

VYUŽITÍ VÍCEROZMĚRNÝCH STATISTICKÝCH METOD PŘI HODNOCENÍ KRAJŮ

USING OF MULTIVARIATE STATISTICAL METHODS FOR REGIONAL EVALUATION

Jakub Hloušek

Anotace:

Problematika diferenciacie regionů podle ekonomických nebo sociálních úrovní patří v současnosti se vstupem České republiky do Evropské unie k hlavním tématům ekonomického i politického spektra. Smyslem a cílem regionálního přístupu je zajištění stejných možností rozvoje regionů tak, aby byl maximálně využíván hospodářský, přírodní a populační potenciál při respektování pravidel trvale udržitelného rozvoje.

Hodnocení regionálních rozdílů lze provádět celou řadou technik a postupů. Účelné je však využít ty, které umožňují si udělat obecnější názor na celkovou úroveň regionu a které shrnují nejvýznamnější dílčí charakteristiky. Specifickou skupinou metod zabývajících se vícekritériálním vyhodnocením určitých skupin jsou vícerozměrné statistické analýzy. Sem patří hned několik užitečných metod, které přinášejí souhrnné hodnocení. Z pohledu rychlého a názorného vyhodnocení se jeví jako nejužitečnější shluková analýza.

Klíčová slova:

Sociální diferenciacie, regionální úroveň, vícerozměrné statistické metody, shluková analýza, kartogram.

Abstract:

Actually in connection with the Czech Republic admission to the European Union belong questions of regional differentiation according to economic or social level to main subjects of economic and political sphere. Sense and goal of regional access is to guarantee the same possibility of regional development so that economic, natural and population potential was used at most as possible in the frame of respecting the rules of sustainable development. The methods of the multivariate statistical analysis form wide and important set of methods and the cluster analysis is one of the most used of them.

Keywords:

Social differentiation, regional level, multivariate statistical methods, cluster analysis, cartogram.

ÚVOD A CÍL PRÁCE

V současné podobě je uspořádání regionů ČR poměrně novou záležitostí. Kraje jako vyšší územní samosprávné celky byly ustanoveny již v dřívější době, kdy ekonomický systém centrálního plánování aplikoval postupy vedoucí ke snižování regionálních disparit. Uplatnění aktivní politiky přerozdělování a alokace příjmů vedlo k prvotnímu snižování krajových nerovností, avšak druhotná reakce na tyto aktivity byla nevhodná koncentrace těžebního a elektrárenského průmyslu, což disparity na úrovni krajů naopak prohloubilo.

K vyzdvížení důležitosti krajského členění došlo až s plánovaným vstupem do Evropské unie, kdy na podzim roku 2000 došlo ke zvolení prvních krajských zastupitelstev. Od roku 2001 operují v ČR tzv. regiony soudržnosti (úrovně NUTS II) – seskupení jednoho až tří krajů

– vybavené vlastními pravomocemi a orgány na základě zákona. Ty byly vytvořeny v rámci kompatibilního územního systému pro účely čerpání prostředků EU v rámci hospodářské a sociální soudržnosti. Vymezování nového regionálního uspořádání ČR bylo předmětem dlouhých diskusí.

Aby bylo zamezeno možnému přesouvání obyvatelstva do ekonomicky či sociálně silnějších regionů začala se Evropská unie silněji soustředit na regionální politiku. Regionální politika má za cíl zlepšit situaci v méně příznivých regionech a zabezpečit tak vývoj, jenž přenesse smazávání propastných rozdílů mezi některými regiony. Nezbytným zdrojem informací pro všechny subjekty vyhodnocující aktuální situaci by pak měla být sociálně-statistická analýza dat, která by měla zajistit objektivní informace pro racionální rozhodování o podpoře jednotlivých regionů. Vícerozměrné statistické analýzy jsou skupinou metod zabývajících se hodnocením jednotlivých regionů pomocí velkého množství proměnných, umožňují tak sdružování regionů do navzájem příbuzných celků, které se pak dají snáze objektivně ohodnotit při přidělování finančních podpor.

