

**Univerzita Karlova v Praze**  
**Přírodovědecká fakulta**

Studijní program: Demografie

Studijní obor: Demografie



**Bc. Gabriela Myšáková**

**Metody analýzy sezónnosti demografických jevů**

**Methods of analysis of seasonality in demography**

Typ závěrečné práce

**Diplomová**

Vedoucí závěrečné práce: Mgr. Klára Hulíková Tesárková

Praha, 2011

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně 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 Praze, 1. 5. 2011

Podpis

**Poděkování**

Děkuji Mgr. Kláře Hulíkové Tesárkové za vstřícnost, odborné vedení a užitečné připomínky k mojí diplomové práci.

Také bych chtěla poděkovat své rodině a nejbližším za vytvoření vhodných podmínek pro práci na diplomové práci a věcné a lingvistické připomínky k mé práci.

## **Metody analýzy sezónnosti demografických jevů**

### **Abstrakt**

Předkládaná práce představuje statistické metody, pomocí nichž je možné analyzovat sezónnost v časových řadách. Podrobně jsou popsány tři metody, dekompozice časové řady, metoda X12-ARIMA a kointegrace časových řad. Krátce jsou představeny i další metody vhodné pro obdobné analýzy. Tři hlavní metody jsou následně aplikovány na měsíční demografická data za Česko a vybrané evropské státy, o porodnosti, sňatečnosti a úmrtnosti. Sezónnost byla odhalena u všech třech demografických jevů, a aplikací metody dekompozice časové řady a metody X12-ARIMA byly časové řady rozloženy na jednotlivé složky, u kterých byl pomocí grafického znázornění a slovní interpretace popsán jejich vývoj. Časové řady sňatečnosti a úmrtnosti evropských zemí byly také podrobeny shlukové analýze za účelem odhalení podobnosti a odlišnosti jednotlivých států. Všechny metody byly na data aplikovány pomocí odpovídajících procedur v prostředí statistického softwaru SAS.

**Klíčová slova:** dekompozice časové řady, X12-ARIMA, kointegrace časových řad, sezónnost, porodnost, sňatečnost, úmrtnost

## **Methods of analysis of seasonality in demography**

### **Abstract**

The thesis presents statistical methods suited for analysis of seasonality in time-series. Three statistical methods have been thoroughly described, namely the time-series decomposition, the X12-ARIMA method and the cointegration of time-series. Further methods applicable for similar analysis have been briefly discussed as well. Three main methods have been subsequently applied to monthly demographic data for the Czech Republic and chosen European countries by natality, nuptiality and mortality. Seasonality has been discovered in all three demographic events and by using the time-series decomposition and the X12-ARIMA method for time-series lay out to separate units, and those progressions have been track by graphic and verbal interpretation. Cluster analysis has been applied to European countries nuptiality and mortality time-series in order to reveal similarities and dissimilarities among particular countries. All the methods were used onto data set by using appropriate procedures in statistical software SAS.

**Keywords:** time-series decomposition, the X12-ARIMA, cointegration of time-series, seasonality, natality, nuptiality, mortality

## Obsah

<b>Přehled použitých zkratk</b> .....	<b>7</b>
<b>Seznam tabulek</b> .....	<b>8</b>
<b>Seznam obrázků</b> .....	<b>9</b>
<b>1 Úvod</b> .....	<b>11</b>
1.1 Téma práce .....	11
1.2 Cíl práce .....	12
1.3 Demografické procesy .....	12
1.4 Diskuze s literaturou .....	13
<b>2 Zdroje dat</b> .....	<b>21</b>
2.1 Data Česko .....	21
2.2 Data ostatních evropských států.....	22
<b>3 Metodologie</b> .....	<b>24</b>
3.1 Dekompozice časové řady.....	29
3.1.1 Výpočet v programu SAS .....	34
3.2 Metoda X12–ARIMA .....	35
3.2.1 Výpočet v programu SAS .....	37
3.3 Kointegrace časových řad .....	41
3.3.1 Výpočet v programu SAS .....	43
3.4 Další metody pro analýzu sezónnosti.....	43
3.4.1 Grafická analýza.....	43
3.4.2 Model SARIMA.....	44
3.4.3 Lineární dynamické modely.....	44
3.4.4 Spektrální analýza .....	44
3.5 Shluková analýza .....	45
3.5.1 Výpočet v programu SAS .....	46
3.6 Prezentace výsledků .....	47
<b>4 Porodnost</b> .....	<b>48</b>
4.1 Česko.....	48
4.2 Evropské státy .....	50

<b>5</b>	<b>Sňatečnost .....</b>	<b>54</b>
5.1	Česko.....	54
5.2	Evropské státy .....	57
5.2.1	Období let 1990–2009.....	58
5.2.2	Období let 1996–2009.....	61
5.2.3	Období let 2001–2009.....	63
5.2.4	Shluková analýza .....	65
<b>6</b>	<b>Úmrtnost .....</b>	<b>67</b>
6.1	Česko.....	67
6.1.1	Podle pohlaví.....	67
6.1.2	Podle věku.....	69
6.1.3	Podle příčin úmrtí.....	73
6.2	Evropské státy .....	75
<b>7</b>	<b>Závěr .....</b>	<b>78</b>
<b>8</b>	<b>Seznam použité literatury.....</b>	<b>80</b>
<b>9</b>	<b>Přílohy .....</b>	<b>85</b>

## Přehled použitých zkratk

ACF	Autokorelační funkce
CSBL	Central Statistical Bureau of Latvia
ČSÚ	Český statistický úřad
EG	Enterprise Guide
EU	Evropská unie
INE	National Statistics Institute
NISES	National Institute of Statistics and Economic Studies
PACF	Parciální autokorelační funkce
SD	Statistics Denmark
SE	Statistics Estonia
SF	Statistics Finland
SI	Statistics Iceland
SL	Statistics Lithuania
SN	Statistics Norway
SNL	Statistics Netherlands
SORS	Statistical Office of the Republic of Slovenia
SOSR	Statistical Office of the Slovak Republic
TSI	Turkish Statistical Institute

## Seznam tabulek

Tab. 1 – Seznam analyzovaných událostí a jejich časového vymezení, Česko .....	22
Tab. 2 – Vybrané země Evropy, název a webová adresa příslušného statistického úřadu, ze kterého byla získána data použitá v této práci.....	22



## Seznam obrázků

Obr. 1 – Schéma metod pro analýzu časových řad, které jsou představeny v této práci .....	25
Obr. 2 – Kondratěvovy dlouhé hospodářské cykly .....	31
Obr. 3 – Procesní tok dekompozice časové řady, SAS Enterprise Guide .....	35
Obr. 4 – Ukázka programovacího kódu pro proceduru X12, SAS Base .....	38
Obr. 5 – Schéma kombinovaného testu přítomnosti identifikovatelné sezónnosti, SAS Base .....	40
Obr. 6 – Ukázka programovacího kódu pro proceduru VARMAX, SAS Base .....	43
Obr. 7 – Ukázka programovacího kódu pro proceduru CLUSTER a proceduru TREE, SAS Base .....	47
Obr. 8 – Sezónně očištěná časová řada, trendová složka, živě narození, Česko, 1983–2009..	49
Obr. 9 – Hodnoty úhrnné plodnosti, Česko, 1983–2009 .....	49
Obr. 10 – Proměnlivé sezónní indexy, živě narození, Česko, 1983–2009 .....	50
Obr. 11 – Sezónně očištěná časová řada, trendová složka, živě narození, Estonsko, 1975–2009 .....	51
Obr. 12 – Sezónně očištěná časová řada, trendová složka, živě narození, Norsko, 1975–2009	51
Obr. 13 – Sezónní indexy, živě narození, Island, Norsko, Španělsko, 1975–2009 .....	52
Obr. 14 – Proměnlivé sezónní indexy, živě narození, Slovinsko, 1975–2009 .....	53
Obr. 15 – Sezónně očištěná časová řada, trendová složka, sňatky, Česko, 1983–2009 .....	55
Obr. 16 – Náhodná složka, sňatky, Česko, 1983–2009 .....	55
Obr. 17 – Sezónní indexy, sňatky, Česko, 1983–2009 .....	56
Obr. 18 – Proměnlivé sezónní indexy, sňatky, Česko, 1983–2009 .....	57
Obr. 19 – Sezónně očištěná časová řada, trendová složka, sňatky, Estonsko, 1990–2009 .....	58
Obr. 20 – Sezónně očištěná časová řada, trendová složka, sňatky, Dánsko, 1990–2009 .....	59
Obr. 21 – Sezónní indexy, sňatky, Lucembursko, Slovinsko, 1990–2009 .....	60
Obr. 22 – Sezónní indexy, sňatky, Dánsko, Estonsko, Španělsko, 1990–2009 .....	60
Obr. 23 – Proměnlivé sezónní indexy, sňatky, Estonsko, 1990–2009 .....	61
Obr. 24 – Sezónní indexy, sňatky, Lotyšsko, Nizozemsko, Norsko, 1996–2009 .....	62
Obr. 25 – Proměnlivé sezónní indexy, sňatky, Lotyšsko, Nizozemsko, 1996–2009 .....	62
Obr. 26 – Sezónní indexy, sňatky, Finsko, Litva, Slovensko, Turecko, 2001–2009 .....	63
Obr. 27 – Proměnlivé sezónní indexy, sňatky, Finsko, Turecko, 2001–2009 .....	64
Obr. 28 – Náhodná složka, sňatky, Finsko, 2001–2009 .....	64
Obr. 29 – Dendogram vybraných evropských zemí, sezónní indexy, sňatky, 2001–2009 .....	65
Obr. 30 – Dendogram vybraných evropských zemí, proměnlivé sezónní indexy, sňatky, 2001–2009 .....	66

Obr. 31 – Dendogram vybraných evropských zemí, náhodná složka, sňatky, 2001–2009 .....	66
Obr. 32 – Sezónně očištěná časová řada, trendová složka, zemřelí, muži, Česko, 1983–2009 .	68
Obr. 33 – Proměnlivé sezónní indexy, zemřelí, ženy, muži, Česko, 1983–2009 .....	68
Obr. 34 – Sezónně očištěná časová řada, trendová složka, zemřelí do 1 roku, Česko, 1983–2009 .....	70
Obr. 35 – Kvocient kojenecké úmrtnosti, Česko, 1983–2009 .....	70
Obr. 36 – Sezónně očištěná časová řada, trendová složka, zemřelí, 50–69 let, 70+, Česko, 1995–2009 .....	71
Obr. 37 – Proměnlivé sezónní indexy, zemřelí, 0–14let, 15–29 let, Česko, 1995–2009.....	71
Obr. 38 – Proměnlivé sezónní indexy, zemřelí, 30–49 let, 50–69 let, 70+, Česko, 1995–2009	72
Obr. 39 – Sezónně očištěná časová řada, trendová složka, zemřelí, nemoci dýchací soustavy, Česko, 1995–2009 .....	73
Obr. 40 – Sezónní indexy, zemřelí, nemoci dýchací soustavy, nemoci oběhové soustavy, Česko, 1995–2009 .....	74
Obr. 41 – Proměnlivé sezónní indexy, zemřelí, nemoci dýchací soustavy, nemoci oběhové soustavy, Česko, 1995–2009 .....	75
Obr. 42 – Dendogram vybraných evropských zemí, sezónní indexy, zemřelí, 2000–2007.....	76
Obr. 43 – Dendogram vybraných evropských zemí, proměnlivé sezónní indexy, zemřelí, 2000–2007 .....	76
Obr. 44 – Dendogram vybraných evropských zemí, náhodná složka, zemřelí, 2000–2007 .....	77
Obr. 45 – Sezónní indexy, zemřelí, Dánsko, Lotyšsko, Španělsko, 2000–2007 .....	77

# 1 Úvod

## 1.1 Téma práce

Pozornost v oblasti sezónnosti demografických jevů byla doposud v českém prostředí zaměřena především na sňatečnost (Konečná, 1997; Fialová, 1995; Kačerová, 2004) a v mnohem menším rozsahu také na úmrtnost (Tesárková, 2008; Myšáková, 2009). Metod použitých k odhalení sezónnosti bylo použito několik, od jednoduchých, např. podíl jednotlivých měsíců v roce na celkový počet událostí, až k pokročilejším metodám v podobě dekompozice časové řady. Analýza časových řad, která je statistickým nástrojem k rozpoznání sezónnosti, však nabízí i jiné, často detailnější a propracovanější metody a možnost využití k výpočtům moderního statistického softwaru, ne vždy je jich však uživatel nucen použít. Pro obecnou představu o vývoji vybraného ukazatele během roku lze využít již zmíněných jednoduchých metod. Tyto metody mohou být prvním krokem pro další a podrobnější analyzování časové řady nebo také vodítkem k zjištění, že distribuce událostí během roku nevykazuje žádné výkyvy, které by bylo vhodné hlouběji analyzovat.

Poznatky o sezónním chování a rozložení demografických událostí během roku je možné využít v mnoha oblastech, které spolu často nemusí ani přímo souviset. V oblasti vědy a výzkumu bývá snaha o co nejpodrobnější statisticky podložené zpracování získaných dat z důvodu přesné a bezchybné interpretace získaných výsledků. Informace o sezónním rozložení demografických jevů je však využíváno i v komerční sféře. Zdravotní pojišťovny na základě informací o nemocnosti a úmrtnosti určitých rizikových skupin mohou cíleně zaměřit preventivní péči, například distribucí vitamínových balíčků nebo brožur s radami o ochraně před konkrétní chorobou. S nerovnoměrným rozložením uzavíraných sňatků během roku musí ve svém oboru kalkulovat také firmy a podnikatelé, kteří působí na svatebním trhu.

V odborných kruzích má podrobná analýza časových řad s důrazem na hledání sezónních pravidelností za cíl především odhalení vnitřních zákonitostí vývoje a možnost odhadu ukazatele do budoucna (Tesárková, 2008).

Téma analýzy sezónnosti demografických jevů jsem si pro svoji diplomovou práci zvolila z důvodu prohlubujícího se zájmu o analýzu časových řad v demografii a také možnosti následně využít získaných poznatků v praxi.

## 1.2 Cíl práce

Předkládaný text je rozdělen na dvě stěžejní části: část metodologickou, která podrobně popisuje jednotlivé metody a vhodnost aplikace každé z nich na určité typy časových řad, a část analytickou, ve které jsou dříve představené metody a postupy analýzy časových řad aplikovány na konkrétní demografická data evropských států.

Za cíl si tato práce klade představit vybrané pokročilejší metody analýzy časových řad. V teoretické části jsou jednotlivé metody nejprve podrobně definovány a je nastíněn jejich vývoj. Práce si však neklade za cíl podrobný výpis rovnic a výpočtů, které jsou v těchto metodách využívány, na tyto přesné matematické vzorce však práce odkazuje. Použití představených metod je dále předvedeno na reálných demografických datech, která zachycují vývoj porodnosti, sňatečnosti a úmrtnosti v posledních dekadách. Názorné ukázky postupu jednotlivých výpočtů jsou prováděny pomocí statistického softwaru SAS za využití příslušných procedur a jsou doplněny informacemi o některých možnostech vlastních nastavení v provedených výpočtech. Na samotné analýzy časových řad pak navazuje představení dalších analýz, které mohou dát uživateli ucelený přehled o získaných výsledcích a o možnostech dalšího zkoumání.

## 1.3 Demografické procesy

Každý individuální jedinec prochází během svého života demografickými událostmi, které pro něj znamenají ukončení jednoho stavu a začátek stavu jiného. Tyto události lze rozdělit na nevyhnutelné a vyhnutelné. Nevyhnutelnými událostmi rozumíme narození a úmrtí, vyhnutelnými pak potrat, sňatek, rozvod, stěhování a nemoc. Demografie však nestuduje tyto individuální události v životě jedince, jako například medicína, ale zaměřuje se na ně jako na hromadné jevy a ty následně podrobuje analýze. Mezi hromadné jevy řadíme porodnost, potratovost, sňatečnost, rozvodovost, úmrtnost, migraci a někdy také zahrnujeme do této skupiny demografických procesů nemocnost. Procesem tak rozumíme prožívání změny stavu jedince, např. ze stavu svobodný do stavu ženatý (Demografický informační portál, 2011).

Výběr demografických procesů vstupujících do analýzy v této práci není náhodný, ale je založen na předpokladech o přítomnosti sezónnosti získaných na základě studia odborné literatury. Do analýzy tak vstupují tři demografické procesy, porodnost, sňatečnost a úmrtnost, která je pak dále dělena dle pohlaví, věku zemřelého, včetně kojenecké úmrtnosti, a příčiny úmrtí. Okolnosti a aspekty ovlivňující sezónní rozložení těchto tří konkrétních demografických jevů jsou podrobně popsány dále v textu a jsou opřeny o již provedené a publikované výzkumy. Mezi analyzovanými procesy není zařazena rozvodovost, neboť statistika uvádí v datových záznamech pouze měsíc, ve kterém bylo manželství rozvedeno. Pro analýzu sezónnosti by však bylo zapotřebí informací o měsíci, kdy byla žádost o rozvod podána, nebo kdy se manželé pro ukončení manželství rozhodli, neboť období, kdy bylo manželství oficiálně rozvedeno je ovlivněno především vyřizováním formálních záležitostí s rozvodem spojených a délkou

samotného rozvodového řízení. V této práci do analýzy nevstupuje ani proces migrace, a to z důvodu jakési nespolehlivosti, resp. nepřesnosti, dat vykazovaných statistikou. Tyto nesrovnalosti pramení ze skutečnosti, že obyvatelé, kteří odjíždějí do zahraničí na delší dobu, nebo se natrvalo stěhují do zahraničí, se často neodhlásí z pobytu na našem území, a nejsou tak v informacích o mezinárodní migraci vůbec evidováni. V případě vnitřního stěhování, tedy změny bydliště v rámci republiky, lidé někdy ponechávají registraci svého trvalého bydliště na jednom místě, i když např. během studia a svého mládí změni své bydliště několikrát. I v případě, že si své trvalé bydliště změni, nemusí tato změna vždy odpovídat i měsíci realizovanému stěhování. Ačkoliv tedy data o stěhování lze získat i po měsících, zveřejněné počty odstěhovaných a přistěhovaných nemusí vždy odpovídat skutečnosti, což je pro analýzu sezónnosti nepřijatelné. Nalezení sezónních pravidelností v časové řadě a jejich následná interpretace by tak mohla být zavádějící.

V případě některých z uvedených demografických jevů je očekávání sezónnosti v průběhu sledovaného období neoddiskutovatelné, nalezení sezónních pravidelností u některých ostatních procesů může být naopak až překvapivé.

## 1.4 Diskuze s literaturou

### *Sezónnost porodnosti*

Články a analýzy, zabývající se rozložením počtu narozených dětí, nebo rozložením předpokládaného počtů, podle měsíců v roce, lze rozdělit do dvou skupin podle konkrétní sféry zájmu. Zatímco v české literatuře je možné objevit jen práce zabývající se narozenými dětmi obecně, bez rozdílů, zahraniční literatura nabízí nepřehledné množství analýz s určitým zaměřením, nejčastěji na souvislost mezi měsícem narození a diagnostikování určité choroby v pozdějším věku.

S nejstaršími daty o sezónním rozložení narozených nebo o rozložení počtů na českém území, která mapují období 2. poloviny 17. století i 18. století, se můžeme setkat v publikaci „Dějiny obyvatelstva českých zemí“ (Fialová, 1996). Příložené grafy znázorňují průběh počtu počtů po jednotlivých měsících a v porovnání s křivkou počtu zemřelých ilustrují souvislost mezi úmrtnostními a reprodukčními poměry. V měsících s vysokým počtem zemřelých je hodnota uskutečněných počtů nižší, neboť se často jednalo o měsíce zasažené epidemiemi nebo hladomorem, kdy lidé měli jiné starosti než rozšiřovat rodinu (tamtéž, str. 116 a 177). V grafu sezónních indexů na str. 124, který zachycuje celou 2. polovinu 18. století, můžeme pozorovat lehce nadprůměrné hodnoty sezónních indexů (okolo hodnoty 1,2) v únoru až dubnu, a podprůměrné hodnoty (okolo 0,8) v červenci, říjnu a listopadu (Fialová, 1996). Český statistický úřad vydal jako jednu ze svých publikací o obyvatelstvu, s názvem „Porodnost a plodnost 2001 – 2005“, krátkou analýzu o sezónnosti počtu živě narozených dětí. V rozporu je název s obsahem zprávy, neboť analýze jsou podrobena data z období 1993–2005. Po standardizaci dat na stejnou délku měsíce jsou vypočteny měsíční sezónní indexy za dvě vybraná pětiletá období: 1993–1997 a 2001–2005. Zatímco pro první období je patrná mírná

převaha jarních a letních měsíců, v druhém období se situace vyrovnává a i přes podprůměrné hodnoty zimních měsíců je rozložení živě narozených během roku rovnoměrnější (ČSÚ, 2006). Zatím poslední prací zabývající se sezónností porodnosti, je bakalářská práce s názvem „Statistická analýza vybraných demografických ukazatelů v České republice“, jejíž autorka Michaela Gabrhelíková si pro analýzu zvolila data za období 1992–2008. Po prvním kroku zkoumání, grafické analýze dat, omezuje interval na období let 1993–2008, neboť v tomto úseku se odhadnutý trend jeví jako kvadratická funkce, a tuto hypotézu pak autorka dále zkoumá. Na základě provedené analýzy autorka došla k závěru, že nejvíce dětí se rodí v měsíci červenci, a to pravděpodobně z důvodu „dlouhých podzimních večerů“ v době početí a dlouhodobě nejnižší úroveň porodnosti je vykazována v prosinci (Gabrhelíková, 2010). Autorka této bakalářské práce není z demografického prostředí, na demografických datech pouze ilustruje využití vhodných metod analýzy časových řad, proto jsou některé příčiny sezónního chování popsány velmi laicky.

V zahraniční literatuře se také objevují analýzy sezónnosti narozených nebo početí, ale v mnohem větší míře je zde hledána odpověď na otázku, zda narození v určitém měsíci v roce může ovlivnit další, především zdravotní, vývoj jedince. Důvody a faktory, které ovlivňují rození v lidské populaci, se zabývá převážně teoretický článek autorů z katedry antropologie z Harvardské Univerzity nesoucí název „Human birth seasonality“ (Ellison, 2005). První skupinou faktorů, které jsou uvažovány, jsou sociální faktory zohledňující především vliv náboženských a světských svátků (prázdniny, doba dovolených) na intenzitu vzniku nových těhotenství. Jako příklad uvádějí vrchol v počtu početí v době vánočních svátků a nového roku ve Spojených státech amerických a období letních dovolených v srpnu ve Francii. Jako sociální faktory dále vnímají změnu v životním stylu a poukazují na skutečnost, že zatímco u předchozích generací byl život rodiny během roku ovlivněn zemědělským způsobem obživy, a odvíjel se tak v závislosti na ročním období, na sezónních pracích a zásobách potravy, dnešní populace není těmito faktory už limitována v takové míře. Druhou skupinou faktorů tvoří klimatické podmínky, s důrazem na teplotu ovzduší a fotoperiodicitu (střídání dne a noci, změna délky působení světla, která ovlivňuje fyziologické reakce organismů periodickými změnami dobou osvětlené a tmavé části dne). Jsou tak ovlivňovány hlavně rytmy pohybové a sexuální aktivity nazývané biorytmy (Špérová, 2008). Vliv na počet početí je přisuzován krátkým dnům, které často i ve spojení se špatným a deštivým počasím nutí partnerské páry trávit více času doma, v soukromí. Hledání souvislostí mezi úrovní reprodukce a klimatickými podmínkami je nahlíženo také z biologického hlediska, kdy autoři popisují závislost biologických funkcí (např. tvorbu spermií nebo tvorbu hormonu melatoninu) na jednotlivých fázích roku. Jako třetí jsou uvedeny energetické faktory ovlivňující ženskou plodnost. Váha ženy a mnoho dalších fyziologicko-biologických podmínek je spojováno s obdobím v roce podle toho, zda má žena dostatečný přísun potravy, vitamínů a dalších důležitých látek ovlivňujících schopnost reprodukce. Tuto skupinu faktorů lze dnes aplikovat spíše na rozvojové země, s problémy zásobování potravinami nebo využít v analýze historických populací (Ellison, 2005). O rozsáhlou a podrobnou datovou základnu o denním počtu narozených adolescentním dívkám

(15–19 let) 1964–2000, které jsou v období 1994–2000 dále rozdělené ve třech etnických skupinách (bělošky, Hispánky, Afroameričanky) a podle rodinného stavu (svobodná a vdaná), se opírají autoři v článku „Seasonality od Birth and Conception to Teenagers in Texas“ z časopisu *Social Biology* z roku 2003 (Scafetta, 2003). Z výsledků lze shrnout několik závěrů: během sledovaného období se v americkém státě Texas zvýšil průměrný počet dětí narozených během jednoho dne mladým matkám z hodnoty 144 narozených v roce 1964 na 151 v roce 2000, nejvýraznější rostoucí trend vykazují mladé neprovdané Hispánky, zatímco postupný pokles pozorujeme u neprovdaných Afroameričanek. Z pohledu sezónního rozložení je zajímavé sledovat křivku počtu početí, která vykazuje nejnižší hodnotu v červenci a srpnu, což autoři připisují prázdninám a tedy omezenému styku se spolužáky, kteří jsou nejčastěji rodiče nově počatého potomka, a průběh počtu početí vrcholí v době vánočních prázdnin, od poloviny prosince do začátku ledna. Křivka počtu narozených pak přirozeně kopíruje informaci o počtu početí, s přibližně devítiměsíčním časovým posunem. Návaznost na průběh školního roku je co do počtu početí přibližně dvojnásobný pro svobodné mladé matky než pro skupinu vdaných, z etnického hlediska je pak nejvýraznější reakce na letní a vánoční prázdniny pro hispánské dívky oproti zbývajícím etnickým skupinám, a to jak ve skupině vdaných, tak ve větší míře i ve skupině svobodných. Lze tedy říci, že nejvyšší sezónnost v reprodukčním chování mladých žen vykazují neprovdané Hispánky (Scafetta, 2003). Analýzu celé kanadské populace za dlouhé časové období přibližuje článek „The Seasonality of Birth in Canada and the Provinces, 1881–1989: Theory and Analysis“ uveřejněný v časopise *Canadian Studies in Population* (Trovato, 1993). Aplikací klouzavých průměrů a regresního modelu obsahujícího vysvětlující proměnné v čase (viz např. Šťastná, 2010) na měsíční data o počtu narozených získali autoři jasnou představu o průběhu porodnosti ze sezónního hlediska. Vysokou úroveň porodnosti v časovém úseku 1926–1989 vykazuje období března až července s vrcholem v měsíci květnu a nízké hodnoty sezónních indexů zaznamenané především v období září až únor. Srpen dosáhl hodnoty sezónního indexu 1 a vyznačoval se tedy jako měsíc s průměrným počtem narozených (Trovato, 1993).

Vědecké články, které jsou v souvislosti se sezónností porodnosti zaměřené na nemoci, se zabývají jak tělesnou schránkou jedince, tak i duševními poruchami. Tělesnou hmotnost ovlivněnou měsícem porodu dětí narozených v rozmezí 37. a 41. týdne těhotenství v letech 1987–2007 analyzují autoři článku „Seasonality of Birth Weight in Chile: Environmental and Socioeconomic Factors“, který vyšel v roce 2010 v časopise *Annals of Epidemiology* (Torche, 2010). Z publikovaných výsledků je zřejmé, že nejvyšší hmotnost při narození mají děti narozené v měsících únor, březen, říjen, listopad a nejnižší porodní váhu mají novorozenci, kteří přišli na svět v červenci. Pro šestnáct evropských zemí byl proveden výzkum, zda souvisí měsíc narození s výskytem cukrovky 1. typu (diabetes mellitus 1. typu označovaný také jako inzulin-dependentní diabetes mellitus (IDDM)) u dětí ve věku 0–14 let, založený na datech z let 1974–1994, jehož výsledky byly publikovány v článku „Seasonality of birth in children with diabetes in Europe: multicentre cohort study“ zveřejněném v časopise *British Medical Journal* v roce 1999 (Rothwell, 1999). Pouze pro data za Velkou Británií byly odhaleny

statisticky významné rozdíly mezi celou populací a populací trpící touto chorobou projevující se poruchou metabolismu sacharidů. Více konkrétní výsledky o onemocnění cukrovkou v návaznosti na měsíc narození přináší zjištění o vyšší incidenci chlapců z Nizozemska narozených v dubnu, květnu a listopadu oproti celé populaci. Podobnou souvislost však nebylo možné dokázat na datech za dívky ani na datech za obě pohlaví dohromady, jak vyplývá z článku „Seasonality of birth in patients with childhood diabetes in The Netherlands“ v časopise *Diabetes Care* (Jongbloet, 1998). Podobnou analýzu, avšak zaměřenou na onemocnění celiakií (celoživotní onemocnění způsobené nesnášenlivostí lepku obsaženého v obilí, u citlivých osob vede k postižení sliznice tenkého střeva (Kohout, 2006)), přináší článek „Seasonality of birth month of children with celiac disease differs from that in the general population and between sexes and is linked to family history and environmental factors“ publikovaný v časopise *Journal of Pediatric Gastroenterology & Nutrition* (Lewy, 2009). Stejnou nemocí se zabývá i článek „Children born in the summer have increased risk for coeliac disease“, který vyšel v roce 2003 v časopise *Journal of Epidemiology & Community Health* (Ivarsson, 2001). Z duševních poruch zkoumaných v tomto kontextu, je možné vybrat schizofrenii (onemocnění vyznačující se zkresleným vnímání reality, progresivním rozpadem myšlení, nezvyklým chováním až netečností (Hrdlička, 2005)) a analýzu představenou prostřednictvím článku „Seasonality of Birth in Schizophrenia in Taiwan“ v časopise *Schizophrenia Bulletin* (Tam, 1995). Z výsledků vyplývá, že statisticky významně vyšší je výskyt této duševní nemoci u osob, které se narodily v chladných zimních měsících, tedy v období od listopadu do února.

### ***Sezónnost sňatečnosti***

O sezónním rozložení sňatků bylo jak v české, tak světové literatuře uvedeno mnoho odborných článků a publikací. Skrze měsíc, kdy došlo k uzavření manželství lze vysledovat typické rysy chování dané populace, neboť sňatek je událostí vyhnutelnou a je tedy možné si ji naplánovat, což je zásadní rozdíl oproti další důležité demografické události v životě jedince, úmrtí.

Na toto téma jsou v českém prostředí publikovány příspěvky nejčastěji v časopise *Demografie*, na internetovém portálu *demografie.info* nebo Českým statistickým úřadem. O sňatečnosti historických populací na českém území přináší množství informací publikace *Dějiny obyvatelstva českých zemí*, kde jsou pomocí měsíčních počtů událostí i pomocí vypočtených sezónních indexů přiblíženy zvyklosti v uzavírání sňatků především v průběhu 18. století. Grafické znázornění absolutních i indexovaných hodnot deklaruje koncentraci sňatků do měsíců listopad, leden a únor (Fialová, 1996). Toto rozložení kopíruje roční cyklus a život rodin, kdy na podzim bylo dost zásob jídla po ukončené zemědělské sezóně, v prosinci nebyl takový počet svateb z důvodu oslav Vánoc a po Novém roce bylo stále dostatek zásob na uspořádání svatby. Na jaře už se tolik sňatků neuzavíralo, neboť bylo období pohyblivých Velikonoc a jim předcházejícího půstu a po Velikonocích opět začínala zemědělská sezóna (Fialová, 1995). Zmínku o sezónním rozložení sňatků najdeme také v publikaci „Základy demografie“ (Pavlík, 1986), kde autoři poukazují na souvislost sezónnosti sňatečnosti



s tradicemi a sociálními podmínkami, ve kterých obyvatelstvo žije. Všechny články za celou historii časopisu *Demografie* shrnuje příspěvek Kláry Tesárkové a Evy Karousové „Vývoj sezónnosti sňatečnosti za 50 let časopisu *Demografie*“, který vyšel v rámci Sborníku příspěvků XXXVIII. konference České demografické společnosti jako příloha prvního čísla časopisu *Demografie* v roce 2009. Článek rekapituluje čtyři předcházející příspěvky na téma sezónnosti sňatečnosti a představuje metodu dekompozice časové řady, kterou aplikuje na česká data, a to na celostátní a krajské úrovni (Tesárková, 2009a). Z již zmíněných čtyř článků vyšel první v roce 1977, nesl název „Sezónnost sňatečnosti v ČSSR“ (Konečná, 1977) a mapuje vývoj sňatečnosti v časovém období 1960 až 1973 podle socioekonomické skupiny snoubenců, odděleně pro Českou a Slovenskou republiku (Tesárková, 2009a). Druhým článkem je příspěvek z roku 1990 „Vývoj sezónnosti sňatečnosti v Československu po druhé světové válce a současná situace v mezinárodním srovnání“. Jeho autoři Michal Roučka a Radka Skočdoplová analyzovali sezónní rozložení sňatků v období let 1945–1985 a výsledky sňatkového chování v Československu porovnali se sedmi evropskými zeměmi (Roučka, 1990; Tesárková, 2009a). V roce 1995 vyšel příspěvek „Sezónnost demografických událostí v českých zemích v 17. až 20. století“ (Fialová, 1995), který analyzuje sezónní rozložení nejen sňatků na našem území (Tesárková, 2009a). Článek s názvem „Mezinárodní srovnání sezónnosti sňatečnosti“ (Kačerová, 2004) vyšel v roce 2004 a ve svém obsahu se nezabývá jen samotným sezónním rozložením sňatků, ale také vlivem tradic a svátků na sňatkové chování (Tesárková, 2009a). Tímto tématem se zabývá i analýza zveřejněná na Demografickém informačním portálu *demografie.info* v roce 2005, a jejímž autorem je Daniel Hůle a byla uvedena pod názvem „Pověra v srdci globálního ateismu“ (Hůle, 2005). Tento autor považuje vedle klimatických (teplota vzduchu, úhrn srážek) a historických podmínek (sezóna úrody) za dominantní faktor, který ovlivňuje rozhodování při výběru data sňatku, pověru. A i když připouští, že dnešní třicátníci asi těžko znají některé pověry spojené s manželstvím, předpokládá, že se vždy najde v příbuzenstvu či blízkém okolí někdo, kdo takovou pověru pronese. Za nejvýznamnější, což dokazují česká data už po desítky let, je třeba považovat pověru „svatba v máji, do roka máry“, která si i dnes najde mezi snoubenci příznivce. Nízký počet květnových svateb se pak následně projeví v kumulaci svateb do předcházející dubna nebo následující června (Hůle, 2005). Eva Kačerová se sezónním vlivům na sňatečnost věnovala i ve své bakalářské práci s názvem „Sezónnost sňatečnosti“ z roku 2001 (Kačerová, 2001). S využitím standardizace na stejný počet dní v měsíci a s výpočtem měsíčních sezónních indexů se můžeme setkat v analýze zveřejňované Českým statistickým úřadem každoročně v rámci publikace „Vývoj obyvatelstva České republiky“, a to v oddíle věnovanému sňatečnosti (ČSÚ, 2010).

Podobně jako v Česku, vznikají i v zahraničí práce zabývající se výběrem vhodného období, a v poslední době i konkrétního „magického“ data, pro vstup do manželského svazku a hledání odlišností mezi jednotlivými populacemi. Z evropského kontinentu lze jmenovat článek „Biodemographic study of a central Apennine area (Italy) in the 19th and 20th centuries: marriage seasonality and reproductive isolation“ z časopisu *Journal of Biosocial Science*, který vystihuje rozdíl v 19. století mezi venkovským obyvatelstvem, které z důvodu

letních venkovních prací a častou migrací za prací, koncentrovalo uzavírání sňatků do zimních měsíců, a ostatním, městským obyvatelstvem, které preferovalo svatby v létě (Danubio, 2001). Pro Itálii existuje také analýza zachycující vývoj sňatečnosti během čtyř po sobě jdoucích století, od 16. až po 19. V článku s názvem „Seasonality of Marriages and Ecological Contexts in Rural Communities of Central-Southern Italy (Abruzzo), 1550–1871“ z časopisu *Collegium Antropologicum* jsou pomocí sezónních indexů porovnávána čtyři dlouhá období a publikovány překvapivé výsledky, např. posun velmi výrazné převahy měsíce ledna v 16. století (hodnota sezónního indexu nad 2) k podprůměrným hodnotám pro tento měsíc v 19. století (Coppa, 2001). Návaznost sezónnosti porodnosti na sezónnost sňatečnosti na relativně malé populaci Malty hledají autoři článku „Seasonality of Births is associated with Seasonality of Marriages in Malta“, který vyšel v časopise *Journal of Biosocial Science* (Grech, 2003). Odhalení přenosu zvyklostí v sezónním chování při uzavírání sňatků v rámci kolonizování amerického kontinentu je výsledkem bádání autora článku „Marriage Seasonal in Early America“, který vyšel v roce 1987 v časopise *Journal of Interdisciplinary History* (Wells, 1987).

### **Sezónnost úmrtnosti**

U úmrtí, které je ve většině případů neovlivnitelnou demografickou událostí, je analýza sezónnosti spojována především s hledáním faktorů, které mají na sezónní průběh počtu zemřelých vliv. Hlavním z těchto faktorů je vliv střídání ročních období a tedy vliv počasí. Neboť je proces úmrtnosti úzce spojen s medicínou, objevují se články na toto téma častěji v lékařských časopisech než v demografických, a to především v zahraničí.

