

STRUCTURILE LATENTE ALE DISTANȚEI SOCIALE

CORNELIA RADA

Scopul acestui articol este să aducă argumente solide pentru utilizarea analizei claselor latente ca un instrument alternativ de investigare în științele sociale și comportamentale. Pornesc de la prezentarea tehnică a metodei de analiză prin clase latente, limitată la atât cât este necesar pentru înțelegerea elementelor sale de bază, continuând apoi cu un exemplu de cercetare empirică. Cu ajutorul programului Latent Gold¹ mi-am propus să identific patternuri ale distanței sociale manifestate prin dorința de a nu avea ca vecini 11 categorii de persoane stigmatizate social, cum ar fi dependenții de droguri, imigranții, romii, alcoolicii, foști condamnați etc. Datele utilizate au fost puse la dispoziție de Fundația Soroș România și fac parte din Barometrului de Opinie, Octombrie 2007. Setul inițial de 11 itemi a fost despărțit în două subseturi, folosind o formă de analiză factorială cu componente latente de tip continuu, care apoi au fost explorate utilizând analiza de clusteri prin clase latente. Pentru primul subset de variabile a fost identificat ca model optim un model cu patru clase latente, iar pentru al doilea subset modelul optim identificat a fost unul cu trei clase latente. În clusterelor intolerant și distant predomină femeile și persoanele din mediul rural. Totuși, femeile acceptă mai ușor vecini infectați cu HIV, iar persoanele din mediul rural sunt mai tolerante cu romii. În clusterelor tolerant și tolerant cu romii predomină bărbații. În urban, toleranța la vecinii infectați cu HIV este mult mai mare. Persoanele cu un nivel de educație ridicat sunt cu precădere în clusterul HIV tolerant, dar distanțe față de persoanele dependente de droguri sau alcool.

Cuvinte-cheie: clase latente, analiză factorială, distanță socială.

INTRODUCERE

Analiza claselor latente s-a desprins din analiza structurilor latente, metodă menționată pentru prima dată într-o vastă lucrare de psihologie socială, numită *American Soldier: Studies in Social Psychology in WW II*, apărută în patru volume între 1949 și 1950 (Henry, 1999). Structurile latente, descrise de Lazarsfeld în volumul 4, intitulat *Measurement and Prediction*, au la bază variabile latente a căror natură a fost conceptualizată de sociologi ca fiind una discretă, în clase. Din

Adresa de contact a autorului: Cornelia Rada, Institutul de Antropologie „Francisc I. Rainer”, Bd. Eroii Sanitari, Nr. 8, O.P. 35, C.P. 13, Sector 5, Cod 050474, București, România, e-mail: corneliarada@yahoo.com.

¹ Latent Gold este un program statistic specializat pentru analiza claselor latente produs de Statistical Innovations Inc. Versiunea 4.5 demo precum și documentația tehnică necesară utilizării programului este disponibilă pe site-ul <http://www.statisticalinnovations.com>.

potrivă, psihologii care operau cu noțiuni cum ar fi abilitatea, sensibilitatea sau inteligența au considerat latența ca având o natură continuă, sub forma unor trăsături sau caracteristici individuale numite *traits*.

Statisticienii au avut abordări ambivalente, formalizând matematic ambele tipuri de modele. Analiza claselor latente, care a cunoscut o dezvoltare spectaculoasă în ultimii 10–15 ani, aproape că a devenit un instrument standard de analiză în cercetarea sociologică, biomedicală, psihologie sau în cercetarea de piață. Pentru prima dată introdusă de Lazarsfeld și Henry în 1968 (Dayton, 1998) metoda a fost folosită inițial pentru a explica variabilele latente atitudinale în sondaje cu itemi dihotomici binari. Ulterior, a fost extinsă de Goodman în 1974 pentru variabile nominale (Goodman, 2002). În prezent, metoda are un domeniu de aplicare vast pentru variabile de toate tipurile și pentru combinații ale acestora. Formele de aplicare au fost de asemenea diversificate la analizele factoriale pentru variabile nominale, regresii în populații heterogene (mixturi), dependențe locale, lanțuri latente Markov etc.

Analiza claselor latente (*latent class analysis*) este o tehnică multivariată exploratorie de identificare a segmentelor, clusterilor sau, mai general spus, a claselor latente, care sunt categorii ale unei variabile latente discrete de tip categorial. Variabilele latente sunt variabile neobservate, teoretice, conceptuale, care ar putea explica gruparea în clase a modurilor de răspuns la un anumit stimul, a atitudinilor și percepțiilor. Ele structurează cazurile în raport cu un set de variabile observate, numite variabile manifest sau indicator. Astfel, *patternurile* de răspuns ale subiecților la un set de itemi sunt condiționate de apartenența subiecților la una sau alta din clasele unei variabile latente, de control, care nu este prezentă explicit în setul de itemi analizați.

Măsurătorile în științele sociale sunt date observate, însă conceptele pe care acestea le măsoară, cum sunt empatia, toleranța, abilitatea, nu sunt. Variabilele, sau factorii latenți și respectiv clasele latente sunt constructe teoretice, care pot căpăta un nume și un sens explicit după ce au fost identificate pe baza modelului analizat. Variabilele-manifest sunt măsurătorile de bază ale studiului empiric. Acceptarea ipotezei că variabilele latente există și că sunt într-o anumită relație cu variabilele-manifest permite cercetătorului să tranșeze datele și să aibă un discurs de teoretician și nu unul de statistician.

Din punct de vedere tehnic, analiza claselor latente are, în principal, trei domenii de aplicabilitate, care decurg din cele trei modele teoretice propuse: analiza de clusteri prin clase latente (*Latent Class Cluster Analysis*), analiza factorială discretă prin clase latente (*Latent Class Dfactor*) și analiza de regresie în clase latente (*Latent Class Regression Analysis*). Toate aceste modele sunt în mare măsură similare modelelor și analizelor tradiționale: analiza de clusteri, analiza factorială și respectiv, analiza de regresie. Avantajul modelelor în clase latente este că restricțiile din ipotezele analizelor tradiționale sunt mult relaxate. De exemplu, utilizarea în analiza statistică a variabilelor de tip nominal sau ordinal nu mai este o problemă, acest lucru fiind valabil chiar și pentru tipurile de analize, care, din punct de vedere

tehnic, sunt aplicabile numai în cazul variabilelor de tip interval, așa cum ar fi analiza factorială sau analiza de *clusteri*. Acesta face din analiza de clase latente, susținută de un program de calcul flexibil, un instrument de cercetare redutabil.

Scopul acestui articol este ca, pornind de la o prezentare tehnică a metodei, limitată la atât cât este necesar pentru înțelegerea elementelor de bază, și apoi continuând cu o exemplificare practică pe un set de date reale, să aducem un argument suplimentar în utilizarea claselor latente ca instrument de cercetare.

PRECIZĂRI CONCEPTUALE

Metoda de analiză de *clusteri* prin clase latente

Analiza de *clusteri* clasică organizează un set de itemi în grupuri (*clusteri*), în așa fel încât gradul de similaritate este maximizat în interiorul grupurilor și minimizat între grupuri. Grupurile se formează pe baza unor distanțe definite (Euclidiană, *Mahalanobis*, *city block* etc.) între itemi care trebuie să fie de tip interval.

