

Jaroslav KRAUS

GEOSTATISTICKÁ ANALÝZA DEMOGRAFICKÝCH JEVŮ PODLE VÝSLEDKŮ CENSU 2001

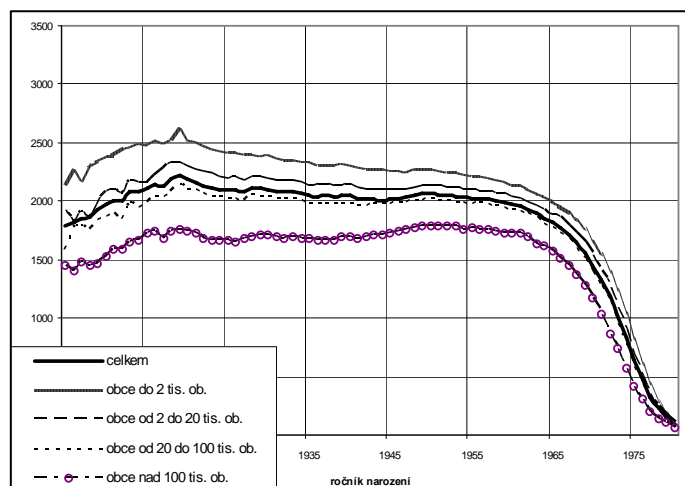
Kraus J.: Geostatistic Analysis of Demographic phenomena in according to Census 2001.
Kartografické listy 2008, 16, 5 figs., 4 refs.

Abstract: The article discusses the new methods and possibilities for the analysis of regional differences in demographic data. It sets out from the idea that changes in spatial distribution are also connected with changes in the reproduction regime (fertility). These changes are studied using the methods of geostatistical analysis and the results are summarised as a spatial model of fertility. Errors occur not only in texts (linguistic errors) but also in graphic works.

Keywords: spatial differentiation, fertility, geostatistics, modelling

Úvod

Údaje populačních censů patří mezi základní zdroje demografických dat a jsou východiskem řady analýz. Proti jiným mají tu výhodu, že poskytují výsledky ve velkém územním detailu a umožňují tak mimo jiné analýzu dat z hlediska územních rozdílů. Definice území vždy souvisí s logickou strukturou analyzovaných dat a s určitou minimální velikostí souborů událostí, které jsou předmětem prostorové analýzy. Mezi základní a v dnešní době silně sledované demografické jevy patří nesporně plodnost žen (obr. 1). Mezi základní ukazatele plodnosti žen pak patří generační (konečná) plodnost žen – ta vypovídá o počtech narozených dětí a tedy o úrovni reprodukčního procesu. Při sčítání v roce 2001 v České republice (ČR) všechny 15leté a starší ženy odpovídaly na otázku, kolik se jim narodilo živých dětí. Tento údaj je – společně s informacemi o místě trvalého bydliště – zdrojem analýzy a modelového řešení konečné plodnosti žen na území ČR v tomto příspěvku.



Obr. 1 Průměrný počet dětí na 1000 žen celkem podle velikostních skupin obcí

Prostorová diferenciace plodnosti

Úroveň konečné plodnosti žen lze vyjádřit průměrným počtem žijících dětí. V posledním desetiletí se v ČR plodnost dramaticky snížila. Hodnota úhrnné plodnosti poklesla z 1,89 v roce 1990 na minimum 1,13 v roce 1999 s následným mírným zvýšením na 1,17 v roce 2002. Tímto trendem se populace ČR zařadila do kategorie lowest low fertility (Rychtaříková 2004) a význam zkoumání změn tohoto poklesu se nabyt na významu. Vedle sledování příčin poklesu je důležitá i otázka zkoumání změn z územního hlediska. Určitou odpověď poskytují výsledky průměrného počtu dětí podle krajů a podle velikostních skupin místa bydliště. Z nich vyplývá (Kraus 2004), že průměrný počet dětí je nejvyšší v nejmenších obcích a naopak nejnižších ve velkých městech, zejména pak v hl.m. Praze. Z hlediska jednotlivých regionů České republiky nejsou rozdíly výrazné, se zřetelnou výjimkou hl. m. Prahy, kdy průměrný počet dětí je výrazně nižší.