V Česku byla sezónnost úmrtnosti popsána u historických populací na území českých zemí v publikaci *Dějiny obyvatelstva českých zemí*. Vývoj úmrtnosti je analyzován v souvislosti s demografickými krizemi, které jsou typické pro populace předindustriálního období a označují situace, kdy došlo k náhlému zvýšení úmrtnosti v důsledku epidemie nebo hladomoru (Fialová, 1996, s. 116). Pomocí sezónních indexů je popisována úmrtnost v 2. polovině 18. století, kdy jsou pro toto období typické nadprůměrné hodnoty na začátku roku, v lednu, únoru a březnu. Znatelná je i nadprůměrná hodnota v září, tedy na konci léta, což je autory připisováno vysoké kojenecké a dětské úmrtnosti spojené se střevními potížemi. Vysoké hodnoty počtu zemřelých na konci zimy pak opisují úmrtnost starých lidí především na nemoci dýchacích cest (tamtéž, str. 124). Další práce zabývající se analýzou sezónnosti úmrtnosti uvedla Klára Tesárková ve svém příspěvku „Sezónnost demografických událostí“ uvedeném na Přírodovědecké fakultě (Tesárková, 2008), kde mj. vznesla otázku souvislosti mezi výkyvy v úmrtnosti a výkyvy v počasí. Tuto hypotézu pak podrobněji analyzovala Gabriela Myšáková ve své bakalářské práci s názvem „Vývoj sezónnosti úmrtnosti podle příčin v Česku, Norsku a Španělsku“ (Myšáková, 2009). Společně pak tyto dvě autorky shrnuly poznatky o sezónnosti úmrtnosti v článku „Analýza sezónnosti v demografii se zaměřením na úmrtnost“, který vyšel v roce 2010 v časopise *Demografie*, a který představuje dvě metody analýzy časových řad, a to tradiční dekompozici časové řady a metodu X12. Jejich aplikací

na vybraná data byla objevena sezónnost u dvou skupin příčin úmrtí, úmrtí na nemoci oběhové soustavy a nemoci dýchací soustavy. U úmrtnosti podle věkových skupin byla sezónnost potvrzena jako statisticky významná u věkové skupiny 15–29 let a u osob starších 50 let (věkové skupiny 50–69 let a 70 a více let). V sezónnosti těchto dvou skupin jsou však výrazné rozdíly. Zatímco mladí lidé zaznamenávají maximum v letních měsících, což je následkem především tragických dopravních nehod a úrazů v době teplého počasí, vrchol úmrtnosti starších lidí připadá do zimních měsíců, především na začátek roku. Pomocí korelační analýzy je sledována souvislost mezi odchylkami v počasí a výkyvy v úmrtnosti, a to na českých datech v období let 1999–2007. Slabá záporná závislost na výkyvech počasí byla odhalena u zemřelých na nemoci oběhové soustavy (Myšáková, 2010).

V zahraničí literatuře nalezneme monografii o sezónnosti úmrtnosti, vydanou v rámci ústavu Max Planck Institute for Demographic Research, který je i významným vydavatelem článků s tematikou sezónních vlivů v demografii. Monografie, která se zabývá rozložením úmrtnosti během roku, vyšla v roce 2007, jmenuje se „Seasonality in Human Mortality: A Demographic Approach“ a jejím autorem je Roland Rau. V knize se na konkrétních datech dvou odlišných států snaží nalézt potenciální dopad sociálních a kulturních rozdílů na sezónnost úmrtnosti. Prvním objektem studie jsou Spojené státy americké, které reprezentují zemi, jejíž populace se vyznačuje velkými sociálními rozdíly. Dalším typickým znakem je variabilita přírodních podmínek na rozlehlém území celého státu. Druhá země, charakterizovaná jako homogenní jak z hlediska populace, tak z hlediska přírodních podmínek, je Dánsko (Rau, 2007). Mezi publikovanými výsledky je odhalení rostoucího vlivu úmrtnosti spojené s chladným počasím na osoby vyššího věku, nepotvrzení rozdílů mezi pohlavími v rozložení zemřelých během roku, potvrzení rozdílů v sezónnosti podle dosaženého vzdělání ve Spojených státech amerických, a to ve smyslu: čím kratší doba strávená vzděláváním, tím větší rozdíly mezi jednotlivými měsíci v roce. Souvislost se socio-ekonomickým statutem, který je ovlivněn úrovní dosaženého vzdělání, se nepotvrdil v Dánsku, což autoři přičítají vyšší homogenitě a životní úrovni populace Dánska (Rau, 2007). V publikaci využil autor i několika svých předchozích článků, mezi které patří „Seasonal mortality in Denmark: the role of sex and age“ (Rau, 2003), který publikoval v časopise *Demographic Research*, „Winter mortality in elderly people in Britain: lack of social gradient in winter excess mortality is obvious in Denmark“ (Rau, 2004) zveřejněný v časopise *British Medical Journal* nebo „Modulation models for seasonal time series and incidence tables“ (Eilers, 2008), který vyšel v časopise *Statistics in medicine*. Rozdíly v úmrtnosti během roku podle věku zemřelého se zabývá článek „Seasonality of Deaths in the U. S. by Age“ (Feinstein, 2002), který vyšel v *Demographic Research*. Odborné časopisy s tematikou lidského zdraví, které publikují články zabývající se sezónností míry úmrtnosti, jsou například: *International Journal of Biometeorology* s příspěvkem „Changes in the seasonality of mortality in Germany from 1946 to 1995: the role of temperature“ (Lerchl, 1998). Specifickou oblast věnovanou sezónnosti úmrtnosti diferencované podle příčiny úmrtí se věnují například články „Seasonality of coronary artery deaths in New South Wales, Australia“, který vyšel v časopise *Heart* (Weerasinghe, 2002), nebo

příspěvek „Seasonal variations in coronary heart disease“ od autorů J. P. Pell a S. M. Cobbe, který vyšel v časopise *QJM: An Interantional Journal of Medicine*. Článek mapuje již provedené výzkumy o souvislosti úmrtnosti na nemoci oběhové soustavy s měsíční teplotou. Popisuje také rizikové faktory životního stylu, které mohou souviset s ročním obdobím. Jedním z nich je snížení pohybové aktivity a cvičení během zimních měsíců, dalším pak zvýšený příjem tuků v této části roku (Pell, 1999). Nemocemi oběhové a dýchací soustavy se v časopise *The Lancet* v příspěvku s názvem „Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe“ zabývali v rámci odborné skupiny vědci z několika států Evropské unie – Finska, Itálie, Německa, Nizozemí, Řecka a Velké Británie (The Eurowinter Group, 1997). Již zmíněnými dvěma skupinami příčin úmrtí a navíc skupinou novotvarů se zabývá článek „The Impact of Heat Waves and Cold Spells on Mortality Rates in the Dutch Population“ (Huynen, 2001).

Další příčinou úmrtí, na kterou je zaměřena pozornost demografů je sebevražda, která je ovlivněna rozhodnutím člověka a tedy i vnějšími okolnostmi, mezi kterými může být i roční doba, a o které se můžeme dočíst v článku „Sezónní aspekty sebevražednosti“ uvedeném na Demografickém informačním portálu (Kasalová Daňková, 2005).

## 2 Zdroje dat

Pro analýzu časových řad za účelem detekování sezónnosti je vždy zapotřebí podrobných časových řad, v této práci byly použity měsíční časové řady. Možnosti získání vhodných dat z volně přístupných databází jsou omezené, proto byla pro tuto práci z velké části využita data za Česko, která jsou v podrobné formě dostupná na internetu. V případě počtu živě narozených, sňatků a zemřelých celkem, která patří mezi nejčastěji studované demografické události z hlediska ročního rozložení událostí, jsou data za Česko porovnána i s dalšími evropskými zeměmi, pro které jsou tyto statistiky veřejně přístupné prostřednictvím webových stránek příslušných statistických úřadů a jejich internetových databází nebo prostřednictvím statistického úřadu Evropské unie EUROSTAT.

### 2.1 Data Česko

Data za Česko použitá k analýzám byla získána z webových stránek Českého statistického úřadu<sup>1</sup>, kde jsou zveřejněna data z každoročního pramenného díla o obyvatelstvu. Pramenné dílo je základním zdrojem podrobných dat o pohybu obyvatelstva, také o jeho početní velikosti a složení podle pohlaví, věku a rodinného stavu v daném roce (ČSÚ, 2011). Obsažena jsou data za všechny demografické události (pokud byla v daném roce zjišťována) – sňatky, rozvody, narození, potraty, úmrtí a stěhování (vnitřní i zahraniční). Aktuálně dostupná jsou data z let 1983–2009. Od roku 1983 se jednalo o publikace „Pohyb obyvatelstva v České socialistické republice“ do roku 1990, „Pohyb obyvatelstva v České a Slovenské Federativní republice“ v období let 1990–1992, „Pohyb obyvatelstva v České republice“ pro roky 1993–2005 a od roku 2006 vychází publikace pod názvem „Demografická ročenka České republiky“ (ČSÚ, 2011)<sup>2</sup>.

S ohledem na skutečnost, že se jedná o dlouhé časové období, je pochopitelné, že se obsah statistických zjišťování během let měnil a ne všechny podrobněji specifikované proměnné (např. úmrtnost dle věkových skupin) lze sledovat po celých 27 let. Přesný seznam do analýzy vstupujících demografických jevů, vybraných ukazatelů a jejich časové vymezení znázorňuje tabulka 1.

---

<sup>1</sup> Český statistický úřad – [www.czso.cz](http://www.czso.cz)

<sup>2</sup> Demografie - pramenná díla – [www.czso.cz/csu/redakce.nsf/i/casova\\_rada\\_demografie](http://www.czso.cz/csu/redakce.nsf/i/casova_rada_demografie)

**Tab. 1 – Seznam analyzovaných událostí a jejich časového vymezení, Česko**

Demografický proces	Ukazatel	Časové období
Porodnost	Počet živě narozených	1983–2009
Sňatečnost	Počet uzavřených sňatků	1983–2009
Úmrtnost		
podle pohlaví	Počet zemřelých mužů	1983–2009
	Počet zemřelých žen	1983–2009
podle věku	Počet zemřelých do 1 roku věku	1983–2009
	Počet zemřelých ve věkové skupině 0–14 let	1995–2009
	Počet zemřelých ve věkové skupině 15–29 let	1995–2009
	Počet zemřelých ve věkové skupině 30–49 let	1995–2009
	Počet zemřelých ve věkové skupině 50–69 let	1995–2009
	Počet zemřelých ve věkové skupině 70+	1995–2009
podle příčiny úmrtí	Počet zemřelých na nemoci dýchací soustavy	1995–2009
	Počet zemřelých na nemoci oběhové soustavy	1995–2009

## 2.2 Data ostatních evropských států

Pro mezinárodní srovnání dosažených výsledků o porodnosti, sňatečnosti a úmrtnosti byla získána a podrobena analýze data za některé další evropské státy, uvedené v tabulce 2. Vybrané státy pokrývají svoji geografickou polohou celé území Evropy a nacházejí se v různé demografické i ekonomické situaci. Liší se i svoji pozicí vzhledem k Evropské unii, neboť jsou v seznamu reprezentováni jak členové EU, tak i jedna nečlenská země (Norsko) a jedna země kandidátská (Turecko). Měsíční data o celkové úmrtnosti populací vybraných evropských zemí byla získána ze statistického úřadu Evropské unie, EUTOSTATu, který však zveřejňuje statistické údaje i za nečlenské státy EU.

**Tab. 2 – Vybrané země Evropy, název a webová adresa příslušného statistického úřadu, ze kterého byla získána data použitá v této práci**

Země	Název statistického úřadu v angličtině	Webová adresa
Dánsko	Statistics Denmark	<a href="http://www.dst.dk">www.dst.dk</a>
Estonsko	Statistics Estonia	<a href="http://www.stat.ee">www.stat.ee</a>
Evropské země	Eurostat	<a href="http://epp.eurostat.ec.europa.eu">epp.eurostat.ec.europa.eu</a>
Finsko	Statistics Finland	<a href="http://www.stat.fi">www.stat.fi</a>
Island	Statistics Iceland	<a href="http://www.static.is">www.static.is</a>
Litva	Statistics Lithuania	<a href="http://www.stat.gov.lt">www.stat.gov.lt</a>
Lotyšsko	Central Statistical Bureau of Latvia	<a href="http://www.csb.gov.lv">www.csb.gov.lv</a>
Lucembursko	National Institute of Statistics and Economic Studies	<a href="http://www.statistiques.public.lu">www.statistiques.public.lu</a>

<b>Země</b>	<b>Název statistického úřadu v angličtině</b>	<b>Webová adresa</b>
Nizozemsko	Statistics Netherlands	<a href="http://www.cbs.nl">www.cbs.nl</a>
Norsko	Statistics Norway	<a href="http://www.ssb.no">www.ssb.no</a>
Slovensko	Statistical Office of the Slovak Republic	<a href="http://www.statistics.sk">www.statistics.sk</a>
Slovinsko	Statistical Office of the Republic of Slovenia	<a href="http://www.stat.si">www.stat.si</a>
Španělsko	National Statistics Institute	<a href="http://www.ine.es">www.ine.es</a>
Turecko	Turkish Statistical Institute	<a href="http://www.turkstat.gov.tr">www.turkstat.gov.tr</a>

Zdroj: *International Statistical Agencies*, [www.census.gov/aboutus/stat\\_int.html](http://www.census.gov/aboutus/stat_int.html)

Všechna zdrojová data, která byla v této diplomové práci použita pro výpočty a analýzy, jsou dostupná na příloženém CD.

### 3 Metodologie

Základním cílem analýzy časových řad je snaha porozumět principům, na základě kterých se generují hodnoty řady. Pokud se nám podaří tento mechanismus odhalit, získáváme šanci, že budeme schopni předpovídat budoucí vývoj systému a v některých případech tento systém také řídit a optimalizovat jeho vývoj (Kvasnička, 2001). Aby bylo možné podrobně porozumět vybrané časové řadě a spolehlivě odhalit sezónnost v jejím průběhu, je třeba rozložit tuto časovou řadu na jednotlivé složky, kterými je tvořena. K tomuto rozložení lze využít některou z mnoha metod vyvinutých pro analýzu časových řad.

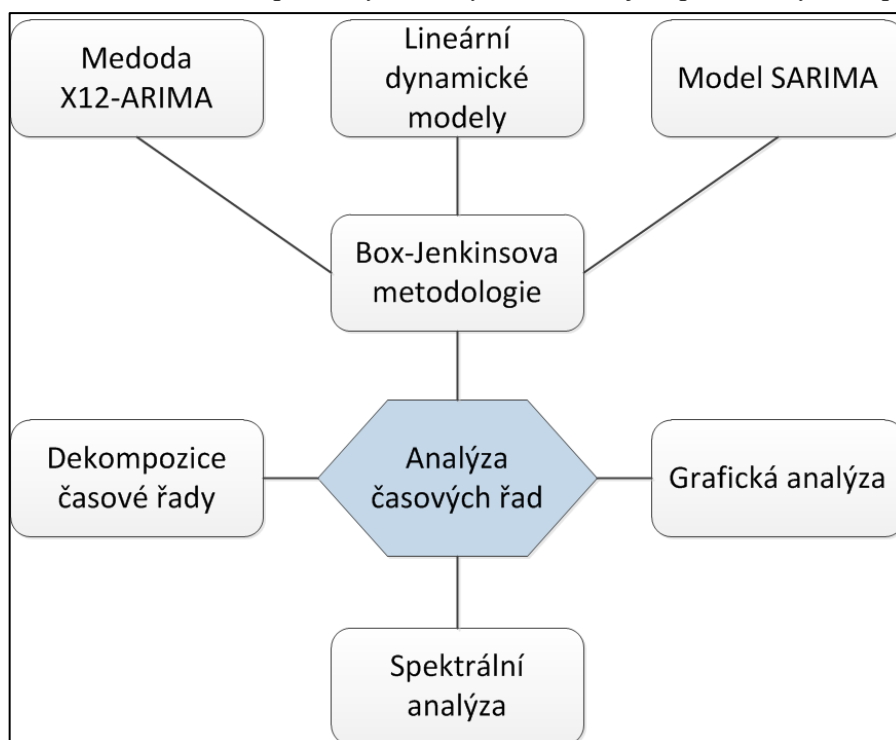
Myšlenka, že časová řada je složená z několika nepozorovatelných složek s rozdílnými periodicitami byla původně nezbytnou součástí mnoha teorií vytvořených v astronomii. Pro ekonomy (v Anglii) se stala velmi populární v období let 1825–1875. První, kdo studoval sezónní modely v ekonomických časových řadách, byl bankéř James W. Gilbert v polovině 19. století, jenž přišel na to, že poptávka po penězích Bank of England byla v určitých měsících roku vysoká a v jiných nízká (Arlt, 1995).

Pro tuto diplomovou práci bylo vybráno několik metod vhodných pro analýzu časových řad, jejichž více či méně podrobný popis obsahuje tato kapitola. Podrobně jsou popsány tři komplexní metody, a to dekompozice časové řady, která rozkládá časovou řadu na jednotlivé složky, metoda X12–ARIMA, která také rozkládá časovou řadu, navíc počítá proměnlivé sezónní indexy a testy sezónnosti. Třetí metodou je kointegrace časových řad, která hledá vztah mezi nejméně dvěma časovými řadami. Stručně představeny jsou další čtyři metody využitelné při analýze časových řad. Jedná se o grafickou analýzu, model SARIMA Box-Jenkinsovy metodologie, lineární dynamické modely a spektrální analýzu. Schéma metod, které jsou popsány v této práci, je znázorněno na obrázku 1.

Na data o sňatečnosti a úmrtnosti v evropských zemích jsou nejprve aplikovány výše zmíněné metody a následně je na získané výsledky aplikována shluková analýza. Tato vícerozměrná statistická metoda třídí jednotlivé jednotky (v tomto případě evropské státy) do shluků na základě vzájemné (ne)podobnosti. S využitím shlukové analýzy tak získáváme další rozměr poznání, neboť po odhalení sezónnosti ve sňatkovém chování a úmrtnosti evropských populací, můžeme najít skupiny států, které si jsou podobné na základě určitého ukazatele, např. sezónních indexů, proměnlivých sezónních indexů nebo náhodné složky.



**Obr. 1 – Schéma metod pro analýzu časových řad, které jsou představeny v této práci**



### **Časová řada**

Časová řada je posloupnost hodnot určitého statistického znaku (ukazatele) uspořádaná v čase, zpravidla ve směru od minulosti do přítomnosti. Výchozím předpokladem je, že vybraný ukazatel je věcně a prostorově shodně vymezen po celé sledované období. Z uvedené posloupnosti lze tedy usuzovat pouze na působení časových změn (Kozák, 1994).

Pro analýzu vývoje časové řady je třeba odlišovat dva typy časových posloupností ukazatele. Časovou řadu okamžikového ukazatele a časovou řadu intervalového ukazatele. U okamžikové časové řady se hodnota ukazatele plynule mění v čase, ukazatel je tedy definován spojitě v čase (Kozák, 1994). Příkladem takové časové řady může být např. řada teplot ovzduší na hydrometeorologické stanici odečítaná každou hodinu. Sčítání hodnot okamžikové časové řady nemá logiku. Někdy je okamžiková časová řada také nazývána jako stavová časová řada. Intervalovou časovou řadu tvoří hodnoty intervalového ukazatele, který je definován pro určitý časový interval (týden, měsíc, čtvrtletí) a je tedy v čase nespojitý. Pro časovou řadu je charakteristické, že s rostoucí délkou časového intervalu hodnota intervalového ukazatele řádově roste (Kozák, 1994). Počet narozených dětí za jeden měsíc je příkladem intervalové časové řady. S ohledem na skutečnost, že měsíční počty událostí získáváme kumulací událostí za jednotlivé dny, je zřejmé, že u tohoto typu časových řad sčítání hodnota smysl má.

Časové řady lze dále dělit na deterministické a stochastické. Deterministické v sobě neobsahují žádný prvek náhody a můžeme je tedy dokonale a bezchybně předpovídat. Stochastické časové řady jsou ovlivněny, a tedy v sobě obsahují, prvek náhody. Frekvence

hodnot časové řady je dalším kritériem, podle kterého je lze dělit na dlouhodobé, krátkodobé a vysokofrekvenční. Hodnoty dlouhodobých časových řad jsou sledovány v ročních či delších časových úsecích. Krátkodobé časové řady vykazují hodnoty z úseků kratších než jeden rok, a to řady čtvrtletní a měsíční. Tyto řady bývají obvykle delší než dlouhodobé časové řady a jejich průběh je variabilnější. Vysokofrekvenční časové řady sledují týdenní a denní hodnoty ukazatele (Arlt, 2002). Krátkodobé a vysokofrekvenční časové řady lze souhrnně označit za časové řady področní. V této práci jsou všechny analyzované časové řady intervalové, s časovým intervalem jeden měsíc.

Především při analýze ekonomických časových řad je důležité rozlišovat časové řady stacionární a nestacionární. Stacionární časové řady se vyznačují kolísáním hodnot časové řady kolem konstanty, kdy se střední hodnota a rozptyl tak v čase nemění. Tyto časové řady se označují jako řady s krátkou pamětí, neboť se ve stacionárním modelu vliv šoku (informace) z minulého období časem postupně vytrácí. S těmito řadami se standardně pracuje jako s prostorově uspořádanými daty a při zkoumání jejich vztahu, např. pomocí regresní analýzy, nevzniká zpravidla žádný problém, který by výsledky analýzy znehodnocoval (Arlt, 2010). U nestacionárních časových řad se střední hodnota procesu, který časovou řadu tvoří, v čase mění. Základním modelem nestacionárních časových řad je proces náhodné procházky, který obsahuje i stochastický trend, který se projevuje cyklickým vývojem. Nestacionární časové řady jsou tedy označovány jako řady s dlouhou pamětí, protože se v průběhu času informace nevytrácí, ale kumuluje. Kumulací těchto informací (šoků) pak tvoří trend. Při analýze nestacionárních časových řad může vzniknout problém tzv. zdánlivé regrese, jejímž zdrojem jsou rozdílné stochastické trendy vysvětlované a vysvětlující proměnné (Arlt, 1998).

### ***Základní charakteristiky časových řad***

Abychom si o vybrané časové řadě a jejích hodnotách dokázali udělat jasnou představu, je možné zvolit několik metod, nebo využít jejich kombinaci. Prvním krokem může být vykreslení časové řady do jednoduchého grafu, který je podkladem pro grafickou analýzu dat. Pro přesné vyjádření charakteristik časové řady je vhodné využít některý z následujících ukazatelů. Zajímají-li nás charakteristiky polohy, vyjádřené průměrnými hodnotami, nabízí se výpočet několika průměrů, a to od prostého aritmetického, přes vážený aritmetický průměr až k váženému chronologickému průměru, který zohledňuje ve výpočtu délku jednotlivých časových intervalů (viz Arlt, 2002). Dvě hlavní charakteristiky variability tvoří rozptyl a směrodatná odchylka, jejichž výpočet je na sobě závislý. Další, méně často užívanou charakteristikou je variační koeficient, k jehož výpočtu přistupujeme, pokud chceme porovnat rozptýlenost dat skupin stejné proměnné s různým průměrem, nebo tehdy, mění-li se velikost směrodatné odchylky tak, že je přímo závislá na úrovni měřené proměnné (Hendl, 2006). Charakteristiky variability určují odlišnost jednotlivých hodnot znaku od aritmetického průměru a umožňují tak porovnávat navzájem několik statistických souborů. Míry dynamiky jsou třetí skupinou ukazatelů, pomocí nichž jsou popisovány základní rysy chování časové řady. Mezi

tyto ukazatele patří absolutní přírůstek, koeficient (tempo) růstu, meziroční koeficient růstu a relativní přírůstek (více viz Arlt, 2002).

### ***Problémy časových řad***

Při analýze časových řad dochází k přeměně modelů oborových disciplín na modely matematické (Bříza, 2010). V případě demografie se tedy jedná o snahu popsat vývoj konkrétního demografického procesu (např. měsíčních počtů uzavřených sňatků) pomocí matematických funkcí tak, aby výsledný nadefinovaný model spolehlivě popisoval průběh časové řady po celé sledované období. Při transformaci těchto dvou typů modelů (oborového a matematického) mohou vzniknout problémy, kterým musí statistik čelit. Mezi tyto problémy řadíme (Bříza, 2010):

#### ***a) problémy s volbou časových bodů pozorování***

Diskrétní časové řady (řady tvořené pozorováními v určitých nespojitých časových bodech) mohou vznikat třemi způsoby. Jsou buď diskrétní svou povahou (hodnota ukazatele za jednotlivé roky) nebo vznikají diskretizací spojité časové řady (např. teplota v danou denní dobu na daném místě) nebo vznikají agregací (akumulací) hodnot za dané časové období (např. počet zemřelých za měsíc). Ne vždy se nabízí možnost volby časových bodů pozorování. Pokud však tato možnost existuje, je třeba věnovat výběru pozornost a najít kompromis mezi protichůdnými požadavky. Na straně jedné je vhodné z důvodu numerické jednoduchosti výpočtů příliš „nezhušťovat“ počet pozorování. Na straně druhé však nesmí být časová řada natolik řídká, že by mohl uživateli uniknout charakteristický rys časové řady. V případě jako je tato práce, kdy nás zajímají změny časové řady během roku, tedy sezónní fluktuace, musíme mít k dispozici několik pozorování během každého roku vstupujícího do analýzy. Interval mezi jednotlivými sousedními pozorováními by měl být ekvidistantní, tzn. stejné délky (Cipra, 1986).

#### ***b) problémy s kalendářem***

Dnešní kalendář ovlivňuje hodnoty časových řad svými nepravidelnostmi, mezi které patří různá délka kalendářních měsíců, proměnlivý počet víkendů v měsíci a s tím úzce spojený počet pracovních dní v měsíci, nebo pohyblivé svátky, kterými jsou Velikonoce. Na tyto nepravidelnosti existují metody, pomocí kterých můžeme časovou řadu očistit a tím odstranit vliv specifických vlastností jednotlivých měsíců (Cipra, 1986). I v této práci byly takové metody využity, konkrétně standardizace na stejný počet dní v měsíci a standardizace na stejný počet pátků a sobot v měsíci, kterou jsou popsány dále.

#### ***c) problémy s délkou časových řad***

Délku časové řady vyjadřuje počet měření, nikoliv časové rozpětí mezi prvním a posledním měřením v časové řadě. I v případě délky časové řady je třeba najít kompromis mezi dvěma pohledy. Zatímco některé metody vyžadují minimální

délku řady, na druhé straně dochází u velmi dlouhých časových řad k nebezpečí, že se v průběhu času významně změní charakteristika modelu, který příslušnou řadu generuje, čímž je odhad modelu s rostoucí délkou řady obtížnější (Cipra, 1986).

**d) problém „ceteris paribus“**

Název tohoto problému vyjadřuje latinské sousloví „za jinak stejných podmínek“, a poukazuje na skutečnost, že získané modely a výsledky analýz jsou platné pouze za stejných okolností a podmínek, za kterých byly odvozeny. Naprosté shody vnějších podmínek však nelze v reálném světě dosáhnout. Z toho důvodu je pro statistické účely pojem rozšířen na „za jinak stejných nebo podobných podmínek“ (Bříza, 2010). Dojde-li k tedy k významné změně okolností, je třeba podrobit časovou řadu novému zkoumání, někdy s využitím odlišné metody.

**Standardizace vstupních dat**

Prvním, a pro všechny metody společným, krokem analýzy je nezbytná úprava získaných dat metodou standardizace na stejný počet dní (a stejný počet pátků a sobot v případě sňatečnosti) v měsíci. Tato transformace je nezbytná pro možnost porovnávání jednotlivých měsíců, které se jinak vyznačují různou délkou. Standardizace na stejný počet pátků a sobot v měsíci při analýze sňatkového chování populace je vhodná z důvodu nejčastějšího uzavírání sňatků v těchto dvou dnech. Nezhlednění této skutečnosti by mohlo výsledky o koncentraci uzavřených sňatků do jednotlivých měsíců v roce značně zkreslit.

Statistikou vykázané, tedy skutečné, počty událostí v příslušných měsících byly upraveny přepočtovým koeficientem. Tento koeficient přepočtu odpovídá skutečnému počtu dní daného měsíce dělenému počtem dní průměrného měsíce, jehož délku lze vyjádřit jako  $365/12$  (resp.  $366/12$  v případě přestupného roku). V případě měsíců, které jsou kratší než průměrný měsíc, tak dojde ke zvýšení počtu standardizovaných událostí, než vykazovaly původní statistiky. U měsíců, jejichž počet dní přesahuje počet dní průměrného měsíce, naopak standardizované počty událostí oproti původním hodnotám poklesnou (Myšáková, 2009).

Standardizace na stejný počet dní v měsíci je dostačující u procesů, které předpokládají rovnoměrné rozložení událostí během měsíce, z analyzovaných jevů v této práci tedy porodnost a úmrtnost. U sňatečnosti je nutné provést také standardizaci na stejný počet pátků a sobot v měsíci z důvodu již zmíněné obliby těchto dvou dnů pro uzavření sňatku, oproti zbyvajícím dnům v týdnu.

Postup standardizace je založen na stejném principu jako při standardizaci na jednotnou délku měsíce. Za celé studované období je třeba zvlášť zohlednit počet pátků a počet sobot každého měsíce. S ohledem na skutečnost, že do této standardizace vstupují dvě proměnné, je možné zvolit jeden ze dvou způsobů výpočtu. Prvním a jednodušším je standardizace bez ohledu na podíl pátků a sobot na celkovém počtu těchto dvou dní v měsíci. Bereme tedy oba dny jako rovnocenné a nezohledňujeme, zda je například pátků pět a soboty jen čtyři nebo naopak. Druhou možností je zohlednění podílového zastoupení těchto dvou dní, a následným přidělením váhy každému z nich. Přepočtový koeficient odpovídá skutečnému počtu pátků

a skutečnému počtu sobot v každém měsíci dělenému počtem pátků a sobot průměrného měsíce. Pro získání hodnoty průměrného počtu pátků a sobot v průměrném měsíci lze využít již uvedeného vzorce pro délku průměrného měsíce jako  $\left(\frac{365}{12}\right)/7$  (resp.  $\left(\frac{366}{12}\right)/7$  v případě

přestupného roku). S ohledem na nestejnou významnost pátku a soboty jako dne zvoleného pro uzavření sňatku, bylo v této práci na data použito různých vah pro jednotlivé dny, a to v poměru 0,205 pro páteční hodnoty a 0,723 pro sobotní údaje o počtu uzavřených sňatků. Tyto hodnoty byly zvoleny na základě poznatků o sňatečnosti z počátku 21. století, kdy bylo v pátek uzavřeno 20,5 % sňatků a v sobotu 72,3 % (viz Polášek, 2005). Aplikací zvolených vah na přepočítací koeficienty obou dnů byl získán jeden výsledný přepočítací koeficient, který byl v posledním kroku standardizace aplikován na data již standardizovaná na stejný počet dní v měsíci, a tím bylo dosaženo nové, dvakrát standardizované, časové řady. Pro nové hodnoty pak platí, že na ně působí dvě pomyslné síly. První z nich, která byla popsána výše, že u kratších měsíců je standardizovaná hodnota vyšší, zatímco u měsíců delších než průměrný měsíc, dojde po přepočtu ke snížení počtu událostí. Obdobně je tomu i s počtem pátků a sobot, neboť je-li počet pátků a sobot nižší než v průměrném měsíci, standardizovaná hodnota časové řady se zvýší. Je-li naopak počet těchto dvou dnů nad průměrnou hodnotu, přepočtem se hodnota zmenšuje. Pokud jsou navíc zvoleny váhy, ovlivňuje přepočítanou hodnotu i to, zda je v daném měsíci více pátků nebo více sobot. V případě zde zvolených vah znamená více sobot po standardizaci výsledné nižší hodnoty, zatímco při více pátcích je zpravidla hodnota výsledné časové řady přepočtem zvýšena. Varianty, které mohou v rozložení pátků a sobot během jednoho měsíce nastat jsou čtyři: 4 pátky a 4 soboty, 4 pátky a 5 sobot, 5 pátků a 4 soboty nebo 5 pátků a 5 sobot (Tesárková, 2009a).

### 3.1 Dekompozice časové řady

Dekompozici časové řady lze chápat dvěma způsoby, první plyne ze samotného názvu a znamená rozklad časové řady na jednotlivé složky, což je proces, který využívá velké množství metod zabývajících se časovými řadami. Druhý úhel pohledu má význam stejný, tedy rozložení časové řady na několik částí, ale chápe dekompozici časové řady jako název pro konkrétní metodu výpočtu, podloženou konkrétní teorií a odvozenými vzorci. Někdy je název této metody doplňován o přívlastek tradiční, nebo klasická, dekompozice časové řady, čímž je snaha odlišit název metody od obecného vyjádření rozkladu časové řady. V pojetí konkrétní metody ji představuje i tato podkapitola.

Předem standardizovaná časová řada je aplikací metody dekompozice rozložena na čtyři samostatné složky, tři systematické a jednu nesystematickou. Mezi systematické složky patří trendová složka, sezónní složka a cyklická složka. Náhodná, někdy také označovaná jako reziduální, složka je nesystematickou částí časové řady (Cipra, 1986). Tyto složky se označují následovně:

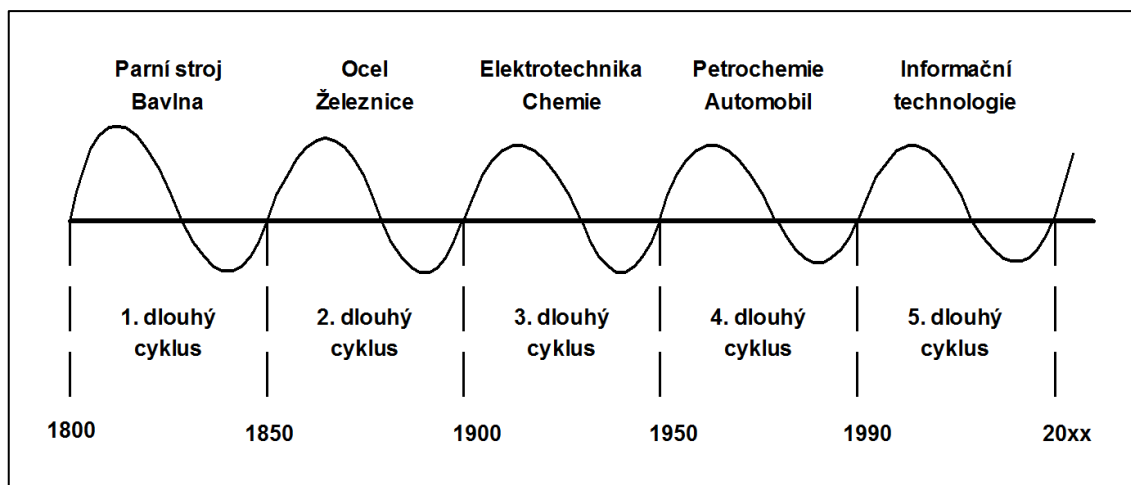
- Trendová složka ( $Tr_t$ )
- Sezónní složka ( $Sz_t$ )
- Cyklická složka ( $C_t$ )
- Náhodná složka ( $E_t$ )

Trendová složka neboli trend vyjadřuje obecnou tendenci vývoje zkoumaného jevu za dlouhé období. Je výsledkem dlouhodobých a stálých procesů. Trend může být rostoucí, klesající nebo může existovat řada bez trendu. Trendová složka se většinou modeluje pomocí klouzavých průměrů nebo s využitím matematických křivek (Hančlová, 2003). Podle průběhu trendu a jeho vyjádření vybranou matematickou funkcí lze rozlišit například trend konstantní, lineární, kvadratický, exponenciální a logistický trend (Kozák, 1994). Sezónní složka odráží pravidelné, periodicky se opakující fluktuační souvislosti se střídáním ročních období, množstvím denního světla, atd. Perioda této složky je kratší než celková velikost analyzovaného časového úseku. Sezónní složka může, podobně jako trend, měnit svůj charakter během období (Arlt, 2002). Sezónní složku lze popsat dvěma způsoby. První, a jednodušší na výpočet, jsou sezónní indexy zprůměrované za celé sledované období, získáváme tedy dvanáct hodnot, pro každý měsíc jednu. Vykreslením těchto hodnot do paprskového grafu získáme přehlednou a ucelenou informaci o sezónním rozložení sledované časové řady. Sezónnost se však může v průběhu doby měnit, obzvláště analyzujeme-li dlouhé časové období. V takovém případě je možné vybrat ze dvou metod výpočtu a znázornění. Máme-li dobrou znalost vývoje sledovaného jevu, můžeme dlouhou časovou řadu rozdělit na několik kratších, homogenních úseků, a pro každý tento úsek vypočítat sezónní indexy, které lze následně navzájem porovnávat. Druhou možností je výpočet v čase proměnlivých sezónních indexů, který nabízí např. metoda X12-ARIMA, jež je popsána níže. Cyklická složka udává kolísání okolo trendu v důsledku dlouhodobého cyklického vývoje, kdy dochází ke střídání fází růstu a poklesu. Jednotlivé cykly se vytvářejí zpravidla za období delší než jeden rok a mohou mít nepravidelný charakter, tzn. různou amplitudu (Hančlová, 2003). Příkladem cyklické složky v ekonomii mohou být teorie hospodářských cyklů, například Kondratěvy dlouhodobé cykly (vlny)<sup>3</sup>, znázorněné na obrázku 2.

