

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Analýza střední délky života ve vybraných
zemích



Katedra matematické analýzy a aplikací matematiky
Vedoucí bakalářské práce: **RNDr. Ondřej Pavlačka, Ph.D.**
Vypracoval(a): **Natálie Mynarčíková**
Studijní program: B1103 Aplikovaná matematika
Studijní obor Matematika–ekonomie se zaměřením na bankovníctví/pojišťovnictví
Forma studia: prezenční
Rok odevzdání: 2022

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Natálie Mynarčíková

Název práce: Analýza střední délky života ve vybraných zemích

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: RNDr. Ondřej Pavlačka, Ph.D.

Rok obhajoby práce: 2022

Abstrakt: Tématem bakalářské práce je analýza střední délky života ve vybraných zemích. V práci je vysvětlena samotná demografie a problematika střední délky života. Pro námi zvolené země je provedena analýza vývoje středních délek života, které jsou následně srovnány. Pomocí korelační analýzy je zkoumána závislost střední délky života na hrubém domácím produktu na osobu a na úhrnné plodnosti.

Klíčová slova: Demografie, střední délka života, hrubý domácí produkt na osobu, úhrnná plodnost, korelační analýza, hierarchická shluková analýza

Počet stran: 52

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Natálie Mynarčíková

Title: Analysis of life expectancy in selected countries

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: RNDr. Ondřej Pavlačka, Ph.D.

The year of presentation: 2022

Abstract: The topic of this bachelor thesis is the analysis of life expectancy at birth in selected countries. The thesis explains demography and life expectancy at birth issues. For the selected countries an analysis of the development of life expectancy at birth is performed, and these are then compared. Using correlation analysis, the dependence of life expectancy at birth on gross domestic product per capita and on fertility rate is examined.

Key words: Demographics, life expectancy at birth, gross domestic product per capita, fertility rate, correlation analysis, hierarchical cluster analysis

Number of pages: 52

Number of appendices: 1

Language: Czech

Prohlášení

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

V Olomouci dne

.....

podpis

Obsah

Úvod	7
1 Demografické a makroekonomické pojmy	9
1.1 Demografie	9
1.1.1 Úmrtnostní tabulky a střední délka života	9
1.1.2 Úhrnná plodnost	13
1.2 Makroekonomie	13
1.2.1 Hrubý domácí produkt	14
2 Aplikovaný matematický aparát	15
2.1 Korelační analýza	15
2.1.1 Spearmanův korelační koeficient	16
2.1.2 Parciální korelační koeficient	18
2.2 Shluková analýza	19
2.2.1 Hierarchická shluková analýza	20
3 Analýza střední délky života	22
3.1 Vývoj střední délky života v průběhu času	23
3.1.1 Hierarchická shluková analýza vývoje střední délky života pro muže	23
3.1.2 Hierarchická shluková analýza vývoje střední délky života pro ženy	27
3.2 Vývoj střední délky života ve vybraných letech - vykreslení do mapy	31
3.2.1 Svět	31
3.2.2 Evropa	34
3.3 Vývoj závislosti mezi střední délkou života a hrubým domácím produktem	36
3.4 Vývoj závislosti mezi střední délkou života a úhrnnou plodností	41
3.5 Korelační analýza pro vybrané faktory	46
Závěr	49
Literatura	50

Poděkování

Ráda bych na tomto místě poděkovala vedoucímu své bakalářské práce panu RNDr. Ondřeji Pavlačkovi, Ph.D. za odborné vedení práce, cenné rady a poznámky, a především za veškerý čas, který mi věnoval.

Úvod

Tématem této bakalářské práce je analýza střední délky života ve vybraných zemích. Budeme analyzovat vývoj střední délky života v porovnání ve vybraných zemích dle dostupných demografických dat. Dále budeme zkoumat vztah mezi vybranými faktory, a to mezi střední délkou života, hrubým domácím produktem na osobu a úhrnnou plodností. Zvyšující se střední délka života je v posledních letech často diskutovaným tématem. Z ekonomických a sociálních důvodů je tento trend částečně považován za problém, na který musí reagovat ekonomové, politici, zdravotníci, ale v podstatě celá společnost. Problém spočívá v přibývání lidí v důchodovém věku, kteří jsou odkázáni na pobírání transferových plateb.

Práce je rozdělena do tří částí. V první části se seznámíme se samotnou demografií a demografickými a makroekonomickými pojmy používanými v bakalářské práci. Představíme si úmrtnostní tabulky, ze kterých dostáváme pro tuto bakalářskou práci podstanou veličinu, střední délku života. Dále si vysvětlíme pojmy úhrnnou plodnost a hrubý domácí produkt na osobu.

Ve druhé části se zaměříme na popis korelační analýzy, vysvětlení rozdílu mezi Pearsonovým a Spearmanovým korelačním koeficientem, a uvedeme si problematiku parciálního korelačního koeficientu. Dále si vysvětlíme v čem tkví podstata shlukové analýzy, a jak se vytváří hierarchická shluková analýza. Uvedené matematické metody budou aplikovány v praktické části bakalářské práce.

V poslední části podrobně rozebereme vývoj střední délky života v čase ve vybraných zemích a zmíníme možné příčiny poklesů nebo naopak vzrůstů střední délky života. Pomocí hierarchické shlukové analýzy provedeme porovnání jednotlivých států. Vývoj střední délky života zaznameneáme do mapy, ve které pře-

hledně uvidíme trend růstu. Dále se zaměříme na vztah mezi vybranými faktory, a to střední délkou života, hrubým domácím produktem na osobu a úhrnnou plodností. Pomocí korelační analýzy zjistíme závislost ukazatelů.

Kapitola 1

Demografické a makroekonomické pojmy

1.1. Demografie

Pojem demografie má původ v řeckém jazyce a doslova znamená popis obyvatelstva (*démos* - obyvatelstvo, *grafein* - popisovat). Demografie je vědní obor, a jak již napovídá význam slova, zabývá se lidskou populací. Dále její velikostí, vývojem, strukturou, a z velké části se věnuje reprodukci, která je chápána jako přirozená obnova lidské populace. Mezi nejvýznamnější události patří narození a úmrtí, od těchto dvou pojmů jsou odvozovány procesy porodnost a úmrtnost. Dále se zabývá sňatečností, rozvodovostí a potratovostí. V neposlední řadě obor demografie navazuje spolupráci s geografii obyvatelstva a do populačního vývoje zahrnují i prostorovou mobilitu - migraci. [18] [31]

1.1.1. Úmrtnostní tabulky a střední délka života

Problematika úmrtnostních tabulek je zpracovaná dle zdrojů [6] [7] [15] [16].

Úmrtnostní tabulky jsou nástrojem, pomocí kterého lze zkonstruovat praktický model úmrtnosti v jednotlivých letech života. Sestavují se pro každé pohlaví zvlášť, protože pravděpodobnost dožití mužů a žen se liší. Úmrtnostní tabulky jsou nedílnou součástí výpočtu pojistného v životním a penzijním pojištění, z důvodu, že smrt je brána jako náhodný jev.

Dle profesora Cipry rozlišujeme několik druhů úmrtnostních tabulek: úplné, ve kterých pracujeme s věkovými intervaly o délce jednoho roku ($0, 1, 2, \dots, \omega$ roků), zkrácené, které mají věkové intervaly delší než jeden rok (např. $0, 1-4, 5-9, 10-14, \dots$ roků), běžné, které vychází z úmrtnostní zkušenosti populace během krátkého časového úseku, a generační, které představují skutečný záznam průběhu života konkrétní populace od narození až po smrt posledního jedince. [6] [7]

V České republice jsou vydávány úmrtnostní tabulky každý rok institucí Český statistický úřad, který používá úplné úmrtnostní tabulky s celočíselnými roky života.

První zmínky o zhotovení modelu úmrtnostních tabulek se datuje do 17. století, kdy známý anglický astronom Edmond Halley zkoumal po dobu pěti let počet narozených a počet zemřelých obyvatel města Vratislav. [2]

Popis a výpočet úmrtnostní tabulky

Úmrtnostní tabulky, které vydává Český statistický úřad, jsou s věkovým intervalem o délce jednoho roku. Tabulka se skládá z deseti sloupců.

V prvním sloupci se nachází věk, který nabývá hodnot $0, 1, \dots, \omega$, kde ω je poslední věková skupina a obsahuje jedince, kteří se dožijí věku ω a více.

Druhý sloupec D_x představuje celkový počet zemřelých jedinců ve věku x a třetí sloupec P_x popisuje střední stav populace České republiky nebo menších uzemních celků, jako jsou kraje a okresy, k 1.7. daného roku.

Dva předchozí sloupce jsou základními daty ke konstrukci pravděpodobnosti úmrtí, která je v pátém sloupci a představuje pravděpodobnost toho, že se jedinec věku x nedožije věku $x + 1$. Ve čtvrtém sloupci se nachází tabulková míra úmrtnosti. Ve chvíli, kdy máme vypočítanou pravděpodobnost úmrtí, můžeme dopočítat zbytek úmrtnostní tabulky, neboť všechny ostatní ukazatele vychází právě z ní. Doplnkem pravděpodobnosti úmrtí je pravděpodobnost dožití se p_x věku x , kterou vypočítáme pomocí vztahu

$$p_x = 1 - q_x. \tag{1.1}$$

Další sloupec, tedy šestý, reprezentuje tabulkový počet dožívajících se osob l_x věku x . V tomto ukazateli pracujeme s fiktivním souborem lidí. Zvolíme si počáteční číselný stav osob l_0 , který nazýváme kořen (radix), nejčasteji 100 tisíc, a poté v posloupnosti $l_0, l_1, \dots, l_\omega$ pozorujeme počty lidí dožívajících se jednotlivých roků ze zvoleného kořene. Jednotlivé prvky posloupnosti počítáme postupně vztahem

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

Posloupnost, kterou dostaneme je nerostoucí, neboť bereme v potaz pouze vymírání populace, nikoli narození nových jedinců a je ve tvaru

$$l_0 \geq l_1 \geq l_2 \geq \dots \geq l_\omega. \quad (1.3)$$

Následující sloupec d_x , počet zemřelých ve věku x , nese informaci o počtu zemřelých osob v dokončeném věku x z fiktivní populace l_0 . Hodnoty počítáme jako rozdíl dvou po sobě jdoucích hodnot dožívajících se osob, tedy

$$d_x = l_x - l_{x+1}. \quad (1.4)$$

Osmý sloupec L_x představuje počet let prožitých jedinci ve věku x . Jedná se o tzv. střední počet „člověkoroků“, který ve věku x celkem prožije l_x jedinců. Při výpočtu se zde využívá aproximace

$$L_x = l_{x+1} + \frac{1}{2} \cdot d_x = \frac{l_x + l_{x+1}}{2} \quad (1.5)$$

kromě $x = 0$, z důvodu, že v roku $x = 0$ není splněn předpoklad o področní linearitě úmrtnosti, neboť po narození je kojenecká úmrtnost nakumulovaná do prvních dnů a týdnů života.

V předposledním sloupci se nachází pomocný ukazatel T_x , který vyjadřuje počet zbylých let života jedinců ve věku x z naší zvolené fiktivní populace. Jedná se o střední počet „člověkoroků“ a vysvětlit to můžeme jako počet let, které zbývají prožít l_x jedincům do konce života. Hodnotu T_x vypočítáme součtem

$$T_x = L_x + L_{x+1} + \dots + L_\omega. \quad (1.6)$$

V posledním sloupci se konečně dostáváme ke zmiňované střední délce života. Počítáme ji jako podíl počtu zbylých let života jedinců ve věku x z naší zvolené fiktivní populace a tabulkovým počtem dožívajících se osob l_x

$$e_x = \frac{T_x}{l_x}. \quad (1.7)$$

Střední délka života, jinak řečeno naděje na dožití, je jedním z ukazatelů úmrtnostních tabulek. Vyjadřujeme pomocí něj průměrný počet let, které by měl jedinec ve svém životě prožít. Ukazatel platí za dané podmínky neměnnosti poměrů úmrtnostních tabulek z doby, kdy jsou tvořeny. Střední délku života uvádíme buď při narození jedince (e_0) a udává průměrný věk, kterého by se měl právě narozený jedinec dožít, nebo od určitého věku x , za podmínky, že se jedinec dožil věku x , a udává průměrnou délku života, kterou má jedinec stále před sebou (e_x). [6] [14]

Střední délka života se v průběhu let mění a její dlouhodobý trend je rostoucí. Dlouhodobé zvyšování hodnot střední délky života má mnoho opodstatnění, např. lepší zdravotní péče a k ní jednodušší přístup, účinnější medikamenty, nové moderní zdravotnické technologie, které mohou včas diagnostikovat závažné onemocnění, dále zvyšující se zájem o zdravý životní styl a s ním spojená kvalitnější strava a pravidelný pohyb, očkování, hygiena a mnoho dalšího. Negativní vliv a následný pokles střední délky života může zapříčinit pandemie nebo války, které evidujeme jak v minulosti, tak v současnosti. Například na pandemii koronaviru lze prakticky ukázat nynější pokles střední délky života. Dle zveřejněných dat DataBank na webu The World Bank, odkud jsou v bakalářské práci data čerpány, v České republice klesla střední délka života v roce 2020 oproti roku 2019 pro ženy o 0,9 roku a pro muže o 1,1 roku.

Fakt, že se lidé dožívají vyššího věku než dříve, s sebou nese i určité negativní důsledky - stárnutí populace. Tento problém je spojen se snižující se porodností a má velký vliv na sociální politiku zemí.

Obrázek 1.1 zobrazuje příklad úmrtnostních tabulek pro ženské pohlaví v České republice v roce 2020. Úmrtnostní tabulka je čerpána ze stránek Českého statis-

2020		CZ		Česká republika					
Ženy Females									
věk (x) age	D_x	P_x	m_x	q_x	l_x	d_x	L_x	T_x	e_x
0	98	54222	0,001807	0,001804	100000	180	99838	8137689	81,38
1	6	55427	0,000139	0,000139	99820	14	99813	8037851	80,52
2	7	56154	0,000088	0,000088	99806	9	99801	7938039	79,53
3	5	55947	0,000106	0,000106	99797	11	99792	7838237	78,54
4	6	55325	0,000098	0,000098	99786	10	99782	7738446	77,55
5	4	54886	0,000077	0,000077	99777	8	99773	7638664	76,56
6	2	54179	0,000062	0,000062	99769	6	99766	7538891	75,56
7	5	53798	0,000058	0,000058	99763	6	99760	7439125	74,57
⋮									
100	148	238	0,535876	0,422636	749	317	591	1299	1,73
101	57	106	0,574300	0,446179	432	193	336	708	1,64
102	30	67	0,611910	0,468554	240	112	183	372	1,56
103	21	51	0,648298	0,489596	127	62	96	189	1,49
104	17	28	0,683107	0,509191	65	33	48	93	1,43
105	14	60	0,716050	1,000000	32	32	45	45	1,40

Obrázek 1.1: Úmrtnostní tabulka

tického úřadu. [17]

1.1.2. Úhrnná plodnost

Jako další demografický pojem si uvedeme úhrnnou plodnost, kterou budeme využívat v praktické části bakalářské práce.