Analiza de *clusteri* prin clase latente poate determina dacă asocierea dintre un set de variabile manifest de tip categorial $A, B, C, D \dots$ etc. poate fi explicată de o variabilă latentă neobservată, discretizată în segmente (clase, *clusteri* sau grupuri), notată convențional X , care are un număr de T categorii. Fără să afectez generalitatea modelului, pentru mai multă claritate, fac o prezentare a modelului de *clusteri* limitată pentru patru variabile-manifest (sau indicator). În forma sa probabilistică, ecuația modelului poate fi exprimată prin (Magidson și Vermunt, 2004):

$$\pi_{ijklt} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X}$$

unde

- π_t^X este probabilitatea de apartenență la clasa (*clusterul*) $t = 1, 2, \dots, T$ al variabilei X ;
- $\pi_{it}^{A|X}$ reprezintă probabilitatea condiționată de a răspunde în categoria $i = 1, 2, \dots, I$ la itemul A pentru clasa t ;
- $\pi_{jt}^{B|X}, \pi_{kt}^{C|X}, \pi_{lt}^{D|X}$ sunt probabilitățile corespunzătoare pentru itemii B, C și D , iar $j = 1, 2, \dots, J; k = 1, 2, \dots, K$ și $l = 1, 2, \dots, L$;
- π_{ijklt} este probabilitatea obținerii unui răspuns cu profilul $\{i, j, k, l\}$ la cei patru itemi, în *clusterul* t .

Modelul are ca ipoteză independența locală a variabilelor-manifest A, B, C și D în fiecare *cluster*. Aceasta înseamnă că nu există relații de asociere între oricare dintre cele patru variabile în fiecare din cei T *clusteri*. Din punct de vedere statistic, independența locală presupune ca valorile testului Pearson χ^2 (Pearson chi pătrat) împărțite la numărul gradelor de libertate, cunoscute ca reziduale bivariate, să nu

difere semnificativ de valoarea 1. Valorile Pearson χ^2 se calculează din tabelele de contingență bivariate, pe baza frecvențelor observate și estimate corespunzător fiecărui model estimat.

Pentru a face modelul identificabil se impun câteva restricții logice. În cazul prezentat mai sus, cu patru variabile manifest și o variabilă latentă, acestea se exprimă prin următoarele ecuații:

$$\sum_t \pi_t^X = \sum_i \pi_{it}^{A|X} = \sum_j \pi_{jt}^{B|X} = \sum_k \pi_{kt}^{C|X} = \sum_l \pi_{lt}^{D|X} = 1.0$$

Cu ajutorul teoremei *Bayes*, se poate determina apoi probabilitatea posteroară de apartenență la un anumit *cluster*, cunoscând *patternul* de răspuns (i, j, k, l):

$$\pi_{ijkl}^{X|ABCD} = \frac{\pi_{ijkl}^{ABCDX}}{\sum_t \pi_{ijkl}^{ABCDX}}, \quad t = 1, 2, \dots, T$$

În *Latent Gold*, fiecare caz este alocat *clusterului* pentru care probabilitatea este maximă.

O altă reprezentare cunoscută a modelului este forma sa log-liniară de parametrizare, în care notăm cu f_{ijkl} frecvența corespunzătoare într-un tabel de contingență în care am introdus toate cele cinci variabile A, B, C, D și X :

$$\begin{aligned} \ln(f_{ijkl}) = & \lambda + \lambda_t^X + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{it}^{A|X} + \lambda_{jt}^{B|X} + \lambda_{kt}^{C|X} + \lambda_{lt}^{D|X} \\ i = & 1, 2, \dots, I; j = 1, 2, \dots, J; k = 1, 2, \dots, K; l = 1, 2, \dots, L \text{ și } t = 1, 2, \dots, T. \end{aligned}$$

Modelul presupune, în acest caz, cinci efecte principale $\lambda_t^X, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_l^D$, patru interacțiuni între variabila latentă și variabilele manifest $\lambda_{it}^{A|X}, \lambda_{jt}^{B|X}, \lambda_{kt}^{C|X}, \lambda_{lt}^{D|X}$ și o constantă λ . Ipoteza de independență locală mutuală dintre variabilele-manifest A, B, C și D este asumată prin absența efectelor de interacțiune dintre variabile. Pentru identificarea parametrilor este necesar, fără a pierde din generalitatea soluției, un sistem de restricționare a parametrilor, cum ar fi „dummy coding” sau „effect coding”. Ca exemplu, în „dummy coding” se impun următoarele restricții:

$$\begin{aligned} \lambda_{1t}^{A|X} = \lambda_{1t}^{B|X} = \lambda_{1t}^{C|X} = \lambda_{1t}^{D|X} = 0; & \text{ pentru } t = 2, 3, \dots, T; \\ \lambda_1^X = \lambda_1^A = \lambda_1^B = \lambda_1^C = \lambda_1^D = 0; & \\ \lambda_{i1}^{A|X} = \lambda_{j1}^{B|X} = \lambda_{k1}^{C|X} = \lambda_{l1}^{D|X} = 0 & \text{ pentru } i = 1, 2, \dots, I; j = 1, 2, \dots, J; k = 1, 2, \dots, K; \\ & l = 1, 2, \dots, L. \end{aligned}$$

În „effect coding”, restricționarea se referă la suma efectelor, care este constrânsă să fie zero. McCutcheon abordează pe larg complementaritatea celor două forme de

exprimare a modelului de clase latente, avantajele și dezavantajele acestora (McCutcheon, 2002). Probabilitățile condiționate din ecuația de parametrizare probabilistică pot fi calculate cu ajutorul parametrilor din forma log-lineară prin următoarea formulă:

$$\pi_{it}^{A|X} = \exp(\lambda_i^A + \lambda_{it}^{A|X}) / \sum_i \exp(\lambda_i^A + \lambda_{it}^{A|X})$$

Prevalența în fiecare clasă latentă sau mărimea *clusterului* se determină printr-o formulă similară:

$$\pi_i^X = \exp(\lambda_i^X) / \sum_i \exp(\lambda_i^X)$$

Pentru modelul de referință (H_0) cu o singură clasă $T = 1$, un model în care *patternul* de răspuns este independent de apartenența la *cluster*, forma ecuației se rezumă, evident, la:

$$\ln(f_{ijkl}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D$$

Numărul de parametri distincți ai modelului, care corespund situației în care $T = 1$, se calculează astfel:

$$NPAR(\text{indep}) = (I - 1) + (J - 1) + (K - 1) + (L - 1).$$

În cazul general în care avem T clase, numărul de parametri este calculat cu formula:

$$NPAR(T) = (T - 1) + NPAR(\text{indep}) \times [1 + (T - 1)],$$

iar numărul gradelor de libertate corespunzătoare modelului testat, $DF(T)$ este:

$$DF(T) = IJKL - NPAR(T) - 1$$

Analiza în clase latente este, inițial, un proces de identificare a celui mai „potrivit” model (*the best fit*). Aceasta înseamnă că trebuie, începând cu modelul de referință H_0 , să estimăm, incrementând T cu 1, câteva modele care conțin $T = 1$, $T = 2$, $T = 3$, sau mai multe clase latente, până când modelul satisface un criteriu ales de acceptare. Un astfel de criteriu poate fi statistica L^2 (*likelihood ratio chi-squared statistic*), calculată pe baza frecvențelor observate și a celor estimate din tabela de contingență multivariată:

$$L^2 = 2 \sum_{ijkl} F_{ijkl} \ln(F_{ijkl} / f_{ijkl})$$

Unde F_{ijkl} este frecvența observată, iar f_{ijkl} reprezintă frecvența așteptată (calculată).

