HYBRIDNÍ LOGARITMICKO-LINEÁRNÍ MODELY (APLIKACE NA MIGRAČNÍ DATA)

Jaromír Běláček

Idea hybridních logaritmicko-lineárních modelů byla navržena v článku Willekens-Baydar(1983) jako prostředek pro analýzu kontingenčních tabulek vázaných jistou speciální doplňující informací. Článek shrnuje obecnou metodologii a ukázky některých elementárních aplikací. Je střechou nad některými speciálními modely zmíněnými v pracech Goodman(1972), Bishop-Fienberg-Holland(1975), Haberman(1974) nebo podrobněji v Gokhale-Johnson(1978) . Tento příspěvek inspirovaný myšlenkou systematické softwareové realizace je zaměřen na aplikaci hybridních modelů při analýze migračních tabulek, kde faktor "sousedství" regionů hraje stejně závažnou roli jako "původní" a "koncové" místo přestěhovavších se migrantů.

Standardní loglineární modely jsou užitečnější pro vícerozměrné kontingenční tabulky. Na těch je možno rozvinout obvyklé testovací strategie pro výběr optimálního loglineárního modelu ze spektra marginálních-rektangulárních alternativ. Hybridní loglineární modely znamenají lepší přizpůsobení objektivním skutečnostem, které jsou vztaženy a priori ke struktuře dané kontingenční tabulky. Jsou použitelné samozřejmě také pro případ více dimenzí. Ale již ve dvoucestné tabulce lze obvykle specifikovat takové množství hybridních alternativ, které nelze z časových a technických důvodů v plném rozsahu verifikovat. Problém nalezení takové třídy hybridních modelů, jejichž testy dobré shody s daty by pokryly kritické hodnoty asymptotického χ^2 rozdělení na požadované hladině významnosti, lze řešit redukcí na systém submodelů, určený právě svojí věcnou interpretací. V tomto příspěvku je učiněn pokus navrhnout jako tento systém právě hybridní třídu "sousedských" alternativ.

Ve vztahu k metodologii tradiční geografické analýzy zobecňují hybridní modely ideu hledání teoretických počtů přemístění z místa jednoho regionu do druhého, které jsou počítány za účelem testování hypotéz o nezávislosti v migrační tabulce (ať již při znalosti nebo bez znalosti diagonály). V tomto příspěvku je výklad zaměřen na testování modelů, což je přístup ve statistickém smyslu ekvivalentní k testování hypotéz. Hybridní hypotézy jsou vlastně hypotézy o struktuře parametrů hybridních modelů, kterým je v dalším výkladu věnována pouze minimální pozornost. Prostřednictvím IPF (iterativ proportional fitting) algoritmu lze totiž v loglineárních modelech testovat hypotézy o vhodnosti modelů bez explicitní znalosti jim příslušných odhadů parametrů. Bude-li tento příspěvek shledán pro geografy jako užitečný, může být této problematice věnováno ještě více místa.

1. Elementární analýza

Migrační tabulka je kontingenční tabulka (n_{ij}) počtů migrantů, kteří mění svoje bydliště z regionu A_i (místa počátku) do regionu B_j (koncového místa) daného územního celku za určité časové období. V našem případě byl za počáteční i koncovou proměnnou zvolen systém všech deseti okresů severočeského kraje znázorněných na obrázku ($A_i = B_i$ pro $i=1,\dots,10$). Četnosti migrantů, kteří se přestěhovali z okresu do okresu v roce 1985 jsou uvedeny v následující tabulce. Zkrácené názvy kategorií v tabulce byly uspořádány tak, aby respektovaly geografické rozmístění okresů od západu na východ s týmiž okresy na hlavní diagonále:

A \ B	CHOM	MOST	LOUN	TEPL	LITC	USTL	DECN	CLIP	LIBE	JABN
CHOM	2393	329	103	65	59	68	40	58	59	13
MOST	241	2447	128	199	75	57	51	44	46	13
LOUN	184	214	957	77	71	51	46	33	22	14
TEPL	80	285	97	2694	77	320	83	87	50	15
LITC	77	81	77	119	1876	331	64	190	71	22
USTL	57	68	38	313	109	1276	211	109	85	43
DECN	55	68	23	71	81	190	1908	394	116	42
CLIP	20	42	22	40	108	61	170	1717	222	36
LIBE	26	49	15	37	48	63	81	318	1439	203
JABN	5	16	5	30	30	22	40	78	316	906

(1)

Konzervativní loglineární analýza nabízí pro tuto tabulku pouze dvě alternativy - - saturovaný model:

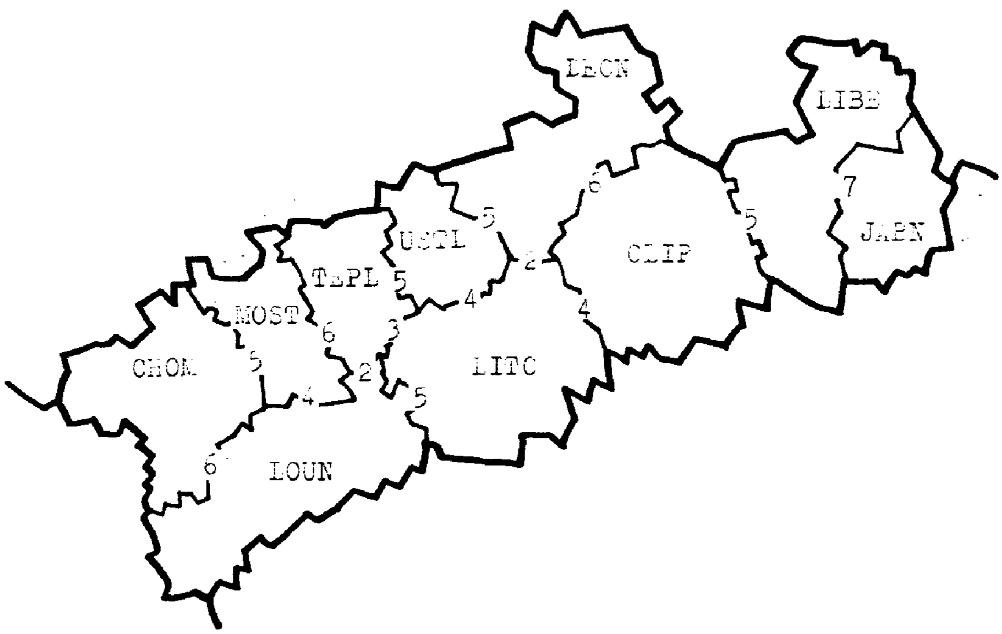
$$m_{ij} = w.w_i^A.w_j^B.w_{ij}^{AB} \quad , \quad i,j=1,\dots,10 \quad , \qquad (2)$$

kde m_{ij} jsou očekávané četnosti za platnosti modelu a w , w_i^A , w_j^B , w_{ij}^{AB} neznámé parametry. Saturovaný model reprezentuje tabulku prostřednictvím součinového-nezávislého vlivu všech rektangulárních efektů, které jsou nejpřirozenější z hlediska provedeného třídění (A resp.B - hlavní efekty vyjadřující čistý vliv počátečního resp. koncového okresu migrace; AB - interakční efekt párového vlivu dvojic okresů). Fakticky jde však pouze o formální rozklad, který doplněn o soustavu reparametrizačních podmínek

$$\prod_i w_i^A = \prod_j w_j^B = \prod_i w_{ij}^{AB} = \prod_j w_{ij}^{AB} = 1 \quad , \quad i,j=1,\dots,10 \quad , \qquad (3)$$

umožňuje explicitně odhadovat neznámé parametry. Protože odhady očekávaných četností jsou rovny $\hat{m}_{ij}=n_{ij}$ pro všechna i a j , je vyhlazovací schopnost tohoto modelu bezcenná;

