

Jihočeská univerzita v Českých Budějovicích

Ekonomická fakulta

Katedra matematiky a aplikované informatiky

Explorační prostorová analýza dat

Autor diplomové práce: **Jan Králík**

Vedoucí diplomové práce: **RNDr. Renata Klufová, Ph.D.**

Rok obhajoby: **2010**

Prohlášení

Prohlašuji, že jsem bakalářskou práci na téma »**Explorační prostorová analýza dat**«
vypracoval samostatně na základě vlastních zjištění a materiálů, které uvádím v seznamu
použité literatury.

Frymburk, duben 2010

.....
podpis studenta

Poděkování

Za cenné rady, náměty a inspiraci

bych chtěl poděkovat paní RNDr. Renatě Klufové, Ph.D., z Jihočeské univerzity v Českých

Budějovicích, Ekonomické fakulty.

Obsah

Úvod	1
1. Teoreticko - metodologická část	3
1.1. Základní pojmy	3
1.1.1. Analýza	3
1.1.2. Data	3
1.1.3. Deskriptivní statistika	4
1.1.4. Induktivní statistika	4
1.2. Měření a typy škál	5
1.2.1. Nominální škála	5
1.2.2. Ordinální (pořadová) škála	5
1.2.3. Intervalová škála	6
1.2.4. Podílová škála	6
1.3. Historie	6
1.4. Explorační analýza dat	8
1.4.1.1. Medián	9

1.4.1.2.	Modus.....	9
1.4.1.3.	Krabicový graf (Box plot)	9
1.4.1.4.	Kvartil.....	10
1.4.1.5.	Mezikvartilové rozpětí	10
1.4.2.	ESDA analýza	10
1.4.2.1.	Smoothing (Vyhlažování)	11
1.4.2.2.	Identifikace trendů a gradientů na mapě	12
1.4.2.3.	Detekce odlehlých prostor.....	12
1.4.3.	Prostorová autokorelace	12
1.4.3.1.	Určení vztahů sousedství.....	13
1.4.3.2.	Způsoby definování sousedství.....	13
1.4.3.3.	Míry prostorové autokorelace.....	14
1.4.3.3.1.	Globální míry.....	14
1.4.3.3.2.	Lokální míry	17
1.5.	GIS, GeoDA	18
1.5.1.	GIS	18
1.5.2.	GeoDa.....	18
1.5.2.1.	Luc Anselin.....	19

2.	Demografická data.....	20
2.1.	Demografické jevy.....	21
2.1.1.	Přehled a členění demografických ukazatelů	22
2.2.	Analýza struktury obyvatelstva	23
2.2.1.	Struktura obyvatelstva podle pohlaví.....	24
2.2.1.1.	Sekundární index maskulinity – sima	24
2.2.1.2.	Mužská nadúmrtnost	25
2.2.1.3.	Migrace.....	25
2.2.2.	Struktura obyvatelstva podle věku	25
2.2.2.1.	Biologické generace	26
2.2.2.2.	Typy populace.....	27
2.2.2.3.	Index stáří	28
2.2.2.4.	Průměrný věk.....	30
2.2.3.	Ekonomická a sociální struktura.....	31
2.2.3.1.	Ekonomické generace	31
2.2.3.2.	Podíl obyvatel nad 65 let	31
2.2.4.	Celkové charakteristiky reprodukce.....	33
2.2.4.1.	Celkový přírůstek.....	34

2.2.4.2. Přirozený přírůstek (PP)	34
3. Analýza demografických dat pomocí explorační analýzy	38
3.1. Příprava dat pro GeoDa.....	38
3.2. Vytvoření vah v GeoDa	38
3.3. Aplikace dat pomocí programu GeoDa.....	40
3.4. Praktická ukázka aplikace dat GeoDa.....	42
3.4.1. Index stárí	42
3.4.2. Průměrný věk.....	47
3.4.3. Podíl osob nad 65 let	49
3.4.4. Přirozený přírůstek.....	52
Závěr.....	56
Summary.....	57
Seznam grafů	59
Seznam obrázků	60
Seznam tabulek	61
Seznam map.....	62
Seznam použité literatury a zdrojů.....	63

Úvod

Tato práce se zabývá explorační prostorovou analýzou dat a cílem práce je zhodnocení možností využití explorační prostorové analýzy při analýze socio-ekonomických dat.

Explorační prostorová analýza dat (ESDA) je souhrn metod používaných pro průzkum **prostorových dat** různého charakteru, kde se užívá celá řada grafických či vizuálních metod a map. Pro prostorová data je charakteristické, že se vztahují k určitým místům v prostoru, tzn. je již známo, kde se data vyskytují a jakou mají polohu. Explorační prostorová analýza dat se pak snaží odpovídat na otázky typu: „*Jsou zde nějaké zajímavé vztahy nebo zákonitosti?*“.¹

V první kapitole je uveden teoreticko-metodologický popis práce, především teoretické metody a přístupy k dané problematice a stručná historie. Také jsou představeny 2 základní počítačové programy, ve kterých lze provádět explorační prostorovou analýzu dat – ArcGIS a GeoDa. Ve druhé kapitole je uveden základní teoretický přehled a popis demografických ukazatelů. Vybrané demografické ukazatele (index stárí, průměrný věk, podíl osob nad 65 let a přirozený přírůstek) jsou pak ve třetí (praktické) kapitole aplikovány do map a grafů, pomocí zvoleného programu GeoDa. Tyto mapy jsou následně zhodnoceny a podrobně popsány.

Demografická data mají široké využití a jejich analýza má pro společnost zásadní význam. Proto je nutné používat takové programy, metody a analýzy, které napomohou k snadné a co nejpřesnější analýze demografických dat. Cílem této práce zhodnotit aplikaci demografických dat pomocí počítačového programu GeoDa. Hodnotíme-li demografická data na regionální úrovni, pak se jedná o data prostorová a je tedy vhodné aplikovat metody ESDA. V poslední době se rozvíjí i samostatná vědní disciplína – geodemografie, která se zabývá analýzou demografických dat v prostorovém kontextu. Geodemografie studuje přirozený a mechanický pohyb obyvatelstva

¹ CENTRUM INFORMAČNÍCH TECHNOLOGIÍ OSTRAVSKÉ UNIVERZITY. GIS laboratoř. [online]. URL: <<http://cit.osu.cz/gis/pages/coJeToGis.php>> [cit. 2010-05-18].

a také demografické procesy struktur (věková, ekonomická). Tato demografická data studuje geodemografie v jejích regionálních rozdílnostech.²

² MARYÁŠ, J. – VYSTOUPIL, J. *Geografie obyvatelstva a geografie sídel*. [online].

URL: < http://www.gisaci.upol.cz/filesftp/GEOGRAFIE_OBYVATELSTVA_A_SIDELO.pdf > [cit. 2010-05-18].

1. Teoreticko - metodologická část

1.1. Základní pojmy

1.1.1. Analýza

Tento pojem lze chápat jako postup rozdělení celku na součásti, tyto součásti pak zkoumáme, abychom pochopili celek.³

1.1.2. Data

Data chápeme jako **vyjádření části reálného světa**, ve většině případů se jedná o číselné vyjádření. Tato data lze pak zobrazovat různými způsoby nebo v různých formách (fotografie, mapy, kresby atd.). Ve statistice jsou data vyjádřena formou číselných hodnot, která se vyjadřují pomocí tabulek. V tabulce každý řádek odpovídá jednotlivému zkoumanému objektu a sloupce znázorňují sledované veličiny. V tabulce jsou pak znázorněny hodnoty veličin, které byly zjištěny a přiřazeny ke každému ze sledovaných objektů. Číselné hodnoty uvnitř tabulky tvoří homogenní datovou strukturu o n řádcích a p sloupcích, která se v matematice nazývá matice.

Data poskytující informace o **celé populaci** nazýváme **základním souborem**. Populace ale může být v některých případech příliš rozsáhlá (např. změřit všechny objekty je časově nebo finančně náročné nebo celkově nemožné). V tomto případě pak zobrazují pouze část objektů populace.⁴

³ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

⁴ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

Statistika

Statistika je vědní obor, který se zabývá studiem zákonitostí hromadných jevů (jevů, které se vyskytují v masovém měřítku). Cílem statistiky je popis vlastností těchto jevů a odhalování souvislostí mezi nimi.

Statistická data jsou číselná zobrazení takového výseku reálného světa, ve kterém se zobrazované objekty vyskytují hromadně. Tyto výseky z reálného světa, které zahrnují více objektů se stejnými vlastnostmi, nazýváme populace.⁵

1.1.3. Deskriptivní statistika

Deskriptivní (popisná) statistika je postup, kdy uspořádáváme soubory a následně je popisujeme a sumarizujeme. Jedná se o analýzu obrazu **celé populace** (nazývaný základní soubor).

Prvním krokem je vymezení souboru prvků, na kterých se bude zkoumat daný jev. Následuje vyšetření všech prvků z hlediska studovaného jevu. Výstupem z této analýzy je přehledný popis dat a popis sledovaného výseku reálného světa (jak kvalitativní, tak kvantitativní výsledky).^{6, 7, 8}

1.1.4. Induktivní statistika

Pokud není k dispozici základní soubor dat, zobrazují data pouze část objektů populace. Může ale nastat případ, že je potřeba učinit si obraz o celé populaci. Je tedy nutné učinit tzv. výběr dat, který je zobrazen v tzv. výběrovém souboru. Metody správného usuzování z výběru na populaci

⁵ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

⁶ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

⁷ HOMOLA, V. Úvod do statistiky. [online]. URL: <<http://home1.vsb.cz/~hom50/SLBSTATS/UST/GS02.HTM>> [cit. 2010-05-18].

⁸ MARTÍNEK, J. Statistické metody v hodnocení léčby. [online]. URL: <<http://www.hpb.cz/index.php?pId=04-1-2-12>> [cit. 2010-05-18].

nám poskytuje matematická statistika. Soubor těchto metod se označuje jako statistická indukce a aplikace těchto metod se nazývá induktivní statistika.⁹

1.2. Měření a typy škál

Měřením zjišťujeme vyjádření číselné hodnoty sledované veličiny pro určitý objekt. Měřicí procedury musí mít řadu jasně definovaných vlastností, např. reprodukovatelnost, ověřitelnost, atd.

Výsledky měření se vyjadřují pomocí tzv. škály nebo stupnice. Škála vyjadřuje veškeré možné hodnoty, kterých může měřená veličina nabývat. Podle typu škály jsou definovány vztahy mezi hodnotami na škále. Rozeznáváme 4 typy škál – nominální, ordinální, intervalovou a podílovou.

¹⁰

1.2.1. Nominální škála

Nominální škála klasifikuje zkoumané objekty do vymezených tříd či kategorií. Hodnoty jsou vyjádřeny slovně. O veličinách měřených v nominální škále hovoříme jako o nominálních veličinách (např. barva očí - modrá, hnědá, černá, ...).¹¹

1.2.2. Ordinální (pořadová) škála

Ordinální škála umožňuje sledované jevy nejen třídít nebo rozlišovat, ale současně také uspořádávat – např. „je větší“, „je menší“ nebo „předchází“, „následuje“. Škála ale není schopná

⁹ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

¹⁰ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

¹¹ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

číselně vyjádřit vzdálenost mezi „větším“ a „menším“ či mezi „předcházejícím“ a „následujícím“. Veličiny měřené v ordinálních škálách se nazývají ordinální veličiny.¹²

1.2.3. Intervalová škála

Intervalová škála má definovanou jednotku měření, a proto umožňuje stanovení vzdálenosti mezi hodnotami měřené veličiny. Intervalová škála nemá jednoznačně stanovenou nulu, takže je možné počítat s rozdíly naměřených hodnot, ale ne s jejich podíly (např. teplota).¹³

1.2.4. Podílová škála

Podílová škála má jednoznačně stanovenou nulu a proto je možné počítat nejen s rozdíly, ale také s podíly hodnot. Veličiny měřené v podílových škálách označujeme jako kardinální veličiny (např. rozměry, objem a hmotnost těles, ...).¹⁴

1.3. Historie

Počátky matematické statistiky sahají do 17. a 18. století. Jako první jí začal používat k astronomickým účelům Galileo Galilei.

Již roku 1710 byl použit statistický test, kterému však nebyla dlouhou dobu věnována pozornost přírodovědci. V této době byly také položeny základy teorie pravděpodobnosti. Prvními představiteli byli P. de Fermat, B. Pascal a Ch. Huygens. Mezi další představitele patří J. Bernoulli, T. Bayes a P. S. Laplace. Bayes a Laplace se zasloužili o teorii induktivního pravděpodobnostního uvažování, která je základem dnešního bayesovského přístupu ke statistice.

¹² Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

¹³ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

¹⁴ Základní statistické pojmy. Popisná statistika. [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].

Další rozvoj teorie pravděpodobnosti a nově vznikající matematické statistiky přišel v 19. století. C. F. Gauss jako první použil metodu nejmenších čtverců a to při výpočtu dráhy planety Ceres.

V 19. století se metody matematické statistiky dostávají do podvědomí věd biologických i humanitních. A. Quételet je aplikoval na demografii, G. Fechner tyto metody použil pro psychologické experimenty a F. Galton zavedl pojem korelace a regrese, ze které vznikla v dnešní době dobře známá regresní analýza.

Ch. S. Peirce formuloval základní prvky matematické statistiky, např. testování hypotéz, což dále rozvíjí K. Pearson, který položil základy chí-kvadrát testu.

W. S. Gosset se zasadil o Studentův t-test.

Za zakladatele moderní statistiky můžeme považovat britského biologa R. Fishera. Publikoval velké množství učebnic a studijních materiálů a prohloubil celou řadu metod a pojmů, např. analýzu rozptylu, diskriminační analýzu, metodu maximální věrohodnosti atd.

Od poloviny 20. století je matematická statistika samostatným vědním oborem na pomezí čisté a aplikované matematiky. Teorií statistických odhadů se zabývali W. G. Cochran nebo L. Kish.

S vývojem výpočetní techniky nastalo zjednodušení a urychlení výpočtů matematické statistiky. Bleskový vývoj výpočetní techniky na konci 20. století umožnil vznik výpočetně náročných metod, často založených na myšlence Monte Carlo (opakované generování náhodných jevů) a dále nástup bayesovských metod, které byly do této doby limitovány jak nedostatečnou výpočetní technikou, tak i nepřítomností numerické metodologie. Podpora matematické statistiky rostla i díky rozvíjejícímu se programovému vybavení.¹⁵

¹⁵ WIKIPEDIA. *Matematická statistika*. [online]. URL: <http://cs.wikipedia.org/wiki/Matematick%C3%A1_statistika> [cit. 2010-05-04].

1.4. Explorační analýza dat

„Explorační analýza dat (*Exploratory data analysis, EDA*) je souhrn metod používaných pro průzkum dat a hledání hypotéz, které jsou vhodné k testování.“¹⁶

Explorační analýzu dat lze užít v situaci, kdy není dostatečně jasné, co vše může být výsledkem. Cílem je tedy vybrat z rozsáhlých dat maximum informací k určité problematice. Tímto se odlišuje explorační analýza od konfirmační analýzy, která testuje předem přesně formulované hypotézy. Konfirmační analýza dat odpovídá na otázky typu „Je pravda, že?“, zatímco explorační analýza odpovídá na otázky typu „Jsou zde nějaké zajímavé vztahy?“.

Explorační analýza se snaží identifikovat vlastnosti dat – vyhledává vzorce v datech, vyvozuje hypotézy o příčinách pozorovaných jevů a aspekty hodnotících modelů. Dále ověřuje předpoklady statistických metod, které je vhodné použít a v neposlední řadě poukazuje na výběr statistických nástrojů a technik.

Tato analýza je založena na užití grafických, vizuálních a numerických metod, které jsou statisticky dané (metody, které nejsou ovlivnitelné atypickými nebo extrémními hodnotami).

Explorační analýzu zavedl jako samostatný podobor statistiky John Tukey.

Metody využívané v explorační analýze dat jsou medián, modus, kvartilové rozpětí, mezikvartilové rozpětí a krabicový graf.^{17, 18, 19, 20}

¹⁶ WIKIPEDIA. *Matematická statistika*. [online]. URL: <http://cs.wikipedia.org/wiki/Matematick%C3%A1_statistika> [cit. 2010-05-04].

¹⁷ WIKIPEDIA. *Matematická statistika*. [online]. URL: <http://cs.wikipedia.org/wiki/Matematick%C3%A1_statistika> [cit. 2010-05-04].

¹⁸ HAINING, R. – WISE, S. *Exploratory Spatial Data Analysis*. [online].

URL: <<http://www.ncgia.ucsb.edu/giscc/units/u128/u128.html>> [cit.2010-05-18].

¹⁹ WIKIPEDIA. *Matematická statistika*. [online]. URL: <http://cs.wikipedia.org/wiki/Matematick%C3%A1_statistika> [cit. 2010-05-04].

²⁰ Role statistiky ve výzkumu. [online]. URL: <www.finman.gzk.cz/.../AS_mat_prednaska_role_statistiky_ve_vyzkumu.doc> [cit. 2010-05-05].

1.4.1.1. Medián

Medián je metoda, která se řadí mezi míry centrální tendence. Tato metoda rozděluje řadu výsledků (seřazených podle velikosti) na dvě poloviny, které jsou stejně početné. „*Platí, že nejméně 50 % hodnot je menších nebo rovných a nejméně 50 % hodnot je větších nebo rovných mediánu.*“²¹

Při určování mediánu dat je nutné seřadit hodnoty dat podle velikosti a poté určit hodnotu, která se nachází uprostřed seznamu. Pokud soubor obsahuje sudý počet hodnot, za medián lze považovat aritmetický průměr hodnot, které se nachází na místech $n/2$ a $n/2+1$.²²

1.4.1.2. Modus

Modus metoda vybírá ze souboru dat takovou hodnotu, která se v daném souboru vyskytuje nejčastěji (je to hodnota znaku s největší relativní četností). Např. v souboru čísel 2, 3, 5, 1, 5, 3, 7, 5 bude modus číslo 5.²³

1.4.1.3. Krabicový graf (Box plot)

Krabicový graf se používá pro grafické znázornění tvaru rozdělení, jeho střední hodnoty a variability. Střední čárka v krabicovém grafu představuje hodnotu mediánu. Hranice krabicového grafu představují 1. a 3. kvartil, oblast mezi těmito intervaly je nazývána interkvartilový interval (IQR). Body, které se nacházejí ve větší vzdálenosti než $1,5 \times \text{IQR}$ lze považovat za extrémní hodnoty. Tyto body jsou zobrazeny v podobě koleček a zobrazují možná odlehlá měření.

