SBORNÍK ČESKOSLOVENSKÉ GEOGRAFICKÉ SPOLEČNOSTI
ROČNÍK 1990 • ČÍSLO 3 • SVAZEK 95

JAN ŘEHÁK

VARIABILITA PROSTOROVÉHO ROZLOŽENÍ ČETNOSTÍ

J. Řehák: *Variability of Spatial Frequency Distribution.* — Sborník ČSGS, 95, 3, p. 186—194 (1990). — A measure of spatial variability (called geostructural variance) is defined for a frequency distribution on a finite set of places in a space whose geographical relations are assessed by a matrix of (generally conceived) distances. A set of measures stemming from the same approach describe the positions and properties of individual places in the geostructure. This complex of characteristics provides a clear-cut way of an analytical diagnostic reflection of the spatial properties of frequency distributions.

Studium vlastností statistických distribucí je závažným krokem v poznávání geografické reality. V tvarech a vlastnostech datových rozložení se promítají a projevují mechanismy vzniku dat, procesy vzniku jevů, jejich organizace a závislosti, prostorové i jiné zákonitosti ap. Nejdůležitějším pojmem ve statistické analýze dat je *variabilita* (rozmanitost, rozptýlenost, vnitřní heterogenita, diferencovanost) datového souboru. Je velmi důležité určit míry variability tak, aby *odpovídaly analytické situaci* (typ dat, návaznost metod) a odrážely potřebné aspekty rozptýlenosti, které chceme zachytit. Z měr variability odvozujeme i další charakteristiky jako míry polohy, asociace, podobnosti apod. V praxi používáme celou řadu takových měr, z nichž nejznámější pro číselná data je *rozptyl (variance)*, resp. *směrodatná odchylka* a pro kvalitativní data (četnosti v kategoriích nominální proměnné) pak *entropii* nebo *Giniho míru koncentrace (nominální variance).* Závažnost těchto měr spočívá v tom, že smyslem explanačních statistických postupů je charakterizovat rozmanitost světa pomocí dat a tuto rozmanitost vysvětlit, nalézt faktory, které ji působí, odlišit je, uspořádat co do intenzity působení a vlivu.

V geografické analýze dat je nutno běžné přístupy obohatit o míry, které odrážejí *prostorové aspekty rozptýlenosti dat.* Pro četnosti výskytu jevů v prostoru (a extensitní ukazatele) je vhodné (a někdy nutné) vztáhnout údaje k ploše území, na němž jevy byly zjištěny, resp. k jiné míře charakterizující velikost. Tímto problémem se zabýval například Batty (1, 2), který zavedl *míru inversní prostorové entropie,* vycházející z testových charakteristik dobré shody teorie informace. Nejde však o míru variability v běžném striktním smyslu, ale o míru podobnosti mezi zjištěnými četnostmi a zvolenými měrami velikosti územních jednotek.

Cílem této práce je zavést míru prostorové variability rozložení četností pro situaci, v níž bereme v úvahu *vzdálenosti* (nebo jiné *prostorové relace* či *obecnější geografické míry nepodobnosti míst).* Navržený model vychází z prací Řehák (7), Řehák—Řeháková (8—11) a poskytuje **řešení těchto statistických úloh:**

a) charakteristika prostorové variability rozložení četností, měření variability, odstupňování jednotlivých míst podle centrality, hodnocení stupně autokorelovanosti vzhledem k dané vzdálenostní struktuře;

b) komparace variabilit různých rozložení (různých jevů, resp. různých kontextů vzniku) v daném území a komparace vlivu různých vzdálenostních struktur vzhledem k variabilitě výskytu určitého jevu;

c) měření asociace výskytovosti různých jevů v daném území a komparace asociací v různých územních soustavách (resp. strukturách), a to včetně měření parciálních asociací;

d) faktorová analýza, typologie, uspořádání vlivů u prostorově rozložených četností včetně grafických vyobrazení vztahů mezi jednotlivými distribucemi navzájem a zobrazení vlivu odvozených faktorů na typy jevů i na místa územní struktury;

e) hierarchizace prostorové struktury uvažovaných míst vzhledem k vnitřní homogenizaci částí;

f) statistické testování hypotéz o autokorelovanosti, shodě četností s předpokladem (apriorně formulovanou hypotézou), shodě dvou i více distribucí, shodě závislých distribucí (např. v časovém posunu, jsou-li známy četnosti přechodů nebo v párovaných jevech dvou kontextů — např. migrace).