CÍL A METODY

Cílem práce je zhodnotit možnosti využití vícerozměrných statistických metod při hodnocení regionální diferenciace v jednotlivých oblastech České republiky. Uvedené hodnocení je zřejmě nejlépe uplatnitelné právě při hodnocení sociálních rozdílů. Úkolem je pouze nalézt vhodné proměnné pro hodnocený model.

K provedení zmíněných analýz je nutné získat vhodná data. Data byla získávána z publikací Českého statistického úřadu a Ministerstva pro místní rozvoj. Důležitou podmínkou prováděných analýz je komplexnost. Výsledky shrnující celou řadu proměnných umožňují objektivnější posouzení současného stavu i možností další potencionálního rozvoje regionů. Pro tyto účely je v regionální analýze možno využít široké palety tzv. vícerozměrné statistické metod. Vícerozměrné statistické metody zahrnují celou řadu algoritmů, mezi něž patří například metoda hlavních komponent, faktorová analýza nebo další skupina zahrnující shlukovou a diskriminační analýzu. Široce uplatnitelnou a vhodnou metodu pro posuzování rozdílnosti regionů se stává shluková analýza. Jejím účelem je rozložení souboru objektů na několik relativně stejnorodých podmnožin tak, aby objekty patřící do téhož shluku si byly co nejvíce navzájem podobné, zatímco objekty z různých shluků se od sebe co nejvýrazněji odlišovaly. Každý objekt (kraj) je popsán více znaky (proměnnými). Podmínkou je pak, aby jednotlivé zvolené proměnné byly nezávislé na jednotkách měření.

Výsledek shlukové analýzy závisí na zvoleném počtu proměnných (znaků), na zvoleném vyjádření podobnosti jednotek a na použitém způsobu shlukování.

Zvolený počet proměnných závisí na rozsahu dostupných dat a na jejich struktuře. Zde je nutno zmínit, že shluková analýza požaduje statistickou nezávislost (nekorelovanost) zvolených proměnných. Avšak na druhé straně stojí všeobecné vztahy platné v rámci reálného působení sil skutečného světa. Zde je nutné sledovat duplicitně zvolené proměnné, které mohou přinést značné zkreslení analýzy.

Vyjádření podobnosti jednotek se označuje jako měření vzdáleností mezi objekty. I zde lze uplatnit celou řadu použitelných metrik, kde mezi nejpoužívanější patří euklidovská vzdálenost nebo Pearsonův koeficient korelace.

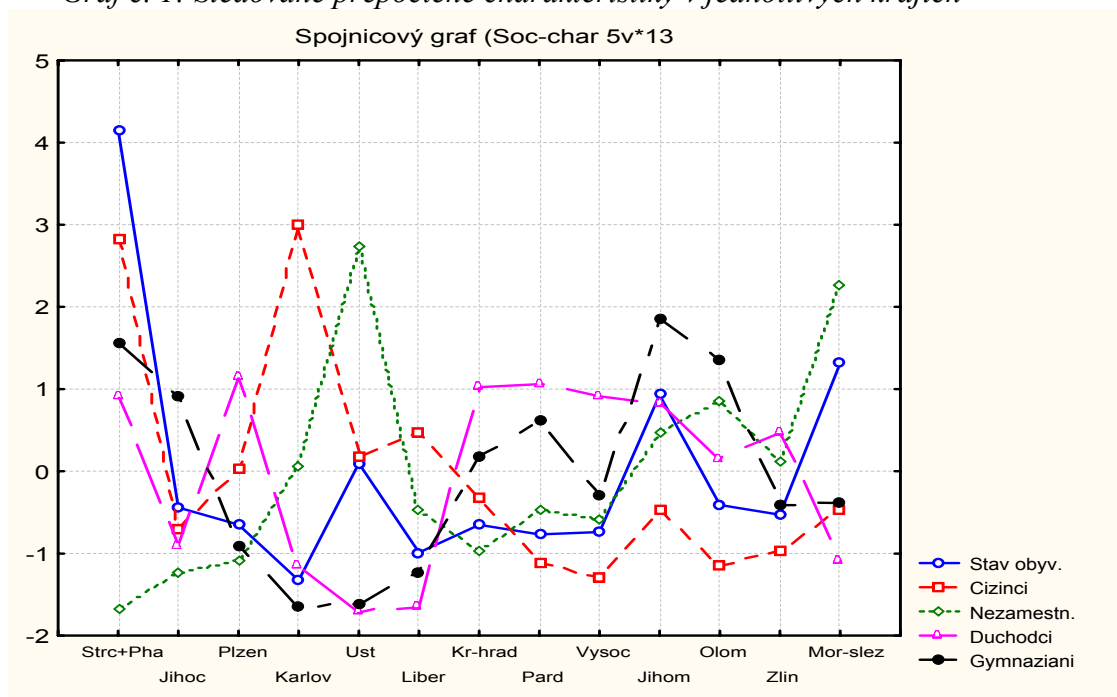
Způsobem shlukování je zamýšlena volba vhodné shlukovací procedury. Sem patří metoda průměrová, metoda centroidní, metoda nejbližšího souseda, metoda nejvzdálenějšího souseda, metoda mediánová, Wardova metoda a další.