Plodnost je jeden z demografických pojmů související s porodností. Plodnost, neboli fertilita, vyjadřuje počet živě narozených dětí, které by se narodily každé ženě během celého jejího reprodukčního věku (15 - 49 let), za podmínky, že se dožije konce reprodukčního období. Ukazatel úhrnné plodnosti je vztažen k jedné ženě. Plodnost je výsledný efekt plodivosti neboli fekundity, což je schopnost muže a ženy plodit resp. родit děti. [12] [27] [32]

1.2. Makroekonomie

Makroekonomie je jedna z částí obecné ekonomické teorie, která se zabývá problémy týkajícími se větších celků. Zaobírá se státem a ekonomickou integrací - zkoumá ekonomické hodnoty státu jako celek. Zabývá se hospodářskou politikou vlády, inflací, ekonomickým růstem či naopak stagnací, celkovou zaměstnaností, měnou, množstvím peněz. [33] [3]

1.2.1. Hrubý domácí produkt

Hrubý domácí produkt si zmiňujeme k pozdějšímu používání v praktické části bakalářské práce.

Hrubý domácí produkt neboli HDP (*GDP - Gross Domestic Product*) je jeden z makroekonomických ukazatelů, který popisuje vývoj a výkonnost ekonomiky v dané zemi, a také vypovídá o průměrné životní úrovni. Je definován jako tržní hodnota finálních statků, které jsou vyrobeny v zemi během určitého časového období. Do HDP zahrnujeme zboží a služby vyprodukované domácími firmami, ale i firmami zahraničního vlastnictví, které jsou na území daného státu. [9] [8]

Hrubý domácí produkt můžeme vypočítat pomocí několika metod. První metodou je výdajová metoda, která měří národohospodářský produkt nepřímou. Každá domácnost, podnik, vláda daného státu a také zahraniční subjekty uskutečňují nákupy výrobků a služeb vyprodukované v dané zemi za určité období, v důsledku toho jim vznikají náklady. Při výpočtu HDP pomocí výdajové metody, tyto výdaje sečteme a dostaneme hodnotu, která vypovídá o celkovém produktu ekonomiky. Další metodou je důchodová (příjmová) metoda. Ekonomické subjekty poskytující služby dostávají důchod (příjem) a při sečtení těchto důchodů, získáme přibližnou hodnotu produktu. Poslední metodou je metoda založená na sumarizaci hodnot přidaných zpracováním. Hodnota přidaná zpracováním neboli přidaná hodnota je hodnota, která se přidává k hodnotě nakupovaných surovin nebo polotovarů, popřípadě služeb. Každý výrobce si ji stanovuje sám a je to výše úsilí výrobců při výrobním procesu. Při použití této metody se sečtou přidané hodnoty a výsledkem je hodnota hrubého domácího produktu (musí se rovnat HDP vypočtenému součtu cen finálních statků). [11]

Jednotlivé státy světa jsou různě velké s různým počtem obyvatel a celkové HDP by nebylo vypovídající pro mezinárodní srovnání, proto se používá pro srovnávání HDP na obyvatele. Tento ukazatel je poměrový a vypovídá o průměrné životní úrovni obyvatelstva dané země. [3]

Kapitola 2

Aplikovaný matematický aparát

V následující kapitole si definujeme a popíšeme matematické metody, pomocí kterých budeme provádět analýzu v praktické části bakalářské práce.

2.1. Korelační analýza

Předpokládejme dvě číselné veličiny X a Y , tyto veličiny mohou být spolu ve vzájemném vztahu, tzv. korelaci. Pokud mezi veličinami existuje vztah, jsou korelované, a mohou být závislé. Sílu závislosti můžeme kvantitativně ohodnotit. V případě lineárního vztahu mezi veličinami X a Y používáme korelační koeficient. Teoretický koeficient korelace lze počítat při znalosti rozdělení pravděpodobností náhodných veličin X a Y . Tento koeficient je, za podmínky nenulových rozptylů $var(X)$ a $var(Y)$, definovaný vztahem

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var(X) \cdot var(Y)}}, \quad (2.1)$$

kde $cov(X,Y)$ je kovariance dvou náhodných veličin X a Y . [10]

Korelační koeficient nabývá hodnot z intervalu $\langle -1, 1 \rangle$, a pokud se rovná přímo nule, znamená to, že mezi sledovanými veličinami není lineární vztah a jsou nekorelované. Naopak čím více se hodnota blíží ke krajům intervalu, tím silnější lineární závislost mezi sledovanými veličinami existuje. V krajních bodech intervalu nastává absolutní korelace. V praxi se může stát, že veličiny spolu korelují, ale není mezi nimi souvislost. [10] [4] [1]

Jelikož v praxi se stává pouze zřídka, že známe konkrétní rozdělení pravděpodobností náhodných veličin, používáme k praktickému výpočtu výběrový korelační koeficient založený na náhodném výběru z rozdělení veličin. Jestliže náhodné veličiny X a Y splňují předpoklady normálního rozdělení, a také mezi nimi můžeme předpokládat linearitu, pro měření závislosti náhodných veličin budeme používat Pearsonův korelační koeficient. Necht' máme náhodný výběr $(X_1, Y_1)', \dots, (X_n, Y_n)'$ z dvourozměrného rozdělení. Označíme

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i,$$

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2,$$

kde \bar{X} a \bar{Y} jsou výběrové průměry náhodných veličin X , Y a S_X^2 , S_Y^2 jsou výběrové rozptyly náhodných veličin X , Y . Dále

$$S_{X,Y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (2.2)$$

nazýváme výběrovou kovariancí náhodných veličin X a Y . Pokud $S_X^2 > 0$ a $S_Y^2 > 0$, potom je Pearsonův korelační koeficient definován vztahem

$$r_{X,Y} = \frac{S_{X,Y}}{\sqrt{S_X^2 S_Y^2}} = \frac{S_{X,Y}}{S_X S_Y}. \quad (2.3)$$

[10] [5]

2.1.1. Spearmanův korelační koeficient

V případě nesplněné podmínky normálního rozdělení náhodných veličin, používáme Spearmanův korelační koeficient, který taktéž nabývá hodnot z intervalu $\langle -1, 1 \rangle$. U tohoto typu výběrového korelačního koeficientu nám stačí pouze předpoklad monotonie, tudíž závislost nemusí být lineární. Spearmanův korelační koeficient počítá s využitím pořadové korelace, a proto jej můžeme nazvat Spearmanův korelační koeficient pořadové korelace. Jednotlivé hodnoty obou proměnných jsou nahrazeny pořadovými čísly, podle toho, jaké místo v číselné řadě určítá

hodnota zaujímá. Pořadí budeme značit x_i a y_i pro veličiny X_i a Y_i . Koeficient definujeme vztahem

$$r_{X,Y} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x} \cdot \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)}}. \quad (2.4)$$

Pomocí následujících vztahů

$$\begin{aligned} \bar{x} = \bar{y} &= \frac{1}{n} \sum_{i=1}^n i = \frac{n+1}{2}, \\ \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2 = \sum_{i=1}^n i^2 &= \frac{n(n+1)(2n+1)}{6}, \\ \sum_{i=1}^n x_i y_i &= \frac{1}{2} \left(\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 \right) - \frac{1}{2} \sum_{i=1}^n (x_i - y_i)^2, \end{aligned}$$

lze odvodit častěji užívaný vzorec

$$r_{X,Y} = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n (x_i - y_i)^2. \quad (2.5)$$

Může nastat situace shodných pořadí, potom platí $r_{X,Y} = 1$. Vzhledem k tomu, že Spearmanův korelační koeficient pracuje s pořadími veličin a ne vzdálenostmi, neovlivňují jej odlehlé hodnoty.

Při testování hypotéz si zvolíme kritický obor W , při kterém budeme nulovou hypotézu H_0 zamítat. W volíme pomocí čísla α ($0 < \alpha < 1$), které nazýváme hladina významnosti testu. Platnou hypotézu budeme zamítat s pravděpodobností nejvýše α . Při testování hypotéz pomocí statistického softwaru bude výstupem nejmenší hladina významnosti testu neboli p -hodnota (p -value). Pokud je p -value menší nebo rovna α , pak naši nulovou hypotézu H_0 zamítneme na hladině testu α . Pokud je p -value větší než α , potom nulovou hypotézu H_0 na zvolené hladině testu nemůžeme zamítnout.

U náhodných veličin s jiným než normálním rozdělením, u kterých známe pouze hodnoty nahrazené pořadím, budeme tedy počítat Spearmanův korelační koeficient. Pro testování používáme test nezávislosti. Testujeme nulovou hypotézu

H_0 : veličiny X, Y jsou nezávislé

proti alternativě

H_A : veličiny X, Y nejsou nezávislé.

Jestliže vyjde číslo $|r_{X,Y}|$ větší nebo rovno kritické hodnotě, budeme zamítat nulovou hypotézu o nezávislosti veličin X a Y . Kritickou hodnotu na hladině významnosti α můžeme vypočítat pomocí statistického softwaru (například software R, který budeme používat v praktické části). [10] [4] [5]

2.1.2. Parciální korelační koeficient

Teorie týkající se parciálního korelačního koeficientu je čerpána ze zdrojů [10] [1].

Korelační koeficient veličin Y a Z sice může být roven vysoké hodnotě, ale obě hodnoty mohou být ovlivňovány dalšími veličinami X_1, \dots, X_p . Tudíž závislost mezi Y a Z může být klamná. Proto zavádíme parciální korelační koeficient, který počítá závislost při vyloučení vlivu \mathbf{X} . Zajímá nás tedy, jak vysoká bude korelace mezi Y a Z , když vyloučíme vliv \mathbf{X} , tzn. \mathbf{X} bude konstantní.

Parciální korelační koeficient pro závislost pouze na jednom externím faktoru definujeme vztahem

$$\rho_{Y,Z.X} = \frac{\rho_{Y,Z} - \rho_{Y,X}\rho_{X,Z}}{\sqrt{(1 - \rho_{Y,X}^2) \cdot (1 - \rho_{Z,X}^2)}}. \quad (2.6)$$

Výběrový parciální korelační koeficient v případě závislosti na jednom externím faktoru definujeme jako

$$r_{Y,Z.X} = \frac{r_{Y,Z} - r_{Y,X}r_{X,Z}}{\sqrt{(1 - r_{Y,X}^2) \cdot (1 - r_{Z,X}^2)}}. \quad (2.7)$$

Pro testování hypotézy použijeme test hypotézy $H_0 : \rho_{Y,Z.X} = 0$ proti alternativě $H_A : \rho_{Y,Z.X} \neq 0$.

2.2. Shluková analýza

Teorie ze sekce shlukové analýzy bylo čerpáno ze zdroje [13].

Princip shlukové analýzy spočívá v seskupení resp. zařazení objektů do jednotlivých shluků. Dva objekty jednoho shluku si mají být co nejvíce podobné a dva objekty různých shluků co nejméně podobné. Vstupním údajem pro shlukování je datová matice o rozměru $n \times m$. V matici je obsaženo m -rozměrné pozorování n objektů. Každý řádek představuje vektor údajů o jednom objektu a celkový počet vektorů je tedy n . Každé pozorování je znázorněno ve sloupcovém vektoru a jsou to hodnoty proměnných neboli statistických znaků.

Při shlukování jsou používány například míry podobnosti. Více však bývá využíváno měření v opačném směru - míra nepodobnosti. Podobnost objektů x_i a x_j značíme $S(x_i, x_j)$, případně S_{ij} . Platí vztah $S_{ij} = S_{ji}$. Pokud budeme uvažovat míru podobnosti $S_{ij} \in \langle 0, 1 \rangle$, platí, že při hodnotě 0 budou objekty maximálně rozdílné a při hodnotě 1 totožné. Pro míru nepodobnosti D_{ij} pak bude platit rovnost $D_{ij} = 1 - S_{ij}$.

Při vztahu dvou objektů danými kvantitativními daty, se pro vyjádření vztahu používá především míra vzdálenosti. Souřadnice objektů představují jednotlivé proměnné. V případě splnění trojúhelníkové nerovnosti

$$D_{iy} + D_{yj} \geq D_{ij} \quad (2.8)$$

($i, j, y \in \langle 1, n \rangle$), pak mluvíme o metrice. Mezi nejznámější metriky patří metrika euklidovská

$$D_E(x_i, x_j) = \sqrt{\sum_{l=1}^m (x_{il} - x_{jl})^2}, \quad (2.9)$$

a dále také metrika manhattanská

$$D_B(x_i, x_j) = \sum_{l=1}^m |x_{il} - x_{jl}|. \quad (2.10)$$

V případě nominálních dat rozlišujeme pouze, zda jsou hodnoty shodné či ni-

koli. U těchto dat zjišťujeme podobnost pomocí koeficientu prosté shody

$$S_{ij} = \frac{\sum_{l=1}^m S_{ijl}}{m}, \quad (2.11)$$

kde v čitateli máme počet proměnných, u kterých se hodnoty dvou sledovaných objektů shodují a ve jmenovateli máme celkový počet proměnných.

Při shlukování můžeme chtít, aby se objekty zařadily do určitého počtu shluků (ploché shlukování), nebo aby byla vytvořena hierarchie shluků (hierarchické shlukování). Dále rozlišujeme shlukování disjunktí a překrývající se. U disjunktího shlukování dochází k přiřazení objektu pouze do jednoho shluku, naopak u překrývajícího se shlukování objekt zařazujeme do více shluků.

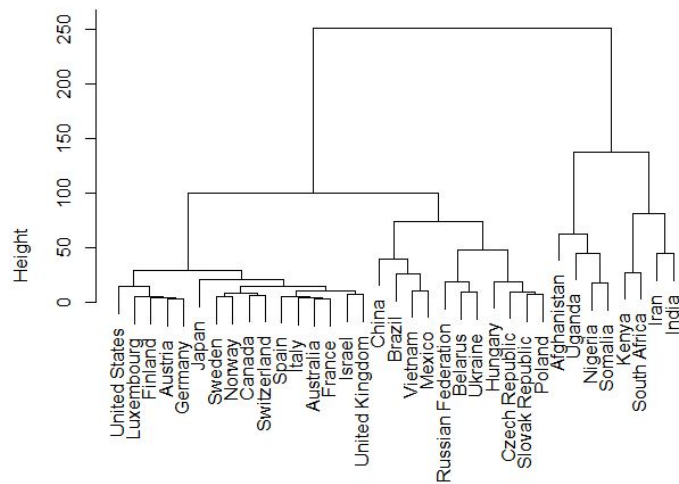
Jednou z metod přiřazení objektů do předem stanoveného počtu shluků je metoda rozkladu. Pomocí metody k -průměrů bychom získali jednoznačné přiřazení pro metodu rozkladu. Je to metoda založená na principu iterací. Nejprve určíme k počátečních centroidů, neboli bodů, jenž tvoří „střed“ shluků, například zvolíme prvních k objektů z daného souboru. Následně se pomocí euklidovské metriky spočte vzdálenost každého objektu ke každému počátečnímu centroidu. Postupně jsou objekty přiřazeny ke svému nejbližšímu centroidu. Při každé iteraci nám vznikne nový shluk, pro který vypočteme nový centroid. Znovu spočítáme vzdálenosti k nově vytvořeným centroidům a pokračujeme tak dlouho, dokud nepřestane docházet k přesunům.