V současné době existuje nová možnost, který umožňuje se na prostorovou diferenciaci plodnosti podívat jiným způsobem – prostřednictvím geostatistické analýzy s využitím nástrojů geografického informačního systému (GIS). Východiskem je v tomto případě geostatistická analýza, která pracuje se známými hodnotami v prostoru a nad nimi vytváří (interpoluje) souvislou plochu. V rámci geostatistické analýzy se tak predikuje hodnota v každém bodě plochy z měření získaných ve výběrových bodech do plošného zobrazení zkoumaného jevu. Apriorním předpokladem tohoto postupu je úvaha, že demografické procesy v prostoru mají spojitý charakter (Tuček 1998).

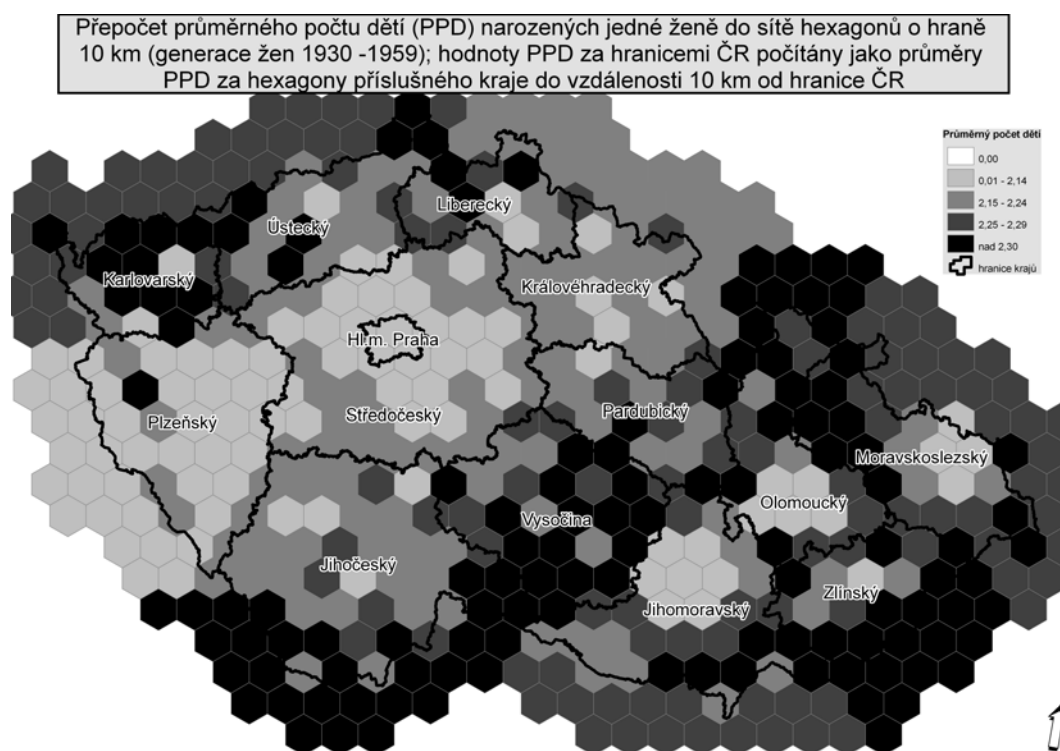
Výchozí situace je patrná z obr. 2 (, kde se zobrazuje průměrný počet dětí žen v obcích s více než 2 tisíci obyvateli. Problémem je výpočet plodnosti žen v obcích do 2 tisíc obyvatel – vzhledem k velikosti souborů a tedy kolísání hodnot. Jedním z možných řešení je, že hodnoty v obcích do 2 tisíc obyvatel se nahradí hodnotou za kraj (NUTS3) pro všechny obce daného kraje, dále pak hodnoty v obcích ležících v zázemí měst nad sto tisíc obyvatel (t.j. v okruhu 20 km) se nahradí hodnotou právě těchto měst. Jedná se o modelové řešení, které je pochopitelně možné podrobit diskusi. Plodnost v obcích nad 2 tisíce obyvatel pak byla vypočtena standardně z individuálních dat sčítání lidu.



