danou zemi nejvhodnější (hodnota Akaikeho informačního kritéria byla nejnižší), hodnota 6 připadla modelu, který se jevil jako nejméně vhodný. Z těchto hodnot za jednotlivé země byla v rámci skupin spočítána **průměrná známka**.

4.2 Analýza vybraných států podle období

Vývoj úmrtnosti hraje ve výběru modelu značnou roli. V průběhu celého 20. století se charakter úmrtnosti značně měnil. Zatímco na v první polovině 20. století se rostla naděje dožití ve vyspělých zemích vlivem poklesu kojenecké úmrtnosti, v druhé polovině největší příspěvky k naději dožití vykazovaly osoby ve středním a vyšším věku (SZÚ, 2006, s. 37–42). Z důvodu značných změn v charakteru úmrtnosti lze předpokládat, že i výběr **nejvhodnějšího modelu**, který vyrovnává empirické hodnoty a extrapoluje je do vyššího věku, se bude na začátku a na konci sledovaného období lišit.

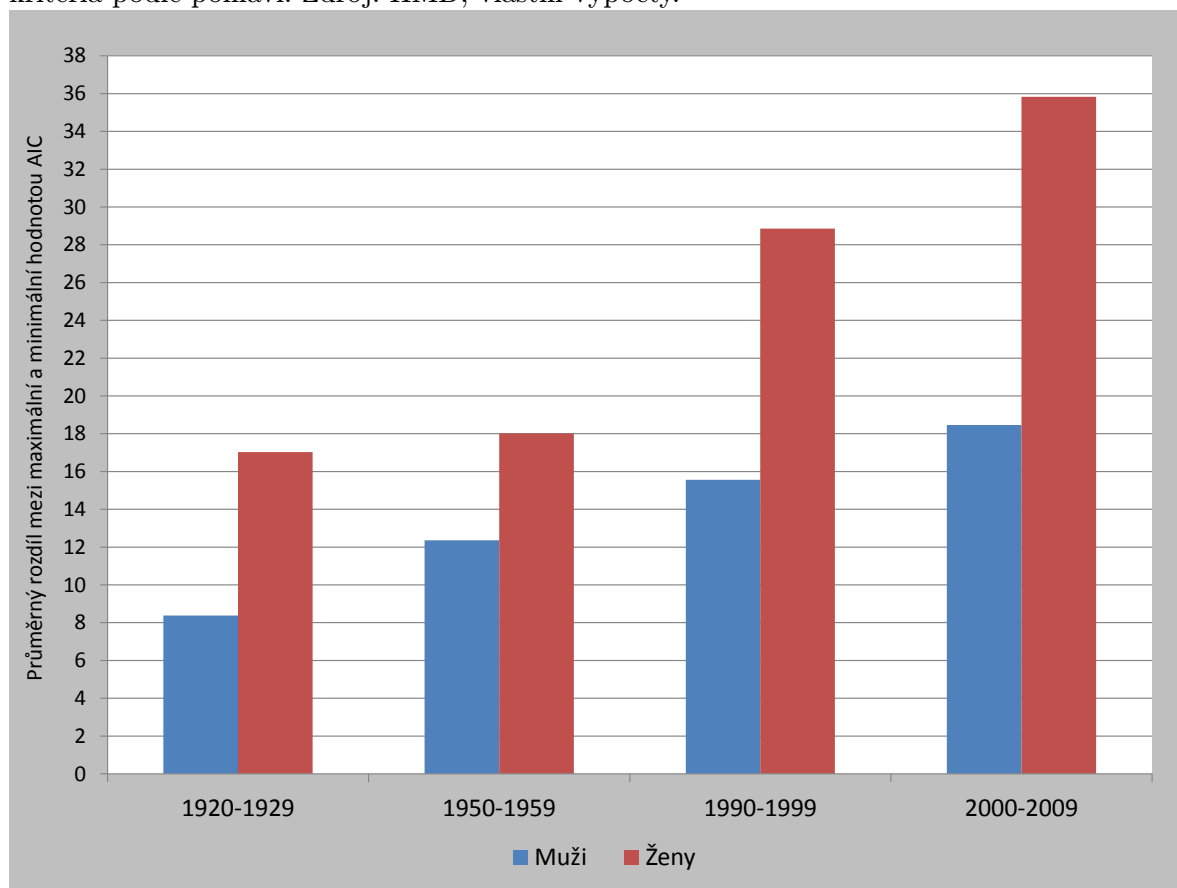
Skupina	Stát
A	Belgie, Dánsko, Finsko, Francie, Island, Itálie, Nizozemsko, Norsko, Španělsko, Švédsko, Švýcarsko
B	Anglie a Wales, Austrálie, Bulharsko, Česko, Irsko, Japonsko, Kanada, Maďarsko, Nový Zéland, Portugalsko, Rakousko, Slovensko, USA
C	Bělorusko, Estonsko, Izrael, Litva, Lotyšsko, Lucembursko, Německo, Polsko, Rusko, Slovinsko, Taiwan, Ukrajina

Tabulka 4.1: Skupiny států rozdělené podle období, za které jsou poskytována data. Zdroj: HMD.

Pro účely analýzy jsme zvolili **čtyři dekády**: 1920–1929, 1950–1959, 1990–1999 a 2000–2009. První dvě období (1920–1929, 1950–1959) byla vybrána jako poválečná, která následovala po značných geopolitických změnách. Poslední dvě dekády (1990–1999, 2000–2009) jsme zvolili pro jejich aktuálnost. Z analýzy bylo **vyřazeno Chile**, za které jsou data publikována pouze od roku 1992 do roku 2005, a tím pádem úplně nepokrývá žádné ze zvolených období.

4.2.

Obrázek 4.2: Rozdíl mezi maximální a minimální průměrnou hodnotou Akaikeho informačního kritéria podle pohlaví. Zdroj: HMD, vlastní výpočty.



V rámci dekád byly utvořeny skupiny (viz tabulka 4.1), a to podle období, za které jsou data poskytována HMD. Do skupiny A patří země, za které jsou na HMD k dispozici data nejpozději od roku 1920. Skupina B je složena ze států, za které jsou poskytnuta data od roku 1921–1950. Třetí skupinu tvoří státy, jejichž data jsou k na HMD k dispozici v období 1951–1990. V prvním období (1920–1929) je jen jedna skupina pro muže a jedna pro ženy, do které spadá 11 států. V období od roku 1950 do roku 1959 byly pro každé pohlaví vytvořeny dvě skupiny. V dalších dvou obdobích (1990–1999 a 2000–2009) byly státy rozděleny na skupiny tři.

V čase se zvyšuje průměrný **rozdíl** mezi hodnotou nejméně vhodného (maximální) a nejvhodnějšího (minimální) modelu. Čím starší je sledované období, tím jsou hodnoty modelů méně diferencovány a naopak, jak ukazuje obrázek

4.2.1 1920–1929

V průběhu 20. let rostla díky se ve většině vyspělých států snižovala intenzita úmrtnosti. HMD poskytovala za toto období data pro 11 států.

V případě mužů (viz tabulka 4.2) získal nejmenší průměrnou známku (2,0) **Heligman-Pollardův** model, který byl nejvhodnějším pro čtyři státy. Jako poměrně dobrý se jevil i Kannistův model. Z jedenácti států získal v pěti známku 1, průměrná známka byla 2,5. Páté a šesté místo bylo přiřazeno Gompertz-Makehamovu (4,3) a Coale-Kiskerovu (4,5) modelu.

Země/model	g	m	k	t	h	c
Belgie	2	4	5	6	1	3
Dánsko	1	3	5	6	2	4
Finsko	4	5	1	3	2	6
Francie	5	6	3	4	1	2
Island	3	5	2	4	1	6
Itálie	6	5	1	2	3	4
Nizozemsko	6	5	1	3	2	4
Norsko	1	3	5	6	2	4
Španělsko	6	2	1	3	4	5
Švédsko	4	5	1	2	3	6
Švýcarsko	3	4	2	6	1	5
Průměrná známka	3,7	4,3	2,5	4,1	2,0	4,5
Pořadí	3	5	2	4	1	6

Tabulka 4.2: Pořadí analyzovaných modelů v období 1920–1929, muži. Zdroj: HMD, vlastní výpočty.

Velice podobné pořadí bylo modelům přiřazeno v případě žen (tabulka 4.3). Jako nejvhodnější v tomto období se jevily modely Hillegman-Pollard a Kannistö, jako nejméně vhodné Coale-Kiskerův a Gompertz-Makehamův model. Kannistův model sice získal 7krát hodnocení 1, celkový průměr (2,5) byl ale ovlivněn známkou 5, která byla modelu přiřazena 4 státy. Celkově nabýval Kannistův model v tomto období pouze známek 1 a 5. V časovém úseku 1920–1929 se celkově pro obě pohlaví jako nejvhodnější jevil Heligman-Pollardův model (2,4), nejméně vhodný byl Coale-Kisker (4,3).

Země/model	g	m	k	t	h	c
Belgie	2	3	5	6	1	4
Dánsko	4	6	1	3	2	5
Finsko	5	6	1	3	2	4
Francie	4	2	5	6	1	3
Island	4	5	1	2	3	6
Itálie	6	4	1	2	5	3
Nizozemsko	6	5	1	2	3	4
Norsko	1	3	5	6	2	4
Španělsko	5	3	1	4	2	6
Švédsko	5	6	1	2	3	4
Švýcarsko	1	3	5	6	2	4
Průměrná známka	3,9	4,2	2,5	3,8	2,4	4,3
Pořadí	4	5	2	3	1	6

Tabulka 4.3: Pořadí analyzovaných modelů v období 1920–1929, ženy. Zdroj: HMD, vlastní výpočty

4.2.2 1950–1959

Po druhé světové válce prošla Evropa značnými geopolitickými změnami. Tyto změny – rozdělení na východní (východní blok) a západní část (Západ) – se projeví ve vývoji úmrtnosti v dalších letech. V tomto období poskytuje HMD data pro 24 zemí, které jsme rozdělili do dvou skupin. Skupinu A tvořily všechny státy, které vstupovaly do analýzy ve sledovaném období 1920–1929 (viz tabulka 4.1). Skupina B se skládala z ostatních států, za které začala být poskytována data až v letech 1921–1950.