2.2.1. Hierarchická shluková analýza

Hierarchické shlukování budeme využívat v praktické části bakalářské práce. Jeden z přístupů je aglomerativní přístup (analýza podobnosti). Na počátku tvoří shluky o jednom objektu a postupně spojuje dva shluky, které si jsou nejvíce podobné, dokud nevznikne pouze jeden shluk. Dalším z přístupů je divizní přístup (analýza nepodobnosti), který je opak aglomerativního. Tudíž na počátku je pouze jeden shluk tvořený všemi objekty a následně se dělí, dokud nebude každý objekt samostatným shlukem.

Výsledkem hierarchické shlukové analýzy je grafický výstup - speciální graf

dendrogram, na kterém lze výstižně vidět postup shlukování, viz obrázek 2.1. Jedná se o stromový diagram znázorňující postupné shlukování objektů (tzv. listů), které splývají do shluků (tzv. větví). Může mít horizontální nebo vertikální podobu. Při horizontální podobě začínají listy zleva a směrem doprava se shlukují, při vertikální podobě se listy nacházejí ve spodní části diagramu a směrem nahoru se shlukují. Ke shlukování dochází dle nejkratších vzdáleností mezi objekty, později shluky, například metodou nejbližšího souseda. Metoda je založena na algoritmu, který počítá vzdálenost mezi shluky a vybere nejmenší ze vzdáleností každých dvou objektů z dvou různých shluků.



Obrázek 2.1: Dendrogram - příklad

Kapitola 3

Analýza střední délky života

Data ke zpracování praktické části bakalářské práce jsou čerpány ze zdroje [34], dostupné na webu The World Bank - Data Bank. Pro výpočty budeme používat data od roku 1960 do roku 2020. Veškerá data jsou zaznamenána v excelovském souboru s názvem *Data_bakalarska_prace* a jsou přiloženy k bakalářské práci. Grafy pro vývoj střední délky života jsou vytvořeny, stejně jako mapové grafy, v Excelu. Dendrogramy a bodové grafy znázorňující závislost mezi střední délkou života a jednotlivými faktory jsou vytvořeny v softwaru R.

V této kapitole se budeme zabývat vývojem střední délky života a rozdíly ve vývoji ve zvolených zemích. Bude zkoumána závislost mezi střední délkou života a jednotlivými faktory. Spočítáme si příslušné korelační koeficienty a otestujeme jejich významnost. Kvůli možnosti falešné závislosti si spočítáme i parciální korelační koeficienty. Mezi vybrané faktory patří hrubý domácí produkt a úhrnná plodnost. Do analýzy zahrneme určité země z celého světa, jak vyspělé země, tak rozvojové. V Africe se zaměříme na Nigérii, Somálsko, Ugandu, Keňu a Jihoafrickou republiku. Asii bude zastupovat Afghánistán, Irán, Čína, Vietnam, Japonsko a Izrael, na pomezí Asie a Evropy Rusko. Z Evropy bude provedena analýza pro nejvíce zemí, a to ze severní Evropy Švédsko, Norsko a Finsko, z jižní Evropy Španělsko a Itálie, z východní Evropy Bělorusko a Ukrajina, ze západní Evropy Velká Británie, Lucembursko a Francie, a ze střední Evropy Česká republika, Slovenská republika, Polsko, Maďarsko, Švýcarsko, Rakousko a Německo. Americký kontinent bude reprezentovat Brazílie, Mexiko, Kanada a Spojené státy

americké. A také do analýzy zahrneme samostatnou Austrálii.

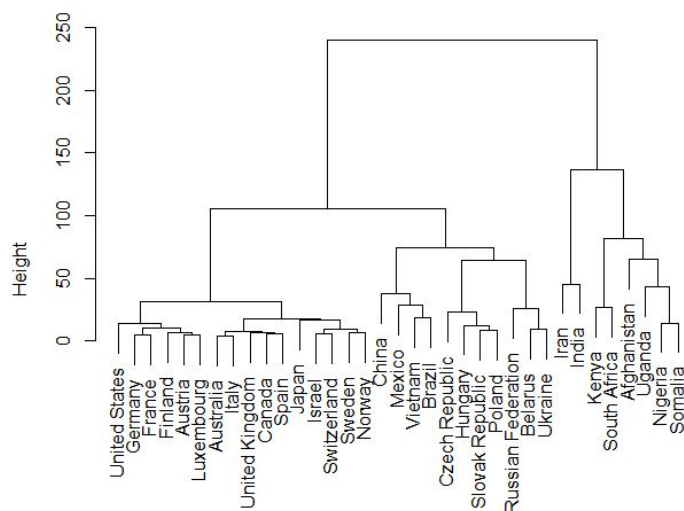
3.1. Vývoj střední délky života v průběhu času

Obecný dlouhodobý lineární trend vývoje střední délky života je rostoucí.

Při hierarchické shlukové analýze budeme, vzhledem k odlišnostem střední délky života pro jednotlivé pohlaví, uvádět data pro muže a ženy zvlášť. U obou pohlaví u většiny zemí lze vidět následek covidové pandemie, a to pokles střední délky života v roce 2020.

3.1.1. Hierarchická shluková analýza vývoje střední délky života pro muže

Z uvedených dat jsme si pomocí hierarchické shlukové analýzy v softwaru R sestrojili následující dendrogram pro muže.

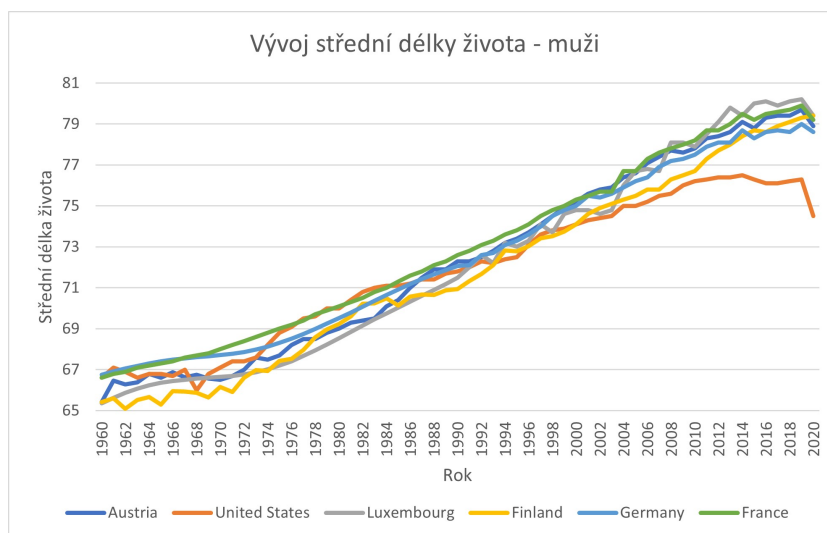


Obrázek 3.1: Vývoj střední délky života pro muže uspořádaná do shluků

V dendrogramu 3.1 lze vidět uspořádání do několika viditelných shluků uspořádaných dle podobnosti vývoje střední délky života pro mužské pohlaví. Na první pohled si lze všimnout obsazeností shluku prvního zleva, který tvoří Spojené státy americké, Kanada, Izrael a Japonsko, což jsou státy velice vyspělé po různých

směrech, ku příkladu rozvinutá věda a technika, energetika a vysoké HDP na osobu. Dále shluk obsahuje převážně země severní Evropy spolu se západní Evropou. K nim svým vývojem střední délky života se přidává Německo, Rakousko a Švýcarsko. Na východě Německa a Rakouska vznikla pomyslná čára a východně od této čáry se země řadí do dalšího shluku. Z většinové míry tento shluk tvoří postkomunistické země Evropy. Poslední shluk je tvořen zeměmi z afrického kontinentu, dále Irán, Afghánistán a Indie. Jsou to státy jinak nazývané země třetího světa, čili rozvojové státy.

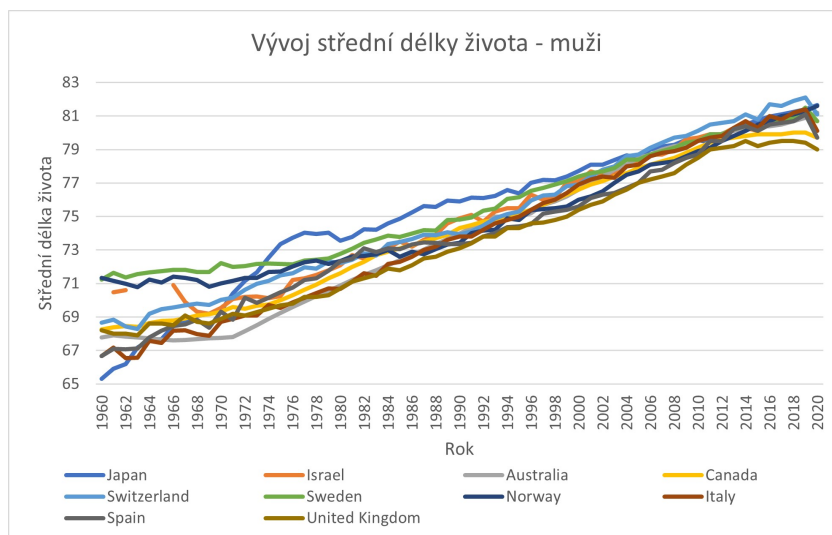
V následujících grafech jsme vykreslili vývoj střední délky života v průběhu času od roku 1960 do roku 2020. První shluk je pro lepší přehlednost rozdělen do dvou samostatných grafů.



Obrázek 3.2: Vývoj střední délky života pro muže - první shluk, část A

V grafu 3.2 můžeme vidět postupný růst střední délky života prvních šesti zemí. Například zde můžeme pozorovat pomalejší růst střední délky života Spojených států amerických oproti jiným zemím. Jedním z důvodů může být jejich zdravotní systém, který se neliší žádným způsobem od zdravotního systému v Evropě. USA má sice jeden z nejkvalitnějších zdravotních systémů na světě, ale vzhledem k tomu, že je pojištění pouze dobrovolné, ne každá nepojištěná osoba si pak může uhradit lékařskou péči. [28] Dále zde můžeme pozorovat jeden z největších poklesů

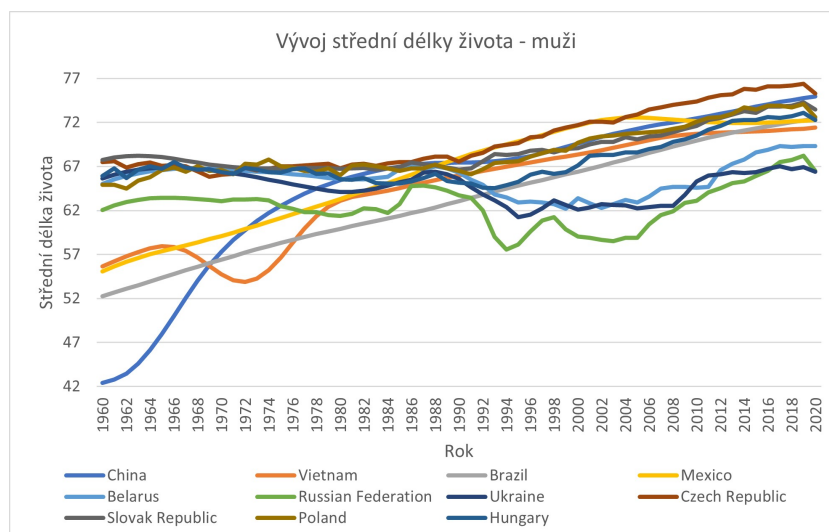
střední délky života v historii USA v roce 2020 z důvodu covidové pandemie, která USA značně zasáhla. Střední délka života u mužů se v USA dostala na úroveň roku 2003. Ostatní země v tomto shluku mají velice podobný, v určitých letech až téměř shodný, vývoj střední délky života.



Obrázek 3.3: Vývoj střední délky života pro muže - první shluk, část B

Z našich vybraných zemí nejvyšší hodnotu střední délky života pro muže mělo v roce 1960 Norsko a dosahovalo hodnoty 71,34 let. O rok později se na první příčku dostalo Švédsko. Avšak od roku 1974 nejvyšší hodnotu neslo Japonsko, vyobrazené v grafu 3.3. V roce 2005 střední délka života ve Švýcarsku převýšila střední délku života v Japonsku. O 15 let později se opět nejvyšší střední délky života dožívají muži v Japonsku, protože ve Švýcarsku klesla střední délka života z důvodu covidové pandemie. Střední délka života pro muže v Japonsku dosahovala v roce 2020 81,64 let. Vysoká hodnota střední délky života v Japonsku má více důvodů, a to pestré stravování s vysokým obsahem živočišného proteinu obsaženo v rybách, rýže, sója a produkty z nich vytvořené, vyspělé všeobecné zdravotní pojištění a také pozitivní přístup ke stáří. [24] V roce 1988 nastal výrazný pokles střední délky života v Izraeli, způsoben s největší pravděpodobností válkou a migrací z Izraele. [20]

Ve vývoji střední délky života Číny vidíme na grafu 3.4 asi největší skok,

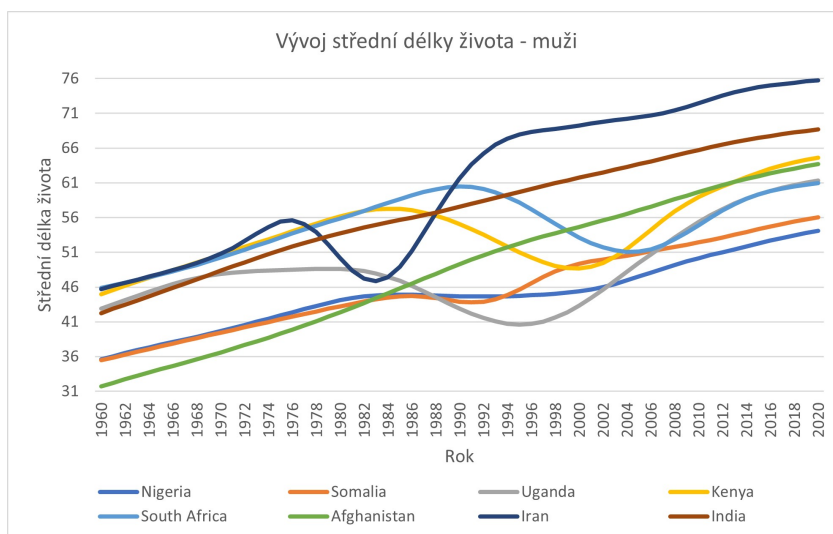


Obrázek 3.4: Vývoj střední délky života pro muže - druhý shluk

jistým důvodem je rozsáhlá reforma, kterou prošlo čínské zdravotnictví. Jeden z důvodů také může být brzký odchod do důchodu čínských obyvatel. [29] Česká a Slovenská republika mají téměř shodné vývoje střední délky života až do roku 1993, kdy došlo k rozpadu Československa. Od té doby má o něco vyšší střední délku života Česká republika. Ve Vietnamu můžeme pozorovat dlouhodobější pokles z důvodu války trvající v letech 1964 až 1975. [21] Netypická věc nastává u vývoje střední délky života na Ukrajině a v Bělorusku. Lineární trend od roku 1960 do roku 2020 bychom popsali jako pomalu klesající. Příčinou je dlouhodobý pokles střední délky života od devadesátých letech 20. století až do prvního desetiletí 21. století, který může způsoben rozpadem SSSR a poměry s touto situací spojené.