---

<sup>3</sup> Kondratěvy hospodářské cykly – „V cyklech jde o výměnu generací techniky a technologií, zásadní změnu v pracovních nástrojích a zdrojových energiích, o vznik principiálně nových směrů techniky a o všestranné techniky přezbrojení znamenající novou kvalitu. Začátek cyklu přináší konjunkturu a iniciuje dalekosáhlé změny v celé společnosti. Dlouhé cykly závisejí na inovacích a investicích do inovací, tedy na činitelích jak vědeckotechnických, tak ekonomických“ (Dufková, 2001).

Obr. 2 – Kondratěvovy dlouhé hospodářské cykly



Zdroj: Dufková, 2001: *Inovace a nová ekonomika*.

Náhodná složka představuje výkyvy, které jsou dynamické, nepředvídatelné a nemají systematický charakter (Cipra, 1986; Tesárková, 2009a). Předpokládá se, že náhodná složka je tvořena tzv. bílým šumem s normálním rozdělením. Pod pojmem bílý šum rozumíme nekorelované (vzájemně nezávislé) náhodné veličiny s nulovou střední hodnotou a konstantním rozptylem (Arlt, 2002).

Jak již bylo nastíněno, metoda dekompozice časové řady rozkládá vybranou časovou řadu na jednotlivé složky, které ji tvoří. Vlastní rozklad časové řady může být proveden třemi způsoby, vyjádřen třemi modely, a to aditivním, multiplikativním nebo smíšeným (viz Cipra, 1986). Aditivní dekompozice má tvar:

$$y_t = Tr_t + Sz_t + C_t + E_t$$

Při aditivním rozkladu jsou jednotlivé složky uvažovány ve svých skutečných absolutních hodnotách a jsou měřeny v jednotkách jako původní řada  $y_t$ . Aditivní dekompozice se používá v případě, že variabilita hodnot časové řady je přibližně konstantní v čase (Kvasnička, 2001). Multiplikativní dekompozice časové řady, aplikovaná na časové řady v této práci, se používá v případě, že variabilita časové řady roste v čase, nebo se v čase mění. V multiplikativním modelu je trendová složka časové řady ve stejných měrných jednotkách jako původní časová řada (v našem případě počet sňatků, počet zemřelých,...), ale ostatní složky, cyklická, sezónní a náhodná, jsou v relativním vyjádření (Cipra, 1986). Sezónní složku tvoří sezónní indexy, které udávají hodnoty pro jednotlivé měsíce celého sledovaného období, a vyjadřují, zda počet událostí byl v příslušném měsíci nadprůměrný (hodnota vyšší než jedna), podprůměrný (hodnota nižší než jedna) nebo odpovídal průměrnému měsíci, čímž hodnota indexu dosáhla jedné. Součet indexů za jednotlivé měsíce je roven hodnotě 12. Nejspornější složkou časové řady je cyklická složka. Délka jednotlivých cyklů (tj. vzdálenost mezi dvěma sousedními horními nebo dolními body zvratu) je obvykle proměnlivá a stejně tak může být proměnlivá intenzita jednotlivých fází cyklického průběhu (Cipra, 1986). Modelování

cyklické složky je proto poměrně obtížné. V krátkém období je možné její vliv zanedbat (Kvasnička, 2001). V této práci je cyklická složka zanedbána, tj. zahrnuta v trendu. Multiplikativní model je vyjádřen:

$$y_t = Tr_t \times Sz_t \times C_t \times E_t$$

Smíšený model je kombinací aditivního a multiplikativního modelu. Některé složky mohou být v součtu, jiné v součinu. Příkladem takového modelu může být vyjádření  $y_t = Tr_t \times Sz_t + C_t + E_t$ . Identifikace smíšených časových řad je nejobtížnější (Kvasnička, 2001).

### **Odhad trendu**

Trendovou složku časové řady je možné popsat pomocí klouzavých průměrů a trendových funkcí. Modelování trendu pomocí klouzavých průměrů se používá, jestliže je vývoj časové řady v důsledku silného vlivu nesystematické složky nerovnoměrný nebo má extrémní hodnoty. Modelování s využitím trendových funkcí se používá v případě, kdy trend odpovídá určité funkci, např. lineární, kvadratické, exponenciální, S-křivky (Hančlová, 2003).

První z metod, která byla využita při výpočtech v programu Microsoft Excel, je mechanická metoda vyrovnávání s využitím klouzavých průměrů, které jsou lineární kombinací původní řady. Vhodná je tato metoda na modely časových řad s proměnlivými parametry, které vycházejí z hypotézy, že k popisu trendu časové řady lze využít jednoduchého systému funkcí časové proměnné. Předpokladem je, že tyto modely mění v čase hodnoty svých parametrů a jejich odhad tedy nelze provést naráz, nýbrž postupně, po částech, a podle této schopnosti reagovat na změny charakteru trendu jsou řazeny mezi tzv. adaptivní přístupy (Kozák, 1994). S ohledem na postupné vyhlazování časové řady je metoda klouzavých průměrů aplikována na delší časové řady, u kterých se předpokládá proměnlivost parametrů. Nejčastěji se volí klouzavé průměry liché délky (s lichým počtem členů), aby byla prostřední hodnota vyrovnaná. Je-li zvolen sudý počet členů, vyrovnaná hodnota spadá mezi původní dvě a je nutné přistoupit k tzv. centrování. Často se v praxi využívají tzv. jednoduché klouzavé průměry, které jsou vypočítány jako běžné aritmetické průměry, možností je ale i vážení hodnot, kdy jsou váhy symetrické kolem prostřední hodnoty, s větší vahou hodnot blíže středu, a jejich součet je jedna (Tesárková, 2009b). Pro měsíční časové řady je vhodné využití klouzavých průměrů o délce 12 (měsíců) a poté výsledné hodnoty upravit centrováním do finálního odhadu trendové složky.

Analytické vyrovnávání je aplikovatelné na časové řady s konstantními parametry, které vycházejí z předpokladu, že trend časové řady postačí v rámci celého období interpolace popsat jediným systémem parametrů, tedy že tyto parametry postačí v rámci celé časové řady odhadnout najednou (Kozák, 1994). Jedná se tedy často o kratší časové řady, kdy je větší pravděpodobnost dodržení podmínky konstantních parametrů. Při tomto postupu se předpokládá, že časová řada má následující tvar nebo byla na tento tvar převedena:  $y_t = Tr_t + E_t$ , kde  $Tr_t$  je systematická složka a představuje deterministický trend, který lze vyjádřit matematickou funkcí časové proměnné  $t$  a  $E_t$  je nesystematická složka s vlastnostmi



procesu bílého šumu, což znamená, že v čase  $t$  platí, že náhodné veličiny  $E_t$  mají v čase  $t$  nulovou střední hodnotu, konstantní rozptyl, jsou vzájemně lineárně nezávislé a mají normální rozdělení (Arlt, 2002). Základními trendovými funkcemi, pro které platí  $t = 1, 2, \dots, n$ ; jsou (viz. Arlt, 2002):

- a) **konstantní trend** má tvar funkce  $Tr_t = \beta_0$
- b) **lineární trend** má tvar funkce  $Tr_t = \beta_0 + \beta_1 \times t$
- c) **kvadratický trend** (parabola) má tvar funkce  $Tr_t = \beta_0 + \beta_1 \times t + \beta_2 \times t^2$
- d) **exponenciální trend** má tvar funkce  $Tr_t = \beta_0 \times \beta_1^t$ , kde parametry  $\beta_0, \beta_1 > 0$  se odhadují metodou nejmenších čtverců, protože trendová funkce se po logaritmické úpravě převede na lineární funkci
- e) **S-křivka** má tvar funkce  $Tr_t = e^{(\beta_0 + \beta_1 \times (1/t))}$ , který se po logaritmické transformaci dá převést na tvar hyperboly  $\ln Tr_t = \beta_0 + \beta_1 \times (1/t)$ . Parametry odhadujeme opět metodou nejmenších čtverců
- f) **modifikovaný exponenciální trend** má tvar funkce  $Tr_t = \gamma + \beta_0 \times \beta_1^t$ , kde  $\beta_0 < 0$ ,  $0 > \beta_1 < 1$  a  $\gamma > 0$ . Konstanta  $\gamma$  je asymptotou (úroveň saturace, hladinou nasycení), ke které trend časové řady pro  $t \rightarrow \infty$  konverguje. Přírůstek exponenciálního trendu  $\beta_1$  je pomalejší, než přírůstek lineárního trendu. Jedná se o nelineární funkci, kterou není možné linearizovat žádnou transformací, a proto se její parametry odhadují iterativní metodou. Tato metoda je matematický postup založený na hledání nejvěrnějšího obrazu postupnými aproximacemi výchozího, počátečního odhadu
- g) **logistický trend** je uveden ve tvaru Pearlovy-Reedovy trendové funkce  $Tr_t = 1/(\gamma + \beta_0 \times \beta_1^t)$ , jejíž inverzní funkce  $1/Tr_t = \gamma + \beta_0 \times \beta_1^t$  má tvar modifikovaného exponenciálního trendu. Parametry se po inverzní transformaci odhadují stejným způsobem jako pro modifikovaný exponenciální trend
- h) **Gompertzův trend** má tvar  $Tr_t = \gamma \times \beta_0 \times \beta_1^t$ , resp.  $Tr_t^* = \gamma^* + \beta_0^* \times \beta_1^t$ . Křivka má horní asymptotu  $\gamma^* = \ln \gamma$  a vyjadřuje hranici nasycení  $t \rightarrow \infty$ . Parametry původního trendového modelu se po transformaci odhadují jako u modifikovaného exponenciálního trendu nebo jednoduchého exponenciálního trendu

Jedním ze základních účelů modelování časových řad v čase  $t$  je využití těchto modelů, v případě jejich statistické významnosti, k předpovídání budoucího vývoje pomocí extrapolace. Extrapolací se rozumí kvantitativní odhady budoucích hodnot časové řady, které vznikají

prodloužením vývoje z minulosti a přítomnosti do budoucnosti za předpokladu, že se tento vývoj nezmění (Arlt, 2002).

### 3.1.1 Výpočet v programu SAS

Výpočet tradiční dekompozice časové řady není složitý v běžně dostupném softwarovém vybavení, např. Microsoft Excel pomocí jednoduchých postupů, popř. s využitím základních matematických funkcí. Pro dlouhé časové řady a zrychlení procesu výpočtu je však vhodné zvolit sofistikovanější program, např. SAS. Tento statistický software nabízí výpočet tradiční dekompozice časové řady nejen ve svém základním rozhraní SAS Base, ale i v uživatelsky přívětivém („klikacím“) a graficky zpracovaném rozhraní SAS Enterprise Guide. Toto rozhraní je vhodné pro začínající uživatele tohoto softwaru, neboť umožňuje provést výpočet bez nutnosti psaní programovacího kódu. Fungování rozhraní EG je založeno na principu „volání rozhraní SAS Base v pozadí výpočtu“, provedení samotného výpočtu v tomto základním rozhraní a následném převedení výsledků zpět do graficky zpracovaného rozhraní Enterprise Guide.

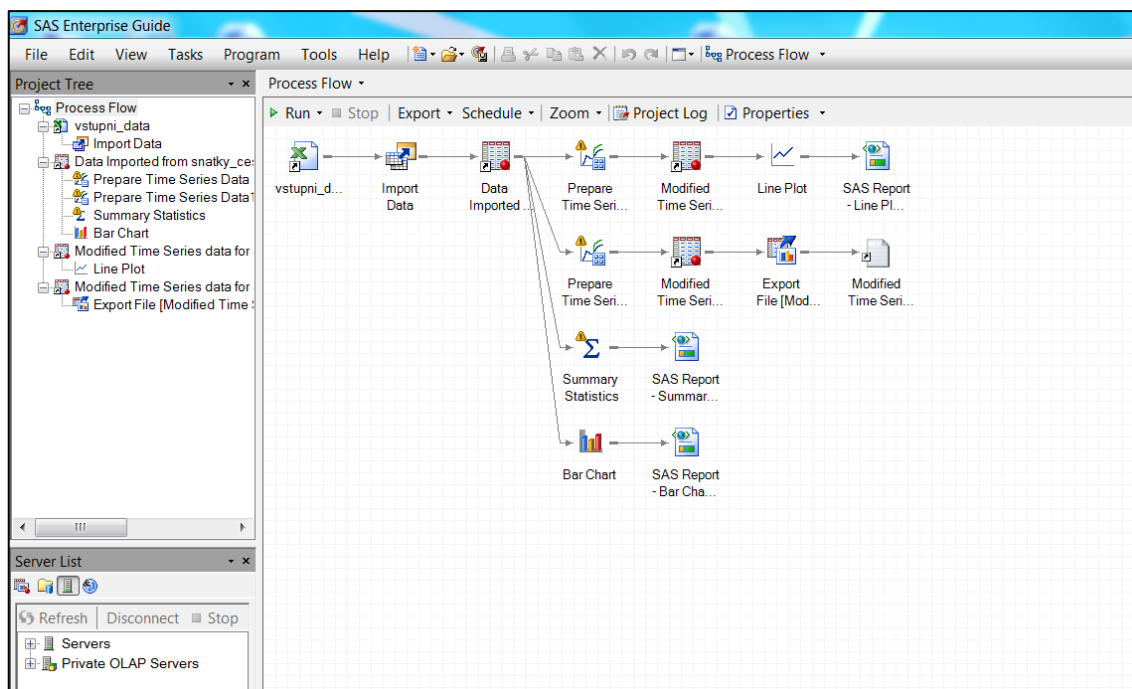
Celý výpočet (procesní tok) se skládá z několika předem nadefinovaných kroků zachycených na obrázku 3 (import dat, aplikace metody, výpočet popisné charakteristiky souboru, vykreslení různých typů grafů a export dat) a může být dle potřeby v jednotlivých krocích zastavován, upravován a opět spouštěn. Zároveň je nadefinovaný proces výpočtu převeden do programovacího kódu, který lze využít v rozhraní SAS Base. Rozhraní Enterprise Guide, však umožňuje i napsat programovací kód, v nové verzi Enterprise Guide 4.3 dokonce pomáhá uživateli pomocí řetězové nabídky procedur a podrobného popisu možností vybrané procedury. Prostředí Enterprise Guide je také vhodné pro proniknutí do pravidelností a zvyklostí psaní programovacího kódu.

Pro výpočet klasické dekompozice časové řady je třeba použít předem standardizovaná data na stejnou délku měsíce, popřípadě i na stejný počet pátků a sobot v měsíci. Na tomto vstupním datovém souboru pak rozhraní Enterprise Guide nabízí výpočet dekompozice pomocí aditivního nebo multiplikativního modelu<sup>4</sup>, a to s volitelnou délkou sezóny (v případě této práce je vždy volena délka sezóny 12). Výpočet je prováděn po jednotlivých složkách, což je oproti rozhraní SAS Base, kde je možný výpočet všech složek najednou, jistou nevýhodou. Z nabídky si tak lze vybrat trendová složku, která zahrnuje i složku cyklickou, sezónní složku, náhodnou složku nebo výpočet sezónně očištěné časové řady. Získané výsledky lze vyexportovat do standardně užívaných formátů (např. CSV, DBF, HTML, XLS) nebo využít možnosti vykreslení výsledků do grafu v prostředí Enterprise Guide a následně export grafu do některého z nabízených formátů (HTML, PDF nebo XML).

Z vygenerovaného kódu vidíme, že software SAS využívá k výpočtu dekompozice časové řady proceduru EXPAND a v záložce *Code* je možné si formát kódu prohlédnout a popř. jej využít pro další výpočty v rozhraní SAS Base.

---

<sup>4</sup> K výpočtu dekompozice časové řady se dostaneme následující cestou: Tasks → Time Series → Prepare Time Series Data → po vybrání časové proměnné a proměnné se standardizovanou časovou řadou → Transformation → Classical decomposition

**Obr. 3 – Procesní tok dekompozice časové řady, SAS Enterprise Guide**

Zdroj: SAS Enterprise Guide 4.2

### 3.2 Metoda X12–ARIMA

První metodou vyvinutou za účelem sezónního očišťování měsíčních a čtvrtletních časových řad v rámci U. S. Census Bureau byla metoda X11. Metoda umožňuje nejen sezónní očištění časových řad, ale také odhad hodnot jednotlivých komponent časové řady. V praxi začala být využívána ve Spojených státech amerických v roce 1965. Během krátké doby metodu převzalo několik dalších statistických úřadů a její výpočet byl integrován do komerčního statistického softwaru, např. SAS a Statistica. Metoda je také obsažena ve volně šiřitelném softwaru Demetra, který pro práci s časovými řadami, především pro jejich sezónní očišťování, používá i Český statistický úřad. Metoda X12–ARIMA vznikla vylepšením metody X11–ARIMA, která zahrnuje možnost aplikace symetrických klouzavých průměrů na časovou řadu s cílem odhadnout trend, sezónní a náhodnou složku časové řady. Obě tyto metody patří do skupiny metod třídy X–11. Kromě již dvou zmíněných patří do této skupiny ještě metoda CENSUS–X11 (Australian Bureau of Statistics, 2011)

Vylepšení metody X12–ARIMA oproti metodě X11–ARIMA spočívá v nových diagnostických možnostech, které umožňují lepší práci s kalendářem a sezónou v časové řadě. Pomocí nových nástrojů k sezónnímu očišťování lze také zvětšit rozsah časových řad, které lze touto cestou očistit (U. S. Census Bureau, 2011). Obě zmíněné metody v sobě zahrnují iterační způsob odhadu jednotlivých složek časové řady s několika cíli, např. vypočítat sezónní indexy, které se v čase mění, vyřadit nebo aproximovat extrémní hodnoty v časové řadě (Grivalská, 2008).

### **Box-Jenkinsova metodologie**

Metoda X12-ARIMA se nejen ve svém názvu, ale i ve svém výpočtu, odkazuje na metodologii, kterou nelze v žádném přehledu možností analýzy časových řad opomenout, Box-Jenkinsovu metodologii, která je základem moderní ekonometrické analýzy časových řad. Tento přístup je založen na myšlence, že časová řada může být chápána jako řada stochastického charakteru (Kozák, 1994). Na jejím základě lze modelovat systematickosti v náhodné, reziduální, složce a tak doplnit dekompoziční analýzu. Soustředí se tedy na tuto zdánlivě nesystematickou složku, která bývá v analýze s využitím dekompozice časové řady brána pouze jako zbytková a důraz je kladen na deterministické části, tedy trendovou a sezónní složku. Na druhé straně je možné každou časovou řadu chápat jako realizaci stochastického (prvek náhody obsahujícího) procesu a Box-Jenkinsova metodologie se pak stává samostatným přístupem k analýze časových řad (Kozák, 1994). Box-Jenkinsova metodologie považuje za základní prvek konstrukce modelu časové řady náhodnou složku, která může být tvořena korelovanými náhodnými veličinami. Mezi modely časových řad, které tato metodologie používá, patří modely klouzavých průměrů (MA), autoregresní modely (AR), smíšené modely (ARMA), speciální také tzv. smíšené integrované modely (ARIMA) a sezónní smíšené integrované modely (SARIMA), (Arlt, 2002).

Mezi typické vlastnosti časových řad patří vzájemná závislost jednotlivých hodnot. Chceme-li u časové řady správně použít metodu nejmenších čtverců pro odhad parametrů matematické funkce popisující trendovou složku a následně vyhodnotit jejich přijatelnost, je nezbytné splnit podmínky tzv. klasického lineárního modelu. Mezi těmito podmínkami je, kromě dalších, i nutnost vzájemné nezávislosti jednotlivých reziduí. Právě tato podmínka není zpravidla u časových řad splněna, což může mít za následek několik nepřesností. Například podhodnocení směrodatných chyb odhadů parametrů, což vede ke zdání, že odhady parametrů funkce popisující trendovou složku jsou přesnější, než je tomu ve skutečnosti. To následně vede ke špatné vypovídací hodnotě testů pro jednotlivé odhadnuté parametry. Proto je nezbytnou podmínkou pro kvalitní modelování deterministických složek časové řady zkoumat také vlastnosti stochastické části (Kozák, 1994). Základním reprezentantem stochastického procesu je proces bílého šumu (viz např. Arlt, 2005).

Box-Jenkinsova metodologie pracuje s několika základními pojmy, které je třeba si před samotnou aplikací tohoto postupu přiblížit. Stochastický proces a stochastický charakter, stejně jako stacionarita časové řady byly již popsány dříve. Nastane-li situace, kdy není zvolená časová řada stacionární, lze tuto řadu na stacionární převést, tedy stacionarizovat, pomocí vhodných transformací. Stacionarizace časových řad však neznámá, že by došlo k odstranění vzájemné závislosti jednotlivých hodnot, pokud je v čase tato závislost přítomna. Pokud je tedy k dispozici konkrétní reálná stacionární (stacionarizovaná) časová řada, lze ji chápat jako závislý výběr z jednorozměrného pravděpodobnostního rozdělení, resp. normálního rozdělení (Kozák, 1994). Dalšími důležitými pojmy v Box-Jenkinsově metodologii jsou autokorelační (ACF) a parciální autokorelační funkce (PACF). Autokorelační funkce vyjadřuje charakter lineární provázanosti stacionárního stochastického procesu (Arlt, 2010). Lze tedy říci, že hodnoty autokorelační funkce zobrazují vztah mezi dvěma sledovanými pozorováními

v časové řadě tak, jak je ovlivněn náhodnými vlivy a pozorováními, která v časové řadě leží mezi sledovanými dvěma (Zatloukal, 2008). Korelace mezi dvěma náhodnými veličinami je často způsobena tím, že obě tyto veličiny jsou korelovány s veličinou třetí. V tomto případě není informace autokorelační funkce o síle závislosti mezi dvěma veličinami dostatečná, proto se přistupuje k určení parciální autokorelační funkce. Tato funkce podává informaci o korelaci dvou sledovaných veličin očištěné o vliv veličin ležících mezi nimi (Arlt, 2010). Obě zmíněné funkce, tedy ACF i PACF, lze vyjádřit pomocí grafického znázornění. Graf těchto dvou funkcí se nazývá korelogram (Kozák, 1994).

Cílem Box-Jenkinsovy metodologie je určení modelu, kterému analyzovaná časová řada svými parametry odpovídá. Podrobnější informace o Box-Jenkinsově metodologii lze nalézt v publikacích o metodách modelování, především ekonomických, časových řad, např. Arlt (2002), Arlt (2005) nebo Cipra (1986).

### 3.2.1 Výpočet v programu SAS

Procedura X12 v softwaru SAS, obdoba programu *U. S. Bureau of The Census X12-ARIMA Seasonal Adjustment*, provádí sezónní očištění měsíčních a čtvrtletních časových řad. Procedura nabízí aditivní a multiplikativní výpočet a součástí výstupů z provedených výpočtů je nejen sezónně očištěná časová řada jako primární výsledek, ale také mnoho dalších, uživatelem definovatelných, složek a charakteristik časové řady. Uživatel má možnost měnit nastavení procedury od základního formátu ve všech krocích analýzy. A to od načtení a rozpoznávání dat, přes práci s proměnnými, jejich modifikaci a kombinace, úpravu kontrolních a informačních nastavení, která určují další směr výpočtu, specifikace hodnot modelů až po výběr jen těch výstupů, v tabulkové nebo grafické podobě, která uživatele v dané situaci zajímají (SAS, 2010).

Zaměříme-li se na výpočet procedury X12 využitý v této práci, lze jej popsat programovacím kódem, a na jeho základě přiblížit průběh výpočtu i slovně. Následující praktický příklad vychází z aplikace metody X12 na evropská data o sňatečnosti v letech 1990–2009. Programovací kód, zobrazený na obrázku 4, pracuje s tabulkou, jejíž název je „*snatky\_1990\_2009*“ a která obsahuje data o šesti evropských zemích, a to ve formátu měsíčních počtů sňatků standardizovaných na stejný počet dní v měsíci a na stejný počet pátků a sobot v měsíci. V první části programovacího kódu dochází k nadefinování kalendáře a doplnění dat o časovou proměnnou, nová verze tabulky je uložena pod názvem „*snatky\_X12\_1990\_2009*“. S touto tabulkou pak pracuje samotná procedura X12, která je v uvedeném příkladu součástí makra. Makro je použito, neboť pomocí něj je možné, podobně jako např. v programu Microsoft Excel, provádět stejný, jednou nadefinovaný výpočet, na datech za více objektů, v tomto případě tedy států. Makro má v tomto případě název „*sezonnost*“ a jeho spouštěním je přes jednoduchý příkaz na jeden řádek spočítána celá procedura X12 pro Česko, Dánsko, Estonsko, Lucembursko, Slovinsko a Španělsko. Samotná procedura X12 pak definuje výpočet pomocí dat z tabulky „*snatky\_X12\_1990\_2009*“, s použitím nadefinovaného kalendáře a s proměnnou „*zeme*“. X11 označuje metodu samotného výpočtu, včetně testování dat na přítomnost sezónnosti. Na dalším řádku jsou vypsány tabulky,

kteřé mají být další, zde nevykreslenou procedurou EXPORT, uloženy do souboru v počítači. Jedná se o 4 tabulky, které obsahují hodnoty pro sezónní proměnlivé indexy (d10), sezónně očištěnou časovou řadu (d11), trendovou složku (d12) a náhodnou složku (d13). Přimo v programu však k těmto výsledkům získáme mnoho dalších charakteristik, jejichž plný výčet by vytvořil dlouhý seznam. Namátkou se jedná o původní časovou řadu, testy sezónnosti, procentuální meziměsíční změnu počtu událostí, logaritmickou transformaci původní i očištěné časové řady, i s vykreslením do grafu nebo projekci sezónních indexů na dalších 12 měsíců.

**Obr. 4 – Ukázka programovacího kódu pro proceduru X12, SAS Base**

```

data snatky_X12_1990_2009;
  set snatky_1990_2009;
  date = intnx( 'month', '01jan90'd, _n_-1 );
  format date monyy.;
run ;

%macro sezonnost (udaje, d, zeme);
  proc x12 data=&udaje date=&d;
    var &zeme;
    x11;
  output out=&zeme d10 d11 d12 d13;
  run;
%mend sezonnost;

%sezonnost (snatky_X12_1990_2009, date, cesko);
%sezonnost (snatky_X12_1990_2009, date, dansko);
%sezonnost (snatky_X12_1990_2009, date, estonsko);
%sezonnost (snatky_X12_1990_2009, date, lucembursko);
%sezonnost (snatky_X12_1990_2009, date, slovinsko);
%sezonnost (snatky_X12_1990_2009, date, spanelsko);

```

Zdroj: SAS Base 9.2

### **Testy sezónnosti**

Procedura X12 má ve svém výpočtu v základním nastavení také testování časové řady na přítomnost sezónnosti, a to pomocí třech jednotlivých testů nebo jedním složeným testem, který je kombinací zmíněných tří. Mezi testy, které SAS na datech v rámci procedury X12 provádí, patří F-test přítomnosti stabilní sezónnosti, neparametrický Kruskal-Wallisův test na přítomnost sezónnosti a F-test přítomnosti proměnlivé sezónnosti, nazývaný Moving Seasonality Test. Souhrnný test se nazývá kombinovaný test přítomnosti identifikovatelné sezónnosti a schéma tohoto kombinovaného testu vidíme na obrázku 5.

### **F-test přítomnosti stabilní sezónnosti**

Tento test je založený na analýze rozptylu dvou proměnných, v případě sezónnosti se jedná o rozptyl sezónní a náhodné složky časové řady. Tyto složky jsou odhadnuty pomocí aplikace klouzavých průměrů, a to v případě měsíčních hodnot časové řady pomocí váženého klouzavého průměru délky 13. Testujeme nulovou hypotézu  $H_0$ : signifikantní stabilní sezónnost

není přítomna v časové řadě, zpravidla na 0,1% hladině významnosti. Testovací kritériem je podíl  $F_s = \frac{\sigma_m^2}{\sigma_r^2}$ , kde  $\sigma_m^2$  je meziměsíční rozptyl, který je tvořen zejména sezónní složkou

a  $\sigma_r^2$  je reziduální rozptyl, který je způsoben zejména náhodnou složkou. Podrobný popis výpočtu rozptylů popisuje např. Grivalská, 2008. Získaný podíl F porovnáme s kritickou hodnotou F-statistiky při požadované hladině významnosti, kde počet stupňů volnosti je  $s-1$  a  $n-s$ , kde  $s$  je počet sezón a nabývá hodnot  $s=1, 2, \dots, s$ ; a  $n$  je počet roků a nabývá hodnot  $n=1, 2, \dots, n$ . S ohledem na skutečnost, že některé základní předpoklady F-testu mohou být porušeny, je obvykle hladina významnosti u tohoto testu volena na 0,1 % (SAS, 2010).

#### ***Kruskal-Wallisův test na přítomnost sezónnosti***

Kruskal-Wallisův chí-kvadrát test je neparametrickou obdobou analýzy rozptylu jednoduchého třídění a k jeho výpočtu se přistupuje ve chvíli, kdy nemůžeme vycházet z předpokladu, že pozorování jsou normálně rozdělená (Hendl, 2006). V rámci procedury X12 nepracuje test s daty původními časové řady, ale vypočítává statistiku s využitím již vypočítaných sezónních indexů (SAS, 2010).

#### ***F-test přítomnosti proměnlivé sezónnosti***

Podobně jako u stabilní sezónnosti je F-test přítomnosti proměnlivé sezónnosti založen na dvou složkách časové řady, sezónní a náhodné, a analýze jejich rozptylů. I nulová hypotéza má obdobné znění, tedy  $H_0$ : signifikantní proměnlivá sezónnost není přítomna v časové řadě.

$F_m = \frac{\sigma_m^2}{\sigma_r^2}$ , kde  $\sigma_r^2$  je meziroční rozptyl, který je způsoben pohyby sezónních odchylek mezi

jednotlivými roky. Oproti F-testu přítomnosti stabilní sezónnosti, ve kterém vysoké hodnoty podílu F poukazovaly na přítomnost měřitelné sezónnosti, v případě tohoto testu vysoké hodnoty podílu F indikují proměnlivou sezónnost, což snižuje pravděpodobnost spolehlivého odhadu sezónních indexů (Grivalská, 2008).

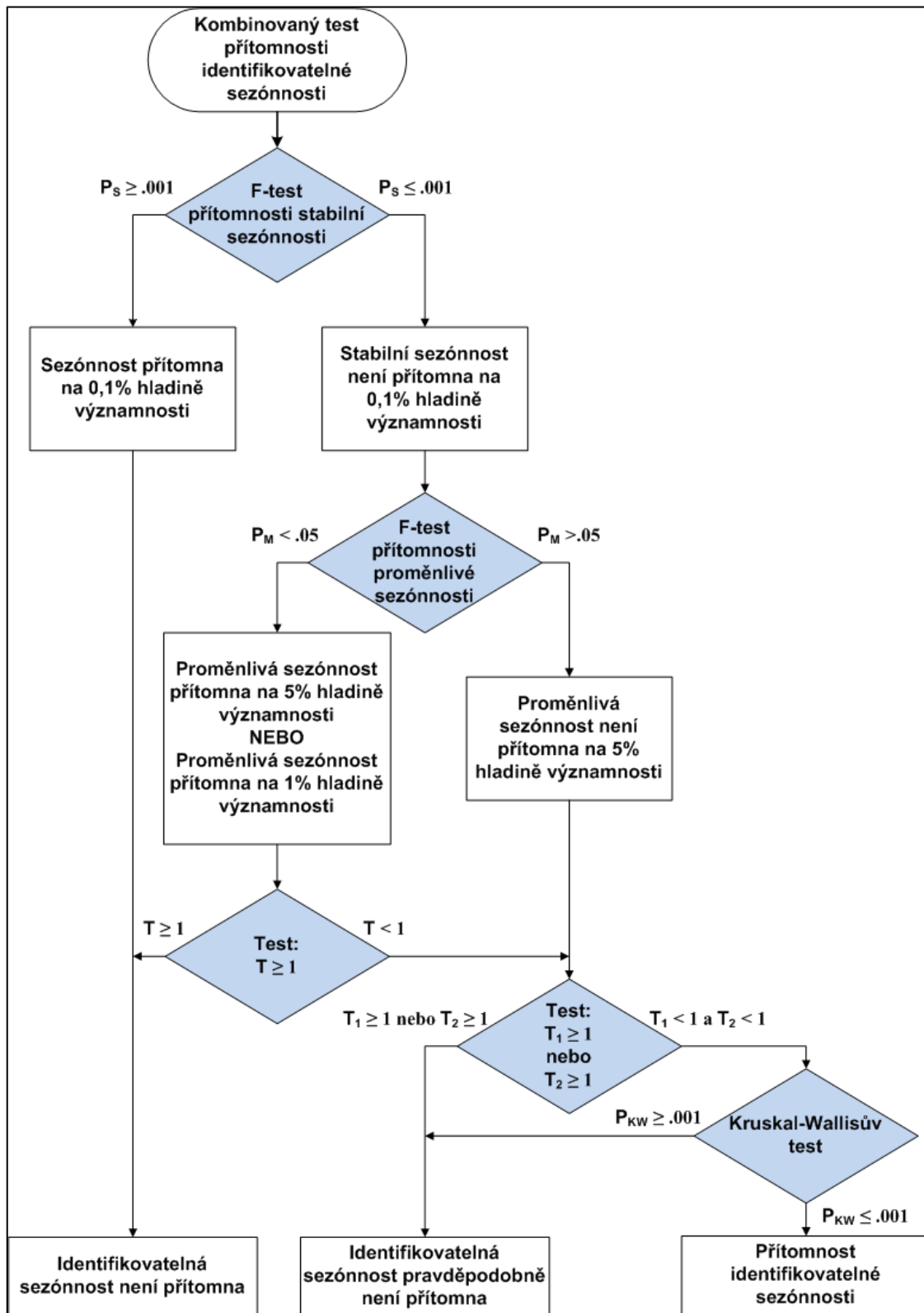
#### ***Kombinovaný test přítomnosti identifikovatelné sezónnosti***

Proces testování dat kombinovaným testem přítomnosti identifikovatelné sezónnosti se sestává z několika logicky za sebou uspořádaných kroků, které testují vždy stejnou nulovou hypotézou  $H_0$ : signifikantní sezónnost není v časové řadě přítomna. V prvním kroku testovacího řetězce jsou data otestována F-testem přítomnosti stabilní sezónnosti. Rozhodneme-li se na základě tohoto testu přijmout nulovou hypotézu  $H_0$ , získáváme výsledek, že v analyzované časové řadě není sezónnost přítomna. Pokud nulovou hypotézu  $H_0$  zamítáme, tedy připouštíme přítomnost sezónnosti v časové řadě, je dalším testovacím krokem F-test přítomnosti proměnlivé sezónnosti. Zároveň s ním vypočítáme hodnoty  $T_1$ ,  $T_2$  a  $T$  pomocí následujících vzorců:

$$T_1 = 7 / (F_m - F_s) \quad T_2 = 3 \times F_m / F_s \quad T = (F_m + F_s) / 2 \quad (\text{SAS, 2010}).$$

V případě, že nepřijmeme nulovou hypotézu  $H_0$  F-testu přítomnosti proměnlivé sezónnosti na 1% a 5% hladině významnosti a vypočtená charakteristika T (aritmetický průměr  $T_1$  a  $T_2$ )

Obr. 5 – Schéma kombinovaného testu přítomnosti identifikovatelné sezónnosti, SAS Base



Zdroj: SAS Institute Inc. 2010. SAS/ETS® 9.22 User's Guide



je větší nebo rovna 1, pak přijímáme původní nulovou hypotézu  $H_0$ , že identifikovatelná sezónnost není v časové řadě přítomna. V případě, že na 1% a 5% hladině významnosti nulovou hypotézu  $H_0$  nepřijímáme a hodnota charakteristiky  $T$  je menší než 1, dostaneme se do fáze testování pomocí vypočtených charakteristik  $T_1$  a  $T_2$ . V této fázi se testování spojuje s výsledkem, kdy na 5% hladině významnosti přijímáme nulovou hypotézu  $H_0$  u F-testu přítomnosti proměnlivé sezónnosti. Jsou-li hodnoty obou ukazatelů  $T_1$  a  $T_2$  větší nebo rovné 1, pak identifikovatelná sezónnost pravděpodobně není přítomna v časové řadě. V případě že hodnoty  $T_1$  a  $T_2$  jsou menší než 1 a Kruskal-Wallisův test nepotvrdí přítomnost stabilní sezónnosti, lze na základě kombinovaného testu tvrdit, že identifikovatelná sezónnost pravděpodobně není v časové řadě přítomna. Pokud jsou obě hodnoty  $T_1$  a  $T_2$  menší než 1 a Kruskal-Wallisův test přítomnost stabilní sezónnosti potvrdí, můžeme konstatovat, že výsledkem kombinovaného testu přítomnosti signifikantní sezónnosti je přítomnost této sezónnosti v časové řadě (SAS, 2010; Grivalská, 2008).

### 3.3 Kointegrace časových řad

Při modelování vícerozměrných časových řad je účelné rozlišovat mezi krátkodobými a dlouhodobými vztahy. Krátkodobý vztah mezi časovými řadami existuje pouze v relativně krátkém období, s postupem času tyto vztahy mizí. Druhý typ vztahů má dlouhodobé trvání, a tedy s postupujícím časem nemizí. Problematika dlouhodobých vztahů velice úzce souvisí s pojmem ekvilibrium, rovnovážný stav (popř. stabilní ekvilibrium, tedy rovnovážný stav, který se v čase nemění). Můžeme jej chápat jako stav, ke kterému je systém neustále přitahován. Protože je systém vystaven neustálým šokům, v ekvilibriu se nenachází nikdy, může však být v dlouhodobém ekvilibriu, tedy ve stavu, který k rovnovážnému stavu konverguje v čase (Arlt, 1997).