Nejlépe hodnoceným modelem pro muže byl v obou skupinách Kannistův model, který získal podstatně lepší průměrnou známku ve skupině A (průměrná známka 1,3), kde byl hodnocen známkou 1 v 9 z 11 států (viz tabulka 4.4). Ve skupině B (průměrná známka 2,1) byl na první místo zařazen 8 ze 13 států (viz tabulka 4.5). Druhé místo v obou skupinách patřilo Heligman-Pollardovu modelu (průměrná známka 2,5 ve skupině A a 2,8 ve skupině B) a třetí Thatcherovu (průměrná známka 2,7 ve skupině A a 3,2 ve skupině B). Oproti období 1920–1929, kde byl jako třetí nejvhodnější vybrán Gompertzův model, se u mužů v této dekádě na všech třech nejvyšších místech umístily logistické modely. Poslední místo ve skupině A i B patřilo Gompertz-Makehamovu modelu.

U žen se v obou skupinách nejlépe umístil Kannistův model (průměrná známka 1,4 ve skupině A a 2,0 ve skupině B). Jako druhý nejvhodnější byl ve skupině A hodnocen Thatcherův

Země/model	g	m	k	t	h	c
Belgie	2	5	3	4	1	6
Dánsko	4	6	1	2	3	5
Finsko	3	5	2	4	1	6
Francie	4	6	1	2	3	5
Island	3	5	1	4	2	6
Itálie	5	6	1	2	3	4
Nizozemsko	4	6	1	2	3	5
Norsko	4	5	1	2	3	6
Španělsko	6	5	1	2	4	3
Švédsko	4	6	1	3	2	5
Švýcarsko	4	6	1	3	2	5
Průměrná známka	3,9	5,5	1,3	2,7	2,5	5,1
Pořadí	4	6	1	3	2	5

Tabulka 4.4: Pořadí analyzovaných modelů v období 1950–1959, skupina A, muži. Zdroj: HMD, vlastní výpočty

Země/model	g	m	k	t	h	c
Anglie a Wales	4	6	2	3	1	5
Austrálie	6	5	1	2	3	4
Bulharsko	6	4	1	2	5	3
Česko	2	4	5	6	1	3
Irsko	1	3	5	6	2	4
Japonsko	6	5	1	2	3	4
Kanada	4	5	1	2	3	6
Maďarsko	2	3	5	6	4	1
Nový Zéland	5	6	1	2	3	4
Portugalsko	6	5	1	2	4	3
Rakousko	4	5	1	2	3	6
Slovensko	3	5	2	4	1	6
USA	4	5	1	2	3	6
Průměrná známka	4,1	4,7	2,1	3,2	2,8	4,2
Pořadí	4	6	1	3	2	5

Tabulka 4.5: Pořadí analyzovaných modelů v období 1950–1959, skupina B, muži. Zdroj: HMD, vlastní výpočty

Země/model	g	m	k	t	h	c
Belgie	5	3	4	6	1	2
Dánsko	6	5	1	2	4	3
Francie	6	5	1	2	3	4
Finsko	6	5	1	2	4	3
Island	5	3	1	2	4	6
Itálie	6	5	1	2	4	3
Nizozemsko	6	5	1	2	4	3
Norsko	6	5	1	2	4	3
Španělsko	6	5	1	2	4	3
Švédsko	6	5	1	2	4	3
Švýcarsko	6	5	2	3	1	4
Průměrná známka	5,8	4,6	1,4	2,5	3,4	3,4
Pořadí	6	5	1	2	3–4	3–4

Tabulka 4.6: Pořadí analyzovaných modelů v období 1950–1959, skupina A, ženy. Zdroj: HMD, vlastní výpočty

model (2,5), viz tabulka 4.6, a ve skupině B Heligman-Pollardův (2,9), viz tabulka 4.7. V této dekádě byl oběma skupinami poměrně dobře hodnocen Coale-Kiskerův model, ve skupině B mu bylo přiřazeno 4. místo, ve skupině A dokonce 3. až 4. místo.

Pátý byl u obou skupin Gompertz-Makehamův model (4,6 ve skupině A a 4,2 ve skupině B), šesté místo se známkou 5,8 ve skupině A a 5,3 ve skupině B připadlo Gompertzově modelu. Analýza období 1950–1959 ukázala, že nejvhodnějším modelem je jednoznačně Kan-nistů a druhé místo patří Heligman-Pollardovu modelu. Mírně lepší průměrnou známku získal Thatcherův model, který u žen ve skupině A získal dokonce druhé místo.

V daném období se celkově nejhůře umístil Gompertz-Makehamův model, který byl v obou mužských skupinách šestý, ve skupinách žen pátý. Celkově pátý se umístil druhý exponenciální model – Gompertzův, který byl skupinami A i B u žen v období 1950–1959 vyhodnocen jako nejméně vhodný.

Země/model	g	m	k	t	h	c
Anglie a Wales	6	5	1	2	4	3
Austrálie	5	2	4	6	1	3
Bulharsko	6	4	2	1	5	3
Česko	6	4	3	5	1	2
Irsko	3	2	4	5	1	6
Japonsko	6	5	1	2	3	4
Kanada	6	5	1	2	3	4
Maďarsko	6	5	1	2	4	3
Nový Zéland	6	5	1	2	3	4
Portugalsko	6	5	1	2	4	3
Rakousko	6	5	1	2	4	3
Slovensko	1	3	5	6	2	4
USA	6	5	1	2	3	4
Průměrná známka	5,3	4,2	2,0	3,0	2,9	3,5
Pořadí	6	5	1	3	2	4

Tabulka 4.7: Pořadí analyzovaných modelů v období 1950–1959, skupina B, ženy. Zdroj: HMD, vlastní výpočty

4.2.3 1990–1999

Konec 20. století se nesl ve znamení změn v politickém uspořádání světa. Po rozpadu Sovětského svazu se začal východ a západ pozvolna demograficky sblížovat. Ve většině vyspělých států se v průběhu 90. let snižovala úroveň úmrtnosti ve středním a vyšším věku u mužů i u žen. To mělo za následek růst naděje dožití (SZÚ, 2006, s. 38–42). V tomto časovém úseku bylo analyzováno 36 států, které byly rozděleny do 3 skupin. První dvě skupiny byly stejné jako v analýze předchozího období. Nově byla přidána skupina C, do které patřily státy, za která byla HMD poskytnuta data až v období 1951–1990.

U mužů ve skupině A byla v letech 1990–1999 křivka úmrtnosti nejlépe vyrovnávána pomocí **Kannistova** modelu, který byl hodnocen průměrnou známkou 1,1. Jako nejvhodnější byl v případě 10 států z 11 (viz tabulka 4.8). Na druhé a třetí místo byl zařazen Heligman-Pollardův a Thatcherův model. Nejhůře z vybraných modelů vyrovnával v tomto období ve skupině A křivku úmrtnosti Gompertz-Makehamův model, jehož průměrná známka byla 5,4.

Ve skupině B byl v tomto desetiletí nejlépe hodnocen **Heligman-Pollardův** model s průměrnou známkou 2,3 (tabulka 4.9). Kannistův model byl hodnocen průměrnou známkou o čtyři desetiny horší (2,7). Další logistický model, Thatcherův, skončil třetí.

Země/model	g	m	k	t	h	c
Belgie	5	6	2	3	1	4
Dánsko	4	6	1	3	2	5
Finsko	4	6	1	3	2	5
Francie	4	5	1	2	3	6
Island	4	5	1	3	2	6
Itálie	4	5	1	2	3	6
Nizozemsko	6	5	1	2	3	4
Norsko	6	5	1	2	3	4
Španělsko	4	5	1	2	3	6
Švédsko	4	6	1	2	3	5
Švýcarsko	4	5	1	2	3	6
Průměrná známka	4,5	5,4	1,1	2,4	2,5	5,2
Pořadí	4	6	1	2	3	5

Tabulka 4.8: Pořadí analyzovaných modelů v období 1990–1999, skupina A, muži. Zdroj: HMD, vlastní výpočty

Země/model	g	m	k	t	h	c
Anglie a Wales	6	5	2	4	1	3
Austrálie	4	5	1	2	3	6
Bulharsko	1	3	6	5	2	4
Česko	1	3	6	5	2	4
Irsko	6	5	1	2	4	3
Japonsko	4	5	1	2	3	6
Kanada	4	6	2	3	1	5
Maďarsko	4	5	1	3	2	6
Nový Zéland	5	6	1	2	3	4
Portugalsko	4	3	5	6	1	2
Rakousko	4	5	2	3	1	6
Slovensko	3	2	6	5	4	1
USA	4	5	1	2	3	6
Průměrná známka	3,8	4,5	2,7	3,4	2,3	4,3
Pořadí	4	6	2	3	1	5

Tabulka 4.9: Pořadí analyzovaných modelů v období 1990–1999, skupina B, muži. Zdroj: HMD, vlastní výpočty

Nejméně vhodný z analyzovaných modelů je ve skupině B, stejně jako ve skupině A, **Gomperz-Makehamův** model, jehož průměrná známka byla 4,5, skončil dvě desetiny za Coale-Kiskerovým modelem.

Skupina C se značně odlišovala ve výběru modelu od předchozích dvou. **Kannistovu** modelu, který v předchozích obdobích ve všech skupinách končil nejhůře jako druhý, bylo v tomto případě přiřazeno až páté místo s průměrnou známkou 3,9 (tabulka 4.10).