Například v České republice byla v roce 1960 střední délka života 67,5 let, v roce 2019 byla 76,4 let, a v roce 2020 vlivem covidové pandemie klesla na 75,3 let. Můžeme tedy říct, že dlouhodobý trend je sice rostoucí, ale v určitých letech může dojít i ke snížení střední délky života.

V zemích třetího světa sice není zatím znatelný pokles kvůli covidové pandemii, ale střední délka života je tam znatelně nižší oproti zbytku vyspělejšího světa. Od roku 1993 v Jihoafrické republice došlo k obrovskému poklesu střední délky

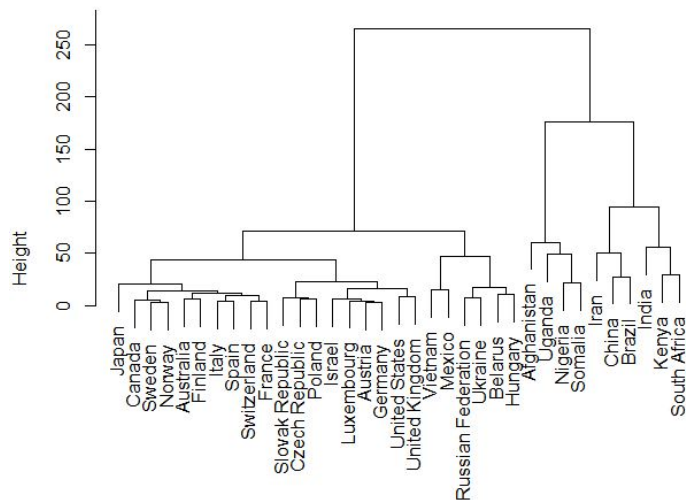


Obrázek 3.5: Vývoj střední délky života pro muže - třetí shluk

života, kvůli epidemii nákazy virem HIV. Epidemie přetrvává dodnes, v roce 2019 v zemi žilo 7,5 milionů lidí nakažených virem HIV. [25] [30] V sedmdesátých a osmdesátých letech 20. století také pozorujeme znatelný pokles střední délky života u mužů v Iránu, kde v té době probíhala Islámská revoluce. Demonstrace byly násilně potlačovány a nechávaly za sebou velké množství mrtvých. Tehdejší vláda byla nahrazena novou a v praxi začalo docházet k odstraňování odpůrců. Přišli také o armádní elity, což následně oslabilo armádu ve válce s Irákem v letech 1980-1988. Během těchto nepokojů také došlo k masivní emigraci. [22] V Ugandě v sedmdesátých letech nastoupil do čela generál Idi Amin Dada, který svrhl tehdejší vládu, zvolil se prezidentem a v Ugandě nastal brutální režim, ve kterém bylo zavražděno odhadem 300 000 Ugandů. To mělo za následek také pokles střední délky života. [23]

3.1.2. Hierarchická shluková analýza vývoje střední délky života pro ženy

U vývoje střední délky života pro ženy v dendrogramu 3.6 lze vidět menší rozdíly oproti mužům. Ženy České a Slovenské republiky se vývojem střední délky života řadí do shluku, který obsahuje vyspělejší země, jako jsou Izrael,



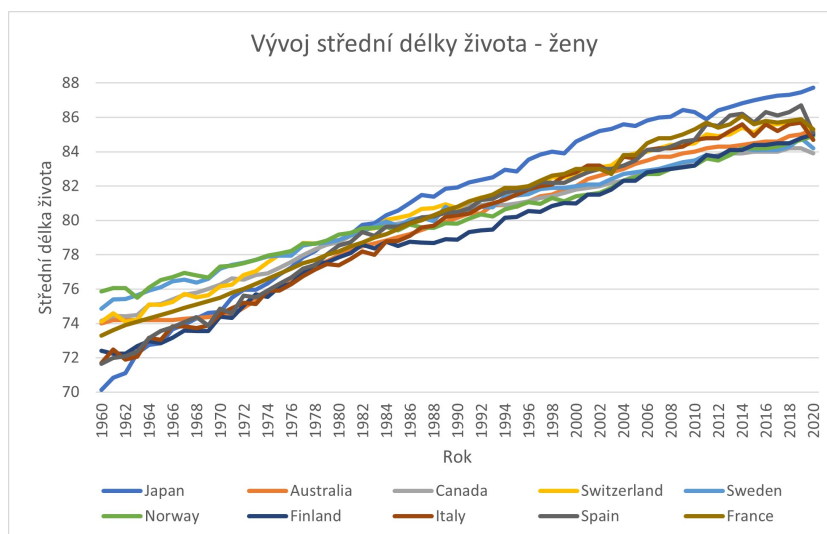
Obrázek 3.6: Vývoj střední délky života pro ženy uspořádaná do shluků

Lucembursko, Rakousko, Německo, Spojené státy americké a Velká Británie. Naopak ženy z Číny a Brazílie si oproti mužům ze stejných zemí pohoršily. Obecně platí, že střední délka života pro ženy bývá delší než pro muže, pouze v Indii a Iránu na začátku šedesátých let byla vyšší střední délka života pro muže.

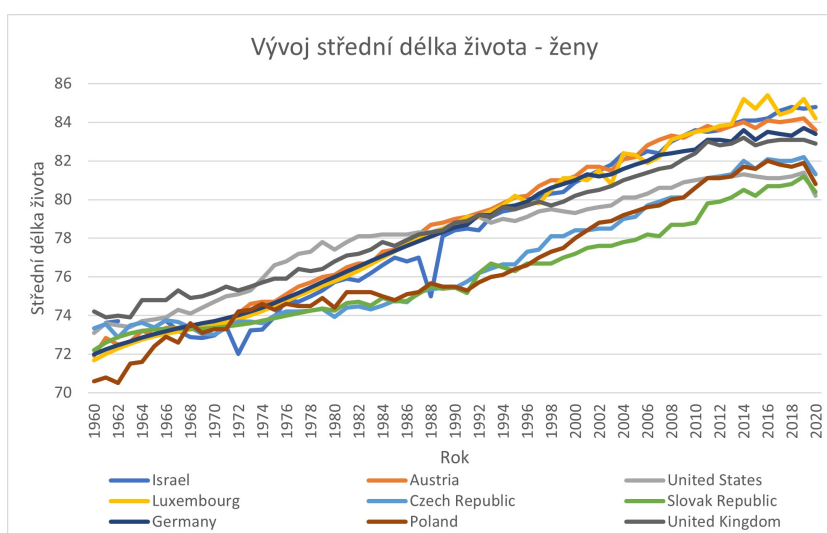
Nejvyšší hodnotu střední délky života pro ženy mělo na počátku šedesátých let Norsko, stejně jako u mužů. U žen byla tato hodnota vyšší než u mužského pohlaví a dosahovala už v šedesátých letech 75,87 let. Například Česká republika se na tuto hodnotu dostala až mezi roky 1991 a 1992. Dlouholeté prvenství nejvyšší hodnoty střední délky života u žen si od roku 1982 drží Japonsko. V roce 2020 dosáhla 87,74 let.

V grafu 3.7 můžeme vidět velice podobný průběh vývoje střední délky života u žen v Itálii, Španělsku a Francii. Ku příkladu na konci španělského vývoje vidíme razantní pokles střední délky života kvůli covidové pandemii. Hodnota se v roce 2020 vrátila na úroveň mezi roky 2010 a 2011.

Při pozorování vývoje střední délky života v České a Slovenské republice v grafu 3.8 lze vidět oddělení vývoje a menšího růstu střední délky života na Slovensku, stejně jako u mužů. V Izraeli, taktéž stejně jako u mužů, pozorujeme v roce 1988 pokles způsobený nejspíš válkou a migrací z Izraele.



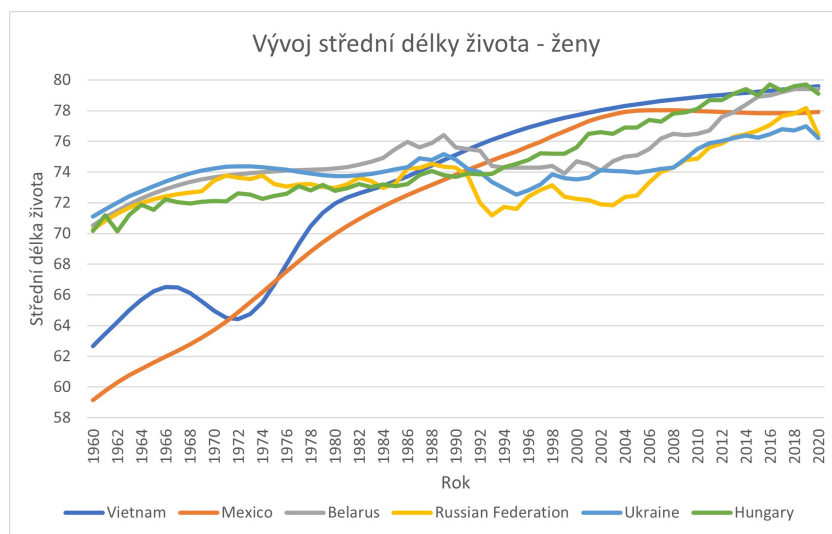
Obrázek 3.7: Vývoj střední délky života pro ženy - první shluk, část A



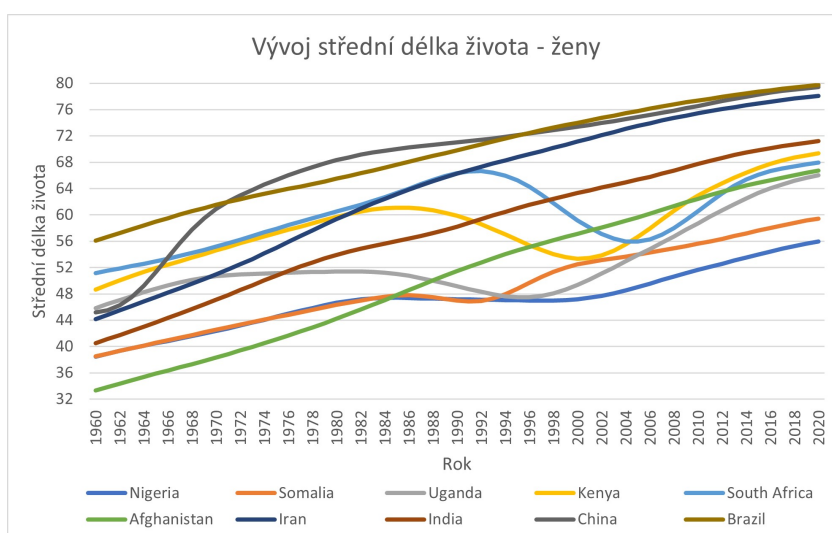
Obrázek 3.8: Vývoj střední délky života pro ženy - první shluk, část B

V šedesátých a sedmdesátých letech ve Vietnamu v grafu 3.9 opět vidíme pokles z důvodu války ve Vietnamu. I přestože důvodem snížení byla válka, ve které bojují převážně muži, ve Vietnamu se pokles týkal i žen, protože v této válce umíraly milióny civilistů. V Rusku, Bělorusku i Ukrajině, ve státech bývalého SSSR, vidíme v roce 1991 při rozpadu SSSR obdobné snížení střední délky života jako u mužského pohlaví.

V Jihoafrické republice na grafu 3.10 můžeme, stejně jako u mužů, vidět pokles



Obrázek 3.9: Vývoj střední délky života pro ženy - druhý shluk



Obrázek 3.10: Vývoj střední délky života pro ženy - třetí shluk

z důvodu epidemie nákazy virem HIV. Narozdíl od mužů v Iránu, u žen nepozorujeme žádný pokles střední délky života, tudíž se můžeme domnívat, že vraždění byli převážně odpůrci systému mužského pohlaví. V Číně střední délka života u žen taktéž skokově vzrostla. Ostatní země třetího světa mají obdobný vývoj střední délky života v čase jako u mužů.

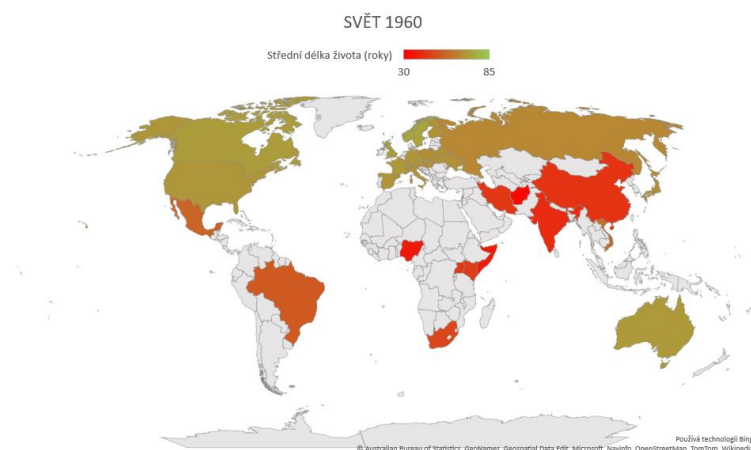
3.2. Vývoj střední délky života ve vybraných letech - vykreslení do mapy

V této kapitole jsme pro lepší představivost vykreslili vývoj střední délky života po deseti letech do mapy pro obě pohlaví dohromady. Na obrázcích jde zřetelně vidět postupné zvyšování střední délky života v průběhu času.