Obr. 2 Průměrný počet dětí narozených jedné ženě v obcích s 2000 a více obyvateli

Modelové řešení

Další krok spočívá v řešení toho, co by se dalo nazvat reprezentativnost území z hlediska plodnosti. Zvolený způsob spočíval v konstrukci dělení prostoru do pravidelné sítě buněk (mozaiky) a výsledky za jednotlivé obce a města do této mozaiky přepočíst. V praxi připadají do úvahy dvě mozaiky: čtvercová a hexagonální. Výhodou první je kompatibilita se strukturami datových posloupností používaných ve výpočetní technice a kompatibilita s karteziánskými souřadnicovými systémy. Hlavní výhodou druhé (hexagonální) je, že středy všech sousedních buněk od středu dané buňky jsou stejně vzdálené. Tato symetrie činí hexagonální mozaiku metodicky správnější z hlediska prostorové analýzy a byla použita v tomto příspěvku (obr. 3). Konečná plodnost žen ve všech obcích byla přepočtena do sítě hexagonů o hraně 10 km. Velikost hrany hexagonu nebyla zvolena náhodně a souvisí s průměrnou velikostí území okresů (NUST4), se kterými se v demografické analýze často pracuje.

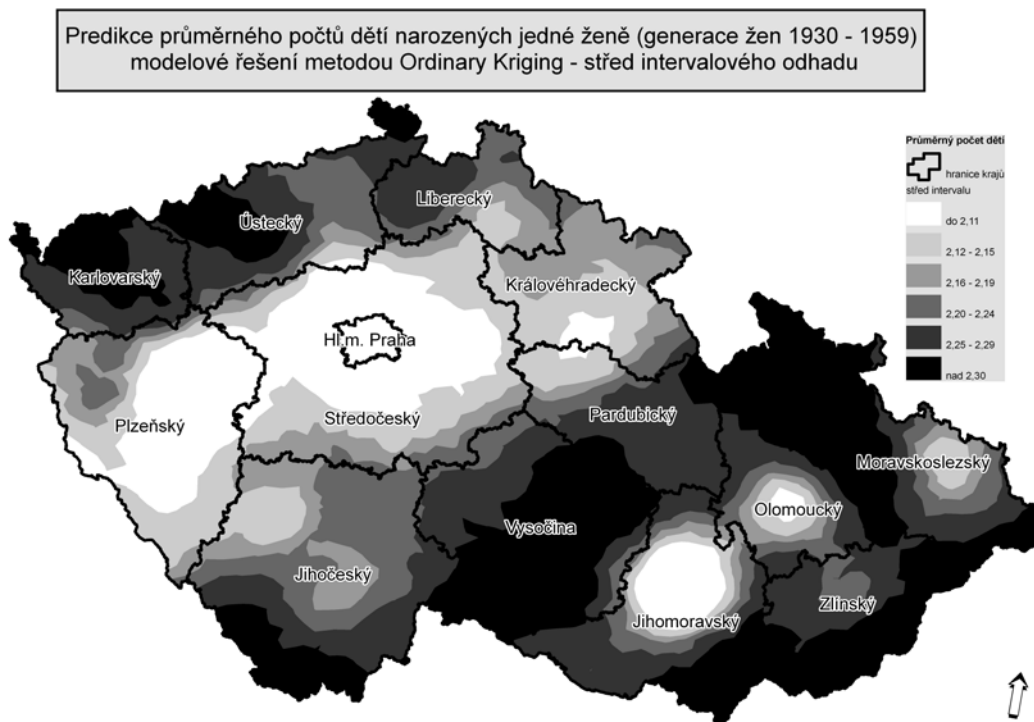


Obr. 3 Přepočet průměrného počtu dětí narozených jedné ženě do sítě hexagonů o hraně 10 km

Základem výpočtu průměrného počtu dětí za obce byly individuální (anonymizované) soubory censu, ze kterých byl vypočten průměrný počet dětí všech žen standardním způsobem. Výpočet průměrného počtu dětí v hexagonální mozaice byl počítán jako vážený průměr, když vahami byl počet žen příslušných generací. Tímto způsobem byl vypočteny hodnoty průměrného počtu dětí za 428 hexagonů mozaiky, které pokrývají území ČR. Pro zjednodušení úlohy byla brána ČR jako uzavřený celek, t.j. hodnoty plodnosti za hranicemi státu byly brány jako nulové. Tím byl vytvořen prostor k práci s geostatistickým modelem. Prvním vhodným krokem při tvorbě modelu území je analýza dat. Většina (ne všechny) datových modelů vychází z předpokladu normality dat nebo alespoň přibližné normality dat. Pokud tato normalita dat splněna není, je nutné data transformovat. Prostorové údaje průměrného počtu dětí podle území však pochází z přibližně z normálního rozdělení a datová transformace nebyla nezbytná.