Nejvhodnějším modelem pro vyrovnávání měr úmrtnosti byl v tomto období ve skupině C Gompertzův model s průměrnou známkou 2,3. Druhé místo bylo přiřazeno Heligman-Pollardovu modelu. Třetí místo obsadil, stejně jako ve skupině B, Thatcherův model. Nejméně vhodným modelem je pro tuto skupinu Coale-Kiskerův model.

Země/model	g	m	k	t	h	c
Bělorusko	2	4	5	3	1	6
Estonsko	1	2	6	5	4	3
Izrael	1	2	5	6	3	4
Litva	1	2	6	5	4	3
Lotyšsko	3	2	6	5	4	1
Lucembursko	3	5	2	4	1	6
Německo	4	5	1	2	3	6
Polsko	4	5	1	2	3	6
Rusko	1	4	5	3	2	6
Slovinsko	4	5	1	3	2	6
Taiwan	2	4	5	3	1	6
Ukrajina	1	5	4	3	2	6
Průměrná známka	2,3	3,8	3,9	3,7	2,5	4,9
Pořadí	1	4	5	3	2	6

Tabulka 4.10: Pořadí analyzovaných modelů v období 1990–1999, skupina C, muži. Zdroj: HMD, vlastní výpočty

U žen vykazovaly všechny tři skupiny podobné výsledky, nejvhodnějším modelem byl Kannistö. Zatímco ale ve skupině A byl mezi tímto a druhým Thatcherovým modelem rozdíl jedné (tabulka 4.11), ve skupině B se rozdíl mezi prvním a druhým pořadím snížil na šest desetin (tabulka 4.12) a ve skupině C byl jen dvě desetiny (tabulka 4.13). Sestupnou tendenci měla i průměrná známka. První skupina mu přiřadila hodnotu 1,3, druhá 1,6 a třetí dokonce 2,3. Ve všech třech skupinách žen první tři místa obsadily logistické modely.

Země/model	g	m	k	t	h	c
Belgie	4	5	1	2	3	6
Dánsko	4	5	2	3	1	6
Finsko	5	6	1	2	3	4
Francie	4	5	1	2	3	6
Island	2	5	3	4	1	6
Itálie	4	5	1	2	3	6
Nizozemsko	4	5	1	2	3	6
Norsko	4	5	1	2	3	6
Španělsko	4	6	1	2	3	5
Švédsko	4	5	1	2	3	6
Švýcarsko	4	5	1	2	3	6
Průměrná známka	3,9	5,2	1,3	2,3	2,6	5,7
Pořadí	4	5	1	2	3	6

Tabulka 4.11: Pořadí analyzovaných modelů v období 1990–1999, skupina A, ženy. Zdroj: HMD, vlastní výpočty

Země/model	g	m	k	t	h	c
Anglie a Wales	4	5	2	3	1	6
Austrálie	4	5	1	2	3	6
Bulharsko	6	3	4	5	2	1
Česko	5	6	2	3	1	4
Irsko	6	5	1	2	3	4
Japonsko	4	5	1	2	3	6
Kanada	4	5	1	2	3	6
Maďarsko	4	6	1	2	3	5
Nový Zéland	4	5	1	3	2	6
Portugalsko	6	5	3	4	1	2
Rakousko	4	5	1	2	3	6
Slovensko	4	6	2	3	1	5
USA	4	5	1	3	2	6
Průměrná známka	4,5	5,1	1,6	2,8	2,2	4,8
Pořadí	4	6	1	3	2	5

Tabulka 4.12: Pořadí analyzovaných modelů v období 1990–1999, skupina B, ženy. Zdroj: HMD, vlastní výpočty

Země/model	g	m	k	t	h	c
Bělorusko	6	5	1	2	3	4
Estonsko	1	3	6	5	2	4
Izrael	1	2	5	6	3	4
Litva	4	6	2	3	1	5
Lotyšsko	1	3	6	5	2	4
Lucembursko	4	5	1	2	3	6
Německo	4	5	1	2	3	6
Polsko	5	6	1	2	3	4
Rusko	4	6	1	2	3	5
Slovinsko	4	6	1	2	3	5
Taiwan	6	5	2	3	1	4
Ukrajina	6	5	1	2	3	4
Průměrná známka	3,8	4,8	2,3	3,0	2,5	4,6
Pořadí	4	6	1	3	2	5

Tabulka 4.13: Pořadí analyzovaných modelů v období 1990–1999, skupina C, ženy. Zdroj: HMD, vlastní výpočty

Ve skupině A byl v období 1990–1999 zařazen na šesté místo Coale-Kiskerův model s průměrnou známkou 5,7. Jen ve 2 z 11 států se umístil lépe než na šestém místě. Ve skupinách B a C šesté místo obsadil Gompertz-Makehamův model vždy jen o několik desetin za Coale-Kiskerovým modelem.

4.2.4 2000–2009

V období od roku 2000 do roku 2009 bylo analyzováno 35 států. Oproti předchozí dekádě (1990–1999) nevstupoval do analýzy Nový Zéland.

Ve skupině A se i v této dekádě nejlépe umístil **Kannistův** model s průměrnou hodnotou 1,5 (viz tabulka 4.14). Druhé a třetí místo si stejně jako v období 1950–1959 a 1990–1999 rozdělily modely Thatcher a Heligman-Pollard.

Nejméně vhodným modelem k vyrovnávání křivky úmrtnosti v tomto období byl ve skupině A **Coale-Kisker**, který získal průměrnou známku 5,6. O čtyři desetiny lepší průměrná známka byla přiřazena Gompertz-Makehamovu modelu, který se umístil pátý.

Země/model	g	m	k	t	h	c
Belgie	4	5	3	2	1	6
Dánsko	4	6	1	3	2	5
Finsko	4	5	1	2	3	6
Francie	4	5	3	2	1	6
Island	4	5	1	3	2	6
Itálie	4	6	1	2	3	5
Nizozemsko	6	5	2	3	1	4
Norsko	4	5	1	2	3	6
Španělsko	4	5	1	2	3	6
Švédsko	4	5	1	2	3	6
Švýcarsko	4	5	2	3	1	6
Průměrná známka	4,2	5,2	1,5	2,4	2,1	5,6
Pořadí	4	5	1	3	2	6

Tabulka 4.14: Pořadí analyzovaných modelů v období 2000–2009, skupina A, muži. Zdroj: HMD, vlastní výpočty

U mužů ve skupině B jsou také nejlépe hodnoceny logistické modely, mění se ale pořadí. Nejvhodnějším modelem v této skupině byl **Heligman-Pollardův**. Na druhém místě s rozdílem pouze jedné desetiny byl Kannistův model s průměrnou známkou 2,8 (viz 4.15). Na pátou příčku byl zařazen Gompertz-Makehamův model, šesté místo bylo přiřazeno **Coale-Kiskerovu** modelu.

Stejně jako ve skupině B byl i ve skupině C **Heligman-Pollardův** model vyhodnocen jako nejvhodnější z analyzovaných modelů, jak ukazuje tabulka 4.16. Druhé místo bylo přiřazeno

Země/model	g	m	k	t	h	c
Anglie a Wales	4	6	1	2	3	5
Austrálie	4	5	3	2	1	6
Bulharsko	4	5	1	3	2	6
Česko	1	4	5	3	2	6
Irsko	6	5	1	2	3	4
Japonsko	4	5	1	2	3	6
Kanada	1	3	6	4	2	5
Maďarsko	1	2	6	5	4	3
Portugalsko	4	6	1	2	3	5
Rakousko	4	5	1	3	2	6
Slovensko	1	2	6	5	4	3
USA	4	5	1	2	3	6
Průměrná známka	3,2	4,4	2,8	3,0	2,7	5,1
Pořadí	4	5	2	3	1	6

Tabulka 4.15: Pořadí analyzovaných modelů v období 2000–2009, skupina B, muži. Zdroj: HMD, vlastní výpočty

Gompertzově modelu. V této skupině byl jako nejméně vhodný vyhodnocen **Coale-Kiskerův** model, stejně jako ve skupině C v předchozím časovém úseku.

U žen se ve všech skupinách A, B i C jevil jako nejlepší **Kannistův model** (z 35 států vykazoval v případě 25 států nejnižší hodnoty Akaikeho informačního kritéria – byl nejvhodnější (viz tabulky 4.17, 4.18 a 4.19). Ve skupině A byl druhý Thatcherův model, ve skupině B a C byl druhý Heligman-Pollardův model. Opět platí, stejně jako ve všech ostatních obdobích, že první tři místa patří logistickým modelům. Nejméně vhodným modelem pro skupinu A a C je **Coale-Kiskerův** model. Ve skupině A byla jeho průměrná známka 6,0. Ve skupině B získal nejvyšší průměrnou známku Gompertz-Makehamův model. Stejně jako v minulé dekádě byl jako nejméně vhodný vyhodnocen Coale-Kiskerův model, 25 států mu přiřadilo 6. místo. Gompertz-Makehamovu modelu bylo v tomto období 6. místo přiřazeno 7krát.