²¹ WIKIPEDIA. *Medián*. [online]. URL: <<http://cs.wikipedia.org/wiki/Medi%C3%A1n>> [cit. 2010-05-04].

²² WIKIPEDIA. *Medián*. [online]. URL: <<http://cs.wikipedia.org/wiki/Medi%C3%A1n>> [cit. 2010-05-04].

²³ WIKIPEDIA. *Modus*. [online]. URL: <<http://cs.wikipedia.org/wiki/Modus>> [cit. 2010-05-04].

1.4.1.4. Kvartil

Míra rozptylu hodnot kolem mediánu odděluje ze statistického souboru čtvrtiny. Rozlišuje se dolní kvartil $Q_0,25$ (tj. taková hodnota, která soubor rozděljuje tak, že 25% hodnot souboru je nižších) a horní kvartil $Q_0,75$ (tj. taková hodnota, která soubor rozděljuje tak, že 75% hodnot souboru je vyšších).²⁴

1.4.1.5. Mezikvartilové rozpětí

Pomocí rozdílu horního a dolního kvartilu lze definovat mezikvartilové rozpětí – hodnota mezikvartilového rozpětí je $Q_0,75 - Q_0,25$. Reprezentuje oblast hodnot, která má středních 50% hodnot proměnné.^{25, 26}

1.4.2. ESDA analýza

Prostorová explorační analýza dat je rozšířením explorační analýzy dat o prostorové vlastnosti dat. Toto rozšíření je možné díky přídavným technikám (detekce prostorových vzorců v datech, formulace hypotéz na základě geografických dat, vyhodnocení prostorových modelů). ESDA analýza propojuje numerické a grafické postupy s podkladovou mapou. Díky analýze je tak možné zobrazit, vyhledat nebo zvýraznit určité oblasti na mapě.

„Prostorová data mají mnoho specifických vlastností, které znesnadňují jejich analýzu a vyžadují použití odlišného souboru statistických metod, modelovacích přístupů i velmi citlivou interpretaci výsledků kvantitativních analýz. Standardní statistické metody vyvinuté pro analýzu neprostorových dat jsou tak v mnoha případech pro analýzu prostorových dat nevhodné. Za nejvýznamnější problémy či specifika analýzy prostorových dat lze považovat závislost výsledků

²⁴ WIKIPEIDA. Kvartil. [online]. URL: <<http://cs.wikipedia.org/wiki/Kvartil>> [cit. 2010-05-04].

²⁵ WIKIPEIDA. Kvartil. [online]. URL: <<http://cs.wikipedia.org/wiki/Kvartil>> [cit. 2010-05-04].

²⁶ WIKISKRIPTA. Míra variability. [online]. URL: <http://www.wikiskripta.eu/index.php/M%C3%ADry_variability> [cit. 2010-05-18].

analýz na agregaci dat neboli na způsobu vymezení prostorových jednotek ve spojení s ekologickou chybou, prostorovou autokorelací a prostorovou nestacionaritou.“²⁷

Tabulka č. 1 – Specifika analýzy prostorových dat

Aspekt	Význam	Příklad metod
Ekologická chyba (agregace dat a určenost prostorových jednotek)	vztah získaný při jednom způsobu agregace dat není univerzální – nemožnost odvození individuálního chování jedinců z analýzy agregátních dat	ekologická regrese a metoda hranic, metoda maximalizace entropie, Kingova metoda, víceúrovňové modely
Prostorová autokorelace (závislost)	závislost výskytu určitého jevu v prostoru na výskytu tohoto jevu v blízkém okolí – porušení základního předpokladu obecného lineárního modelu a řady standardních parametrických statistických testů, že jednotlivá pozorování jsou navzájem nezávislá či nekorelovaná	Moranovo I kritérium, Gearyho C kritérium, analýza LISA, obecná G statistika, lokální G statistika
Prostorová nestacionarita a heterogenita	nestabilita zkoumaných jevů a vztahů v prostoru – odlišnost funkčních vztahů mezi proměnnými v různých prostorově vymezených oblastech, nestabilita regresních funkcí	metoda regrese pohyblivých oken (moving window regression), metoda geograficky vážené regrese (GWR)

Zdroj: SPURNÁ, P. *Prostorová autokorelace – všudypřítomný jev při analýze prostorových dat?*. Sociologický časopis / Czech Sociological Review. 2008. Vol. 44. No. 4. Str. 768

Metody, které ESDA využívá, jsou smoothing, identifikace trendů a gradientů na mapě, prostorová autokorelace a detekce odlehlých prostor.²⁸

1.4.2.1. Smoothing (Vyhlazování)

Při krátkodobé predikci trendů je vhodné využít smoothing, neboli tzv. vyhlazování. Tato metoda je časově nenáročná a je založena na vyhlazování pomocí klouzavého průměru. Vhodné je

²⁷ SPURNÁ, P. *Prostorová autokorelace – všudypřítomný jev při analýze prostorových dat?*. Sociologický časopis / Czech Sociological Review. 2008. Vol. 44. No. 4. Str. 768

²⁸ HAINING, R. – WISE, S. *Exploratory Spatial Data Analysis*. [online]. URL: <<http://www.ncgia.ucsb.edu/giscc/units/u128/u128.html>> [cit.2010-05-18].

použití této metody v případě, že je mapa složena z mnoha malých oblastí, protože metoda vyhlazení odhalí přítomnosti obecných vzorců v analyzovaných datech.²⁹

1.4.2.2. Identifikace trendů a gradientů na mapě

V distribučních mapách lze identifikovat obecné trendy nebo tendence z analyzovaných hodnot.

1.4.2.3. Detekce odlehlých prostor

Je situace, kdy jednotlivé hodnoty atributů jsou extrémní z hlediska atributů v přilehlých oblastech, ne z hlediska distribučního. Atributem rozumíme grafická prostorová digitální data.

1.4.3. Prostorová autokorelace

Podstatou prostorové autokorelace je jev, ve kterém jeden jev koreluje se sebou samým v prostoru a platí, že proměnné na bližších lokalitách dosahují podobnějších hodnot, než proměnné ve vzdálenějších lokalitách. Hodnoty atributů, které spolu korelují, nejsou náhodné – jsou ovlivněny jejich vzájemnou polohou (např. v důsledku podobných přirozených podmínek, přirozené spojitosti jevů apod.).

Mezi prostorově autokorelovanými daty existuje prostorová závislost – pozitivní, negativní nebo nulová. Pozitivní prostorová autokorelace nastává, pokud se vysoké hodnoty určité proměnné shlukují v některé části oblasti a nízké hodnoty zase v jiné části oblasti. Negativní prostorová autokorelace nastává, pokud se naopak vysoké hodnoty určité proměnné nacházejí v těsné blízkosti nízkých hodnot – sledovaná proměnná má pak tendenci v prostoru vytvářet tzv.

²⁹ Úvod do analýzy časových řad. [online]. URL: <http://home1.vsb.cz/~tvr12/PUBLIKACE/2003_Analyza_CR.pdf> [cit. 2010-05-04].

šachovnicový vzorek. Pokud mezi uspořádáním neexistuje žádný vztah, nazýváme tento jev nulovou prostorovou autokorelací a jde tedy o čistě náhodné uspořádání hodnot v prostoru.³⁰

1.4.3.1. Určení vztahů sousedství

Prostorovou autokorelací lze vyjádřit stupeň podobnosti atributů mezi danou plochou a sousedními plochami. Je proto nutné definovat vztahy sousedství a to pomocí matic prostorových vah (Spatial Weights Matrix). Každá prostorová jednotka je v matici prezentována jedním řádkem a sloupcem. Jednotlivé buňky matice pak vyjadřují hodnotu, která určuje prostorový vztah mezi danými jednotkami. Hodnota buněk je ovlivněna způsobem definování sousedství.³¹

1.4.3.2. Způsoby definování sousedství

Vztahy sousedství jednotlivých ploch je možno definovat celou řadou způsobů. Běžně se využívá faktor vzdálenosti mezi jednotlivými objekty, protože intenzita vztahu je větší mezi blízkými jednotkami a naopak. Určit vzdálenost je možné pomocí různých typů matic – např. BCM matice nebo stochastická matice.

Binární BCM matice (Binary Connectivity Matrix) je symetrická matice, kde suma řádku vyjadřuje informaci o počtu sousedů dané jednotky. Matice nabývá hodnot 1 nebo 0, a to 1 pro jednotky sousedící a 0 pro jednotky nesousedící.

Matice stochastická zahrnuje standardizované váhy jednotlivých řádků. Jedničky jsou nahrazeny vahou, která je určena poměrem mezi hodnotou a počtem sousedů, čímž dochází k zohlednění počtu sousedů.³²

³⁰ LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta. Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.

³¹ LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta. Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.

³² LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta. Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.

1.4.3.3. Míry prostorové autokorelace

Výše uvedené způsoby stanovení vztahu sousedství slouží k definování měr prostorové autokorelace. Rozlišujeme globální míry prostorové autokorelace (jedna hodnota pro celou studovanou oblast) a lokální míry prostorové autokorelace (různé hodnoty v rámci studované oblasti) – přesněji řečeno: globální míru lze vypočítat ze všech hodnot, lokální míru pro každou samostatně vždy z předem definovaného okolí – lépe tedy vyjadřuje její prostorovou variabilitu. Globální míry prostorové autokorelace jsou Moranův a Gearyho index, u lokálních měr je to pak LISA (Local Indicators of Spatial Association).³³

1.4.3.3.1. Globální míry

Moranův index I

Index se vypočte podle vzorce:

$$I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2}$$

kde x_i je hodnota proměnné v ploše i , w_{ij} jsou váhy, W matice vah.

Index nabývá hodnot od -1 do 1, přičemž hodnota -1 odpovídá negativní prostorové autokorelaci a hodnota +1 vyjadřuje pozitivní prostorovou autokorelaci. Hodnota 0 vyjadřuje náhodné prostorové uspořádání. V případě, že je index I kladný, převažují v analyzované oblasti sousedé s podobnými hodnotami.

³³ LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta. Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.

Obrázek č. 1 – Moranův diagram

vážená hodnota proměnné v blízkých jednotkách	nízká – vysoká <i>negativní prostorová autokorelace</i>	vysoká – vysoká <i>pozitivní prostorová autokorelace</i>
	nízká – nízká <i>pozitivní prostorová autokorelace</i>	vysoká – nízká <i>negativní prostorová autokorelace</i>
hodnota proměnné v prostorové jednotce		

Zdroj: NETRDOVÁ, P. – NOSEK, V. *Přístupy k měření významu geografického rozměru společenských nerovnoměrností*. Geografie – Sborník české geografické společnosti. Rok 2009. Číslo 1. Ročník 114. Strana 58.

Po provedení výpočtu se ověřuje, do jaké míry je odlišná vypočtená hodnota od hodnoty očekávané. Ze zjištěného rozdílu se odvozuje hodnota z-skóre pomocí míry rozptylu. Pokud je velikost z-skóre menší než -1,96 nebo větší než 1,96 je hodnota indexu statisticky významná na hladině $\alpha = 0,05$.³⁴

„Měření prostorové autokorelace pomocí Moranova indexu lze považovat za globální analýzu, neboť jedna výsledná hodnota ukazuje na míru prostorové autokorelace či shlukování určité proměnné v celém zkoumaném území, přičemž dochází k zprůměrování možných výrazných územních rozdílů.“³⁵

³⁴ LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta. Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.

³⁵ Zdroj: NETRDOVÁ, P. – NOSEK, V. *Přístupy k měření významu geografického rozměru společenských nerovnoměrností*. Geografie – Sborník české geografické společnosti. Rok 2009. Číslo 1. Ročník 114. Strana 58.

Gearyho index C

Index se vypočte podle vzorce:

$$C = \frac{(n-1) \sum \sum w_{ij} (x_i - x_j)^2}{2W \sum (x_i - \bar{x})^2}$$

kde x_i je hodnota proměnné v ploše i , w_{ij} jsou váhy, W matice vah.

Index nabývá hodnot od 0 do 2, přičemž 0 odpovídá pozitivní autokorelaci, 1 znamená nulovou prostorovou autokorelaci a 2 vyjadřuje autokorelaci negativní.

Gearyho index porovnává hodnoty atributů přímo mezi sebou, zatímco výše zmíněný Moranův index porovnával hodnoty sousedních ploch pomocí odchylek od průměru.

Po provedení výpočtu indexu je opět nutné vypočítat hodnotu z-skóre pro prokázání statisticky významného rozdílu. Hodnoty z-skóre jsou založené na rozdílu pozorovaných a očekávaných hodnot.³⁶

³⁶ LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta. Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.

1.4.3.3.2. Lokální míry

LISA (Local Indicators of Spatial Association)

Lokální indikátory vyjadřují proměnlivost prostorové autokorelace v rámci studované (lokální) oblasti. Jedná se o modifikované verze Moranova a Gearyho indexu. Analýza LISA byla vyvinuta Lucem Anselinem.

Vzorec pro výpočet **lokálního Moranova indexu** pro jakoukoli prostorovou jednotku i je:

$$I_i = z_i \sum_j w_{ij} z_j$$

kde $z[i]$ a $z[j]$ jsou odchylky od průměru.

Pokud index vykazuje nízké hodnoty, znamená to kumulaci odlišných hodnot atributů, zatímco vysoké hodnoty indexu znamenají kumulaci podobných hodnot v sousedních plochách. Zjištěné hodnoty je nutné porovnat s očekávanými hodnotami a zjistit, jaké je statistická významnost jejich rozdílu pomocí z-skóre.

Lokální verze Gearyho indexu se vypočte podle vzorce:

$$C_i = \sum_j w_{ij} (z_i - z_j)^2$$

Interpretace indexu je obdobná s interpretací globální verze. Kumulace odlišných hodnot atributů vede k vysokým hodnotám indexu, zatímco nízké hodnoty indexu vyjadřují kumulaci podobných hodnot.³⁷

³⁷ LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta, Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.

1.5. GIS, GeoDA

1.5.1. GIS

Geografický informační systém (GIS) integruje hardware, software a data pro pořizování, správu, analýzu a zobrazování všech forem geografických informací.

GIS umožňuje zobrazit, pochopit a interpretovat data mnoha způsoby, které odhalí a vysvětlí vztahy, vzory a trendy formou map, grafů apod.

GIS napomáhá zodpovědět otázky a řešit problémy při rozboru dat, a to snadno pochopitelným způsobem, který je možno jednoduše sdílet s ostatními uživateli.

GIS technologie mohou být integrovány do jakéhokoliv podnikového informačního systému.³⁸

1.5.2. GeoDa

GeoDa centrum vyvíjí a implementuje metody geoprostorové analýzy a aplikuje je v sociálních a environmentálních vědách. Rozsah aktivit zahrnuje tvorbu nových metod pro analýzu prostorových dat, geovizualizace, geosimulace a prostorové modelování procesů k využití těchto technik, jako jsou softwarové nástroje a podpůrné materiály pro stále se rozvíjející celosvětovou skupinu uživatelů.³⁹

GeoDa je volně dostupný software vyvinutý Lucem Anselinem. Je vhodný k provádění prostorové analýzy a také analýzy prostorové autokorelace. Oficiální internetové stránky tohoto programu jsou na adrese <https://www.geoda.uiuc.edu/>

³⁸ What is GIS? [online]. URL: <<http://www.gis.com/content/what-gis>> [cit. 2010-05-18].

³⁹ ARIZONA STATE UNIVERSITY. *GeoDa Center for Geospatial Analysis and Computation at ASU*. [online]. URL: <<http://sharedresources.asu.edu/node/2736>> [cit. 2010-05-18].

1.5.2.1. Luc Anselin

Luc Anselin je celosvětově proslulý díky statistickému bádání a vývoji softwaru GeoDA. Pochází z Belgie, kde získal magisterský titul v oboru ekonomie a ekonometrie na bruselské universitě. V současné době působí v USA jako ředitel University geografických věd (ASU School of Geographical Sciences), kterou také založil. Dále založil a řídí GeoDA centrum pro geoprostorové analýzy a výpočty. V roce 2004 byl jmenován členem mezinárodní vědecké asociace, v roce 2005 získal ocenění Walter Isard Prize a v roce 2006 William Alonso Memorial Prize. V roce 2008 byl zvolen do Národní akademie věd.⁴⁰

⁴⁰ ARIZONA STATE UNIVERSITY. *Luc Anselin*. [online]. URL <<https://webapp4.asu.edu/directory/person/1103870>> [cit. 2010-05-04].

2. Demografická data

Demografie zkoumá lidskou populaci, konkrétně demografickou (populační) reprodukci, její podmínky a prognózy vývoje do budoucna. S demografií se také prolíná geografie obyvatelstva, vědní disciplína, která zkoumá prostorové umístění, strukturu a migraci obyvatelstva.