La un model perfect ($F_{ijkl} = f_{ijkl}$) ar trebui ca L^2 să fie egal cu 0. Măsura în care L^2 depășește valoarea 0 indică cât de mult modelul nu se potrivește cu datele observate, adică în ce măsură asocierea dintre variabile nu este explicată de

modelul de segmentare în clase latente. Statistica L^2 (uneori notat cu G^2), atunci când volumul de date N este suficient de mare, are o distribuție asimptotică χ^2 și astfel poate fi determinată probabilitatea p de acceptare a ipotezei alternative. Regula practică utilizată pentru acceptare este ca L^2 să nu fie substanțial mai mare decât numărul de grade de libertate (DF), ceea ce corespunde la un $p > 0,05$.

Ajustarea modelului la datele analizate, adică estimarea numărului de clase latente, se poate face și pe baza unor criterii de ajustare similare cu cele utilizate în metodele de analiză clasice, cum sunt: *Bayesian Information Criterion* (BIC), *Akaike Information Criterion* (AIC), *Akaike Information Criterion 3* (AIC3), și *Consistent Akaike Information Criterion* (CAIC). Aceste criterii penalizează valorile L^2 , ținând cont de numărul de parametri ai modelului, sau de volumul de date N din eșantion.

De exemplu, BIC se calculează astfel:

$$\text{BIC} = L^2 - DF * \ln(N)$$

Cea mai mică valoare a criteriului, obținută la fiecare model analizat, indică modelul care se potrivește cel mai bine la datele analizate (*the best fit*). Totuși, se ține cont și de faptul că modelul trebuie să fie cât mai simplu, adică cu cât mai puțini parametri estimați pentru a putea fi explicat și înțeles practic. Din acest motiv, uneori, în practică se pot accepta și modele pentru care $p < 0,05$.

Alte statistici utilizate pentru evaluarea modelelor în clase latente similare statisticii L^2 sunt statistica *Pearson* χ^2 , și statistica chi-pătrat *Cressie-Read* CR^2 . Indexul de disimilaritate (*Dissimilarity Index*) (valoarea de referință este $I_D < 0,05$) indică în ce măsură sunt diferențe importante între frecvențele observate și cele estimate ale modelului. Indexul de disimilaritate exprimă proporția din eșantion care ar trebui redistribuită în alte celule pentru a obține un model perfect.

O altă serie de indicatori statistici, cum ar fi eroarea de clasificare (*classification error*), care trebuie să fie apropiată de valoarea 0, sunt folosiți la estimarea calității modelului, adică cât de bine modelul clasifică cazurile în *clusteri* (clase). Alte trei statistici similare sunt reducerea erorilor (*lambda*), *entropia* R^2 și R^2 *standard*, care trebuie să aibă valori apropiate de 1.

În unele programe de analiză a claselor latente, așa cum este *Latent Gold*, pot fi folosite variabilele covariate (*covariates*). Acestea sunt variabile exogene, de regulă, variabile sociodemografice, introduse suplimentar în analiza de clase latente pentru a prezice și explica formarea claselor. De obicei, parametrii modelelor sunt cunoscuți ca *beta* sau *gama*. *Beta* corespund efectelor variabilelor-indicator (manifest) asupra fiecărui cluster. *Gama* corespund efectelor variabilelor covariate asupra variabilei latente. *Beta* și *gama* modelelor analizate trebuie evaluați pentru gradul de semnificație, de regulă prin teste *Wald* sau prin valoarea erorilor standard. Valori ale erorilor standard mai mari ca 2 sau un $p < 0,05$ pentru testul *Wald*, indică faptul că parametrul respectiv este semnificativ diferit de 0. Aceasta poate fi un indiciu util de a elimina, eventual, efectul total sau parțial al unei variabile din model în faza confirmatorie.

Alegerea modelului optim în analiza de clase latente trebuie să se bazeze pe un echilibru între criteriile statistice menționate mai sus și necesitatea de a putea să interpretăm modelul obținut. Modelul de clase latente încearcă să explice asocierea dintre variabile în termeni de independență, în condițiile existenței unei variabile latente de control. Când în realitate două variabilele sunt controlate de o a treia, pe care atunci când nu o putem observa direct, o numim latentă, asocierea marginală dintre variabile poate dispărea sau poate avea o direcție diferită (Agresti, 1990).

Analiza factorială discretă prin clase latente (*Latent Class Dfactor*)

Analiza de *clusteri* prin clase latente presupune existența unei singure variabile latente. Dacă se dorește o analiză în mai multe variabile latente trebuie folosită analiza factorială latentă în factori discreți. *Latent Gold* are posibilitatea, în modulul de bază, să facă analiza factorială în factori discreți (*Dfactor*). Analiza, fie de sine stătător, fie în completare la analiza de *clusteri*, permite o mai clară perspectivă asupra modului de formare și determinare a claselor latente. Poate fi o metodă alternativă la analiza factorială tradițională, atunci când există motive să presupunem că natura factorilor este, mai degrabă, una discretă decât continuă.

Factorii în *Dfactor* au rolul unor variabile latente ordinale, care în loc de clase latente au niveluri latente (până la cinci niveluri în fiecare factor). În principiu, analiza factorială în clase latente este un tip de analiză factorială clasică, cu aplicabilitate în variabile discrete (nominale, ordinale, binare etc.), spre deosebire de cea clasică, ce admite numai variabile de tip *interval* (scală).

Analiza de regresie în clase latente (*Latent Class Regression Analysis*)

Analiza de regresie în clase latente permite ca în eșantioane heterogene să identificăm mixturi de modele de regresie, specifice la nivel de *cluster* (segment). În *Latent Gold* avem posibilitatea să introducem și unele efecte aleatoare (*random effect*), cu distribuție normală atât la nivelul termenului liber cât și la nivelul coeficienților din regresie.

Latent Gold folosește pentru determinarea parametrilor o combinație de doi algoritmi, *expectation-maximization* (EM) și *Newton-Raphson* (NR).

Distanța socială

Scala distanței sociale a fost elaborată, pentru prima dată, de Emory S. Bogardus și evidențiază atitudinea de acceptare sau izolare a unor grupuri etnice (Bogardus, 1925).

Distanța socială se definește ca „diferență percepută și evaluată între persoane sau grupuri, prin raportare la un criteriu (o caracteristică a personalității sau a grupului, poziția în ierarhie, un mod de comportare etc.)” (Vlăsceanu, 1998).

Traian Herseni formulează ipostazele primare ale distanței sociale, prin conceptele de distanță spațială, vecinătate, stratificare, și creionează ipostazele

secundare ale distanței sociale, identificând existența unei distanțe psihosociale (magistru–discipol, părinte–copil) (Herseni, 1931; Stăvărache, 2002).

Septimiu Chelcea adaptează scala pentru prima dată în România și calculează *indicele distanței de contact social și indicele calității contactelor sociale*. Cu acest prilej, se constată că românii nu manifestă atitudini xenofobe, totuși, față de etnia romilor; atitudinea a fost una slab pozitivă. De asemenea, s-a constatat că femeile și persoanele din mediul rural au atitudini etnice mai puțin pozitive (Chelcea, 1995).

O altă contribuție importantă privind distanța socială o găsim în lucrarea intitulată „Interethnic Relations in Romania” (Abraham și alții, 1995).

Conceptul se poate interpreta și în legătură cu cel de discriminare, care presupune o tratare inegală a indivizilor sau grupurilor, în raport cu unele trăsături, cum ar fi apartenența etnică, rasială, religioasă sau de clasă. Generalizarea unei experiențe negative la un întreg grup, stereotipurile, prejudecățile, absența reală a cunoașterii poate conduce la distanțe sociale nejustificate și chiar la discriminare. Chiar dacă, în toate societățile democratice, din punct de vedere legal, este interzisă discriminarea legată de sex, rasă, apartenență, etnică și religioasă, totuși acțiunea socială încalcă uneori prevederile constituționale (Mihăilescu, 1998).

