Tilburg University

Loglineáris elemzés rejtett változókkal Mobilitási táblák

Luijkx, R.

Published in:
A társadalmi struktúra az életmód és a tudat alakulása magyarországon

Publication date:
1985

Ruud LUIJKX

LOGLINEÁRIS ELEMZÉS REJTETT VÁLTOZÓKKAL: MOBILITÁSI TÁBLÁK

Bevezetés

E dolgozatban angliai és dániai mobilitásvizsgálatok adatait dolgozzuk fel bemutatva a loglineáris elemzés rejtett változóra való alkalmazását, ahogy azt CLOGG /1981a,b/ felvetette. A módszer részét képezi a rejtett strukturák elemzési lehetőségeinek, ahogy azt P.F. Lazarsfeld fejlesztette ki /LAZARSFELD-1950a,b, 1954, 1959; LAZARSFELD és HENRY-1968/. Lazarsfeld és Clogg munkáin kivül a dolgozat alapját képezik még Leo Goodman /1974a,b/, Jacques Hagenaars /1976, 1978/, Clifford Clogg /1980, 1981a,b/ dolgozatai és a szerző korábbi publikálatlan munkája /LUIJKX-1983/.
Az első részben megvilágitjuk a loglineáris elemzés alkalmazását rejtett változókra. A rejtett struktura elemzése igy analóg a folytonos változókra alkalmazott faktoranalizis modellekkel /GREEN-1952/, és mint ilyen egy általános feldolgozási metodika /LAZARSFELD-1955; ROSENBERG-1968; HAGENAARS-1980/. A lényeg megvilágitásának egyszerüsitése érdekében a továbbiakban kétdimenziós kontingencia táblára és egyetlen rejtett változóra szoritkozunk, s mivel a szerző társadalmi mobilitással foglalkozik, a táblák a fent emlitett mobilitási táblák lesznek.
Mindeddig lényeges problémát jelentett az egyes modellekben a szükséges számitások végrehajtása. Olyan programok azonban, mint a MLISA /CLOGG-1977/ és a LCAG/HAGENAARS és VAN DER WALLE-1983/ ma már viszonylag egyszerüen lehetővé teszik rejtett osztályozási modellek számitását maximum likelihood becslés alapján. A dolgozatban bemutatjuk a LCAG müködését.
A második részben kvázi-rejtett struktura modelleket mutatunk be a mobilitási táblákra vonatkozóan. Ezek meglehetősen jól irják le a mobilitási adatokat. Clogg /1981b/ kimutatta, hogy a rejtett osztályozási modellek alkalmas eszköznek tekinthetők arra, hogy különböző országok mobilitási adatait összehasonlitsuk. A dolgozatban erre is kitérünk, bemutatván hogy az országot mint változót tekintve hogyan vizsgálható az országok közötti hasonlóság.

1.0. Loglineáris elemzés rejtett változókkal

A módszer a kiindulási -(0) és a cél - (D) változók közötti összefüggést egy harmadik -(X) változó által meghatározottnak tekinti. Két megközelités lehetséges: M/arginális és P/arciális modell. Az M-tipusu model esetében az 0 és D változók közötti asszociáció az X változó minden rögzitett értékénél azonos ($0D.X_a=0D.X_b$), de nem azonos az eredeti asszociációval (0D.X≠0D) ennek a tipusnak alesete a tiszta model, amikor 0 és D asszociációja X minden értékénél o - ekkor az 0 és D közti összefüggés látszólagos. A P-tipusu modelben az 0 és D közti asszociáció nem azonos X minden rögzitett értékénél ($0D.X_a \neq 0D.X_b$). Az M-tipusu modeleken belül megkülönböztethető MA-tipusu és MI-tipusu, az első esetben X oksági kapcsolatban van o és D változókkal egyaránt /magyarázó változó/, a második esetben az X köztes változó 0 és D között. Igy X magyarázhatja ill. értelmezheti az 0 és D közti asszociációt.

A rejtett osztályozási modelekben a teszt-változó /X/ nem megfigyelhető, rejtett változó. A klasszikus alkalmazásokban tiszta MA-tipusu modelekkel dolgoztak, feltételezték, hogy a fejtett változó okozza a két megfigyelt változó közötti összefüggést. Ez sematikusan a következőképpen ábrázolható:

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ 0 & & D \end{tikzcd}$$

Lazarsfeldnek nem sikerült megfelelő statisztikai becslés, és próba eljárásokat kidolgozni, ő és mások is determináns egyenleteket használt a becslések végrehajtásához. Leo Goodman /1974a,b/ és Shelby Haberman /1974/ találtak olyan algoritmust, amely a model paramétereinek maximum likelihood becsléseit /MLE/ előállitja. E cikkben a Goodman féle algoritmust tárgyaljuk.

A következő jelöléseket vezetjük be:

π várható valószinüségek a populációban

$\hat{p}$ várható valószinüségek a mintában

p megfigyelt valószinüségek a mintában

$0, D, X$ felső index a kiindulási 0, cél D, és rejtett X változókra

i, j, t alsó index az 0, D és X változók kategóriáira

A feltételes valószinüségeket a felső index feletti vonás kitételével jelöljük, pl. $\hat{P}_{it}^{\bar{O}X}$ jelöli azt a feltételes valószinüséget, hogy valaki az 0 változót tekintve az i-dik osztályban van/0 értéke i/ feltéve, hogy az X változót tekintve a t osztályban van.

A rejtett osztály elemzés során két alapfeltevést kell elfogadni:

- homogenitás; feltéve, hogy a rejtett változó egy adott osztályába esik /egy t értéket vesz fel az X változót tekintve/ a populáció minden egyede azonos feltételes valószinüséggel veszi fel a megfigyelhető változók/0 és D/ valamely értékét/ az egyedek T diszjunkt rejtett osztályba sorolhatók, az osztályozás teljes/
- lokalis függetlenség; feltéve, hogy a rejtett változó egy adott osztályába esik egy egyed, az 0 és D szerinti osztálybasorolás feltételesen független egymástól.

1.1. Rejtett osztályozási modellmegszoritás nélkül

Egyetlen rejtett X változót feltételezve a megfigyelt arányok X összes értékére való összegzéssel kaphatók meg, X az OD táblában csak indirekt módon figyelhető meg.

(1) $$\pi_{ij}^{OD} = \Sigma_t \pi_{ijt}^{ODX}$$

(1)-ben π_{ijt}^{ODX} annak a valószinüsége, hogy egy egyed az ODX tábla (i, j, t) cellájába esik.

A feltételes valószinüségre vonatkozó számolási szabályokból és a lokális függetlenség feltételezéséből:

(2) $$\pi_{ijt}^{ODX} = \pi_t^X \pi_{it}^{\bar{O}X} \pi_{jt}^{\bar{D}X}$$

A következő kikötés triviális[3]

(3) $$\Sigma_t \pi_t^X = \Sigma_i \pi_{it}^{\bar{O}X} = \Sigma_j \pi_{jt}^{\bar{D}X} = 1$$

(2)-t (1)-be helyettesitve:

(4) $$\pi_{ij}^{OD} = \Sigma_t \pi_t^X \pi_{it}^{\bar{O}X} \pi_{jt}^{\bar{D}X}$$

A megfigyelhető π_{ij}^{ob} arány a fejtett osztályokba tartozás valószinüségei és a megfelelő feltételes valószinüségek szorzatainak az összege. π_t^x a mobilitási táblából indirekt módon megfigyelhető rejtett osztályok eloszlása, $\pi_{it}^{\bar{o}x}$, $\pi_{jt}^{\bar{b}x}$ irják le a kiindulási és cél változók eloszlásait egy-egy rejtett osztályba tartozás esetén.

Számos kompjuterprogram alkalmas a rejtett osztályozási modell paramétereinek a meghatározására. Tudomásunk szerint az első ilyen volt a Clogg által készitett MLLSA/Maximum Likelihood Latent Structure Analysis, CLOGG-1977/, ez a program Goodman interativ skálázási módszerét alkalmazza /GOODMAN-1974b/. Haberman programja /HABERMAN-1979/ a Nexton-Raphson algoritmust használja. Az előbbi algoritmust a LCAG program is használja /Latent Class Analysis Goodman/ /HAGENAARS és VAN DER WALLE-1983/[5]

Haberman /1974 pp 912-915/ megmutatja, hogy ugyanazok a maximum likelihood egyenletek alkalmasak a rejtett strukturák modelljére, mint ha a modell csak a megfigyelhető változók vizsgálatára szoritkozna. A rejtett strukturák esetében a gyakoriságok helyettesitendők feltételes várható értékeik becsléseivel feltéve a megfigyelt marginálisokat.