Procesy s dlouhou pamětí, které byly popsány dříve v této práci, se nazývají integrovanými procesy řádu jedna a označují se jako  $I(1)$ . První diferencí se tyto procesy transformují na procesy s krátkou pamětí (Arlt, 1997). První diference označuje proces převedení nestacionární časové řady na řadu stacionární (Arlt, 2010). Obecně, procesy, které se po  $d$ -té diferenci transformují na procesy s krátkou pamětí, se nazývají integrovanými procesy řádu  $d$  a značí se jako  $I(d)$ . Z toho plyne, že procesy s krátkou pamětí se nazývají integrovanými procesy řádu nula a značí se jako  $I(0)$ . Na základě pravidel vztahů jednotlivých integrovaných procesů (uvedených v Arlt, 1997), kdy existuje lineární kombinace dvou časových řad, můžeme určit kointegrační vektor, který popisuje tento vzájemný vztah. V případě dvou procesů může existovat pouze jeden kointegrační faktor (proces), existuje tedy pouze jedna jejich lineární kombinace, která je stacionární. Obecně pak platí, že máme-li  $n$  časových řad, může být maximálně  $n-1$  kointegračních procesů (Arlt, 1997).

Aby bylo možné se prakticky zabývat kointegrací, je třeba nejprve zjistit, jakého typu jsou časové řady vstupující do analýzy. Existuje několik způsobů jak zjistit řád kointegrace časové řady. První možností je subjektivní posouzení na základě vykreslení grafu časové řady.

Druhou jednoduchou metodou, opět subjektivního charakteru, je posouzení tvaru ACF analyzované časové řady. Je-li první hodnota této funkce blízká jedné a ostatní hodnoty klesají jen velmi pomalu, lze očekávat, že daná řada nebude typu  $I(0)$ . V případě, že subjektivní metody nejsou dostačující pro zjištění typu časové řady, je třeba použít přesnější metody. Existuje několik statistických testů pro zjištění řádu integrace, označují se jako testy jednotkových kořenů. Nejpoužívanějším z nich je pak Dickey-Fullerův test, který porovnává naměřené hodnoty s hodnotami kritickými a určuje tak, zda je časová řada typu  $I(0)$  nebo  $I(1)$  (Arlt, 1997).

Existují nejméně tři závažné důvody, proč lze považovat princip kointegrace za ústřední myšlenku modelování integrovaných časových řad:

- Stacionární lineární kombinaci integrovaných (nestacionárních) časových řad je možné chápat jako odhad ekvilibria, které spojuje uvažované časové řady. Ekvilibrium je v tomto případě střední hodnota stacionárního generujícího procesu, jeho odhadem je pak průměr hodnot složené stacionární časové řady.
- Regrese obsahující integrované časové řady má smysl pouze tehdy, jsou-li tyto časové řady kointegrované. Test kointegrace časových řad je tedy současně metoda pro odlišení mezi pravou regresí a regresí zdánlivou.
- Skupinu kointegrovaných časových řad lze, kromě jiných modelů, popsat modelem „error-correction“, jehož prostřednictvím je možné odlišit dlouhodobé a krátkodobé vztahy mezi časovými řadami. Tento model obsahuje parametry charakterizující míru vychýlení systému od dlouhodobě se prosazujícího ekvilibria. Tato skutečnost je významná jak sama o sobě, tak také může být i prostředkem pro řešení sporu mezi dvěma přístupy modelování nestacionárních časových řad, statistickým a ekonometrickým. Umožňuje totiž spojit metody statistické, tradičně spočívající ve zkoumání vlastností diferencovaných časových řad a metody ekonometrické, které kladou důraz na ekvilibrium časových řad, a proto se zajímají o jejich úrovně analyzování. Použití obou přístupů izolovaně je problematické, neboť statistický přístup se zbavuje důležitých informací obsažených v původních nestacionarizovaných časových řadách a ekonometrický přístup má tendenci přehlížet problém zdánlivé regrese (Arlt, 1997).

V demografických časových řadách a jejich analýze lze najít vztahy jak krátkodobé, tak dlouhodobé. Krátkodobé vztahy by mohly být identifikovány v případě porovnávání teplotních výkyvů a výkyvů v úmrtnosti, neboť lze očekávat, že při velkém výkyvu teploty může dočasně dojít k okamžitému nárůstu/poklesu počtu zemřelých osob a po návratu teploty ovzduší k hodnotám dlouhodobého normálu také návrat počtu zemřelých osob k průměrným hodnotám. Tento vztah byl již zkoumán s využitím korelační analýzy (Myšáková, 2009). Pomocí kointegrace časových řad bude provedena analýza v této práci na empirických datech za Česko. Analýzou dalších demografických procesů, porodnosti a sňatečnosti, nebo porodnosti

a kojenecké úmrtnosti, lze hledat dlouhodobé vztahy, které budou čitelné mezi dvěma časovými řadami po celé sledované období.

### 3.3.1 Výpočet v programu SAS

Výpočet kointegrace časových řad nabízí v prostředí rozhraní SAS Base procedura VARMAX. Součástí této procedury je i možnost určení stacionarity/nestacionarity pomocí testu jednotkového kořene, Dickey-Fullerova testu. V případě, že jsou vybrané časové řady nestacionární, tedy mají jednotkový kořen, je možné přistoupit ke hledání kointegrovaného procesu, je-li mezi analyzovanými řadami přítomen. To je možné s využitím např. Johansenova testu kointegrace. Pomocí něj testujeme nulovou hypotézu  $H_0$ : mezi řadami není žádný kointegrovaný proces. V případě zamítnutí této nulové hypotézy  $H_0$  postupuje test k alternativní hypotéze, která tvrdí, že mezi časovými řadami je alespoň jeden kointegrovaný proces. V případě, že jsou časové řady více než dvě, pokračuje testování k další nulové hypotéze  $H_0$ , která říká, že mezi analyzovanými časovými řadami je právě jeden kointegrovaný proces a další alternativní hypotéze, která říká, že kointegrovaných procesů je mezi řadami více (SAS, 2010).

Podobu jednoduchého programovacího kódu znázorňuje příklad na obrázku 6, který zachycuje testování kointegrace u dvou časových řad, nazvaných „porodnost“ a „úmrtnost“, které vyjadřují počet živě narozených za měsíc a počet zemřelých do jednoho roku věku za měsíc. Nejprve jsou data otestována Dickey-Fullerovým testem a následně Johansenovým testem kointegrace.

**Obr. 6 – Ukázka programovacího kódu pro proceduru VARMAX, SAS Base**

```
proc varmax data=kointegrace;  
  model porodnost umrtnost/p=1 dftest noint cointest=(johansen);  
run;
```

Zdroj: SAS Base 9.2

## 3.4 Další metody pro analýzu sezónnosti

Metody analýzy časových řad uvedené v předcházejících podkapitolách nejsou jedinými metodami, které je možné při studiu chování časové řady využít. Jedná se však o metody, které jsou nejvíce rozšířené a především využitelné v oblasti demografie. Následující tři metody stručně přibližují další možnosti v přístupu k analýze časových řad, především se zaměřením na detekci a analýzu sezónnosti v časové řadě.

### 3.4.1 Grafická analýza

Zatímco většinu zde uvedených metod analýzy časových řad lze vypočítat pomocí sofistikovaného softwarového vybavení, grafickou analýzu časových řad lze provést i bez něj,

tedy např. s využitím běžně dostupného programu Microsoft Excel. Jedním ze základních prostředků prezentace časových řad je jejich vizuální zobrazení ve formě grafu. Původní hodnoty časové řady, nebo jejich kumulaci, lze vykreslit pomocí široké nabídky grafů. Na časovou řadu je vhodné použít spojnicový graf znázorňující vývoj v čase. Z něj lze pak subjektivním zhodnocením určit, zda je vybraná časová řada zajímavá například z pohledu analýzy trendu nebo sezónnosti a na základě prvotní informace o časové řadě určit postup a vhodnou metodu analýzy. Často se ale časové řady zobrazují i tak, aby více vynikly jejich charakteristické vlastnosti a rysy, k tomu pak slouží speciální typy grafů, např. krabičkový graf, nazývaný také box-plot (Arlt, 2002).

### 3.4.2 Model SARIMA

Model SARIMA označuje sezónní smíšené integrované modely Box-Jenkinsovy metodologie. Obdobně jako v případě trendu, rozlišujeme i sezónní složku na stochastickou a deterministickou. Především u sezónních ekonomických časových řad má sezónní složka stochastický charakter a je často zkorelována s ostatními složkami časové řady (Kozák, 1994). Myšlenka tzv. sezónního procesu navazuje na klasický model ARIMA, ve kterém předpokládáme vzájemnou závislost mezi veličinami. Protože však tento proces obsahuje ještě sezónní kolísání, lze očekávat i závislost mezi sobě odpovídajícími veličinami (hodnotami) v jednotlivých sezónách. Délka sezónní periody je označována jako  $s$  a udává počet sezón během roku, tedy pro měsíční časovou řadu  $s=12$ , pro čtvrtletní časovou řadu se  $s=4$  (Kozák, 1994).

### 3.4.3 Lineární dynamické modely

Lineární dynamické modely zobecňují Box-Jenkinsovu metodologii. Dříve uvedené modely Box-Jenkinsovy metodologie se snažily vysvětlit chování časové řady jen na základě jí samotné, tedy na „historických“ záznamech řady, na stejném principu i konstruuji tyto modely předpovědi do budoucna. Tento typ předpovědi označujeme jako autoprojektivní. Lineární dynamické modely pracují s jiným přístupem k časové řadě, kdy dávají chování analyzované časové řady do vztahu k chování jiných časových řad, které jsou následně i použity při odhadu modelu pro původní řadu. Lineární dynamické modely lze různými způsoby modifikovat a je tak možné akceptovat sezónnost časových řad (Cipra, 1986).

### 3.4.4 Spektrální analýza

V analýze časových řad existují v podstatě dva přístupy, analýza v časové doméně a analýza ve spektrální doméně. Zatímco všechny doposud zde představené metody analýzy časových řad využívají první přístup, tedy závislost jednotlivých složek na čase (dekompozice časové řady) nebo závislost ACF na časovém zpoždění (Box-Jenkinsova metodologie), spektrální analýza využívá přístupu druhého. Ten považuje analyzovanou časovou řadu za nekonečnou kombinaci sinusových a kosinusových křivek s různými amplitudami a frekvencemi. Někdy se tato metoda

také označuje jako fourierovská analýza. Pomocí spektrálních statistických nástrojů, tzv. periogramu a spektrální hustoty, lze při této analýze získat obraz o intenzitě zastoupení jednotlivých frekvencí v časové řadě (o spektru řady) a pomocí něj pak určit frekvence, které jsou v dané řadě zastoupeny nejvýrazněji. Spektrální analýza je také vhodná při srovnání chování několika řad, neboť umožňuje nalézt případné časové zpoždění mezi dvěma analyzovanými řadami. Pomocí spektrální analýzy lze také posuzovat, jak kvalitní bylo sezónní očištění vybrané časové řady (Cipra, 1986).

### 3.5 Shluková analýza

Metody shlukové analýzy se vyvinuly z potřeby analyzovat informaci obsaženou v datech, která jsou generována množinou objektů, o jejíž struktuře toho víme jen málo. Strukturou se přitom myslí rozdělení objektů do určitého systému kategorií, jež zachycuje podobnost objektů patřících do téže kategorie na jedné straně a nepodobnost objektů patřících do různých kategorií na straně druhé. Hledáme tedy jakési „přirozené skupiny“, k nimž nám metody shlukové analýzy poskytují algoritmy, které toto hledání provádějí automaticky za pomoci počítačů. Algoritmy můžeme použít k měření starých klasických kategorizačních struktur nebo k nalézání nových (Hendl, 2006).

Shlukovou analýzu lze označit jako skupinu metod, které se liší podle zvoleného výpočtu podobnosti a odlišnosti nebo podle různých postupů seskupování jednotek do shluků. Při aplikaci shlukové analýzy nesnižujeme počet jednotek ani počet proměnných, pouze jednotlivé jednotky shrnujeme do skupin. Shluková analýza jako vícerozměrná statistická metoda splňuje několik požadavků: požadavek nepřekrývání tříd, požadavek úplnosti klasifikace, požadavek jednotného kritéria klasifikace a skladebnost jednotek do jednotek vyššího či nižšího řádu (Heřmanová, 1991). Pro bezproblémovou aplikaci shlukové analýzy na empirická data je třeba, aby tato data splňovala několik podmínek: vzájemně nezávislé a nekorelované proměnné, bezrozměrná čísla, tedy nezávislost na jednotkách měření (v případě, že data tuto podmínku nesplňují, je třeba je nejprve standardizovat a převést tak do normovaného tvaru) a stejnou významnost použitých charakteristik. Pokud není splněna podmínka stejné významnosti, lze data do této roviny přetransformovat aplikací vah (Dzúrová, 2008).

Shlukovací metody jsou většinou založeny na výpočtu měr nepodobnosti (resp. podobnosti) objektů a shluků, jež odrážejí naše požadavky. Jednou z nejpoužívanějších měr nepodobnosti je euklidovská vzdálenost, dále se používá vzdálenost nejbližšího souseda, vzdálenost nejvzdálenějšího souseda nebo průměrná vzdálenost mezi sousedy (Hendl, 2006). Vždy je tedy nutné, aby do analýzy vstupovala množina o alespoň dvou proměnných, aby bylo možné vypočítat vzájemnou vzdálenost. Obvyklý je však vyšší počet proměnných, neboť pro určení navzájem si podobných skupin je třeba několikanásobně více proměnných, než je požadovaný počet shluků. Počet shluků můžeme určit buď subjektivně, např. víme-li

předem, kolik shluků chceme vytvořit, nebo na základě některého z testů kritérií pro určení počtu shluků (viz SAS, 2010).

Shlukování se provádí pro omezený počet objektů. V případě, že nalezneme optimální počet a systém shluků pro výběrovou množinu objektů, vzniká otázka, jak tento výsledek přenést na celou populaci, ze které byly objekty vybrány. V této fázi analýzy je třeba uplatnit klasické prostředky matematické statistiky, jako jsou třeba metody testování hypotéz. Podle cíle můžeme rozeznávat tři druhy úloh shlukové analýzy (Hendl, 2006):

- Cílem je nalezené předem definovaného množství shluků.
- Cílem je nalezení množiny shluků, přičemž jejich počet není specifikován.
- Cílem je vytvořit hierarchický strom.

### 3.5.1 Výpočet v programu SAS

Výpočet shlukové analýzy se v prostředí softwaru SAS provádí v základním rozhraní SAS Base např. s využitím procedury CLUSTER (v rámci softwaru je možné využít i další procedury, které počítají shlukovou analýzu). Zvolená procedura CLUSTER nabízí možnost zvolit pro výpočet jednu z 11 shlukovacích metod, z nichž všechny využívají hierarchické shlukování. Při takovém postupu shlukování vstupuje do analýzy každé pozorování samo za sebe, tedy počáteční počet shluků odpovídá počtu pozorování. V každém kroku analýzy jsou pak dva nejbližší pozorování (shluky) spojeny v jeden a jako jeden společný také pokračují do následujícího kroku. Postupným shlukováním tak získáváme stále menší počet clusterů, které se v posledním kroku analýzy spojí v jeden všeobsahující shluk (cluster). Posouzení vhodného počtu výsledných clusterů je možné buď subjektivně nebo pomocí testů (např. pseudo F statistic).

Získané výsledky je možné znázornit do tabulky, lze tak vypsát např. příslušnost daného pozorování k určitému clusteru nebo vzdálenost (euklidovskou, průměrnou) mezi dvěma naposledy seskupenými clustery. Ukazatelů, které je možné nechat vypsát do souboru je mnohem více a jejich výběr se odvíjí od zvolené metody shlukování (viz SAS, 2010). Další metodou zobrazení získaných výsledků je vykreslení grafu, který se nazývá dendrogram. Pro vykreslení tohoto grafu je třeba použít proceduru TREE, která vypočítané charakteristiky shlukové analýzy převede do grafické podoby. Procedura nabízí opět několik možností, např. výběr horizontální nebo vertikální orientace grafu a pro jeho vykreslení je třeba předem určit počet clusterů.

Ukázka programovacího kódu pro výpočet a vykreslení výsledků shlukové analýzy je znázorněna na obrázku 7. Tabulka, se kterou se od začátku shlukové analýzy pracuje, se jmenuje „*indexy*“ a obsahuje sezónní indexy vypočítané pomocí dekompozice časové řady pro 13 evropských států pro období let 2001–2009. Tyto státy jsou jako pozorování uvedené v řádcích a hodnoty sezónních indexů jsou po jednotlivých měsících uvedeny ve sloupcích, tedy jako proměnné. Výpočet shlukové analýzy se ukládá do tabulky, která nese název „*tree*“. Tento krok je nezbytný pro aplikaci procedury TREE, která z uvedené tabulky vychází

při vykreslování dendogramu. Pro jeho vykreslení je nejprve nutné spustit grafické prostředí Output Delivery System (ods graphics on). Jako v případě jiných grafů je možné nadefinovat celou škálu formátování, např. velikost písma, velikost názvů, které jsou v ukázce uvedené, ale také třeba barvu písma, barvu grafu, barvu pozadí. V příkladu znázorněném na obrázku 7 je zvolena horizontální orientace dendogramu, vykreslení pro 3 clustery, a to podle zemí. Jako titulek grafu je nastaven text „Dendogram vybraných evropských zemí podle rozložení počtu snatku během roku, 2001–2009“.

**Obr. 7 – Ukázka programovacího kódu pro proceduru CLUSTER a proceduru TREE, SAS Base**

```
proc cluster data=indexy outtree=tree method=average pseudo;  
var leden--prosinec;  
id zeme;  
run;  
  
ods graphics on;  
goptions htext=2pct htitle=3pct;  
proc tree data=tree out=new nclusters=3 horizontal;  
id zeme;  
title "Dendogram vybraných evropských zemí podle rozložení  
počtu snatku během roku, 2001–2009";  
run;
```

Zdroj: SAS Base 9.2

### 3.6 Prezentace výsledků

V jednotlivých kapitolách o demografických procesech jsou přiloženy grafy, které ilustrují některé získané výsledky. Ne všechny použité ukazatele v těchto grafech jsou však vhodné k vzájemnému porovnávání jednotlivých států. Těmito ukazateli jsou odhad trendu a sezónně očištěná časová řada. Pro vzájemné srovnání nejsou vhodné z toho důvodu, že vyjadřují absolutní počet událostí během jednotlivých měsíců. Tyto počty nejsou dále nijak upraveny, standardizovány, a jsou tedy ovlivněny jak populační velikostí jednotlivých států, tak i věkovou strukturou jednotlivých populací.

Pro vzájemné porovnávání států, které do analýzy vstupují, jsou vhodné jiné složky časové řady a ukazatele, které získáme aplikací popsaných metod. Jedná se o sezónní indexy, proměnlivé sezónní indexy a náhodnou složku. Tyto tři ukazatele jsou vyjádřeny bezrozměrnými čísly, a proto jsou vhodné pro vzájemné porovnávání.

Všechny publikované výsledky v následujících kapitolách reprezentují časové řady, u kterých byla jednotlivými testy sezónnosti přítomnost sezónnosti potvrzena.

## 4 Porodnost

Analýza sezónnosti časových řad vypovídajících o porodnosti (o počtu narozených celkem, živě narozených) může podrobněji odhalit některé skutečnosti o jejich vývoji, které jsou již všeobecně známé díky analýze ročních časových řad. Takovým příkladem může být nárůst porodnosti v návaznosti na různá populační opatření. Po výpočtu jednotlivých složek časové řady můžeme například vidět, kolik měsíců po zavedení nějakého opatření se zvýšila porodnost, pomocí sezónního indexu určit kolikrát se v některém z exponovaných měsíců hodnota počtu narozených zvýšila oproti hodnotě průměrného měsíce nebo o jak výrazný, oproti průběhu časové řady nesystematický, výkyv se jednalo.

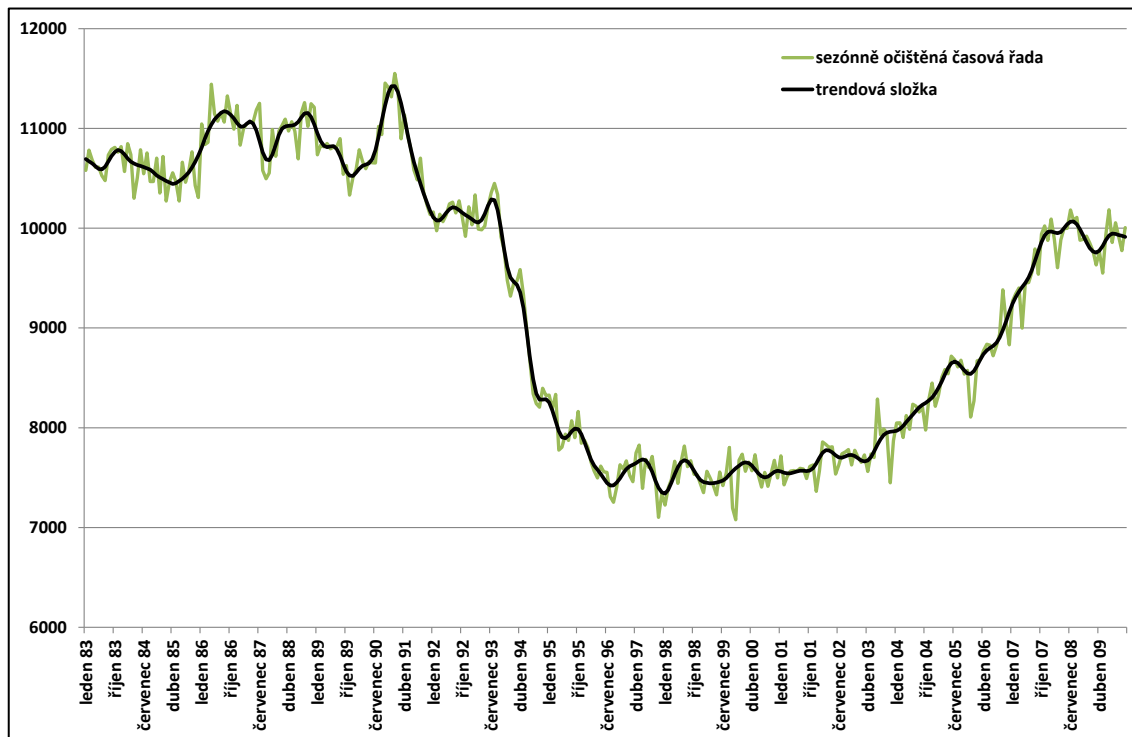
### 4.1 Česko

Statistická data o měsíčním počtu živě narozených za Česko jsou v této práci analyzována od kalendářního roku 1983. Z grafu na obrázku 8 vidíme průběh sezónně očištěné časové řady a křivky trendové složky během celého období. Zobrazený průběh křivek lze rozdělit do tří fází. První fází je období od počátku roku 1983 do poloviny roku 1993, kdy je srpnová hodnota na dlouhou dobu poslední nad hranicí 10000 živě narozených dětí za měsíc. Toto porevoluční období se, podobně jako v ostatních bývalých socialistických zemích, stále ještě vyznačovalo dvoudětným modelem rodiny (Bartoňová, 2008). Následná fáze se vyznačuje prudkým poklesem a stagnací na velmi nízké úrovni počtu živě narozených, což souvisí se změnami v reprodukčním chování, kdy je typ univerzální rodiny nahrazován několika modely životních drah (Bartoňová, 2008). Tyto změny byly způsobeny především novými možnostmi v kariérním životě žen i mužů a s tím spojeným odkladem rodičovství do vyššího věku, menším počtem dětí v rodině nebo až dobrovolnou volbou bezdětnosti. Takováto změna chování moderních populací, kdy se do popředí dostává individualismus a budování kariéry, se označuje jako druhý demografický přechod. Od roku 2003 pak můžeme pozorovat třetí fázi, která se vyznačuje mírným nárůstem a v posledních letech opět stagnací. Tento postupný nárůst počtu živě narozených dětí není způsoben návratem k tradičnímu modelu rodiny, ale jedná se o přechodnou fázi, kdy se do reprodukčního věku dostávají silné ročníky žen narozené v 70. letech minulého století. Tato silná kohorta, která se s přelomem tisíciletí dostala do věku, kdy dnešní ženy nejčastěji rodí děti, tak vyvolala nárůst porodnosti, v masmédiích často označovaný jako „baby boom“ (Bartoňová, 2008).



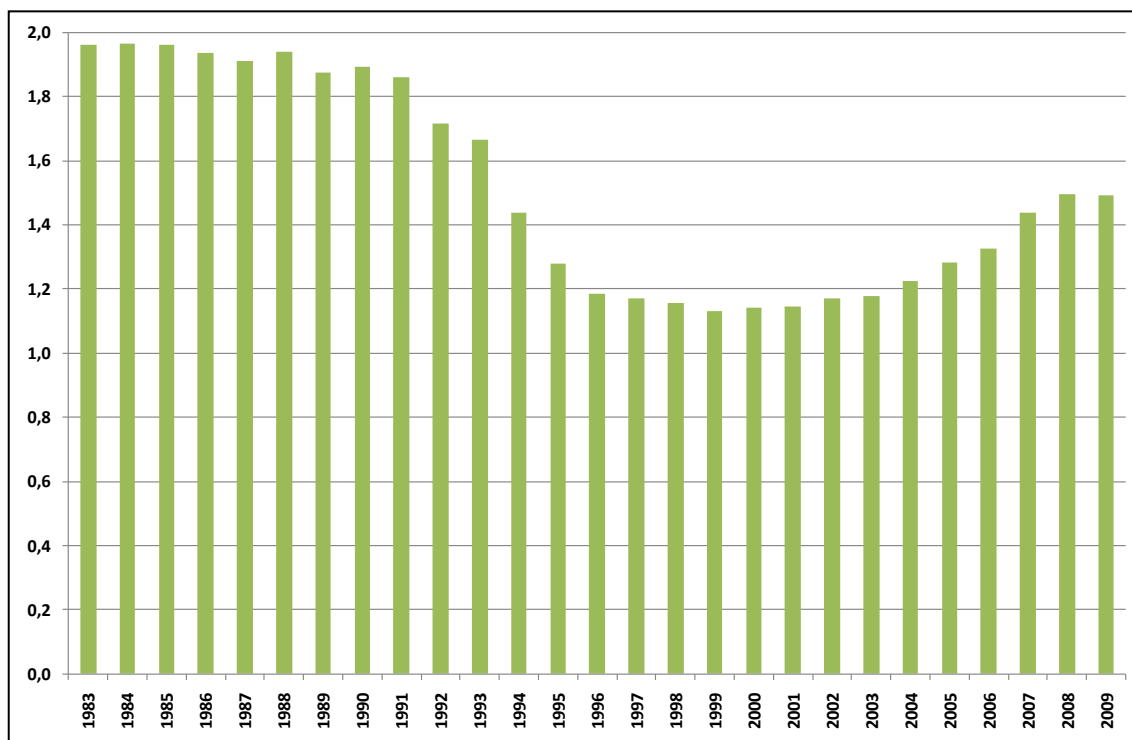
Následující dva grafy na obrázcích 8 a 9 znázorňují paralelní vývoj mezi počtem živě narozených dětí a trendem ve vývoji tohoto ukazatele s hodnotami úhrnné plodnosti.

**Obr. 8 – Sezónně očištěná časová řada, trendová složka, živě narození, Česko, 1983–2009**



Zdroj: ČSÚ, vlastní výpočty

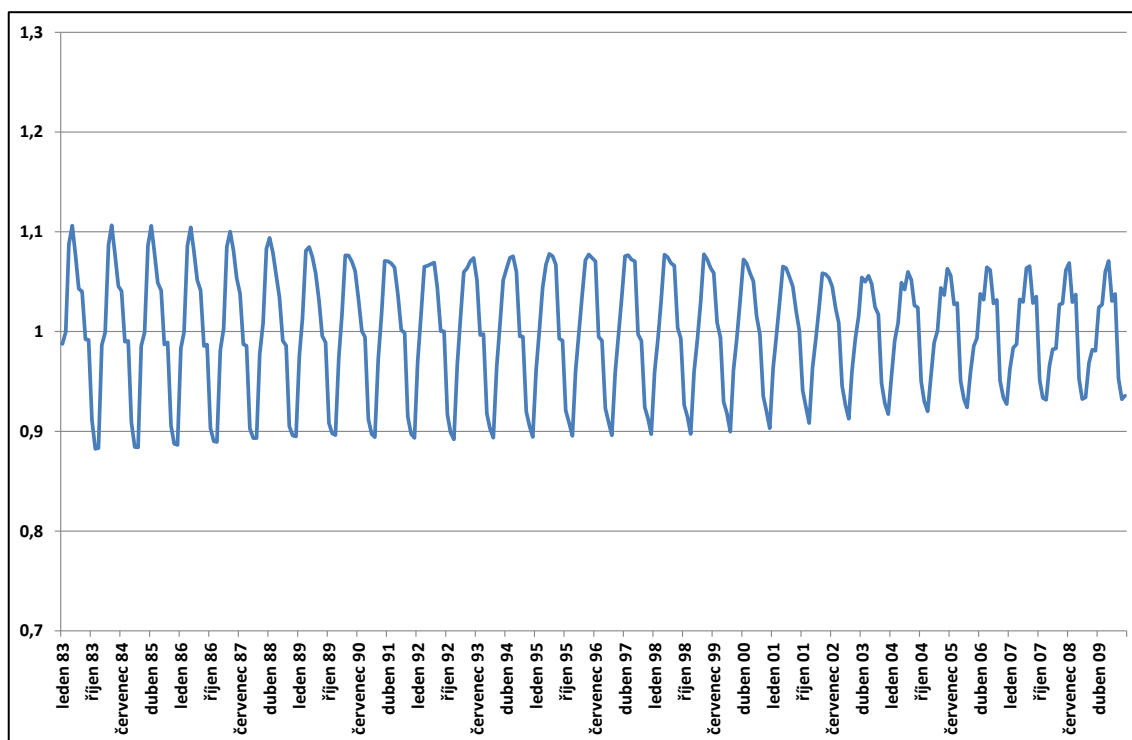
**Obr. 9 – Hodnoty úhrnné plodnosti, Česko, 1983–2009**



Zdroj: ČSÚ

Porodnost se, stejně jako ostatní demografické procesy, vyznačuje sezónností, i když ne v takové míře. Přestože existují rozdíly mezi jednotlivými měsíci v hodnotách příslušných sezónních indexů, není rozpětí těchto hodnot tak široké jako například u sňatečnosti. Proměnlivé sezónní indexy pro Česko ve sledovaném období, znázorněné na obrázku 10, ilustrují změnu v rozložení živě narozených během roku. Během analyzovaných 26 let vidíme nejen zúžení už tak úzkého rozpětí mezi minimální a maximální hodnotou sezónního indexu, ale také obměnu měsíců, které tyto hodnoty vykazují. Zatímco minimální hodnoty se nepravidelně střídají u měsíců listopad a prosinec, maximální hodnoty se proměňují během celého období. Do roku 1989 byl nejčastějším měsícem porodu, s výrazným nárůstem, duben, od roku 1990 se po dobu 14 let střídala nejvyšší hodnota sezónního indexu mezi březnem, dubnem a květnem. Od roku 2004 je pak maximální hodnota indexu vykazována vždy pro měsíc červen.

**Obr. 10 – Proměnlivé sezónní indexy, živě narození, Česko, 1983–2009**



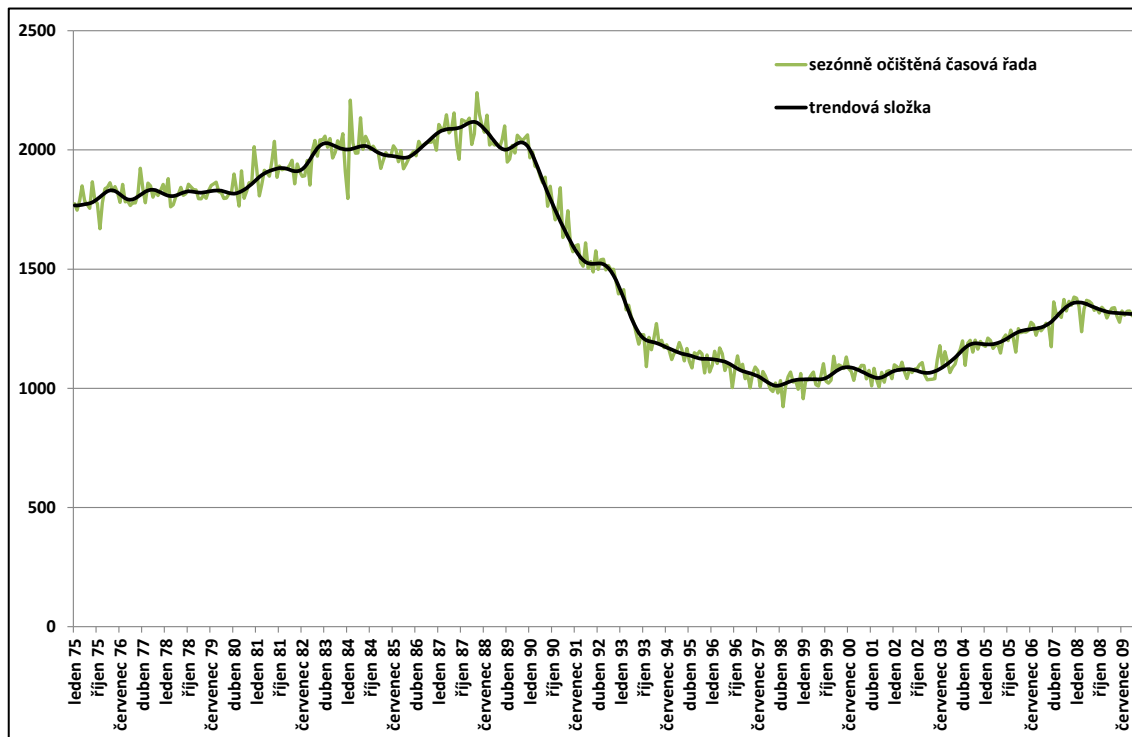
Zdroj: ČSÚ, vlastní výpočty

## 4.2 Evropské státy

Za některé evropské státy jsou dostupné delší časové řady o počtu narozených, ať už celkem narozených nebo živě narozených, než pro Česko. Analyzovány tak byly časové řady pro období let 1975–2009. Jedná se tedy o dostatečně dlouhý časový úsek, který umožňuje u východoevropských zemí, které v roce 1989 prodělaly politickou revoluci, porovnat dvě odlišná období, éru socialistickou a éru demokratickou. A vývoj v celém období také zároveň porovnat se zeměmi, které takovou změnou neprošly a vývoj porodnosti tedy nebyl

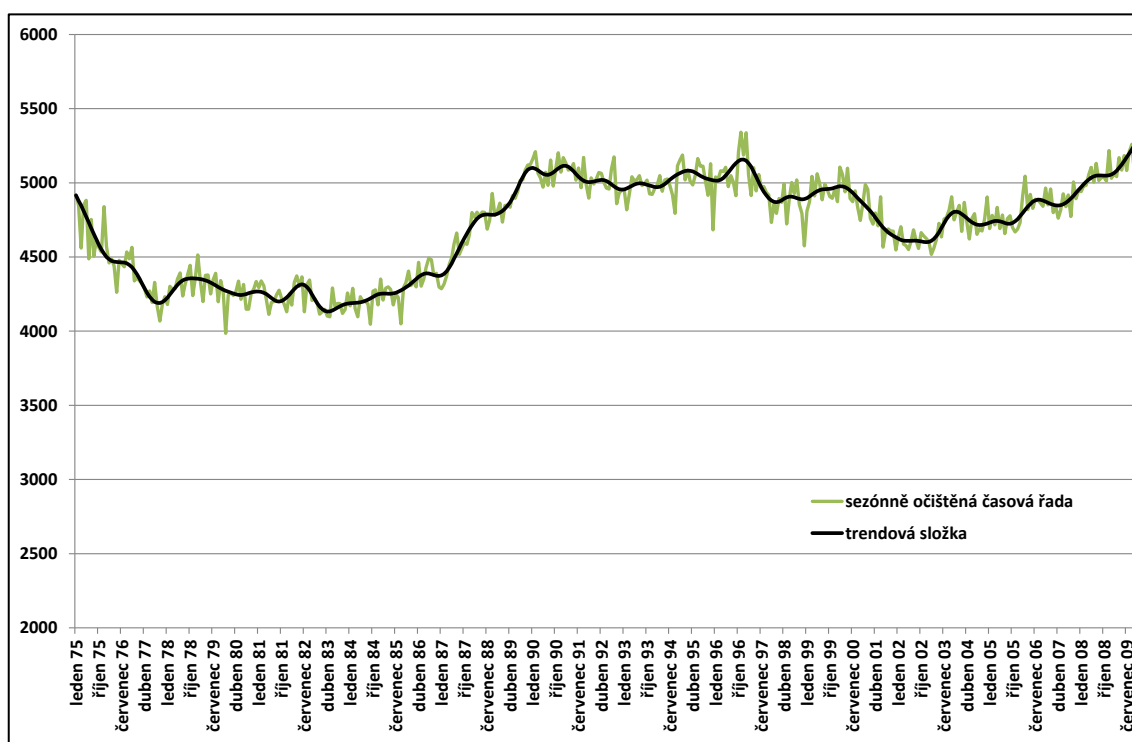
závislý na radikálních změnách v politickém uspořádání, ve způsobu života, možnostech cestování nebo v možnostech budování kariéry.