Dificultatea în evaluarea distanței sociale provine, în principal, din aceea că este o problemă legată de atitudine. Când acționăm, gândim, simțim relativ stabil în raport cu o persoană spunem că avem o atitudine față de aceasta. Așadar, atitudinea are trei dimensiuni: comportamentală, cognitivă și evaluativă. Uneori, chiar dacă nu am fost în contact cu un grup, simțim nevoia de o distanță față de acel grup, doar în baza unor stereotipuri.

DATE ȘI METODĂ

În cadrul Barometrului de Opinie al Fundației Soroș, România, din Octombrie 2007, subiecții au fost chestionați în legătură cu 11 tipuri de vecini (http://www.osf.ro/ro/program_articol.php?articol=107). Respondenții trebuiau să menționeze categoria pe care nu ar fi de dorit să o aibă ca vecină. Pentru fiecare categorie, subiecții au dat răspunsuri codificate binar, indicând dacă respectiva categorie de vecini este acceptabilă (0 = Nemenționat) sau inacceptabilă (1 = Menționat). Categoriile de persoane propuse în chestionar, precum și proporțiile în care acestea au fost sau nu au fost menționate sunt prezentate în *Tabelul nr. 1*.

Primele trei categorii de persoane nedorite ca vecini au fost: persoanele dependente de droguri, alcoolicii și homosexualii. Cele mai acceptabile trei categorii de persoane ca vecini au fost: persoanele având religie diferită, cuplurile necăsătorite care trăiesc împreună, oamenii de altă etnie.

Îmi propun să identific tipologiile, *patternurile* probabile de respingere socială, manifestate prin dorința declarată de a nu avea ca vecini anumite categorii de persoane, care în sistemul de valori autohton sunt stigmatizate. Este o analiză de tip cantitativ și calitativ.

Tabelul nr. 1

Proporțiile grupurilor de persoane nedorite ca vecini, BOP, Octombrie, 2007

Categoria de vecin nedorit	Variabila manifest	Nemenționat %	Menționat %
Persoane dependente de droguri.	Dependenți	24,7	75,3
Persoane de rasă diferită de a dvs.	Rasă	81,9	18,1
Persoane care au SIDA.	Seropozitivi	56,2	43,8
Imigranți.	Imigranți	81,5	18,5
Homosexuali.	Homosexuali	36,4	63,6
Persoane având o religie diferită de a dvs.	Religie	87,6	12,4
Alcoolici.	Alcoolici	29,6	70,4
Cupluri necăsătorite care trăiesc împreună.	Necăsătoriți	87,5	12,5
Oameni de altă etnie.	Etnie	86,9	13,1
Romi, țigani.	Romi	47,5	52,5
Persoane care au suferit condamnări penale.	Condamnați	45,0	55,0

Deoarece eșantionul a cuprins $N = 2\ 000$ de subiecți și a avut o rată a non-răspunsurilor de 4,8%, cele 11 categorii de vecini nu au putut să fie analizate simultan. Folosirea simultan a celor 11 variabile binare ar fi presupus analizarea unei table multidimensionale de $2^{11} = 2\ 048$ celule, care în mod evident ar fi fost subpopulată. Din acest motiv am separat cele 11 variabile-indicator în două subseturi, de șase variabile și respectiv cinci variabile. Criteriile care au stat la baza separării variabilelor în subseturi sunt detaliate în cele ce urmează.

Analiza factorială preliminară în clase latente

Pentru determinarea celor două subseturi de variabile am utilizat una dintre posibilitățile speciale de modelare din *Latent Gold*, și anume, aceea de a face analize asemănătoare analizei factoriale tradiționale. Cunoscută și sub denumirea de analiză de caracteristici latente (*Latent Traits Analysis*) analiza este utilă în situația în care variabilele-indicator sunt de tip *categorie* (așa cum este cazul acum) și variabilele latente sunt presupuse de tip *continuu*. În *Latent Gold*, variabilele latente continue sunt notate implicit cu CFactor și pot fi maxim trei. Toate cele 11 variabile manifest au fost introduse formal într-un model de *clusteri* cu un singur *cluster*, în care am introdus unul până la trei factori de tip *continuu* pentru explorare. Modelul cel mai performant (*the best fit*) a fost identificat pe baza criteriilor informaționale BIC și CAIC care au realizat valori minime pentru un model care include doi factori de tip *continuu*. Un astfel de model reprezintă o extensie bifactorială a modelului *Rasch*².

² Modelele *Rasch*, folosite mai frecvent în psihometrie, sunt modele speciale pentru analizarea datelor de măsurare a abilităților, a atitudinilor sau a caracteristicilor de personalitate. Teoria matematică care stă la baza acestui model este, în multe privințe, aceeași din teoria răspunsurilor la itemi (IRT).

Ținând cont de faptul că variabilele-manifest sunt binare, ecuația modelului pentru o singură clasă și doi factori, este de forma:

$$\frac{P(Y_t = 1 | F_{1i}, F_{2i})}{P(Y_t = 0 | F_{1i}, F_{2i})} = (\beta_{01}^t - \beta_{00}^t) + \lambda_1^t F_{1i} + \lambda_2^t F_{2i}$$

unde F_{1i} și F_{2i} reprezintă scorurile pe fiecare factor, care sunt independente mutual și au distribuții normale $N(0,1)$.

Termenul liber din ecuație, $-(\beta_{01}^t - \beta_{00}^t)$ ar putea fi interpretat, din perspectiva teoriei răspunsului la itemi (*Item Response Theory*), ca reprezentând „dificultatea întrebării” (Vermunt și Magidson, 2006). Pentru detalii privind utilizarea analizelor factoriale cu factori de tip *continuu* (*CFactor*) sau cu factori discreți (*Dfactor*) precum și o comparație cu analiza factorială clasică poate fi consultat (Vermunt și Magidson, 2005).

Analiza interacțiunilor dintre factori și cele 11 variabile-manifest indică o încărcare diferențiată pe cei doi factori. Din cei 33 de parametri ai modelului, 31 sunt semnificativ diferiți de 0, conform testelor *Wald*, la un prag de semnificație de 0,05. Modelul se îmbunătățește puțin din punctul de vedere al criteriilor BIC, AIC și CAIC, dacă constrângem parametrii λ_2^t , corespunzători celor două variabile *Dependenți* și *Alcoolici*, pentru care parametrii au $p > 0,05$ să fie egali cu 0 (*Tabelul nr. 2*).

Tabelul nr. 2

CFactor Loadings – Saturații

<i>Loadings</i>	<i>CFactor 1</i>	<i>CFactor 2</i>	<i>Loadings</i>
Dependenți	0,57	0,00	Dependenți
Rasa	0,33	0,52	Rasa
Seropozitivi	0,52	0,30	Seropozitivi
Imigranți	0,34	0,46	Imigranți
Homosexuali	0,58	0,15	Homosexuali
Religie	0,14	0,52	Religie
Alcoolici	0,49	0,00	Alcoolici
Necăsătoriți	0,14	0,46	Necăsătoriți
Etnie	0,21	0,54	Etnie
Romi	0,40	0,23	Romi
Condamnați	0,52	0,12	Condamnați

Astfel, șase variabile, *Dependenți*, *Seropozitivi*, *Homosexuali*, *Alcoolici*, *Romi* și *Condamnați*, au saturații mai mare sau egală cu 0,400 pe *CFactor 1* iar celelalte cinci variabile: *Rasa*, *Imigranți*, *Religie*, *Necăsătoriți* și *Etnie* corelează mai puternic cu *CFactor 2*. Acest model statistic sugerează faptul că distanța socială poate fi explicată prin două componente care separă cele două subseturi de variabile, după cum urmează.