V tomto článku je uvedeno, vzhledem k omezení místa, pouze řešení úlohy ad a) — jde o základní stavební kámen celého modelu (budeme jej nazývat *distančním modelem* nebo *D-modelem*). Jsou tu uvedeny pouze aplikační vzorce. Situaci vhodnou pro použití navrženého postupu lze charakterizovat stručně: zjišťuje se četnost výskytu určitého jevu (nebo více jevů) v určitých územních útvarech, přičemž se v analýze dat bere v úvahu i vzájemná poloha těchto útvarů vyjádřená vzdálenostmi zvoleného typu. K tomuto cíli budou v první části uvedeny některé pojmy, v druhé části vybrané jednoduché míry s popisem vlastností a ve třetí numerická ilustrace.

1. Základní pojmy modelu

Analýzy dat se účastní tři složky, které vytváří model situace:

1. *Množina prvků zemského povrchu,* kterou označíme $A = \{A_1, A_2, \ldots, A_K\}$, K = počet prvků.

2. *Vzdálenosti dvojic prvků*, $d(A_j, A_k) = d_{jk}$, které jsou uspořádány do matice vzdáleností $D = \|d_{jk}\|$, typu KxK, a které vyjadřují prostorové relace mezi prvky.

3. *Četnosti výskytu* jevu v jednotlivých místech, které označíme: f_k = relativní četnost výskytu $f_k = n_k / \Sigma\, n_a$, n_a = absolutní četnost; rozložení četností v místech se značí $f = (f_1, f_2, \ldots, f_K)$.

Nazveme:

a) *nosičem rozložení* množinu $A = \{A_1, \ldots, A_K\}$, která tvoří soustavu míst;

b) *diskrétní geostrukturou* (geografickou strukturou, diskrétní geografickou, územní, areální ap. proměnnou) dvojici $G = \{A; D\}$, resp. $G = \{A_1, A_2, \ldots, A_K; D\}$, tj. množinu prvků geografické soustavy a prostorové relace na ní;

c) *prostorovým rozložením četností* (rozložením četností na geostruktuře) trojici $\{f, A, D\}$.

Veškeré výsledky D-modelu jsou vždy závislé na třech uvedených složkách a veškerá interpretace tak musí být vztažena k nim a jejich významu. Geostruktura je chápána jako uzavřený systém a vše se hodnotí vzhledem k této uzavřenosti (to je dáno jednak výčtem míst v A a jednak normalizací četností vzhledem k A). Specifikací A, D a f lze model aplikovat na nejrůznější typy problémů. Tyto tři složky zasluhují alespoň stručný rozbor.

A) Množina prvků geografického povrchu A může být (v abstraktním slova smyslu či v mapovém vyjádření) množinou

bodů: sídla, měřící stanice, křižovatky silnic, soutoky řek, železniční stanice, umístění obchodů, stanic metra, závodů apod.;

křivek: síť silnic, železnic, řek, izočáry, hranice územních celků, trasy technické infrastruktury apod.;

ploch: administrativně-správní celky určité úrovně, soustava vodních ploch, soustava pásů povodí, lesů, určených ekosystémů, chráněná území, kvadráty náhodného výběru, kvadrátová úplná síť, města se spádovými oblastmi.