Země/model	g	m	k	t	h	c
Bělorusko	1	3	5	4	2	6
Estonsko	1	5	4	3	2	6
Izrael	3	5	2	4	1	6
Litva	4	5	1	2	3	6
Lotyšsko	1	3	6	2	4	5
Lucembursko	4	5	1	3	2	6
Německo	4	5	3	2	1	6
Polsko	4	5	1	2	3	6
Rusko	1	4	5	3	2	6
Slovinsko	3	5	2	4	1	6
Taiwan	1	3	5	4	2	6
Ukrajina	1	4	6	3	2	5
Průměrná známka	2,3	4,3	3,4	3,0	2,1	5,8
Pořadí	2	5	4	3	1	6

Tabulka 4.16: Pořadí analyzovaných modelů v období 2000–2009, skupina C, muži. Zdroj: HMD, vlastní výpočty

Země/model	g	m	k	t	h	c
Belgie	4	5	1	2	3	6
Dánsko	2	5	4	3	1	6
Finsko	4	5	1	2	3	6
Francie	4	5	1	2	3	6
Island	4	5	1	2	3	6
Itálie	4	5	1	2	3	6
Nizozemsko	4	5	1	2	3	6
Norsko	4	5	1	2	3	6
Španělsko	4	5	1	2	3	6
Švédsko	4	5	1	2	3	6
Švýcarsko	4	5	1	2	3	6
Průměrná známka	3,8	5,0	1,3	2,1	2,8	6,0
Pořadí	4	5	1	2	3	6

Tabulka 4.17: Pořadí analyzovaných modelů v období 2000–2009, skupina A, ženy. Zdroj: HMD, vlastní výpočty

Země/model	g	m	k	t	h	c
Anglie a Wales	4	5	3	2	1	6
Austrálie	4	5	3	2	1	6
Bulharsko	6	5	1	2	3	4
Česko	4	6	1	2	3	5
Irsko	5	6	1	3	2	4
Japonsko	4	5	1	2	3	6
Kanada	4	5	3	2	1	6
Maďarsko	4	6	1	2	3	5
Portugalsko	5	6	1	2	3	4
Rakousko	4	5	1	2	3	6
Slovensko	6	5	3	4	1	2
USA	4	5	3	2	1	6
Průměrná známka	4,5	5,3	1,8	2,3	2,1	5,0
Pořadí	4	6	1	3	2	5

Tabulka 4.18: Pořadí analyzovaných modelů v období 2000–2009, skupina B, ženy. Zdroj: HMD, vlastní výpočty

4.2.5 Shrnutí

Z analýzy podle období vyplývá, že existuje hned několik faktorů, které mají na volbu nejvhodnějšího modelu pro daný stát vliv. Těmito faktory jsou pohlaví, doba, od které byla poskytována data, a období.

V rámci skupin byla nejméně proměnlivá skupina A. Ve všech analyzovaných obdobích mimo prvního sledovaného období 1920–1929 byl na základě průměrné známky mužů i žen jako nejvhodnější vyhodnocen Kannistův model. V období 1920–1929 se u mužů i žen jevil jako nejvhodnější Heligman-Pollard. Obecně se nejlépe umisťovaly logistické modely. U mužů se mírně lépe umisťoval Heligman-Pollardův model, u žen Thatcherův.

Pro ženy ve skupině A bylo nejméně vhodnou variantou užití Coale-Kiskerova modelu. Mimo období 1950–1959 byl vždy šestý. Pro muže z této skupiny nebyly vhodnými modely Coale-Kiskerův a Gompertz-Makehamův. Tyto modely ani jednou nezískaly lepší než páté místo.

Skupina B se příliš nelišila od skupiny A. U mužů se jako nejvhodnější jevil v dekáдах 1990–1999 a 2000–2009 Heligman-Pollardův model, u žen ve všech třech obdobích ve skupině B dominoval model Kannistö.

Země/model	g	m	k	t	h	c
Bělorusko	5	6	1	2	3	4
Estonsko	4	5	2	3	1	6
Izrael	6	5	1	3	2	4
Litva	2	5	4	3	1	6
Lotyšsko	4	5	2	3	1	6
Lucembursko	4	5	1	3	2	6
Německo	4	5	1	2	3	6
Polsko	4	5	1	2	3	6
Rusko	4	6	1	2	3	5
Slovinsko	4	5	1	2	3	6
Taiwan	2	5	4	3	1	6
Ukrajina	4	6	1	2	3	5
Průměrná známka	3,9	5,3	1,7	2,5	2,2	5,5
Pořadí	4	5	1	3	2	6

Tabulka 4.19: Pořadí analyzovaných modelů v období 2000–2009, skupina C, ženy. Zdroj: HMD, vlastní výpočty

V letech 1950–1959 byl v případě žen nejhůře hodnocen Gompertzův model, v dalších dvou vybraných desetiletích (1990–1990 a 2000–2009) to byl Gompertz-Makehamův model, který v analyzovaném období 1950–1959 skončil pátý.

Nejvyšší hodnoty pořadí u mužů ve skupině B v období 1950–1959 a 1990–1999 vykazoval Gompertz-Makehamův model. V posledním sledovaném časovém úseku dosáhl nejvyšší průměrné známky Coale-Kiskerův model.

Poslední skupina se ve výběru modelu velmi lišila od ostatních dvou. Značný vliv na výběr modelu měly v této skupině země bývalého východního bloku, kterých ve skupině 12 zemí bylo 7. V případě mužů byl ve skupině C v obou sledovaných obdobích (1990–1999 a 2000–2009) nejvhodnějším modelem Gompertzův a druhým nejvhodnějším Heligman-Pollardův model. U žen ze skupiny C se stejně jako u ostatních skupin nejlépe umístily logistické modely, zejména model Kannistö. Druhý nejvhodnější byl stejně jako u mužů Heligman-Pollard. Nejhorším z analyzovaných modelů byl u mužů v obou obdobích, za která byla skupina C analyzována, Coale-Kisker. U žen v prvním sledovaném období (1990–1999) bylo šesté místo přiřazeno Gompertz-Makehamovu modelu, v druhém (2000–2009) Coale-Kiskerovu modelu.

Analýza ukázala, že více než čas ovlivňuje výběr nejvhodnějšího modelu charakter úmrtnosti v daném státu. Z toho důvodu jsme se rozhodli analyzovat data dalším způsobem – podle úrovně naděje dožití při narození.

4.3 Analýza vybraných států podle úrovně naděje dožití při narození

Na základě úrovně **naděje dožití** při narození bylo sestaveno celkem osm skupin, čtyři pro muže, čtyři pro ženy, jak ukazuje tabulka 4.20. Informace o tomto ukazateli byly čerpány z webových stránek *Population Reference Bureau*. Nabízelo se i dělení podle naděje dožití při narození, kterou poskytuje na svých stránkách HMD, nicméně země se různí v posledním roce, pro který je naděje dožití při narození počítána. Státy byly do skupin rozděleny na základě e_0 v roce 2011.

Skupina	e_0	Muži	e_0 let	Ženy
I	79–80 let	Austrálie, Itálie, Izrael, Japonsko, Nizozemsko, Norsko, Nový Zéland, Španělsko, Švédsko, Švýcarsko	84–86 let	Japonsko, Španělsko, Francie, Izrael, Island, Švédsko, Švýcarsko, Itálie, Austrálie
II	77–78 let	Anglie a Wales, Belgie, Dánsko, Finsko, Francie, Irsko, Kanada, Lucembursko, Německo, Rakousko	83 let	Finsko, Kanada, Lucembursko, Německo, Nizozemsko, Norsko, Nový Zéland, Rakousko, Slovinsko
III	71–76 let	Česko, Chile, Polsko, Portugalsko, Slovensko, Slovinsko, Taiwan, USA	80–82 let	Anglie a Wales, Belgie, Česko, Dánsko, Estonsko, Chile, Irsko, Polsko, Portugalsko, Taiwan, USA
IV	63–70 let	Bělorusko, Bulharsko, Estonsko, Litva, Lotyšsko, Maďarsko, Rusko, Ukrajina	75–79 let	Bělorusko, Bulharsko, Litva, Lotyšsko, Maďarsko, Rusko, Slovensko, Ukrajina

Tabulka 4.20: Skupiny států podle naděje dožití. Zdroj: PRB

Akaikeho informační kritérium bylo vypočteno jako průměr všech informačních kritérií pro jednotlivé kalendářní roky za celé období, ve kterém byla data státu poskytována. Do analýzy mohlo být zařazeno Chile. Z důvodu malého počtu událostí a malé velikosti exponované populace v jednotlivých letech byl z analýzy vyřazen Island.

4.3.1 Analýza skupin mužů

Skupina I

Skupina I se vyznačuje **nejvyšší nadějí dožití**. Patří sem státy, jejichž úroveň naděje dožití při narození se v roce 2011 pohybovala v intervalu 79,00–80,99 let. Obecně lze říct, že tyto země jsou řazeny mezi **nejrozvinutější** země světa, mají dobrou úroveň zdravotnictví a vyznačují se nízkou úrovní úmrtnosti (World Population Data Sheet, 2011).

Země/model	g	m	k	t	h	c
Austrálie	4	5	1	3	2	6
Itálie	6	4	1	3	2	5
Izrael	1	3	5	6	2	4
Japonsko	6	5	1	2	3	4
Nizozemsko	5	6	1	3	2	4
Norsko	4	6	2	3	1	5
Nový Zéland	4	6	1	2	3	5
Španělsko	6	5	1	2	3	4
Švédsko	1	3	5	6	2	4
Švýcarsko	4	5	1	3	2	6
Průměrná známka	4,1	4,8	1,9	3,3	2,2	4,7
Pořadí	4	6	1	3	2	5

Tabulka 4.21: Pořadí analyzovaných modelů, skupina I, muži. Zdroj: HMD, vlastní výpočty.

Jak ukazuje tabulka 4.21, **nejnižší** průměrná známka byla v této skupině přiřazena Kanistovu modelu, jehož celková známka je 1,9. Průměr silně ovlivnila druhá nejvyšší průměrná známka v případě Švédky a Izraele. V případě Švédska je to způsobeno již zmíněnou délkou sledovaného období. Pokud by bylo analyzováno Švédsko ve dvou obdobích, první úsek od roku 1751–1914, nejnižší průměrnou známku by v něm získal Gompertzův model a těsně za ním by byl Gompertz-Makehamův populační model. V druhém úseku (1915–2011) se jevil

jako nejvhodnější model Kannistö. Gompertz-Makehamův je naopak hodnocen známkou 6, Gompertzův 4.

Druhé místo ve skupině I pak náleží Heligman-Pollardovu modelu. Naopak **nejhůře** hodnocen byl Gompertz-Makehamův model s průměrnou známkou 4,8 v těsném závěsu za Coale-Kiskerovým lineárním modelem.