Prima componentă, care include subșetul de variabile observate: Dependenți, Seropozitivi, Homosexuali, Alcoolici, Romi și Condamnați, identifică diferențierea prin raportul individului cu valori sociale cum ar fi *legea* și *ordinea* în comunitate. Reprezintă o distanțare față de persoanele care au suferit condamnări penale, față de romi, care în general sunt percepuți ca având probleme cu legea și ordinea, față de homosexualitate, care cândva era condamnată de lege. De asemenea, consumul de droguri, care este adesea asociat cu infectarea HIV și homosexualitatea, este în opoziție cu legea. Dependenții de alcool sunt și ei percepuți ca aducători de dezordine și scandal în societate.

A doua componentă, care include subșetul de variabile observate: Rasa, Imigranți, Religie, Necăsătoriți și Etnie reprezintă distanțarea pentru conservarea intimității culturale și confesionale. Persoanele necăsătorite care trăiesc împreună sunt asociate acestei componente, deoarece sunt în opoziție cu valoarea socială și morală a căsătoriei ca act legal și religios.

Analiza de clusteri în clase latente

Pentru primul subșet, am selectat aceleași șase variabile ca la analiza precedentă, și anume: Dependenți (Persoane dependente de droguri), Seropozitivi (Persoane care au SIDA), Homosexuali, Alcoolici, Romi și Condamnați (Persoane care au suferit condamnări penale).

În primă fază de explorare am cerut, în *Latent Gold*, să fie evaluate toate modelele de la un cluster până la șapte clusteri. Performanțele modelelor analizate, criteriile de evaluare sunt prezentate în Tabelul nr. 3. În conformitate cu criteriul L^2 , modelul cu cinci clusteri reprezintă modelul optim (*the best fit*), deoarece $p = 0,37 > 0,05$ și putem observa că valoarea lui L^2 este apropiată de numărul gradelor de libertate ($L^2 = 30,80$ iar $df = 29$). Valorile criteriilor informaționale BIC (13393,45) și CAIC (13420,45) sunt minime pentru modelul cu patru clusteri. AIC (13226,30) este minim pentru modelul cu cinci clusteri. Eroarea de clasificare pentru modelul cu patru clusteri este 0,1953. Din motive practice vom face un compromis și vom alege modelul cu patru clusteri, deși $p = 0,0045 < 0,05$ la un $L^2 = 62,05$ și $df = 36$.

Tabelul nr. 3

Criterii de evaluare a modelului optim (*the best fit*)

Modelul	BIC(LL)	AIC(LL)	CAIC(LL)	Npar	L^2	df	p-value	Eroarea de clasificare
1-Cluster	14850,81	14817,50	14856,81	6	1678,011	57	0,000	0
2-Cluster	13551,79	13479,62	13564,79	13	326,1238	50	0,000	0,0982
3-Cluster	13405,93	13294,89	13425,93	20	127,4017	43	0,000	0,1566
4-Cluster	13393,45	13243,55	13420,45	27	62,0567	36	0,0045	0,1953
5-Cluster	13415,06	13226,30	13449,06	34	30,8059	29	0,37	0,2109
6-Cluster	13460,77	13233,14	13501,77	41	23,6471	22	0,37	0,1947
7-Cluster	13503,94	13237,45	13551,94	48	13,96	15	0,53	0,2548
4 Cluster + Efect direct	13386,05	13230,6	13414,05	28	47,1081	35	0,083	0,1833

În următoarea etapă vom analiza valorile rezidualelor bivariate (*Tabelul nr. 4*) la modelul cu patru *clusteri*, pentru a verifica dacă ipoteza independenței locale nu a fost violată. Observăm că rezidualele indică o asociere apreciabilă (>2) între variabilele Dependenți și Romi (6,2198) și respectiv între Dependenți și Condamnați (2,4501). Din acest motiv vom adopta o strategie recomandată pentru *Latent Gold*, care constă în introducerea în model a efectului direct de interacțiune între variabilele Dependenți și Romi.

Tabelul nr. 4

Valorile rezidualelor bivariate pentru modelul cu patru clusteri

Indicatori	Dependenți	Seropozitivi	Homosexuali	Alcoolici	Romi
Dependenți
Seropozitivi	0,1002
Homosexuali	0,0087	0,0308	.	.	.
Alcoolici	0,4283	0,0403	0,0948	.	.
Romi	6,2198	0,3533	0,2082	0,0159	.
Condamnați	2,4501	0,0012	1,1071	0,5124	1,5052

În noul model obținut (patru Cluster + Efect direct) $L^2 = 47,10$ și $p = 0,083 > 0,05$, iar valorile BIC, AIC și CAIC sunt mai mici decât la modelul cu patru Clusteri fără efecte directe ceea ce indică o ameliorare a modelului. De asemenea se observă scăderea erorii de clasificare la 0,1833. Indexul de disimilaritate este $0,0448 < 0,05$. Reanalizarea rezidualelor bivariate nu mai indică violări ale ipotezei de independență locală, toate valorile fiind apropiate sau sub 1. *Tabelele nr. 3 și 4* prezintă parametrii modelului. Analiza parametrilor modelului indică faptul că toți coeficienții de interacțiune sunt statistic semnificativ diferiți de 0 ($p < 0,05$), conform testului *Wald* (*Tabelul nr. 5*). Deoarece am folosit o codificare de tip *efect*, suma pe rânduri a coeficienților din *Tabelul nr. 4* este 0.

Tabelul nr. 5

Parametrii modelului – interacțiunile dintre cluster și variabile

Variabile	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Wald	p-value
Dependenți	-0,30	3,23	0,34	-3,26	123,92	0,000
Seropozitivi	-2,05	1,79	2,58	-2,31	95,30	0,000
Homosexuali	-0,40	2,61	0,74	-2,94	131,72	0,000
Alcoolici	0,62	2,50	-0,47	-2,64	88,95	0,000
Romi	0,03	2,18	-0,25	-1,96	110,26	0,000
Condamnați	0,10	2,30	0,20	-2,59	127,33	0,000

Coefficienții efectelor principale prezentați în *Tabelul nr. 6* sunt numai în parte diferiți semnificativ de 0, sugerând modificarea modelului astfel încât acești coeficienți, pentru care $p > 0,05$, să fie constrânși la valoarea 0. Suma coeficienților pe verticală (pentru toate categoriile fiecărei variabile-indicator) este, de asemenea, 0.

Tabelul nr. 6

Parametrii modelului – efectele principale

Variabile și categorii	Intercepts	Wald	p-value
Dependenți			
Nemenționat	-0,826	44,884	0,000
Menționat	0,826		
Seropozitivi			
Nemenționat	0,210	0,529	0,47
Menționat	-0,210		
Homosexuali			
Nemenționat	-0,231	7,247	0,0071
Menționat	0,231		
Alcoolici			
Nemenționat	-0,310	9,041	0,0026
Menționat	0,310		
Romi			
Nemenționat	-0,135	2,426	0,12
Menționat	0,135		
Condamnați			
Nemenționat	0,113	2,278	0,13
Menționat	-0,113		

În Tabelul nr. 7 sunt prezentate profilele în *cluster*. Pe primul rând avem dimensiunea fiecărui *cluster*. De exemplu, *clusterul* 1 cuprinde 37,7% din eșantion iar *clusterul* 4 este cel mai puțin populat, cu 14,0% din eșantion. Pentru fiecare variabilă și fiecare cluster, în tabel sunt indicate probabilitățile ca subiectul să răspundă cu menționat sau nemenționat. Suma probabilităților pe verticală pentru fiecare variabilă-indicator este egală cu 1. Pentru o mai ușoară și rapidă înțelegere, în *Latent Gold* profilele sunt reprezentate și grafic, alături de prezentarea tabelară.