Plochy navíc mohou vznikat dekompozicí určitého území nebo mohou být nespojené v souvislý územní celek. Předpokládáme, že jednotlivé prvky A jsou pevně dané, v případě náhodného vzniku (ať už přirozenou cestou nebo výběrem) od aspektu náhodnosti odhlížíme.

Většinou se v analýzách budou vyskytovat vždy prvky jen jednoho z uvedených typů (i když to není nutnou podmínkou).

B) Prostorové relace jsou charakterizovány vzdálenostmi $D = \|d_{ik}\|$, které mohou mít velmi rozmanitý charakter. Určením typu vzdálenosti přijímáme model struktury, geografické hledisko a přístup k analýze. Přijetí matice D je apriorní — je to přijetí pohledu na reálnou hodnocenou skutečnost.

Požadavky na matici D:

(a) identita: $d_{ii} = 0$

(b) nezápornost: $d_{ij} \geqslant 0$

(c) symetrie: $d_{ij} = d_{ji}$

(d) interpretabilita: $d_{ij} < d_{mk}$ značí, že dvojice prvků A_k, A_m je od sebe vzdálenější než dvojice A_i, A_j.

Není tu požadována trojúhelníková nerovnost, to znamená, že matice D může vyjadřovat vztahy z podstatně širší třídy možností, než jsou geometrické vzdálenosti — pro naše aplikace to jsou především funkční vzdálenosti, náklady, ztrátové funkce, čas, bariéry psychologického rázu, subjektivně určované vzdálenosti atp. Od geometrických představ je možno ustoupit ještě dále tím, že uvolníme předpoklady: (a) povolením kladných vzdáleností jednotky samy k sobě (např. u plošných jednotek či křivkových prvků); (c) povolením asymetrie (např. v situaci jednosměrných dopravních tahů, objížděk, převládajících dopravních tahů v době špiček, vzdáleností po a proti proudu, převládajících směrů větru ap.). Tyto odchylky jsou možné, avšak je nutno změnit některé vzorce a rozšířit statistické míry (vzhledem ke vzdálenostem k a od); v tomto článku předpokládáme, že předpoklady (a) — (d) jsou splněny.

C) *Rozložení četností* $f = (f_1, f_2, \ldots, f_K)$ nepředpokládá pro deskriptivní využití modelu žádné specifické vlastnosti. Pro testování hypotéz vycházíme v některých případech z multinomického rozložení absolutních četností $(n_1, \ldots n_K)$, resp. nezávislých poissonovských rozložení veličin n_k, $k = 1, 2, \ldots, K$, majících parametry λ_k, resp. obecněji $\lambda_k = \lambda(a, b_k, \ldots)$. Místo rozložení četností mohou vystupovat jako f jakákoli rozložení spojitých proměnných, u nichž má podíl na celkové sumě interpretační význam (spotřeba potravin nebo surovin, podíl na výrobě, výstavbě, vývozu, dovozu, na investicích, službách, bytech, ploše lesů, vod, orné půdy, úsporách, výdajích ap.).

2. Měření prostorové variability — geostrukturní variance a centrality míst

A. Pojem geostrukturní variance

Geostrukturní variance je geografická specifikace pojmu *zobecněné variance*, Gvar f, která byla zavedena v práci Řehák (7) jako zobecnění myšlenky Corranda Giniho, který definoval rozptyl jako průměrný součet čtverců rozdílů pro všechny dvojice souboru, $\mathrm{var}X = 2\Sigma\Sigma(x_i - x_j)^2 / (n(n-1))$. Využití pojmu bylo rozpracováno v pracech Řehák—Řeháková (8—11). Později byl pojem Gvar f zaveden C. R. Raem (5—6) a nazván kvadratickou entropií. V geostrukturním pojetí budeme pokládat za vzdálenost výskytu dvou jevů v prostoru (na A) vzdálenost míst, ve kterých se vyskytly, tj. d_{ij}, vyskytly-li se v A_i, A_j. Dva jevy ve stejném místě mají nulovou vzdálenost.