**Obr. 11 – Sezónně očištěná časová řada, trendová složka, živě narození, Estonsko, 1975–2009**



Zdroj: SE, vlastní výpočty

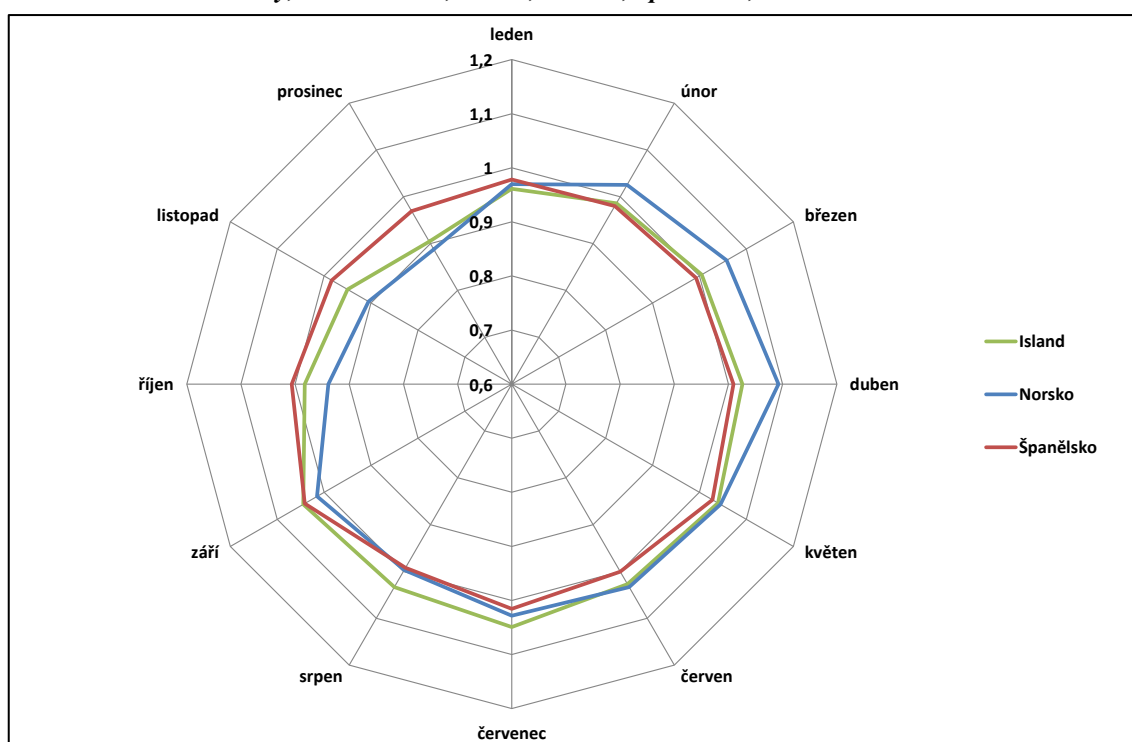
**Obr. 12 – Sezónně očištěná časová řada, trendová složka, živě narození, Norsko, 1975–2009**



Zdroj: SN, vlastní výpočty

Grafy na obrázcích 11 a 12 znázorňují vývoj počtu živě narozených dětí a křivku trendu tento vývoj popisující. Estonsko, znázorněné na obrázku 11, je zástupcem východoevropské země, která byla po dlouhá desetiletí součástí Sovětského svazu a s rokem 1989 tak přišly kromě demokratizace i nové možnosti realizace jedince. Podobně jako v Česku, dochází i v Estonsku, k poklesu úrovně porodnosti, i když zde je pokles zaznamenán již od prvního roku 90. let, dá se tedy předpokládat, že prudká změna poměrů začala v Estonsku dříve než v Česku. Ke konci sledovaného období nedochází v Estonsku k takovému nárůstu počtu narozených, pravděpodobně z důvodu absence početných kohort rozených v Česku v důsledku propopulačních opatření. Odlišný vývoj porodnosti sledujeme v Norsku, které je zobrazeno na obrázku 12. Zde naopak s počátkem 90. let dochází k mírnému nárůstu měsíčních počtů živě narozených dětí. Tato situace odpovídá již proběhnutému druhému demografickému přechodu a návratu k vyšší plodnosti, kterou dnes vykazují především země severní a některé státy západní Evropy (např. Francie).

**Obr. 13 – Sezónní indexy, živě narození, Island, Norsko, Španělsko, 1975–2009**



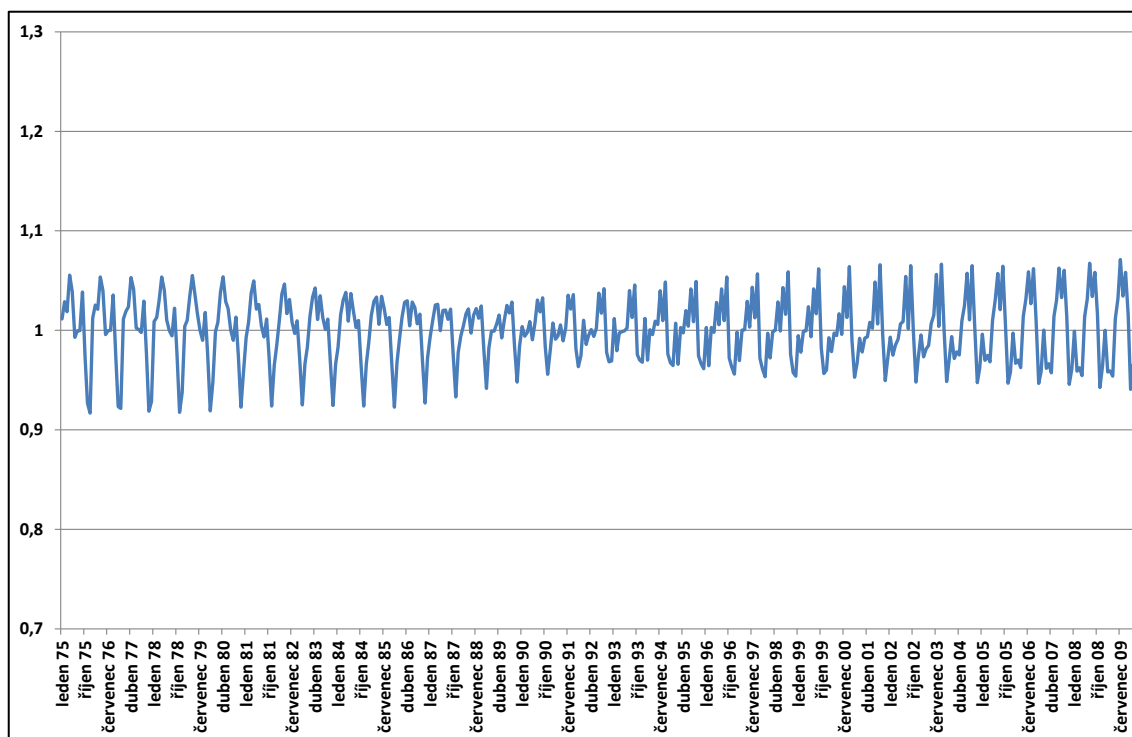
Zdroj: INE, SI, SN, vlastní výpočty

Sezónní indexy pro tři vybrané evropské země, Island, Norsko a Španělsko, znázorňuje paprskový graf na obrázku 13. Z něj můžeme vidět, že i když jsou rozdíly v hodnotách jednotlivých indexů malé, lze označit islandské živě narozené děti jako letní, norské děti jako jarní, a španělské děti jako podzimní. V těchto obdobích totiž vykazují jednotlivé státy vyšší hodnoty sezónních indexů než zbylé dvě země.

Zcela odlišný průběh proměnlivých sezónních indexů oproti Česku pozorujeme u Slovinska, které je znázorněno na obrázku 14. I přes skutečnost, že je znázorněn delší časový

úsek než u Česka, vidíme, že se rozložení živě narozených během roku zcela změnilo. Ani v tomto případě není rozpětí mezi minimální a maximální hodnotou široké, přesto lze ze získaných výsledků vyvodit některé závěry. Zatímco na počátku období byl odstup minimální hodnoty označující listopadové počty narozených od ostatních měsíců velký, na počátku 90. let 20. století došlo k přiblížení všech hodnot sezónních indexů, rozložení živě narozených během roku tedy bylo více rovnoměrné. V dalších letech se naopak zvýšily hodnoty indexů pro zářijové hodnoty, které byly od roku 2006 dosaženy a od roku 2008 i převyšeny hodnotami červencových indexů. Listopadové minimum však zůstává po celou dobu.

**Obr. 14 – Proměnlivé sezónní indexy, živě narození, Slovinsko, 1975–2009**



*Zdroj: SORS, vlastní výpočty*

Na časové řady o porodnosti v Estonsku, Finsku, Norsku, Slovinsku a Španělsku v období let 1975–2009- byla aplikována metoda kointegrace časových řad. Na základě Dickey-Fullerova testu bylo odhaleno, že dvě časové řady, pro Finsko a Norsko, jsou stacionární a nemohou tedy do testu kointegrace vstoupit. Johansenův test kointegrace byl tedy použit na tři zbývající časové řady a i přes jejich nestacionaritu mezi nimi nebyl nalezen žádný kointegrační proces.

## 5 Sňatečnost

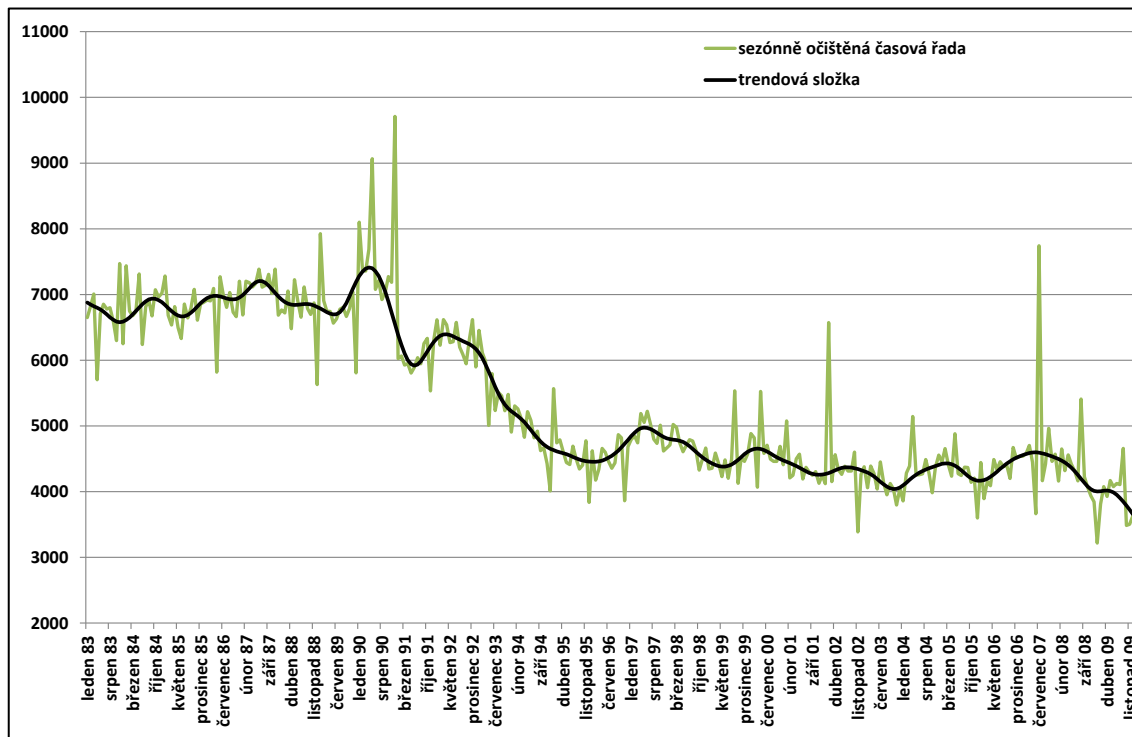
Sňatečnost je z hlediska sezónnosti pravděpodobně nejvíce sledovaným demografickým procesem. Sezónní výkyvy v oblasti sňatečnosti jsou významné především z toho důvodu, že rozložení sňatků do jednotlivých měsíců v roce je ovlivněno lidským rozhodnutím (Tesárková, 2009a). Na sezónnost sňatečnosti tak má vliv mnoho vnějších faktorů, z minulosti je možné zmínit např. rozložení hospodářského roku, a tedy zásoby jídla po podzimních sklizních nebo návaznost na nepřítomnost mužů v době sezónních prací. V současnosti přetrvává vliv církevních nebo světských svátků, preference teplého počasí a stále také pověr, i když jejich vliv se zmenšuje, a to především ve velkých městech. Novým fenoménem v rozhodování o datu uspořádání svatby se stala tzv. „magická data“, která si snoubenci stále častěji vybírají, a to ať už z důvodu atraktivity data, anebo z pověrčivosti, kdy věří na šťastná čísla v datu svatby, například 7. 7. 2007.

### 5.1 Česko

Známou skutečnost, že především po roce 1990 dochází ke snižování sňatečnosti, dokumentuje i graf na obrázku 15. Přesto je v sezónně očištěné časové řadě zřejmých několik výrazných výkyvů. Tři vrcholy dosahují hodnot nad 8000 sňatků během roku 1990, z nichž jeden je květnový, a tedy vyčnívá ze zvyklosti minimalizovat počet svateb v tomto jarním měsíci. Tyto vysoké počty uzavřených manželství odkazují na kumulaci svateb do roku 1990 před zrušení státní garance novomanželských půjček, platné od ledna 1991 (Tesárková, 2009a). Následně dochází k trvalému poklesu počtu svateb, který je v několika jednotlivých měsících narušen jednorázovým nárůstem těchto událostí. Prvním takovým výrazným výkyvem je únorová hodnota počtu uzavřených manželství okolo 6500, což je přibližně o 2000 svateb více, než byl průměr v daném období. Důvodem tohoto nárůstu, navíc v měsíci, který není pro pořádání svateb mezi těmi oblíbenými, je již zmíněné magické datum. Jednalo se především o pátek 22. 2. 2002 a sobotu 2. 2. 2002, některé páry však využily i středu 20. 2. 2002. Největší odchylku od okolního průběhu má pak jednoznačně hodnota 7741 uzavřených svateb během července roku 2007. V sobotu 7. 7. 2007, tedy v den, kdy se během týdne nejčastěji uzavírají svatby, navíc v oblíbeném letním měsíci a s datem plným „šťastných sedmiček“ bylo podle ČSÚ oddáno rekordních 4 406 párů. Z grafu je však zřejmé, že se nejednalo o počátek zvyšování sňatečnosti, ale pouze o kumulaci svateb do jednoho okamžiku. Výrazněji se do průběhu časové řady zapsal i srpen 2008, kdy bylo k uzavření sňatku využito pátku

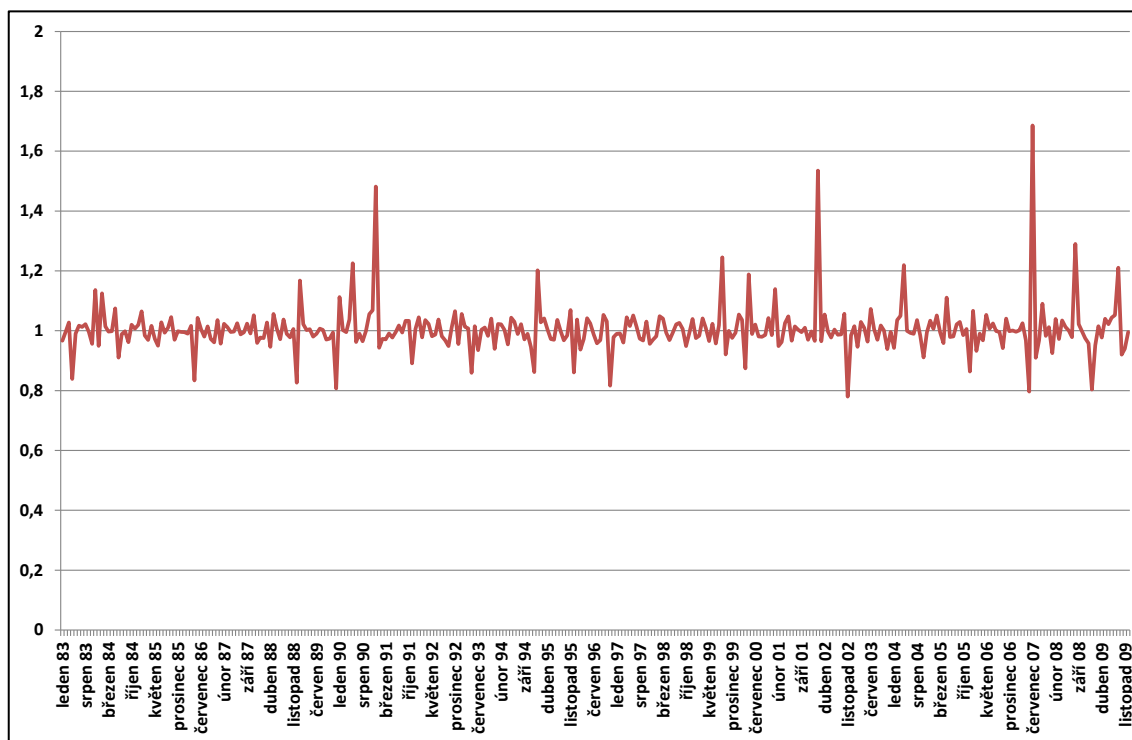
8. 8. 2008. Graf náhodné složky časové řady o sňatečnosti v Česku, zobrazený na obrázku 16, koresponduje s výkyvy v sezónně očištěné časové řadě.

**Obr. 15 – Sezónně očištěná časová řada, trendová složka, sňatky, Česko, 1983–2009**



Zdroj: ČSÚ, vlastní výpočty

**Obr. 16 – Náhodná složka, sňatky, Česko, 1983–2009**

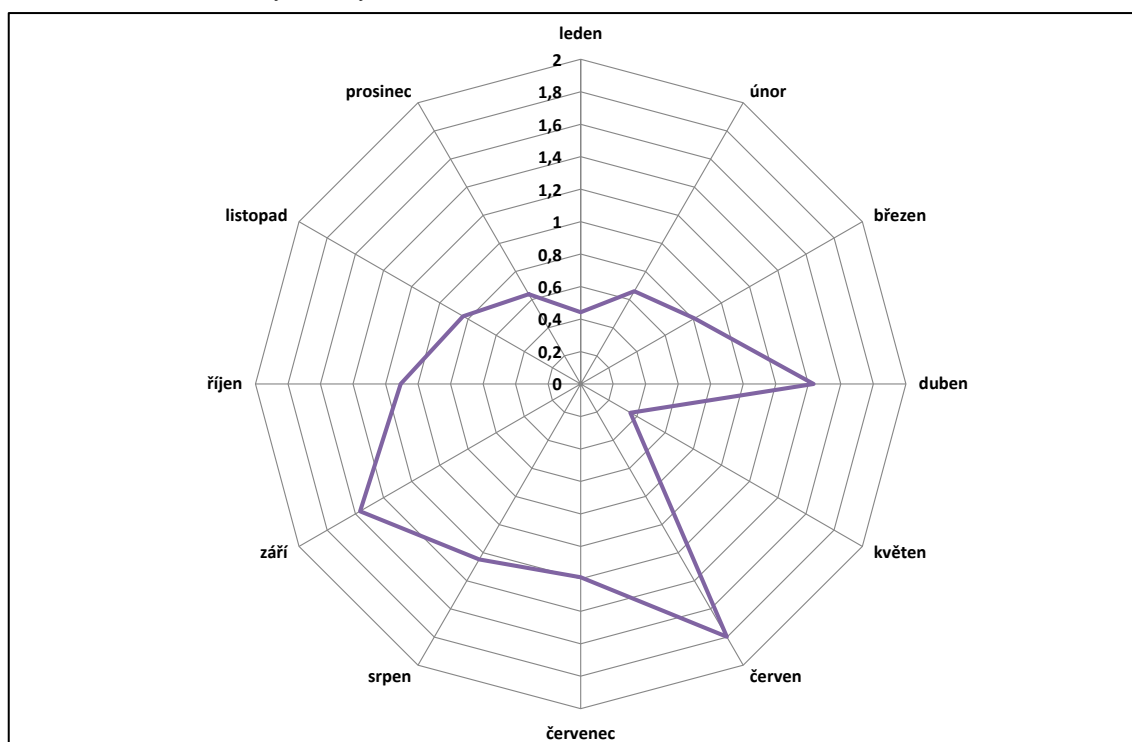


Zdroj: ČSÚ, vlastní výpočty

Kromě již výše zvýšených výkyvů do kladných hodnot lze na křivce náhodné složky pozorovat také výkyvy pro nízký počet uzavřených sňatků, které ve většině případů odpovídají měsícům listopad a prosinec. Zajímavý je výkyv k hodnotě 0,8 v červnu 2007, který předchází hodnotě indexu 1,7 pro měsíc červenec a ilustruje tak odložení potenciálních červnových sňatků na další měsíc, který nabízel velmi atraktivní sobotní datum. Poslední výkyv v časové řadě směrem k vyšším hodnotám vykazuje září 2009, které ale ani zdaleka nedosáhlo věhlasu července 2007 a srpna 2008.

Sezónní indexy sňatečnosti mají na rozdíl, např. od indexů o porodnosti, široké rozpětí, jak můžeme vidět na grafu na obrázku 17. Zatímco nejnižší hodnotu indexu za celé sledované období vykazuje květen (0,35), po něm následující červen dosahuje až k hodnotě 1,80. Lze tedy říci, že zatímco v květnu je uzavřeno 35 % průměrného měsíčního počtu svateb, v červnu je to 180 %. Spolu s květnem jsou podprůměrnými měsíci z hlediska počtu uzavřených manželství ještě listopad, prosinec, leden a únor, tedy zimní měsíce.

**Obr. 17 – Sezónní indexy, sňatky, Česko, 1983–2009**



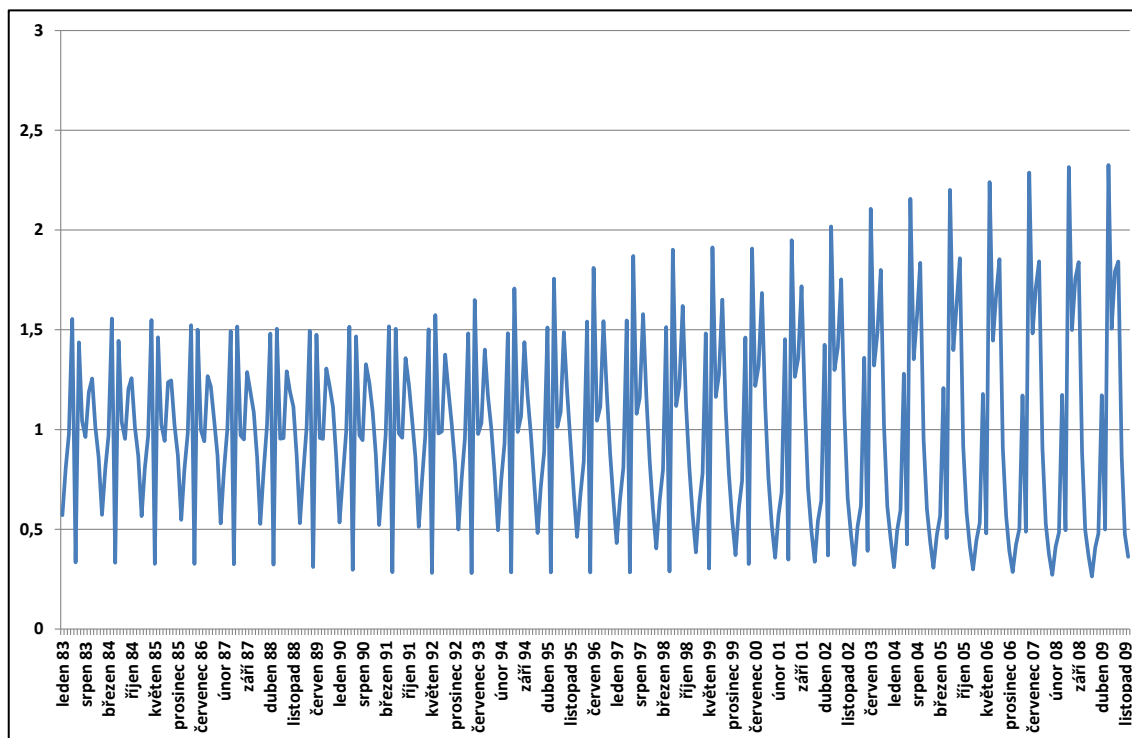
Zdroj: ČSÚ, vlastní výpočty

Pro podrobnější analýzu sezónnosti sňatkového chování v Česku je vhodné zobrazit proměnlivé sezónní indexy. Tento ukazatel je vykreslen v grafu na obrázku 18 a je z něj zřejmé, že počátkem 90. let došlo ke zvětšení rozdílů v už tak nerovnoměrném rozložení sňatků během roku. Minimální hodnoty sezónních indexů pro květen pozorujeme od počátku sledovaného období až do roku 2001, kdy pozici nejméně oblíbeného měsíce pro pořádání svateb přebírá leden. I v případě maximálních hodnot došlo během analyzovaných 26 let ke změně. Do roku 1991 byl nejvíce oblíbeným měsícem pro uzavření manželství duben, od následujícího roku pak



ho na pomyslné první příčce vystřídal červen. Rokem 1992 se také začínají prudce zvětšovat rozdíly mezi jednotlivými měsíci a obliba června víceméně plynule roste. Během následujících let tak vzroste z hodnoty 1,57 v roce 1992 na hodnotu 2,33 v roce 2009.

**Obr. 18 – Proměnlivé sezónní indexy, sňatky, Česko, 1983–2009**



*Zdroj: ČSÚ, vlastní výpočty*

V roce 2010, v této práci neanalyzovaném, byla pomyslným atraktivním datem neděle 10. 10., o kterou ale nakonec nebyl v Česku takový zájem, jak očekávaly matřiky. Naproti tomu například v Číně vyvolalo toto magické datum svatební horečku, kdy chtěly podle deníku China Daily vstoupit do manželství desetitisíce párů. Pro umocnění už tak mimořádného oddacího termínu byl největší zájem o rezervaci doby oddání na 10:00 a 10:10 (China Daily, 2010).

V letošním roce 2011 by tak podle minulých zkušeností bylo nasnadě očekávat zvýšený zájem o datum 11. 11. 2011 (pátek) a 20. 11. 2011 (neděle). Ačkoliv listopad patří mezi měsíce s nízkým počtem uzavíraných svateb, skutečnost že první magické datum připadá na pátek, by mohlo situaci pro letošní rok změnit. Jen těžko pak lze odhadnout zájem o nedělní atraktivní datum, obzvláště po zkušenostech matřik s nedělí 10. 10. 2010.

## 5.2 Evropské státy

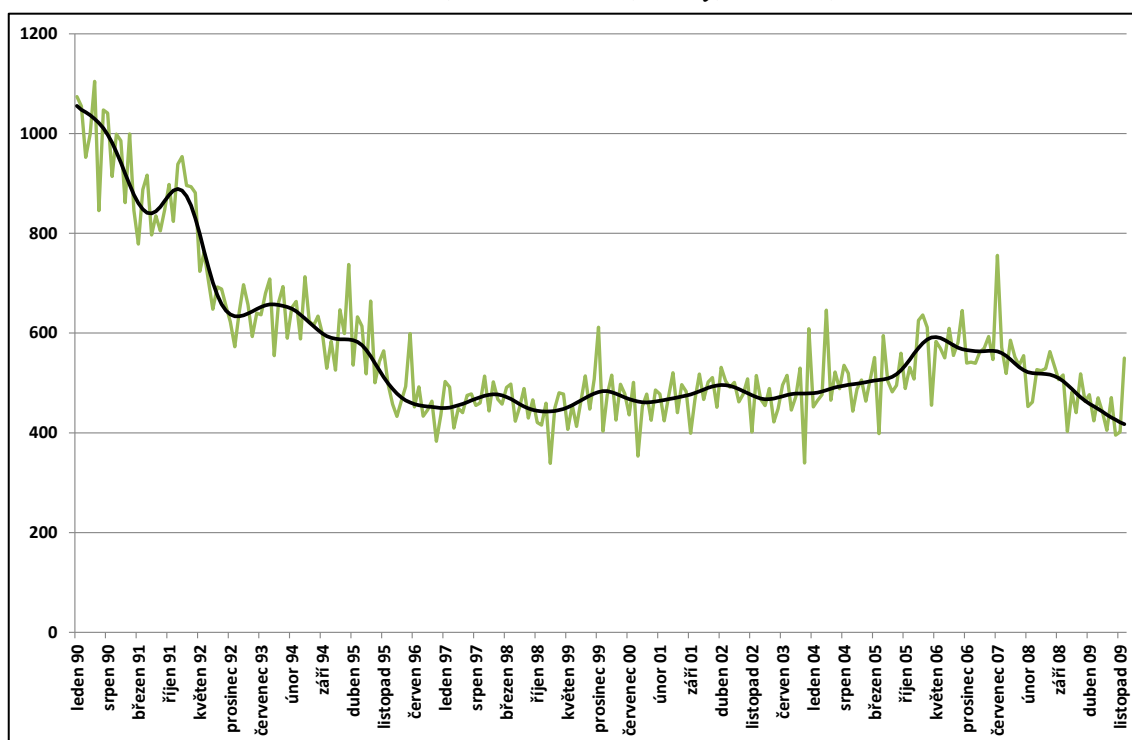
Analýza sezónnosti sňatečnosti států z rozdílných geografických poloh může být zajímavá i z hlediska laické veřejnosti, neboť přiblíží preference jednotlivých populací pro různá roční období a měsíce v roce. A na základě získaných výsledků lze založit podrobné studium náboženských tradic a lidových zvyků dané oblasti.

Měsíční data o počtu uzavíraných sňatků nejsou jednotně dostupná pro všechny státy Evropské unie, jako tomu je například u počtu zemřelých. Ze statistických úřadů jednotlivých zemí tak byly vygenerovány časové řady různé délky a výsledky rozděleny do tří časových období. Pro srovnání nejhodnější sezónní indexy jsou zobrazeny v grafech s jednotným měřítkem, aby bylo možné jednotlivá období a země snadno porovnávat.

### 5.2.1 Období let 1990–2009

V časovém období let 1990–2009 bylo podrobena analýze 5 států, Dánsko, Estonsko, Lucembursko, Slovinsko a Španělsko. Podobně jako u porodnosti mezi zeměmi nalezneme různé modely vývoje měsíčních počtů událostí. Na obrázcích 19 a 20 znázorněné časové řady a trendové složky pro Estonsko a Dánsko ilustrují, že zatímco se v Estonsku během sledovaného období počet sňatků snížil na méně než polovinu počáteční hodnoty, u Dánska pozorujeme mírný nárůst a následnou stagnaci okolo hodnoty stejné jako na počátku období. V časové řadě Estonska je viditelný nárůst počtu svateb v červenci 2007, související s již zmíněným atraktivním datem, není tento výkyv tak výrazný jako v případě časové řady Česka. V časové řadě Dánska pak vidíme tři takové výkyvy během druhé poloviny sledovaného období, která signalizují zvýšený počet uzavřených sňatků v únoru 2002, květnu 2005 a červenci 2007.

**Obr. 19 – Sezónně očištěná časová řada, trendová složka, sňatky, Estonsko, 1990–2009**

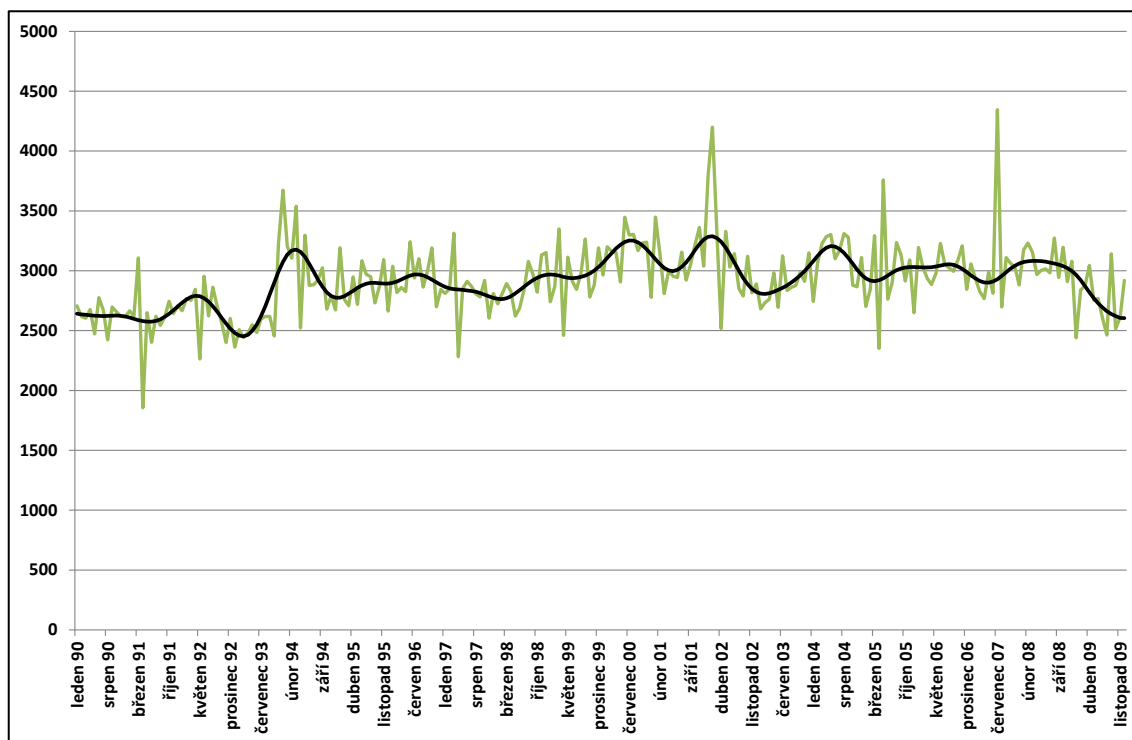


*Zdroj: SE, vlastní výpočty*

S ohledem na skutečnost, že se nám vývoj v časové řadě porodnosti a sňatečnosti v Estonsku subjektivně jeví jako podobný, lze na tato data v období let 1990–2009 použít metodu kointegrace časových řad. Po otestování obou standardizovaných řad

Dickey-Fullerovým testem jednotkového kořene však bylo zjištěno, že časová řada sňatečnosti nemá jednotkový kořen, tedy nespĺňuje podmínku nestacionarity. S ohledem na tuto skutečnost tak není možné hledat mezi dvěma vybranými časovými řadami kointegrační proces pomocí Johansenova testu kointegrace.

**Obr. 20 – Sezónně očištěná časová řada, trendová složka, sňatky, Dánsko, 1990–2009**



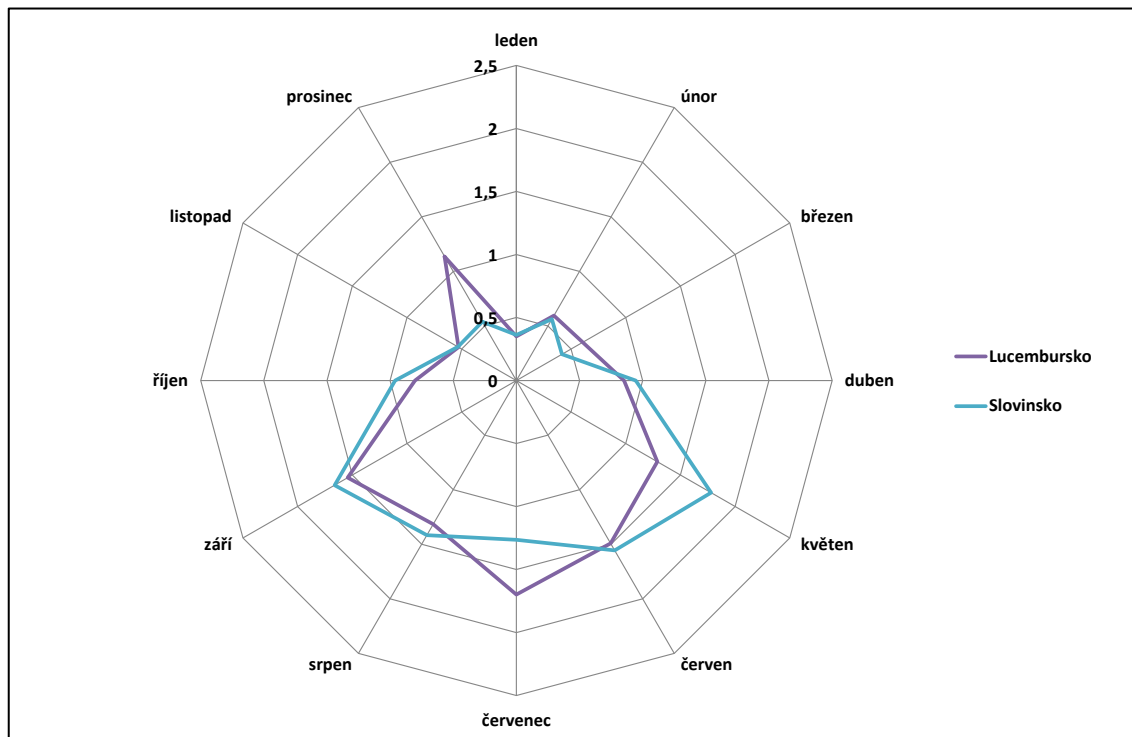
Zdroj: SD, vlastní výpočty

Získané sezónní indexy byly z důvodu přehlednosti vykresleny do dvou grafů, které jsou znázorněny na obrázcích 21 a 22. Na prvním zmíněném grafu lze zaznamenat rozdílné rozložení sňatků, zatímco v Lucembursku jsou nejčastěji svatby konány v měsíci červenci, ve Slovinsku výrazně převyšují ostatní měsíce hodnoty v květnu a září. V obou zemích je pak nejméně oblíbeným měsícem pro uzavření sňatku leden. Tento zimní měsíc je i podle následujícího grafu nejméně často vybírán jako měsíc pro uzavření sňatku, a to ve všech třech znázorněných zemích, Dánsku, Estonsku a Španělsku. Naopak nejfrekventovanější měsíc se u jednotlivých států liší. Zatímco obyvatelé Dánska preferují srpen, Španělé září, obyvatelé Estonska mají shodné hodnoty pro dva letní prázdninové měsíce, červenec a srpen.