Se identifică, astfel, două profile total opuse: *clusterul* 2, care include 32,4% din subiecți și *clusterul* 4, care cuprinde 14% din eșantion. *Clusterul* 2 poate fi clasificat ca *intolerant*, deoarece probabilitățile de a menționa ca indezirabili pentru toate categoriile de vecini menționați în chestionar sunt mari, între 0,797 (seropozitivi) și 0,986 (dependenți). *Clusterul* 4 poate fi definit *tolerant*, probabilitatea de menționare ca vecin indezirabil aici fiind scăzută: între 0,056 (condamnați) și 0,156 (dependenți).

*Cluster*ele 1 și 3 sunt mai nuanțate în privința vecinilor. În *clusterul* 1 există o toleranță specială pentru seropozitivi, care sunt menționați ca nedoriți cu o probabilitate de numai 0,078. Este un *cluster tolerant HIV*, dar care respinge ca vecini persoanele dependente de droguri (0,734) sau alcool (0,775). De asemenea, *clusterul* 1 este destul de permisiv cu vecinii romi sau care au suferit condamnări. *Clusterul* 3, care include 15,8% din subiecți este mai tolerant cu romii, probabilitatea de respingere a acestora ca vecini fiind de 0,365. Ei sunt, de asemenea, relativ toleranți cu cei care au suferit condamnări, probabilitatea fiind de 0,494). Este un cluster *tolerant cu romii*, dar care se distanțează destul de categoric de dependenții de droguri, de seropozitivi și de homosexuali.

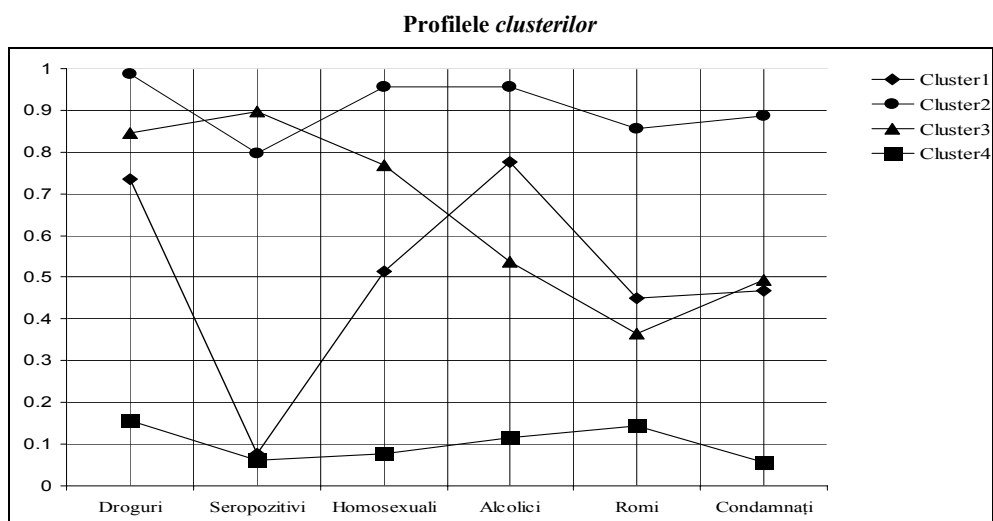
Tabelul nr. 7

Profilul clusterilor – variabile-manifest

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Mărime clusteri	0,377	0,324	0,158	0,140
Dependenți				
Nemenționat	0,266	0,014	0,154	0,844
Menționat	0,734	0,986	0,846	0,156
Seropozitivi				
Nemenționat	0,922	0,203	0,104	0,939
Menționat	0,078	0,797	0,896	0,061
Homosexuali				
Nemenționat	0,486	0,044	0,232	0,923
Menționat	0,514	0,956	0,768	0,077
Alcoolici				
Nemenționat	0,225	0,043	0,464	0,883
Menționat	0,775	0,958	0,536	0,117
Romi				
Nemenționat	0,550	0,144	0,635	0,855
Menționat	0,450	0,856	0,365	0,145
Condamnați				
Nemenționat	0,533	0,112	0,506	0,944
Menționat	0,467	0,888	0,494	0,056

Diferențele de profile dintre cele patru clusterare sunt prezentate grafic în *Figura 1*. Pe axa verticală a graficului, în acest caz, sunt reprezentate probabilitățile de a menționa ca nedorită o persoană din categoria reprezentată pe axa orizontală.

Figura 1



Introducerea în model a covariatelor (*Tabelul nr. 8*) genul, educația (nivelul de instruire) și mediul (mediul de rezidență) permite explicații privind structura sociodemografică din clusteri.

Tabelul nr. 8

Distribuția marginală a variabilelor covariate în eșantion

Variabila covariate	Categoria	Proporția %
Genul – Sexul respondentului	Masculin	49,7
	Feminin	50,3
Educația – Nivel de instruire	Scăzut	46,7
	Mediu	40,7
	Ridicat	12,6
Mediul – Mediul de rezidență	Urban	57,3
	Rural	42,7

Se poate observa, din *Tabelul nr. 9*, că femeile sunt preponderente în *clusterule* 1 și 2, unde probabilitatea ca sexul subiectului să fie feminin este de 0,556 și respectiv de 0,582, comparativ cu proporția generală din eșantion, de 50,3%. Femeile sunt preponderente în *clusterul* 2 (0,582), deci pot fi considerate mai distante față de categoriile propuse ca vecini. Totuși preponderența acestora în *clusterul* 1 ne arată că acceptă mai ușor vecini infectați cu HIV. Bărbații sunt preponderenți în *clusterule* 3 și 4 și putem trage concluzia că, în general, bărbații sunt mai toleranți.

Tabelul nr. 9

Profilul clusterilor – covariate

Covariate	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	HIV tolerant	Intolerant	Etnic tolerant	Tolerant
Genul				
Masculin	0,444	0,418	0,683	0,581
Feminin	0,556	0,582	0,317	0,419
Educația				
Scăzut	0,282	0,529	0,657	0,481
Mediu	0,470	0,405	0,309	0,430
Ridicat	0,248	0,066	0,034	0,089
Mediul				
Urban	0,739	0,477	0,433	0,547
Rural	0,261	0,523	0,567	0,453

Toleranța la vecinii infectați cu HIV este mult mai mare în urban (0,739). În mediul rural, populația este în general mai intolerantă (0,523, în *clusterul* 2), dar este mai tolerantă cu romii (0,567, în *clusterul* 3), comparativ cu mediul urban, ținând cont că, la nivelul eșantionului, mediul rural are o pondere de 42,7%.

Nivelul de educație nu explică cum se formează *clusterii* decât în măsura în care persoanele cu un nivel de educație ridicat au o probabilitate mai mare de

apartenență la clusterul *HIV tolerant*. Acest cluster este, totuși, în același timp, un cluster distant față de persoanele dependente de droguri sau alcool.