Okresy severočeského kraje



(číslice identifikují stejný typ sousedství okresů podle přibližně stejné délky společné hranice)

- model nezávislosti:

$$m_{ij} = w.w_i^A.w_j^B \quad , \quad i,j=1,\dots,10 \quad . \tag{4}$$

Parametry w , w_i^A , w_j^B mají analogický význam jako v (2) posílený absencí interakčních parametrů ($w_{ij}^{AB}=1$). Odhady očekávaných četností se vypočtou podle

$$\hat{m}_{ij} = \frac{n_{i+}.n_{+j}}{n_{++}} \quad , \quad i,j=1,\dots,10 \quad , \tag{5}$$

pro $n_{i+}=\sum_i n_{ij}$, $n_{+j}=\sum_j n_{ij}$, $n_{++}=\sum_i\sum_j$, odhady dalších parametrů jsou (stejně jako u ostatních níže uvedených modelů) z hlediska modelového vyhlazení nepodstatné.

Testy kvality vyhlazení modelu jsou založeny na statistikách (viz Prášková(1985), str.30-31)

$$\chi^2 = \sum_i\sum_j \frac{(n_{ij}-\hat{m}_{ij})^2}{\hat{m}_{ij}} \quad \text{resp.} \quad G^2 = 2\sum_i\sum_j n_{ij}.\log(n_{ij}/\hat{m}_{ij}) \tag{6}$$

obou srovnatelných s kritickými hodnotami asymptotického χ_q^2 rozdělení na zvolené hladině významnosti. Počet stupňů volnosti q lze vyjádřit jako $q=h-r$, kde h je počet platných polí tabulky a r počet lineárně nezávislých parametrů. V případě modelu (4) tedy $q=81$ (viz Anděl(1978),str.211) a hodnoty testových statistik ze (6) jsou rovny $\chi^2=95536.1$ resp. $G^2=53703.9$, což jsou absurdně vysoké hodnoty statisticky významné na libovolné hladině významnosti. Smyslem tohoto článku je najít takové třídy hybridních loglineárních modelů, které by po formální i obsahové stránce korektně redukovaly hrubou nepřizpůsobivost modelu nezávislosti reálným datům.

2. Modely kvazinezávislosti

Letmý pohled na migrační tabulku (1) indikuje nápadně vyšší počty migrantů ve všech polích na hlavní diagonále. Tento nejvýraznější zdroj heterogenity charakterizuje migrační aktivity závislé na vymezení registračních obvodů a na jejich geografických vzdálenostech. Proto tato pole nemají právo vstupovat do analýzy stejnocenným způsobem jako výše, ale musí být ošetřena jinak. Běžně jsou pokládána za tzv.strukturální nuly v modelu kvazinezávislosti: položme jako $S_1=\{(i,j),i\neq j,i,j=1,\dots,10\}$ množinu všech navzájem různých indexů i,j . Očekávané četnosti v modelu jsou vyjádřeny stejným předpisem jako v (4) omezeným pouze na podmnožinu S_1 tj.:

$$m_{ij} = \begin{cases} w.w_i^A.w_j^B & \text{pro } (i,j)\in S_1 \quad , \\ 0 & \text{jinak.} \end{cases} \tag{7}$$

Odhady $\hat{m}_{ij}$ (neexistuje explicitní vyjádření) jsou hledány prostřednictví IPF algoritmu (viz Prášková(1985),str.80) následovně -

- krok 0: nechť $s=0$; položme $\hat{m}_{ij} \begin{cases} 1 & \text{jestliže } (i,j)\in S_1 \\ 0 & \text{jinak} \end{cases}$;

- krok A: $$\hat{m}_{ij}^{(2s+1)} = \frac{n_{i+}}{\sum_j \hat{m}_{ij}^{(2s)}} . \hat{m}_{ij}^{(2s)} \quad ;$$