Geostrukturní varianci (resp. *kvadratickou entropii*) definujeme jako průměr všech vzdáleností dvojic jevů:

(1) $$\mathrm{Gvar}\, f = \Sigma\Sigma\, f_j\, d_{jk}\, f_k\,.$$

Tato míra má vlastnosti (za předpokladů (a) — (d) pro D):

1. Jsou-li všechny jevy soustředěny v jednom místě, je Gvar f = 0;
2. Jsou-li všechny vzdálenosti d_{jk} pro různá místa A_j, A_k nenulové, platí: Gvar f = 0 implikuje soustředění všech jevů v jednom místě;
3. Gvar f = 0 implikuje v obecném případě, že všechny jevy jsou rozmístěny jen v místech, mezi nimiž byly zavedeny nulové vzdálenosti;
4. Pro rovnoměrné rozložení je hodnota $\mathrm{Gvar}\, f = \Sigma\Sigma\, d_{jk} / K^2$.

Geostrukturní variance odstupňuje variabilitu (tj. vnitřní statistickou heterogenitu) výskytu jevů. Tím umožňuje porovnávat různé distribuce na téže geostrukturní soustavě. Prostorová koncentrace jevů odpovídá nízkým hodnotám Gvar f, vysoké hodnoty Gvar f odpovídají maximálnímu prostorovému oddělení výskytů po obvodu geostruktury. Všimněme si, že D-model je v „inversním" vztahu ke gravitačním modelům, u nichž roli d_{jk} v obdobném součtu součinů má $1/d_{jk}$. Interpretace a použití je ovšem v obou případech různé.

Výpočet max Gvar f má význam v tom, že poskytuje horní hranici, tj. nejvýše možnou numerickou hodnotu měření, ke které lze variabilitu poměřovat. Zároveň umožňuje *normalizaci geostrukturní variance:*

(2) $$\mathrm{norm\ Gvar}\, f = \frac{\mathrm{Gvar}\, f}{\max \mathrm{Gvar}\, f}$$

Ta umožňuje komparaci relativní variability vzhledem k maximu pro různé geostruktury ať již definované různými maticemi D pro stejný nosič A, nebo pro různé nosiče (a tím samozřejmě i pro různá D).

Dalším důležitým a zajímavým referenčním bodem škály je Gvar rovnoměrného rozložení, $\Sigma\Sigma d_{ij}/K^2$: nižší hodnoty indikují *geografické soustředění*, vyšší *geografickou polarizaci.*

Variabilita rozložení četností v místech $A = \{A_1, A_2, \ldots, A_K\}$ bez ohledu na přijatou distanční strukturu D je měřena entropií, nebo (v našem kontextu lépe) *Giniho mírou koncentrace:*

$$(3) \qquad G = 1 - \Sigma f_k^2,$$

jež má $\max G = (K-1)/K$, dosažené pro rovnoměrné rozložení a jež lze normovat jako $KG/(K-1)$.

Vztažením Gvar f ke G je možno měřit, obdobně jako Moranovým I a Gearyho c, v nichž se používá místo D matice W, jejímiž prvky jsou zvolené míry blízkosti (indikace sousedství, indikace okolí, určená klesající funkce vzdálenosti apod.) *prostorovou autokorelovanost* indexem:

$$(4) \qquad g = \frac{\text{Gvar f}}{(\Sigma\Sigma d_{ij})(1-\Sigma f_k^2)/(K(K-1))} = \frac{\text{Gvar f}}{\bar{d}(1-\Sigma f_k^2)}$$

kde $\bar{d}$ je průměrná nediagonální vzdálenost z matice D.