Pokud bychom vyřadili ze skupiny I země Švédsko a Izrael, celkové pořadí by se příliš nezměnilo, jak ukazuje tabulka 4.22. Znatelně by se ale zvýšil rozdíl mezi maximální a minimální průměrnou známkou.

Země/model	g	m	k	t	h	c
Průměrná známka	4,9	5,3	1,1	2,6	2,3	4,9
Pořadí	4–5	6	1	3	2	4–5

Tabulka 4.22: Pořadí analyzovaných modelů po vyřazení Švédska a Izraele, skupina I, muži. Zdroj: HMD, vlastní výpočty.

Logistické modely v našem hodnocení Kannistö, Heligman-Pollard a Thatcher (v tomto pořadí) obsadily první tři příčky. Potvrdila se dominance modelu Kannistö, který místo průměrné známky 1,9 získal známku 1,1. Poslední zůstává i nadále Gompertz-Makehamův model a o 4. a 5. místo se dělí Coale-Kisker a Gompertz shodně s průměrem 4,9. Po vyřazení výše zmíněných zemí dostáváme větší škálu hodnot (od 1,1 až po 5,3). Vyřazení modelů slouží jen ilustraci vlivu Švédska a Izraele. Ke konečné analýze je použita tabulka 4.21, která obě země zahrnuje.

Skupina II

Skupinu II tvoří státy, jejichž naděje dožití při narození se v roce 2011 pohybovala v intervalu **77,00–78,99 let**. Do této skupiny patří celkem 10 zemí.

Pořadí modelů se oproti skupině I dramaticky nemění. Nejlépe hodnocen byl opět model **Kannistö** s průměrným pořadím 1,5, jak ukazuje tabulka 4.23. **Druhé místo** získal Heligman-Pollardův model a třetí Thatcherův. Oproti skupině I se ve skupině II jako **nejméně vhodný** ukazuje Coale-Kiskerův model s průměrnou známkou 5,4, který si vyměnil pořadí s Gompertz-Makehamovým modelem.

Země/model	g	m	k	t	h	c
Finsko	4	5	1	3	2	6
Francie	5	4	1	3	2	6
Kanada	4	5	1	3	2	6
Dánsko	3	5	2	4	1	6
Lucembursko	3	5	2	4	1	6
Německo	4	5	2	3	1	6
Belgie	2	5	3	6	1	4
Anglie a Wales	6	5	1	2	3	4
Irsko	6	5	1	2	3	4
Rakousko	4	5	1	2	3	6
Průměrná známka	4,1	4,9	1,5	3,2	2,1	5,4
Pořadí	4	5	1	3	2	6

Tabulka 4.23: Pořadí analyzovaných modelů, skupina II, muži. Zdroj: HMD, vlastní výpočty.

Skupina III

Země další skupiny představují z hlediska celkové intenzity úmrtnosti velmi heterogenní soubor, naděje dožití při narození se pohybovala v roce 2011 v intervalu **71,00–76,99 let**. Zatímco v předchozích dvou skupinách, I i II, byl tento interval 2 roky, v tomto případě to je **6 let**. Je to z toho důvodu, aby skupiny byly přibližně stejně velké. Heterogenita se projevuje i v nejednoznačnosti při výběru nejlepšího modelu a je dobře vidět na rozdílu mezi maximální a minimální hodnotou. Průměrné známky mají menší rozptyl.

Země/model	g	m	k	t	h	c
Česko	2	5	3	4	1	6
Chile	1	3	4	6	2	5
Polsko	4	5	1	2	3	6
Portugalsko	6	5	1	2	3	4
Slovensko	1	2	6	5	4	3
Slovinsko	4	5	1	3	2	6
Taiwan	3	2	4	5	1	6
USA	4	5	1	2	3	6
Průměrná známka	3,1	4,0	2,6	3,6	2,4	5,3
Pořadí	3	5	2	4	1	6

Tabulka 4.24: Pořadí analyzovaných modelů, skupina III, muži. Zdroj: HMD, vlastní výpočty.

Nejlépe hodnocený model pro skupinu III je oproti předchozím skupinám **Heligman- - Pollardův** s průměrnou známkou 2,4. **Kannistův** model je s malým rozdílem na druhém místě, dosahuje hodnoty 2,6. Zatímco v předchozích skupinách byly vždy na prvních třech místech modely s logistickým principem výpočtu, v tomto případě je třetí exponenciální Gompertzův populační model. Mezi zeměmi ve skupině jsou čtyři, které byly v minulosti součástí východního bloku². Tyto země vykazují v porovnání se západními státy **vyšší úroveň úmrtnosti**.

Nejhůře hodnoceným modelem v této skupině je **Coale-Kiskerův**. V tomto případě dosahuje hodnoty 5,3 a dostal jinou známku než 6 pouze třikrát.

Skupina IV

Poslední skupina vykazovala v roce 2011 **nejnižší hodnoty naděje dožití při narození – 63,00–70,99 let**. Celkem čítá 8 zemí. Všechny země náležely před rokem 1989 k východnímu bloku, avšak i přes více než 40 let společné minulosti se po rozpadu východního bloku státy vyvíjely demograficky odlišně.

V rámci této skupiny lze podle naděje dožití při narození nalézt další **dvě podskupiny**. První se skládala ze států, ve kterých byla naděje dožití při narození v roce 2011 větší než 68 let. Tato skupina (IVA) se skládala z Estonska, Lotyšska, Litvy, Bulharska a Maďarska, do druhé skupiny (IVB) s nadějí dožití nižší než 65 let patřilo Bělorusko, Rusko a Ukrajina. Absolutně nejnižší hodnoty naděje dožití při narození vykazovalo Rusko, přibližně 63 let. Tyto dvě skupiny jsou natolik liší ve výběru modelu, že považuji za důležité uvést základní odlišnosti u každé z nich, než budou analyzovány jako jeden celek.

Země/model	g	m	k	t	h	c
Bulharsko	6	5	1	2	4	3
Estonsko	1	3	5	6	2	4
Litva	2	5	3	4	1	6
Lotyšsko	2	1	6	5	4	3
Maďarsko	2	4	3	5	1	6
Průměrná známka	2,6	3,0	4,0	4,4	2,4	4,4
Pořadí	2	3	4	5–6	1	5–6

Tabulka 4.25: Pořadí analyzovaných modelů, skupina IVA, muži. Zdroj: HMD, vlastní výpočty.

²Neoficiální označení Sovětského svazu a jeho politických satelitů.

V rámci první podskupiny je nejlépe hodnocen Heligman-Pollardův model, o pouhé dvě desetiny bodu za ním Gompertzův. Tabulka 4.25 ukazuje, že se státy Estonsko, Litva, Lotyšsko a Maďarsko v preferencích modelů příliš neliší. Oproti tomu Bulharsko vykazuje značně odlišné hodnoty v rámci této skupiny. Pokud by byl eliminován jeho vliv, v celkovém hodnocení by Heligman-Pollardův model skončil jako druhý. Gompertzův jako první. Další změny v pořadí by neproběhly.

Země/model	g	m	k	t	h	c
Bělorusko	3	5	2	4	1	6
Rusko	4	5	2	1	3	6
Ukrajina	4	5	3	2	1	6
Průměrná známka	3,7	5,0	2,3	2,3	1,7	6,0
Pořadí	4	5	2–3	2–3	1	6

Tabulka 4.26: Pořadí analyzovaných modelů, skupina IVB, muži. Zdroj: HMD, vlastní výpočty.

Druhá podskupina byla více homogenní. Jako nejvhodnější model se jevil Heligman-Pollardův. Na 2. až 3. místě byly další dva logistické modely. Poslední byl Coale-Kiskerův model, který nabývá hodnoty 6,0.

Po sloučení podskupin zůstane na prvním místě **Heligman-Pollardův** model. Převáží vliv početnější podskupiny a druhé místo připadne **Gompertzovu** modelu. Naproti tomu šesté místo zůstane Coale-Kiskerovu modelu.

Země/model	g	m	k	t	h	c
Bělorusko	3	5	2	4	1	6
Bulharsko	6	5	1	2	4	3
Estonsko	1	3	5	6	2	4
Litva	2	5	3	4	1	6
Lotyšsko	2	1	6	5	4	3
Maďarsko	2	4	3	5	1	6
Rusko	4	5	2	1	3	6
Ukrajina	4	5	3	2	1	6
Průměrná známka	3,0	3,8	3,4	3,6	2,1	5,0
Pořadí	2	5	3	4	1	6

Tabulka 4.27: Pořadí analyzovaných modelů, skupina IV, muži. Zdroj: HMD, vlastní výpočty.

4.3.2 Analýza skupin žen

Skupina I

Skupina I se stejně jako u mužů vyznačuje dlouhověkostí. Světové prvenství patří japonským ženám, které dosahovaly v roce 2011 naděje dožití při narození 86 let (World Population Data Sheet, 2011). Do skupiny I patří státy, u nichž se naděje dožití při narození pohybuje v uzavřeném intervalu **84,00–86,99 let** (viz tabulka 4.20).

Země/model	g	m	k	t	h	c
Austrálie	6	4	2	3	1	5
Francie	6	5	1	3	2	4
Itálie	6	4	1	2	3	5
Izrael	2	3	4	6	1	5
Japonsko	4	6	1	2	3	5
Španělsko	5	6	1	2	3	4
Švédsko	4	5	1	3	2	6
Švýcarsko	4	5	1	3	2	6
Průměrná známka	4,6	4,8	1,4	3,0	2,1	5,0
Pořadí	4	5	1	3	2	6

Tabulka 4.28: Pořadí analyzovaných modelů, skupina I, ženy. Zdroj: HMD, vlastní výpočty

Stejně jako u mužů je nejvíce vyhovujícím modelem Kannistö, který získal známku 1,4. Druhé místo patří Heligman-Pollardovu modelu. Třetí je Thatcherův model.