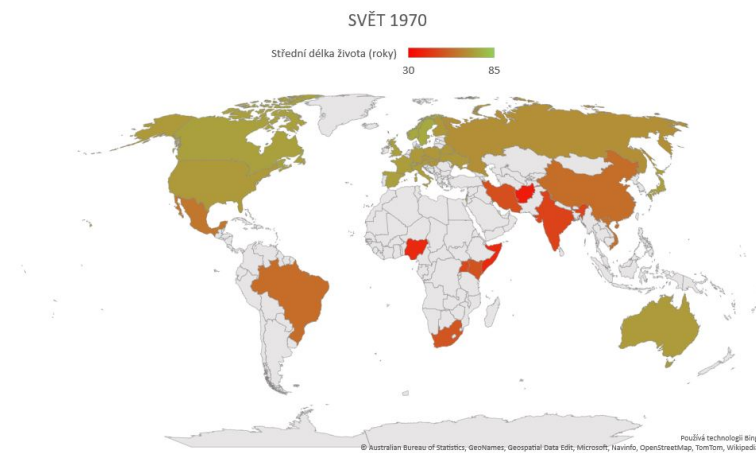
Země vykreslené šedou barvou nejsou zahrnuty do provedení analýzy střední délky života. Červená barva znázorňuje nejnižší střední délku života a zelená barva nejvyšší.

3.2.1. Svět

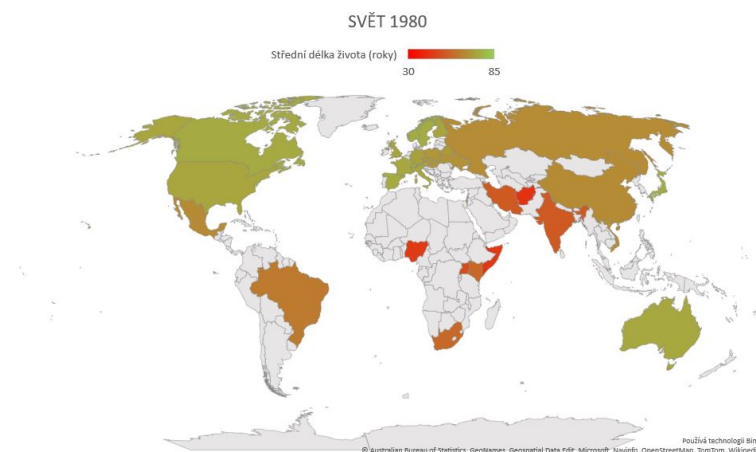
Na začátku sledovaného období můžeme pozorovat dle barev na vykreslených mapách nižší střední délku života ve Střední a Jižní Americe, Africe a jihovýchodní Asii. Avšak i země na těchto kontinentech se s každým dalším desetiletím čím dál tím víc „zabarvují do zelena“, tzn. zvyšuje se střední délka života. Z grafu 3.17 můžeme vyčíst, že až na africké státy Nigérii, Jihoafrickou republiku, Somálska, Keni a Ugandy a asijského státu Afghánistánu, všechny země překročily hranici zhruba 70 let střední délky života, nebo se ji velice blíží.



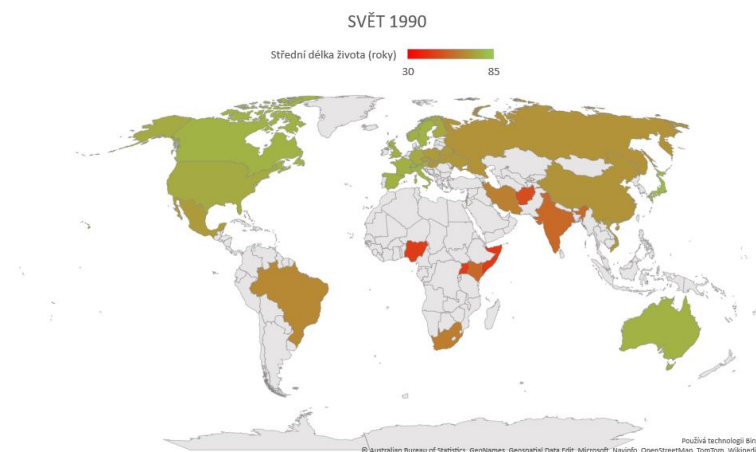
Obrázek 3.11: Střední délka života ve světě v roce 1960



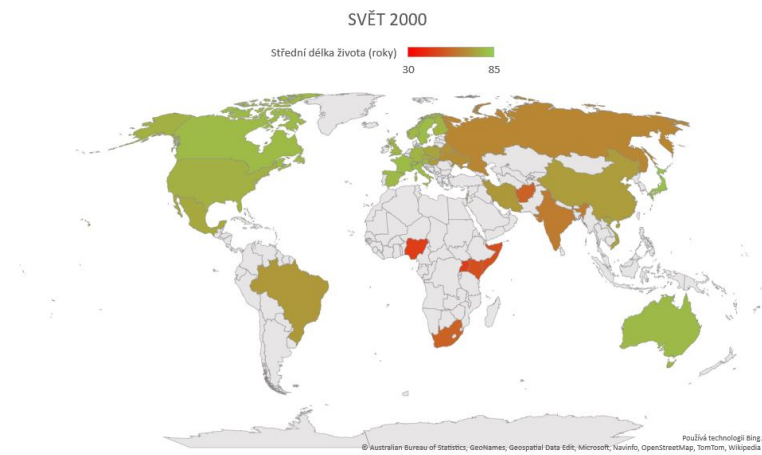
Obrázek 3.12: Střední délka života ve světě v roce 1970



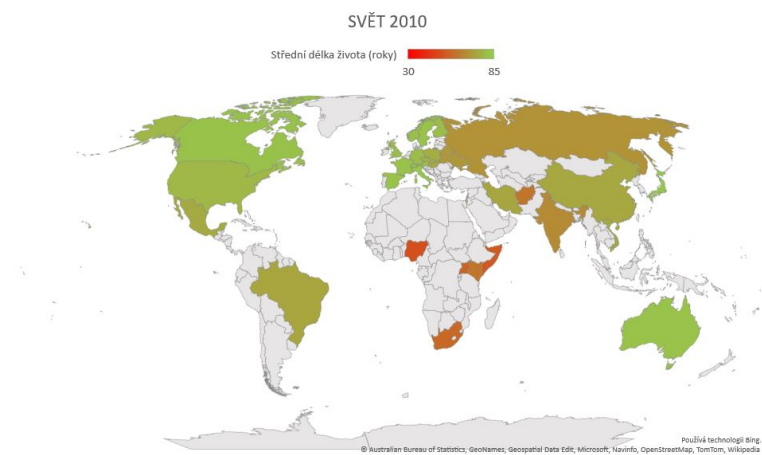
Obrázek 3.13: Střední délka života ve světě v roce 1980



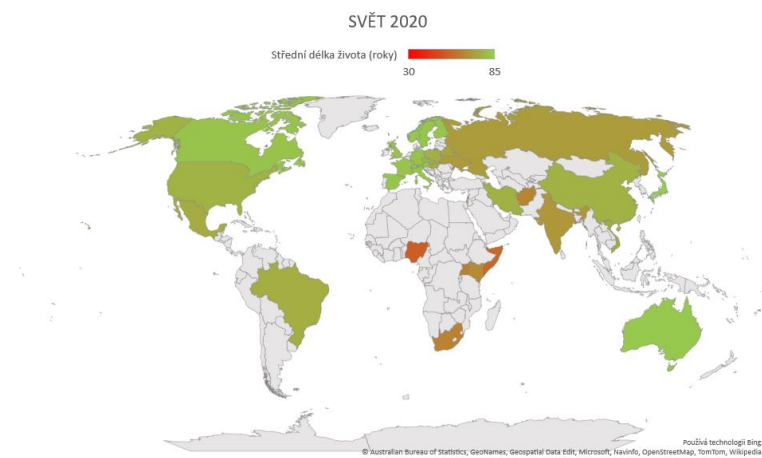
Obrázek 3.14: Střední délka života ve světě v roce 1990



Obrázek 3.15: Střední délka života ve světě v roce 2000



Obrázek 3.16: Střední délka života ve světě v roce 2010

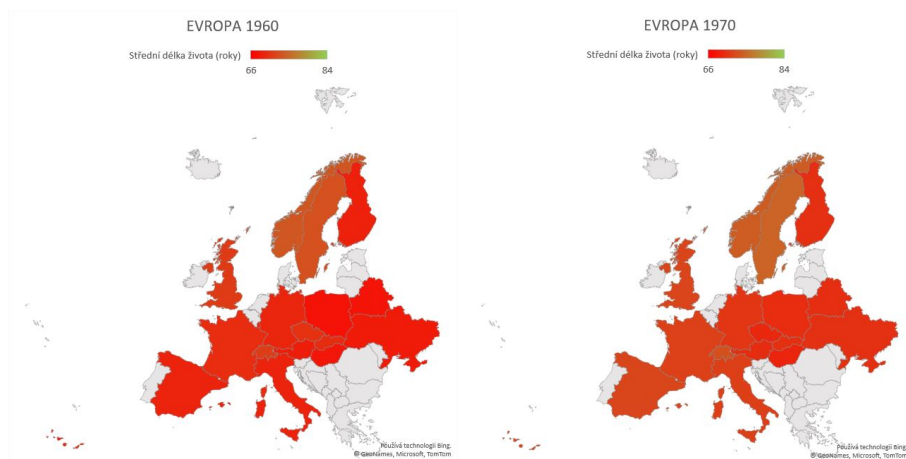


Obrázek 3.17: Střední délka života ve světě v roce 2020

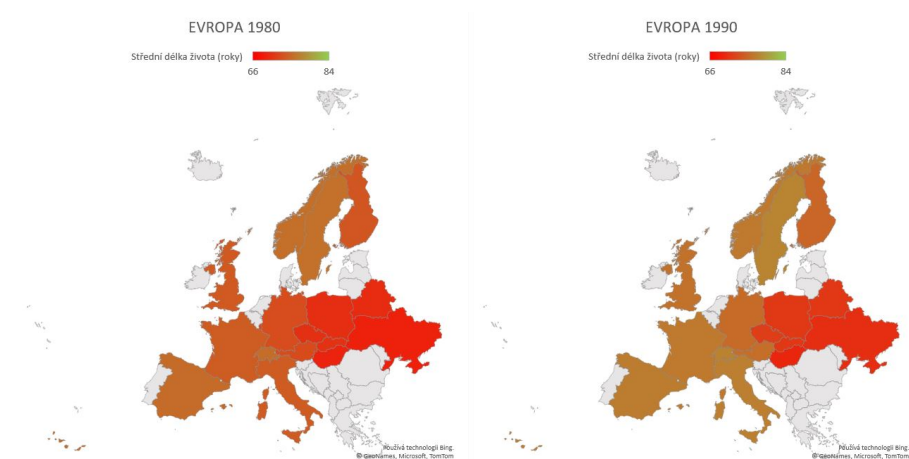
3.2.2. Evropa

Na grafech 3.11 až 3.17 není zřetelně vidět vývoj pro samotnou Evropu, proto jsme si pro lepší vizualizaci vykreslili Evropu zvlášť.

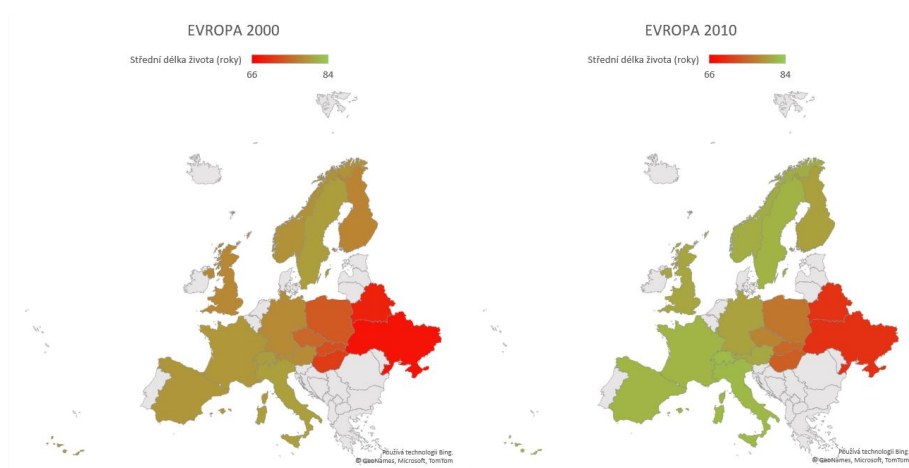
Celá Evropa prošla od roku 1960 do roku 2020 zjevným navýšením střední délky života. Na začátku sledovaného období všechny země spadaly do evropských nižších hodnot střední délky života a žádná z nich nijak nevybočovala, avšak o dalších 10 let později můžeme pozorovat první změny, kdy trend navýšení střední délky života přichází převážně ze severu a západu Evropy. Začínají se odlišovat postkomunistické země, ve který je trend opožděný. Nicméně v posledním sledovaném období, v roce 2020, vidíme, že od západu postkomunistických zemí se začíná střední délka života přibližovat zemím s nejvyšší hodnotou střední délky života.



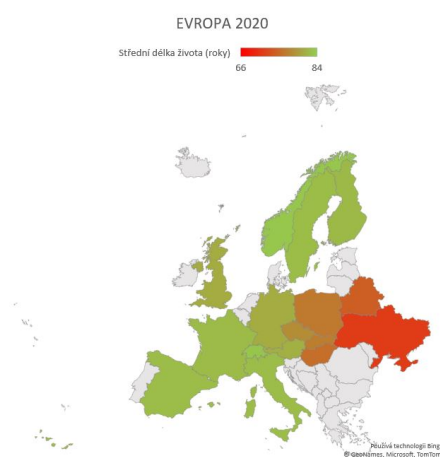
Obrázek 3.18: Střední délka života v Evropě v roce 1960 (vlevo) a 1970 (vpravo)



Obrázek 3.19: Střední délka života v Evropě v roce 1980 (vlevo) a 1990 (vpravo)



Obrázek 3.20: Střední délka života v Evropě v roce 2000 (vlevo) a 2010 (vpravo)



Obrázek 3.21: Střední délka života v Evropě v roce 2020

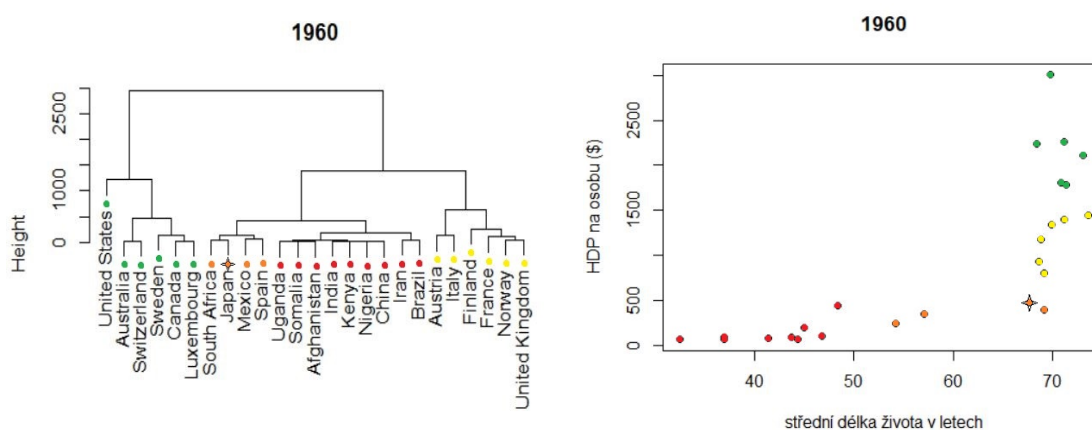
3.3. Vývoj závislosti mezi střední délkou života a hrubým domácím produktem

V následující kapitole budeme pozorovat vývoj závislosti mezi střední délkou života a hrubým domácím produktem na osobu. V pozorovaném období budeme sledovat roky po jednotlivých desetiletích od roku 1960 do roku 2020. Hodnota hrubého domácího produktu na osobu je uváděna v amerických dolarech.