Pentru al doilea subset s-au selectat cinci variabile: Rasa (persoane de rasă diferită), Imigranți (imigranți sau muncitori din altă țară), Religie (persoane având o religie diferită), Necăsătoriți (cupluri necăsătorite care trăiesc împreună) și Etnie (oameni de altă etnie).

Analiza exploratorie a modelelor de unul până la cinci clusteri indică faptul că un minim pentru criteriile BIC și CAIC se obține pentru modelul cu trei clusteri (Tabelul nr. 10), la care corespunde o statistică $L^2 = 40,8094$ și respectiv $p = 0,0002 < 0,05$. În conformitate cu criteriul AIC și L^2 , cel mai bun model este cel cu patru clusteri ($L^2 = 10,894$, $df = 8$ și $p = 0,21 > 0,05$).

Tabelul nr. 10

Criterii de evaluare a modelului optim (*the best fit*)

Modelul	BIC(LL)	AIC(LL)	CAIC(LL)	Npar	L^2	df	p-value	Eroarea de clasificare
1-Cluster	7999,97	7972,21	8004,97	5	1861,345	26	0,0000	0,0000
2-Cluster	6375,04	6313,97	6386,04	11	191,101	20	0,0000	0,0321
3-Cluster	6270,06	6175,67	6287,06	17	40,809	14	0,0002	0,1009
4-Cluster	6285,45	6157,76	6308,45	23	10,894	8	0,2100	0,0877
5-Cluster	6322,45	6161,44	6351,45	29	2,578	2	0,2800	0,1211
3-Cluster + Efect Direct	6256,02	6156,09	6274,02	18	19,2308	13	0,1200	0,1099

Spre a avea mai multă simplitate, am optat pentru modelul cu trei *clusteri*, în care am introdus pentru respectarea ipotezei de independență locală un efect direct al interacțiunii dintre variabilele Rasa și Imigranți. Performanțele noului model obținut (3-Cluster + Efect Direct) sunt mai bune decât ale modelului cu numai trei clusteri: $L^2 = 19,2308$ și $p = 0,12 > 0,05$, iar valorile BIC, AIC și CAIC sunt mai mici. *Dissimilarity Index* este 0,0153, destul de apropiat de 0.

În Tabelul nr. 11 sunt prezentate profilele în *cluster*. Pe primul rând sunt indicate dimensiunea fiecărui *cluster*. De exemplu, *clusterul* 1 cuprinde 70,98% din eșantion, *clusterul* 2 cuprinde 22,24%, iar *clusterul* 3 este cel mai puțin populat, cu 6,78% din eșantion. Pentru fiecare variabilă în tabel sunt indicate probabilitățile ca subiectul să răspundă la o categorie sau alta, respectiv 0 = Nemenționat și 1 = Menționat. Suma probabilităților pe verticală pentru fiecare variabilă-indicator este egală cu 1.

Clusterul 1 este caracterizat de un grad de apropiere socială ridicat. Probabilitățile de a menționa ca vecin nedorit o persoană din cele cinci categorii sunt scăzute, între 0,0005, pentru Etnie și 0,0618, pentru Imigranți.

La polul opus *clusterul* 3 este caracterizat de un grad ridicat de distanțare socială. Probabilitățile de a menționa ca vecin nedorit o persoană din cele cinci categorii sunt ridicate, între 0,7759 pentru Necăsătoriți și 0,9946, pentru Etnie.

În *clusterul* 2, atitudinea subiecților este ușor rezervată, probabilitatea maximă de respingere fiind de 0,3927 pentru vecinii de altă rasă.

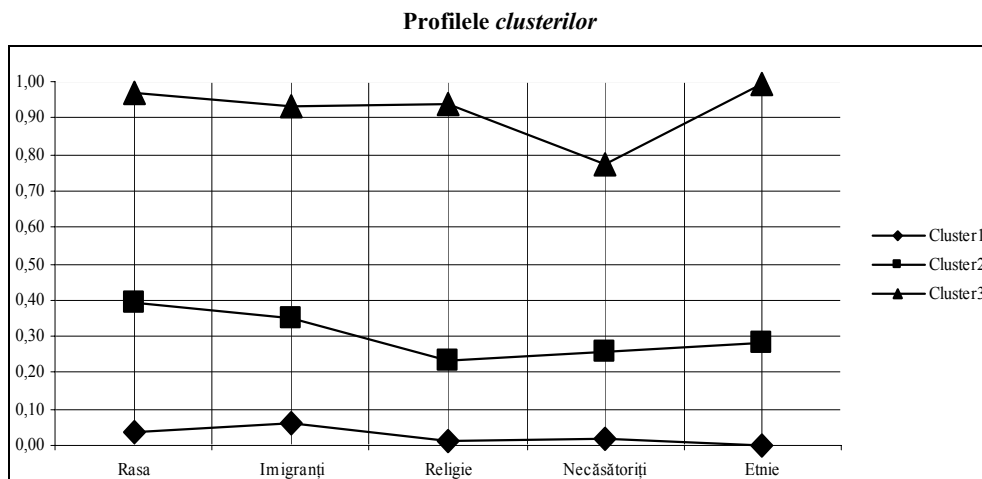
Tabelul nr. 11

Profilul clusterilor – variabile-manifest

	Cluster 1	Cluster 2	Cluster 3
Mărime clusteri	0,7098	0,2224	0,0678
Variabile			
Rasa			
Nemenționat	0,9605	0,6073	0,0337
Menționat	0,0395	0,3927	0,9663
Imigranți			
Nemenționat	0,9382	0,6513	0,0690
Menționat	0,0618	0,3487	0,9310
Religie			
Nemenționat	0,9882	0,7682	0,0583
Menționat	0,0118	0,2318	0,9417
Necăsătorii			
Nemenționat	0,9792	0,7403	0,2241
Menționat	0,0208	0,2597	0,7759
Etnie			
Nemenționat	0,9995	0,7161	0,0054
Menționat	0,0005	0,2839	0,9946

Profilele din fiecare *cluster* sunt prezentate grafic în *Figura 2*. Se observă că separarea profilelor este evidentă, fără intersecțiuni. Pe axa verticală din grafic este indicată probabilitatea de a menționa ca vecin nedorit o persoană cu atributul variabilei-manifest de pe axa orizontală.

Figura 2



În *Latent Gold* sunt prezentate și probabilitățile condiționate de apartenență la un anumit cluster, dacă răspunsul este dat pentru o anumită categorie (*Tabelul nr. 12*). Pe

fiecare rând, suma probabilităților este 1. Astfel, de exemplu, probabilitatea ca un subiect să aparțină la *clusterul* 3, dacă a menționat imigranții ca o categorie nedorită de vecini, este de 0,3426. Dacă nu a făcut această mențiune, probabilitatea este de 0,0056.

Tabelul nr. 12

Probabilitățile de apartenență la *clusteri* (*ProbMeans*)

Variabila	Cluster 1	Cluster 2	Cluster 3
Rasa			
Nemenționat	0,8325	0,1649	0,0026
Menționat	0,1546	0,4826	0,3628
Imigranți			
Nemenționat	0,8168	0,1776	0,0056
Menționat	0,2374	0,4200	0,3426
Religie			
Nemenționat	0,8007	0,1950	0,0043
Menționat	0,0672	0,4163	0,5165
Necăsătoriți			
Nemenționat	0,7946	0,1882	0,0172
Menționat	0,1178	0,4612	0,4209
Etnie			
Nemenționat	0,8165	0,1833	0,0002
Menționat	0,0027	0,4817	0,5156

În *Latent Gold*, rezultatele asociate cu probabilitățile de apartenență sunt prezentate și sub formă grafică, în mod specific modelului de clase latente ales. Modul de prezentare și indicațiile privind interpretarea grafică sunt detaliate în (Magidson și Vermunt, 2001) sau (Van der Ark și Van der Heijden, 1998).