Pro modelování prostorového řešení průměrného počtu dětí na území ČR byla zvolena jedna z metod kriging, kterých je celá škála. Metody jsou založeny na autokorelaci jako funkci vzdálenosti. Jinými slovy: jevy v dané lokalitě jsou více ovlivňovány jevy v lokalitách sousedních, než v oblastech vzdálených. Tento předpoklad lze modelově aplikovat i na data plodnosti. Informace o hodnotě daného jevu v určitém prostoru je počítána jako vzdálenosti mezi jednotlivými pozorováními a modelovou autokorelací jako funkcí vzdálenosti: $Z(s) = \mu(s) + \varepsilon(s)$, kde $Z(s)$ je zkoumaná proměnná (t.j. průměrný počet dětí) dekomponovaná na deterministický trend $\mu(s)$ a náhodnou autokorelovanou chybu $\varepsilon(s)$. Hodnota s indikuje umístění v prostoru: x, y souřadnice určitého území. Podobně jako v jiných stochastických modelech se vychází z nedokonalosti předpovědi deterministické složky, hodnota $\varepsilon(s)$ by tak měla být v průměru rovna 0. Autokorelace mezi $\varepsilon(s)$ a $\varepsilon(s+h)$ nezávisí na aktuálních x, y hodnotách s , ale na posunu h mezi nimi.

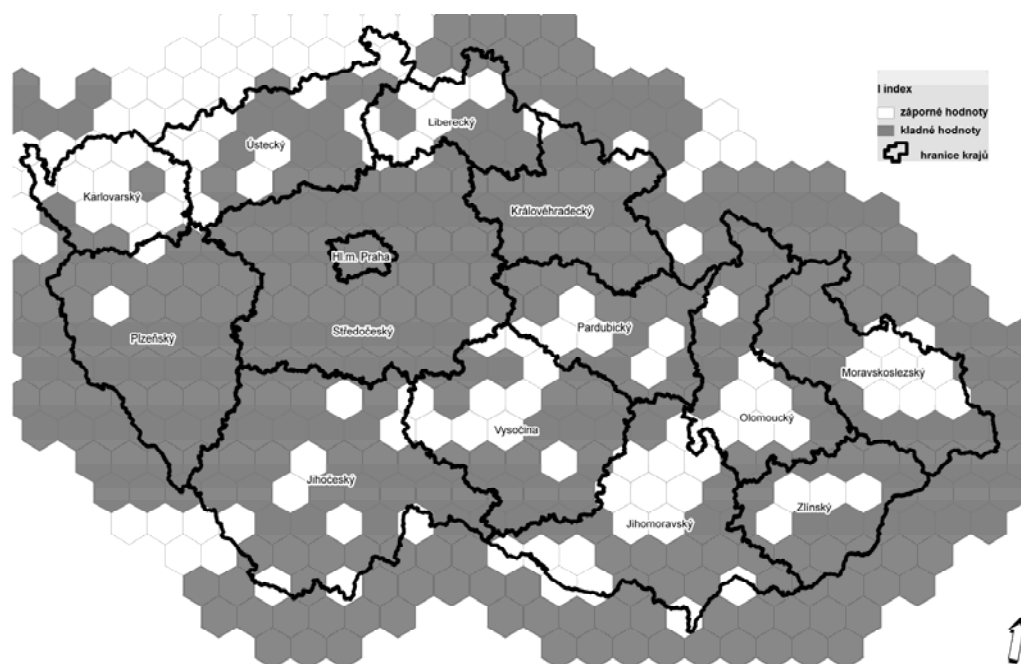
Další problém při řešení úlohy tohoto typu spočívá ve zkoumání trendu: jestli existuje (resp. jestli ho lze se znalostí dat a priori předpokládat). Pokud existuje, tak jakého je typu – konstanta nebo lineárního funkce, která pak může být vyjádřena polynomem n -tého řádu. Metody kriging jsou predikčními metodami a jejich konečným cílem je předpověď zkoumané plochy (různými metodami). Z toho vyplývá možnost výpočtu standardizované chyby. Jak je patrné z obr. 4, lze v datech průměrného počtu dětí vysledovat určité prostorové schéma: počet dětí je v západní a střední části ČR (Čechy) zřetelně nižší než ve východní části (Morava). Lze tedy předpoklad trendu do modelu zabudovat, je však problematické zabudovat do modelu jeho odhad $\mu(s)$. Do úvahy by snad připadalo nahradit tuto hodnotu výpočtem průměrného počtu dětí za celou ČR. Z těchto důvodů byla zvolena metoda *Ordinary kriging*, kde se předpokládá spojitý trend v datech s neznámou hodnotou.