Rozdíly mezi muži a ženami ve skupině I jsou pouze v posledních dvou místech, kdy pro ženy je nejhorší variantou Coale-Kiskerův lineární model. Těsně před ním, se známkou jen o 0,25 nižší, se umístil Gompertz-Makehamův model. U mužů je rozdíl mezi 5. a 6. místem ještě o něco menší – patý skončil Coale-Kisker, který předběhl Gompertz-Makehamův model pouze o jednu desetinu. Ve skupině I u mužů i žen je poměrně velký rozdíl mezi modelem v druhém a třetím pořadí (vždy skoro 1), naopak rozdíly mezi posledními dvěma jsou relativně malé.

Ve skupině I v případě mužů se v pořadí modelů v rámci skupiny výrazně lišily státy Švédsko a Izrael. V případě žen se výrazněji lišil jen Izrael.

Skupina II

Skupinu II tvoří 9 států, které měly úroveň naděje dožití při narození v roce 2011 mezi **83,00–83,99 roky**. Tato skupina je geograficky velice rozmanitá, patří sem země ze severní, střední i jižní Evropy, Kanada a Nový Zéland. Avšak i přes geografickou rozmanitost je v této skupině vhodný model vybrán poměrně jednoznačně.

Země/model	g	m	k	t	h	c
Finsko	6	5	1	2	3	4
Kanada	4	5	1	3	2	6
Lucembursko	4	5	2	3	1	6
Německo	4	5	2	3	1	6
Nizozemsko	5	6	1	2	3	4
Norsko	4	5	1	3	2	6
Nový Zéland	4	5	1	2	3	6
Rakousko	5	6	1	2	3	4
Slovinsko	4	6	1	2	3	5
Průměrná známka	4,4	5,3	1,1	2,3	2,6	5,2
Pořadí	4	6	1	2	3	5

Tabulka 4.29: Pořadí analyzovaných modelů, skupina II, ženy. Zdroj: HMD, vlastní výpočty.

První tři místa obsadily, stejně jako v minulém případě, logistické modely Kannistö, Thatcher a Heligman-Pollard. Poslední dva jmenované modely si oproti skupině I pouze vyměnily místa, rozdíl mezi nimi je 0,3. Coale-Kiskerův lineární model se umístil jen velmi těsně před Gompertz-Makehamovým modelem (5,3).

Nejvhodnějším modelem pro všechny státy této skupiny, s výjimkou Lucemburska, je poměrně jednoznačně Kannistö. Celkově tak získává známku 1,1, což je nejnižší hodnota v rámci všech zpracovaných dat.

Skupina III

Velice rozmanitou skupinou, jejíž naděje dožití při narození se pohybuje mezi **80,00–82,99** roky, je skupina III. Oproti předchozím dvěma skupinám je více heterogenní co se týče výběru nejlepšího modelu. To vysvětluje i malé rozdíly ve známkování, kdy celkové rozpětí známek je pouze 2,9, zatímco u skupiny I to bylo 3,5 a u skupiny II dokonce 4,2.

Země/model	g	m	k	t	h	c
Anglie a Wales	6	5	1	2	3	4
Belgie	5	6	2	3	1	4
Česká republika	6	5	1	3	2	4
Dánsko	3	5	2	4	1	6
Estonsko	3	5	2	4	1	6
Irsko	6	5	1	2	3	4
Polsko	6	5	1	2	3	4
Portugalsko	6	5	1	2	4	3
USA	4	5	1	2	3	6
Průměrná známka	4,5	4,6	1,9	3,0	2,1	4,8
Pořadí	4	5	1	3	2	6

Tabulka 4.30: Pořadí analyzovaných modelů, skupina III, ženy. Zdroj: HMD, vlastní výpočty.

Mezi těmito 11 zeměmi je opět nejnižší průměrnou známkou hodnocen **Kannistův model** (1,9), ovšem s výrazně menším odstupem než tomu bylo u předchozí skupiny. Akaikeho informační kritérium tomuto modelu v případě Taiwanu určuje čtvrté místo, Chile dokonce až páté. Pokud bychom tyto státy vyřadili, první tři místa by se více lišila průměrnou známkou, v pořadí by se tato změna nepromítla. Jiné je to se zbývajících třemi modely, kdy vyřazení Chile a Taiwanu znamená změnu pořadí. Původně byl nejhůře hodnocen Coale-Kiskerův model. Po vyřazení států Chile a Taiwanu by byl jako nejméně vhodný zvolen Gompertz-Makehamův model, který byl nejlépe známkován hodnotou 5.

Skupina IV

Vůbec nejvíce geograficky homogenní skupinou je skupina IV, jejíž naděje dožití při narození se pohybuje v rozmezí **75,00–79,99 let**. Skupina IV se skládá pouze ze států bývalého **východního bloku**, stejně jako skupina IV v oddíle 4.3.1. Na rozdíl od mužské se ženská skupina více shoduje ve výběru modelu. Zcela patrné je to na posledních dvou místech, jak ukazuje tabulka 4.31.

První tři místa jsou obsazena podobně jako u ostatních modelů. Prvenství si opět drží logistické modely. Nejvhodnější pro data skupiny IV je Kannistův model s hodnotou 1,5, následovaný Thatcherovým a Heligman-Pollardovým populačním modelem. Zajímavá jsou ale poslední dvě místa, na kterých se umístily exponenciální modely se známkou 6,0 (Gompertzův) a 5,0 (Gompertz-Makehamův). Ačkoli až na výjimky patří do skupiny IV u mužů i žen stejné státy, pořadí modelů se ve skupinách liší.

Země/model	g	m	k	t	h	c
Bělorusko	6	5	1	2	3	4
Bulharsko	6	5	2	1	4	3
Litva	6	5	1	2	3	4
Lotyšsko	6	5	2	3	1	4
Maďarsko	6	5	1	2	3	4
Rusko	6	5	1	2	3	4
Slovensko	6	5	3	4	1	2
Ukrajina	6	5	1	2	3	4
Průměrná známka	6,0	5,0	1,5	2,3	2,6	3,6
Pořadí	6	5	1	2	3	4

Tabulka 4.31: Pořadí analyzovaných modelů, skupina IV, ženy. Zdroj: HMD, vlastní výpočty.

4.3.3 Shrnutí

V tabulce 4.32 a 4.33 jsou shrnuty celkové průměrné známky z jednotlivých skupin. Na základě průměru těchto známek bylo každému modelu přiřazeno pořadí.

Dalším možným způsobem analýzy bylo přiřadit známky průměrnému pořadí. Jednotná forma by byla jednou (a snad i jedinou) výhodou tohoto postupu, od kterého bylo nakonec upuštěno z toho důvodu, že do souhrnné analýzy vstupovaly vždy jen čtyři skupiny, průměry by často měly naprosto stejné skóre. Nebylo tedy možné jednoznačně určit pořadí.

Skupina	g	m	k	t	h	c
I	4,1	4,8	1,9	3,3	2,2	4,7
II	4,1	4,9	1,5	3,2	2,1	5,4
III	3,1	4,0	2,6	3,6	2,4	5,3
IV	3,0	3,8	3,4	3,6	2,1	5,0
Průměrná známka	3,6	4,4	2,4	3,4	2,2	5,1
Celkové pořadí	4	5	2	3	1	6

Tabulka 4.32: Shrnutí pořadí ve všech skupinách, muži. Zdroj: HMD, vlastní výpočty.

Výsledné tabulky 4.32 a 4.33 ukazují velmi zajímavé výsledky. Zejména analýza mužů odhalila několik výrazných tendencí. Mezi muži a ženami se v rámci skupin objevily poměrně značné rozdíly ve výběru vhodných modelů. Zatímco ve vybraných 36 státech byl mezi muži z nejvyspělejšího a nejméně vyspělého státu rozdíl hodnot naděje dožití při narození přibližně 17 let, mezi ženami to bylo podstatně méně, přibližně 11 let. Formování skupin hraje v analýze

Skupina	g	m	k	t	h	c
I	4,6	4,8	1,4	3,0	2,1	5,1
II	4,4	5,3	1,1	2,3	2,6	5,2
III	4,6	4,6	1,9	3,0	2,1	4,8
IV	6,0	5,0	1,5	2,3	2,6	3,6
Průměrná známka	4,90	4,93	1,5	2,6	2,3	4,7
Celkové pořadí	5	6	1	3	2	4

Tabulka 4.33: Shrnutí pořadí ve všech skupinách, ženy. Zdroj: HMD, vlastní výpočty.

velmi důležitou roli. U obou pohlaví mají první dvě skupiny rozpětí naděje dožití při narození menší, muži maximálně 3 roky, ženy 2 roky. U mužů je v případě skupin III a IV toto rozpětí větší, 6 let ve skupině III a 8 let ve skupině IV, u žen je nejširší interval, ve kterém se pohybují hodnoty naděje dožití při narození 5 let ve skupině IV. Značné rozdíly naděje dožití při narození v rámci skupiny jsou příčinou malé difference mezi extrémy – mezi nejlepším a nejhorším modelem u skupiny III a IV u mužů a u skupiny III u žen je rozdíl jen 2,9 – a horšího rozlišení mezi jednotlivými modely. V absolutních číslech je tento rozdíl stejný i ve skupině I, kde se opět projevuje vliv Izraele a Švédska, jak je uvedeno výše. Po vyřazení těchto států by se rozdíl extrémů razantně zvýšil, jak ukazuje tabulka 4.22. Obrázek 4.3 ilustruje rozdíly mezi extrémními hodnotami u mužů a žen. Největší rozdíl je mezi skupinami IV, u žen je tento rozdíl způsoben průměrnou známkou 6,0. Ve všech čtyřech skupinách jsou o něco lépe diferencovány ženy. Podobně jako v případě mužů jsou ve skupině C malé rozdíly mezi extrémy právě z důvodu heterogenity skupiny.