VÝSLEDKY A DISKUSE

Shlukovací procesy krajů České republiky byly provedeny pomocí počítačového zpracování. Uživatelsky nejpříjemnější prostředí v tomto ohledu nabízí statistický software Statistica 6.0. Ještě před samotnou analýzou je nutné provést výběr ukazatelů z oblasti sociálních ukazatelů a zhodnotit vhodnost použití jednotlivých charakteristik.

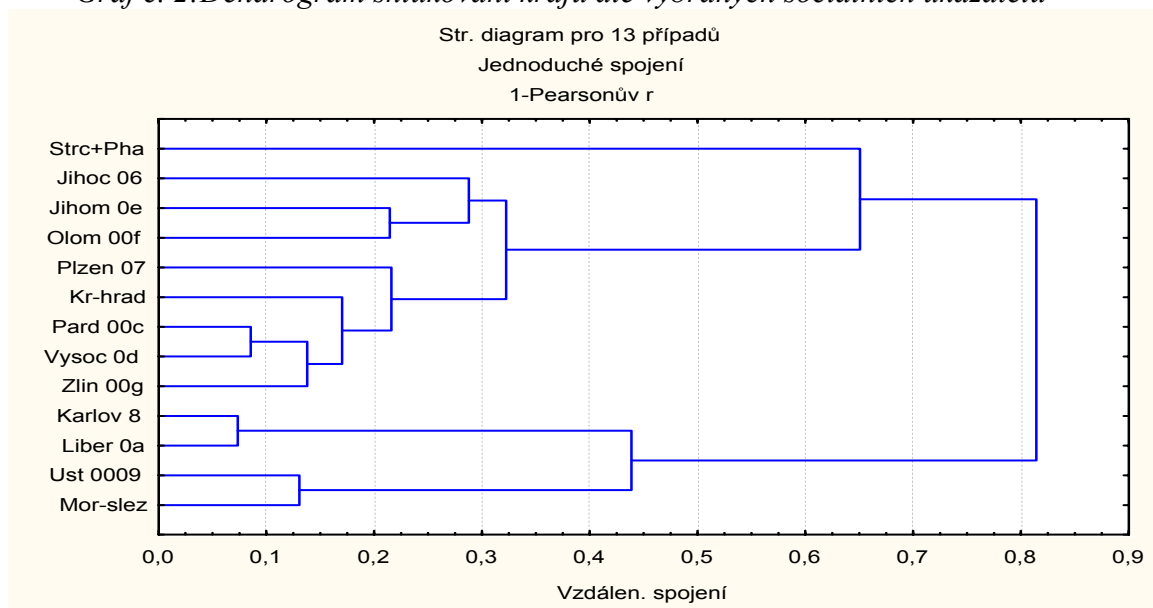
Statistické databáze v České republice nabízejí širokou paletu ukazatelů sledujících sociální úroveň a to i na úrovni krajů. Při výběru vhodných proměnných je nutné dodržovat zmíněnou nekorelovanost, která se sleduje pomocí vypočtu korelační matice, kde vhodné ukazatele by neměly přesahovat hodnotu 0,5. Z vybraných ukazatelů bylo asi polovinu nutno vyřadit pro silnou vzájemnou závislost. Pro analýzu byly zvoleny následující ukazatele: střední stav obyvatel (tis. osob), cizinci s délkou pobytu nad jeden rok, registrovaná míra nezaměstnanosti (%), počet příjemců starobních důchodů a žáci denního studia na gymnáziích. Všechny ukazatele jsou sledovány za rok 2002. Jednotlivé ukazatele bylo nutno nejdříve upravit. V první fázi došlo u absolutních ukazatelů k přepočtení na 1000 obyvatel a v druhé fázi pak byla provedena standardizace pomocí průměru a průměrné odchylky (tzv. transformace), kde jednotlivé údaje byly odečteny od průměru a poté vyděleny jejich průměrnou odchylkou. Uvedené přepočtené charakteristiky jsou uvedeny v následujícím grafu č. 1.