Obr. 4 Predikce průměrného počtu dětí narozených jedné ženě – modelové řešení metodou Ordinary kriging, střed intervalového odhadu

Jak je patrné, výsledky odpovídají očekávání: že míra prosté reprodukce 2,15 dítěte na jednu ženu bude s vyšší pravděpodobností překročena na Moravě, naopak nejnižší pravděpodobnost vychází pro Prahu a středočeský region. Současně se však ukazuje slabina tohoto řešení: tím, že se do

úvahy berou pouze data za ČR hexagony ležící v příhraničí mají nižší četnost dat, protože část hexagonu leží za hranicí ČR (to je problém řešení a nesouvisí se strukturou osídlení ČR), což ovlivňuje výpočet průměrné hodnoty za celý hexagon. V tomto příspěvku bylo zvoleno řešení problému přidáním jedné řady hexagonů ležících za hranicemi ČR. Hodnota v těchto hexagonech bylo vypočtena jako průměr hexagonů za každý kraj ležících v 10 km od hranice republiky. Toto řešení je patrné z obr. 5.



Obr. 5 Prostorová analýza – výpočet lokálního prostorového (Moranovo I) indexu

Na základě zvoleného postupu je dále možné spočítat pravděpodobnost, že hodnota v daném území převyší určitou předem zvolenou hodnotu (konstantu), např. průměrný počet dětí v ČR. Při zpracování predikční mapy je nutné pracovat s řadou vstupních parametrů jako je počet sousedících hexagonů, které jsou brány při výpočtu hodnoty v dané lokalitě, případné členění celkové plochy do dílčích regionů, výběr vlastní modelové funkce. Na konci postupu je pak kartodiagram jako modelové řešení dané úlohy. Vypočtené hodnoty je nutné dále zpracovat: z prostorového řešení v hodnotách x_{min} , x_{max} , y_{min} a y_{max} vybrat prostorovým dotazem hodnoty za území ČR. Tyto výsledky pak dále prostorově škálovat – např. kvartilovým rozložením zkoumaného jevu a bichromatickou barevnou škálou.

V předchozím postupu byl po transformaci údajů za obce do sítě hexagonů (plochy) vytvořen prostorový predikční model – pomocí jedné datové proměnné metodou ordinary kriging. Tímto přístupem bylo možné stanovit oblasti s vyšší, průměrnou nebo nižší intenzitou jevu – průměrného počtu dětí připadajících na jednu ženu (generace žen 1930-1959) a vypočítat modelové charakteristiky.

Další krok, který se po zpracování prostorové predikce nabízí, je vyhodnocení intenzity prostorové závislosti pomocí statistiky, která je určena pro analýzu prostorové a věčné variability zkoumaného jevu. K tomu se používá řada statistik, např. tzv. Moranovo I. (Maguire 2005), Jedná se o prostorovou aplikaci shlukové analýzy, jejímž výstupem je hodnota indexu (Moranovo I) a hodnota statistiky z , počítaná pro každý prvek (hexagon). Jestliže je hodnota indexu pozitivní, pak má prvek (hexagon) hodnotu podobnou té, jaká je zaznamenána v hexagonech okolních. Jestliže je hodnota negativní, pak se hodnota v tomto hexagonu liší od okolních – posuzováno na základě toho, co je a co není statisticky významné. Hodnota z skóre pak vypovídá o statistické vý-

znamnosti Moranova I indexu. Aby bylo možné určit, zda je jev (v dané lokalitě) statisticky významný, je nutné jej porovnat s intervalem spolehlivosti. Např. na hladině významnosti 95% jsou rozdíly mezi sousedícími hexagony statisticky významné, jestliže hodnota z skóre je menší než (-1,96), resp. větší než (+1,96).