Pro index platí:

a) v případě absence prostorové autokorelovanosti je $g = 1$, $g > 1$ indikuje *prostorově podmíněnou zvýšenou heterogenitu*, $g < 1$ naopak indikuje *prostorovou koncentraci* výskytů jevů,

b) minimum závisí na vlastnostech matice D,

c) jmenovatel g je permutační očekávaná hodnota Gvar f za předpokladu absence prostorové autokorelovanosti a Gvar f je speciálním případem permutační statistiky (Hubert et al., 4),

d) permutační testy prostorové nezávislosti výskytů lze provést obvyklým způsobem buď přesným zjištěním permutačního rozdělení pro malá K nebo pomocí simulačního odhadu rozložení statistiky (metoda Monte Carlo), popřípadě aproximací této distribuce.

B. Pozice míst v geostruktuře

Vzhledem k roli v prostorovém rozložení četností je pozice míst modelem hodnocena v rámci systému geostruktury $G = \{A; D\}$. Místa jsou posuzována podle *f* a podle určených distancí D. Proto tu není uvažován vztah mimo hranice A a není řešen tzv. hraniční problém, který bere v úvahu efekty autokorelace působící z vnějších sousedních územních útvarů. (Jednou z možností rozšíření je zavedení kategorie „vnější prostředí“ mezi A_k a určení vhodné vzdálenosti pro každou z vlastních kategorií systému — to může být užitečné např. pro problémy migrace, zásobování apod.) Vztah jednotlivých A_k vzhledem k celku G, tak jak je dán rozložením četností f, můžeme charakterizovat *mírou centrality* A_k *(resp. mírou koncentrace rozložení kolem* A_k), která vyjadřuje průměrnou vzdálenost všech výskytů od místa A_k:

$$(5) \qquad d_k = \sum_j d_{kj} f_j$$

Tuto charakteristiku zjišťujeme pro všechny prvky A_k a dostaneme tím *vektor centralit*

$$(6) \qquad \mathbf{d} = (d_1, \ldots, d_K).$$

Ten prvek, pro nějž je d_k nejmenší, má v průměru nejblíže ke všem výskytům jevů. Nazveme jej *centrem rozložení* f na geostruktuře G = = {A, D}.

(7) $C = A_c$ je místo, pro něž $d_c = \min d_k$.

Center rozložení může k danému f existovat několik, dosahujeme-li centralita d_k svého minima ve více případech. Vektor centralit poskytuje první rychlý obraz o prostorových koncentracích výskytu jevu v rámci dané geografické soustavy. Míra $d^* = d_c$ může zároveň sloužit jako paralelní charakteristika geostrukturního rozptýlení (v interpretaci bereme G jako chybu v modelu proporcionální predikce a d^* jako chybu při optimální predikci).

Vlastnosti centralit:

a) d_k má hodnotu nula, jsou-li všechny jevy soustředěny v A_k (pak je též Gvar f = 0, $d_j = d_{kj}$ pro všechna ostatní A_j), nebo jsou-li jevy v místech, kterým byla přisouzena nulová vzdálenost od A_k;

b) d_k nabývá maximální hodnoty, jsou-li jevy plně soustředěny u prvků A_j s maximální vzdáleností od A_k (může jich být více);

c) platí vztah $d^* \leq$ Gvar f;

d) geostrukturní variance je průměrná koncentrace: Gvar f = $\Sigma f_k d_k$.

Míry d_k můžeme také normalizovat vzhledem k jejich maximálně dosažitelné hodnotě:

(8) $$\bar{d}_k = \frac{d_k}{\max\limits_{j} d_{kj}}$$

Tyto míry můžeme nazvat *měrami separability* A_k *od výskytu jevů.* Souhrnnou informaci shrnuje *vektor separability* $\bar{d} = (\bar{d}^1, \ldots, \bar{d}_K)$. Min $\bar{d}_k$ určuje místo s nejmenší separabilitou (nemusí to být centrum), max $\bar{d}_k$ ukazuje na místo s nejvyšším stupněm dosažení separability.