- krok B: $$\hat{m}_{ij}^{(2s+2)} = \frac{n_{+j}}{\sum_i \hat{m}_{ij}^{(2s+1)}} . \hat{m}_{ij}^{(2s+1)} \quad ;$$

nechť s=s+1 ; návrat do kroku A, dokud neplatí, že

$$\max_{i,j} |\hat{m}_{ij}^{(2s+2)} - \hat{m}_{ij}^{(2s)}| < \varepsilon \quad \text{pro některé } s \text{ a dané } \varepsilon .$$

V konzistenci s dalším značením lze model (7) charakterizovat identifikační maticí

$$\begin{Vmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{Vmatrix} \tag{8}$$

pouze náhodou totožnou s maticí počátečních hodnot v kroku 0 . Testové statistiky pro model kvazinezávislosti nabývají hodnot $\chi^2=5980.6$ resp. $G^2=4859.7$ tj.řádově desetkrát nižších než u předchozího modelu. Porovnání s kritickými hodnotami χ^2 rozdělení se 71 stupni volnosti je ovšem stejně bezcenné jako dříve.

Kromě diagonály však migrační tabulka (1) obsahuje ještě jeden systematický zdroj zvýšeného výskytu četností zřetelně se kumulující v okolních polích. Jde o dvojice vázané vzájemným sousedstvím okresů. Použití modelu kvazinezávislosti s identifikační maticí

$$\begin{Vmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{Vmatrix} \tag{9}$$

vyhlazuje data na hodnoty $\chi^2=207.0$ resp. $G^2=198.6$ při 41 stupních volnosti. Množina S_1 ve vzorci (7) je ovšem nyní definována jako množina všech dvojic (i,j) , kterým v identifikační matici (9) přísluší hodnota 1 . Problém však spočívá v tom, že nuly na diagonále identifikační matice jsou jiného typu nežli strukturální nuly mimo ni. I z formálního hlediska tento kvazinezávislý model vyhlazuje pouze tu část tabulky, která je pro analýzu migrací vlastně nejméně podstatná (v tomto případě právě polovina buněk). Korektnější řešení navrhují právě hybridní modely.

3. Hybridní identifikace sousedství

Migranti stěhující se mezi sousedními okresy představují tzv.přirozenou složku, která spočívá v přelivu obyvatelstva v blízkém okolí (za účelem zakládání rodin, nalezení nových pracovních příležitostí, rušení starých) a která je přímým důsledkem demografických, ekonomických a sociálních aspektů života dané lokality a jejího okolí (viz Řehák a kol.(1989),str. 6). Tato skupina migrantů činí v běžné analýze migrací značné problémy, neboť znečišťuje výsledky všech analýz z okresu do okresu právě nezohledněním vlivu jejich sousedství. Úloha tedy spočívá v takové matematické formalizaci, která adekvátní měrou zahrnuje tento aspekt do modelů. Vzhledem k úspěšné reprezentaci diagonály prostřednictvím strukturálních nul je vhodné toto ideu zahrnout též do hybridních specifikací.

Loglineární model schopný charakterizovat sousedství okresů jako celek lze formalizovat následujícím způsobem - model hybridní "sousedské" kvazinezávislosti:

$$m_{ij} = w.w_i^A.w_j^B.w_k^C \quad , \quad i,j=1,\dots,10 \quad , \tag{10}$$

kde
$$w_k^C \left\langle \begin{array}{ll} \bar{w}_k & \text{pro } (i,j)\in S_k \quad , \ k=1,2 \ , \\ 0 & \text{jinak} \ . \end{array} \right.$$

Tento model je určen identifikační maticí

$$\left\| \begin{array}{cccccccccc} 0 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 0 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 0 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 0 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 0 & 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 0 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 2 & 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 & 2 & 0 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 0 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 0 \end{array} \right\| \ , \tag{11}$$

kde hodnota 0 , 1 nebo 2 pro danou dvojici indexů (i,j) charakterizuje příslušnost k množině S_o , S_1 nebo S_2 (S_o identifikuje formálně množinu strukturálních nul na hlavní diagonále).