A LCAG program alkalmas bármely identifikálható hierarchikus loglineáris modell paramétereienk becslésére rejtett változók esetében sőt több nem-hierarchikus modell esetén is alkalmazható. Goodman eljárásának egyfajta továbbfejlesztését tartalmazza, amennyiben egyrészt lehetőség van rejtett változók közötti relációk hierarchikus modellezésére, másrészt a modellekbe bevehetők olyan rejtett változók, melyeknek paraméterei ugy számittatnak ki, ahogy a módositott ut-elemzés Goodman féle tárgyalásában. Mindkét kiterjesztés felhasználásra kerül.[6]

A program számitási menete Goodman cikkén alapul; /1974b p 217(9)-(136) /HAGENAARS-1976/

1. lépés: kezdeti értékek felvétele - a π vektor $/\pi_t^x, \pi_{it}^{\bar{o}x}, \pi_{jt}^{\bar{b}x}/$ elemei kapnak értéket /pl. véletlenszám generátor használatával/

2. lépés: az aktuális paraméter-értékekkel kiszámitódnak az ODX tábla celláihoz tartozó valószinüségek:

(5) $\hat{P}^{ODX}_{ijt} = \hat{P}^{X}_{t}\,\hat{P}^{\bar{O}X}_{it}\,\hat{P}^{\bar{D}X}_{jt}$

3. lépés: a megfigyelt /0D/ tábla celláihoz tartozó valószinüségek számidódnak:

(6) $\hat{P}^{OD}_{ij} = \Sigma_t\,\hat{P}^{ODX}_{ijt}$

4. lépés: a megfigyelt tábla cellák feltételes valószinüségei feltéve X valamennyi rögzitett értékét kiszámitódn ak:

(7) $\hat{P}^{OD\bar{X}}_{ijt} = \hat{P}^{ODX}_{ijt} / \hat{P}^{OD}_{ij}$

5. lépés: uj becslések a paraméter vektor elemeire / P^{OD}_{ij} a megfigyelt valószinüségek! /:

(8) $\hat{P}^{X}_{t} = \Sigma_{ij}\,P^{OD}_{ij}\,\hat{P}^{OD\bar{X}}_{ijt}$

(9a) $\hat{P}^{\bar{O}X}_{it} = [\Sigma_j\,P^{OD}_{ij}\,\hat{P}^{OD\bar{X}}_{ijt}]\,\hat{P}^{X}_{t}$

(9b) $\hat{P}^{D\bar{X}}_{jt} = [\Sigma_i\,P^{OD}_{ij}\,\hat{P}^{OD\bar{X}}_{ijt}]\,\hat{P}^{X}_{t}$

A ciklus a 2. lépéstől folytatódik. Egy rejtett osztály akkor szünik meg / azaz kihagyódik a számitásból/, ha a megfelelő becslés 0-hoz tart. A számitás addig ismétlődik, amig valamilyen előre adott megállási szabály szerint tekintve el nem ér egy konvergencia-pontot - ez a megállási szabály lehet az iterációk számának megadott felső korlátja, vagy két egymásutáni iteráció során kiszámitott statisztikus próbák értéke különbségének előre adott korlátja. Minden iteráció után az aktuális paraméter-vektor ujra rendeződik - megfelelően egy esetlegesen adott /ld. 1.2 §/ kezdeti feltételnek - továbbá minden iteráció után a megfelelő való szinüségek összege 1-re normalizálódik.[9]

Mint statisztikai próbák a log-likelihood arány (L) és a Pearson féle χ-négyzet kerül kiszámitásra az egyes iterációk 3. lépésének eredményeiből:[10]

(10a) $L = 2n\sum_{ij} P_{ij}^{OD} * \ln(P_{ij}^{OD}/\hat{P}_{ij}^{OD})$

(10b) $\chi^2 = n\sum_{ij} (P_{ij}^{OD} - \hat{P}_{ij}^{OD})^2 / \hat{P}_{ij}^{OD}$

A χ-négyzet szabadsági fokát a következő képlet adja - kezdeti feltétel nélküli modellről lévén szó:

(11) $df = IJ - (I + J - 1)T$

Előfordulhat, hogy az iteráció nem vezet egy globális maximumhoz, ilyenkor csak terminális maximumról beszélhetünk.[11]

Annak eldöntésére, hogy melyik maximum a maximum-likelihood megoldás az eljárást többször meg kell ismételni kissé eltérő kezdeti értékekkel. A maximum-likelihood megoldás a statisztikai próbák legalacsonyabb értékét adja.

Nem minden modell eredményez identifikálható paramétereket. Ez azt jelenti, hogy esetleg több paraméter halmaz eredményezheti ugyanazokat a várható gyakoriságokat. A rejtett /feltételes/ valószinüségek becslései nem egyértelmüen határozhatók meg a megfigyelhető gyakoriságokból. Azt hogy egy modell nem identifikálható paraméterekkel rendelkezik, onnan lehet látni, hogy különböző kiindulási értékekkel különböző paraméterértékeket kapunk, a χ-négyzet statisztika azonban azonos. A modell identifikálhatósága érdekében ilyenkor szükség lehet megszoritások bevezetésére. Ekkor azoban gondosan kell eljárni, mert bizonyos megszoritások a modellt nem-identifikálhatóvá tehetik /GOODMAN-1974b pp. 225-226 és 1.2 §/. Célszerübb tehát a modell identifikálhatóságának kérdését a megszoritások bevezetése előtt eldönteni. Goodman /1974b/ bemutat néhány nem-identifikálható modellt és ötletet ad azok identifikálhatóvá tételére. Egy modell paraméterei lokálisan identifikálhatók ha a 2 által meghatározott transzformáció nem szinguláris.[13] A LCAG program teszteli a paraméterek lokális identifikálhatóságát.[14]

1.2. Rejtett osztályozási modell megszoritásokkal

Az 1.1§-ban láttuk, hogy sok esetben szükséges lehet néhány rejtett /feltételes/ valószinüséget ismrtnek tekinteni /értékét előirni/ annak érdekében, hogy a paraméterek becslései meghatározhatók legyenek. A következő tipusu megszoritások lehetségesek:

- Előirhatjuk a rejtett osztályok valószinüségeinek π_t^X vektorát, vagy annak egy elemét, pl. $\pi_2^X = 0{,}25$ /szükséges, hogy nem kisebb mint o, és kisebb mint 1 legyen./
- Előirhatunk egy vagy több $\pi_{it}^{\bar{D}X}$ vagy $\pi_{jt}^{\bar{D}X}$ feltételes valószinüséget, pl. $\pi_{51}^{\bar{D}X} = 0{,}20$
- Megkövetelhető egyenlőség a rejtett osztályok valószinüségei között, pl. $\pi_2^X = \pi_3^X$
- Megkövetelhető egyenlőség a $\pi_{it}^{\bar{D}X}$ és a $\pi_{jt}^{\bar{D}X}$ feltételes valószinüségek között tetszőleges kombinációban: pl:

$$\pi_{51}^{\bar{D}X} = \pi_{52}^{\bar{D}X}, \quad \pi_{51}^{\bar{D}X} = \pi_{41}^{\bar{D}X}, \quad \pi_{52}^{\bar{D}X} = \pi_{52}^{\bar{D}X}$$

Annak eldöntésére, hogy egy rejtett strukturát magában foglaló modell. az adott megszoritásokkal identifikálható paraméter becslésekkel rendelkezik-e a fentebb emlitett kiértékelési eljárásnak egy módositott formája használható.[15]

Ha a megszoritott modell identifálható, akkor a hozzátartozó szabadsági fok a megfelelő megszoritás nélküli modellhez tartozó szabadsági fokból és a nem-redundáns megszoritások számából tevődik össze:

(12) $df = IJ - I+J-1\ T + d$

Itt a d a nem-redundáns kezdeti feltételek száma.

Amint már emlitettük az 1.1§-ban, olyan megszoritások is megadhatók, amelyek a paraméter-becsléseket nem-identifikálhatókká teszik. Példának okáért ha a 13a és 13b megszoritásokat vesszük figyelembe, akkor az X=1 és az X=2 rejtett osztályok összeesnek /hacsak nem alkalmzunk további alkalmas megszoritást a π_1^x vagy π_2^x paraméterekre is/[16]

(13a) $\pi_{i1}^{\bar{o}x} = \pi_{i2}^{\bar{o}x} \quad (i = 1, \ldots I)$

(13b) $\pi_{j1}^{\bar{b}x} = \pi_{j2}^{\bar{b}x} \quad (j = 1, \ldots J)$

Ha csak a 13b megszoritásokat tekintjük, a paraméter-becslések akkor is nem-identifikálhatók lesznek/hacsak nem alkalmazunk megfelelő megszoritásokat ez esetben a π_1^x, π_2^x, $\pi_{j1}^{\bar{b}x}$ vagy $\pi_{j2}^{\bar{b}x}$ paraméterekre is./[17]

2.o A mobilitás táblák vizsgálata

Clogg /198la,b/ a rejtett osztályozás modelljét két klasszikusnak mondható mobilitási táblán mutatta be, az egyik Anglia és Wales adatait tartalmazza /GLASS-1954/, a másik pedig Dánia adatait /SVALAGOSTA-1959/. Ebben a dolgozatban ugyanezeket használjuk a továbbiakban. A nyers adatokat az alábbi 1. tábla tartalmazza.[18]