Z analyzovaných pěti států v tomto časovém období bylo vybráno Estonsko pro znázornění vývoje proměnlivých sezónních indexů. Křivka tohoto ukazatele byla vybrána z důvodu velice rychle se rozšiřujícího rozpětí mezi minimálními a maximálními hodnotami sezónních indexů v jednotlivých letech, jak je vidět na obrázku 23. Zatímco rozdíl mezi lednovou a červnovou hodnotou v prvním analyzovaném roce, tedy v roce 1990 byl 0,39, rozdíl na konci časového intervalu, v roce 2009, mezi minimálním lednovým indexem a maximálním srpnovým indexem dosáhl hodnoty 1,45. Srpnové a zářijové indexy jsou navíc od roku 2003

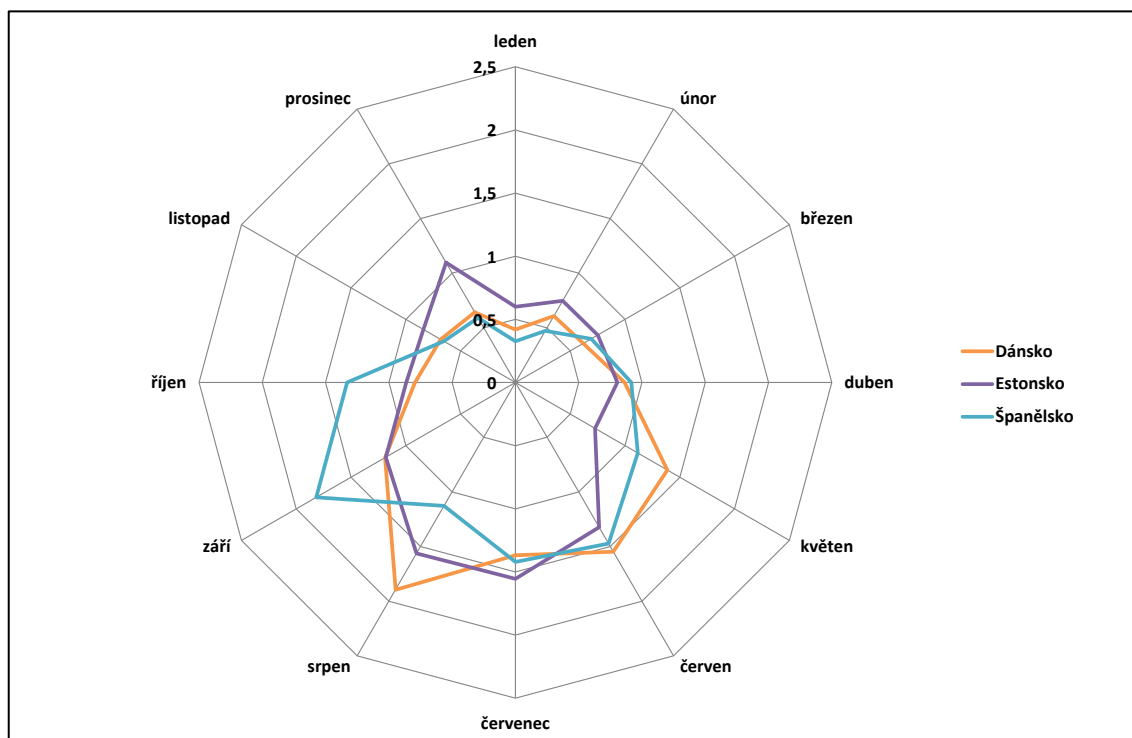
jedinými hodnotami vyššími než 1, jejich podíl na celkovém součtu sezónních indexů 12 je tak nejvýraznější.

**Obr. 21 – Sezónní indexy, sňatky, Lucembursko, Slovinsko, 1990–2009**



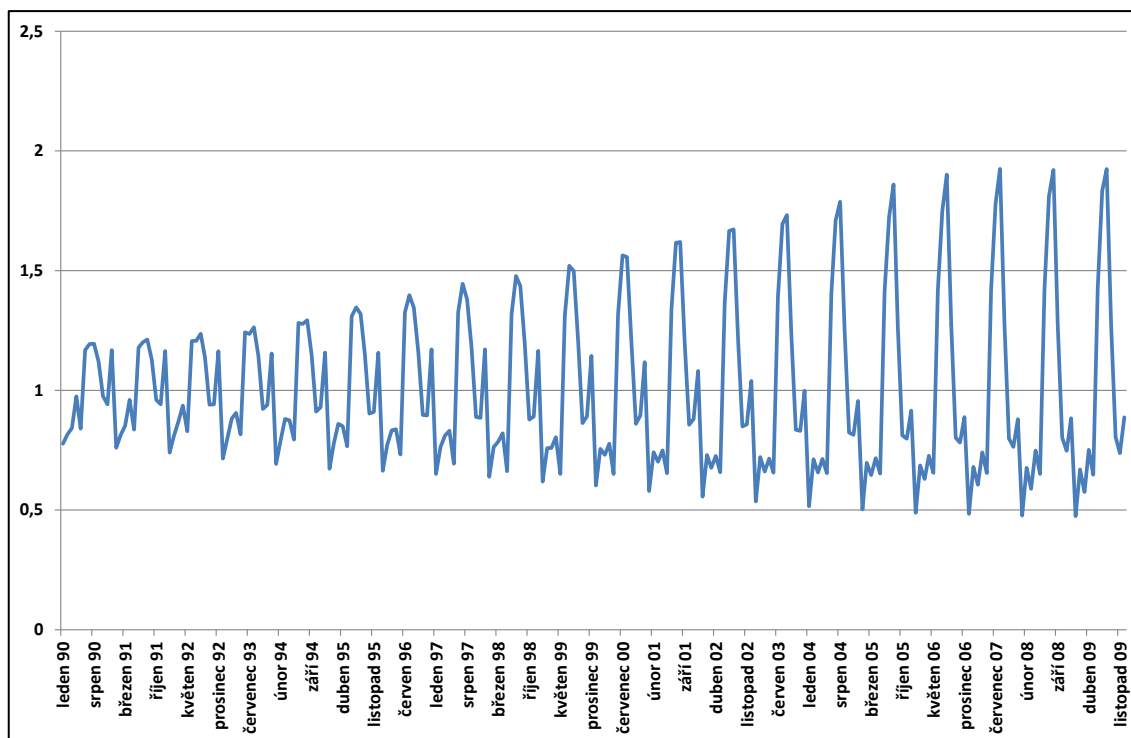
Zdroj: NISES, SORS, vlastní výpočty

**Obr. 22 – Sezónní indexy, sňatky, Dánsko, Estonsko, Španělsko, 1990–2009**



Zdroj: INE, SE, SD, vlastní výpočty

Obr. 23 – Proměnlivé sezónní indexy, sňatky, Estonsko, 1990–2009



Zdroj: SE, vlastní výpočty

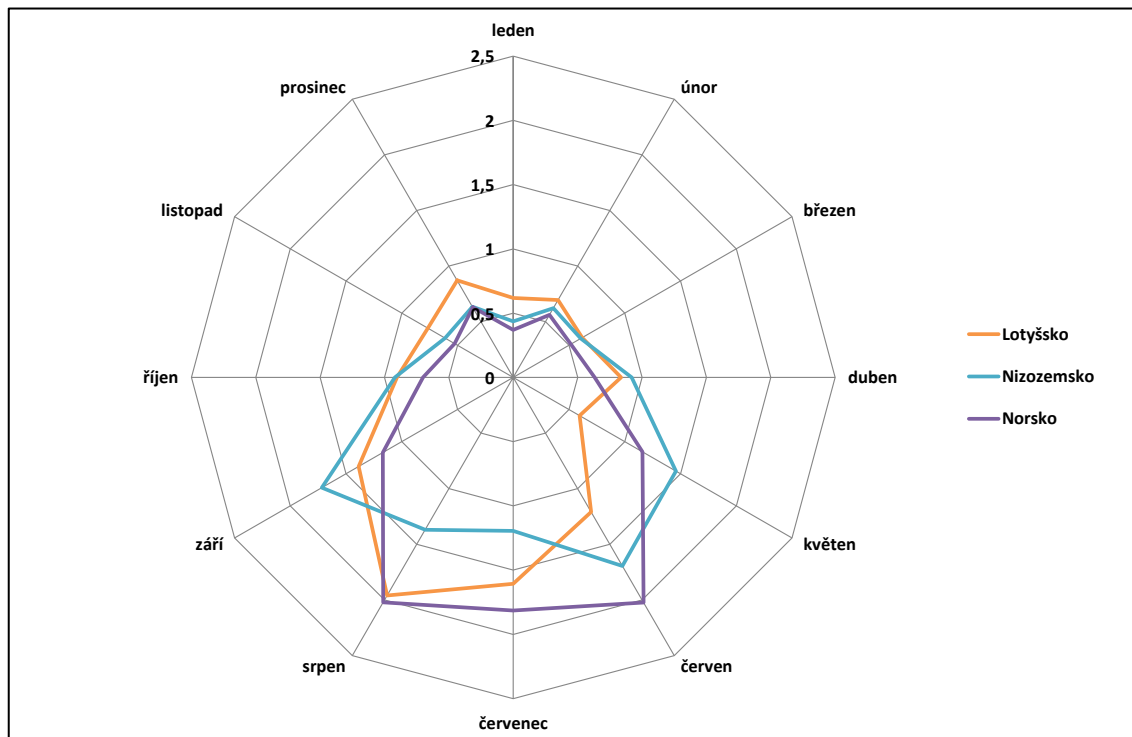
### 5.2.2 Období let 1996–2009

V období dlouhém 14 let, od roku 1996 do roku 2009 byly analyzovány časové řady o sňatečnosti pro tři evropské země, Lotyšsko, Nizozemsko a Norsko. Rozložení sňatků v těchto zemích do jednotlivých měsíců znázorňuje graf sezónních indexů na obrázku 24. Z něj vidíme oblibu měsíce srpna pro uzavírání sňatků jak v Lotyšsku, tak i v Norsku, ve kterém je na podobné úrovni i sezónní index pro měsíc červen. I v případě Nizozemska jsou hodnoty indexů na skoro stejné vysoké úrovni u dvou měsíců, a to u června a září. Leden je v případě těchto tří států měsícem nejméně oblíbeným pro konání svateb, v případě Lotyšska je stejně nízká hodnota zaznamenána i pro měsíc květen.

Z grafu proměnlivých sezónních indexů, znázorněných na obrázku 25, můžeme odhalit zajímavé, a navzájem odlišné, pravidelnosti v rozložení počtu sňatků během roku u populace Lotyšska a Nizozemska. Na první pohled je zřejmé zvyšování variability sezónních indexů pro Lotyšsko, neboť zatímco na počátku období byly nadprůměrné hodnoty na podobné úrovni, ke konci sledovaného intervalu převyšují maximální hodnoty lotyšských indexů ty nizozemské o 0,5. Narůstá tak odstup oblíbenosti svateb v Lotyšsku v měsíci srpnu od ostatních měsíců v roce. Z křivky popisující proměnlivé sezónní indexy pro nizozemskou populaci vidíme po celé období přetrvávající minimální hodnotu v měsíci lednu, zatímco maximální hodnoty indexů jsou nepravidelně rozloženy mezi červnem a zářím. Zajímavá je skutečnost, že po celé sledované období je vždy červencová hodnota sezónního indexu o 0,5 „bodu“ níže než okolní

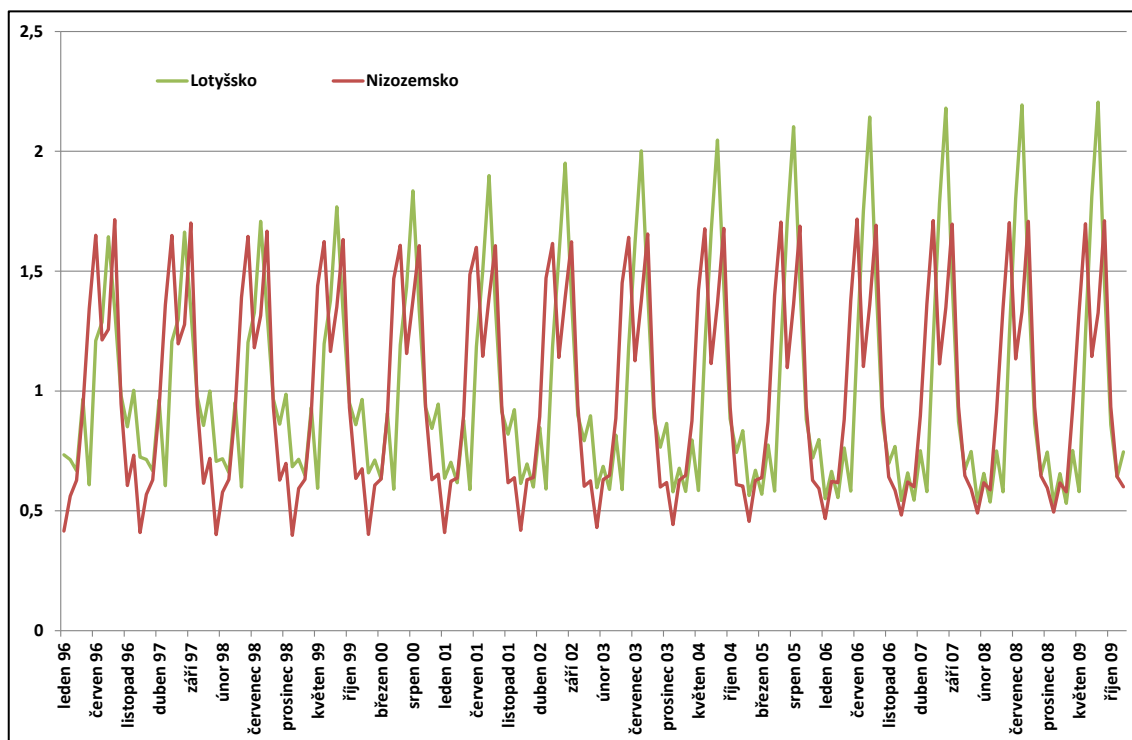
dvě hodnoty. Tento prudký výkyv má jistě své důvody, ať už např. pověru jako v Česku pro květen, nebo sezónní migraci na léto do zahraničí, které tato práce neodhalila.

**Obr. 24 – Sezónní indexy, sňatky, Lotyšsko, Nizozemsko, Norsko, 1996–2009**



Zdroj: CSBL, SN, SNL, vlastní výpočty

**Obr. 25 – Proměnlivé sezónní indexy, sňatky, Lotyšsko, Nizozemsko, 1996–2009**

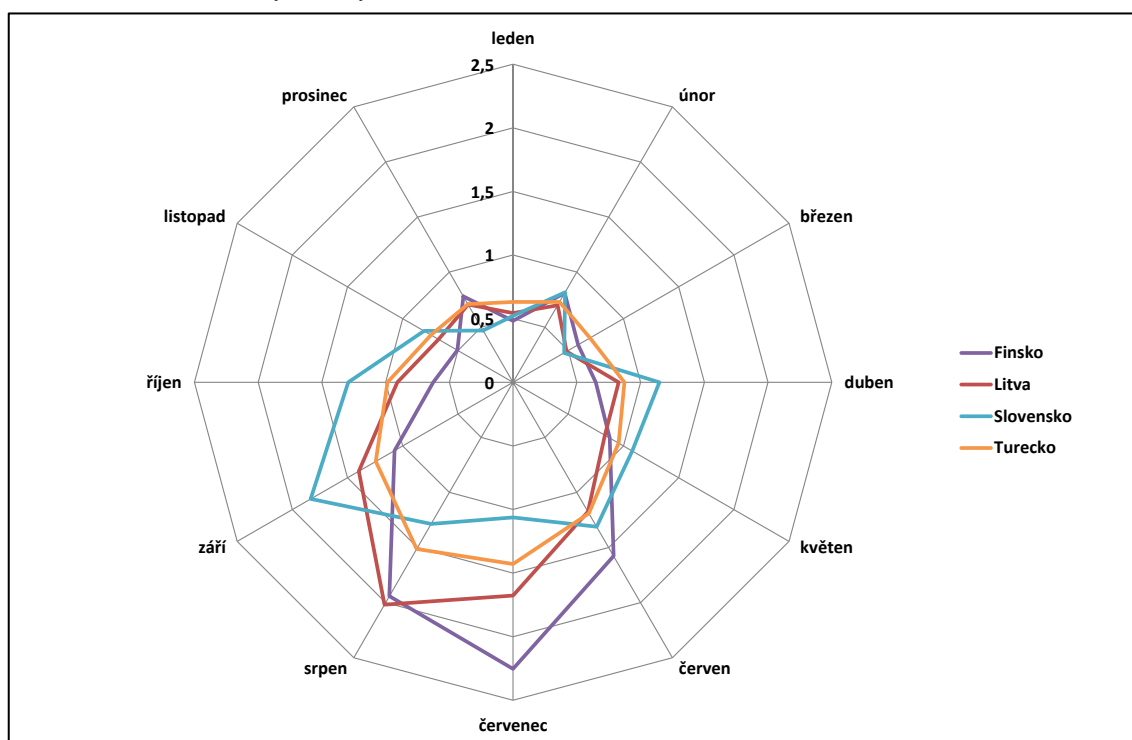


Zdroj: CSBL, SNL, vlastní výpočty

### 5.2.3 Období let 2001–2009

Pro nejkratší časové období let 2001–2009 byla analyzována data čtyř evropských zemí, Finska, Litvy, Slovenska a Turecka. Z grafu sezónních indexů, znázorněného na obrázku 26, jasně vidíme nejvyšší hodnotu indexu ze všech analyzovaných zemí (i v ostatních časových obdobích), a to v případě července pro Finsko, kdy tato hodnota dosáhla 2,25. Vysokou oblibu i hodnotu sezónního indexu registrujeme u litevské populace v srpnu. U obou těchto zemí pak lze pozorovat velmi nízké hodnoty indexů pro zimní měsíce, které klesají až k hodnotě 0,5. I Slovensko se vyznačuje jedním výrazně převládajícím měsícem co do počtu konaných svateb, a to zářím, zatímco Turecko vykazuje rozložení nadprůměrných hodnot indexů během všech letních měsíců, a to vcelku rovnoměrně.

Obr. 26 – Sezónní indexy, sňatky, Finsko, Litva, Slovensko, Turecko, 2001–2009



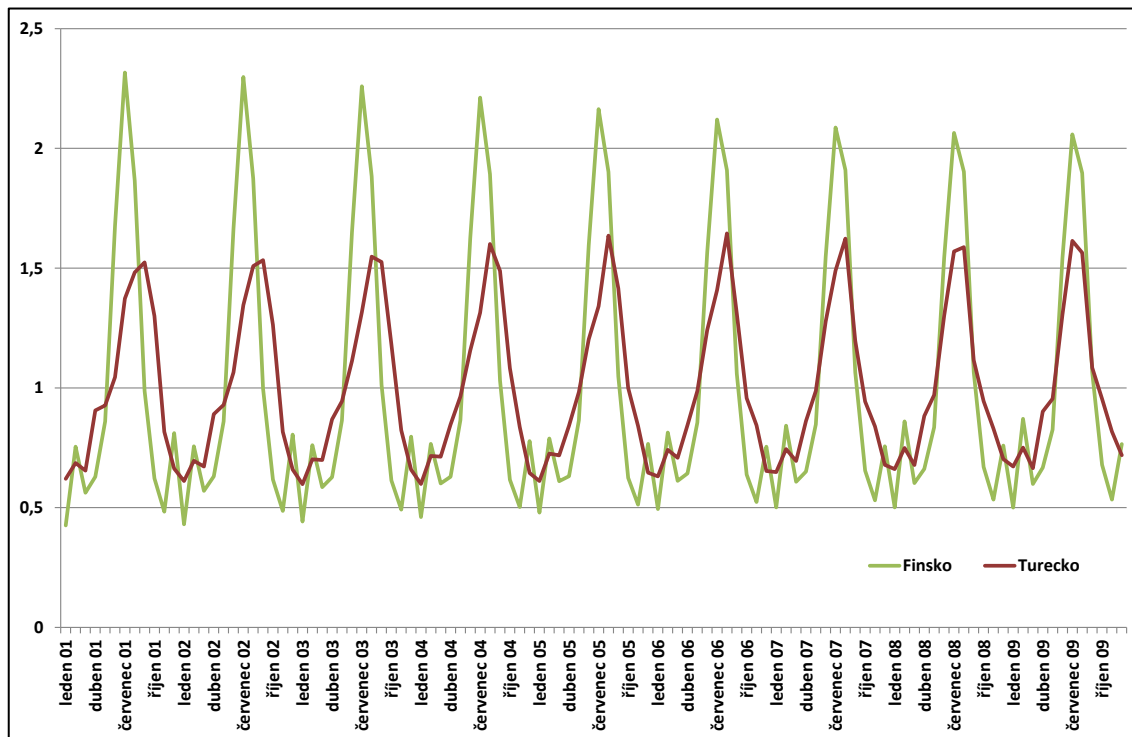
Zdroj: SF, SL, SORS, TSI, vlastní výpočty

Obrázek 27, který vykresluje proměnlivé sezónní indexy pro Finsko a Turecko, se vyznačuje podobným tvarem, avšak odlišnou variabilitou hodnot jednotlivých křivek. Finská červencová maxima se během období přibližují indexům ostatních měsíců, čímž se zmenšuje rozpětí hodnot. Pro křivku Turecka vidíme naopak mírné zvětšení rozpětí, kdy minimální lednové hodnoty se mění jen pozvolna, zatímco roste odstup srpnových a zářijových sezónních indexů.

Z grafu náhodné složky, znázorněné na obrázku 28, vidíme sedm výrazných výkyvů do nadprůměrných hodnot, a často před, nebo za, těmito kladnými výkyvy lze pozorovat propad do nízkých hodnot. Tento průběh je opět způsoben kumulací svateb z několika sousedních měsíců do měsíce, které nabízí atraktivní datum. Popíšeme-li výkyvy postupně, jak se udály v čase, jedná se o leden 2001, únor 2002, duben 2004, květen 2005, červenec 2007, srpen 2008

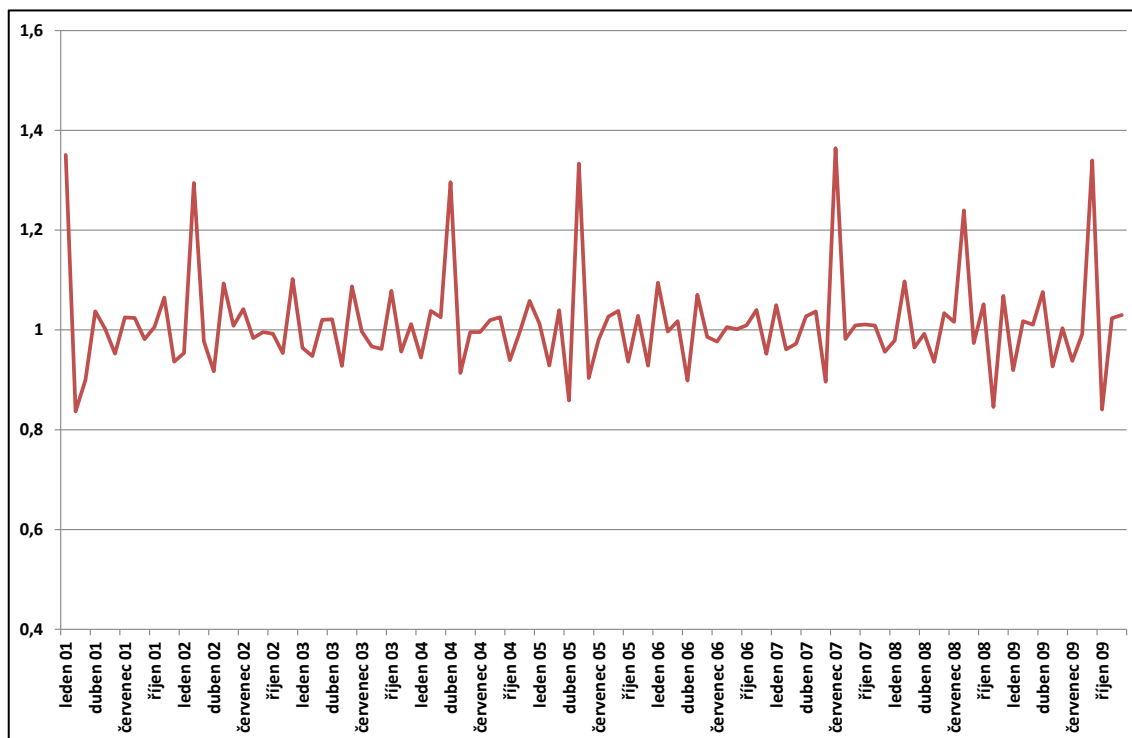
a září 2009. I přes vynechání dvou března 2003 a června 2006, se zdá, že obliba magického data je ve Finsku ještě vyšší než v Česku.

**Obr. 27 – Proměnlivé sezónní indexy, sňatky, Finsko, Turecko, 2001–2009**



Zdroj: SF, TSI, vlastní výpočty

**Obr. 28 – Náhodná složka, sňatky, Finsko, 2001–2009**



Zdroj: SF, vlastní výpočty

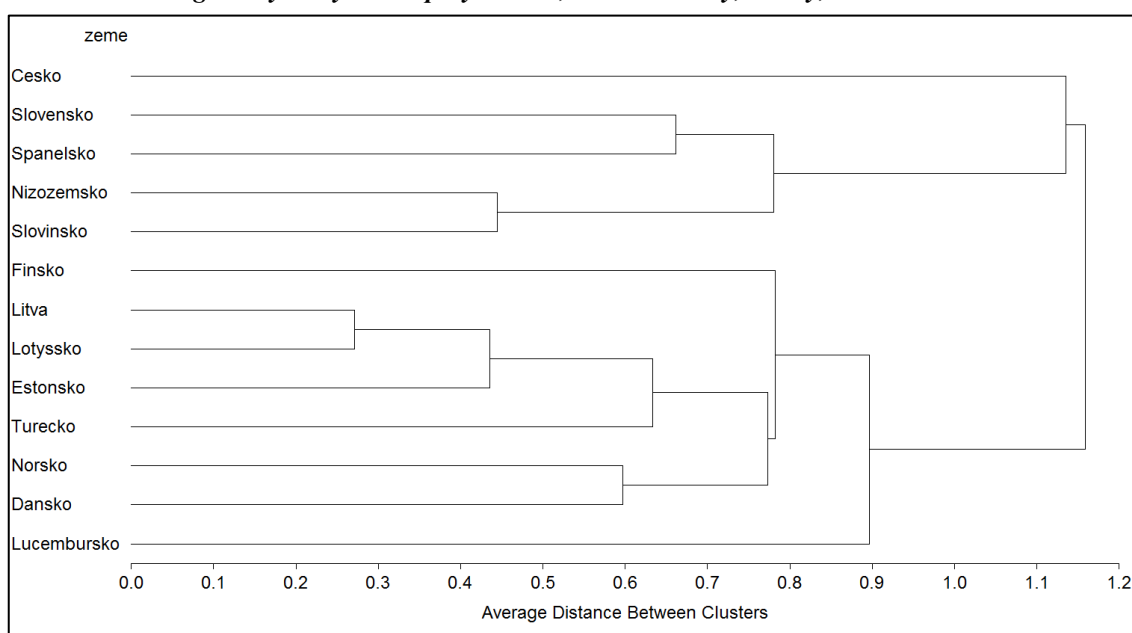


### 5.2.4 Shluková analýza

Pro vzájemné porovnání sňatkového chování v jednotlivých evropských zemích byla použita shluková analýza. Aby byla dosažena co největší podobnost vstupních dat, byly časové řady za všechny země vstupující do analýzy sjednoceny z hlediska délky sledovaného období, a to na období let 2001–2009. Následně byly pro tyto zkrácené řady vypočítány jednotlivé složky časových řad a tyto bezrozměrné charakteristiky byly následně využity jako proměnné pro výpočet shlukové analýzy. Těmito charakteristikami tak jsou sezónní indexy, proměnlivé sezónní indexy a náhodná složka. Pro všechny tyto tři složky byla provedena shluková analýza a její výsledky je možné vidět na dendogramech, které jsou vykresleny na obrázcích 29, 30 a 31.

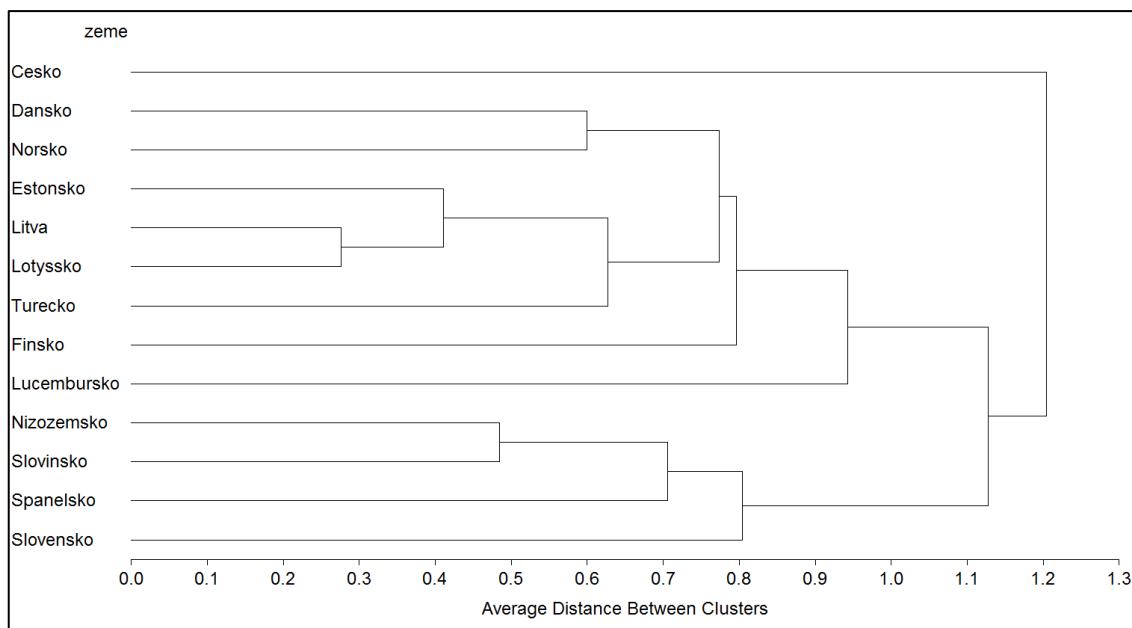
Z přiložených grafů můžeme vidět, že výsledky shlukové analýzy na datech za sezónní indexy a proměnlivé sezónní indexy se významně neliší. V obou případech jsou stejná složení jednotlivých shluků, ať už vezmeme jako počet shluků tři nebo čtyři, a hodnoty průměrných vzdáleností mezi jednotlivými shluky jsou velmi podobné. V případě tří shluků je jeden shluk tvořen pouze jedním státem, Českem, což souvisí s odmítáním květnových svateb u české populace. Druhý shluk je pak tvořen zeměmi Nizozemsko, Slovensko, Slovinsko, Španělsko a třetí shluk zeměmi zbývajícími. Pokud bychom ze všech zemí chtěli vytvořit shluky (clustery) čtyři, vydělí se v obou případech z nejpočetnějšího clusteru Lucembursko. Jak v případě dendogramu vykresleného na základě jednoduchých sezónních indexů, tak určeného podle dat proměnlivých sezónních indexů, si jsou nejbližší Litva a Lotyšsko, a následně tyto dva státy jako jeden shluk s třetím pobaltským státem, Estonskem. Tento fakt není tak překvapivý, vezmeme-li v úvahu geografickou i historickou sounáležitost. Podobný vztah vidíme z dendogramu i mezi Norskem a Dánskem. O to více pak může být překvapivé, že tak malá vzdálenost není například mezi Českem a Slovenskem. V tomto případě je však nepodobnost způsobena májovou povětří, která je známa a dodržována v Česku, zatímco na Slovensku nikoliv.

**Obr. 29 – Dendogram vybraných evropských zemí, sezónní indexy, sňatky, 2001–2009**



*Zdroj: ČSÚ, SD, SE, SF, SL, CSBL, NISES, SNL, SN, SOSR, SORS, INE, TSI, vlastní výpočty*

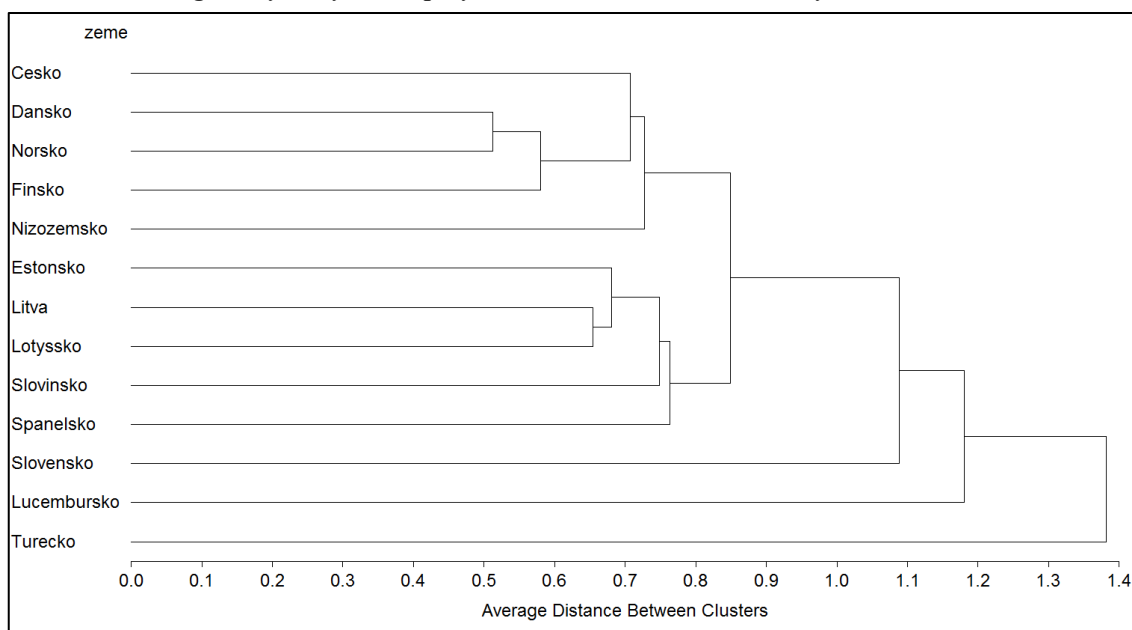
**Obr. 30 – Dendrogram vybraných evropských zemí, proměnlivé sezónní indexy, sňatky, 2001–2009**



Zdroj: ČSÚ, SD, SE, SF, SL, CSBL, NISES, SNL, SN, SOSR, SORS, INE, TSI, vlastní výpočty

Odlišný tvar dendrogramu i složení jednotlivých shluků znázorňuje graf, který je výsledkem aplikace shlukové analýzy na hodnoty náhodné složky. Z tohoto grafu vidíme, že jsou státy rozděleny do dvou velkých, vnitřně homogenních, shluků a tři státy, které se od ostatních výrazně vydělují. Těmito třemi státy jsou Slovensko, Lucembursko a Turecko. První homogenní shluk je tvořen Českem, Dánskem, Norskem, Finskem a Nizozemskem, a druhý potom pobaltskými zeměmi (Estonsko, Litva, Lotyšsko), Slovinskem a Španělskem.