Rozeznáváme několik typů demografických oborů:

- **kvantitativní demografie** – zkoumá kvantitativní stránku stavu a vývoje lidských populací, - u výčtů se dávají čárky nebo středníky
- **popisná demografie** – popisuje stav a vývoj populace, založena na empirických datech,
- **teoretická demografie** – zabývá se obecným studiem populačních problémů a dále formálními vztahy mezi různými demografickými jevy,
- **matematická demografie** – podobor teoretické demografie, aplikuje matematické modely na demografická data (např. vytváření populačních prognóz),
- **kvalitativní demografie** – hlavním oborem zkoumání je rozložení kvalitativních znaků v populaci (intelektuální, fyzické, sociální znaky atd.),
- **ekologická a sociální demografie** – studium vztahů mezi populačními jevy, jevy ekonomickými a sociálními.⁴¹
- **regionální demografie** – zabývá se vývojem populací v jednotlivých regionech, studuje demografické procesy z hlediska regionu (regionálních podobností a rozdílů). Regionální demografie je vhodná k aplikaci prostorových metod.⁴²

⁴¹ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

⁴² DEMOGRAFICKÉ INFORMACE, ANALÝZY A KOMENTÁŘE. *Vztah subdisciplín k vědním disciplínám mimo demografie*. [online]. URL: < http://www.demografie.info/?cz_discipliny=> [cit. 2010-05-18].

2.1. Demografické jevy

Demografie úzce spolupracuje s historií, sociologií, filosofií a také statistikou. Statistické údaje jsou totiž klíčovým zdrojem pro demografické údaje. Demografické údaje lze rozdělit na **stavové** (stav populace k určitému časovému datu) a **pohybové** (události nebo jevy, které nastanou během určitého časového úseku, zpravidla jednoho roku).

Zdrojem dat demografických údajů jsou různé metody sběru dat:

- **sčítání lidu**
- **soupisy**

Sčítání lidu má své kořeny již v dávné historii, záznamy o primitivních registrech obyvatelstva jsou již z roku 3800 př. n. l. V současné době má sčítání lidu svou přesnou podobu, je řízeno speciálním zákonem a jedná se o velice důležitý proces, zásadně důležitý je proces přípravy i analýzy. Určen tzv. „rozhodný okamžik“, což je okamžik, ke kterému se zaznamenají platná data (většinou se jedná o půlnoc k určitému datu). Poslední sčítání lidu proběhlo v České republice v roce 2001.

Soupisy obyvatelstva jsou akce menšího a jednoduššího charakteru.

- **běžná evidence přirozené měny**

Jedná se o sledování porodnosti a úmrtnosti. Toto sledování provádí matriky.

- **běžná evidence migrací**

Běžná evidence migrace sleduje pohyb obyvatelstva, změnu bydliště dané (migrující) osoby. Migrace se dělí na **vnitřní migraci** (změna bydliště uvnitř sledované oblasti) a **migraci vnější** (změna bydliště přes hranice sledované oblasti). Smyslem evidence je přehled o počtu obyvatel v jednotlivých oblastech.

- **populační registr**

Populační registr provádí neustálou evidenci obyvatelstva, identifikačním údajem je rodné číslo každého občana. V případě správného fungování by mohl nahradit sčítání lidu. V České republice provádí registraci populace Centrální registr občanů (CRO).

- **zvláštní šetření**

Zvláštní šetření, tzv. výběrové šetření, shromažďuje údaje o vybraném souboru obyvatel. Zvláštní šetření se zaměřují na informace, které je potřebné zjistit pouze u některého souboru obyvatel nebo na postoje lidí k určité problematice.

Demografické události musí být jednoznačně přesně vymezeny (**věcně, místně, časově**), aby mohly být statisticky sledovány. Časové vymezení zjišťuje, jaké události proběhly během určitého časového období. Místní vymezení zase sleduje, v jakém místě události proběhly.

Demografická analýza srovnává demografické jevy a procesy, avšak v jejich absolutních hodnotách. K hlubšímu poznání těchto jevů je nutné vypočítat z nich analytická data – tzv. **demografické ukazatele**.⁴³

2.1.1. Přehled a členění demografických ukazatelů

Demografická data lze rozdělit na **základní** a **analytická**. Základní data jsou prosté řady absolutních čísel a vznikají z výsledků sčítání obyvatel či výsledků běžných evidencí. Analytická data vznikají po zpracování základních dat.

Demografické ukazatele se dělí do tří kategorií – poměrná čísla extenzivní, poměrná čísla intenzivní a indexy.

⁴³ KLUFŮVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

- **Poměrná čísla extenzivní** – jedná se o podíl nebo proporce celku v procentech. Tento ukazatel se vypočte vydělením dvou stejnorodých ukazatelů, které proběhly ve stejném čase a na shodném území. Používá se k vyjádření struktury.
- **Poměrná čísla intenzivní** – tento ukazatel se vypočte vydělením různorodých údajů, z nichž jmenovatel je nositelem události nebo jevu, který je vyjádřen v čitateli.
- **Poměrná čísla srovnávací, tzv. indexy** – index lze vypočítat vydělením dvou absolutních čísel, které proběhly v odlišném čase nebo v odlišném území.

Další dělení ukazatelů je pak **celkové / specifické** (podle toho, zda vyjadřují ukazatel za celou populaci nebo její část), **definitivní / předběžné** (vypočteny na základě neúplných dat), **hrubé / srovnávací**.⁴⁴

2.2. Analýza struktury obyvatelstva

Demografická statika zkoumá okamžikový stav populace, konkrétně její **velikost** a **strukturu**. **Demografická dynamika** zkoumá vše, co je spojeno s procesem **reprodukce** (plodnost, úmrtnost, sňatečnost, rozvodovost). Demografická statika a demografická dynamika pak spolu tvoří demografickou statistiku.

Struktura obyvatelstva se dělí podle demografických, sociálních, geografických či ekonomických kategorií. Základní kritéria pro sledování struktury jsou pohlaví a věk, což jsou jednoznačné objektivní charakteristiky jedince. Dalšími kritérii jsou rodinný stav, národnost, vzdělání, sociální skupina a náboženské vyznání.⁴⁵

⁴⁴ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

⁴⁵ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

2.2.1. Struktura obyvatelstva podle pohlaví

Struktura obyvatelstva podle pohlaví patří k základním demografickým charakteristikám. Nejčastěji se vyjadřuje pomocí indexů **maskulinity** (podíl mužů v populaci) a **feminity** (podíl žen v populaci) a dále **index maskulinity** (poměr mužů a žen v populaci vyjádřený na 100 nebo 1000 žen).

Index maskulinity

$$I_M = \frac{M}{F} * C$$

Index feminity

$$I_F = \frac{F}{M} * C$$

Struktura populace podle pohlaví je ovlivněna třemi procesy – **sekundárním indexem maskulinity** (tzn., že počet narozených chlapců je větší než počet narozených dívek), dále **mužskou nadúmrtností a migrací**.⁴⁶

2.2.1.1. Sekundární index maskulinity – sima

V každé populaci si rodí více chlapců než dívek. Index sima říká, kolik narozených chlapců připadá na 1000 narozených dívek, průměrně je to 1040 – 1070.⁴⁷

⁴⁶ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

⁴⁷ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

2.2.1.2. Mužská nadúmrtnost

Úmrtnost mužů a žen v průběhu celého jejich života je velice specifická. Zatímco mezi novorozenci převažují chlapci a tato převaha trvá zhruba do 40 let věku, okolo 45 let věku se poměr četnosti pohlaví (tzv. věk pohlavní rovnováhy) vyrovnává a ve starších věkových skupinách převažuje ženské pohlaví. Tento jev se nazývá mužská nadúmrtnost – muži všech věkových kategorií mají vyšší intenzitu úmrtnosti než ženy.⁴⁸

2.2.1.3. Migrace

Migrace je velice proměnlivá a významně závisí na věku migrujících obyvatel. Migrace může mít vliv na poměr počtu mužů a žen v populaci zejména při analýze malého území nebo regionu. Např. lokality s převahou pracovních příležitostí pro muže (těžký průmysl atd.) zapříčiňují disproporce mezi počtem mužů a žen.⁴⁹

2.2.2. Struktura obyvatelstva podle věku

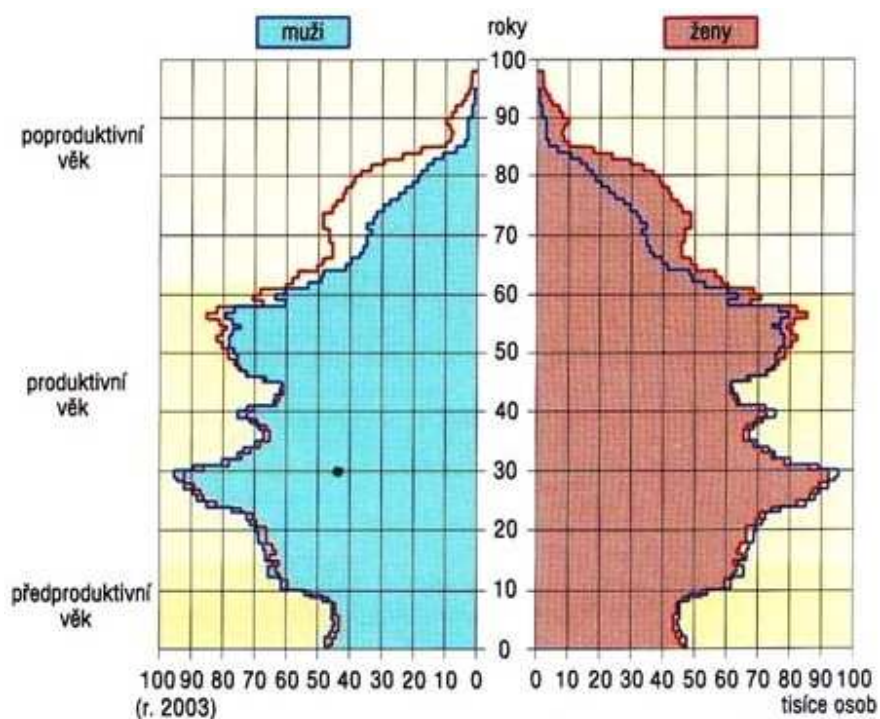
Věk je další základní, velice objektivní a jednoznačnou, charakteristikou jedince. Nejvýstižnější charakteristikou demografické struktury je grafické zobrazení, tzv. **strom života** či **věková pyramida**. Strom života je tzv. polygon a vyjadřuje relativní hodnoty ukazatelů, zatímco věková pyramida je tzv. dvojitý histogram, který vyjadřuje absolutní hodnoty ukazatelů.⁵⁰

⁴⁸ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

⁴⁹ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

⁵⁰ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

Obrázek č. 2 – Věková pyramida



Zdroj: Geografie obyvatelstva a sídel – Demografie. [online]

URL: <http://dsvoboda.eu/obrazky/vyuka/kvinta/socioekonomicka_%20sfera/obyvatelstvo/vek_pyramida.jpg> [cit. 2010-05-07]

2.2.2.1. Biologické generace

Populaci lze rozčlenit z hlediska biologického do 3 základních generací:

- I. Biologická generace - **předreproduktivní generace**, 0 – 14 let (tzv. dětská složka)
- II. Biologická generace - **reproduktivní generace**, 15 – 49 let (tzv. složka ekonomicky aktivních).
- III. Biologická generace - **poreproduktivní generace**, 50+ (tzv. Složka osob v důchodovém věku) ⁵¹

⁵¹ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

2.2.2.2. Typy populace

Biologická generace č. II tvoří v každé populaci zhruba 50%. Podle vztahu biologických generací I. a III. rozdělujeme populace do tří typů:

- **Progresivní typ**

I. biologická generace převažuje nad III. biologickou generací. V této populaci roste počet narozených dětí a celkově se zvětšuje podíl obyvatel v dětském (předreproduktivním) věku. Tento typ populace se vyskytuje zejména u rozvojových států nebo u prehistorických a historických populací, je zde vysoký podíl dětí a nízká naděje na dožití.

- **Stacionární typ**

V tomto typu populace je stejný podíl I. a III. generace. Jedná se o stagnující populace, kde je stabilní počet narozených i zemřelých.

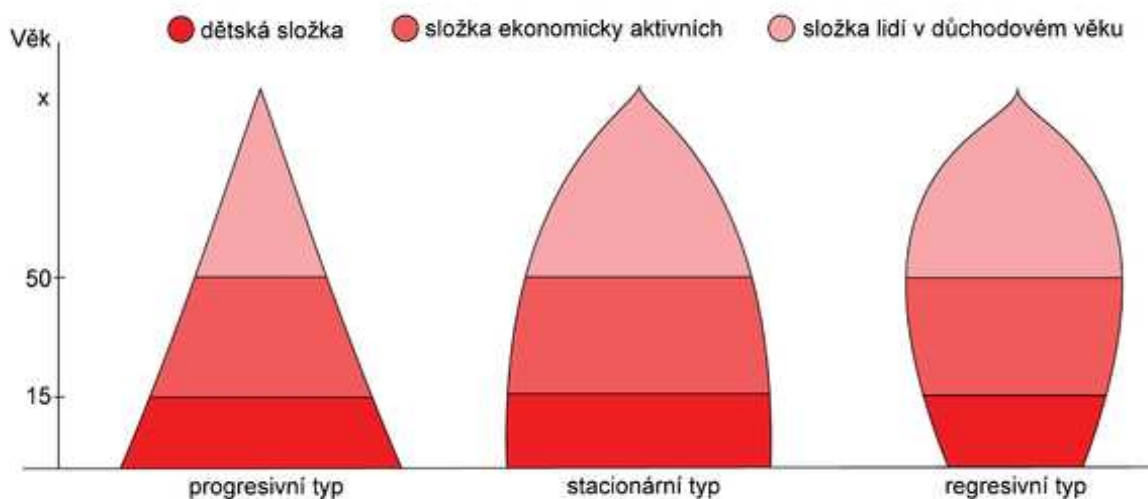
- **Regresivní typ**

III. biologická generace je početnější než I. biologická generace a dochází tak k ubývání obyvatelstva. Do této kategorie patří Česká republika a většina vyspělých evropských států.

Každá z populací postupně prochází všemi třemi fázemi populačního typu a skončí u regresivního typu populace. Tento proces se nazývá **demografické stárnutí**.⁵²

⁵² KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

Obrázek č. 3 – Populační typy



Zdroj: http://upload.wikimedia.org/wikipedia/commons/thumb/b/b9/Typy_vekovych_pyramid.png/600px-Typy_vekovych_pyramid.png

2.2.2.3. Index stárí

Index stárí vyjadřuje vztah mezi I. biologickou složkou a III. biologickou složkou populace. Index stárí je využíván při mezinárodním srovnávání věkových struktur a stárnutí populací a je označován také jako Sauvyho index.⁵³

$$is = \frac{\text{III. biologická generace}}{\text{I. biologická generace}}$$

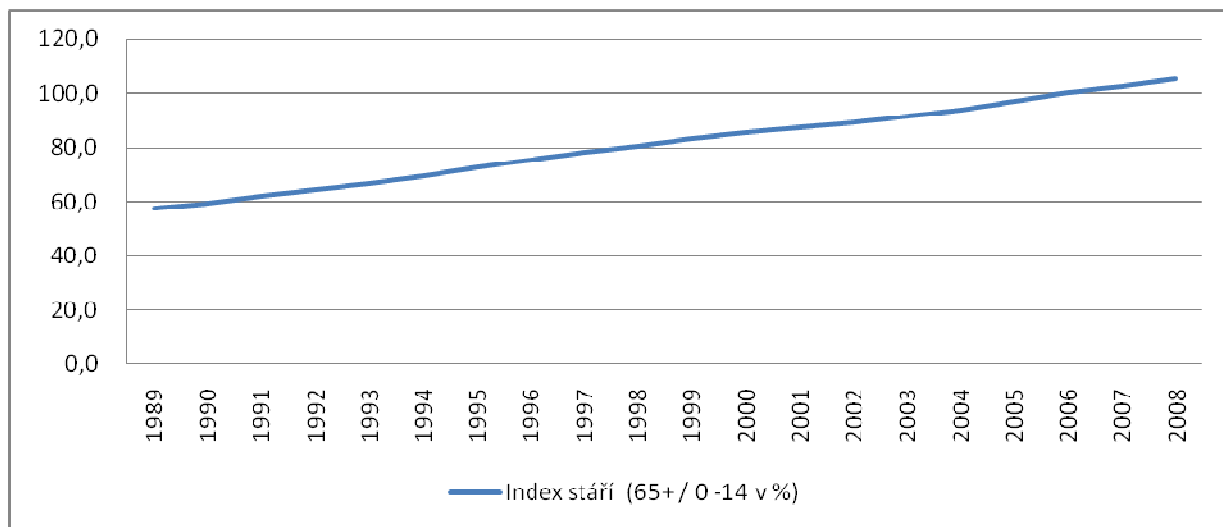
Populace progresivního typu nabývají indexu stárí hodnoty okolo 25, populace stacionárního typu hodnoty okolo 87 a populace regresivního typu hodnoty okolo 150.

V České republice se index stárí v posledních letech razantně zvyšuje. Zatímco v roce 1993 připadalo na 100 dětí 67 osob nad 65 let, v roce 2000 to bylo již 86 osob starších 65 let na 100 dětí a v roce 2006 se poměr otočil a počet osob nad 65 let převážil počet dětí. Lze to odůvodnit především neustále se snižující porodností (výjimku tvoří roky 2006 – 2008, kdy se porodnost

⁵³ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

dočasně zvýšila, neboť se do fáze odložené plodnosti dostaly silné populační ročníky 70. let 20. stol.) a stále se prodlužující střední délkou života, která je ovlivněna zejména zvyšující se úrovní lékařské péče. Očekává se, že do roku 2030 vzroste index stáří na hodnotu 243, resp. 296 do roku 2050.⁵⁴

Graf č. 1 – Vývoj indexu stáří v České republice v letech 1989 - 2008



Zdroj: Vlastní grafmzpracovaný na základě dat dostupných na: Český statistický úřad: Vybrané demografické údaje v České republice [online]. URL: <http://www.czso.cz/cz/cr_1989_ts/0101.pdf> [cit. 2010-05-12]

⁵⁴ DEMOGRAFICKÉ INFORMACE, ANALÝZY A KOMENTÁŘE. Stárnutí - vývoj. [online]. URL: <http://www.demografie.info/?cz_demstarnutivvyvoj> [cit. 2010-05-18].