Odhady očekávaných četností $\hat{m}_{ij}$ lze získat aplikací IPF algoritmu na tabulce rozšířené o "hybridní" indexy k (viz Willekens-Baydar(1983),str.24-25) -

- krok 0:

necht' $s=0$, $n_{i++}=n_{i+}$, $n_{+j+}=n_{+j}$, $n_{++k}=\sum_{(i,j)\in S_k} n_{ij}$ a

definujme $$\hat{m}_{ijk} \left\langle \begin{array}{ll} 1 & \text{pro } (i,j)\in S_k \\ 0 & \text{jinak} \end{array} \right. ;$$

- krok A:

$$\hat{m}_{ijk}^{(3s+1)} = \frac{n_{i++}}{\sum_j \sum_k \hat{m}_{ijk}^{(3s)}} \cdot \hat{m}_{ijk}^{(3s)} \quad ;$$

- krok B:

$$\hat{m}_{ijk}^{(3s+2)} = \frac{n_{+j+}}{\sum_i \sum_k \hat{m}_{ijk}^{(3s+1)}} \cdot \hat{m}_{ijk}^{(3s+1)} \quad ;$$

- krok C:

$$\hat{m}_{ijk}^{(3s+3)} = \frac{n_{++k}}{\sum_i \sum_j \hat{m}_{ijk}^{(3s+2)}} \cdot \hat{m}_{ijk}^{(3s+2)} \quad ;$$

nechť $s=s+1$ a návrat do kroku A, pokud $\max_{i,j,k} |\hat{m}_{ijk}^{(3s+3)} - \hat{m}_{ijk}^{(3s)}| \geq \varepsilon$.

Konečné odhady dostaneme položením $\hat{m}_{ij} = \sum_k \hat{m}_{ijk}$. Aplikace tohoto modelu na migrační tabulce (1) redukuje testy dobré shody na $\chi^2=883.7$ resp. $G^2=895.1$ se 62 stupni volnosti.

Uvedená možnost specifikace sousedství není jediná. Pokud sousedství okresů považujeme za symetrické můžeme za nejpodrobnější hybridní model považovat ten, který pojímá každou dvojici sousedních okresů jako zcela autonomní - model hybridní "maximální sousedské" kvazinezávislosti:

$$m_{ij} = w.w_i^A.w_j^B.w_k^C \quad , \quad i,j=1,\dots,10 \quad , \qquad (12)$$

kde

$$w_k^C < \begin{matrix} \bar{w}_k & \text{pro } (i,j) \in S_k \ , \ k=1,2,\dots,16 \ , \\ 0 & \text{jinak} \ . \end{matrix}$$

Ve srovnání s (10) se model (12) odlišuje pouze specifikací množin S_k pro $k=2,\dots,16$ specifikující všechny sousední okresy. V identifikační tabulce

$$\begin{Vmatrix}
0 & 2 & 3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
2 & 0 & 4 & 5 & 1 & 1 & 1 & 1 & 1 & 1 \\
3 & 4 & 0 & 6 & 7 & 1 & 1 & 1 & 1 & 1 \\
1 & 5 & 6 & 0 & 8 & 9 & 1 & 1 & 1 & 1 \\
1 & 1 & 7 & 8 & 0 & 10 & 11 & 12 & 1 & 1 \\
1 & 1 & 1 & 9 & 10 & 0 & 13 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 11 & 13 & 0 & 14 & 1 & 1 \\
1 & 1 & 1 & 1 & 12 & 1 & 14 & 0 & 15 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 15 & 0 & 16 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 16 & 0
\end{Vmatrix} \qquad (13)$$

mohou být symetrické dvojice libovolně permutovány, neboť mají pouze symbolický význam. Kriteriální hodnoty vylepšují dobrou shodu na $\chi^2=308.8$ resp. $G^2=303.1$ při 56 stupních volnosti.