V některých letech pro pár vybraných zemí bohužel chybí data (viz data v příloze), a proto tyto země nebudou zahrnuty do celého časového horizontu.

Pro lepší přehlednost jsme si pomocí grafického editoru vykreslili shluky barevně. Do bodového grafu jsme přenesli zbarvení, aby bylo jasné, kde přesně se shluky nacházejí a zmíněné země v textu jsou označeny symbolem. Jednotlivé země nejsou v grafu popsány z důvodu přehlednosti, avšak veškeré informace najdeme v příloženém excelovském souboru pod názvem *Data_bakalarska_prace*.

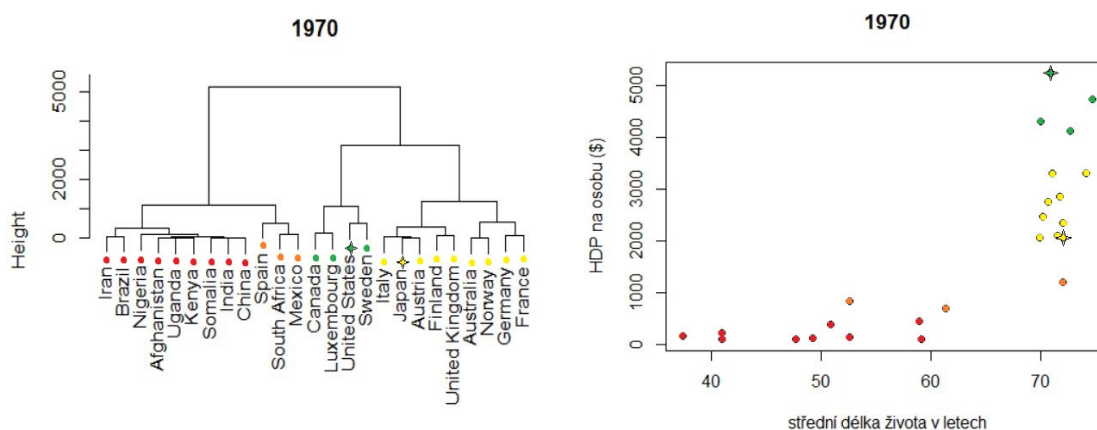
Na začátku sledovaného období, v roce 1960, můžeme vidět pomocí provedené shlukové analýzy tři větší shluky. První skupinu tvoří státy, které mají nejvyšší HDP na osobu, a také jejich střední délka života patří k nejvyšším ve své době. Mezi tyto státy patří Spojené státy americké, Kanada, Austrálie a z Evropy Lucembursko, Švédsko a Švýcarsko. Druhý větší shluk naopak tvoří státy, které



Obrázek 3.22: Vztah mezi střední délkou života a hrubým domácím produktem na osobu v roce 1960

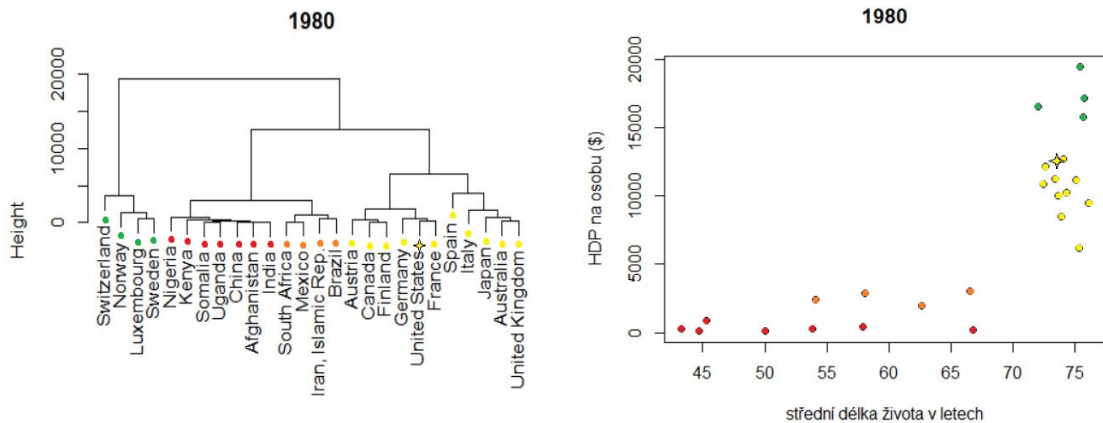
mají jedno z nejnižších HDP na osobu, a dále se dělí na dva shluky podle střední délky života, která také nedosahuje vysokých hodnot. Můžeme vidět, že veškeré státy třetího světa, spadají právě do tohoto shluku. Průměrné hodnoty hrubého domácího produktu mají státy, které patří do třetího shluku, a jsou to například Finsko, Francie a Itálie. Hrubý domácí produkt na osobu mají sice nižší než státy prvního shluku, ale střední délku život mají podobnou, jako státy v prvním shluku.

O deset let později, v roce 1970, můžeme pozorovat velice podobné shlukování, ale zvýšilo se HDP na osobu. Například v Japonsku se téměř zčtyřnásobilo.



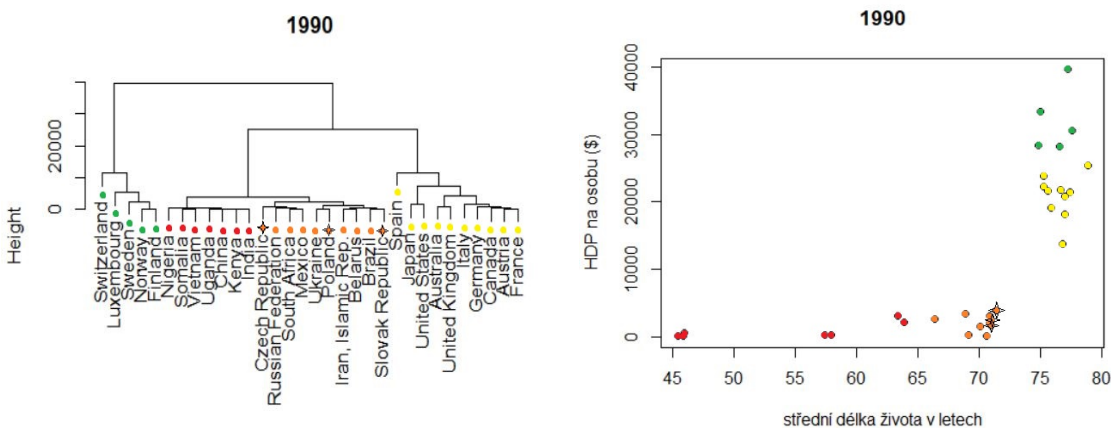
Obrázek 3.23: Vztah mezi střední délkou života a hrubým domácím produktem na osobu v roce 1970

V roce 1980 pozorujeme změnu umístění Spojených států amerických. Můžeme vidět, že se během uplynulých deseti let zvýšilo HDP na osobu u Lucemburska, Švýcarska, Švédska a Norska mnohem rychleji než v USA, i přestože se v USA zvýšilo HDP na osobu více jak dvojnásobně.



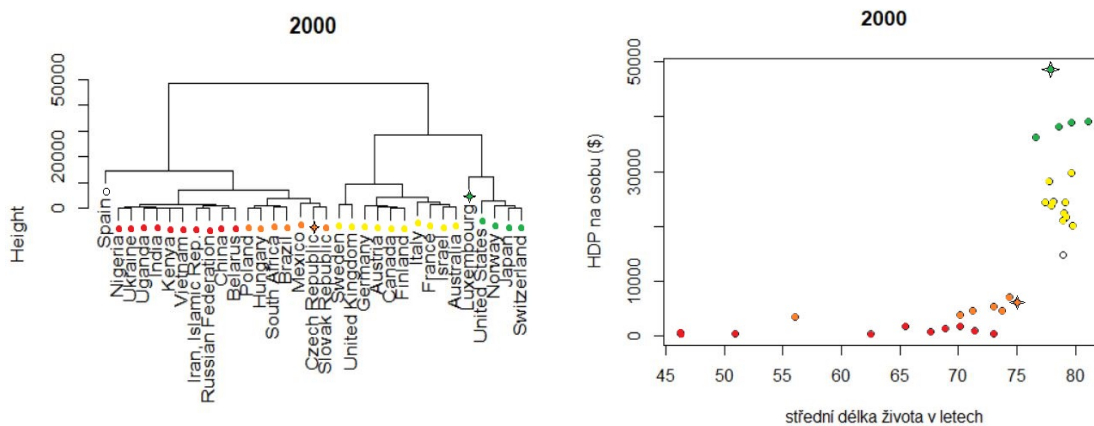
Obrázek 3.24: Vztah mezi střední délkou života a hrubým domácím produktem na osobu v roce 1980

Od roku 1990, kdy již máme hodnoty pro podstatnou část Evropy, můžeme vidět její rozdělení. Postkomunistické země jako je Česká republika, Slovenská republika a Polsko nabývají podobných hodnot. Na západ od naší republiky mají státy jak vyšší střední délku života, tak vyšší HDP na osobu, tudíž jsou vyspělejší státy ve stejných shlucích.

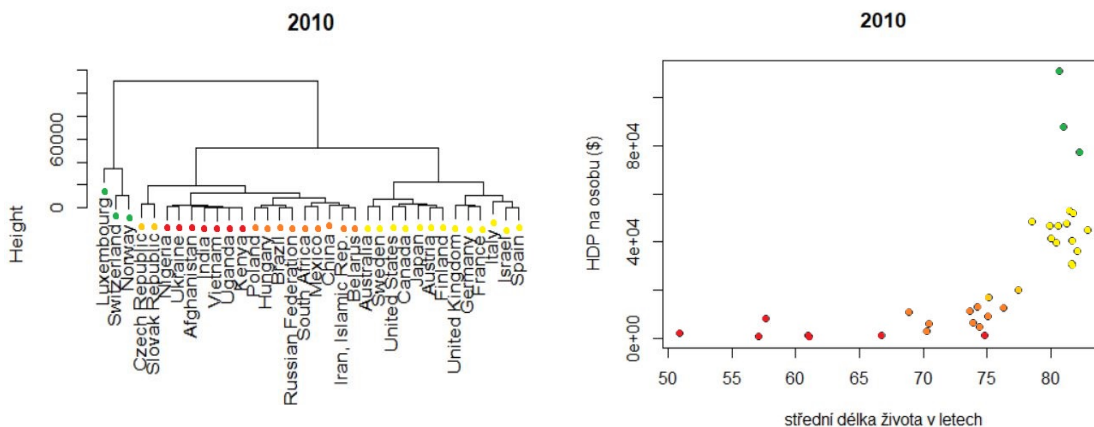


Obrázek 3.25: Vztah mezi střední délkou života a hrubým domácím produktem na osobu v roce 1990

V roce 2000 se shluky oproti minulému desetiletí nijak zásadně nezměnily. Pouze vidíme, že Lucembursko se začíná svým HDP na osobu čím dál tím více vzdalovat ostatním státům.

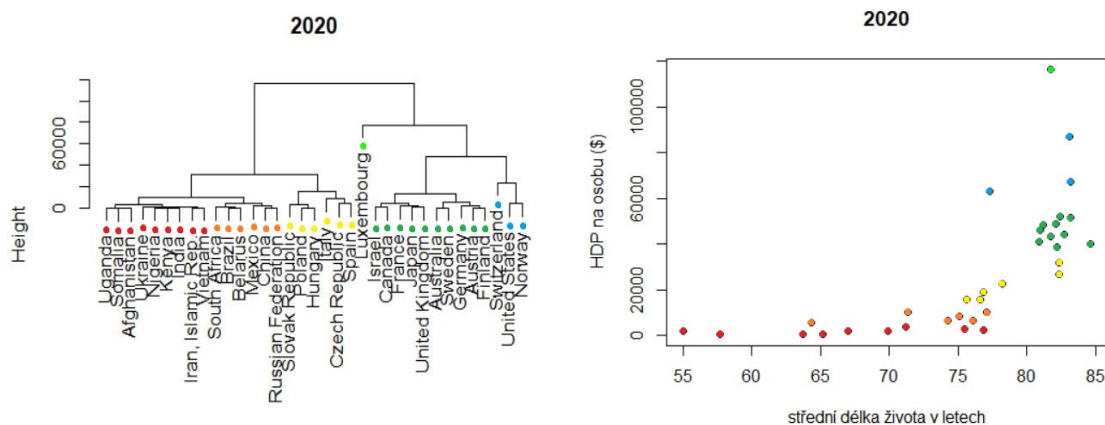


Obrázek 3.26: Vztah mezi střední délkou života a hrubým domácím produktem na osobu v roce 2000



Obrázek 3.27: Vztah mezi střední délkou života a hrubým domácím produktem na osobu v roce 2010

Lucembursko bylo v roce 2020 země s nejvyšším HDP na osobu a dosahovalo částky 116 356 dolarů na osobu. V roce 1960 však měla 2 242 dolarů HDP na osobu, tzn. že svoje HDP na osobu dokázali zvýšit během 60 let na padesáti násobek.



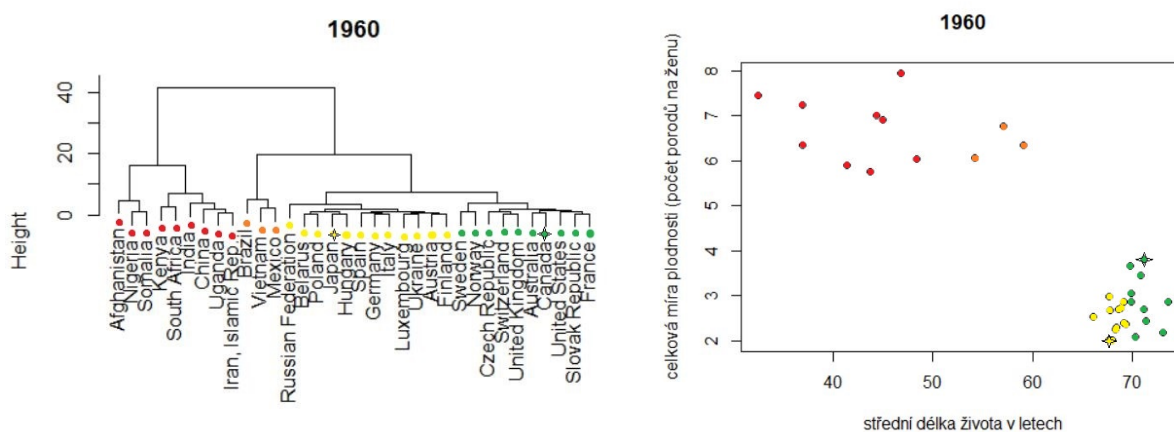
Obrázek 3.28: Vztah mezi střední délkou života a hrubým domácím produktem na osobu v roce 2020

Během celého vývoje v čase si lze všimnout rostoucího trendu závislosti - se zvyšujícím se HDP na osobu roste střední délka života. Také zde můžeme předpokládat, že ekonomická úroveň jednotlivých států roste s životní úrovní.