1. táblázat

Intergenerációs társadalmi mobilitás: Dánia - Anglia és Wales

		Fiu (Dánia)						Fiu (Anglia és Wales)					
		1	2	3	4	5		1	2	3	4	5	
A	1	18	17	16	4	2	57	58	45	8	18	8	129
P	2	24	185	189	59	21	318	28	174	84	154	55	495
A	3	23	84	289	217	95	708	11	78	110	223	96	518
	4	8	49	175	348	198	778	14	150	185	714	447	1510
	5	6	8	69	201	246	530	0	42	72	320	411	845
		79	263	658	829	562	2391	103	489	459	1429	1017	3497

Clogg a problémát a kvázi-függetlenség /GOODMAN-1969b/ alapján tárgyalta. Clogghoz hasonlóan mind Goodman mind Hauser /GOODMAN-1969b/ /HAUSER 1978, 1979a,b/ vizsgálataik során feltételezték a rejtett struktura meglétét, anélkül azonban, hogy explicit módon használták volna a rejtett strukturák modelezési lehetőségét. A szerző ugy véli, hogy mind Goodman, mind Hauser elsősorban a megfigyelhető strukturák vizsgálatára szoritkozott. A dolgozatban látni fogjuk, hogyan fejezhető ki a Goodman féle kvázi-függetlenség a rejtett strukturák analizise segitségével. Ugyanez elmondható Hauser szint-paraméter modelljére is /jóllehet ez esetben számos megszoritást kell alkalmazni/.

Clogg az X rejtett változót mint köztes változót tekintette:

$$O \longrightarrow X \longrightarrow D$$

Ezzel eltér Lazarsfeld eredeti elgondolásától, amely szerint a rejtett strukrura MA-tipusu, hiszen a vázolt elgondolás MI-tipusu modellt jelent. Elméletileg ez a megközelités ugy interpretálható, hogy a rejtett X változót a kiindulási O változó határozza meg, másrészt a rejtett változó meghatározza a célváltozónak, D-nek az alakulását.

Ez a tárgyalásmód összevethető Boudon /1973/ felfogásával, aki vizsgálatában nem alkalmaz rejtett változót, de a "képzettség szintjét" mint köztes változót feltételezi.

Clogg a rejtett X változót "osztály" változónak, a kiindulási O változót, és a D célváltozót pedig "státusz" változónak nevezi. Rámutat, hogy X valójában több változóból állhat. X osztályai az OD táblában feltünően különböző státusz-kombinációk formájában nyilvánulnak meg. Clogg szerint a valószinüségi eloszlást tekintve, X és O részben koegzisztensnek részben X az O következményének is tekinthető, de X nem tekinthető pusztán a "bekerülési valószinüségnek", ahogy az egy rejtett struktura modellben természetes lenne. Clogg álláspontját védve rámutat arra az izomorfizmusra, amelyet a rejtett változó és a Max Weber által bevezetett osztály fogalma között vélt megtalálni. Mindazonáltal ha van is hasonlóság, nincs /és nem is lehet/ létjogosultsága annak, hogy az osztályt mint köztes változót definiáljuk. Elképzelésünk szerint Weber elgondolása egy olyan rejtett struktura modellbe illeszthető, amelyben a rejtett osztályok változója ok-okozati viszonyban van mind a kiindulási, mind a cél változóval.

$$X \begin{array}{l} \nearrow O \\ \searrow D \end{array}$$

A rejtett osztályozási modellek elméleti alapját és a formalizmust átgondolva, nincs jelentősége annak, hogy X logikailag megelőzi-e az O és D változókat, vagy köztes változóként lehet felfogni. Az egyedüli lényeges pont az, hogy O-nak és D-nek legalább bizonyos osztályai definiálják /valószinüségileg/ az X-nek egy osztályozását, amelyben O és D kölcsönösen függetlenek egymástól. Mivel e dolgozat elsősorban módszertani jellegü, a továbbiakban nem taglaljuk Clogg értelmezéseit az O, D és X változók és azok összefüggése vonatkozásában.

2.1. Az egyes táblák külön elemzése

Összehasonlitási alapként Clogg a tiszta mobilitási modellt használja, ez esetben O és D függetlenek, a modell a következőképpen irható fel:

$$(14) \qquad \pi_{ij}^{OD} = \pi_{1}^{X}\, \pi_{i1}^{\bar{O}X}\, \pi_{j1}^{\bar{D}X}$$

Az X változónak egyetlen osztálya van igy $\pi_{1}^{X} = 1$ továbbá:

$$(15) \qquad \pi_{ij}^{OD} = \pi_{i}^{O}\, \pi_{j}^{D}$$

Ekkor nincsen rejtett struktura, a mobilitás véletlenszerü, csak O és D marginális eloszlásától függ.

Ezután Clogg két rejtett osztályt feltételezett. A mobilitás ez esetben mindkét rejtett osztályban egyedül a $\pi_{it}^{\bar{O}X}$ és $\pi_{jt}^{\bar{D}X}$ marginális eloszlásoktól függ, a mobilitás "osztály korlátjáról" beszélhetünk. Ez a modell nem identifikálható paramétereket eredményez. Clogg elméletileg megalapozott a priori megszoritásokat javasol, javaslata az, hogy bizonyos feltételes valószinüségeket tekintsünk o-nak. Az első rejtett osztályba tartozó mobilok számára a kiinduló változó első osztályba tartozás, a másik rejtett osztály mobiljai számára a kiindulási változó utolsó /5./ osztályába való tartozás legyen o valószinüségü. Igy tehát a mobilok első rejtett osztályába tartozók nem indulhatnak a kiindulási változó első státuszcsoportjából a másik osztályba tartozók pedig az utolsó /5./ státuszcsoportból:[21]

$$(16) \qquad \pi_{11}^{\bar{O}X} = \pi_{52}^{\bar{O}X} = \phi$$

Még logikusabbnak tünne ugyanilyen megszoritást alkalmazni a célváltozóra vonatkozó feltételes valószinüségekre. Clogg ezt triviális matematikai okokból nem tette meg, ugyanis ugyanazokat a kezdeti feltételeket alkalmazva a kiindulási és a cél változókra a kapott modell o várható értéket eredményezne; ez a következőkből belátható. /ld. 1. ábra/

1. ábra diagram /0 a 0-valószinüségeket, 1 a 0-nál nagyobb valószinüségeket jelöli/

0	0	0	0	0	1	1	1	1	0	1	1	1	1	0
0	1	1	1	1	1	1	1	1	0	1	1	1	1	1
0	1	1	1	1	1	1	1	1	0	1	1	1	1	1
0	1	1	1	1	1	1	1	1	0	1	1	1	1	1
0	1	1	1	1	0	0	0	0	0	0	1	1	1	1
		(1)					(2)					(3)		

(1) : a mobilok első rejtett osztálya

(2) : a mobilok második rejtett osztálya

(3) : (1) és (2) összege

Az (1) diagramon ábrázoltuk a mobilok első rejtett osztályát azzal a megszoritással, hogy nem lehetnek benne sem a kiindulási, sem a cél változó első státuszcsoportjában, a (2) diagramon a második rejtett osztály mobiljait ábrázoltuk, ezek a kiindulási és a cél változó utolsó státuszcsoportjaiban nem lehetnek. Az utolsó diagram az első kettő összege, látható rajta, az /1,5/ és /5,1/ cellában a várható gyakoriság o, ez pedig az algoritmus 4. lépésében o-val való osztást eredményezne.

Clogg három rejtett osztályos modellt is bemutatott ugyanilyen és más additiv megszoritásokkal,[23] egyikük sem hozott kielégitő eredményt.

Térjünk rá a kvázi-rejtett struktura modelljére. Először is irjuk le Goodman kvázi-független modelljét ilyennek. Ebben az esetben a rejtett osztályok száma 1-gyel nagyobb a kiindulási és a célváltozó osztályainak közös számánál. A fődiagonális minden celljára definiáljon egy rejtett osztályt /v. ö. pl. Clogg/1980 p. 254/. Az alábbi megszoritásokat alkalmazzuk:

(17) $\pi_{it}^{\bar{o}x} = \pi_{it}^{\bar{b}x} = 1$ ha i=t és i=1....I

Ez olyan feltételes valószinüségeket ad meg, hogy bárki aki a rejtett változó t osztályában van, egyben a kiindulási és a cél változó szerint is a t-be kerül. Clogg ezeket a rejtett osztályokat, melyek a rejtett változó determinisztikus státuszosztályai, rejtett "maradóknak" nevezte. Ezeken kivül van még egy rejtett osztály, amelybe a mobilok kerülnek.