**Obr. 31 – Dendrogram vybraných evropských zemí, náhodná složka, sňatky, 2001–2009**



Zdroj: ČSÚ, SD, SE, SF, SL, CSBL, NISES, SNL, SN, SOSR, SORS, INE, TSI, vlastní výpočty

## 6 Úmrtnost

Studium sezónnosti úmrtnosti může být velkým přínosem nejen v teoretické rovině, ale především v aplikované sféře, v tomto případě medicíně. Při znalosti vyšších rizik úmrtnosti pro některé subpopulace (diferencované podle věku, pohlaví nebo podle výskytu nemoci) lze v kritickém období věnovat např. zvýšenou pozornost zdravotnímu stavu pacienta nebo pokusit se snížit rizikové vnější faktory (např. vystavení organismu prudkým teplotním změnám). Jak již bylo zmíněno dříve, informace mohou být také cenné pro zdravotní pojišťovny a správné zacílení preventivních opatření na exponovanou skupinu osob. Diferenciace úmrtnosti je v této práci limitována dostupností dat, za Česko lze tedy odděleně zkoumat populaci podle pohlaví, věkových skupin nebo příčiny úmrtí. Zajímavé by jistě bylo i analyzovat sezónnost úmrtnosti podle dosaženého vzdělání, rodinného stavu nebo vykonávané profese. Pro evropské státy jsou srovnatelné údaje dostupné pouze za celkový počet zemřelých v období let 2000–2007, a oproti údajům o sňatečnosti, chybí údaje za jeden stát, za Turecko.

### 6.1 Česko

Český statistický úřad zveřejňuje podrobné statistiky o zemřelých podle několika diferenčních znaků. Ne vždy jsou však časové řady k analýze vhodné, a to především z důvodu malého počtu událostí v jednotlivých měsících. I takové řady je možné zkoumat, získané výsledky však mohou být nesignifikantní.

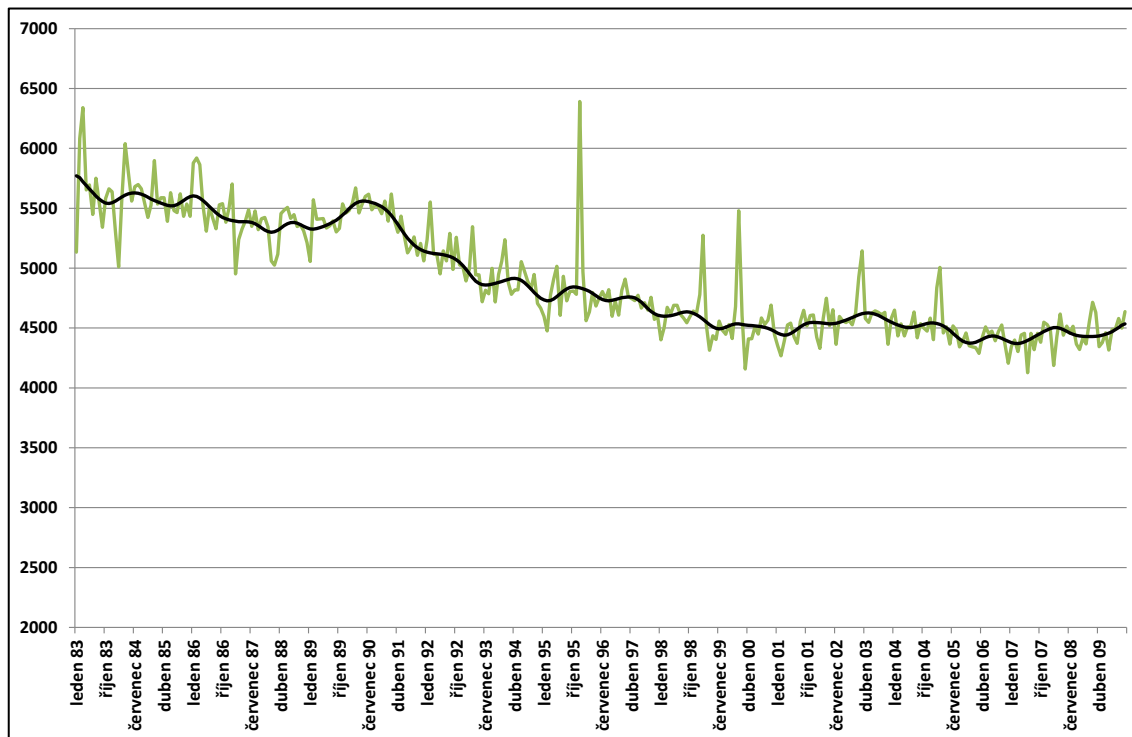
#### 6.1.1 Podle pohlaví

Rozdíly v úmrtnosti české populace jsou minimální, bereme-li v úvahu všechny zemřelé, nijak dále specifikované. Pro ukázkou je na obrázku 32 vykreslen graf sezónně očištěné časové řady a trendové složky pro muže, který znázorňuje tendenci zemřelých během celého období. V časové řadě jsou však některé nezanedbatelné výkyvy, které upozorňují na nečekaný nárůst počtu zemřelých. V samém počátku časové řady můžeme zaznamenat chřipkovou epidemii z období 16. 2. – 18. 3. 1983 (Kyselý, 2002), nejvýrazněji se pak do úmrtnosti mužů zapsala velká chřipková epidemie z prosince 1995 (9. 12. 1995–1. 1. 1996, Kyselý 2002). I následující čtyři výkyvy odkazují na období zvýšeného výskytu respiračních onemocnění, konkrétně se jedná o únor 1999, leden 2000, březen 2003 a březen 2005.

Proměnlivé sezónní indexy pro obě pohlaví jsou znázorněny v grafu na obrázku 33, ze kterého je zřejmé, že muži umírají rovnoměrněji během celého roku, a rozpětí sezónních

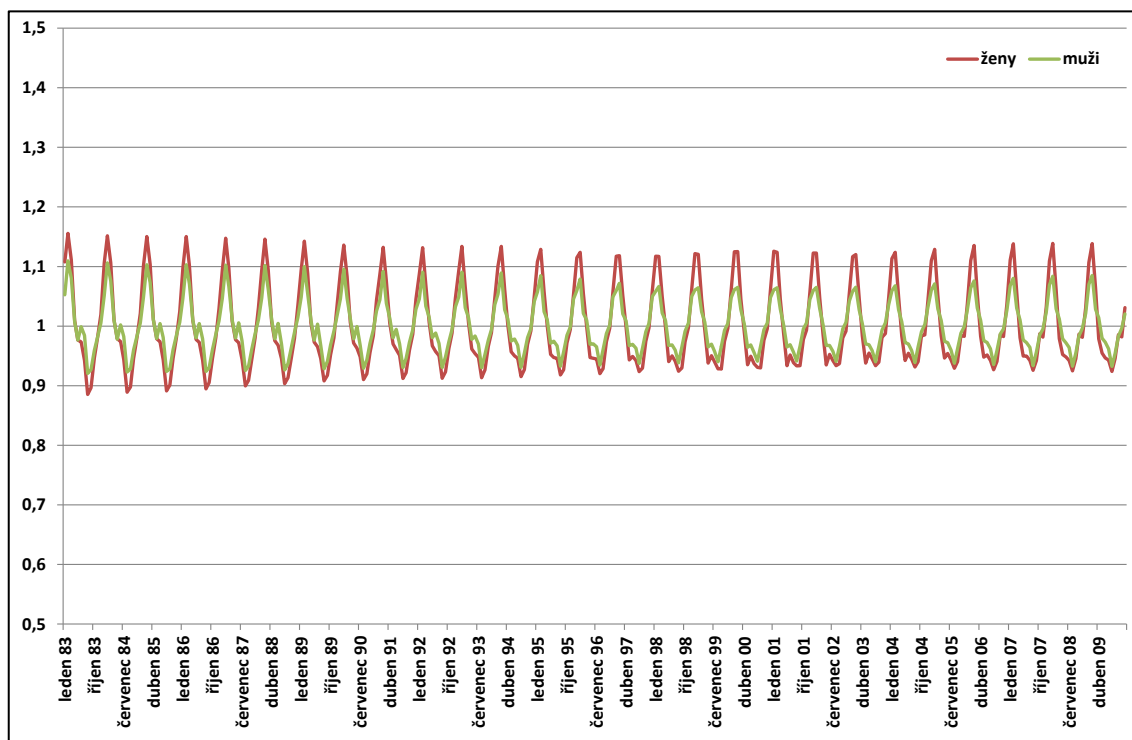
indexů pro muže tedy nemá takové rozpětí, zatímco ženy mají o poznání vyšší hodnoty indexů pro zimní měsíce, především pro měsíce leden a únor.

**Obr. 32 – Sezónně očištěná časová řada, trendová složka, zemřelí, muži, Česko, 1983–2009**



Zdroj: ČSÚ, vlastní výpočty

**Obr. 33 – Proměnlivé sezónní indexy, zemřelí, ženy, muži, Česko, 1983–2009**



Zdroj: ČSÚ, vlastní výpočty

### 6.1.2 Podle věku

Sezónnost úmrtnosti diferencované podle věku byla analyzována pro skupinu zemřelých do 1 roku věku a dále pro věkové skupiny 0–14 let, 15–29 let, 30–49 let, 50–69 let a skupinu osob starších 70 let. Kromě věkové skupiny zemřelých 0–14 let, byla u všech ostatních skupin potvrzena přítomnost sezónnosti, a to pomocí již dříve popsanych testů sezónnosti v časové řadě. Přesto je tato skupina nejmladších zemřelých vykreslena do grafu proměnlivých sezónních indexů na obrázku 35, a to z důvodu srovnání s další věkovou skupinou.

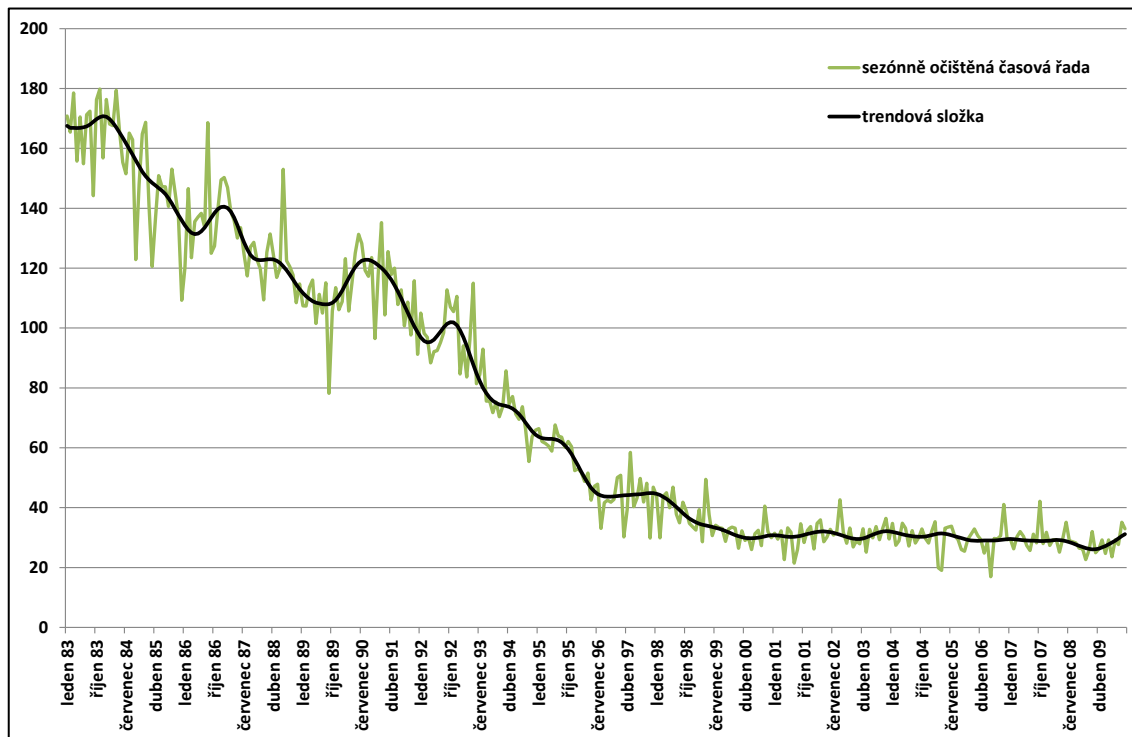
V grafu na obrázku 34 je znázorněn průběh časové řady a trendové složky pro zemřelé do 1 roku. Z něj můžeme vidět trvalý pokles měsíčního počtu zemřelých od počátku sledovaného období až do konce 90. let 20. století. Tento vývoj souvisí se stále se zlepšujícími podmínkami v lékařské péči, kdy Česko dohánělo ztrátu na západní vyspělé země způsobenou životem v socialismu a s tím související nedostupností moderních přístrojů a léčebných metod. Od roku 2000 se počet zemřelých stabilizoval na velmi nízké úrovni. Porovnáme-li tento vývoj s vývojem kojenecké úmrtnosti, který je znázorněn na obrázku 35, vidíme, že hodnoty kojenecké úmrtnosti nestagnují, ale dále se snižují. Tento fakt souvisí s metodou výpočtu kvocientu kojenecké úmrtnosti, který udává počty zemřelých dětí do 1 roku věku (0. dokončeného věku) na 1000 živě narozených dětí v dané populaci (pro Česko znázorněný na obrázku 8). V tomto případě, kdy roste počet živě narozených dětí a počet zemřelých do 1 roku věku se udržuje na stejné úrovni, tak neustále dochází ke snižování hodnoty kojenecké úmrtnosti. V současnosti patří Česko mezi země s nejnižší úrovní kojenecké úmrtnosti a hodnoty, které jsou v posledních letech dosahovány, je možné označit za přirozené absolutní minimum úrovně kojenecké úmrtnosti. Tento stav byl dosažen především velmi kvalitní prenatální a neonatální lékařskou péčí (Bartoňová, 2008).

Metoda kointegrace časových řad byla využita i na časovou řadu o úmrtnosti, konkrétně na počet zemřelých do 1 roku věku v letech 1983–2009, a byl hledán kointegrační proces s časovou řadou živě narozených dětí v Česku ve stejném období let 1983–2009. Aplikací Dickey-Fullerova testu bylo odhaleno, že obě vybrané časové řady jsou nestacionární, a tudíž je možné na ně aplikovat Johansenův test kointegrace. Na základě výsledků tohoto testu přijímáme nulovou hypotézu  $H_0$ , která říká, že mezi časovými řadami vstupujícími do analýzy není přítomen kointegrovaný proces. Časové řady živě narozených a zemřelých do 1 roku věku v české populaci tedy nejsou kointegrované.

Dvě nejstarší věkové skupiny (50–69 let a 70+) jsou porovnávány v grafu na obrázku 36, který vykresluje průběh sezónně očištěné časové řady a trendové složky. Z průběhu křivek je na první pohled zřejmé, že osoby ve vyšší věkové skupině jsou více náchylné k výkyvům v úmrtnosti. Nárůsty počtu zemřelých odpovídají již zmíněným obdobím s vyšším výskytem respiračních onemocnění, označovaných jako chřipkové epidemie. I když je úroveň úmrtnosti v obou skupinách odlišná, na příkladu chřipkové epidemie z prosince roku 1995 můžeme dokázat vyšší náchylnost starších osob výpočtem. Zatímco u věkové skupiny 50–69 let se zvýšil měsíční počet zemřelých z průměru okolních měsíců 2600 úmrtí na prosincových přibližných

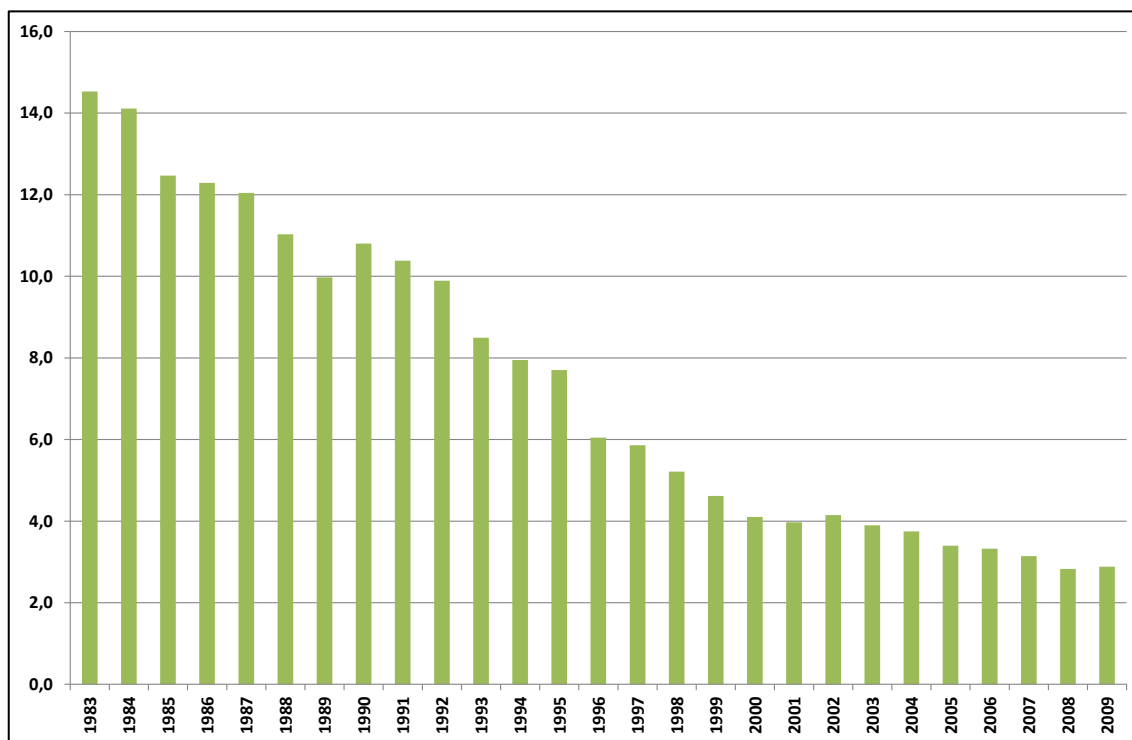
3200, tedy se jedná o nárůst 23 %, u skupiny osob starších 70 let došlo k nárůstu o 67 % (z průměrné hodnoty tohoto období 6000 na 8900 zemřelých osob v prosinci 1995).

**Obr. 34 – Sezónně očištěná časová řada, trendová složka, zemřelí do 1 roku, Česko, 1983–2009**



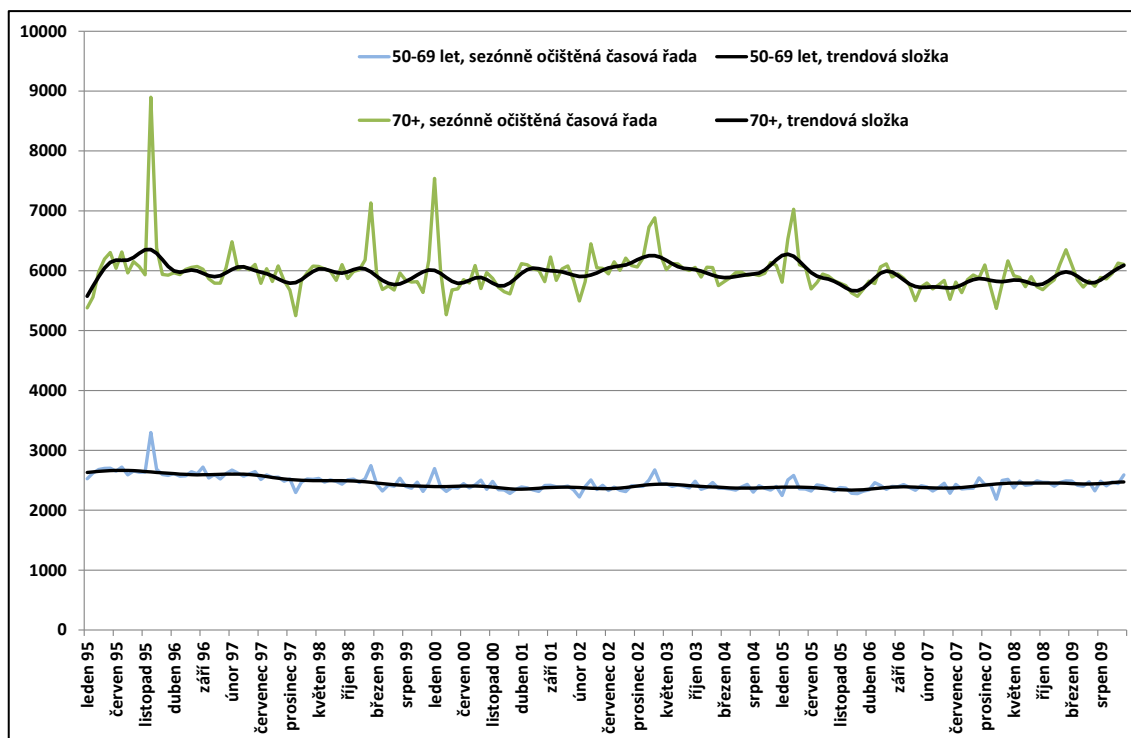
Zdroj: ČSÚ, vlastní výpočty

**Obr. 35 – Kvocient kojenecké úmrtnosti, Česko, 1983–2009**



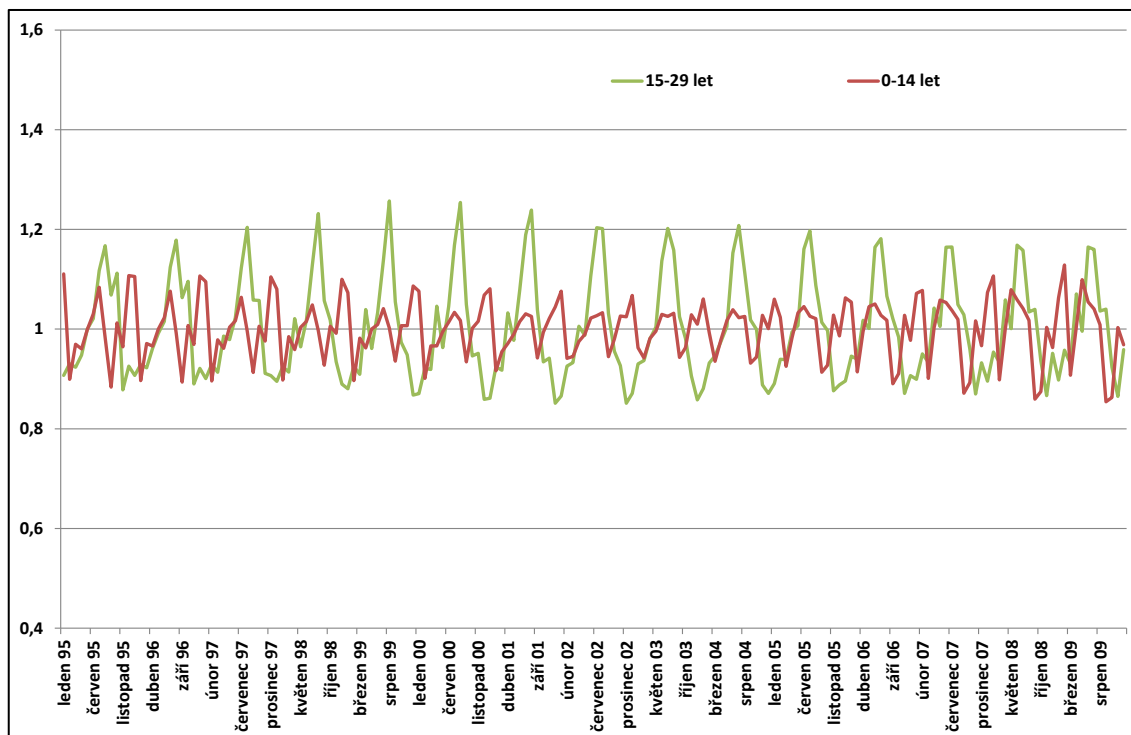
Zdroj: ČSÚ

Obr. 36 – Sezónně očištěná časová řada, trendová složka, zemřelí, 50–69 let, 70+, Česko, 1995–2009



Zdroj: ČSÚ, vlastní výpočty

Obr. 37 – Proměnlivé sezónní indexy, zemřelí, 0–14let, 15–29 let, Česko, 1995–2009

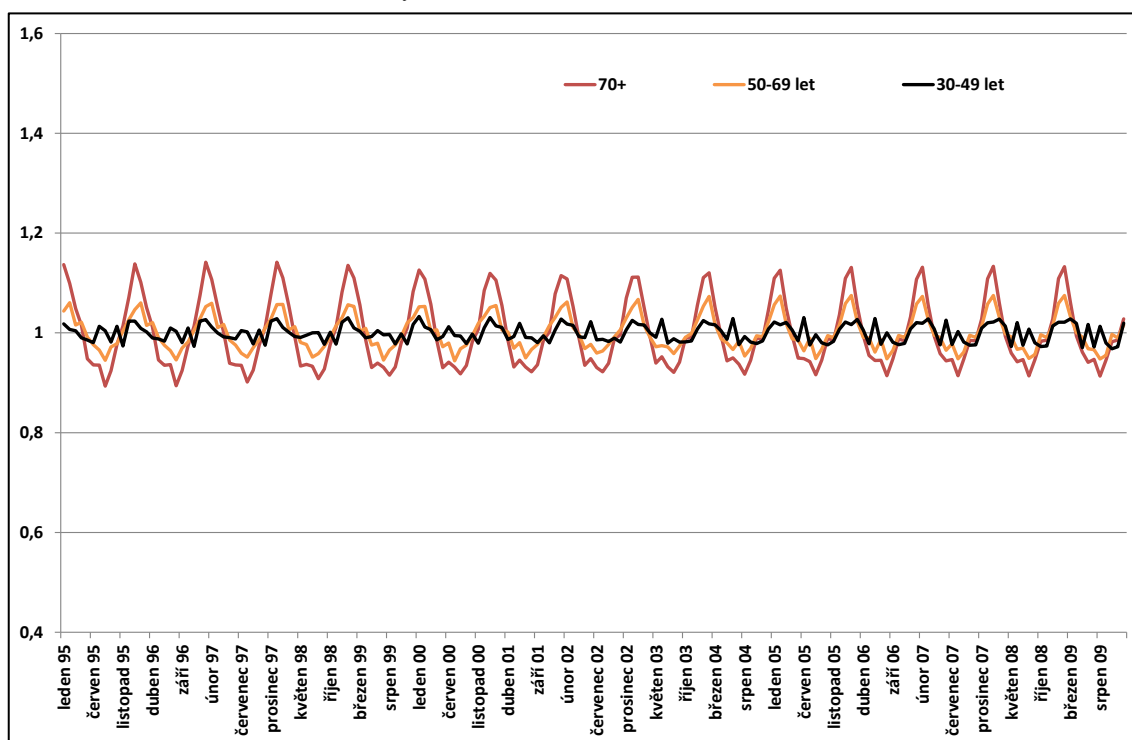


Zdroj: ČSÚ, vlastní výpočty

Proměnlivé sezónní indexy pro věkové skupiny 0–14 let a 15–29 let jsou znázorněné na obrázku 37. Zatímco u mladší věkové skupiny vidíme, že indexy nevykazují jasné

pravidelnosti, což je způsobeno absencí sezónnosti v časové řadě, u věkové skupiny 15–29 let vidíme jasnou převahu úmrtnosti v letních měsících, do roku 2001 v srpnu, a od roku 2002 střídající se maximální hodnotu mezi červencem a srpnem. Minimální hodnoty sezónních indexů vykazují měsíce zimní, prosinec a leden. Letní maximum úmrtnosti u mladých osob by pravděpodobně lépe ilustrovalo rozdělení podle příčiny úmrtí, tato data však nejsou dostupná. Úmrtí mladých lidí v letních měsících totiž souvisí především se smrtelnými úrazy a tragickými dopravními nehodami, které jsou často spojeny s teplým počasím a obdobím letních prázdnin a dovolených. Rozložení úmrtnosti věkové skupiny 15–29 let je naprosto odlišné od rozložení u starších věkových skupin, pro něž vidíme proměnlivé sezónní indexy vykreslené v grafu na obrázku 38. Liší se nejen rozložení, kdy oproti letnímu maximum u mladých osob jsou u osob ve vyšším věku maximální hodnoty indexů v zimních měsících, lednu a únoru, ale také hodnoty indexů. Zatímco červencové a srpnové indexy na obrázku 37 několikrát překročily hranici hodnoty 1,2, zimní maxima ani u nejstarších osob této hranice nedosáhly v žádném roce. Z tohoto grafu lze vyčíst zvyšování rozdílů mezi minimálními a maximálními hodnotami sezónních indexů s přibývajícím věkem zemřelých. Zatímco pro věkovou skupinu 30–49 let je v sezónních indexech zřejmé nízké rozpětí, se zvyšujícím se věkem roste i tato charakteristika. Především vysoké hodnoty nadprůměrných indexů souvisí s již zmiňovanou vyšší náchylností starších osob k nepříznivým vnějším vlivům.

**Obr. 38 – Proměnlivé sezónní indexy, zemřelí, 30–49 let, 50–69 let, 70+, Česko, 1995–2009**



Zdroj: ČSÚ, vlastní výpočty



### 6.1.3 Podle příčin úmrtí

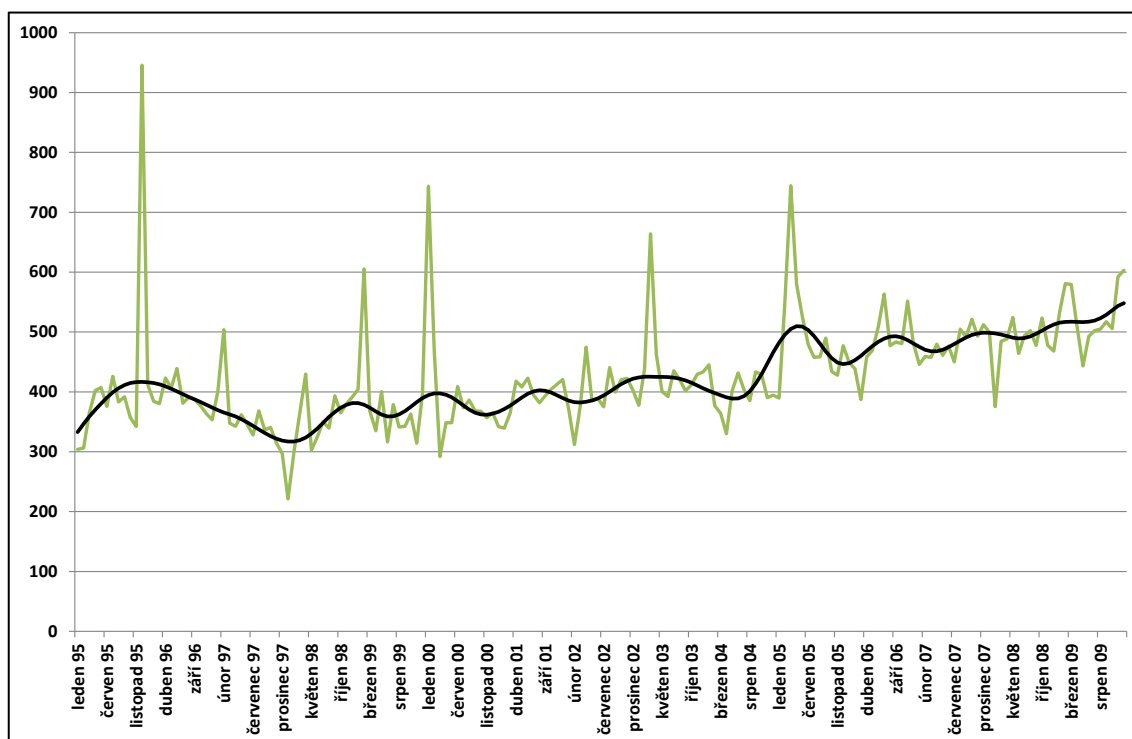
Pro ilustraci analýzy sezónnosti úmrtnosti podle příčin byly vybrány měsíční počty zemřelých pro dvě skupiny příčin úmrtí klasifikované podle 10. revize Mezinárodní statistické klasifikace nemocí a přidružených zdravotních problémů, zkratkou 10. revize MKN:

sk. IX. Nemoci oběhové soustavy (I00 - I99)

sk. X. Nemoci dýchací soustavy (J00 - J99)

Tyto dvě skupiny byly vybrány z několika důvodů. Prvním z nich je skutečnost, že nemoci oběhové soustavy patří v současnosti v Česku mezi nejčastější příčinu úmrtí, a nemoci dýchací soustavy bývají nejčastější příčinou úmrtí v důsledku vyššího výskytu respiračních onemocnění. Podle údajů ČSÚ zemřelo v roce 2009 na nemoci oběhové soustavy 56,3 % žen a 44,5 % mužů. Návaznost úmrtnosti z důvodu onemocnění dýchací soustavy na chřipkové epidemie ilustruje graf na obrázku 39. Dalším důvodem pro výběr těchto dvou příčin úmrtí byla jejich dostatečné zastoupení (počet zemřelých za měsíc), které je nezbytné k identifikaci jednotlivých složek časové řady. Dále se u těchto skupin dá očekávat sezónnost, tedy návaznost na teplotu a roční období (viz např. Myšáková, 2009).

**Obr. 39 – Sezónně očištěná časová řada, trendová složka, zemřelí, nemoci dýchací soustavy, Česko, 1995–2009**

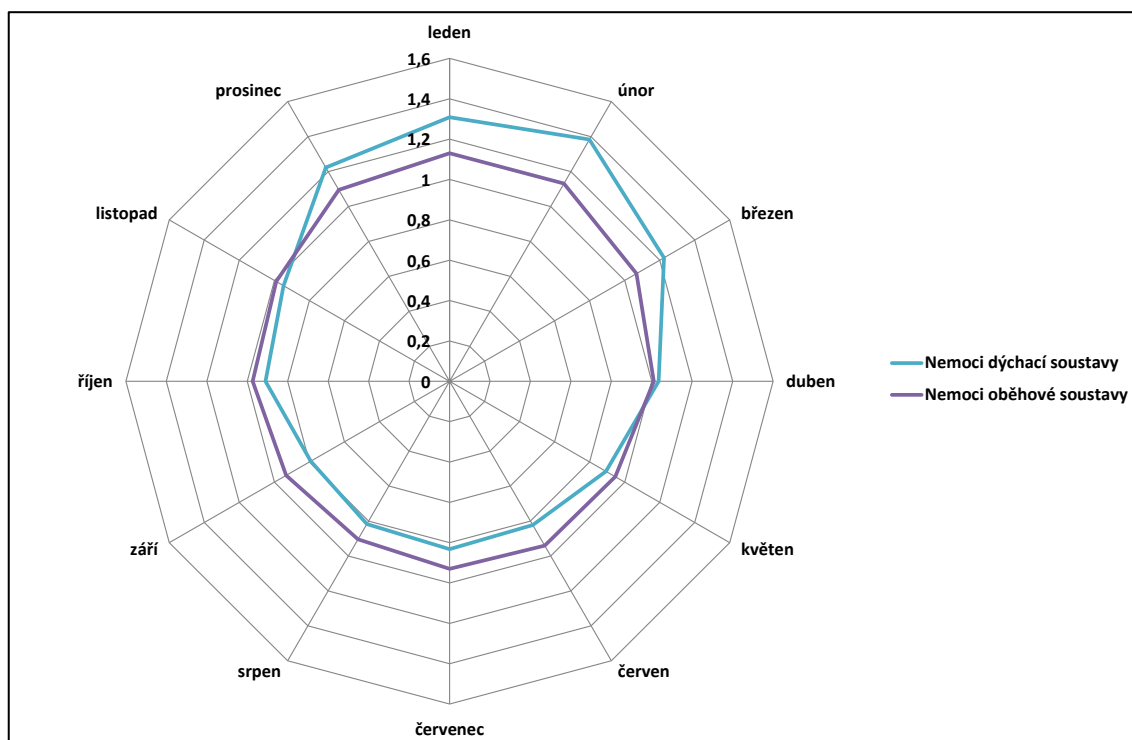


Zdroj: ČSÚ, vlastní výpočty

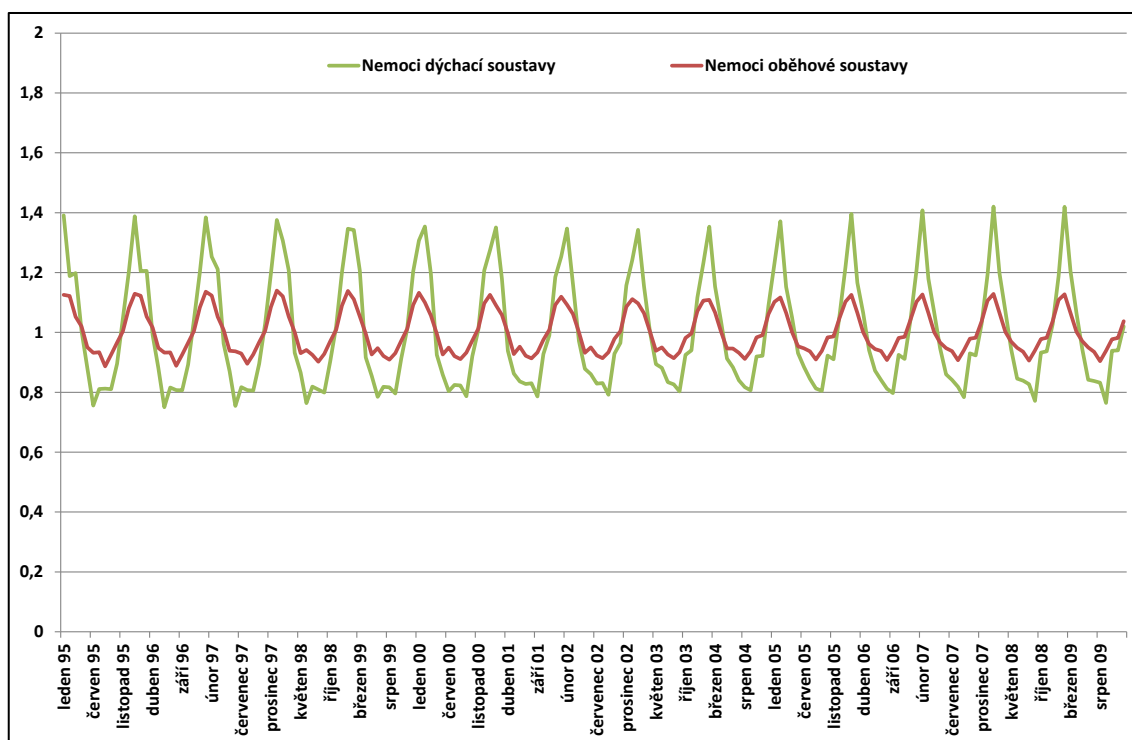
Rozložení počtu zemřelých do jednotlivých měsíců v roce ilustrují následující dva grafy, znázorněné na obrázcích 40 a 41. Na prvním z nich vidíme, že je úmrtnost na nemoci oběhové

soustavy rozložena během roku rovnoměrněji, než úmrtnost na nemoci dýchací soustavy. Zatímco mezi letními podprůměrnými hodnotami a zimními maximy v případě zemřelých na nemoci oběhové soustavy je úzké rozpětí, u druhé analyzované skupiny vidíme výraznou převahu nadprůměrných hodnot indexů v zimních měsících nad minimálními letními hodnotami. Tento zmiňovaný rozdíl je mnohem lépe patrný z grafu proměnlivých sezónních indexů na obrázku 41. U indexů pro úmrtí na nemoci dýchací soustavy vidíme změnu měsíce s minimální hodnotou sezónního indexu, který byl na začátku analyzovaného období pro červen a na konci intervalu jej vykazuje měsíc září, maximální hodnota je po celou dobu pro měsíc únor. V případě nemoci oběhové soustavy je po celé období nejčastějším měsícem úmrtí leden. Z hlediska nejnižšího sezónního indexu se o pomyslné prvenství střídavě dělí letní měsíce červenec, srpen a září.