3.4. Vývoj závislosti mezi střední délkou života a úhrnnou plodností

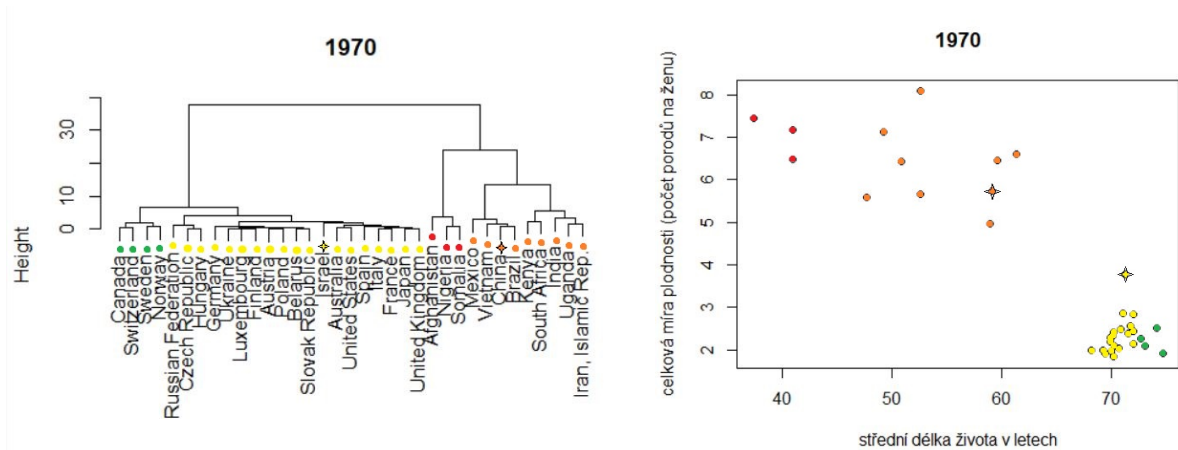
V této kapitole budeme pozorovat taktéž závislost, a to mezi střední délkou života a úhrnnou plodností. Opět bude naše sledované období od roku 1960 do roku 2020 po deseti letech. Úhrnnou plodnost uvádíme počtem dětí na jednu ženu v reprodukčním období.

Nejvyšší počet dětí už v tuto chvíli můžeme očekávat v rozvojových zemích, tedy v zemích třetího světa.



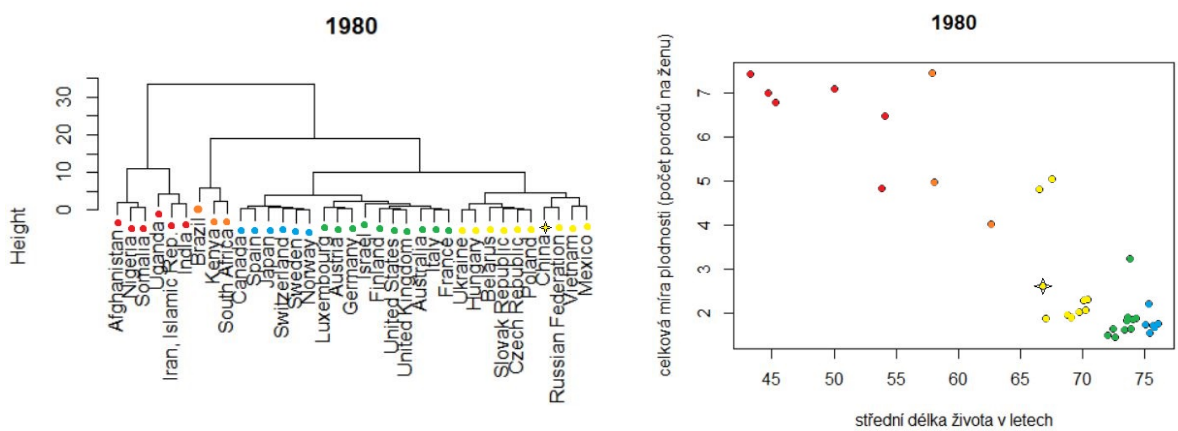
Obrázek 3.29: Vztah mezi střední délkou života a úhrnnou plodností v roce 1960

Z dendrogramu 3.29 vidíme rozdělení do několika shluků. Při rozdělení na dva největší shluky vidíme, že první shluk obsahuje země třetího světa. A z prvního bodového grafu ihned potvrzujeme náš předpoklad, a to, že nejvyšší počet dětí mají země jako Nigérie, Keňa, Indie a Somálsko. Z grafu lze dále vyčíst, že s nejvyšším počtem dětí souvisí nejnižší střední délka života. Můžeme předpokládat, že v těchto zemích není tak vysoká životní úroveň. Nejnižší úhrnnou plodnost v roce 1960 mělo Japonsko s hodnotou 2 děti na jednu ženu.



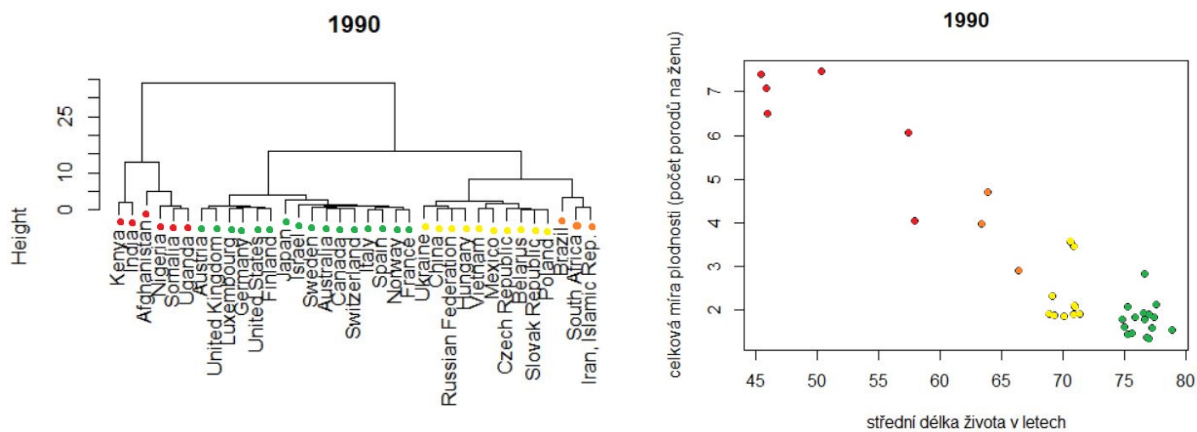
Obrázek 3.30: Vztah mezi střední délkou života a úhrnnou plodností v roce 1970

V roce 1970 Keňa přesáhla hodnotu 8 dětí na jednu ženu. Izrael v roce 1970 patřil k zemím s vyšší střední délkou života a zároveň měl z těchto zemí nejvyšší úhrnnou plodnost, a to 3,8 dítěte na jednu ženu.

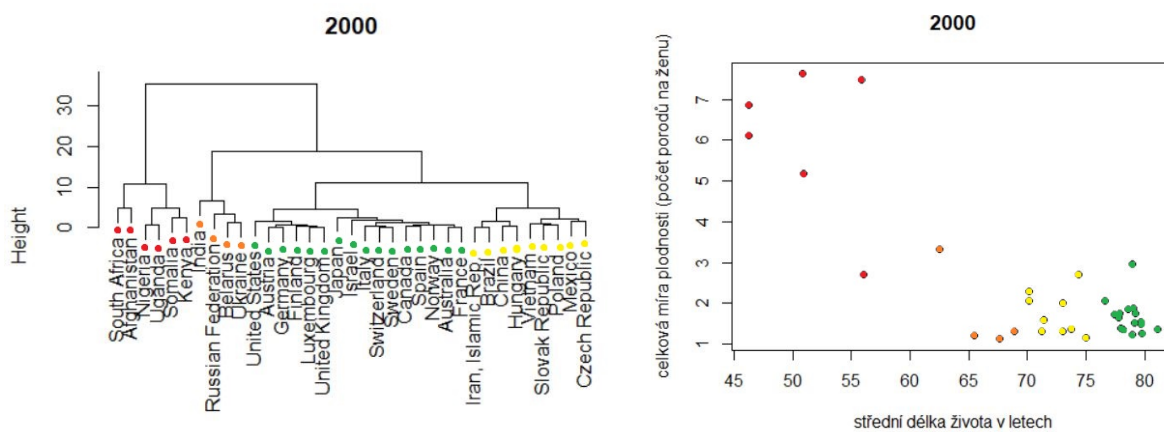


Obrázek 3.31: Vztah mezi střední délkou života a úhrnnou plodností v roce 1980

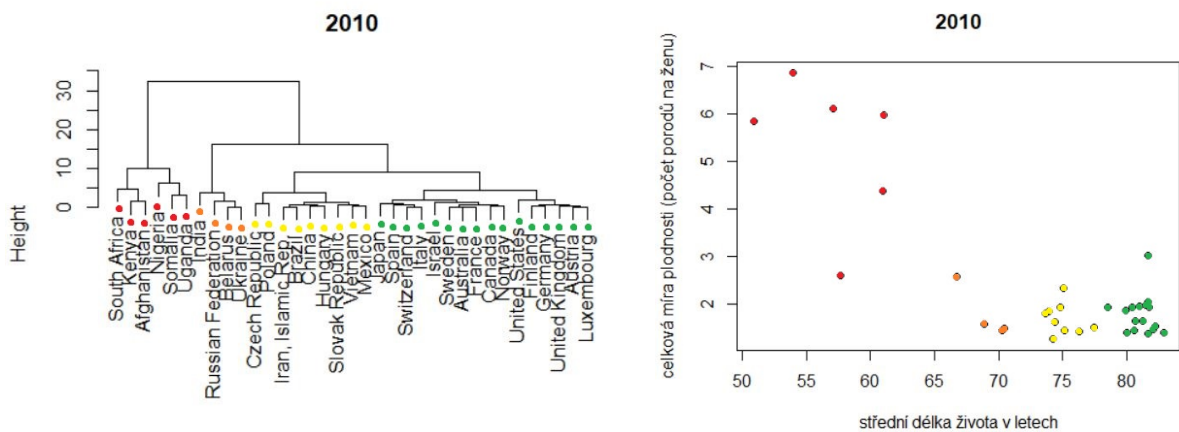
Mezi lety 1970 a 1980 došlo v Číně k největšímu poklesu úhrnné plodnosti ze všech zemích v celém sledovaném období. Tento pokles byl zřejmě způsoben zavedením tzv. Politiky jednoho dítě, což znamenalo přísný zákaz mít více dětí. Zákaz byl reakcí na situaci po Kulturní revoluci, kdy hospodářství v Číně zcela krahovalo a nebylo schopno uživit ani tehdejší populaci. [29]



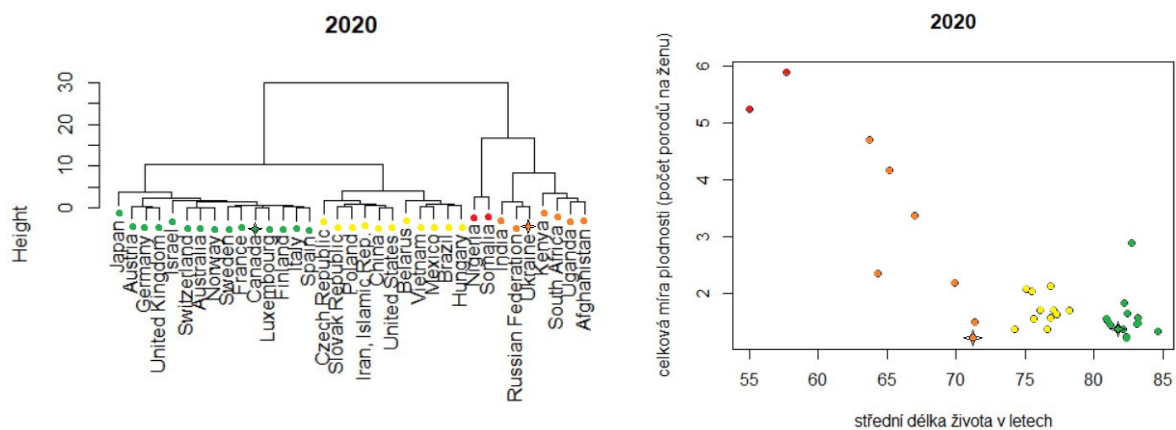
Obrázek 3.32: Vztah mezi střední délkou života a úhrnnou plodností v roce 1990



Obrázek 3.33: Vztah mezi střední délkou života a úhrnnou plodností v roce 2000



Obrázek 3.34: Vztah mezi střední délkou života a úhrnnou plodností v roce 2010



Obrázek 3.35: Vztah mezi střední délkou života a úhrnnou plodností v roce 2020

Velký pokles úhrnné plodnosti z vyspělých zemí můžeme pozorovat u Kanady, kde v roce 1960 byla úhrnná plodnost 3,8 a v roce 2020 pouze 1,4 dítěte na jednu ženu. Jestliže se celkově podíváme na země světa, během sledovaného období došlo ve všech zemích ke snížení úhrnné plodnosti a zvýšení střední délky života. Nejvyšší úhrnné plodnosti, přes 5 dětí na jednu ženu, v roce 2020 dosahovalo Somálsko s Nigérií, ve kterých byla v daném roce také nejnižší střední délka života. Nejnižší úhrnné plodnosti v roce 2020 měla Ukrajina s hodnotou 1,21 dětí na jednu ženu, a hned za Ukrajinou následovalo Španělsko s Itálií s hodnotami 1,23 a 1,24 dětí na jednu ženu.

I ve vztahu střední délky života a úhrnné plodnosti můžeme pozorovat životní úroveň jednotlivých zemí, kdy země s nejvyšší střední délkou života a nejnižší úhrnnou plodností patří mezi ekonomicky nejvyspělejší země světa.

3.5. Korelační analýza pro vybrané faktory

Při tvorbě korelační analýzy nás bude zajímat, zda spolu střední délka života a úhrnná plodnost nebo střední délka života a hrubý domácí produkt na osobu ve sledovaných letech souvisí. Pro měření závislostí budeme používat Spearmanův korelační koeficient. Spearmanův korelační koeficient využívá pořadí realizací náhodných veličin a neovlivňují jej odlehle hodnoty, tudíž je pro nás vhodnou volbou. Následně otestujeme, zda jsou vybrané faktory závislé. Korelační analýzu budeme provádět pomocí softwaru R.