Az egyetlen megszoritás a fődiagonálissal kapcsolatos. Ez a modell kvázi-rejtettnek nevezhető, mivel nem tartalmaz rejtett strukturát a szó szoros értelmében véve, a rejtett változó teljes mértékben a megszoritásokkal adott. A modell a következőképpen irható fel:

(18) $\pi_{ij}^{ob} = \Sigma_t \pi_t^x \pi_{it}^{\bar{o}x} \pi_{jt}^{\bar{b}x}$ t=1...6

ahol $\pi_{ii}^{\bar{o}x} = \pi_{ii}^{\bar{b}x} = 1$

Clogg definiált egy megfelelő immobilitási indexet:

(19) $$\gamma_i = \frac{\pi_T^x \pi_{iT}^{\bar{o}x} \pi_{iT}^{\bar{b}x} + \pi_i^x}{\pi_T^x \pi_{iT}^{\bar{o}x} \pi_{iT}^{\bar{b}x}}$$

A következő modell, amelyet Clogg definiált olyan kvazi-rejtett strukturáju, amelyben a rejtett osztályok száma I+2. Hasonlóan az előbbihez I osztály a fődiagonális celláinak státuszosztályaival lesz definiálva, a fennmaradó két osztály itt is a mobilok két rejtett valószinüségi osztálya:

(20) $\pi_{ij}^{ob} = \Sigma_t \pi_t^x \pi_{it}^{\bar{o}x} \pi_{jt}^{\bar{b}x}$ t=1...7

ahol $\pi_{ii}^{\bar{o}x} = \pi_{ii}^{\bar{b}x} = 1$

Ehhez a kvázi-rejtett modellhez is definiálható egy (19)-hez hasonló immobilitási index. Ez az index az i-statuszban lévők között lévő "maradók" azon többletét méri, amely nem tudható be annak, hogy a két rejtett osztályt feltételező modell esetében a várható immobilitás alatta marad az előzőnek. /CLOGG-1981b p. 843/ Clogg szerint mindeddig az volt a probléma az immobilitási indexszel, hogy a státuszcsoport gyakorisága a tiszta mobilitás esetében várható mobilitási gyakorisággal volt összehasonlitva. Az, hogy a Clogg féle index jobb interpretációs lehetőséget biztosit-e, attól függ, hogy modellje hogyan értelmezhető és hogyan illeszkedik a mobilitás jelenségéhez.

Clogg /191a, b/ a brit és dán mobilitási adatokat külön elemzi és összehasonlitja eltéréseiket. Eközben nem használ formális tesztet és nem hasonlitja össze az országokat egy modellben /azaz az országot mint változót használva/. A következőkben pótoljuk ezeket a hiányosságokat.

Clogg mindkét kvázi-rejtett struktura modell egy változatát definiálja. Nem definiálja valamennyi determinisztikus státusz osztályt, csak hármat az öt közül /az 1, 2, és 3 diagonális cellákat/. Ez a szabadságfokok számának emelkedéséhez vezet. Helyesebb eljárásnak tünik az egyes tényezők bevételéről vagy kihagyásáról elméleti alapon dönteni, de az összehasonlithatóság érdekében, Clogghoz hasonlóan öt rejtett osztályt fogunk használni. Az első rejtett osztály a legfelső és $\pi^{\bar{o}x}_{51}$ egyenlő nullával; a második rejtett osztály a legalacsonyabb azzal a megkötéssel, hogy $\pi^{\bar{o}x}_{12}$ egyenlő nullával. A többi rejtett osztály az /1, 1/, /3, 3/ és az /5, 5/ celláknak felel meg. Formálisan a következőképpen definiálhatjuk a modellt:

$$(21) \qquad \pi^{OD}_{ij} = \Sigma_t \pi^{x}_{t}\, \pi^{\bar{o}x}_{jt}\, \pi^{\bar{D}x}_{jt} \qquad t = 1, \ldots, 5$$

$$\text{ahol} \quad \pi^{\bar{o}x}_{13} = \pi^{\bar{o}x}_{34} = \pi^{\bar{o}x}_{55} = \pi^{\bar{D}x}_{13} = \pi^{\bar{D}x}_{34} = \pi^{\bar{D}x}_{55} = 1$$

$$\pi^{\bar{o}x}_{51} = \pi^{\bar{o}x}_{12} = 0$$

Goodman jelölésével a 21 modell igy irható: /XO, XD/. Pontosabban mondva ez nem igaz, mert ez a jelölés nem utal a megszoritásokra. A 21 alatti megszoritások az ezután következő valamennyi modellre érvényesek. A rejtett változó osztályait minden esetben ugyanigy definiáljuk. Az /XO, XD/ modell szabadsági fokainak száma 6. A szabadságfokok száma egyenlő a cellák száma minusz egy minusz a függetlenül becsült paraméterek száma. A becsült paraméterek száma 18.[29]

2.2 A táblák szimultán elemzése

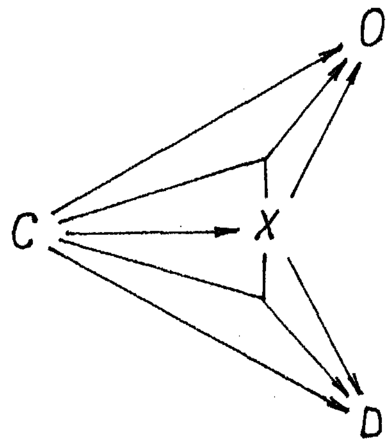
Vezessük be a vizsgálatba az ország-változót, jelöljök C-vel. A rejtett változó igy 1o osztályban vehet fel értéket /5 az egyik és 5 a másik ország esetében/. A rejtett változót ekkor felirhatjuk /C, X/ formában is, ahol X a 2.1 §-ban taglalt X változó és C az ország változója.[30] A C változó megfigyelhető és mint ilyen pontosan mutatja a rejtett C változó értékét, a következő megszoritások alapján:

(22) $$\pi^{\bar{C}X}_{1t} = \pi^{\bar{C}X}_{2t} = 1 \qquad t=1...5$$

Végeredményben az összetett (C, X) változó szintjei egyrészt az egyik, másrészt a másik országbeli X rejtett változó 5-5 szintje. Ennek a modellnek /CXO, CXD/ lehet a jelölése. Azok a tesztek, amelyeket a 2.1 §-ban érintettünk, érvényesek mindkét ország /XO, XD/ modelljére. A továbbiakban a modelleket csak sematikusan vázoljuk, az esetleges külön megszoritásokat csak megemlitjük:[31]

/CXO, CXD/ modell:

(23)



A /CXO, CXD/ modell a két országra vonatkozó /XO, XD/ modellek összege. Külön megszoritásokat nem definiálunk, igy a szabadsági fokok száma $2 \cdot 6 = 12$. A 2. táblázatban az illeszkedési mutatókat közöljük.[32] Ebben a χ -négyzeten és az L-en kivül még a következők is szerepelnek:

- különbözőségi index[33]
- a normált illeszkedési index /BONETT és BENTLER-1983, p.157/

$$(24) \qquad R_L = \frac{L_b - L_u}{L_b}$$

ahol L_b a megszoritott modell Loglikelihood hányadosa

L_u a megszoritás nélküli modell Loglikelihood hányadosa

- módositott X-négyzet /WONNACOTT és WONNACOTT-1979, - p.366/[34]

$$(25) \qquad F = L/_{df}$$

2. táblázat

Különböző modellek illeszkedési mutatói

Modell	df	L	χ^2	Δ	R_L^+	F
/CXO, CXD/	12	21.24	20.69	1.73	.99	1.77
/C, X//CXO, CXD/	16	50.78	50.74	3.02	.97	3.17
/CX, CO, CD, XO, XD/	20	50.29	47.61	2.57	.97	2.51
/CO, CD, XO, XD/	24	94.87	92.25	4.38	.94	3.95
/CX, XO, XD/	26	216.54	216.36	6.76	.85	8.33
/C, XO, XD/	30	406.48	404.90	9.57	.72	13.55

+ Az összehasonlitási alapot jelentő független modellben: L=1465; df=32

E modellnek az eredményei a 3. táblázatban láthatók. A paraméterbecslések ugyanazok, mint a Clogg félék, az illeszkedési mutatók a különkülön Dániára és Angliára számolt mutatók összegei.