Vzhledem k tomu, že Gvar f je průměrem centralit, můžeme zavést další užitečné charakteristiky prvků geostruktury vzhledem k rozložení f:

a) index příspěvků místa k prostorové heterogenitě

(9) $$t_k = \frac{d_k}{\text{Gvar } f}$$

vyjadřuje vztah koncentrace prvků kolem A_k a prostorové vzdálenosti. Platí pro něj: $1/2 \leq t_k < \infty$. Je-li

$t_k = 1$, A_k přispívá průměrně

$t_k > 1$, A_k přispívá nadprůměrně k heterogenitě

$t_k < 1$, A_k přispívá nadprůměrně ke koncentraci, jestliže A_k z G vyjmeme, tak se podle hodnoty t_k normalizovaný Gvar f buď nemění, sníží nebo zvýší. Vektor $t = (t_1, t_2, \ldots, t_K)$ tak určuje užitečnou klasifikaci míst pro hodnocení distribuce.

b) *normalizovaný index příspěvku místa k prostorové heterogenitě*

(10) $$u_k = \frac{\text{Gvar } f}{2d_k}$$

vyjadřuje inversně informaci indexu t_k; je normalizován na interval (0,1).

c) *podíl příspěvku místa na geostrukturní variabilitě*

$$(11) \qquad u_k = \frac{f_k\, d_k}{\mathrm{Gvar}\, f}$$

(popřípadě v %) ukazuje na část heterogenity, vážící se k A_k.

Podle analytických potřeb můžeme zavést další míry a charakteristiky míst geostruktury, například *míru stupně úměrnosti mezi vzdálenostmi míst od A_k a četnostmi:*

$$(12) \qquad \bar{r}_k = \frac{\Sigma d_{kj} f_j}{\sqrt{\Sigma d_{kj}^2\, \Sigma f_j^2}} = \frac{d_k}{\sqrt{\Sigma d_{kj}^2\, \Sigma f_j^2}}$$

nebo korelační (Pearsonův nebo neparametrický) koeficient r_k pro proměnné „vzdálenosti od A_k" a „četnosti". Hodnoty vektoru $\bar{r} = (\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_k)$ nabývají hodnoty $\bar{r}_k = 1$, jsou-li všechny četnosti f_j úměrné hodnotám d_{kj}; naopak $\bar{r}_k = 0$, jsou-li všechny jevy od A_k vzdáleny nulově. Vysoká hodnota tak indikuje, že A_k je nízko obsazeno a četnosti stoupají ve vzdálenosti od A_k. To může znamenat buď umístění v prázdném středu" geostruktury, u níž se četnosti rozmísťují na okrajích, nebo unimodální, prostorově asymetrickou distribuci se špičkou u opačné hranice od A_k.

Uvedené míry poskytují charakterizaci prostorového rozložení četnosti: Gvar f a popřípadě d* vyjadřují stupeň variability (analogie rozptylu, entropie, Giniho míry koncentrace), centrum označuje střed rozložení (analogie průměru, mediánu, módu) a míry centrality a z nich odvozené indexy charakterizují postavení jednotlivých míst (obdoba prostého rozložení četností). Pomocí těchto měr lze provádět analýzy různých typů:

a) prostého popisu vlastností četnostní distribuce,
b) porovnání prostorových míst v rámci geostruktury,
c) porovnání prostorových četnostních vlastností pro danou soustavu míst a různé distanční struktury,
d) porovnání prostorových četnostních vlastností pro geostruktury s různými nosiči.

Tyto úkoly mohou být dovedeny až do klasifikačních stádií (klasifikace míst v geostruktuře, klasifikace jevů na geostruktuře podle vlastností různých typů).