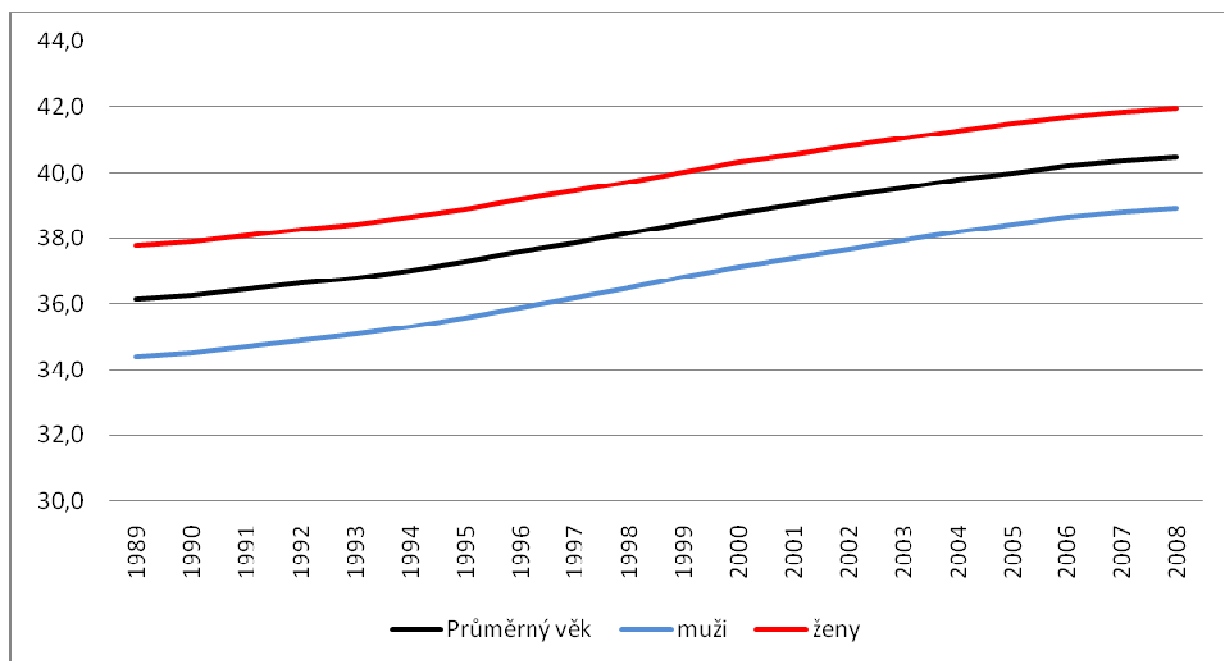
2.2.2.4. Průměrný věk

Průměrný věk populace je dalším jednoduchým ukazatelem, pomocí kterého lze vyjádřit stáří populace. Jedná se o vážený průměr počtu let prožitých příslušníky dané populace (průměr věků všech obyvatel v populaci).⁵⁵

$$\bar{x} = \frac{\sum_{x=0}^{\omega-1} (x + 0,5) \bar{S}_x}{\sum_{x=0}^{\omega-1} \bar{S}_x}$$

V České republice se průměrný věk pohybuje okolo 40 let. Analýzy věkové struktury jsou velice důležité z hlediska státní politiky v oblasti školství a dětské péče (podíl dětské populace), dále také v oblasti sociální a důchodové péče (podíl důchodců v populaci).

Graf č. 2 – Průměrný věk v České republice v letech 1989 - 2008



Zdroj: Vlastní graf zpracovaný na základě dat dostupných na: Český statistický úřad: Vybrané demografické údaje v České republice [online].
URL: <http://www.czso.cz/cz/cr_1989_ts/0101.pdf> [cit. 2010-05-14].

⁵⁵ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

Hodnota průměrného věku vzrostla v analyzovaném období z původních 36 let na 40,5 let. Ženy vykazují vyšší hodnoty průměrného věku, v současné době 42 let, zatímco muži nižší hodnoty průměrného věku, v současné době 38,9 let.

2.2.3. Ekonomická a sociální struktura

Ekonomická struktura obyvatelstva vyjadřuje, do jaké míry jsou obyvatelé dané populace ekonomicky aktivní.

*„**Ekonomicky aktivní obyvatelstvo (EAO)** jsou ti, kteří přispívají svou činností k hospodářskému výsledku společnosti (zaměstnaní lidé) a ti, kteří museli z nějakého důvodu touto svou činnost pozastavit – tzv. zdržení od povolání (nezaměstnaní, vojáci v základní vojenské službě, vězni a ženy na mateřské dovolené).“⁵⁶*

Podíl ekonomicky aktivních osob z celkového počtu (z celé populace) charakterizuje úroveň ekonomické aktivity.

2.2.3.1. Ekonomické generace

- I. Ekonomická generace – **předreproduktivní**, 0 – 19 let,
- II. Ekonomická generace – **produktivní**, 20 – 64 let,
- III. Ekonomická generace – **poproduktivní**, 65 let a více

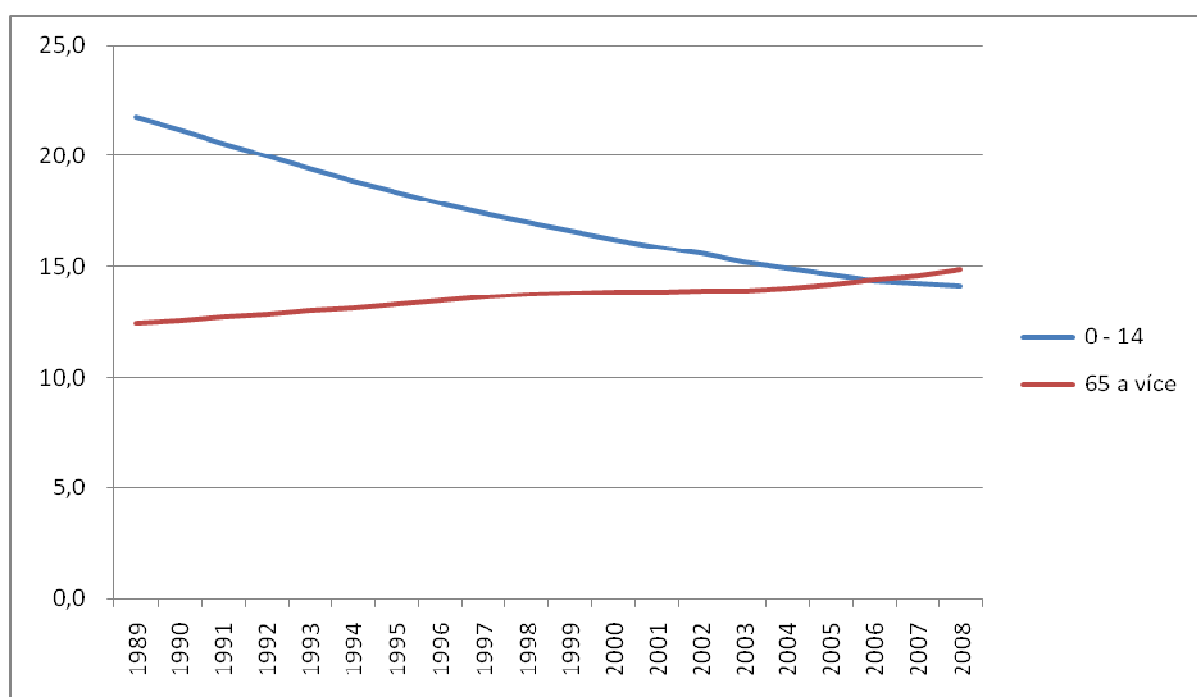
2.2.3.2. Podíl obyvatel nad 65 let

Analýza podílu obyvatel nad 65 let sleduje poproduktivní složku obyvatelstva – tj. obyvatele v důchodovém věku. U této věkové skupiny je často spekulovaná věková hranice, od které se jedná o poproduktivní složku. V České republice to bylo donedávna 60 let pro muže a 55 pro

⁵⁶ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3. Strana 46.

ženy (tedy věk odchodu do důchodu). V České republice je věk odchodu do důchodu v současné době nižší než 65 let, ale ve většině evropských zemí se jedná o věkovou hranici 65 let u mužů i žen a i v České republice postupně vzroste. Navíc existuje velké množství osob, které odchod do důchodu odkládají nebo pracují i v důchodovém věku. Proto se v demografických analýzách sjednotila věková hranice na 65 let.⁵⁷

Graf č. 3 – Podíl obyvatel ve věkové skupině 0-14 a 65 a více let v České republice v letech 1989 - 2008



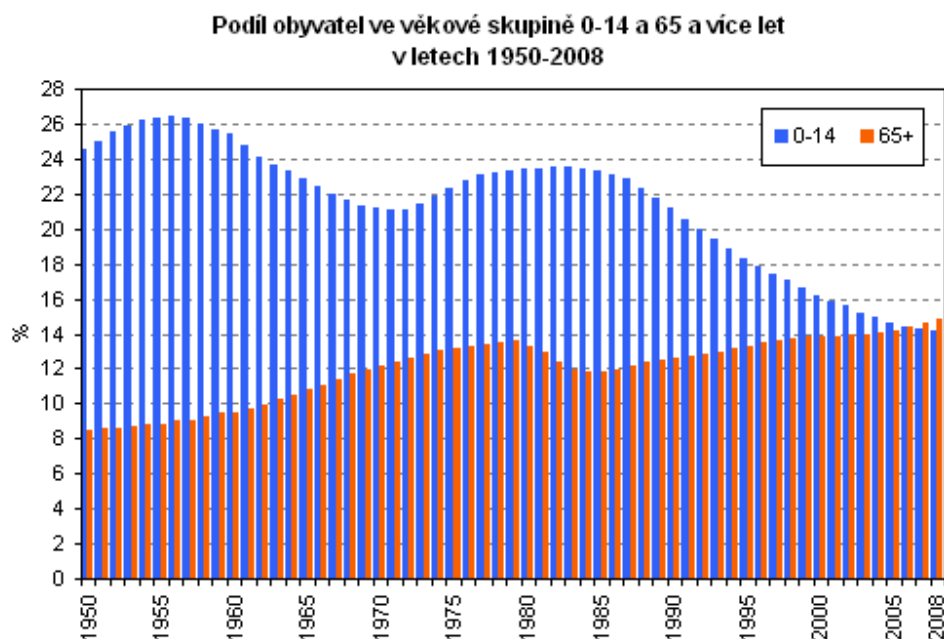
Zdroj: Vlastní grafmzpracovaný na základě dat dostupných na: Český statistický úřad: Vybrané demografické údaje v České republice [online]. URL: <http://www.czso.cz/cz/cr_1989_ts/0101.pdf> [cit. 2010-05-14].

Zatímco v roce 1989 můžeme pozorovat výraznou převahu dětské populace (22% dětská populace, 12,5 % populace nad 65 let), v následujících letech můžeme pozorovat výrazné přibližování těchto dvou hodnot. Zvrat nastal v roce 2006, kdy obě hodnoty dosáhly stejného

⁵⁷ KLUFVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

procenta a to 14,4%. Od té doby podíl dětí v populaci stále klesá a naopak podíl osob od 65 let výše narůstá.

Graf č. 4 – Podíl obyvatel ve věkové skupině 0-14 a 65 a více let v České republice v letech 1950 - 2008



Zdroj: Český statistický úřad: Podíl ve věkové skupině 0-14 a 65 a více let v letech 1950-2008 [online].

URL: <http://www.czso.cz/csu/redakce.nsf/i/podil_obyvatel_ve_vekove_skupine_0_14_a_65_a_vice let_v letech_1950_2008> [cit.2010-05-14].

2.2.4. Celkové charakteristiky reprodukce

Na populaci lze pohlížet jako na **uzavřenou** a relativně izolovanou, kde nově narození jedinci jsou produktem žijících členů. V tomto případě se reprodukce obyvatel označuje jako **přirozená reprodukce**. Opačným pohledem je zahrnutí migrace k nahlížení na populaci. V tomto případě se pak jedná o **otevřenou populaci**, kde dochází k **populačnímu vývoji**.⁵⁸

⁵⁸ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

2.2.4.1. Celkový přírůstek

Celkový přírůstek vyjadřuje celkový pohyb obyvatelstva (výsledek přirozeného i mechanického pohybu obyvatelstva).⁵⁹

Celkový přírůstek = porodnost – úmrtnost + imigrace - emigrace

2.2.4.2. Přirozený přírůstek (PP)

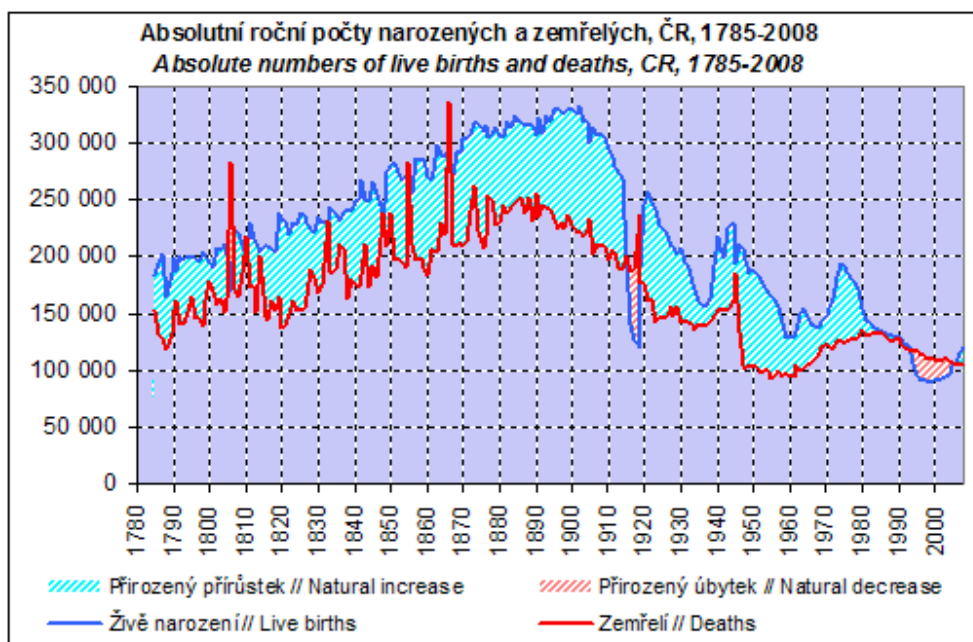
Přirozený přírůstek vyjadřuje rozdíl mezi živě narozenými a zemřelými v daném časovém období. Vykazuje-li tento ukazatel záporné hodnoty, jedná se o přirozený úbytek. Tato hodnota se vyjadřuje v promilích.

Roční světový přírůstek se pohybuje okolo 70 mil lidí. Velký nárůst je patrný v 19. století díky rozvoji vědy a lékařství. Přirozené úbytky jsou pak zaznamenávány v současné době u vyspělých států a i prognózy do budoucna předpovídají přirozené úbytky.⁶⁰

⁵⁹ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

⁶⁰ KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.

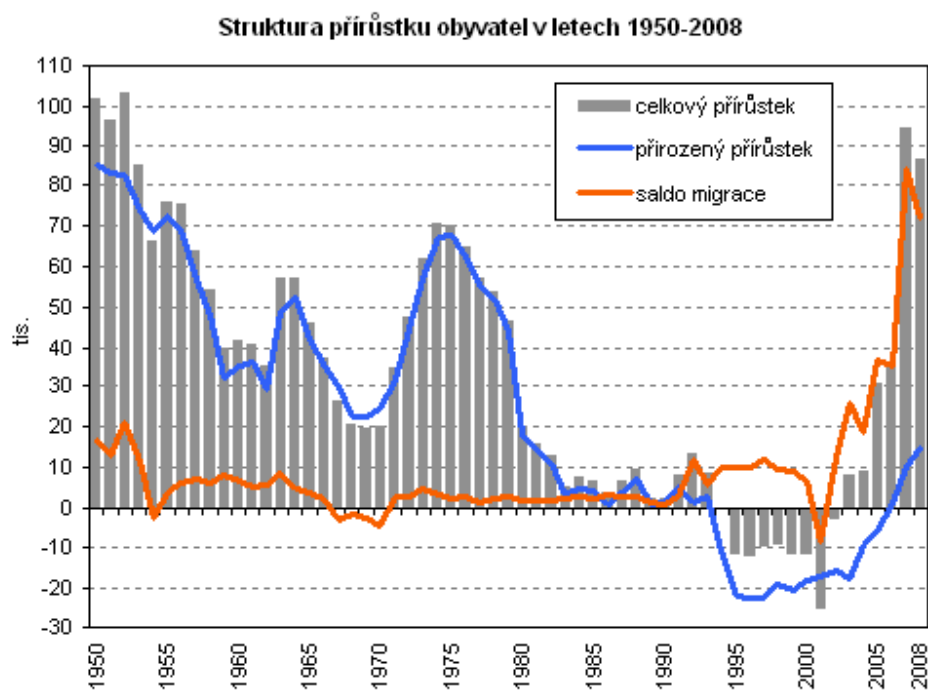
Graf č. 5 – Struktura přirozeného přírůstku v České republice v letech 1780 - 2000



Zdroj: Český statistický úřad: Roční časové řady. [online]. URL: <http://www.czso.cz/csu/redakce.nsf/i/obyvatelstvo_hu> [cit. 2010-05-14].

Jak je patrné z grafu, v průběhu historie se v grafu vyskytuje po celou dobu přirozený přírůstek, se dvěma výjimkami. První z nich je patrná v letech přibližně 1915 – 1920, což je období první světové války. Po roce 1920 je ale opět patrný obrovský přirozený přírůstek. Druhá je zřejmá před rokem 2000 a i do budoucna lze očekávat trend přirozeného úbytku. Z grafu je také patrné, že se na celkovém přírůstku bude stále významněji podílet migrace.