4. Identifikace typu sousedství

Řešení geografických úloh však vyžaduje poněkud pestřejší paletu modelů; sousedství okresů mohou být agregována do skupin podle typu jejich vzájemné podobnosti. Podrobnější pohled na geografii severočeského kraje nabízí myšlenku specifikovat "podobnost sousedství okresů" kupř. na základě délky společných hranic. Níže uvedená matice identifikuje hybridní loglineární model, pro který množiny $S_2,\dots,S_7$ odpovídají typům zakresleným na obrázku:

$$\begin{Vmatrix}
0 & 5 & 6 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
5 & 0 & 4 & 6 & 1 & 1 & 1 & 1 & 1 & 1 \\
6 & 4 & 0 & 2 & 5 & 1 & 1 & 1 & 1 & 1 \\
1 & 6 & 2 & 0 & 3 & 5 & 1 & 1 & 1 & 1 \\
1 & 1 & 5 & 3 & 0 & 4 & 2 & 4 & 1 & 1 \\
1 & 1 & 1 & 5 & 4 & 0 & 5 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 2 & 5 & 0 & 6 & 1 & 1 \\
1 & 1 & 1 & 1 & 4 & 1 & 6 & 0 & 5 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 5 & 0 & 7 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 7 & 0
\end{Vmatrix} \qquad (14)$$

stejná čísla v rámci horního (resp.dolního) trojúhelníku matice přísluší dvojicím okresů s přibližně stejnou délkou společné hranice. (Nejkratší hranice jsou symbolizovány číslicí 2 , nejdelší číslem 7 .) Shoda modelu s daty odpovídá hodnotám χ^2=616.4 resp. G^2=618.2 se 62 stupni volnosti.

Uspořádejme dosažené výsledky z proložení migrační tabulky (1) sestupně podle hodnot testovacích statistik:

model:	χ^2:	G^2:	q:	diference:	s:	hladina významnosti:	
nezávislost	95536.1	53703.9	81			0.000	
				48844.2	10		0.000
kvazinezávislost	5980.6	4859.7	71			0.000	
				3964.7	1		0.000
sousedská kvazinezávislost	883.7	895.1	70			0.000	
				276.9	8		0.000
délkově-sousedská kvazinezávislost	616.4	618.2	62			0.000	
				315.0	6		0.000
maximální sousedská kvazinezávislost	308.8	303.1	56			0.000	
? ? ? HIC SUNT LEONES ? ? ?							
saturovaný	0.0	0.0	0			1.000	

Systematický pokles hodnot χ^2 a G^2 není náhodný, protože jde o vnořené-zahnízděné modely (každý následující model je submodelem resp.podrobnější specifikací předchozího). Tuto vlastnost lze ověřit na základě matice plánu experimentu transformující logaritmy očekávaných četností na součet logaritmovaných parametrů (viz Willekens-Baydar (1983),str.40-42). Hodnost této matice určuje počet nezávislých parametrů modelu a byla použita pro výpočet hodnoty q . Diferenci mezi sousedními statistikami G^2 lze interpretovat jako statistiku testující nulovost všech rozdílových parametrů obsažených v modelu následujícím. Počta stupňů volnosti s mezi druhým až pátým řádkem odpovídají rozdílům v nárůstu počtu hybridních parametrů.

Skutečnost, že všechny diference jsou statisticky významné, inklinuje k prověření ještě podrobnější sestavy modelů umožňující testovat jednotlivé parametry zvlášť'. Problém je ovšem v tom, že akceptovatelných hybridních modelů (mezi kvazinezávislostí a maximální sousedskou kvazinezávislostí) je právě tolik, kolik existuje podmnožin vybraných z množiny patnácti dvojic sousedských okresů tj. 2^{15} modelů (!) . Navíc všechny modely nejsou navzájem vnořené, takže existuje celá řada optimalizačních cest mezi oběma koncovými sousedskými alternativami. V daném případě však ani nejpodrobnější hybridní model nedosahuje takových hladin významnosti, které by byly použitelné pro uplatnění optimalizace. Hledání dalších modelů ve vyznačené neprobádané oblasti, kde formálně končí geografie založená na sousedství okresů, může být problémem dalšího výzkumu.