3. táblázat

Paraméterbecslések a /CXO, CXD/ modellre+

k, t	prob.	$\hat{p}^{x}_{t}$	$\hat{p}^{\bar{c}x}_{1t}$	$\hat{p}^{\bar{c}x}_{2t}$	$\hat{p}^{\bar{o}x}_{1t}$	$\hat{p}^{\bar{o}x}_{2t}$	$\hat{p}^{\bar{o}x}_{3t}$	$\hat{p}^{\bar{o}x}_{4t}$	$\hat{p}^{\bar{o}x}_{5t}$	$\hat{p}^{\bar{D}x}_{1t}$	$\hat{p}^{\bar{D}x}_{2t}$	$\hat{p}^{\bar{D}x}_{3t}$	$\hat{p}^{\bar{D}x}_{4t}$	$\hat{p}^{\bar{D}x}_{5t}$
1,1	0.251	0.102	1.000	0.000	0.071	0.450	0.342	0.137	0.000	0.084	0.384	0.375	0.125	0.032
1,2	0.624	0.254	1.000	0.000	0.000	0.032	0.242	0.466	0.260	0.010	0.022	0.195	0.505	0.269
1,3	0.006	0.003	1.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
1,4	0.059	0.024	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
1,5	0.059	0.024	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000
2,1	0.221	0.131	0.000	1.000	0.111	0.507	0.178	0.204	0.000	0.077	0.438	0.165	0.253	0.067
2,2	0.691	0.411	0.000	1.000	0.000	0.043	0.137	0.560	0.260	0.000	0.062	0.117	0.510	0.310
2,3	0.012	0.007	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2,4	0.014	0.008	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
2,5	0.062	0.037	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000

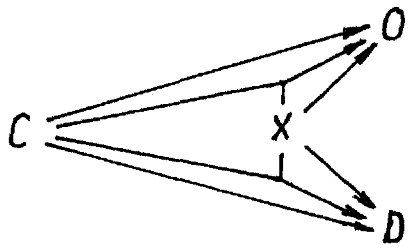
+ A különböző mintanagyságok miatt a rejtett valószinüségeket ujra kell számitani a Clogg-féle adatokkal való összehasonlitáshoz.

A következő modellben semmiféle asszociációt nem feltételezünk a C és X változók között. Más szóval a rejtett változó relativ gyakoriságát azonosnak tekintjük mindkét országban. Egy további megszoritást kell kizárni:

(26) $\pi^{\bar{x}c}_{t1} = \pi^{\bar{x}c}_{t2}$ t=1,....5

Ez a modell /C,X/ /CXO,CXD/. Sematikusan:

(27)



A /C,X//CXO,CXD/ modellben nincs összefüggés az ország és a rejtett változó osztályainak relativ nagysága között. A rejtett változót megadva az eredet- és célosztályok relativ gyakoriságainak eloszlása különböző lehet. Nincs egyszerü módszer a modell becslésére.
A $\pi^{\bar{x}c}_{tk}$ valószinüségek nem paraméterek a rejtett osztály modellben. Feljebb láttuk, hogy a rejtett változó az ország (C) és az osztály (X) kombinációja. A LCAG programban lehetőség van arra, hogy a rejtett változót több változóból állónak tekintsük és alkotó részei között függetlenséget tételezzünk fel. Ilyen eset az is amikor a kiindulási O, és a cél változó D feltételes eloszlása adott az X köztes változó feltételével az egyes országokban, ezek a feltételes eloszlások különbözők lehetnek. A (21) kezdeti feltétel értelmében lehetséges, hogy mindkét országra vonatkoztatva azonos nagyságu meghatározó státuszosztályokat definiáljunk, A modell, melynek szimboluma /C,X//CXO,CXD/ olyan kezdeti feltételeket feltételez, melyek a rejtett változó gyakorisági eloszlására vonatkoznak. Nem 2•4, hanem csak 4 paramétert kell megbecsülni, a szabadsági fokok száma ezért 12+4=16.

A loglikelihood hányados 50,78, ez nem mutat tul jó illeszkedést. Pusztán tájékoztatásul közöljük a paraméterbecsléseket a 4. táblázatban.
Már első ránézésre, de hosszasabb vizsgálódással is megállapitható, hogy a modell azért viselkedik ennyire előnytelenül, mert mindkét ország esetében a 3-as státuszcsoportban nagy öröklődési különbség lép fel.

4. táblázat

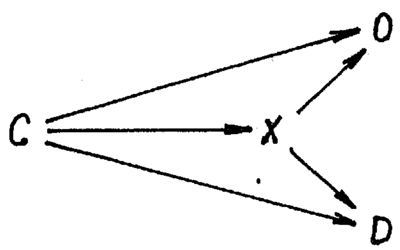
Paraméterbecslések a /C,X/ /CXO, CSD/ modellben+

k,t	prob.	$\hat{p}^{x}_{t}$	$\hat{p}^{\bar{c}x}_{1t}$	$\hat{p}^{\bar{c}x}_{2t}$	$\hat{p}^{\bar{d}x}_{1t}$	$\hat{p}^{\bar{d}x}_{2t}$	$\hat{p}^{\bar{d}x}_{3t}$	$\hat{p}^{\bar{d}x}_{4t}$	$\hat{p}^{\bar{d}x}_{5t}$	$\hat{p}^{\bar{b}x}_{1t}$	$\hat{p}^{\bar{b}x}_{2t}$	$\hat{p}^{\bar{b}x}_{3t}$	$\hat{p}^{\bar{b}x}_{4t}$	$\hat{p}^{\bar{b}x}_{5t}$
1,1	0.255	0.104	1.000	0.000	0.066	0.408	0.415	0.111	0.000	0.077	0.362	0.447	0.093	0.021
1,2	0.655	0.266	1.000	0.000	0.000	0.042	0.255	0.456	0.247	0.010	0.025	0.210	0.496	0.260
1,3	0.010	0.004	1.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
1,4	0.019	0.008	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
1,5	0.061	0.025	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000
2,1	0.255	0.152	0.000	1.000	0.098	0.503	0.177	0.223	0.000	0.068	0.400	0.159	0.286	0.087
2,2	0.655	0.389	0.000	1.000	0.000	0.022	0.132	0.571	0.275	0.000	0.059	0.113	0.511	0.316
2,3	0.010	0.006	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2,4	0.019	0.011	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
2,5	0.061	0.036	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000

+A különböző mintanagyságok miatt a rejtett valószinüségeket ujra kell számitani a Clogg-féle adatokkal való összehasonlitáshoz.

A következő modellt tárgyaljuk: /CS, CO, CD, XO, XD/.
Egyetlen magasabbrendü interakciót sem tételezünk fel a C, X, O és D változók között. Az a lényeg, hogy a kiindulási és a cél változó eloszlása különböző a különböző országokra és a különböző X változó kategóriákra nézve, de nem a CX kereszt-osztályozásra nézve / azaz nincs CXO és CXD interakció/. Amilyen könnyü ezt a modellt definiálni, olyan nehéz becsülhető rejtett struktura modellként felirni. Az alkalmazott stratégia a következő: definiáljunk egy 250 értékü rejtett változót, ez tkp. a C', O' és X rejtett változók kereszt-osztályozása, itt az első három egyértelmüen megfeleltethető a megfigyelhető C, O és D változókhoz.

(28)



A loglineáris modell a C, X, O, D rejtett változókkal definiálható.[37]
A χ - négyzet statisztikai értéke 50.29, a szabadsági fokok száma 12 + 2 x 4 = 20
A 3. és 4. sz. táblázathoz hasonlóan az eredmények az 5. sz. táblázatban láthatók:

5. táblázat

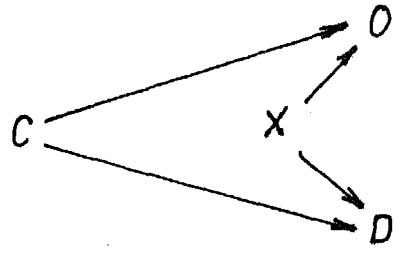
Paraméterbecslések a /CX, CO, CD, XO, XD/ modellben+

k,t	prob.	p^{x}_{t}	$\hat{p}^{\bar{C}X}_{1t}$	$\hat{p}^{\bar{C}X}_{2t}$	$\hat{p}^{\bar{O}X}_{1t}$	$\hat{p}^{\bar{O}X}_{2t}$	$\hat{p}^{\bar{O}X}_{3t}$	$\hat{p}^{\bar{O}X}_{4t}$	$\hat{p}^{\bar{O}X}_{5t}$	$\hat{p}^{\bar{D}X}_{1t}$	$\hat{p}^{\bar{D}X}_{2t}$	$\hat{p}^{\bar{D}X}_{3t}$	$\hat{p}^{\bar{D}X}_{4t}$	$\hat{p}^{\bar{D}X}_{5t}$
1,1	0.259	0.105	1.000	0.000	0.069	0.434	0.360	0.137	0.000	0.091	0.331	0.351	0.183	0.045
1,2	0.616	0.250	1.000	0.000	0.000	0.034	0.234	0.470	0.262	0.006	0.040	0.204	0.486	0.265
1,3	0.006	0.004	1.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
1,4	0.059	0.024	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
1,5	0.060	0.024	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000
2,1	0.180	0.107	0.000	1.000	0.135	0.563	0.173	0.129	0.000	0.073	0.511	0.186	0.182	0.047
2,2	0.795	0.433	0.000	1.000	0.000	0.055	0.141	0.561	0.243	0.005	0.065	0.115	0.516	0.299
2,3	0.013	0.007	0.000	1.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
2,4	0.014	0.008	0.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
2,5	0.065	0.038	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000

+ A különböző mintanagyságok miatt a rejtett valószinüségeket ujra kell számitani a Clogg-féle adatokkal való összehasonlitáshoz.