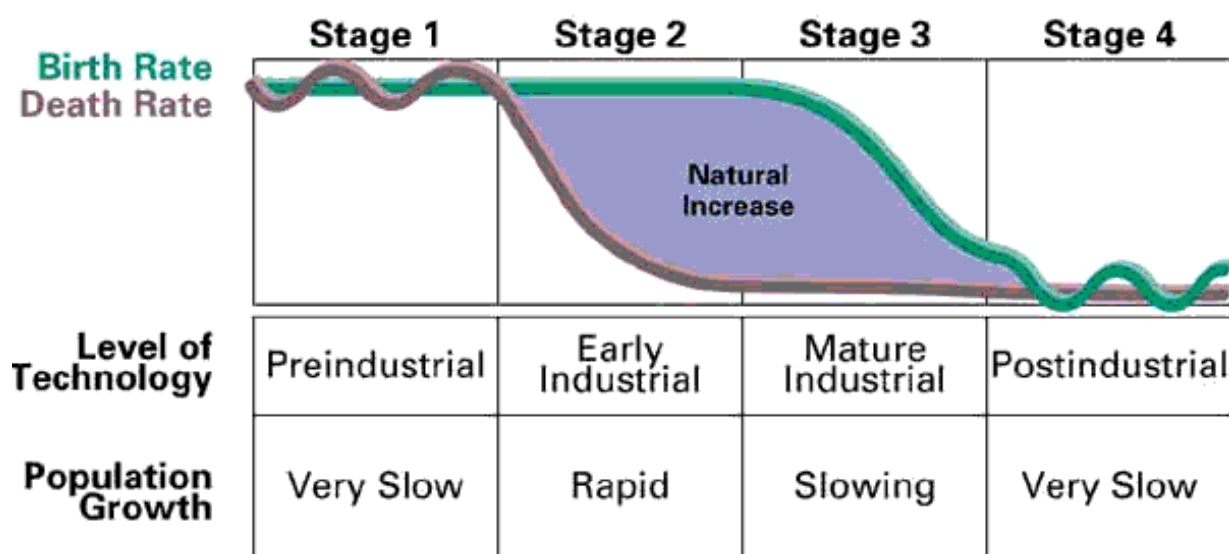
Graf č. 6 – Struktura přírůstku obyvatel v České republice v letech 1950 - 2008



Zdroj: Český statistický úřad: Struktura přírůstku obyvatel v letech 1950 – 2008[online]. URL: http://www.czso.cz/csu/redakce.nsf/i/struktura_pristunku_obyvatel_v letech_1950_2008 [cit. 2010-05-14].

S přirozeným přírůstkem, resp. úbytkem souvisí také **model demografického cyklu**, který rozděluje demografický vývoj na 4 fáze.

Obrázek č. 4 – Model demografického cyklu



Zdroj: Global Sociology [online]. URL: <<http://globalsoc.files.wordpress.com/2008/04/demographic-transition-theory.png>> [cit. 2010-05-08].

V první fázi země vykazuje vysokou porodnost i úmrtnost, celkový přirozený přírůstek je tedy relativně malý. Jedná se o chudé, rozvojové země, s nízkou úrovní lékařské péče. Ve druhé fázi je porodnost stále vysoká, ale úmrtnost se výrazně snižuje díky zlepšující se lékařské péči. Délka života obyvatel dané populace se zvyšuje a celkový přirozený přírůstek je velký. Ve třetí fázi se úmrtnost ustálí na nízkých hodnotách a porodnost klesá výrazně. Celkový přirozený přírůstek se začíná snižovat, ale je stále kladný. Ve čtvrté, závěrečné fázi, demografické revoluce je úmrtnost ustálena na velmi nízkých hodnotách a porodnost kolísá kolem nízkých hodnot a populace začíná trpět demografickým stárnutím. Přirozený přírůstek je velmi malý, nejvyspělejší země vykazují přirozený úbytek.

3. Analýza demografických dat pomocí explorační analýzy

3.1. Příprava dat pro GeoDa

GeoDa je statistický program umožňující analýzu lokalit prostorových dat. GeoDa nabízí sadu nástrojů pro prostorovou explorační analýzu dat. Jde o samostatný program, který je volně přístupný na internetu (tzv. freeware), který funguje pod systémem Windows. Program GeoDa je kompatibilní i s ostatními programy pro prostorové explorační analýzy, např. různými programy GIS.

Primární činností každého projektu je získání kvalitních a relevantních dat. Pro tuto práci byla použita data Českého statistického úřadu, konkrétně Územně analytické podklady obcí a podklady geografické databáze ArcČR 500.

Pro převod do základního formátu byl použit software ArcGis. Data získaná na stránkách ČSÚ byla importována do programu ArcGis a v tomto programu modifikována do souboru ve formátu shapefile. Tento typ souborů lze importovat do programu GeoDa.

3.2. Vytvoření vah v GeoDa

Prostorové váhy jsou nezbytné pro výpočet prostorové autokorelace. V programu GeoDa jsou prostorové váhy také využívány k provádění prostorového vyhlazování (tzv. smoothing). Váhy mohou být vypočteny na základě sousednosti ze souborů polygonů nebo jsou počítány na základě vzdáleností mezi body.

Váhy vytvoříme pomocí nástroje v nabídce „*create weight*“, poté v dialogovém okně vyplníme námi vybrané parametry a definujeme váhy. Lze zvolit mezi možnostmi nazvanými „*queen*“ (královna) nebo „*rook*“ (věž). Toto pojmenování je inspirováno pohybem šachových figurek. Věž se pohybuje pouze horizontálním a vertikálním směrem, zatímco královna pohybuje směrem

úhlopříčným. Věž používá pouze společné hranice pro definování sousedů, zatímco královna zahrnuje všechny společné body (hranice a vrcholy). Královna má hustější propojenost (více sousedů) než věž. Po definování a vytvoření vah jsou tyto váhy k dispozici pro všechny prostorové analýzy.

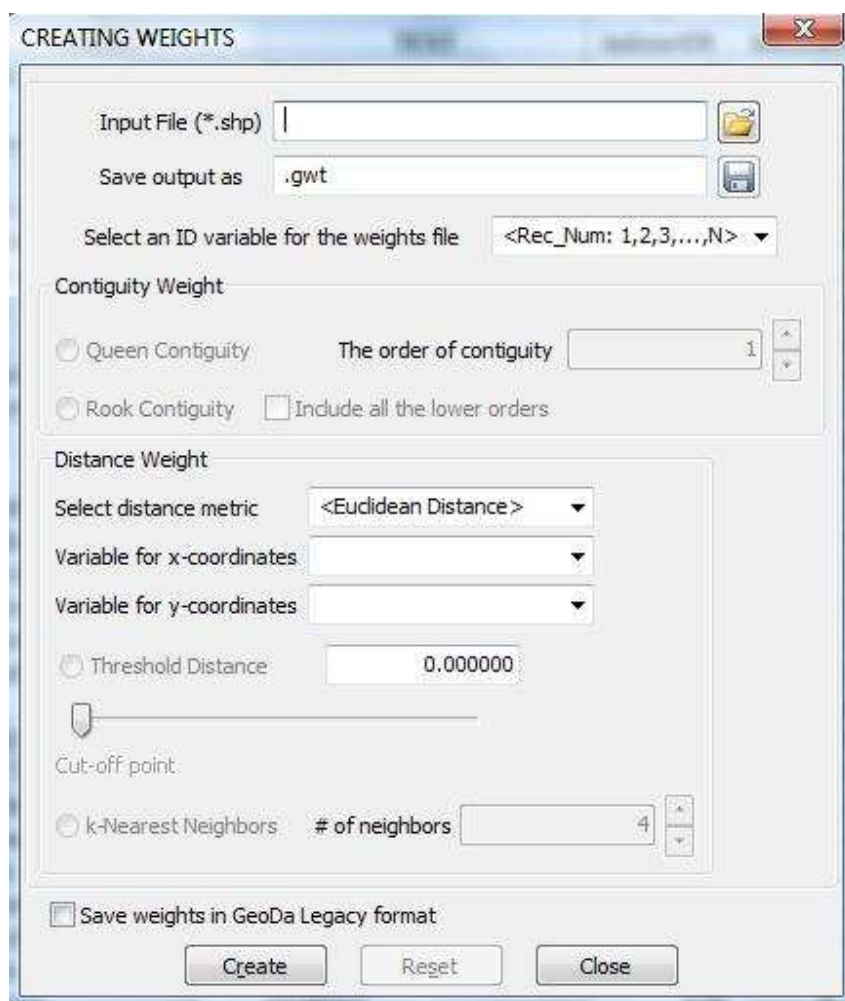
Obrázek č. 5 – Definice vah Rook (věž)

	1	
1	1	1
	1	

Obrázek č. 6 – Definice vah Queen (královna)

1	1	1
1	1	1
1	1	1

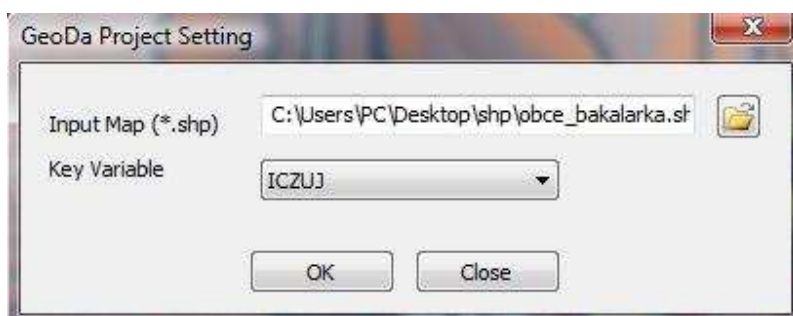
Obrázek č. 7 – Definování vah v programu GeoDa



3.3. Aplikace dat pomocí programu GeoDa

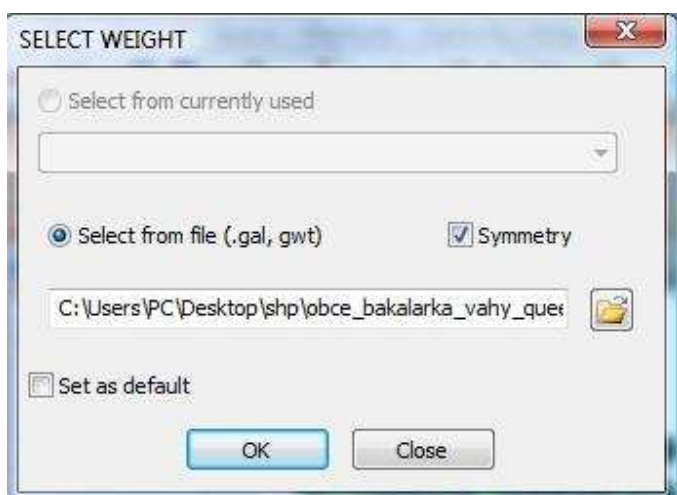
Prvním krokem při aplikaci dat je načtení souboru s příponou shp, tzv. *shapefile*, což je formát prostorových dat, na který se dále zobrazují statistická data. Pro připojení atributových (popisných) dat k vektorovému mapovému podkladu je nutné nejprve zvolit variabilní klíč, což je např. rozloha, kód okresu, ICZUJ atd.

Obrázek č. 8 – Připojení datového souboru shape file



Dalším krokem je připojení vah - tento proces byl již popsán výše v kapitole 3.2. Vytvoření vah v GeoDa.

Obrázek č. 9 – Připojení vah



Po připojení vah lze použít veškeré metody a analýzy, které GeoDa nabízí. V tomto případě použijeme „*univariate LISA*“ a dále zvolíme proměnnou, kterou budeme analyzovat (např. index stárí, přirozený přírůstek, saldo migrace atd.). Lze zvolit několik druhů kartografického zobrazení výsledků explorační prostorové analýzy dat, např. „*cluster map*“, „*significance map*“, „*box plot*“, „*Moran scatter plot*“.

Cluster map (klastrová mapa) zobrazuje zkoumaný jev (zvolenou proměnnou) rozškálovaný do stupnice pěti barev, a to červená – světle červená – světle modrá – modrá - bílá . Červená barva (označená *high-high*) zobrazuje taková místa, kde zkoumaný jev nabývá nejvyšších

(maximálních) hodnot. Světle červená barva (označená high-low) zobrazuje střední až vyšší hodnoty zkoumaného jevu. Světle modrá barva (označená low-high) zobrazuje nižší až střední hodnoty zkoumaného jevu a tmavě modrá barva (značená low-low) zobrazuje nízké (minimální) hodnoty zkoumaného jevu. V místech, kde je barva na mapě bílá, jedná se o statisticky nevýznamné hodnoty.

Significance map (mapa významnosti) zobrazuje statistickou významnost zkoumaného jevu. Na mapě je daný jev zobrazen v barevné škále bílá – světle zelená – zelená – tmavě zelená. Bílá barva zobrazuje území, která jsou statisticky nevýznamná. Světle zelená barva zobrazuje území, kde je zkoumaný jev statisticky signifikantní s 95% pravděpodobností, zelená barva pak označuje 99,5% pravděpodobnost a tmavě zelená barva 99,9% pravděpodobnost.

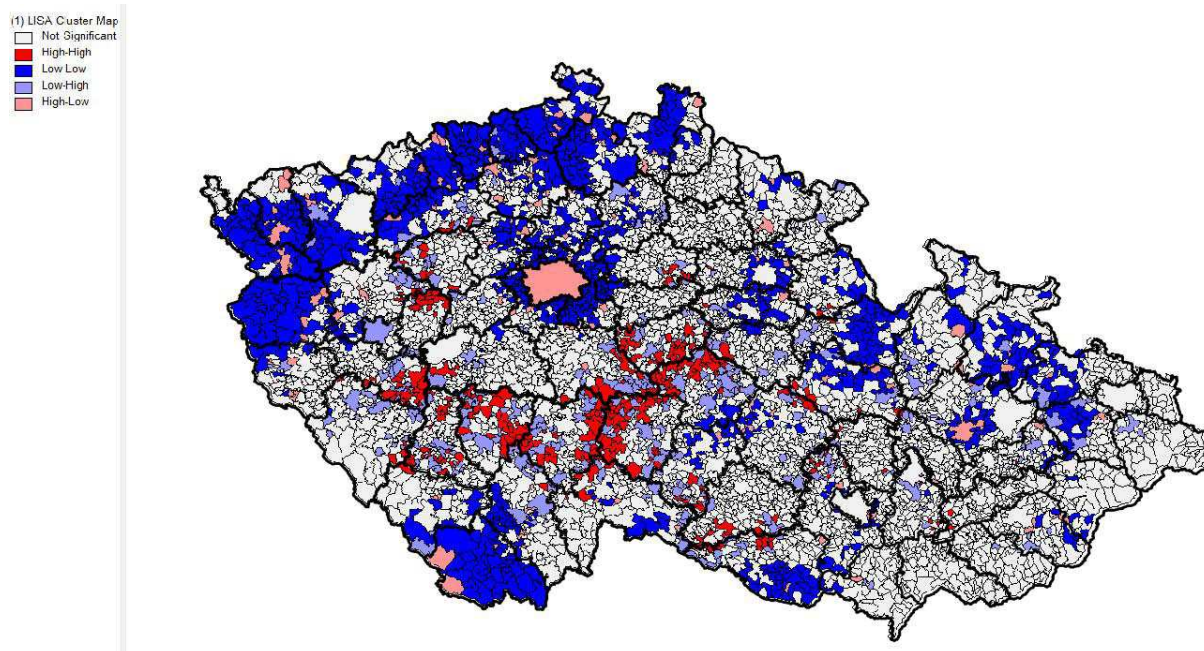
Moran scatter plot (Moranův graf) zobrazuje prostorovou autokorelaci daného jevu podle Moranova indexu. Moranův index byl již popsán v kapitole 1.4.3.3.1.

3.4. Praktická ukázka aplikace dat GeoDa

3.4.1. Index stáří

Jak již bylo zmíněno v kapitole 2.2.2.3., index stáří vyjadřuje vztah mezi I. biologickou složkou a III. biologickou složkou populace. Hodnoty indexu stáří se pohybují okolo 25 u progresivního typu populace, okolo 87 u stacionárního typu populace a okolo 150 u regresivního typu populace.