**Obr. 40 – Sezónní indexy, zemřelí, nemoci dýchací soustavy, nemoci oběhové soustavy, Česko, 1995–2009**



Zdroj: ČSÚ, vlastní výpočty

**Obr. 41 – Proměnlivé sezónní indexy, zemřelí, nemoci dýchací soustavy, nemoci oběhové soustavy, Česko, 1995–2009**

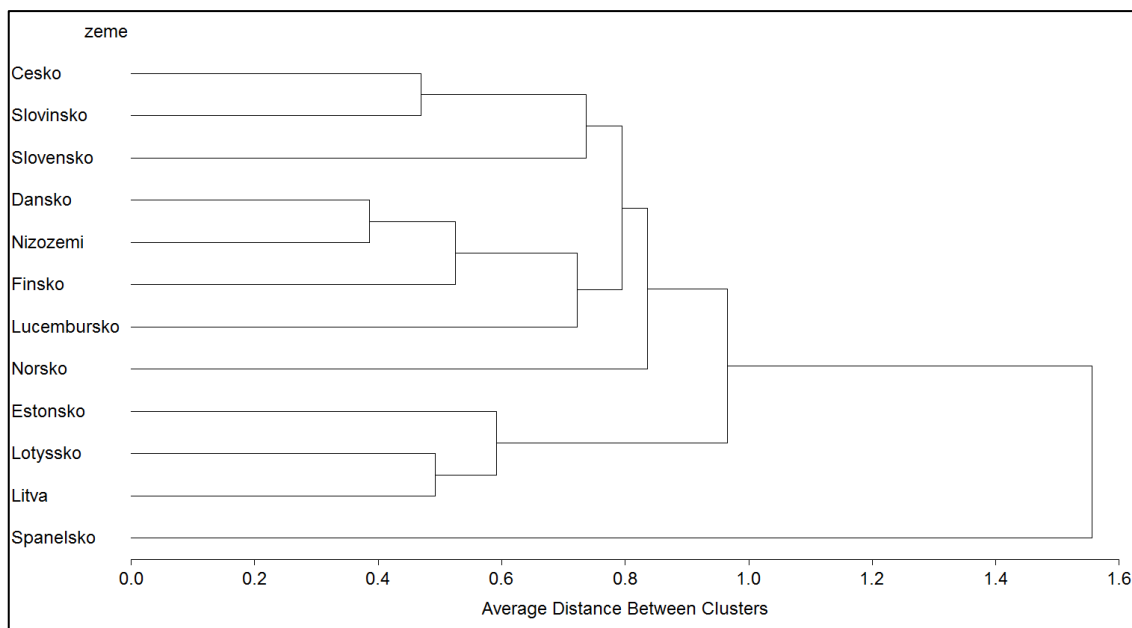
Zdroj: ČSÚ, vlastní výpočty

## 6.2 Evropské státy

Pro analýzu sezónnosti úmrtnosti u vybraných států, kdy jsou k dispozici pouze celkové počty zemřelých, byly na data aplikovány nejprve metody dekompozice časové řady a X12–ARIMA. Tak byly získány hodnoty sezónních indexů, proměnlivých sezónních indexů a náhodné složky. Na základě těchto ukazatelů pak byla, podobně jako v případě sňatečnosti, provedena shluková analýza, jejíž výsledky byly určující pro podrobnější hledání vnitřních zákonitostí vývoje časových řad.

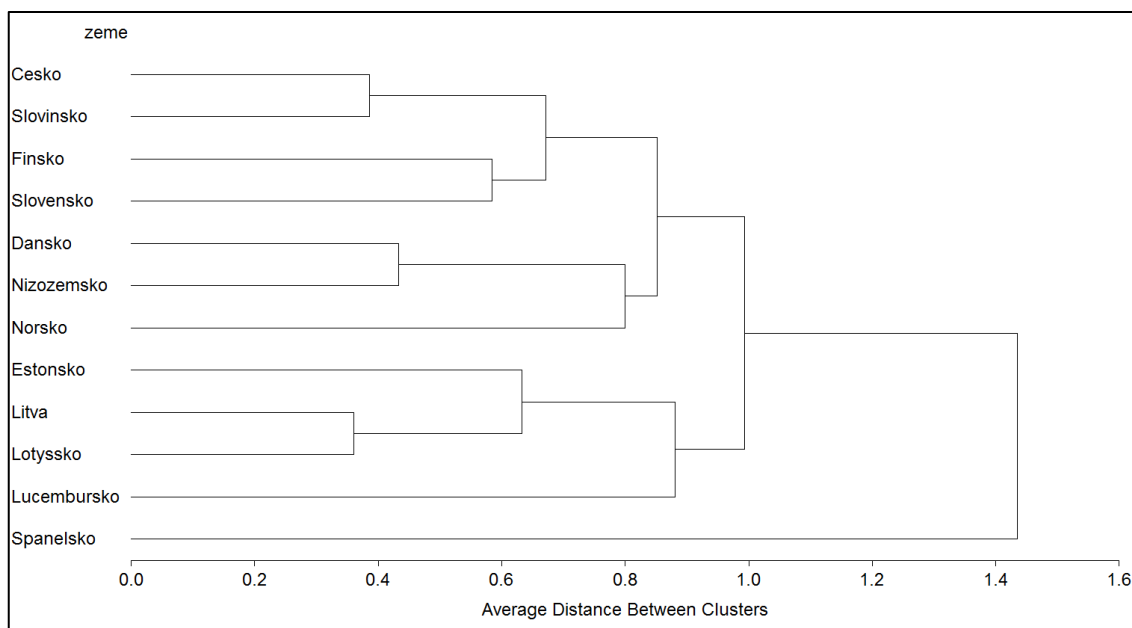
Výsledky shlukové analýzy použité na sezónní indexy úmrtnosti graficky vyjadřuje dendrogram znázorněný na obrázku 42. Na tomto grafu můžeme vidět tři homogenní shluky a dva státy stojící stranou zbývajících shluků. Těmito dvěma státy jsou Norsko a Španělsko, první shluk, který můžeme označit jako středoevropský, tvoří Česko, Slovensko a Slovinsko, druhý, severoevropský, shluk tvořený Dánskem, Finskem, Nizozemskem a třetí, pobaltský, shluk ve kterém je Estonsko, Litva a Lotyšsko. Nečekaně odlišný dendrogram pro proměnlivé sezónní indexy je znázorněný na obrázku 43. Odlehlým státem mimo ostatní shluky je jen Španělsko, zatímco složení ostatních shluků se změnilo. Do prvního shluku se k Česku, Slovensku a Slovinsku připojilo Finsko, v druhém bylo Finsko nahrazeno Norskem a tvoří tak shluk spolu s Dánskem a Nizozemskem, a třetí, pobaltský, shluk doplnilo Lucembursko, i když jeho vzdálenost od Estonska, Litvy a Lotyšska je relativně velká.

**Obr. 42 – Dendrogram vybraných evropských zemí, sezónní indexy, zemřelí, 2000–2007**



Zdroj: EUROSTAT, vlastní výpočty

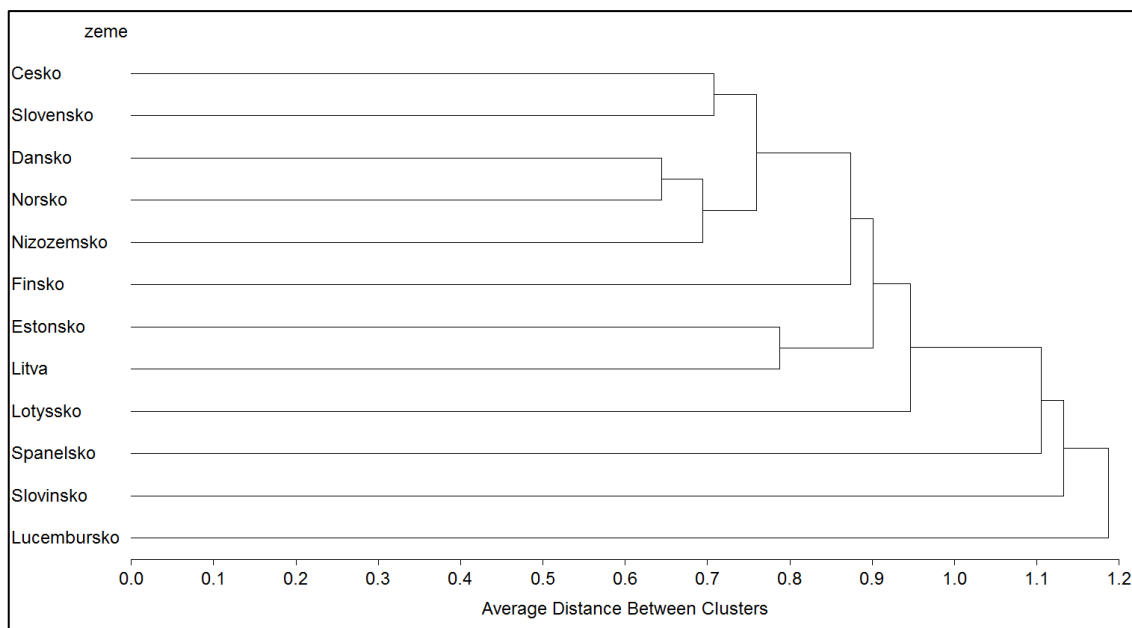
**Obr. 43 – Dendrogram vybraných evropských zemí, proměnlivé sezónní indexy, zemřelí, 2000–2007**



Zdroj: EUROSTAT, vlastní výpočty

Dendrogram vzniklý na základě dat o náhodných složkách jednotlivých států je vyobrazený na obrázku 44. O tomto grafu lze říci, že jej tvoří dva shluky, prvním je shluk se státy Česko, Slovensko, Dánsko, Norsko a Nizozemsko, druhým pak malý shluk zahrnující Estonsko a Litvu. Ostatní státy stojí samostatně a s většími či menšími odstupy se postupně „nabalují“ na první, jakýsi hlavní shluk.

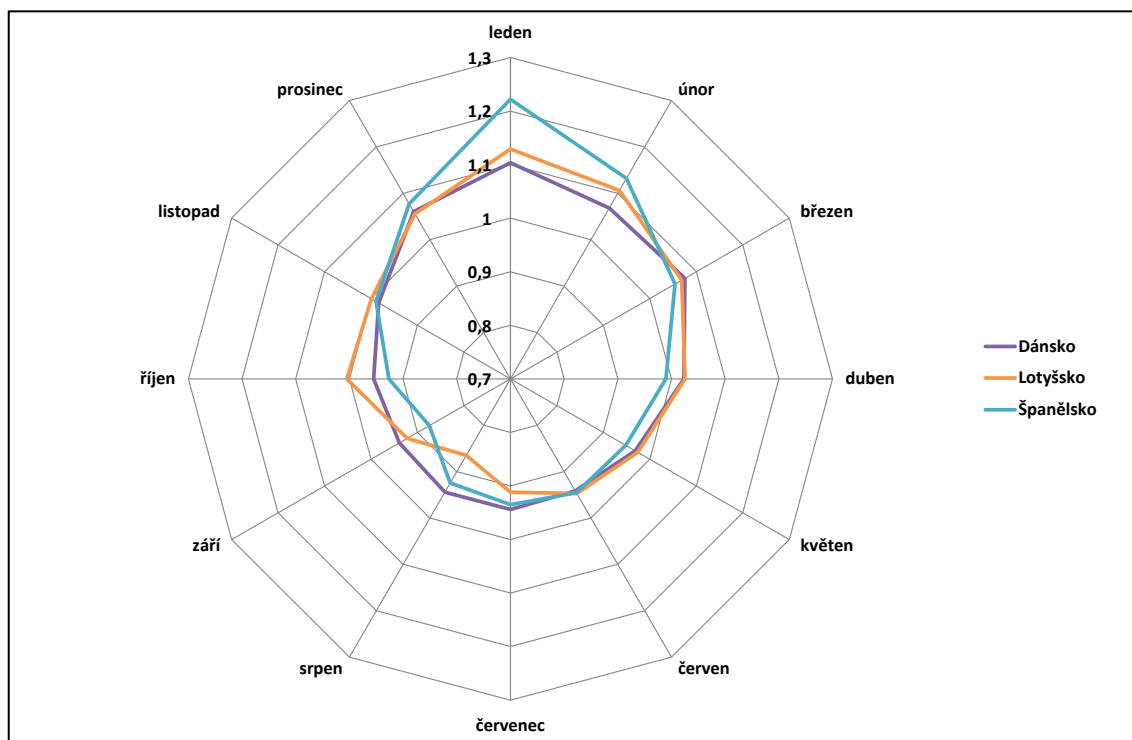
**Obr. 44 – Dendrogram vybraných evropských zemí, náhodná složka, zemřelí, 2000–2007**



Zdroj: EUROSTAT, vlastní výpočty

Na základě výsledků shlukové analýzy založené na sezónních indexech byl do obrázku 45 vykreslen graf sezónních indexů navzájem nejvíce vzdálených zemí. Z něj vidíme, že se Španělsko opravdu odlišuje rozložením zemřelých během roku, neboť lednové maximum vysoce převyšuje hodnoty ostatních zimních indexů.

**Obr. 45 – Sezónní indexy, zemřelí, Dánsko, Lotyšsko, Španělsko, 2000–2007**



Zdroj: EUROSTAT, vlastní výpočty

## 7 Závěr

Cílem této práce bylo představení několika metod analýzy časových řad a názorná ukázka jejich aplikace na časové řady demografických jevů. Porodnost, sňatečnost a úmrtnost populací Česka a vybraných evropských států, byly třemi vybranými demografickými procesy, u kterých byla snaha odhalit a popsat sezónnost. Výběr těchto tří demografických jevů byl založen na důkladném studiu jak české, tak zahraniční literatury o sezónnosti, aby do analýzy zbytečně nevstupovaly jevy, u nichž není ani teoretické odůvodnění pravidelnosti v sezónním rozložení událostí během roku.

Metodologická část popisuje nejen samotné metody analýzy časových řad, ale obsahuje také teoretický základ o problematice časových řad, ať už samotné vymezení časové řady, popis statistických charakteristik řady, problémů s časovými řadami spojených nebo standardizaci hodnot časových řad. Tři vybrané časové řady jsou popsány podrobně, i s ukázkou výpočtu dané metody v prostředí statistického softwaru SAS. Metoda tradiční dekompozice časové řady je popsána pro všechny modely, které lze pro výpočet použít, v rozhraní SAS Enterprise Guide je pak znázorněna multiplikativní dekompozice, která je využita ve všech výpočtech dekompozice v práci. Metoda X12-ARIMA, díky které jsou v analytické části grafy proměnlivých indexů, je metodou zkonstruovanou primárně pro sezónní očišťování dat a její výpočet v softwaru SAS je možný s pouze fungujícím modulem SAS/ETS. Kointegrace časových řad je metodou, která odhaluje krátkodobé nebo dlouhodobé vztahy mezi časovými řadami. V práci byla aplikována několikrát, nikdy však nenašla mezi dvěma vybranými časovými řadami žádný kointegrační proces. V některých případech nemohl být samotný test kointegrace použit, a to z důvodu nesplnění podmínek nestacionarity časových řad, která byla testována pomocí testu jednotkového kořene Dickey-Fullerova testu. Podrobněji je popsána také shluková analýza, která byla v práci aplikována na data evropských států o sňatečnosti a úmrtnosti. Čtyři další metody analýzy časových řad a jejich využití při analýze sezónnosti jsou představeny jen krátce. Jedná se o grafickou analýzu, model SARIMA Box-Jenkinsovy metodologie, lineární dynamické modely a spektrální analýzu.

Analytická část této práce je rozdělena do třech kapitol, které vždy nesou název demografického procesu, jež je v rámci kapitoly analyzován. U všech časových řad vystihujících porodnost v Česku a evropských zemích byla identifikována sezónnost. Vývoj počtu živě narozených v Česku ilustruje prudký pokles během 90. let 20. století a následný nárůst v důsledku reprodukce žen pocházejících ze silných kohort narozených během 70. let. Sezónnost porodnosti však ani v těchto trendově odlišných obdobích není změněna a přetrvává tak koncentrace živě narozených do jarních měsíců březen, duben květen. U jednotlivých

evropských států lze identifikovat odlišný vývoj v porodnosti, který se liší především v závislosti na skutečnosti, zda daná země byla na východní či západní straně železné opony. Stejně jako u porodnosti, i ve všech časových řadách sňatečnosti byla identifikována přítomnost sezónnosti. Analýza časové řady sňatečnosti v Česku odhalila již známé trendy nízké sňatečnosti v zimních měsících, odmítání květnových svateb z důvodu pověry „svatba v máji, do roka máry“ a oblibu atraktivního data, kdy je datum svatby tvořeno co nejvyšším počtem shodných číslic. Tento relativně nový faktor v rozhodování o termínu svatby byl zaznamenán i v některých evropských zemích, např. ve Finsku, ale není to případ všech analyzovaných zemí. Evropské země byly rozděleny do třech různě dlouhých časových období, a to podle dostupnosti dat. Posledním krokem v analýze sňatečnosti pak byla aplikace shlukové analýzy, postupně na sezónní indexy, proměnlivé sezónní indexy a hodnoty náhodné složky jednotlivých států pro období 2001–2009. Výsledné dendogramy odhalily populace, které si jsou svým sňatkovým chováním blízké. Složení jednotlivých shluků bylo shodné při použití jednoduchých a proměnlivých sezónních indexů jako vstupních dat, a odlišovalo se od třetího dendogramu. Analýza úmrtnosti byla opět provedena zvlášť pro Česko a ostatní evropské státy, tentokrát se však lišil obsah časových řad. Zatímco pro evropské země byly použity pouze celkové počty zemřelých, za Česko byla úmrtnost diferencována podle pohlaví, věkových skupin a dvou vybraných příčin úmrtí. Rozložení časových řad na jednotlivé složky pak odhalilo pravidelnosti, a to v ročním rozložení úmrtí především do zimních měsíců, a také některé podobnosti mezi jednotlivými skupinami osob. Evropská data byla opět podrobena shlukové analýze a na základě jejích výsledků vykresleny grafy zástupců jednotlivých shluků. Jedinou časovou řadou v celé práci, u které nebyla potvrzena přítomnost sezónnosti, je počet zemřelých ve věkové skupině 0–14 let v Česku v letech 1995–2009.

Tato práce představila pokročilé statistické metody, které lze pomocí softwarového vybavení aplikovat na demografické časové řady a získat tak nový, detailnější pohled na generování hodnot jednotlivých časových řad. Odhalení těchto vnitřních zákonitostí vývoje může být využito pro vědecké účely, např. populační prognózy, nebo v aplikované sféře.

## 8 Seznam použité literatury

- ARLT, J.; ŠKUTHANOVÁ, M. 1995. Úvod do problematiky sezónního očišťování ekonomických časových řad. *Acta oeconomica pragensia*. 1995, roč. 3, č. 1, str. 15-23.
- ARLT, J. 1997. Kointegrace v jednorovnicových modelech. *Politická ekonomie*. 1997. roč. 45, č. 5, s. 733-746. Praha: VŠE, 1997.
- ARLT, J. 1998. Časové řady typu  $I(0)$  a  $I(1)$ . *Acta oeconomica pragensia*. 1998, roč. 6, č. 2, str. 7-11.
- ARLT, J.; ARLTOVÁ, M.; RUBLÍKOVÁ, E. 2002. Analýza ekonomických časových řad s příklady. Praha: VŠE, 2002. ISBN 80-245-0307-7.
- ARLT, J.; ARLTOVÁ, M.; 2005. Ekonomické časové řady. Praha: Grada Publishing, a. s. ISBN 978-80-247-1319-9.
- ARLT, J. 2010. Přednášky předmětu Časové řady. Praha: Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky. 18. 18. 2010.
- Australian Bureau of Statistics. 2011. Time Series Analysis: Seasonal Adjustment Methods. Canberra, 2008. [cit. 2011-3-12] Dostupný z WWW: <<http://www.abs.gov.au/websitedbs/d3310114.nsf/4a256353001af3ed4b2562bb00121564/c890aa8e65957397ca256ce10018c9d8!OpenDocument#X12ARIMA>>
- BARTOŇOVÁ, D. et al. 2008. Populační vývoj České republiky 2007. Praha: Univerzita Karlova v Praze, Přírodovědecká fakulta, 2008. ISBN 978-80-86561-46-2.
- BŘÍZA, M. 2010. Analýza výkonnosti firmy MIDA, a.s. pomocí časových řad. Bakalářská práce (Bc.). Brno: Vysoké učení technické v Brně. Fakulta podnikatelská, 2010.
- CHINA DAILY 2010. Tens od thousands of Chinese love birds tie knot on 'perfect day'. 11. 10. 2010 [cit. 2011-4-29] Dostupný z WWW: <[http://www.chinadaily.com.cn/china/2010-10/11/content\\_11391539.htm](http://www.chinadaily.com.cn/china/2010-10/11/content_11391539.htm)>
- CIPRA, T. 1986. Analýza časových řad s aplikacemi v ekonomii. Praha: SNTL - Nakladatelství technické literatury, 1986.
- COPPA, A.; DONATO, L. D.; VECCHI, F.; DANUBIO, M. E. 2001. Seasonality of Marriages and Ecological Contexts in Rural Communities of Central-Southern Italy (Abruzzo), 1550–1871. *Collegium Antropologicum*. 2001, roč. 25, č. 2, s. 403-412.
- ČSÚ 2006. Porodnost a plodnost 2001–2005. Praha: ČSÚ, 2006. [cit. 2011-4-6] Dostupný z WWW: <<http://www.czso.cz/csu/2006edicniplan.nsf/p/4008-06>>
- ČSÚ 2010. Vývoj obyvatelstva České republiky v roce 2009. Praha: ČSÚ, 2010. [cit. 2011-4-7] Dostupný z WWW: <<http://www.czso.cz/csu/2010edicniplan.nsf/publ/4007-10->>



- ČSÚ 2011. Demografie – pramenná díla. Praha: ČSÚ, 2011. [cit. 2011-4-9] Dostupný z WWW: <[http://www.czso.cz/csu/redakce.nsf/i/casova\\_rada\\_demografie](http://www.czso.cz/csu/redakce.nsf/i/casova_rada_demografie)>
- DANUBIO, M. E. 2001. Biodemographic study of a central Apennine area (Italy) in the 19th and 20th centuries: marriage seasonality and reproductive isolation. *Journal of Biosocial Science*. 2001, roč. 33, č. 3, s. 427-449.
- Demografický informační portál 2011. Demografické procesy. Praha: Demografický informační portál, 2011. [cit. 2011-2-16] Dostupný z WWW: <[http://www.demografie.info/?cz\\_procesy](http://www.demografie.info/?cz_procesy)>
- DUFKOVÁ, M.; VRÁBLÍK, M. 2001. Inovace a nová ekonomika. Příspěvek z konference TVŮRČÍ ODKAZ TOMÁŠE BATI a současné podnikatelské metody, Zlín, 17. 5. 2001.
- DZÚROVÁ, D. 2008. Shluková analýza. Přednášky předmětu Statistická analýza dat II. Praha: Univerzita Karlova v Praze. Přírodovědecká fakulta, 5. 12. 2008.
- EILERS, P. H. C.; GAMPE, J.; MARX, B. D.; RAU, R. 2008. Modulation models for seasonal time series and incidence tables. *Statistics in Medicine*. 2008, vol. 27, no. 17, s. 3430-3441. [cit.2011-4-10] Dostupný z WWW: <[http://www.demogr.mpg.de/publications/files/2943\\_1213957725\\_1\\_StatMed%2027\(2008\)17%203430.pdf](http://www.demogr.mpg.de/publications/files/2943_1213957725_1_StatMed%2027(2008)17%203430.pdf)>
- ELLISON, P. T.; VALEGGIA, C. R.; SHERRY, D. S. 2005. Human birth seasonality. Kapitola v knize *Seasonality in Primates*. University of North Carolina, 2005. ISBN 9780511542343.
- FEINSTEIN, C. A. 2002. Seasonality of Deaths in the U.S. by Age and Cause. *Demographic Research*. 2002, vol. 6, s. 469-486. [cit.2011-4-10] Dostupný z WWW: <<http://www.demographic-research.org/volumes/vol6/17/>>
- FIALOVÁ, L. 1995. Sezónnost demografických událostí v českých zemích v 17. až 20. století. *Demografie*. 1995, roč. 37, č. 1, s. 9-21.
- FIALOVÁ, L.; HORSKÁ, P.; KUČERA, M.; et al. 1996. Dějiny obyvatelstva českých zemí. Praha: Mladá fronta, 1996. ISBN 80-204-0720-0.
- GABRHELÍKOVÁ, M. 2010. Statistická analýza vybraných demografických ukazatelů v České republice. Bakalářská práce (Bc.). Olomouc: Univerzita Palackého v Olomouci. Přírodovědecká fakulta, 2010.
- GRECH, V.; SAVONA-VENTURA, CH. et al. 2003. Seasonality od Births is associated with Seasonality of Marriages in Malta. *Journal of Biosocial Science*. 2003, roč. 35, s. 95-105.
- GRIVALSKÁ, Z. 2008. Porovnanie prognostickej schopnosti modelov při využití metod sezónneho očšťovania. Diplomová práce (Ing.). Bratislava: Univerzita Komenského v Bratislavě. Fakulta matematiky, fyziky a informatiky, 2008.
- HANČLOVÁ, J.; TVRDÝ, L. 2003. Úvod do analýzy časových řad. Ostrava: Vysoká škola báňská, 2003.
- HENDL, J. 2006. Přehled statistických meto zpracování dat. Praha: Portál, s.r.o., 2006. ISBN 80-7367-123-9.
- HRDLIČKA, M. 2005. Schizofrenie v dětském a adolescentním věku. *Pediatric pro praxi*. 2005, roč. 5, č. 5, s. 240-242.

- HŮLE, D. 2005. Pověra v srdci globálního ateismu. Demografický informační portál. [cit. 2009-3-18] Dostupný z WWW: <[http://www.demografie.info/?cz\\_detail\\_clanku&artclID=41](http://www.demografie.info/?cz_detail_clanku&artclID=41)>
- HUYNEN, M. M. T. E.; MARTENS, P.; SCHRAM, D.; WEIJENBERG, M. P.; KUNST, A. E. 2001. The Impact of Heat Waves and Cold Spells on Mortality Rates in the Dutch Population. *Environmental Health Perspectives*. 2001, vol. 109, no. 5, s. 463-470. [cit. 2011-4-10] Dostupný z WWW: <<http://www.ehponline.org/members/2001/109p463-470huynen/huynen.pdf>>
- IVARSSON, A.; HERNELL, O.; NYSTROM, L.; PERSSON, L. A. 2003. Children born in the summer have increased risk for coeliac disease. *Journal of Epidemiology & Community Health*. 2003, roč. 57, s. 36-39.
- JONGBLOET, P. H.; GROENEWOUD, H. M.; HIRASING, R. A.; VAN BUUREN, S. 1998. Seasonality of birth in patients with childhood diabetes in The Netherlands. *Diabetes Care*. 1998, roč. 21, č. 1, s. 190-191.
- KAČEROVÁ, E. 2001. Sezónnost sňatečnosti. Bakalářská práce (Bc.). Praha: Univerzita Karlova v Praze. Přírodovědecká fakulta, 2001.
- KAČEROVÁ, E. 2004. Mezinárodní srovnání sezónnosti sňatečnosti. *Demografie*. 2004, roč. 46, č. 3, s. 186-198.
- KASALOVÁ DAŇKOVÁ, Š. 2005. Sezónní aspekty sebevraždění. *Demografický informační portál*. [cit. 2011-4-10] Dostupný z WWW: <[http://www.demografie.info/?cz\\_detail\\_clanku&artclID=133](http://www.demografie.info/?cz_detail_clanku&artclID=133)>
- KOHOUT, P. 2006. Diagnostika a léčba celiakie. *Interní medicína pro praxi*. 2006, roč. 7, č. 7, str. 324-326.
- KONEČNÁ, A. 1977. Sezónnost sňatečnosti v ČSSR. *Demografie*. 1977, roč. 19, č. 3, s. 215-222 a s. 302-307.
- KOZÁK, J.; HINDLS, R.; ARTL, J. 1994. Úvod do analýzy ekonomických časových řad. Praha: VŠE, 1994. ISBN 80-7079-760-6.
- KVASNIČKA, M.; VAŠÍČEK, O. 2001. Úvod do analýzy časových řad. [cit. 2008-12-14] Dostupné z WWW: <<http://www.econ.muni.cz/~qasar/vyuka/emm2/skriptaemmii.pdf>>
- KYSELÝ, J.; HUTH, R.; KRÍŽ, B. 2002. Úmrtnost spojená se stresem z horka v ČR – první výsledky. XIV. Česko-slovenská bioklimatologická konference. Lednice na Moravě. 2.-4. září 2002. Dostupné z WWW: <<http://old.chmi.cz/meteo/CBKS/sbornik02/kysely.pdf>>
- LERCHL, A. 1998. Changes in the seasonality of mortality in Germany from 1946 to 1995: the role of temperature. *International Journal of Biometeorology*. 1998, vol. 42, s. 84-88. [cit. 2011-4-10] Dostupný z WWW: <<http://www.springerlink.com/content/fx524e6q92r196ba/>>
- MYŠÁKOVÁ, G. 2009. Vývoj sezónnosti úmrtnosti podle příčin v Česku, Norsku a Španělsku. Bakalářská práce (Bc.). Praha: Univerzita Karlova v Praze. Přírodovědecká fakulta, 2009.

- MYŠÁKOVÁ, G.; TESÁRKOVÁ, K. 2010. Analýza sezónnosti v demografii se zaměřením na úmrtnost. *Demografie*, 2010, roč. 51, č. 2, str. 90-102.
- PAVLÍK, Z.; RYCHTAŘÍKOVÁ, J.; ŠUBRTOVÁ, A. 1986. *Základy demografie*. Praha : Academia, 1986.
- PELL, J. C.; COBBE, S. M. 1999. Seasonal variations in coronary heart disease. *QJM: An International Journal of Medicine*. 1999, vol. 92, s. 689-696. [cit. 2011-4-10] Dostupný z WWW: <<http://qjmed.oxfordjournals.org/cgi/reprint/92/12/689>>
- POLÁŠEK, V. 2005 Silnější než pověra – změny v sezónnosti sňatků. *Zpravodaj České demografické společnosti*, č. 35/2005.
- RAU, R.; DOBLHAMMER, G. 2003. Seasonal mortality in Denmark: the role of sex and age. *Demographic Research*. 2003, vol. 9, s. 197-222. [cit. 2011-4-10] Dostupný z WWW: <<http://www.demographic-research.org/volumes/vol9/9/9-9.pdf>>
- RAU, R. 2004. Winter mortality in elderly people in Britain: lack of social gradient in winter excess mortality is obvious in Denmark. *BMJ*. 2004, vol. 329, s. 976-977. [cit. 2011-4-10] Dostupný z WWW: <<http://www.bmj.com/cgi/content/extract/329/7472/976-b>>
- RAU, R. 2007. *Seasonality in Human Mortality: A demographic Approach*. Rostock : Springer, 2007. [cit. 2011-4-10] Dostupný z WWW: <<http://www.demogr.mpg.de/books/drm/003/index.htm>>. ISBN 978-3-540-44900-3>
- ROTHWELL, P. M.; GUTNIKOV, S. A.; MCKINNEY, P. A.; et al. 1999. Seasonality of birth in children with diabetes in Europe: multicentre cohort study. *British Medical Journal*. 1999, č. 319, s. 887-888.
- ROUČKA, M.; SKOČDOPOLOVÁ, R. 1990. Vývoj sezónnosti sňatečnosti v Československu po druhé světové válce a současná situace v mezinárodním srovnání. *Demografie*. 1990, roč. 32, č. 2, s. 116-125.
- SAS Institute Inc. 2010. *SAS/ETS® 9.22 User's Guide*. Cary, NC: SAS Institute Inc. ISBN 978-1-60764-543-6.
- SCAFETTA, N.; RESTREPO, E.; WEST, B. J. 2003. Seasonality od Birth and Conception to Teenagers in Texas. *Social Biology*. 2003, roč. 50, č. 1-2, s. 1-22.
- ŠPÉROVÁ, L. 2008. *Cirkadiánní rytmy u člověka. Bakalářská práce (Bc.)*. Brno: Masarykova Univerzita v Brně. Přírodovědecká fakulta. 2008.
- ŠŤASTNÁ, A. 2010. Změny reprodukčních vzorců a individuální souvislosti rodičovství. Příspěvek z XL. konference České demografické společnosti „Dvacet let sociodemografické transformace“. Brno, Fakulta sociálních studií, 27. 5. 2010.
- TESÁRKOVÁ, K. 2008. *Sezónnost demografických událostí. Příspěvek ze semináře k vybraným demografickým tématům*. Praha: Univerzita Karlova v Praze. Přírodovědecká fakulta, 3. 12. 2008.

- TESÁRKOVÁ, K.; KAROUSOVÁ, E. 2009a. Vývoj sezónnosti sňatečnosti za 50 let časopisu Demografie. Sborník příspěvku XXXVIII. konference České demografické společnosti „Padesát let časopisu „Demografie“. Praha: ČSÚ. 2009, s. 103-114.5
- TESÁRKOVÁ, K. 2009b. Základy ekonometrie. Přednášky předmětu Základy ekonometrie. Praha: Univerzita Karlova v Praze. Přírodovědecká fakulta, 19. 11. 2009.
- The Eurowinter Group. 1997. Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. *The Lancet*. 1997, vol. 348, s. 1341-1346. [cit. 2011-4-10] Dostupný z WWW: <[http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(96\)12338-2/fulltext](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(96)12338-2/fulltext)>
- TORCHE, F.; CORVALAN, A. 2010. Seasonality of Birth Weight in Chile: Environmental and Socioeconomic Factors. *Annals of Epidemiology*. 2010, roč. 20, s. 818-826.
- TROVATO, F.; ODYNAK, D. 1993. The Seasonality of Births in Canada and teh Provinces, 1881-1989: Theory and Analysis. *Canadian Studies in Population*. 1993, roč. 20, č. 1, s. 1-41.
- U. S. Census Bureau. 2001. X12-ARIMA Reference Manual. Washington DC, Statistical Research Division. Dostupný z WWW: <<http://www.census.gov/srd/www/x12a/>>
- WELLS, R. V. 1987. Marriage Seasonal in Early America. *Journal of Interdisciplinary History*. 1987, roč. 18, č. 2, s. 299-307.
- WEERASINGHE, D. P.; MACINTYRE, C. R.; RUBIN, G. L. 2002. Seasonality of coronary artery deaths in New South Wales, Australia. *Heart*. 2002, vol. 18, s. 30-34. [cit. 2011-4-10] Dostupný z WWW: <<http://heart.bmj.com/cgi/reprint/88/1/30>>
- ZATLOUKAL, R. 2008. Analýza a předpověď časových řad pomocí statistických metod se zaměřením na metodu Box-Jenkins. Diplomová práce (Ing.). Brno: Vysoké učení technické v Brně. Fakulta strojního inženýrství. 2008.

---

<sup>5</sup> Sborník příspěvků XXXVIII. konference České demografické společnosti „Padesát let časopisu „Demografie“ je na CD ROM přílohou časopisu Demografie 1/2009.

## **9 Přílohy**

Všechna zdrojová data, která byla v této diplomové práci použita pro výpočty a analýzy, jsou dostupná na přiloženém CD.