A következő modellben ismét feltételezzük C és X függetlenségét /CO, CD, XO, XD/:

(29)



A χ-négyzet értéke 94.87, a szabadsági fokok száma 20 + 4 = 24. A modell viszonylagos rossz illeszkedését ismét a (3, 3) cellának tudhatjuk be.

A következő modell esetében a rejtett változó gyakorisága eloszlása országonként más és más. A feltételes gyakorisági eloszlások feltéve a rejtett változó egyes osztályaiba tartozást azonban országonként azonosak. A következő kezdeti megszoritásokat adjuk meg:[39]

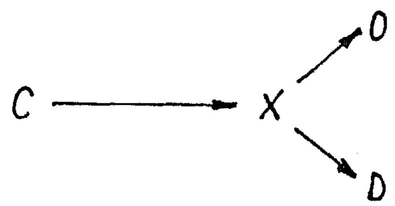
(30a) $\pi^{\bar{O}XC}_{it1} = \pi^{\bar{O}XC}_{it2}$ $\quad$ i=1....5 $\quad$ t= 1,....5

(30b) $\pi^{\bar{D}XC}_{it1} = \pi^{\bar{D}XC}_{it2}$ $\quad$ i=1....5 $\quad$ t= 1,...5

A /CX, XO, XD/ modell megszoritásai a kiindulási és a cél változók feltételes gyakorisági megoszlásaira vonatkoznak. 2x14 paraméter helyett csak 14 paramétert becsülünk, az érvényes szabadsági fok 12+14=26.

Ez a modell sematikusan a következőképpen néz ki.

(31)



Ennek a modellnek az illeszkedése : 216.54. / χ^2 /

A (26) (30a) és (30b) megszoritások együttes előirásával a (27) és (31) modell keverék modelljét állithatjuk fel. Ez mindkét országra azonos gyakorisági megoszlásokat és azonos feltételes gyakorisági megoszlásokat ir elő a rejtett változóra nézve. Az eddigi szimbólumok alkalmazásával ez az uj modell /C,XO,XD/ és a megszoritások a /C,X//CXO,CXD/ és a /CX,XO,XD/ modellek megszoritásai, igy a szabadsági fokok száma: 12+4+14 = 32.

Sematikusan ez a modell a következő:

(32)
$$\begin{tikzcd} & & O \\ C & X \arrow[ur] \arrow[dr] & \\ & & D \end{tikzcd}$$

Ennek a modellnek az illeszkedése: 406.48[40]. /χ^2/

Végeredményben tehát alapul véve Clogg vizsgálatait, bemutattuk, hogy milyen lehetőségek vannak különböző országok társadalmi mobilitási tábláinak összehasonlitására felhasználva a loglineáris elemzést, ill. annak rejtett változó szerinti osztályozásra való alkalmazását.[41]

JEGYZETEK

1. A determináns-egyenleteket felhasználó ilyen módszer az "Anderson-Lazarsfeld-Dudman féle". Ez csak bizonyos feltételek teljesülése esetén szolgáltat konzisztens becsléseket. Ha a változók dichotomok és a rejtett osztályok száma kevesebb mint (1/2) x (m+1) ahol a tábla dimenzioszáma m. /LAZARSFELD és HENRY-1968 4. fejezet/. Továbbá ez a módszer csak aszimptotikusan efficiens paramétereket eredményez, ha a tábla háromdimenziós kontingencia-tábla és a rejtett osztályok száma 2. /ANDERSON-1959/. Az is előfordulhat, hogy a módszer nem megfelelő becsléseket ad, olyan értelemben, hogy azok értéke nincs a /o,1/ intervallumban, sőt esetleg komplex. Lásd még /ANDERSON-1954, GIBSON-1954, MADANSKY-1960/

2. A Haberman féle algoritmust precizebben taglaljuk ennek a dlgozatnak egy későbbi folytatásában.

3. Loglineáris modellként felirva:

(33) $$\log m_{ijt}^{ODX} = \mu + \lambda_t^X + \lambda_i^O + \lambda_j^D + \lambda_{ti}^{XO} + \lambda_{tj}^{XD}$$

A paraméterekre vonatkozó szokásos megszoritásokkal.

4. A Newton-Raphson algoritmus egy változata a "scoring"-algoritmus/ /HABERMAN-1979 p. 542/. A számitás menete hasonlatos egy sulyozott regresszió analizisére.

5. Egyéb eljárások lehetnek: Caroll kanonikus dekompoziciós algoritmusa, amely nem maximum likelihood megoldás /CLOGG-1981a, p. 221/ Formann algoritmusa pedig a modell paramétereinek egy logit-transzformációja után egyfajta gradiens módszert alkalmaz.

6. A LCAG outputja a következőket tartalmazza:

- a rejtett változó osztályainak várható valószinüségei
- a megfigyelhető változók osztályainak várható feltételes valószinüségei feltéve a rejtett változó egy-egy értékének felvételét.
- a megfigyelhető változók kereszt-osztályozásainak /cellák/ m megfigyelt és várható gyakoriságait.
- a rejtett változó osztályainak várható gyakoriságait
- a χ-négyzet és L statisztikákat.

Lehetőség van a rejtett és a megfigyelhető változók kereszttáblázására, azaz az ilyen cellák gyakoriságainak tárolására. Ennek akkor van jelentősége, ha a loglineáris hatásokkal is meg akarjuk határozni /FREQ, GLIM, ECTA felhasználásával/; ez persze akkor értelmes, ha a modell jól illeszkedik.

7. Kezdeti becslésként alkalmazható a determináns-módszerrel nyert eredmény, ha az benne van a /o, 1/ intervallumban. Goodman /1974 pp. 1251-1255/ ennél összetettebb eljárásokat is leir a kezdeti becslések előállitására.

8. Haberman /1979 pp. 545-546/ leir egy megfelelő iterativ algoritmust, amely nem a scoring-algoritmus, amelyhez a kezdő becsléseket az 1. lépésben a megfigyelt táblából nyeri, ekkor a megfelelő loglineáris paramétereket is meg kell határoznia - igen fáradságos módon.

9. Olyan megszoritások, amelyek a paraméterekre o, vagy 1 értékeket irnak elő könnyen teljesithetők a kezdeti értékek megválasztásával, másfajta megszoritások beépitése az algoritmus 5. lépésénél igényel számitásokat az iteráció minden lépésében.

10. Nem könnyü eldönteni, hogy milyen statisztikai próbák alkalmasak leginkább, az L statisztika jól alkalmazható pl. particionálás esetén.

11. Goodman /1974a, p. 1245, 95 jegyzet/ rámutat arra, hogy tapasztalata szerint az iterativ eljárás nem vezet lokális maximumokhoz. Egy globális maximum olyan paramétervektort ad, amely maximalizálja a likelihood-függvényt a megengedhető paraméterek teljes halmazán /a megengedhetőség azt jelenti, hogy az egyes paramétervektor elemek szigoruan a /o, 1/ intervallum belsejébe esnek/ A lokális maximum a kezdeti vektor környezetére ad ilyen valószinüség-maximumot, amely nem szükségképpen globális maximum is. Harmadik lehetőség a terminális maximum, amely lehet lokális vagy globális, de vannak olyan paramétervektor elemek, amelyek a megengedhetőség határán vannak, azaz értékük o, vagy 1.

12. Clogg bevezet egy $\lambda_{X.OD}$ arányos hibacsökkenési mértéket arra, hogy figyelembe vegye a rejtett strukturák valamely loglineáris modelljének érvényességét az általánosan használt statisztikus próbáktól eltérő elvi alapokon. /CLOGG-1981b, p. 840/. A dolgozat egy későbbi változatában erre is ki fogunk térni.

13. Az identifikálhatóság problémájának rövid áttekintése Goodman /1974b/ alapján: A kérdés az, hogy modell paramétereinek maximum likelihood becslései lokálisan identifikálhatók-e. Induljunk ki a becsülendő paraméterek számából: a $\hat{P}_t^X$ paraméterekből T-1 paraméter vehető figyelembe, a $\hat{P}_{it}^{\bar{A}X}$ paraméterekből I-1 minden t-re, a $\hat{P}_{jt}^{\bar{B}X}$ paraméterekből pedig J-1 minden t-re /ld 3/. Igy tehát a becsülendő paraméterek teljes száma:

$$T-1+(I-1)T+(J-1)T=(I+J-1)T-1.$$

A megfigyelhető valószinüségek száma IJ-1.

Ha a (34) formula érvényes, akkor a rejtett (feltételes) valószinüségek száma nagyobb a megfigyelhető valószinüségek számánál:

(34) $\quad IJ - 1 < (I+J-1)\,T - 1$

Ez a nem identifikálhatóság elégséges feltétele.