Mapa č. 1 – Index stáří v České republice, rok 2008



Index stáří je nejvyšší v místech, kde výrazně převažuje podíl osob nad 65 let nad dětskou populací a v mapě je graficky znázorněn červenou barvou. Vytváří tak signifikantní shluky především na hranicích okresů na Plzeňsku, v Jižních Čechách, Středočeském kraji a na Vysočině, konkrétně na hranicích okresů Tábor – Pelhřimov, Tábor – Písek a Havlíčkův Brod - Kutná Hora - Benešov. Je zajímavé, že výskyt červených polí je zřejmý pouze na hranicích okresů, nikoliv uvnitř. Tyto oblasti jsou označovány jako tzv. *vnitřní periferie*. Jedná se o tradiční venkovské oblasti, kde na hranicích okresů leží malé vesničky a uprostřed okresů pak jedno větší okresní město. Pracovní příležitosti jsou ale v malých vesnicích velice omezené a tak mladší generace odchází do měst nebo do okrajových částí měst a vesnice zůstávají osídlené většinou pouze starší populací (proces urbanizace). Ve vnitřních periferiích je také špatná vybavenost jak vzdělanostní, tak infrastrukturní. Intenzita veřejné dopravy je zde velice nízká, což představuje významnou překážku pro stávající, ale i potenciální (nové) obyvatele, a tedy pro další rozvoj těchto regionů.⁶¹

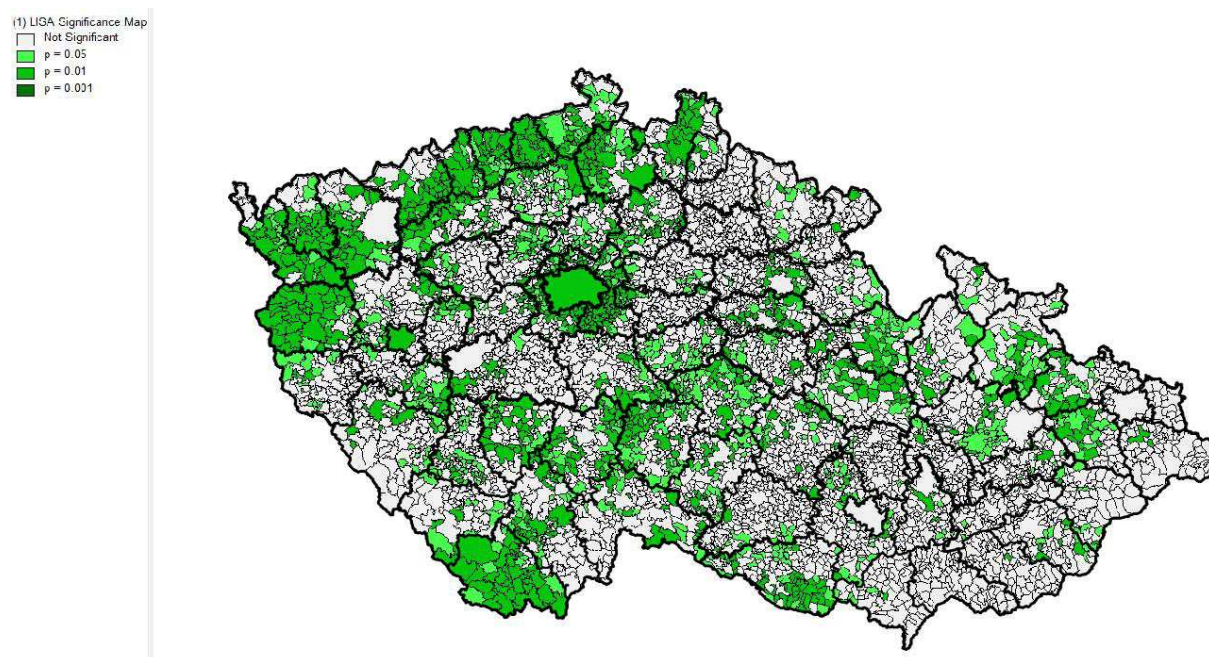
⁶¹PERLÍN, R. *Venkov – typologie venkovského prostoru*. Katedra sociální geografie a regionálního rozvoje PřF UK.

Růžové polygony představují obce s převažujícím podílem osob nad 65 nad dětskou populací obklopené obcemi s nízkými hodnotami indexu stárí. Obvykle se jedná o velká města s migračními úbytky a malé obce v jejich zázemí, které jsou v posledních desetiletích migračně ziskové. Nejpatrnější je tato barva na území hlavního města Prahy, kde dochází k trendu vyliďňování center měst. Mladí obyvatelé se stěhují na okraje měst, kde žijí v rodinných domech a zakládají nové sídelní zóny. Centra měst jsou pak osídlená převážně starší populací. Velká města, v 70. a 80. letech minulého století migračně zisková, zaznamenávají po roce 1989 migrační ztráty, což úzce souvisí s probíhajícím procesem suburbanizace.

Tmavě modrá pole vyjadřují nízký index stárí – tzn., že podíl dětské populace je vyšší než podíl osob nad 65 let. Jedná se především o okrajové části velkých měst (např. Praha). Do těchto míst se stěhují mladí lidé z center měst. Proces suburbanizace je patrný u všech velkých měst.

Index stárí je nízký také v příhraničních oblastech - na Českokrumlovsku a Znojemsku, dále na celé severozápadní hranici České republiky (v okresech Tachov, Cheb, Sokolov, Karlovy Vary, Chomutov, Most, Teplice, Ústí nad Labem, Děčín, Česká Lípa a Liberec), v okrese Ústí nad Orlicí, Bruntál, Opava a Nový Jičín. Tyto příhraniční oblasti byly vysídleny, zejména v poválečném období (jedná se o Sudetské oblasti). V poválečném období zde bylo velké množství prázdných domů a budov a docházelo k osídlování těchto oblastí sociálně slabými skupinami. Nyní v těchto oblastech žijí ve velké míře mladí lidé, kteří mají většinou větší počet dětí. Dále je zde patrná imigrace za prací. Je zde velké množství Poláků, Ukrajinců, Rusů, Vietnamců a jiných menšin, kteří zde pracují a zakládají početné rodiny.

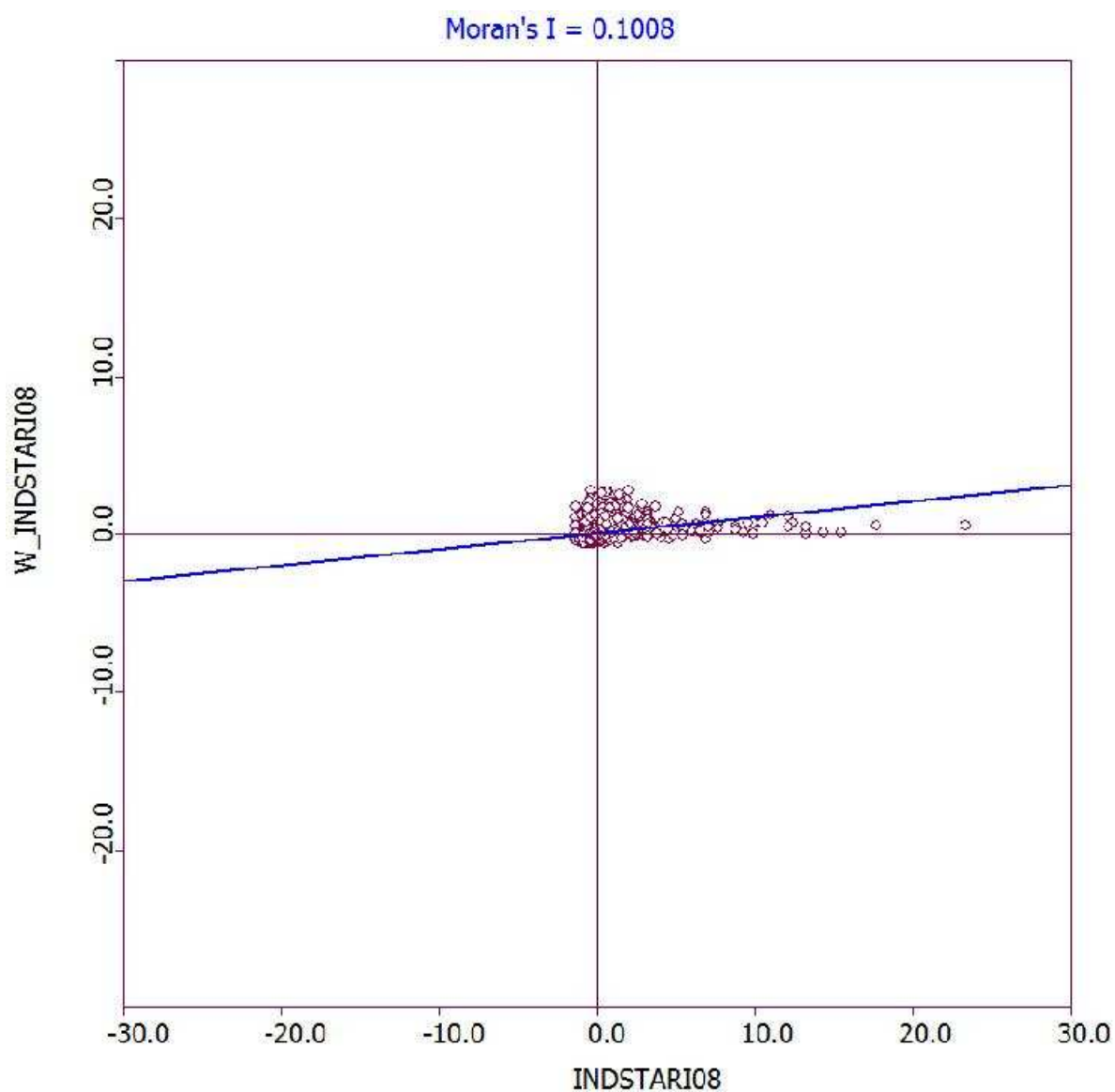
Mapa č. 2 – Index stáří v České republice (statistická významnost), rok 2008



Na mapě lze pozorovat shlukování dat na základě podobnosti hodnot sledovaného jevu. Nelze rozeznat, zda jsou hodnoty vysoké či nízké (to lze pozorovat na předchozí – klastrové – mapě), lze pouze sledovat statistickou významnost na určité hladině.

Na mapě se nevyskytuje tmavě zelená barva, zatímco zelená a světle zelená je jasně patrná. Lze tedy tvrdit, že data, která byla popsána u předchozí klastrové mapy, jsou statisticky signifikantní z 95% (světle zelená barva) či 99% (zelená barva). Např. území Prahy bylo na klastrové mapě znázorněno světle červenou barvou, tj. střední až vysoké hodnoty indexu stáří. Na této mapě významnosti je výskyt této hodnoty potvrzen s 99 % pravděpodobností.

Graf č. 7 – Index stáří dle Moranova indexu v České republice, rok 2008

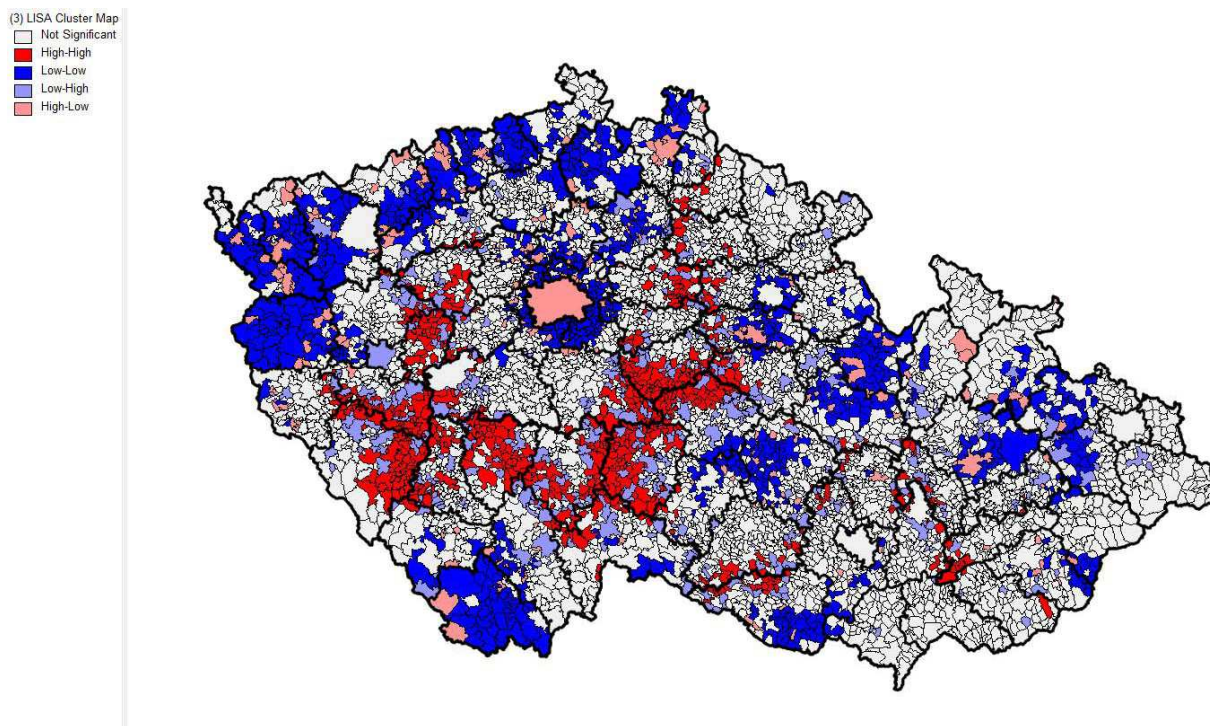


Tento graf zobrazuje prostorovou autokorelaci indexu stáří v České republice za rok 2008. Lze tvrdit, že se jedná o mírně pozitivní prostorovou autokorelaci, protože index vykazuje kladnou hodnotu a data jsou kumulovaná v oblasti grafu high-high (tj. II. kvadrant).

3.4.2. Průměrný věk

Průměrný věk populace se vypočítává váženým průměrem počtu let prožitých příslušníky dané populace a ukazuje stáří populace.

Mapa č. 3 – Průměrný věk v České republice, rok 2008

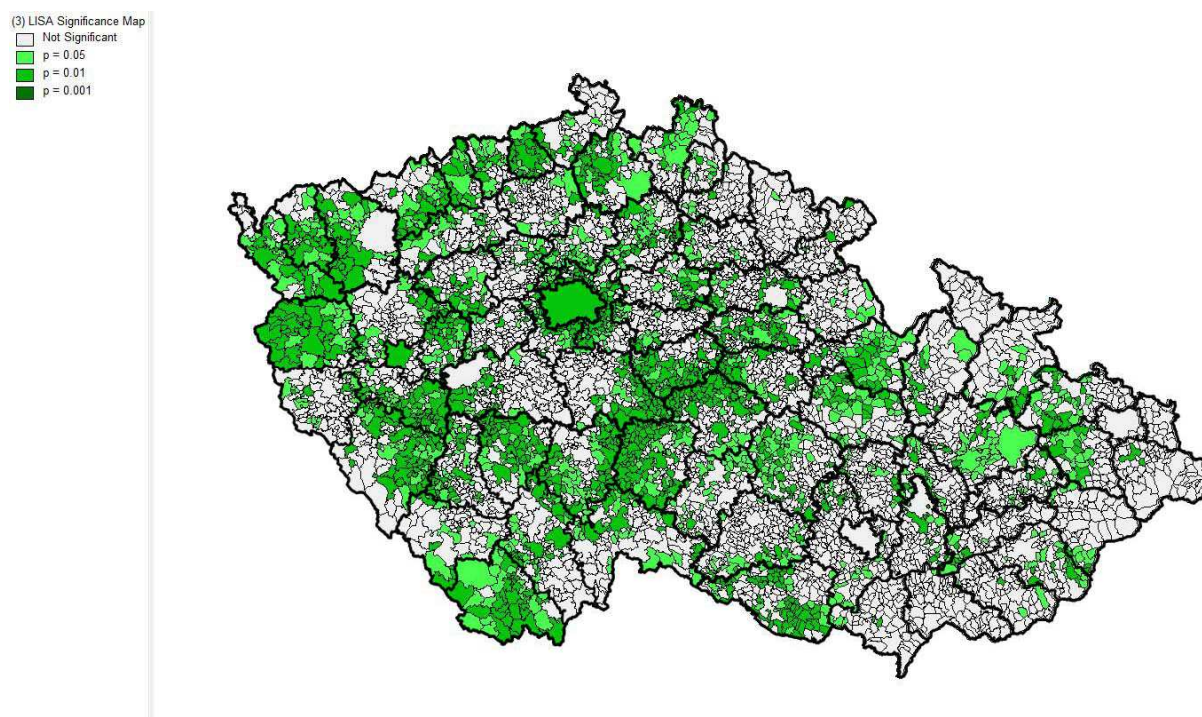


Škála barev zobrazuje stáří populace v dané oblasti – červená barva vyjadřuje shluky obcí s nejstarší populací, zatímco modrá barva vyjadřuje shluky obcí s velice mladou populací.