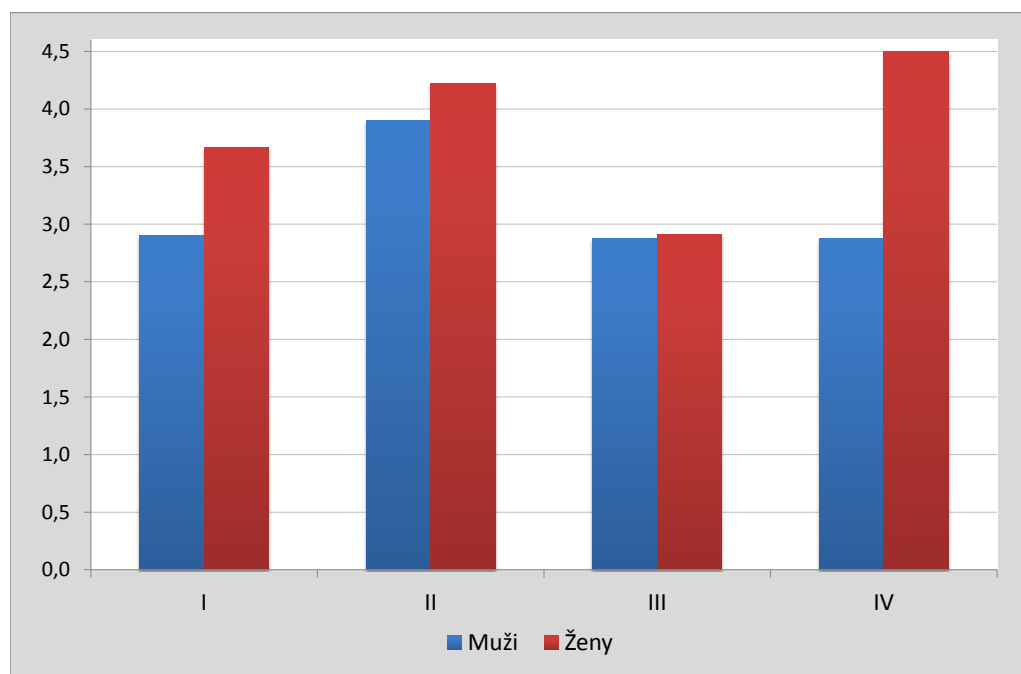
Kannistö (k)

Všechny logistické modely se, nezávisle na pohlaví, umisťovaly více než dobře. Podle srovnání modelů v článku *The Force of Mortality at Ages 80 to 120* (Thatcher, Kannistö a Vaupel, 1998) byl model poměrně spolehlivý u mužů do zhruba 90 let, u žen až do 95 let.

Data za mužskou část populace umisťují model Kannistö až na druhé místo s odstupem jen dvou desetín bodu. V prvních dvou skupinách I a II byl tento model hodnocen jako nejvhodnější. Ve skupině III se umístil jako druhý a ve skupině IV byl hodnocen jako třetí nejvhodnější. U tohoto modelu je naznačena přímá úměrnost – čím je vyšší úroveň intenzity úmrtnosti, tím je vyšší průměrné pořadí.

V tabulce 4.33 jsou shrnuty výsledky skupin žen. Není zde naznačena žádná úměrnost, hodnoty oscilují mezi 1,1 a 1,9. Ve všech čtyřech skupinách byl Kannistův model hodnocen jako nejlepší, jednoznačně mu bylo přiřazeno první místo s celkovým průměrem 1,5.

Obrázek 4.3: Porovnání rozptylu mezi muži a ženami ve skupinách. Zdroj: HMD, vlastní výpočty.



Heligman-Pollard (h)

Další analyzovaný logistický model by se o celkové prvenství mohl dělit s Kannistovým. Stejně jako Kannistö je i Heligman-Pollardův model dvouparametrický.

V případě mužů vystihuje tento model data velice dobře od 80 do 95 let. Poté má tendenci pravděpodobnost úmrtí mírně podhodnocovat (Thatcher, Kannistö a Vaupel, 1998). Celková průměrná známka Heligman-Pollardova modelu u mužů je 2,2, což jej řadí na první místo mezi hodnocenými modely. Ve všech skupinách se jeho průměrná známka pohybuje lehce nad 2. Ve skupinách I a II byl zařazen na druhé místo za Kannistův model. Ve skupinách III a IV se dostal na první místo před Gompertzův a Kannistův model.

U žen byl interval, ve kterém byly hodnoty hodnoceny jako spolehlivé, ještě o něco užší – mezi věky 80–92. Dokonce i v tomto intervalu měl model sklony mírně data nadhodnocovat. Vně intervalu naopak pravděpodobnost úmrtí podhodnocoval (Thatcher, Kannistö a Vaupel, 1998). Celkově je pro ženskou populaci vybraných zemí poměrně vhodný, v analýze podle

intenzity úmrtnosti byl druhý nejvhodnější. Nejlepší hodnocení bylo tomuto modelu přiřazeno skupinami I a III, shodně s průměrnou známkou 2,1. V dalších dvou skupinách byla průměrná známka také shodná – 2,6.

Zatímco u mužů je Heligman-Pollardův model (průměrná známka 2,2) těsně následován Kannistovým (průměrná známka 2,4). V případě žen skončil Kannistův model (s průměrnou známkou 1,5) před Heligman-Pollardovým (s průměrnou známkou 2,3) se značným odstupem.

Thatcher (t)

Třetí a poslední analyzovaný logistický model se zařazuje do první trojice nejvhodnějších modelů. Hodnoty Akaikeho informačního kritéria u mužů i žen tomuto modelu shodně přiřadily třetí místo. Celkově model trochu lépe vystihoval data žen, kdy byl střídavě umisťován na druhé a třetí místo. U mužů platilo, že vyšší naděje dožití při narození u skupin I a II znamenala i lepší známku.

Gompertzův model (g)

Gompertzův model se jevil jako poměrně spolehlivý, nejstarší ze všech modelů byl u mužů celkově hodnocen jako čtvrtý nejvhodnější. V případě žen získal až místo páté, a to pouze o několik setin před Coale-Kiskerovým modelem. Zatímco u mužů je naznačena nepřímá úměrnost mezi nadějí dožití a hodnocením modelu, u žen v prvních třech skupinách kolísá průměrná známka Gompertzova modelu kolem 3. Celkové umístění tohoto modelu v rámci ženské populace silně negativně ovlivňuje skupina IV, kde se jeví jako jednoznačně nejhorší se známkou 6,0. Celkově Gompertzův model dosahuje lepších výsledků v případě mužů a je poměrně vhodný pro muže ze států s nižší nadějí dožití.

Gompertz-Makehamův model (m)

Vyšší průměrné pořadí bylo přiřazeno Gompertzovu modelu modifikovanému o tzn. Makehamovu korekci. Makehamova korekce přidává ke Gompertzově rovnici konstantu značící úmrtí způsobená vnějšími příčinami. V celkovém hodnocení skončil Gompertz-Makehamův model u mužů až pátý, u žen šestý. Stejně jako v předchozím případě pozorujeme nepřímou úměrnost v případě mužů, nebo-li čím vyšší hodnota naděje dožití při narození, tím je model méně vhodný. U žen opět není naznačen žádný směr závislosti, průměrná známka ve všech čtyřech skupinách kolísá mezi 4,6 a 5,3.

Coale-Kisker (c)

Nejméně vhodným srovnávaným modelem byl Coale-Kiskerův, který je dimenzován na věkovou skupinu nad 85 let. Nezávisle na pohlaví byl poměrně spolehlivý od věku 85 až do věku 95 let. Poté se jednotlivé země odchylují od hodnoty jedna směrem nahoru i dolů (Thatcher, Kannistö a Vaupel, 1998).

V analýze 13 zemí se umístil jako pátý před Gompertzovým modelem (Thatcher, Kannistö a Vaupel, 1998). V předkládané práci jej Gompertzův model zanechal za sebou. Coale-Kiskerův model skončil za Gompertz-Makehamovým modelem, který nebyl v rámci příspěvku *The Force of Mortality at Ages 80 to 120* vůbec srovnáván.

Poslední místo tomuto modelu přiřadila analýza mužů, u žen získal čtvrté místo před dvěma exponenciálními modely, což je podobný výsledek jako ve výše zmíněném článku. V analýze žen byl hodnocen výrazně lépe, průměrnou známkou 4,7. Celkově byl Coale-Kiskerův model jediný model, u kterého byla u žen naznačena závislost. Obecně platilo, že čím jsou nižší hodnoty naděje dožití při narození, tím lepší známku Coale-Kiskerův model získal. To ale neplatilo v případě mužů, kde pouze ve skupině I získal tento model průměrnou známku nižší než 5 (4,1), jeho celková průměrná známka byla potom 5,1.

Závěr

Předkládaná práce si kladla za cíl představit vybrané modely vyrovnávání a extrapolace křivky úmrtnosti, aplikovat tyto modely na data vybraných států a zjistit, jak dobře modely vyrovnávaly křivky úmrtnosti ve státech s odlišnou životní úrovní a v různém časovém období. Všechny analyzované státy se řadí mezi vyspělé³, přesto i mezi nimi byly v úrovni úmrtnosti velké rozdíly. Od začátku tak bylo jasné, že bude nutné rozdělit tyto státy do několika relativně homogenních celků a ty pak analyzovat jednotlivě.

Všechna potřebná data (narození podle kalendářního roku a pohlaví, zemřelí a střední stav populace podle věku, kalendářního roku a pohlaví) byla čerpána z *The Human Mortality Database*. Prvotním záměrem bylo analyzovat všech 37 zemí v celém období, pro které jsou dostupná data, avšak situaci komplikovala chybějící data u Belgie, nedostatečná velikost populace ve věku 65 let a více v případě Islandu a krátké sledované období u Chile. Belgie byla zpracována až od roku 1919 místo od roku 1841, Island a Chile byly po propočítání informačního kritéria vyřazeny úplně.