Graf č. 1: Sledované přepočtené charakteristiky v jednotlivých krajích



Zvolené charakteristiky vykazují značnou variabilitu a na první pohled z uvedeného grafu nelze poznat podobnost jednotlivých krajů. Velice snadno a rychle uplatnitelnou metodou pro shlukování je grafický výstup v podobě stromového grafu neboli dendrogramu. Nejvíce si podobné kraje jsou spojeny v dendrogramu nejdříve a postupně se přidávají stále vzdálenější regiony. Přehled spojování je zachycen v následujícím grafu č. 2. Nevíce podobné kraje se spojily jako první - Karlovarský a Liberecký, naopak nejodlišnější až v končných fázích. V předposledním kroku se teprve připojuje Středočeský kraj + Praha. Zde je nutno podotknout, že z čistě praktických důvodů byla Praha sloučena se Středočeským krajem, aby bylo zamezeno výraznému vychýlení hlavního města, které se stejně ač v menší míře projevovalo.

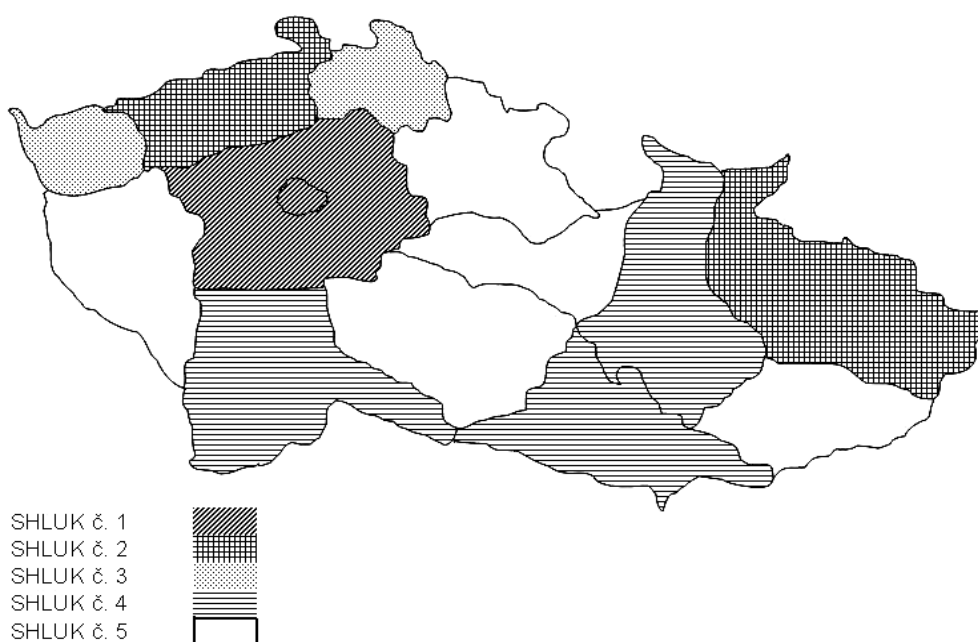
Graf č. 2: Dendrogram shlukování krajů dle vybraných sociálních ukazatelů



Z uvedeného dendrogramu (graf č. 2) lze vyčíst, že jako metrika byl využit místo obvyklé euklidovské vzdálenosti, méně využívaný Pearsonův koeficient korelace.