3. Závěr

Míry zavedené v článku obsahují různou informaci o rozložení i o pozici míst. Kvantitativní stránka sama o sobě je velmi důležitá, neboť modelově zajišťuje sjednocený pohled a exaktní charakterizaci; podstatné však je to, že *kvantifikace umožňuje prohloubení kvalitativního pohledu, obsahového porozumění a přesnější významové interpretace.* Vhodný systém měr může velice usnadnit práci s obsáhlým datovým materiálem, neboť kondenzuje informaci, zpřehledňuje ji a umožňuje uspořádání, klasifikace a komparace z toho plynoucí. Sebevhodnější model však zůstává pouhou abstrakcí a jeho praktické nasazení není efektivní, nejsou-li jeho výsledky využity s vysokou meritorní odborností a citem pro data, ať již jde o závěry analytického či syntetického charakteru, regionalizační úvahy, prognostiku ap. V jednotlivých analýzách nepoužíváme nutně všechny uvedené míry, nýbrž jen ty, které se pro daný účel a typ jevů hodí. Tyto míry mohou dostat také v rámci specifického kontextu úlohy i specifické názvy podle obsahové stránky problému.

Uvedený přístup D-modelu pro analýzu prostorových četnostních distribucí je jedním z možných přístupů k analýze dat v popsané situaci. Svoji užitečnost prokazuje již v článku uvedená část modelu, která se týká popisu vlastností četnostního rozložení. V plném rozsahu má model velmi široké analytické funkce zmíněné v úvodu. K širšímu využití v geografii a ve vědách pracujících s územními aspekty dat jej předurčují některé přednosti:

a) je rozvinut podle specifických potřeb geografických analýz;
b) je jednoduchý a intuitivně snadno pochopitelný — jednotlivé míry mají přirozený a logicky interpretovatelný význam a tudíž využívání nevede k chybám;
c) tvoří jednotný systém vzájemně navazujících a konzistentních metod v celé šíři úloh a) — e) uvedených v úvodu a tudíž eliminuje metodologické inkonzistence, které v běžné praxi působí ad hoc metody a různé základy používaných metodik a které v praxi analýzy dat odvádějí pozornost od vlastního soustředění na předmět;
d) umožňuje popis i testování hypotéz a mohou na něj také navázat skupiny metod jako je seskupovací analýza, mnohorozměrné škálování, analýza komparačních reziduí LINDA-K ad., pro něž model připravuje legitimní vstupy;
e) model ve své obecné formulaci obsahuje jako speciální případy běžné míry pro nominální a číselná kategorizovaná data a tudíž umožňuje porovnání výsledků z různých situací a různých specifikací modelů.

Paralelně k uvedenému diskrétnímu případu geostruktury lze vyvinout model pro spojitý nosič A (souvislé území). Přímá aplikace výsledků je však náročná v případě většího počtu pozorování jak na přípravu sběru, sběr i zakódování dat, tak na zpracování. V praxi je možné provést přechod od spojitého A k diskrétnímu modelu rozdělením plochy na přirozené anebo i uměle definované (např. čtvercová síť) celky a zjištění četností v částech. Tak není nutno zpracovávat územní určení každého jednotlivého jevu a určování všech vzájemných vzdáleností (které je pro případ, že nejsou zadány formulí obtížné).

D-model tak aspiruje na to, aby se stal jednou z běžně používaných metodik analýzy dat. Jeho význam se zvýší také po realizaci možnosti dialogického zpracování dat mikrotechnikou poskytující různé úpravy G (vynechání míst, změny vzdáleností, Monte Carlo simulace pro formulované hypotézy ap.) a zároveň přímou možnost vykreslení výsledků v mapkách.

Literatura:

1. BATTY, M.: Spatial Entropy. Geographical Analysis VI, 1974, s. 1—31.
2. BATTY, M.: Entropy in Spatial Aggregation. Geographical Analysis VIII, 1976, s. 1—21.
3. CLIFF, A. D. — ORD, J. K.: Spatial Autocorrelation. London, Pion Ltd, 1973.