Z primárních dat byly vytvořeny soubory CSV, které byly následně zpracovány softwarem DeRaS. Ze získaných dat bylo spočítáno průměrné Akaikeho informační kritérium pro každou zemi a model. Protože se výsledné hodnoty Akaikeho informačního kritéria mezi zeměmi značně lišily v absolutních hodnotách⁴, bylo třeba zvolit postup, který by eliminoval vliv absolutních hodnot a umožnil srovnání. Tímto postupem se stalo známkování modelů, které je více popsáno v podkapitole 4.1.

Nejprve byly země analyzovány podle časového období. Na základě geopolitického vývoje světa byly vybrány čtyři dekády (1920–1929, 1950–1959, 1990–1999 a 2000–2009). V jednotlivých dekádách jsme rozdělili státy do několika skupin podle období, od kterého HMD poskytovala data. V prvním období se jako nejvhodnější jevil Heligman-Pollardův model u obou pohlaví. Pro modelování úmrtnosti žen byl ve všech ostatních sledovaných obdobích

³Naděje dožití při narození byla v roce 2011 vyšší než 63 let u mužů a 75 let u žen

⁴Například zatímco hodnoty u žen z Anglie a Walesu se pohybovaly v rozpětí od -14 do 1, v případě žen z Chile hodnoty spadaly do intervalu -37– -26.

nejvhodnějším modelem Kannistö, a to nezávisle na úrovni úmrtnosti. U žen se výrazně lépe umísťovaly logistické modely. Nejhubře hodnocen byl Coale-Kiskerův a Gompertz-Makehamův model.

U mužů není volba nejvíce vyhovujícího z vybraných modelů tak jednoznačná. Ve skupinách s nižší úrovní úmrtnosti se nejvhodnějším modelem jevil Heligman-Pollard, druhý byl Kannistö. Ve skupině s relativně vysokou úrovní úmrtnosti byl nejlépe hodnocen Gompertzův model. Nejvyšší hodnoty pořadí získaly modely Coale-Kisker a Gompertz-Makeham. Z analýzy podle období vyplývá, že výběr nejvhodnějšího modelu závisí na dvou faktorech – na pohlaví a na úrovni úmrtnosti, která úzce souvisí s obdobím.

Analýza podle skupin přinesla velmi zajímavé výsledky. Bylo zjištěno, že ačkoli (až na výjimky) byly státy přiřazeny do stejné skupiny u mužů i žen, ve výběru nejvhodnějšího modelu byly značné rozdíly. Zatímco ve skupinách mužů s nižší nadějí dožití se Gompertzův model umístil poměrně dobře, byl zařazen na třetí místo, v případě žen ve skupině s nejnižší nadějí dožití se jevil v porovnání s ostatními modely jako nejméně vhodný. Ukázalo se, že muži a ženy tvoří dva samostatné celky, které při výběru nejvhodnějšího modelu není vhodné slučovat.

U čtyřech z šesti modelů (Gompertz, Gompertz-Makeham, Kannistö, Thatcher) byl u mužů naznačen lineární vztah mezi nadějí dožití při narození a známkou modelu získanou ve skupině. Obecně lze tedy usuzovat, že úroveň úmrtnosti determinuje výběr modelu. Je zde jasná diferenciací mezi skupinami států s vyšší a nižší úrovní naděje dožití při narození. Zatímco ve skupinách s nejvyšší nadějí dožití, I a II, jasně dominují logistické modely (zejména Kannistö), v dalších dvou skupinách si drží prvenství Heligman-Pollardův model. Jako poměrně dobrý se jeví i Gompertzův model. Naopak naprosto nevyhovujícím je v tomto případě Coale-Kiskerův model, který pouze jednou získal jiné než šesté místo.

Z výzkumu plyne, že pro státy, které vykazují relativně vysokou nadějí dožití v případě mužů, se nejlépe hodí Kannistův nebo Heligman-Pollardův model. Naopak státy s nižší nadějí dožití by měly modelovat křivku úmrtnosti pomocí Heligman-Pollardova nebo Gompertzova modelu.

Analýzovaná populace žen vykazuje celkově vyšší hodnoty ukazatele naděje dožití při narození a také se zdá být oproti mužům více homogenní. To může být dáno výrazně menším rozpětím naděje dožití při narození, pouze 11 let⁵. Výzkum ukazuje, že pro vyrovnávání křivky úmrtnosti žen z vyspělých zemí je nejvhodnější Kannistův model extrapolace následován Thatcherovým a Heligman-Pollardovým modelem. Státy přiřazené do skupiny IV by se měly

⁵U mužů je rozdíl mezi maximální a minimální nadějí dožití 17 let.

vyvarovat použití exponenciálních modelů – tedy Gompertzova a Gompertz-Makehamova. Gompertz-Makehamův a Coale-Kiskerův model také není vhodný pro ostatní tři skupiny.

Celkově byl nejvhodnějším modelem Kannistö, následován dalšími dvěma logistickými – Heligman-Pollardovým a Thatcherovým. Nejméně výstižné pak byly modely Coale-Kisker a Gompertz-Makeham.

Téma demografického modelování je poměrně mladé, avšak velmi široké a bude třeba učinit ještě mnoho pro nalezení vhodných postupů k výběru nejlepšího z modelů extrapolace a vyrovnávání křivky úmrtnosti, kterých je dnes celá řada. Možností rozšíření této práce by mohlo být srovnání informačních kritérií, která slouží k srovnání modelů mezi sebou, a nalezení zásadních rozdílů v hodnocení mezi těmito kritérii.

Seznam literatury

- BOZDOGAN, Hamparsum. 2000. Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology* [online]. Roč. 44, č. 1, s. 62-91 [cit. 2014-07-26]. Dostupné na WWW: <http://www.yaroslavvb.com/papers/bozdogan-akaike.pdf>.
- BURCIN, Boris, Klára TESÁRKOVÁ a Luděk ŠÍDLO. 2010. Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. *Demografie: revue pro výzkum populačního vývoje* [online]. Roč. 52, č. 2 [cit. 2014-06-30]. Dostupné na WWW: [http://www.czso.cz/csu/2012edicniplan.nsf/t/1000344DDD/\[dolar\]File/demografie_2_2010.pdf](http://www.czso.cz/csu/2012edicniplan.nsf/t/1000344DDD/[dolar]File/demografie_2_2010.pdf).
- BURCIN, Boris, Klára TESÁRKOVÁ a David KOMÁNEK. 2014. *DeRaS: software tool for modelling mortality intensities and life table construction*. Charles University in Prague, Prague. Dostupné na WWW: <http://deras.natur.cuni.cz>.
- CARRIER, Jacques F. 1992. Parametric Models for Life Table. *Transaction of Society of Actuaries*. Č. 44, s. 77-99.
- COEHLO, Edviges, Maria G. MAGALHÃES a Jorge M. BRAVO. 2007. Mortality and Longevity Projections for the Oldest-Old in Portugal. *Proceedings of the Joint Eurostat/UNECE (United Nations Statistical Commission and Statistical Office of the Economic Commission for Europe)*. Work Session on Demographic Projections, Bucharest, Romania, s. 117–132.[online]. [cit. 2014-06-30]. Dostupné na WWW: <http://epc2008.princeton.edu/papers/80105>.
- ČESKÝ STATISTICKÝ ÚŘAD. 2014. *Úmrtnostní tabulky - metodika*[online]. [cit. 2014-07-25]. Dostupné na WWW: http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_metodika.
- RABUŠIC, Ladislav. 1995. *Česká společnost stárne*. Vyd. 1. V Brně: Georgetown, s. 192. ISBN 80-901-6042-5.

- STÁTNÍ ZDRAVOTNÍ ÚSTAV. 2006. *Hodnocení zdravotního stavu: Vybrané ukazatele demografické a zdravotní statistiky* [online]. Praha [cit. 2014-08-01]. ISBN 80-7071-270-8. Dostupné z: http://www.szu.cz/uploads/documents/chzp/odborne_zpravy/OZ_05/Demo_05.pdf
- ŠUBRTOVÁ, Alena. 2006. *Dějiny populačního myšlení v českých zemích*. Praha: Acta demographica, s. 64–72. ISBN 80-239-8369-5.
- THATCHER, A. R. 1999. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* Published by: Wiley Article Stable [online]. Roč. 162, č. 1, s. 5-43 [cit. 2014-06-30]. Dostupné na WWW: <http://www.jstor.org/stable/2680465>.
- THATCHER, A. R., V. KANNISTO, V. a J. VAUPEL. 1998. The Force of Mortality at Ages 80 to 120. *Odense Monographs on Population Aging* 5. Odense University Press, [online]. [cit. 2014-06-30]. Dostupné na www: <http://www.demogr.mpg.de/Papers/Books/Monograph5/start.htm>.

Seznam použitých datových zdrojů

HUMAN MORTALITY DATABASE. 2014. *Human Mortality Database* [online]. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). [cit. 2014-03-24]. Dostupné na WWW: <http://www.mortality.org/>.

POPULATION REFERENCE BUREAU. 2011. *World Population Data Sheet* [online]. [cit. 2014-07-26]. Dostupné na WWW: http://www.prb.org/pdf11/2011population-data-sheet_eng.pdf.

POPULATION REFERENCE BUREAU. 2012. *World Population Data Sheet* [online]. [cit. 2014-07-26]. Dostupné na WWW: http://www.prb.org/pdf12/2012-population-data-sheet_eng.pdf.

WORLD POPULATION PROSPECTS: The 2012 Revision [online]. 2012. Poslední revize 9. 1. 2014 [cit. 2014-05-20]. Dostupné na WWW: <http://esa.un.org/unpd/wpp/Documentation/glossary.htm>.