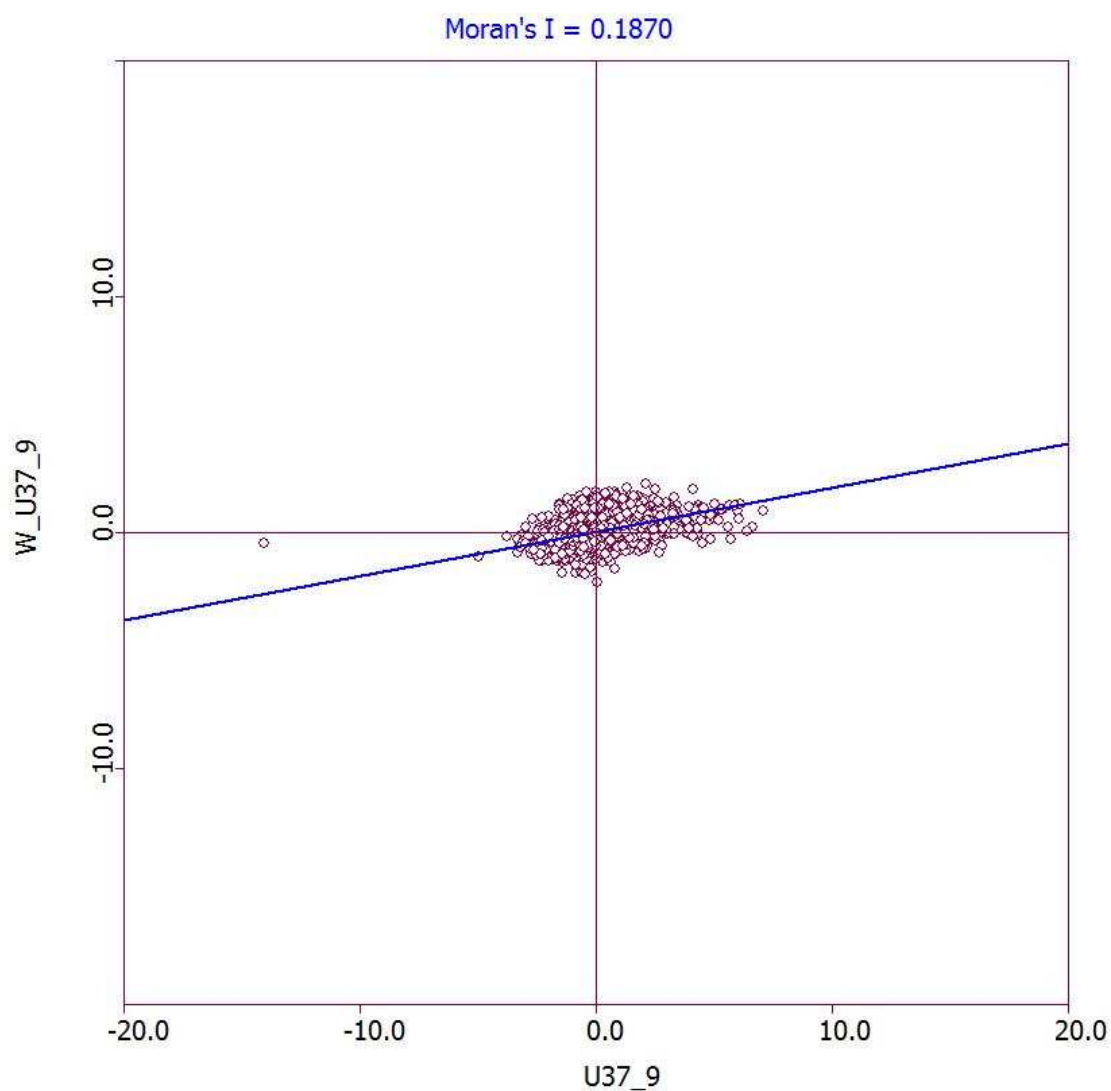
Z mapy je patrné, že nejmladší generace obyvatelstva je koncentrovaná v příhraniční oblasti (Českokrumlovsko, severozápadní příhraniční oblast od Tachovska až po Liberecko, Olomouc, Bruntál, Opava, Ústí nad Orlicí a Svitavy, Znojensko) a dále opět v okrajových částech Prahy. Důvody jsou zde stejné jako u předchozího indexu, indexu stáří, a to zejména vysídlování center měst a stěhování se na okrajové části velkých měst a dále poválečné osídlování Sudetských oblastí sociálně slabými skupinami lidí, které mají většinou početné rodiny. Dalšími důvody může být imigrace do těchto oblastí za práci.

Naopak červená barva (nejstarší populace) je nejvýraznější opět na hranicích okresů, zejména ve středních a jižních Čechách, ale oproti indexu stárí, který byl výrazně červený pouze na hranicích okresů, zasahuje v této mapě červená barva i dovnitř těchto okresů, konkrétně Klatovy, Písek, Tábor, Pelhřimov, Havlíčkův Brod a Kutná Hora.

Mapa č. 4 – Průměrný věk v České republice (statistická významnost), rok 2008



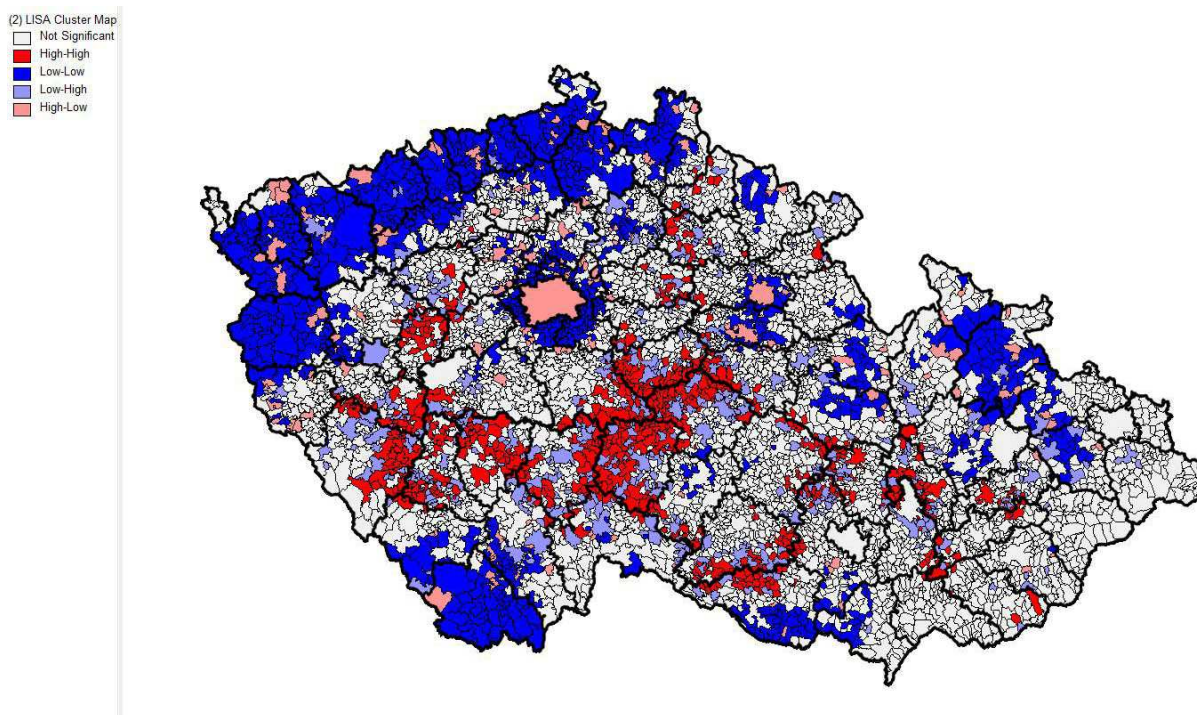
Graf č. 8 – Průměrný věk dle Moranova indexu v České republice, rok 2008



3.4.3. Podíl osob nad 65 let

Tento ukazatel vyjadřuje, jak vysoký podíl III. biologické generace se na daném území nachází v porovnání s I. biologickou generací. Na mapě je graficky znázorněno, ve kterých územích je velký, či naopak malý, podíl osob nad 65 let.

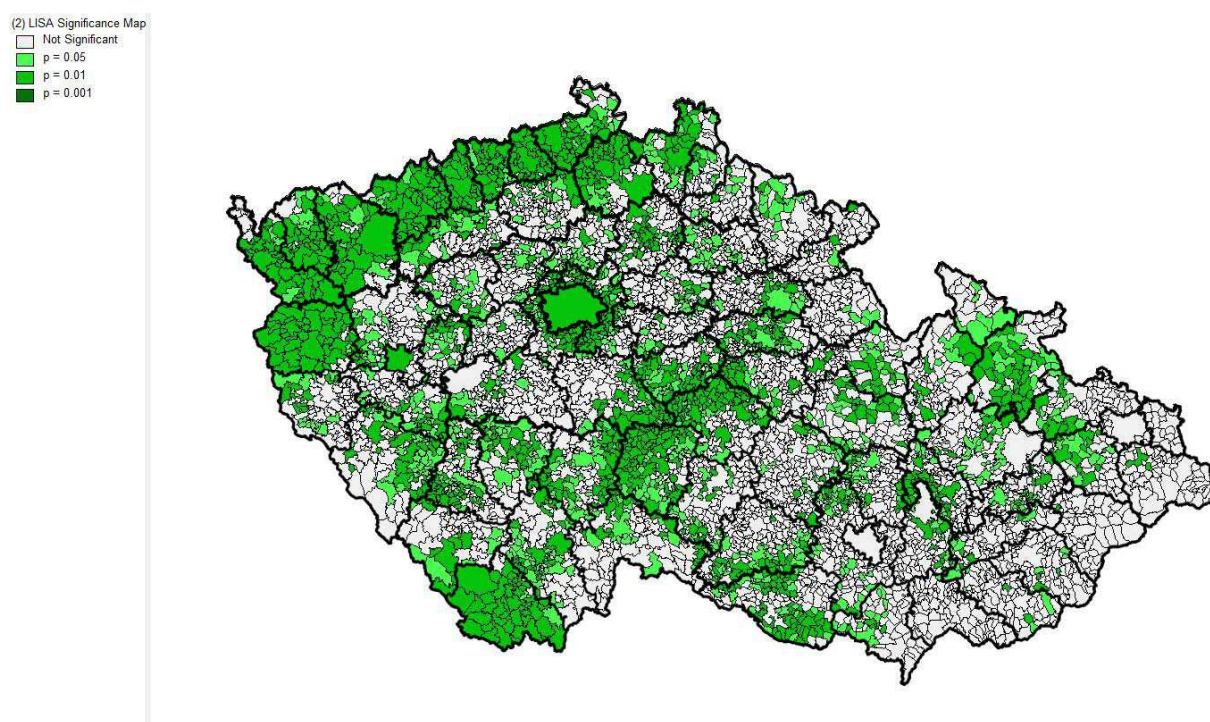
Mapa č. 5 – Podíl osob nad 65 let v České republice, rok 2008



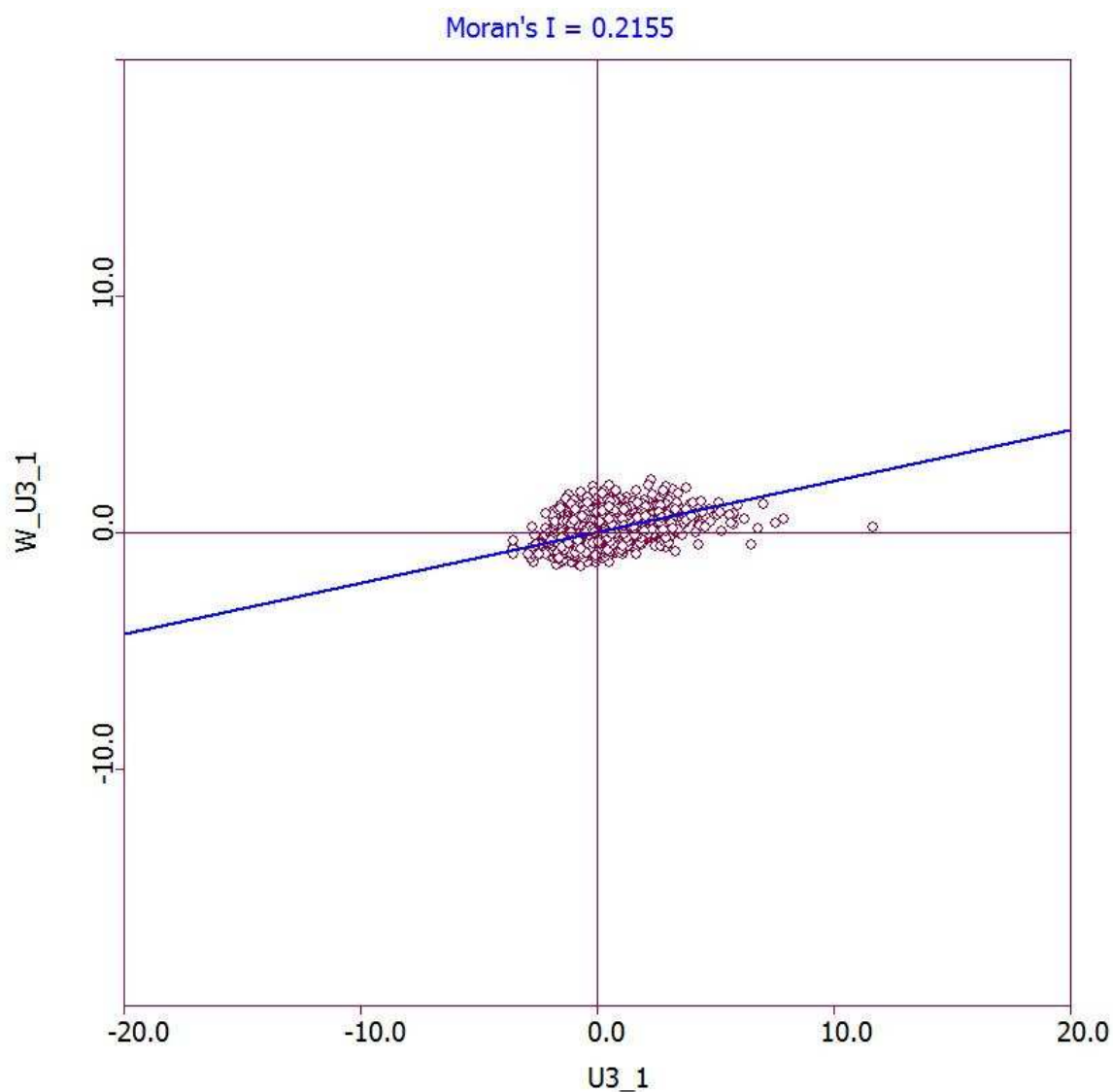
Na mapě je zobrazen vysoký podíl osob nad 65 let červenou barvou. Jedná se opět o hraniční oblasti okresů Klatovy, Strakonice, Plzeň – Jih, Písek, Tábor, Pelhřimov, Benešov, Havl. Brod, Kutná Hora, Znojmo a Třebíč. Jedná se především o vesnické sídelní struktury. Mladé generace díky procesu urbanizace odešly do měst, zatímco starší generace zůstaly v původním osídlení.

Tmavě modrá barva je naopak v místech, kde je podíl populace nad 65 let velice malý. Jedná se o Českokrumlovsko a Prachaticko, Znojemsko, Šumpersko, Bruntálsko a Jesenicko. Dále opět na okraji Prahy a velkých měst. Nejvýraznější souvislý pás modré barvy je pak zřetelný od Domažlicka po Liberecko. Oproti předchozím mapám je modrá naprosto celá oblast, včetně větších měst, která byla v mapě průměrného věku i indexu stáří označena jako statisticky nevýznamná.

Mapa č. 6 – Podíl osob nad 65 let v České republice (statistická významnost), rok 2008



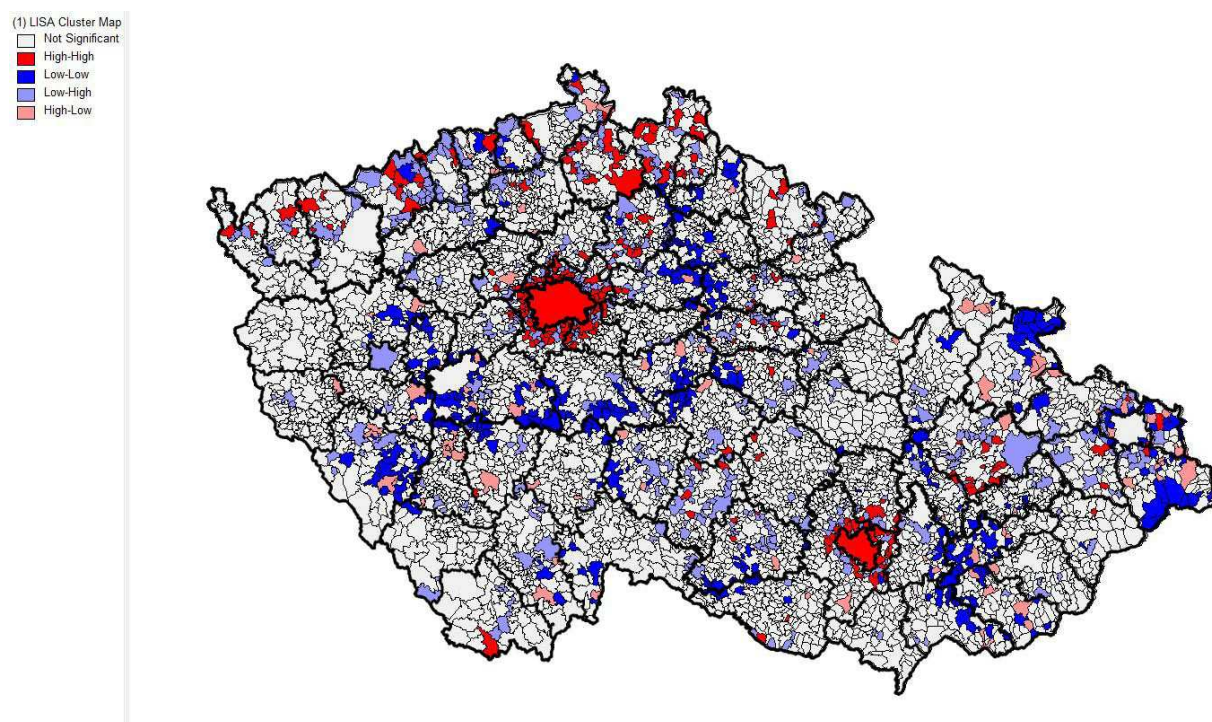
Graf č. 9 – Podíl osob nad 65 let dle Moranova indexu v České republice, rok 2008



3.4.4. Přirozený přírůstek

Ukazatel přirozeného přírůstku vyjadřuje rozdíl mezi živě narozenými a zemřelými v daném časovém období na daném území. V případě kladné hodnoty se jedná o přirozený přírůstek, v případě záporné hodnoty se jedná o přirozený úbytek.