Z výsledků obsažených v obr. 5, který ukazuje hodnoty indexu z (zobrazeného jako kladné a záporné hodnoty) vyplývá, že území krajů Středočeského a Plzeňského a Prahy, charakterizované nízkou hodnotou průměrného počtu dětí (PPD), tvoří homogenní celek, kde se úroveň plodnosti místně významně neliší. K této oblasti se ještě přimyká území Královéhradeckého kraje s průměrnými hodnotami PPD. U dalšího území tato relativní stabilita neplatí. Typickým příkladem je kraj Jižomoravský, Olomoucký a Moravskoslezský, kde je zřejmé, že plodnost, vyjádřená PPD, se výrazným způsobem diferencuje v závislosti na konkrétní lokalitě – rozhodující vliv zde má velikostní skupina místa trvalého bydliště. Podobné kolísání bylo zjištěno i v kraji Karlovarském, Ústeckém, Libereckém, Jihočeském, Pardubickém a Zlínském. Statistická významnost tohoto jevu, zobrazená v kartogramu 20, byla potvrzena pro kraj Středočeský, Plzeňský a Prahu: týká se celého území těchto krajů a neexistují zde místní odlišnosti. Signifikantní (t.j. významná podobnost hodnot PPD) závislost, je rovněž zjištěna v širším zázemí Brna, Olomouce a Ostravy – nikoliv však v krajích těchto měst jako celcích. Nejnižší plodnost (vyjádřenou minimálním počtem dětí) lze očekávat ve především ve velkoměstech a v zázemí těchto velkoměst. V Čechách minimální hodnoty vytváří souvislý pás vycházející z hl.m. Prahy, zahrnují převážnou část Středočeského kraje a zasahují většinu Plzeňského kraje, těžištěm je zde město Plzeň. Morava netvoří jednotný celek a existují zde větší kontrasty: Brno, méně pak Olomouc a Ostrava jsou obklopeny v úzkém pásu regiony, kde je plodnost citelně vyšší.

Závěr

Tento příspěvek je určitým modelovým řešením, které si zasluhuje podrobnější diskusi, ukazuje však možnosti některých nových nástrojů geografického informačního systému v oblasti demografické analýzy.

Příspěvek je součástí výstupů projektů GA ČR 205/06/0965 „Vizualizace, interpretace a percepce prostorových informací v tematických mapách“.

Literatura

- RYCHTAŘÍKOVÁ, J. (2004). Změny generační plodnosti v ČR se zaměřením na vzdělání žen. *Demografie*, č. 2, s. 77-90.
- KRAUS J., RYCHTAŘÍKOVÁ J.(eds) (2005). *Atlas sčítání 2001* (Univerzita Karlova Přírodovědecká fakulta).
- KRAUS, J. (2004). Regionální diference plodnosti. In. *Perspektivy rodinné politiky v ČR – sborník z konference*, s. 123-129.
- TUČEK, J. (1998). *Geografické informační systémy*. Praha, s. 53-162. (Computer Press).

S u m m a r y

Geostatistic Analysis of Demographic phenomena in according to Census 2001

This article focuses on an analysis and conclusions about regional differences based on a spatial model of fertility. It is methodologically grounded in the theory of geostatistics, which has thus far only been used to a limited degree in the social sciences. In spatial analysis the value of the phenomenon at all places across an area is derived from the values measured in selected places. Each place than has the value obtained either by measurement or by estimation.

The solution is based on calculating the average number of children (ANC – average number of children) born to the 1930-1959 generation of women in municipalities with over two thousand inhabitants and on replacing the value of the average number of children in municipalities with fewer than two thousand inhabitants with the value of the region (NUTS 3) for all the municipalities of this size in the given region.

The model solution of the spatial prediction developed using the method of ordinary kriging works with the existence of a certain mean value of the average number of children and with the fact that local conditions in some way further modify this mean value.

The results indicate that the Central Bohemian Region, the Pilsen Region, and the Prague Region, characterised by a low ANC value, form a homogenous unit, where the fertility rate does not significantly differ locally. The Hradec Králové Region also approaches this group with its average ANC values. This relative stability does not apply to other regions. Typical examples are the South Moravian Region, the Olomouc Region, and the Moravian-Silesian Region, where it is evident that fertility, expressed as ANC, is significantly differentiated by locality. The statistical significance of this phenomenon was confirmed for the Central Bohemian Region, the Pilsen Region and Prague Region: it relates to the entire territory of these regions and there are no local differences. A significant dependence (i.e. significant similarity of ANC values) is also ascertained in the wider base of the cities of Brno, Olomouc, and Ostrava – but not in the regions of these towns in their entirety (fig. 1).

Fig. 1 Average number of children per 1000 of women by generation of women and by community size

Fig. 2 Average number of children per 1000 of women (the generation of women born in 1930-1959) in municipalities with 2000 or more inhabitants

Fig. 3 Recalculation of average number of children born per women into network of hexagons with 10 km edge

Fig. 4 Prediction of average number of children born per women, modelling by Ordinary Kriging method, medium of interval

Fig. 5 Spatial Analysis – Local Moran's I index

Lektoroval:

Prof. RNDr. Vít VOŽENÍLEK, CSc.,

Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Olomouc, Česká republika