Poslední fází analýzy je spuštění shlukovacího procesu, při kterém je nutno zvolit metodu pro sdružování. Zde byla vybrána metoda pomocí k-průměrů a počet shluků. Jako nejvhodnější bylo vybráno pět výsledných shluků, které sdruží kraje dle příbuznosti v oblasti sociálních charakteristik. Výsledné shluky jsou pak rozděleny do pěti oblastí: první tvoří samostatně Středočeský kraj + Praha, druhý shluk tvoří velice odděleně kraje s těžebním průmyslem, další shluk tvoří zbytek severních krajů, poslední dva shluky jsou rozděleny na jižní a střední oblast ČR. Výjimku tvoří Olomoucký kraj přiřazený k Jihočeskému a Jihomoravskému. Velice přehledným zobrazením je vynesení výsledných shluků do kartogramu.

Graf č. 3: Kartograf typologie ČR podle sociálních ukazatelů



Uvedenou shlukovou analýzou byla nastíněna možnost rozčlenit kraje české republiky do menšího počtu větších celků. Vzniklé shluky by měly obsahovat sobě nejpodobnější kraje a zároveň dobře rozlišitelné regiony podle zvolených sociálních charakteristik. V zájmu vyšší objektivity by bylo nutno dohodnout a ujednotit výběr ukazatelů, které by kraje do těchto celků rozčlenil, aby tak byly alespoň částečně vytěsněny subjektivní vlivy.

ZÁVĚR

Současný vývoj související se vstupem České republiky do Evropské unie si žádá komplexní přístup k hodnocení diferenciací regionů. Nestačí jen sledovat jednotlivé charakteristiky, ale je nutné najít metody, které souhrnným způsobem zhodnotí diferenciaci mezi regiony. Česká republika nemá snad žádný komplexně zaostalý region. Současné regionální rozdíly plynou z transformačních procesů, směřování k tržní ekonomice a jednotnému evropskému trhu. Hlavními zdroji regionálních disparit zůstávají: restrukturalizace průmyslu v některých krajích, významné rozdíly mezi okresy, prohlubující se nevýhodnost venkovského prostředí proti městskému (podpora podnikání, věková struktura), nedostatečná dálniční síť v moravské oblasti, nerovnoměrné zastoupení vysokoškolsky vzdělaných a silně narušené životní prostředí v některých oblastech. Nové trendy směřují k aktivnímu snižování diferencovaných regionů v Evropě. Strukturální politika by měla pomoci hlavně nově vstupujícím zemím, je však nutné zajistit cílovost a transparentnost požadovaných dotací z evropských fondů.

Literatura:

1. Flury, B., Riedwyl, H.: Multivariate statistics. A practical approach. Chapman and Hall, London, 1998.
2. Hair, J. F., Anderson, R. E. and coll.: Multivariate Data Analysis With Readings. Prentice Hall, New Persey, 1995.
3. Hebák, P., Hustopecký, J.: Vícerozměrné statistické metody s aplikacemi. Praha, SNTL/ALFA, 1987.
4. Hindls, R.-Hronová, S.-Sege, J.: Statistika pro ekonomy. VŠE, Praha 2002.
5. Macháček, O.: Základní výpočetní postupy vícerozměrné analýzy. Praha, ČZU, 1987.
6. Meloun, M., Militký, J.: Kompendium statistického zpracování dat. Praha, Academia, 2002.
7. Moravová, J.: Základy sociální statistiky. VŠE, Praha 1998.
8. ČSU: Porovnání krajů – vybrané ukazatele 1995 – 2000. Praha, 2001.
www.czso.cz
www.mmr.cz
www.statsoft.cz

Kontakt:

Jakub Hlousek, ČZU Praha, hlousek@pef.czu.cz, tel: +420 224383246