Ha a (34) formula nem áll fenn, akkor (4) formulát deriválhatjuk a várható megfigyelhető valószinüségek szerint, pl.:

(35a) $$\frac{\partial \hat{P}^{OD}_{ij}}{\partial \hat{P}^{X}_{t}} = \hat{P}^{\bar{O}X}_{it}\,\hat{P}^{\bar{D}X}_{jt} - \hat{P}^{\bar{O}X}_{iT}\,\hat{P}^{\bar{D}X}_{jT}$$

(35b) $$\frac{\partial \hat{P}^{OD}_{ij}}{\partial \hat{P}^{\bar{O}X}_{it}} = \begin{cases} \hat{P}^{X}_{t}\,\hat{P}^{\bar{D}X}_{jt} & i = 1 \ldots\ldots I-1 \\ -\hat{P}^{X}_{t}\,\hat{P}^{\bar{D}X}_{jt} & i = I \\ 0 & \textit{különben} \end{cases}$$

Ha a fenti mátrix rangja azonos az oszlopok számával, akkor a modell paraméter-becslései lokálisan identifikálhatók.

14. Ha megszoritásokat adunk meg, ez a teszt nem használható /HAGENAARS és VAN DER WALLE-1983, p.7/

15. A mátrixból azok az oszlopok, amelyek ismertnek feltételezett paraméterbecslésekhez tartoznak, elhagyandók. Olyan oszlopok pedig, amelyek azokhoz a paraméterbecslésekhez tartoznak, amelyekre egyenlőséget irtunk elő, pedig hozzáveendők a mátrixhoz. A mátrix oszlopainak a száma tehát (I+J-1)T - 1 + d. Itt d a modellhez felvett nem-redundáns megszoritások száma. Ezek után a mátrix rangjára érvényes a 13. jegyzet.

16. A (4) formula átirható a következőképpen:

(36) $$\pi^{OD}_{ij} = \sum_{t=2}^{T} \theta^{X}_{t}\,\pi^{\bar{O}X}_{it}\,\pi^{\bar{D}X}_{jt}$$

$$\text{ahol} \quad \theta^{X}_{t} = \begin{cases} \pi^{X}_{1} + \pi^{X}_{2} & t = 2 \\ \pi^{X}_{t} & t = 3 \ldots T \end{cases}$$

17. A (4) formula átirható a következőképpen is:

(37)
$$\pi_{ij}^{ob} = \sum_{t=2}^{T} \theta_t^{x}\, \theta_{it}^{\bar{o}x}\, \pi_{jt}^{\bar{b}x}$$

ahol θ_t^x *ld (36)*

$$\theta_{it}^{\bar{o}x} = \begin{cases} (\pi_1^x\, \pi_{i1}^{\bar{o}x} + \pi_2^x\, \pi_{i2}^{\bar{o}x})/\theta_2^x & t=2 \\ \pi_{it}^{\bar{o}x} & t=3\ldots T \end{cases}$$

Goodman /1974b, pp. 225,226/ megmutatja, hogy ez általánositható n-dimenziós kontingencia táblákból való kiindulás esetére is, amikor az 1. és 2. rejtett osztályhoz tartozó feltételes valószinüségek azonosak az n illetve n-1 változóra nézve.

18. A brit adatok foglalkozási kategóriai a következők:

(1) értelmiségi és felső szintü vezetés

(2) . vezető és végrehajtó

. egyéb nem fizikai magas képzettséggel

(3) . egyéb nem fizikai alacsonyabban képzett

(4) . alacsony képzettségi szintü nem fizikai

. fizikai, szakmunka

(5) . fizikai, betanitott

. fizikai, segédmunka

Clogg szerint /1981b, p. 845/ Svalagosta elvégezte az emlitett összevonásokat, hogy az adatok összevethetők legyenek a dániaiakkal.

19. Egy ujabb vizsgálatban, mely az osztályok elméletéhez óhajt adalékokat szolgáltatni, Parkin /1979/ magát a /foglalkozási/ mobilitást tekinti osztálytagozódásnak. Sematikusan ábrázolva:

$$(\mathrm{O} - \mathrm{D}) \longrightarrow \mathrm{X}$$

20. Jóllehet a becsült paraméterek száma 17 és igy kevesebb 24-nél a modellben fellépnek előre nem látható összefüggések.

21. Először is Clogg /1981b, p. 848/ rámutat arra, hogy ha nem tette volna fel, hogy a táblában a legalacsonyabb és a legmagasabb státuszokhoz tartozó értékek a priori ismeretek, akkor más megszoritásokat kellett volna alkalmazniuk. Másrészt a szerző inkább a cél változóra alkalmazna megszoritásokat, mivel ugy tünik, hogy

valószinübb, hogy valaki bizonyos státuszból indulva nem érhet el bizonyos poziciókat, mint az, hogy egy adott státuszban lévén bizonyos osztályból nem indulhatott volna. Az összehasonlitás érdekében a Clogg féle eljárást követtük. Természetesen egy alternativa az, hogy mindkét megszoritási tipust előirjuk.

22. A LCAG nem áll meg a o-val való osztásra, de kihagyja a $\hat{P}^{ob\bar{x}}_{ijt}$ -re való számitást. Eredményül o gyakoriságot kapunk. A χ négyzet érték kiszámitásakor ennek a cellának a hozzájárulása o, ez a Pearson féle χ-négyzet értéket nem befolyásolja, de a loglikelihood hányadost igen. A leginkább ajánlatos ebben az esetben ennek a cellának a megfigyelt gyakoriságát is o-nak tekinteni. Haberman LAT programja hasonlóképpen jár el. A LCAG használatakor nem okoz problémát a o megfigyelt gyakoriságu cellák előfordulása, Clogg vizsgálata során azonban ez prolematikus. Clogg egy önkényes módszerrel megnövelte a megfigyelt gyakoriságokat, 0,10 ill. 1,00 értékre. Clogg ennek indoklásául emliti az országok közötti szabadsági fokok összehasonlitásának lehetőségét.

23. A paraméterbecslések száma a kezdeti feltételek figyelembe vétele nélkül 26.

24. A több-státuszon keresztüli mobilitás le lett tiltva minden egyes státusz-osztályba tartozó mobil részére /CLOGG-1981b, p. 853/

25. Az LCAG vagy az MLLSA program használata esetén a becsült várható gyakoriságok a kvázi-rejtett-struktura modell és a kvázi-független modell esetében nem szükségképpen azonosak. Ez abból következik, hogy a Goodman algoritmusban természetesen nem lehetnek negativ valószinüségek, ezáltal a rejtett osztályokat státuszosztályokkal definiálva azok valószinüsége legalább o. A model specifikációjából logikailag következik, hogy csak az örökölhetőség mérhető és annak ellenkezője nem, azaz csak a diagonális cellákban szereplő többlet /a független modellhez képest/ mérhető.
Haberman LAT programja viszont azonos eredményeket ad a két modellre, mivel ez az algoritmus közvetlenül a loglineáris hatásokat számitja. A várható gyakoriságok nem az X osztályai várható gyakoriságainak, az összegei, hanem a sulyozott loglineáris paramétereknek.

Ezek a loglineáris paraméterek lehetnek negativak is. Mindazonáltal a kvázi-független és a kvázi-rejtett strukturát tartalmazó modell formálisan ekvivalensek.

26. Ez az immobilitási arány az eddig emlitett okokból kifolyólag mindig nagyobb egynél. Ha azonban a számitás a LAT programmal történik, az immobilitási arány a loglineáris paraméterek függvényeként irható fel, következésképpen bármely pozitiv értéket felvehet.

27. Ugy gondoljuk, hogy számos egyéb probléma is felmerül ezzel a mobilitási aránnyal kapcsolatban, igy például az, hogy levásott egyenleteken alapul /HOPE-1981; HAUSER-1981/. Erre azonban más helyen fogunk kitérni.

28. Clogg /1981b, p. 853/ rámutat arra, hogy K statusosztály választható, ahol $K \leq I$

29. A rejtett változónak öt osztálya van, a becsülendő paraméterek száma 4/ mivel a valószinüségek összege 1/ Az 1 és 2 osztályra adott feltételes eloszlások a kiindulási és cél változóra 5 értéküek, igy a becsülendő paraméterek száma a kiindulási és a cél változóra 5-5 osztályonként. Mivel a kiindulási-változó osztályaira osztályonként egy-egy megszoritás lett megadva, a cél-változóra kezdeti feltétel nem lett megadva, ezért a becsülendő paraméterek az iménti 5-5 helyett 3 ill. 4 osztályonként, azaz a feltételes paraméterek közül becsülendő 3+3+4+4=14. Az összes becsülendő paraméter pedig 14+4=18

30. V. ö. Wiggins /1973/; periódus helyett ország értendő

31. Ha az egyértelmü indikátorokon kivül a rejtett osztályokat is figyelembe akarjuk venni, a következő sémákat alkalmazhatjuk:

$$(38) \qquad C = C' \qquad X \qquad \begin{matrix} O' = O \\ \\ D' = D \end{matrix}$$

C; O' és D' ekkor a rejtett változók és C, O és D egyértelmü indikátoraik. Az egyszerüség kedvéért azonban a továbbiakban a rejtett változókat nem jelöljük külön. Ne felejtsük el, hogy a loglineáris modell a rejtett változók értékeire is formálisan definiálható /LISREL/ ahol is a rejtett változók akár mindannyian lehetnek a megfigyelt változók egyértelmü indikátorai.