4. HUBERT, L. J. — GOLLEDGE, R. G. — CONSTANZO, C. M. — GALE, N.: Measuring Association between Spatially Defined Variables: An Alternative Procedure. Geographical Analysis 17, 1985, s. 36—46.
5. RAO, C. R.: Diversity and Dissimilarity Coefficients: A Unified Approach. Theoretical Population Biology 21, 1982, s. 24—43.
6. RAO, C. R.: Analysis of Diversity: A Unified Approach. Statistical Decision Theory and Related Topies III, 2, 1982, s. 233—250.
7. ŘEHÁK, J.: Základní deskriptivní míry pro rozložení ordinálních dat. Sociologický časopis 12, 1976, s. 416—431.
8. ŘEHÁK, J. — ŘEHÁKOVÁ, B.: Basic Characteristics for Finite-Valued Variables and a Distance Analysis of Their Distributions. 9th World Congress of Sociology, Uppsala 1978.
9. ŘEHÁK, J. — ŘEHÁKOVÁ, B.: Základní charakteristiky proměnných s konečným počtem hodnot a distanční analýza a jejich rozložení. Sociologický časopis 12, 1976, s. 416—431.
10. ŘEHÁK, J. - ŘEHÁKOVÁ, B.: Klassifikacija s numeričeskimi sootnošenijami - D-model dlja analiza raspredělenij. 2. sovětsko-československý seminář „Analýza a modelování sociálního a ekonomického rozvoje oblastí". Kemerovo 1984.
11. ŘEHÁK, J. — ŘEHÁKOVÁ, B.: Classifications with Relations: A Model for the Distributions and Their Distances. Kybernetika 22, 1986, s. 158—175.

Summary

VARIABILITY OF SPATIAL FREQUENCY DISTRIBUTION

The measurement of *spatial variability* is derived from a general model called D-model. It is based on a notion of spatial frequency distribution (frequency distribution on geostructure). *Geostructure* is defined as finite set of geographical elements in space (points, lines, areas) of which a numerically assesed relation of distance (separation, dissimilarity) is given as follows: $G = \{A, D\}$, with $A = \{A_1 \ldots A_K\}$ being a set of places and $D = \|d_{jk}\|$ being a matrix of numerical characteristic which expresses spatial (or generally conceived) relations of geographical differentiation of places.

Geostructural variance (or quadratic entropy), Gvarf, represents the average distance of all pairs of events under study that occurred in f. It is defined in (1). The value of Gvarf reflects the distance structure and orders the distributions quite differently from the usual nominal approach of Gini's concentration measure (3) or measures based on entropy. Various characteristics are derived to provide the measurement of various aspects of spatial differentiation: *the normalized geostructural variance* (2) relates Gvarf to its attainable maximum, *the index of autocorrelation* (4) is an analogy to Moran's I or Geary' c and shows the standardized ratio of Gvarf to Gini's measure G, and the *ratio obtained variance to that of a uniform distribution.*

Positions of places (elements of the geostructure) in frequency spatial field are characterized by analogous principles. The measure of *centrality (concentration)* is defined in (5) as the average distance of all ocurrences from the place. *The centre* of the distribution is a place in A attaining the minimum value of centrality. Further properties of places are given by various ways of normalization, such as the proportion of the attained centrality to its available maximum (7), the comparison with Gvarf (9), (10), (11), the Cauchy-Schwartz inequality application yielding a measure of proportionality between frequencies and distances, and the linear (or nonparametric) correlation coefficient for a degree of linearity between frequencies and distances. They can provide a diagnostic information on places as regards the interpretation and practical application.

The model is presented only at its basic level that yields the description of spatial properties of a frequency distribution. Its advantage are the simple applied measures and their natural and clear-cut interpretation. Further steps of the D-model include comparative analyses and factor analyses, the clustering and scaling of spatial distributions, the decomposition of the geostructure into spatially homogenous parts, the hypotheses testing and the confidence intervals.

(Pracoviště autora: Socioekonomický ústav ČSAV, Na Zbořenci 3, 120 00 Praha 2.)
Došlo do redakce 22. 1. 1988. *Lektoroval Zdeněk Pavlík*