H_0 : námi vybrané faktory jsou nezávislé

H_A : námi vybrané faktory jsou závislé

Výsledky ze softwaru R jsme si přepsali do tabulky 3.1, ve které můžeme vidět spočítané hodnoty pro korelační analýzu. V prvním sloupci uvádíme střední délku života (zkr. SDŽ) s vybraným faktorem, pro které jsme provedli korelační analýzu. Ve druhém sloupci máme hodnotu Spearmanova korelačního koeficientu. Do třetího sloupce jsme zaznamenali p -value neboli p -hodnotu. A nakonec v po-

	$r_{X,Y}$	p-value	rozhodnutí
SDŽ a HDP/os 1960	0,886709	$3,613 \times 10^{-09}$	zamítáme
SDŽ a HDP/os 1970	0,8	$3,535 \times 10^{-06}$	zamítáme
SDŽ a HDP/os 1980	0,7702564	$7,831 \times 10^{-06}$	zamítáme
SDŽ a HDP/os 1990	0,8090176	$6,736 \times 10^{-07}$	zamítáme
SDŽ a HDP/os 2000	0,8308824	$3,431 \times 10^{-07}$	zamítáme
SDŽ a HDP/os 2010	0,8352941	$2,087 \times 10^{-07}$	zamítáme
SDŽ a HDP/os 2020	0,8661064	$2,657 \times 10^{-08}$	zamítáme
SDŽ a plodnost 1960	-0,6160312	0,000105	zamítáme
SDŽ a plodnost 1970	-0,5835143	0,0002339	zamítáme
SDŽ a plodnost 1980	-0,7901961	$3,892 \times 10^{-07}$	zamítáme
SDŽ a plodnost 1990	-0,7501926	$2,121 \times 10^{-07}$	zamítáme
SDŽ a plodnost 2000	-0,4406162	0,008599	zamítáme
SDŽ a plodnost 2010	-0,4186568	0,01231	zamítáme
SDŽ a plodnost 2020	-0,5504973	0,0006115	zamítáme

Tabulka 3.1: Hodnoty Spearmanova korelačního koeficientu pro střední délku života spolu s hrubým domácím produktem na osobu a střední délku života s úhrnnou plodností

sledním sloupci máme naše rozhodnutí, zda nulovou hypotézu zamítáme či nikoliv. Hypotézu testujeme na hladině významnosti $\alpha = 0,05$.

Pro závislost mezi střední délkou života a hrubým domácím produktem na osobu nám vyšly korelační koeficienty blízké 1, tudíž bychom mohli předpokládat závislost mezi střední délkou života a hrubým domácím produktem na osobu. Ovšem abychom předpoklad potvrdili, budeme testovat významnost korelačního koeficientu. p -value nám vyšly po celou dobu pozorování, od roku 1960 do roku 2020, mnohem menší než 0,05, a proto jsme ve všech letech nulovou hypotézu mohli zamítnout. Testování tedy prokázalo, že mezi střední délkou života a hrubým domácím produktem na osobu po celou dobu pozorování je závislost.

Korelační koeficient jsme taktéž spočítali pro střední délku života a úhrnnou plodnost. Korelační koeficienty nám vycházely sice blíže 0 než v předchozím případě, ale stále dost vysoké hodnoty, u kterých bychom mohli předpokládat závislost. V tomto případě nám vycházely záporné hodnoty, tzn. s růstem střední délky života bude úhrnná plodnost klesat. Spočítali jsme příslušné p -value, které nám vycházely větší než v případě s hrubým domácím produktem, ale opět nám vyšly mnohem menší než 0,05. Tím pádem jsme nulovou hypotézu také zamítli a testování nám prokázalo i v tomto případě, že je mezi střední délkou života a úhrnnou plodností po celou dobu pozorování závislost.

Abychom zjistili „skutečnou“ závislost mezi střední délkou života a úhrnnou plodností, spočítali jsme si parciální korelační koeficienty při vyloučení vlivu hrubého domácího produktu na osobu.

Parciální korelační koeficienty jsme spočítali pomocí softwaru R a zapsali jsme je do tabulky 3.2. V prvním sloupci nalezneme ukazatele, pro které je provedena parciální korelační analýza, a ve kterých letech. Ve druhém sloupci se nachází hodnota parciálního korelačního koeficientu. X je pro nás střední délka života, Y hrubý domácí produkt na osobu a Z úhrnná plodnost. Ve třetím sloupci je vypsána p -value a v posledním sloupci rozhodnutí o hypotéze.

V roce 1960 pro střední délku života a úhrnnou plodnost byla hodnota Spearmanova korelačního koeficientu rovna -0,616. Jestliže spočítáme parciální korelaci

	r	p-value	rozhodnutí
SDŽ a plodnost 1960	$r_{X,Z.Y} = -0,2348042$	0,2694094	nezamítáme
SDŽ a plodnost 1970	$r_{X,Z.Y} = -0,2834718$	0,1794938	nezamítáme
SDŽ a plodnost 1980	$r_{X,Z.Y} = -0,1899653$	0,3630758	nezamítáme
SDŽ a plodnost 1990	$r_{X,Z.Y} = -0,4928449$	0,0048496	zamítáme
SDŽ a plodnost 2000	$r_{X,Z.Y} = -0,2690549$	0,1364702	nezamítáme
SDŽ a plodnost 2010	$r_{X,Z.Y} = -0,2040405$	0,2547195	nezamítáme
SDŽ a plodnost 2020	$r_{X,Z.Y} = -0,1395183$	0,4313079	nezamítáme

Tabulka 3.2: Hodnoty parciálního Spearmanova korelačního koeficientu

při vyloučení vlivu hrubého domácího produktu na osobu, klesne v absolutní hodnotě parciální korelační koeficient na -0,235. Dá se předpokládat, že při očištění vlivu třetího faktoru nebude střední délka života s úhrnnou plodností závislá. Provedeme testování

$$H_0 : r_{Y,Z.X} = 0$$

$$H_A : r_{Y,Z.X} \neq 0.$$

Hypotézu testujeme na hladině významnosti $\alpha = 0,05$. p -value v tomto případě vyšla 0,269. Je tedy mnohem větší než hladina významnosti, a proto nulovou hypotézu nelze zamítnout. Znamená to, že pro stálý hrubý domácí produkt na osobu nelze předpokládat závislost mezi střední délkou života a úhrnnou plodností. Jednoduchý korelační koeficient mohl ukazovat klamnou závislost.

Z celého sledovaného období pouze v roce 1990 vyšel korelační koeficient vzdálenější od nuly, a i p -value vyšla menší než $\alpha = 0,05$. Tudíž jsme nulovou hypotézu zamítli a mohli bychom říct, že v roce 1990 byla střední délka života závislá spolu s úhrnnou plodností s vyloučením hrubého domácího produktu na osobu. Avšak z dlouhodobého hlediska závislost u střední délky života a úhrnné plodnosti můžeme vyloučit.

Závěr

Na začátku praktické části jsme provedli analýzu vývoje střední délky života v čase pro vybrané země pro jednotlivé pohlaví zvlášť. Z poskytnutých dat jsme zjistili, že střední délka života pro ženy je historicky vyšší než pro muže. Při určování závislostí pomocí Spearmanova korelačního koeficientu mezi střední délkou života a hrubým domácím produktem na osobu, nebo úhrnnou plodností, jsme zjistili, že s rostoucím hrubým domácím produktem na osobu roste střední délka života. A s klesající úhrnnou plodností roste střední délka života. Avšak při počítání parciálního korelačního koeficientu pro střední délku života a úhrnnou plodnost, se nám dříve zjištěná závislost vyvrátila.

Na závěr můžeme říct, že z provedené analýzy je při vývoji střední délky života zřejmý rostoucí trend hodnot střední délky života, avšak během vývoje dochází k výkyvům způsobeným aktuální situací v daném státu. Výkyvy mohou být ovlivňovány válkami, politikou v zemi, nemocemi a jinými faktory. Dále jsme zjistili, že samotný vývoj je závislý na námi zvoleném faktoru hrubý domácí produkt na osobu. Do budoucna se dá předpokládat stálý růst střední délky života, protože hrubý domácí produkt s největší pravděpodobností také stále poroste. I když například během krize se HDP nemusí zvýšit, může se snížit životní úroveň, a tím pádem může nastat stagnace, nebo dokoce pokles střední délky života. Na faktoru úhrnné plodnosti střední délka života závislá není, avšak je zřejmé, že dochází k poklesu úhrnné plodnosti, tj. rodí se méně dětí, dále populace stárne a stát by měl dbát na sociální systém státu v souvislosti s transferovými platbami.

Literatura

- [1] Anděl J.: *Základy matematické statistiky* MATFYZPRESS, Praha, 2007.
- [2] Bacäer N.: *A short history of mathematical population dynamics* Springer, Londýn, 2010.
- [3] Bellová J., Musilová S., Zlámal J.: *Základy ekonomie* Univerzita Palackého v Olomouci, Olomouc, 2013.
- [4] Bílková D., Budinský P., Vohánka V.: *Pravděpodobnost a statistika* Vydavatelství a nakladatelství Aleš Čeněk s.r.o., Plzeň, 2009.
- [5] Budíková M., Lerch T., Mikoláš Š.: *Základní statistické metody* Masarykova univerzita, Brno, 2005.
- [6] Cipra T.: *Pojistná matematika: teorie a praxe*. Ekopress, Praha, 2006.
- [7] Cipra T.: *Finanční a pojistné vzorce* Grada Publishing, Praha, 2006.
- [8] Frank R. H., Bernanke B. S.: *Ekonomie* Grada, Praha, 2003.
- [9] Holman R.: *Ekonomie* C. H. Beck, Praha, 2005.
- [10] Hron K., Kunderová P., Vencálek O.: *Základy počtu pravděpodobnosti a metod matematické statistiky* Univerzita Palackého v Olomouci, Olomouc, 2018.
- [11] Jurečka V., Hlaváček K., Jánošíková I., Kolcunová E., Macháček M., Paličková I., Spáčilová L., Šimek M.: *Úvod do ekonomie* VŠB - Technická univerzita Ostrava, Ostrava, 2015.
- [12] Kalibová K., Pavlík Z., Vodáková A.: *Demografie (nejen) pro demografy* Sociologické nakladatelství, Praha, 1998.
- [13] Řezanková H., Húsek D., Snášel V.: *Shluková analýza dat. 2., rozš. vyd.* Professional Publishing, Praha, 2009.
- [14] Český statistický úřad - Naděje dožití aneb kolika let se můžeme dožít v našem kraji? [online]. Dostupné z: <https://www.czso.cz/csu/xu/nadeje-do-ziti-aneb-kolika-let-se-muzeme-do-zit-v-nasem-kraji>.

- [15] Český statistický úřad - Metodické vysvětlivky [online]. Dostupné z: https://www.czso.cz/csu/czso/4024-07-v_obdobi_2001_az_2005-metodicke_vysvetlivky.
- [16] Český statistický úřad - Metodika úmrtnostních tabulek [online]. Dostupné z: https://www.czso.cz/documents/10180/85591762/metodika_ut_akt2020.pdf/-96681886-68bc-420e-b2de-69b9fe6e7b80?version=1.1.
- [17] Český statistický úřad - Úmrtnostní tabulky za ČR 2020 [online]. Dostupné z: <https://www.czso.cz/documents/10180/142756988/1300632102.pdf/1a207fd7-00dc-412b-907a-0c656a71ca97?version=1.1>.
- [18] Demografie [online]. Dostupné z: http://www.demografie.info/?cz_obecne=.
- [19] EDU Univerzity Hradec Králové [online]. Dostupné z: https://edu.uhk.cz/jindrvo1/files/maek1/texty/01_Uvod_do_makroekonomie.pdf.
- [20] Elektronická učebnice úvodu do politické geografie MUNI - Šestidenní válka [online]. Dostupné z: <https://polgeo.geogr.muni.cz/konflikty-smlouvy-a-organizace/sestidenni-valka/>.
- [21] Elektronická učebnice úvodu do politické geografie MUNI - Vietnamská válka [online]. Dostupné z: <https://polgeo.geogr.muni.cz/konflikty-smlouvy-a-organizace/vietnamska-valka/>.
- [22] Encyklopedie migrace - Íránská islámská revoluce [online]. Dostupné z: <https://www.encyclopediaofmigration.org/iranska-islamska-revoluce-1978-1979/>.
- [23] History - Idi Amin [online]. Dostupné z: <https://www.history.com/topics/africa/idi-amin>.
- [24] Informační centrum bezpečnosti potravin - Japonci se dožívají v průměru nejdelšího věku na světě [online]. Dostupné z: <https://www.bezpecnostpotravin.cz/japonci-se-dozivaji-v-prumeru-nejdelsiho-veku-na-svete.aspx>.
- [25] iRozhlas.cz - HIV v Jihoafrické republice [online]. Dostupné z: https://www.irozhlas.cz/zpravy-svet/jihoafricka-republika-hiv-aids.2201101219_pj.
- [26] MUNI Seminář čínských studií - Politika jednoho dítěte [online]. Dostupné z: <https://cinskastudia.phil.muni.cz/studentске-projekty/cinska-spolecnost/politika-jednoho-ditete>.

- [27] OECD Data [online]. Dostupné z: <https://data.oecd.org/pop/fertility-rates.htm>.
- [28] Peníze.cz - Zdravotnictví USA [online]. Dostupné z: <https://www.penize.cz/pojisteni/29425-zdravotnictvi-usa-zdravi-knezaplaceni>.
- [29] Přírodovědci.cz - Demografické stárnutí Číny [online]. Dostupné z: <https://www.prirodovedci.cz/aktuality/demograficke-starnuti-ciny>.
- [30] Přírodovědci.cz - HIV a AIDS v subsaharské Africe [online]. Dostupné z: <https://www.prirodovedci.cz/geograf/clanky/hiv-a-aids-v-subsaharske-africe-spatne-i-lepsi-zpravy>.
- [31] Přírodovědecká fakulta Univerzity Karlovy - Katedra demografie a geodemografie - Co je to demografie? [online]. Dostupné z: <https://www.natur.cuni.cz/geografie/demografie-a-geodemografie/studium/informace-pro-zajemce-o-studium-2/co-je-to-demografie>.
- [32] Příručka demografické statistiky pro potřebu oblastních statistických orgánů [online]. Dostupné z: <https://web.natur.cuni.cz/ňak/gos/demmetodika/www.czso.cz/cz/cisla/0-02/020100/narozeni.htm>.
- [33] Quick-R [online]. Dostupné z: <https://www.statmethods.net/graphs/scatterplot.html>.
- [34] The World Bank - Data Bank [online]. Dostupné z: <https://databank.worldbank.org/source/world-development-indicators/Type/TABLE/preview/on#>.