32. Clogg /1981b, p. 845/ amlit meglehetősen komplikált mintavételi eljárásokat és azok hatását a χ -négyzet próba szignifikanciaszintjére.

33. A tapasztalat azt mutatja, hogy a különbözőségi index függ a minta nagyságától.

34. Ennek részleteibe a dolgozat későbbi folytatásában fogunk elmerülni.

35. A csekély különbség: 8,2+12,9 = 21,1 és a 21,2 érték között annak köszönhető, a kerekitésen kivül, hogy az (5,1) cella o gyakoriságát /az angol adatokban/ Clogg o,10 gyakorisággal helyettesitette.

36. Viszonylag egyszerü, de meglehetősen gépidő igényes olyan modellt definiálni, ahol minden rejtett osztályra egyenlőségeket irunk elő, a 3. státuszosztályhoz tartozó kivételével. A dolgozat következő részében bevezetünk egy ilyen modellt.

37. Vegyük észre, hogy az összes modellt lehetne ilyen módon definiálni. Valójában ez a legegzaktabb módja a modell definicióinak. Egyetlen oka annak, hogy ezt ha nem szükséges, mégsem igy teszszük, az, hogy az ily módon megadott modell a számitások során kb 1oo százalékos gépidő-többletet okoz.

38. Valójában a tábla 25o rejtett osztályt tartalmaz, amelyeket a C, D és D változók összes lehetséges kereszt-osztályozása - ezek a megfigyelhető változók - definiái.

39. Itt ismét alkalmazható a számitásnak egy könnyebb módja.

4o. A gyenge illeszkedés ismét a (3, 3) cellának tudható be.

41. Elvi következtetéseinket e dolgozat későbbi folytatásában fogjuk öszszefoglalni.

REFERENCIÁK

Anderson, T.W.

1954 "On estimation of paramters in latent structure analysis"
Psychometrika 19: 1-10.

1959 "Some scaling models and estimation procedures in the latent class models"
in: U. Grenander /ed./: Probability and Statistics.
New York: Wiley

Bonett, Douglas C és P.M. Bentler

1975 "Goodness-of-Fit Procedures for the Evaluation and Selection of Log-Linear Madels"
Psychological Bulletin 93: 149-166

Boudon, Raymond

1973 Mathematical Structure of Social Mobility
Amsterdam: Elsevier

Clogg, Clifford C

1977 "Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis: A Manual for Users"
Pennsylvania State University, Population Issues Research Office: Working Paper No.1977-09

1979 "Some Latent Structure Models for the Analysis of Likert-Type Data"
Social Science Research 8: 287-301

1980 "Characterizing The Class Organization Of Labor Market Opportunity. A madified latent structure approach"
Sociological Methods and Research 8:243-72

1981a "New Developments in Latent Structure Analysis"
in: D.J. Jackson and E.F. Bargotta (eds)
Factor Analysis and Measurement, London/Beverly Hills: Sage

1981b "Latent Structure Models of Mobility"
American Journal of Sociology 86: 836-68

Formann, A.K.
1978 "A Note on Parameter Estimation for Lazarsfeld's Latent Class Analysis", Psychometrika 43 : 123-126

Gibson, W.A.
1955 "An extension of Anderson's solution for the latent structure aquations"
Psychometrika 20: 69-73

Glass, David.V (ed)
1954 Social Mobility in Britain. London: Routledge and Kegan Paul.

Goodman, Leo A.
1969b "On the Measurement of Social Mobility: an Index of Status Persistence"
American Sociological Review 34: 831-50

1974a "The Analysis of Systems of Qualitative Variables when Some of the Variables are Unobservable. Part I - a Modified Latent Structure Approach"
American Journal of Sociology 79: 1179-1259

1974b "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models"
Biometrika: 215-231

Green, Bert F.,Jr.
1952 "Latent Structure Analysis and its Relation to Factor Analysis" , Journal of the American Statistical Association 47: 71-76

Haberman, Shelby
1974 "Log-linear Models for Frequency Tables Derived by Indirect Observation: Maximum Likelihood Equations",
The Annals of Statistics 2: 911-924

1977 "Product Models for Frequency Tables Involving Indirect Observation", The Annals of Statistics 5: 287-301

1978 Analysis of Qualittative Data, Volume 1, Introductory Topics. New York: Academic Press

1979 Analysis of Qualitative Data, Volume 2, New Developments. New York: Academic Press

Hagenaars, Jacques A.P.

1976 "Latente $V_a$riabelen" (Latent Variables) Sociale Wetenschappen 19: 250-78

1978 "Latent Probability Models with Direct Effects between Indicators" Quality and Quantity 12: 205-22

1980 "Multivariate Tabelanalyse: Elaberative van Sammenhang" (Multivariate Contingency Table Analysis: Elaboration of Association) Pp 119-148 in: J.H.G. Segers and J.A.P. Hagenaars (eds) Sociologische Onderzoeksmethoden deel II Technieken van Causale Analyse (Sociological Methods of Research volumen II Techniques of Causal Analysis) Assen: Van Gorcum

Hagenaars, Jacques A.P. and Ton G.J.J. Heinen

1980 "Analyse van Contingentietabellen met behulp van het Loglinearie Model" (Analysis of Contingency Tables with the Loglinear Model) Pp 185-258 in: J.H.G. Segers and J.A.P. Hagenaars (eds.) Sociologische Onderzoeksmethoden deel II Technieken van Causale Analyse (Sociological Methods of Research volume II Techniques of Causal Analysis) Assen: Van Gorcum

Hagenaars, Jacques A.P. and Martin A.G. Van der Walle

1983 "LCAG - Een Computerprogramma voor de Analyse van Loglineaire Modellen met Latente Variabelen"
(LCAG- A Computerprogram for the Analysis of Loglinear Models with Latent Variables)
Tilburg University: Department of Sociology.

Hauser, Robert M.

1978 "A Structural Model of the Mobility Table"
Social Forces 56: 919-53

1979a Erratum to Hauser - 1978
Social Forces 57: 777

1979b "Some Exploratory Methods for Modelling Mobility Tables and Other Cross-Classified Data"
Pp. 413-458 in: K.F. Schuessler (ed)
Sociological Methodology 1980.
San Francisco: Jossey-Bass

1981 "Hope for the Mobility Ratio"
Social Forces - 60: 544-556

Lazarsfeld, Paul Felix

1950a "The Logical and Mathematical Foudation of Latent Structure Analysis"
Chapter 10 in: Samuel Stouffer et al
Studies in Social Psychology in World War II.
Vol. IV Measurement and Prediction.
Princeton: Princeton University Press

1950b "The Interpretation and Computation of Some Latent Structures"
Chapter 11 in: Samuel Stouffer et al.
Studies in Social Psychology in World War II.
Princeton: Princeton University Press

1954 "A Conceptual Introduction to Latent Structure Analysis."
Pp. 349-387 in: P.F.Lazarsfeld (ed.)
Mathematical thinking in the social sciences.
Glencoe (Ill.): the Freee Press

1955 "Interpretation of Statistical Relations as a Research Operation." Pp. 115-125 in: P.F. Lazarsfeld and M. Rosenberg (eds), The Language of Social Research
Glencoe (III): the Free Press

1959 "Latent Structure Analysis." Pp. 476-543 in: S. Koch (ed) Psychology: a study of a science, vol. 3. New York.

Lazarsfeld, Paul F. és Neil W. Henry

1968 Een loglineaire analyse van mobiliteitstabellen
(A loglinear analysis of mobility tables) (Master's thesis)
Tilburg University: Department of Sociology.

Madansky, A.

1960 "Determinantal methods in latent class analysis"
Psychometrika 25: 183-198

McHugh, R.B

1956 Efficient estimation and local identification in latent class analysis"
Psychometrika 21: 311-347

Parkin, Frank

1979 Marxism and Class Theory: A Bourgeois Critique.
Cambridge: Tavistock

Rosenberg, M.

1968 The Logical Survey Analysis
New York: Basic Books

Rossi, P.H.

1951 "The application of Latent Structure Analysis to the empirical study of social stratification" (Unpublished doctoral dissertation)
Columbia University

Svalagosta, K.

1959 Prestiqe, Class and Mobility. London: Heinemann

Wiggins, L.M.

1973 Panel Analysis: Latent Probability Models for Attitude and Behavior Processes.
Amsterdam: Elsevier.