5. Závěry a doporučení

Není podstatné, zda uvažovaný typ sousedství okresů nějak souvisí s délkou společných hranic nebo je dán i jinými geografickými specifikacemi (např.vzdáleností geografických středů vztažených okresů nebo dalšími funkčními vztahy). Z tohoto pohledu může být velmi zajímavá úloha, který typ sousedství se jeví jako vhodný z hlediska optimálního vyhlazení migrační tabulky prostřednictvím testů χ^2 dobré shody. Problémem je ovšem kriteriální hladina významnosti, když všechny dosavadní modely byly silně statisticky významné. Tuto záležitost lze řešit dvěma alternativními způsoby:

1/ Nalézt druhý hybridní efekt (analogický hybridnímu efektu sousedství) a jeho prostřednictvím se snažit vysvětlovat danou dvourozměrnou tabulku. Tato možnost je složitá spíše z obsahového hlediska, neboť již pojem "hybridního efektu sousedství" byl vázán na řadu interpretačních obtíží. Technicky lze tuto alternativu zabezpečit zavedením druhé nezávislé identifikační tabulky a úpravou vyhlazovacího algoritmu.

2/ Vytřídit vstupní migrační tabulku navíc podle třetí (nehybridní) proměnné (pohlaví, věku apod.) a teprve poté využít spektrum výše navrhovaných hybridních alternativ již při korektních hladinách významnosti. V tomto případě však připadá v úvahu větší množství výchozích trojrozměrných modelů (úplná, párová, sdružená resp.podmíněná nezávislost), na nichž se hybridní efekty uplatní ve formě čtvrté optimalizovatelné proměnné.

Otázku odhadu parametrů(eventuálně výpočtu jejich asymptotických směrodatných odchylek) lze řešit prostřednictvím matice plánu experimentu s využitím standardních algoritmů pro řešení soustav lineárních (normálních) rovnic s neúplnou hodností. U všech modelů ze stati 4 lze ukázat, že hodnost této matice je rovna $r = (10 + 10 + K - 2)$, kde K je počet všech hybridních parametrů modelu. Tento fakt umožňuje svázat parametry hybridních sousedských modelů podmínkami

$$\prod_i w_i^A = \prod_j w_j^B = \prod_k \bar{\bar{w}}_k^C = 1 \quad , \qquad (15)$$

které zaručují jednoznačnost řešení. Hybridní "sousedské" efekty lze tedy interpretovat v analogickém smyslu jako efekty "počátečního" nebo "koncového" místa migrace.

<u>Literatura</u>:

[1] Anděl J.(1978): Matematická statistika. SNTL, Praha.

[2] Bishop Y.M.-Fienberg S.E.-Holland P.W.(1975): Discrete multivariate analysis - - theory and practise. Cambridge, MIT Press.

[3] Gokhale D.V.-Johnson N.S.(1978): A class of alternatives to independence in contingency tables. JASA 73, str.800-804.

[4] Goodman L.(1972): Some multiplicative models for the analysis of cross-classified data. In: L.Le Cam et al.(eds.), Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, Berkeley: University of California Press, str.649-696.

[5] Haberman S.J.(1974): The analysis of frequency data. Chicago: The University of Chicago Press.

[6] Prášková Z.(1985): Kontingenční tabulky. Skripta MFF UK Praha.

[7] Řehák a kol.(1989): Migrační toky Severočeského kraje - mimokrajová výměna. Výzk. zpráva SEÚ ČSAV.

[8] Willekens F.-Baydar N.(1983): Hybrid loglinear models. Netherlands Interuniversity Demographic Institute, WP 41, Voorburg. (Preprint(1985): Measuring the unmeasurable. In Nijkamp P. et al.(eds.), Dordrecht Nijhof, str.141-176).