Mapa č. 7 – Přirozený přírůstek v České republice, rok 2008



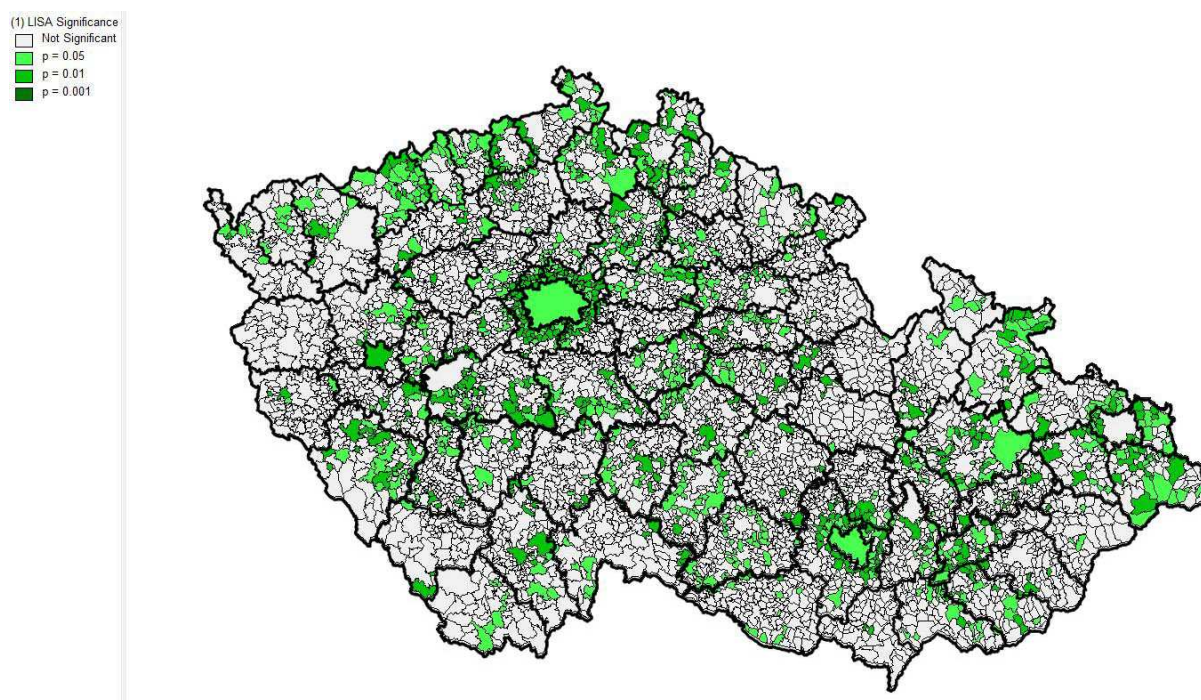
Nejvyšší hodnoty přirozeného přírůstu vykazují lokality, které jsou označeny červenou barvou. Výše přirozeného přírůstu je značně závislá na porodnosti, protože úmrtnost je již dlouhodobě ustálená a výrazně se nemění. Lze tedy tvrdit, že v místech, kde je přirozený přírůstek nejvyšší, je i vysoká porodnost a naopak.

Nejvýraznější je červená barva v lokalitách Praha a okolí, Brno, Kaplice na Českokrumlovsku, v okrese Česká Lípa a Liberec. V těchto oblastech lze tedy předpokládat vysokou porodnost.

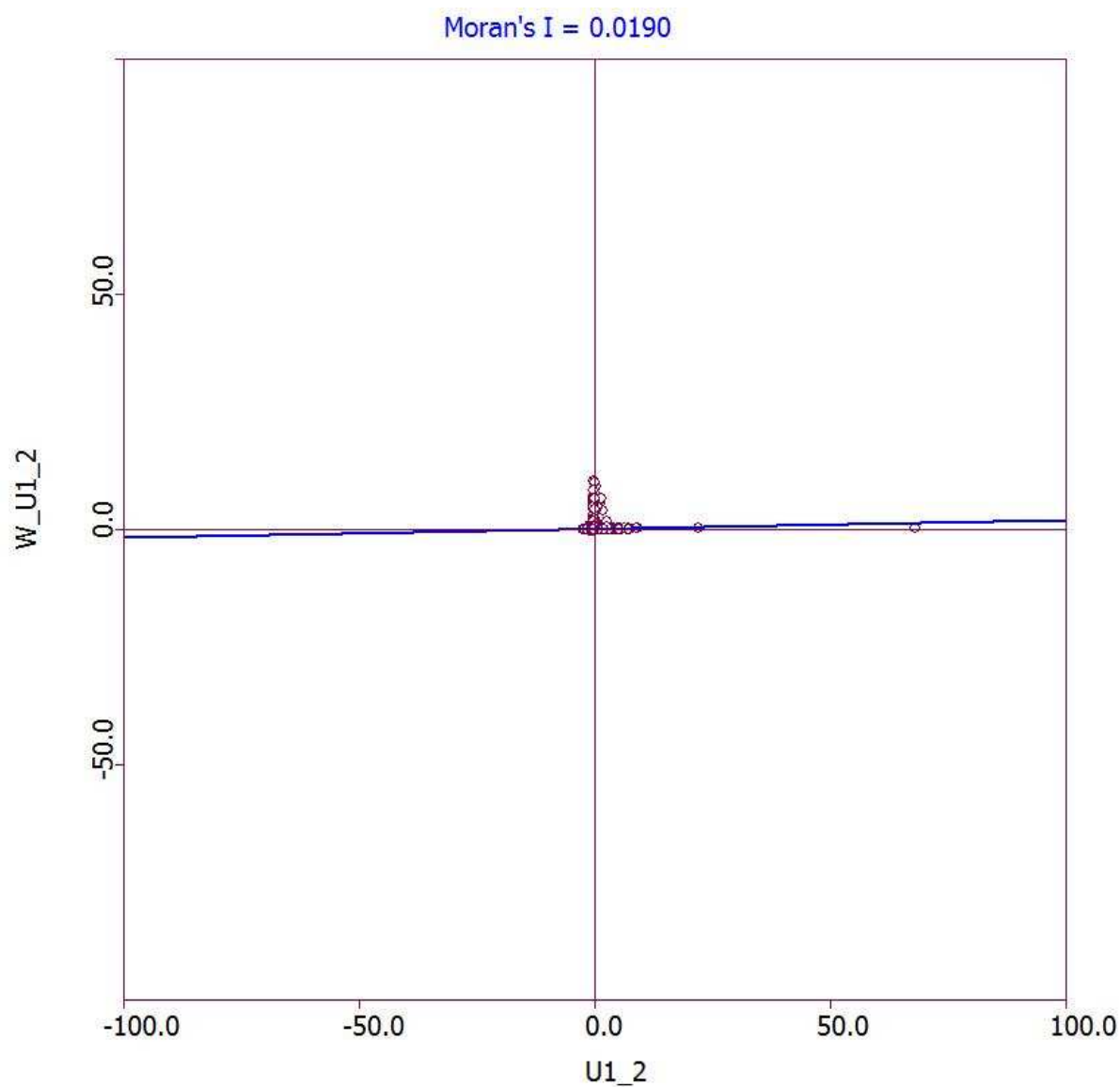
Naopak záporné hodnoty (jedná se tedy o přirozený úbytek) vykazují oblasti označené modrou barvou. Jedná se o severní výběžek Bruntálska, dále o jižní hranici okresu Frýdku – Místku, dále hranice okresů Hodonín, Uherské Hradiště a Kroměříž. V Čechách je to pak oblast hranic Středočeského kraje s ostatními kraji.

Dubí, Frýdlant nad Ostravicí, Sedlčany, Kroměříž, Vrchlabí, Krnov, Vlašim, Karviná, Čáslav, Opava, Karlovy Vary

Mapa č. 8 – Přirozený přírůstek v České republice (statistická významnost), rok 2008



Graf č. 10 – Přirozený přírůstek dle Moranova indexu v České republice, rok 2008



Hodnota Moranova I je téměř nulová, což znamená náhodné prostorové uspořádání. Totéž lze pozorovat i na mapách přirozeného přírůstku – přirozený přírůstek nevykazuje v podstatě žádné významné prostorové uspořádání, jedná se o náhodné prostorové uspořádání.

Závěr

Cílem této bakalářské práce bylo zhodnocení možností využití explorační analýzy dat při analýze socio-ekonomických ukazatelů. V praktické části je popsán postup, jak aplikovat demografická data pomocí počítačového programu GeoDa. Jako vstup do programu GeoDa slouží statistická data (datové řady), které lze získat např. na internetových stránkách Českého statistického úřadu. Tato data se pak musí upravit do požadovaného formátu a načíst v programu GeoDa. Výstupem programu GeoDa jsou různé mapy (v tomto případě České republiky), kde je sledovaný demografický ukazatel rozškálován do několika barev a tyto barvy jsou pak rozmístěné po mapě, dle toho, jakých hodnot nabývá sledovaný jev na konkrétním území. Vzniklé „klastrové“ mapy jsou velice přehledné a snadno analyzovatelné i pro laika. Jejich využití je tudíž širokospektré, dají se použít v různých odborných publikacích, ale také např. ve výuce a výukových materiálech pro studenty. Program nám umožňuje vytvořit si základní představu o prostorové distribuci sledovaného ukazatele, míře prostorové autokorelace a v návaznosti na ně pak použít např. metodu vážené geografické regrese.

Počítačový program GeoDa je vhodný pro analýzu demografických dat, protože nabízí různé způsoby nastavení vah, různé formy výstupních dat (mapy a grafy) a výstupní mapy jsou velice přehledné a snadno analyzovatelné. Tento program je volně ke stažení na internetu.

Summary

The aim of this thesis is to evaluate the use of exploratory spatial data analysis, the analysis of socio-economic activities in the area. This analysis is applied to real data that are imported into the program GeoDa. This program was used for its wide range of applications of spatial analysis, and especially because it is a freely available software. GeoDa is fully compatible with programs dealing with geographic information systems, called GIS program. GIS was used for initial treatment of data.

Exploratory spatial data analysis (ESDA) is a collection of methods used for exploration of spatial data of different type, which is used in a wide range of graphic or visual methods, and maps. Exploratory spatial data analysis tries to answer questions like: "Are there any interesting patterns or relationships?".

The first chapter is the theoretical and methodological description of work, primarily theoretical methods and approaches to the issue and brief history. There are also presented 2 basic computer programs which can perform exploratory spatial data analysis - ArcGIS and GeoDa. This work presents a comparison of ESDA (exploratory spatial data analysis) compared to EDA (exploratory data analysis). In this chapter, it is clear what are the differences in spatial and nonspatial exploratory analysis.

In the second chapter is an overview and basic theoretical description of the demographic indicators. In the third chapter, selected demographic characteristics (age index, average age, proportion of persons over 65 years, natural increase) are applied to the maps and graphs, using the selected program GeoDa. There are analyzed the data that was provided on the website of the Czech Statistical Office. This work assesses the spatial clustering and statistical dependence of these indicators in the Czech Republic.

The program allows you to create a basic understanding of the spatial distribution of the reference indicators, a measure of spatial autocorrelation, and following them, then use a method such as geographical weighted regressions. Analyzed data are due to program clarity and color distribution of the results, relatively easy to interpret.

Demographic data have wide application and analysis is crucial for the society. It is therefore necessary to use such programs, methods and analysis to help the easy and most accurate analysis of demographic data. The aim of this study to evaluate the application of demographic data using a computer program GeoDa. If we evaluate the demographic data at the regional level, then it is a spatial data and is therefore appropriate to apply methods of ESDA. Recently developed as an independent branch of science - geodemography, which deals with the analysis of demographic data in a spatial context. Geodemography is studying natural and mechanical movement of population and demographic processes, structures (age, economic).

Seznam grafů

Graf č. 1 – Vývoj indexu stáří v České republice v letech 1989 - 2008.....	29
Graf č. 2 – Průměrný věk v České republice v letech 1989 - 2008.....	30
Graf č. 3 – Podíl obyvatel ve věkové skupině 0-14 a 65 a více let v České republice v letech 1989 - 2008.....	32
Graf č. 4 – Podíl obyvatel ve věkové skupině 0-14 a 65 a více let v České republice v letech 1950 - 2008.....	33
Graf č. 5 – Struktura přirozeného přírůstku v České republice v letech 1780 - 2000.....	35
Graf č. 6 – Struktura přírůstku obyvatel v České republice v letech 1950 - 2008.....	36
Graf č. 7 – Index stáří dle Moranova indexu v České republice, rok 2008.....	46
Graf č. 8 – Průměrný věk dle Moranova indexu v České republice, rok 2008.....	49
Graf č. 9 – Podíl osob nad 65 let dle Moranova indexu v České republice, rok 2008.....	52
Graf č. 10 – Přirozený přírůstek dle Moranova indexu v České republice, rok 2008.....	55

Seznam obrázků

Obrázek č. 1 – Moranův diagram.....	15
Obrázek č. 2 – Věková pyramida.....	26
Obrázek č. 3 – Populační typy	28
Obrázek č. 4 – Model demografického cyklu.....	37
Obrázek č. 5 – Definice vah Rook (věž)	39
Obrázek č. 6 – Definice vah Queen (královna).....	39
Obrázek č. 7 – Definování vah v programu GeoDa.....	40
Obrázek č. 8 – Připojení datového souboru shape file	41
Obrázek č. 9 – Připojení vah.....	41

Seznam tabulek

Tabulka č. 1 – Specifika analýzy prostorových dat	11
---	----

Seznam map

Mapa č. 1 – Index stáří v České republice, rok 2008	43
Mapa č. 2 – Index stáří v České republice (statistická významnost), rok 2008.....	45
Mapa č. 3 – Průměrný věk v České republice, rok 2008.....	47
Mapa č. 4 – Průměrný věk v České republice (statistická významnost), rok 2008	48
Mapa č. 5 – Podíl osob nad 65 let v České republice, rok 2008.....	50
Mapa č. 6 – Podíl osob nad 65 let v České republice (statistická významnost), rok 2008	51
Mapa č. 7 – Přirozený přírůstek v České republice, rok 2008.....	53
Mapa č. 8 – Přirozený přírůstek v České republice (statistická významnost), rok 2008	54

Seznam použité literatury a zdrojů

- ARIZONA STATE UNIVERSITY. *GeoDa Center for Geospatial Analysis and Computation at ASU*. [online]. URL: < <http://sharedresources.asu.edu/node/2736>> [cit. 2010-05-18].
- ARIZONA STATE UNIVERSITY. *Luc Anselin*. [online]. URL: <<https://webapp4.asu.edu/directory/person/1103870>> [cit. 2010-05-04].
- CENTRUM INFORMAČNÍCH TECHNOLOGIÍ OSTRAVSKÉ UNIVERZITY. *GIS laboratoř*. [online]. URL: < <http://cit.osu.cz/gis/pages/coJeToGis.php>> [cit. 2010-05-18].
- DEMOGRAFICKÉ INFORMACE, ANALÝZY A KOMENTÁŘE. *Stárnutí - vývoj*. [online]. URL: < http://www.demografie.info/?cz_demstarnutivyvoj> [cit. 2010-05-18].
- DEMOGRAFICKÉ INFORMACE, ANALÝZY A KOMENTÁŘE. *Vztah subdisciplín k vědním disciplínám mimo demografii*. [online]. URL: < http://www.demografie.info/?cz_discipliny=> [cit. 2010-05-18].
- HAINING, R. – WISE, S. *Exploratory Spatial Data Analysis*. [online]. URL: <<http://www.ncgia.ucsb.edu/giscc/units/u128/u128.html>> [cit. 2010-05-18].
- HOMOLA, V. *Úvod do statistiky*. [online]. URL: < <http://homel.vsb.cz/~hom50/SLBSTATS/UST/GS02.HTM>> [cit. 2010-05-18].
- KLUFOVÁ, R. *Základy demografie*. České Budějovice: Jihočeská univerzita v Českých Budějovicích, Ekonomická fakulta, 2008. ISBN 978-80-7394-125-3.
- LANGER, T. *Rozdíly v charakteru využití země v krajích ČR – využití metod prostorové autokorelace*. Brno, 2007. Bakalářská práce. Přírodovědecká fakulta. Masarykova univerzita. Vedoucí práce: doc. RNDr. Petr Dobrovolný, CSc.
- MARTÍNEK, J. *Statistické metody v hodnocení léčby*. [online]. URL: < <http://www.hpb.cz/index.php?pId=04-1-2-12>> [cit. 2010-05-18].

- MARYÁŠ, J. – VYSTOUPIL, J. *Geografie obyvatelstva a geografie sídel*. [online]. URL: < http://www.gisaci.upol.cz/filesftp/GEOGRAFIE_OBYVATELSTVA_A_SIDE_L.pdf> [cit. 2010-05-18].
- NETRDOVÁ, P. – NOSEK, V. *Přístupy k měření významu geografického rozměru společenských nerovnoměrností*. Geografie – Sborník české geografické společnosti. Rok 2009. Číslo 1. Ročník 114. Strana 58.
- PERLÍN, R. *Venkov – typologie venkovského prostoru*. Katedra sociální geografie a regionálního rozvoje PřF UK.
- Role statistiky ve výzkumu. [online]. URL: < www.finman.gzk.cz/.../AS_mat_prednaska_role_statistiky_ve_vyzkumu.doc > [cit. 2010-05-05].
- SPURNÁ, P. *Prostorová autokorelace – všudypřítomný jev při analýze prostorových dat?*. Sociologický časopis / Czech Sociological Review. 2008. Vol. 44. No. 4
- Úvod do analýzy časových řad. [online]. URL: <http://home1.vsb.cz/~tvr12/PUBLIKACE/2003_Analyza_CR.pdf> [cit. 2010-05-04].
- *What is GIS?*. [online]. URL: <<http://www.gis.com/content/what-gis>> [cit. 2010-05-18].
- WIKIPEDIA. *Matematická statistika*. [online]. URL: <http://cs.wikipedia.org/wiki/Matematick%C3%A1_statistika> [cit. 2010-05-04].
- WIKIPEDIA. *Medián*. [online] URL: <<http://cs.wikipedia.org/wiki/Medi%C3%A1n>> [cit. 2010-05-04].
- WIKIPEIDA. *Kvantil*. [online]. URL: < <http://cs.wikipedia.org/wiki/Kvantil>> [cit. 2010-05-04].
- WIKIPEDIA. *Modus*. [online]. URL: <<http://cs.wikipedia.org/wiki/Modus>> [cit. 2010-05-04].

- WIKISKRIPTA. *Míra variability.* [online]. URL: <http://www.wikiskripta.eu/index.php/M%C3%ADry_variability> [cit. 2010-05-18].
- *Základní statistické pojmy. Popisná statistika.* [online] URL: <<http://www.informatika.xcars.cz/zakladnipojmy.html>> [cit. 2010-05-04].