Tabelul nr. 13

Profilul *clusterilor* – covariate

Covariate	Cluster 1	Cluster 2	Cluster 3
	Apropiat	Rezervat	Distant
Genul			
Masculin	0,502	0,501	0,442
Feminin	0,498	0,499	0,558
Educație			
Scăzut	0,431	0,499	0,548
Mediu	0,422	0,402	0,397
Ridicat	0,148	0,099	0,055
Mediul			
Urban	0,605	0,513	0,475
Rural	0,395	0,487	0,525

Ca și în subsetul anterior analizat, introducerea în model a covariatelor gen (masculin sau feminin), educație (nivelul de instruire) și mediu (mediul de rezidență)

ne permite să explicăm structura sociodemografică din *cluster*. Pentru analiză, trebuie să comparăm distribuția marginală a variabilelor covariate din *Tabelul nr. 7* cu distribuția corespunzătoare din interiorul *clusterelor* (*Tabelul nr. 13*). Diferențele semnificative de distribuții indică specificul *clusterului*.

Remarcăm și în acest caz că în *clusterul 3 Distant* femeile sunt preponderente (0,558). În *cluster*ele 1, 2 respectiv *Apropiat* și *Rezervat* bărbații și femeile sunt aproximativ egal reprezentate.

În *clusterul Distant* predomină subiecții din mediul rural (0,525), iar în *cluster*ele *Apropiat* și *Rezervat* sunt preponderente persoanele din mediul urban.

Clusterul Distant este mai populat de subiecți cu nivel de educație scăzut (0,548), în timp ce în *clusterul Apropiat* educația respectă distribuția marginală din eșantion.

CONCLUZII ȘI DISCUȚII

Primele trei categorii de persoane nedorite ca vecini au fost: persoanele dependente de droguri, alcoolicii și homosexualii. Cele mai acceptabile trei categorii de persoane ca vecini au fost: persoanele având religie diferită, cuplurile necăsătorite care trăiesc împreună, oamenii de altă etnie.

Analiza factorială preliminară în clase latente a identificat două componente. Prima componentă, care include subsetul de variabile observate: Dependenți, Seropozitivi, Homosexuali, Alcoolici, Romi și Condamnați, identifică diferențierea prin raportul individului cu valori sociale cum ar fi legea și ordinea în comunitate. A doua componentă, care include subsetul de variabilele observate: Rasa, Imigranți, Religie, Necăsătoriți și Etnie reprezintă distanțarea pentru conservarea intimității culturale și confesionale.

Analiza de clusteri în clase latente pentru prima componentă (set de variabile), identifică patru *clusteri*-profile: *clusterul 2, intolerant*, (probabilități mari de a menționa indezirabili toate categoriile de vecini menționați) și *clusterul 4, definit tolerant*, (probabilități mici de a-i menționa indezirabili). *Clusterul 1* este *tolerant HIV, dar respinge ca vecini persoanele dependente de droguri*. *Clusterul 3* este *tolerant cu Romii, dar se distanțează destul de categoric de dependenții de droguri, de seropozitivi și de homosexuali*.

În *clusterul intolerant* predomină femeile, persoanele din mediul rural. Totuși, femeile acceptă mai ușor vecini infectați cu HIV, iar persoanele din mediul rural sunt mai tolerante cu romii. În *clusterul tolerant și tolerant cu romii* predomină bărbații.

În urban, toleranța la vecinii infectați cu HIV este mult mai mare. Persoanele cu un nivel de educație ridicat sunt cu precădere în *clusterul HIV tolerant*, dar distanțează față de persoanele dependente de droguri sau alcool.

Pentru al doilea subset (variabilele): Rasa (persoane de rasă diferită), Imigranți (imigranți sau muncitori din altă țară), Religie (persoane având o religie diferită), Necăsătoriți (cupluri necăsătorite care trăiesc împreună) și Etnie (oameni

de altă etnie) s-au conturat trei clusteri-profile: *clusterul 1 Apropiat* (grad de apropiere socială ridicat), *clusterul 3 Distant* (grad ridicat de distanțare socială) și *clusterul 2 Rezervat* (grad mediu de distanțare socială).

În *clusterul Distant* predomină femeile, persoanele din mediul rural și cu nivel instructiv-educativ scăzut. În *clusterulele Apropiat și Rezervat*, proporția femeilor și a bărbaților este relativ egală și predomină persoanele din mediul urban.

Am abordat aspectele legate de structurile latente ale distanței sociale, ca atitudine față de vecini, dintr-o perspectivă bivalentă, considerând variabilele latente atât în spațiu continuu cât și discret multidimensional. Toți cei 11 itemi au fost inițial analizați în spațiul bidimensional continuu și apoi fiecare dimensiune, separat, a fost analizată în clase latente discrete. Dualitatea „caracteristică (*trait*) – clasă” este benefică în cercetare, deoarece permite o interpretare complementară, așa cum sociologia poate fi completată interesant prin perspectiva psihologiei. Analiza factorială în factori discreți (*DFactor*) în programul *Latent Gold*, în care se introduc cei 11 itemi analizați anterior prin factori de tip continuu (*CFactor*) ne conduce la un rezultat oarecum similar cu cel din analiza factorială în spectru continuu (*Tabelul nr. 14*).

Sunt, totuși, două excepții importante, și anume, variabilele Seropozitivi și Romi, care sunt în această analiză aproximativ la fel reprezentați pe ambii factori discreți, saturațiile (*loadings*) fiind aproximativ egale ca intensitate. Trebuie însă să remarcăm că, din punctul de vedere al criteriilor informaționale BIC, AIC și CAIC, modelul în factori tip *continuu* este un model mai bun decât modelul în factori *discreți*.

Tabelul nr. 14

DFactor Loadings – Saturații

Loadings	DFactor 1	DFactor 2	Loadings
Dependenți	0,138	-0,551	Dependenți
Rasa	0,717	-0,087	Rasa
Seropozitivi	0,375	-0,370	Seropozitivi
Imigranți	0,628	-0,125	Imigranți
Homosexuali	0,215	-0,537	Homosexuali
Religie	0,607	0,042	Religie
Alcoolici	0,134	-0,475	Alcoolici
Necăsătoriți	0,522	0,022	Necăsătoriți
Etnie	0,673	-0,014	Etnie
Romi	0,286	-0,293	Romi
Condamnați	0,254	-0,451	Condamnați

Analiza în clase latente, în formele ei de bază sau avansate, este un instrument deosebit de util în cercetare. Această metodă aduce cu sine beneficiile analizelor tradiționale, permise numai pentru datele de tip continuu, în domenii ale cercetării care operează, de regulă, cu date discrete. În ciuda faptului că în ultimii ani au apărut programe statistice specializate, cu interfețe utilizator prietenoase, care le fac utilizabile chiar în practica de rutină a cercetării sociologice, analiza

claselor latente încă nu este considerată ca o alternativă viabilă la tehnicile clasice de analiză, nici măcar în situații când o asemenea metodă, teoretic, ar putea fi mult mai potrivită sau chiar este singura aplicabilă.

Statisticieni precum Goodman (2002) au atras atenția asupra faptului că și într-un context mai simplu, cum ar fi cel al unei tabele de contingență de 2×2 , pentru, să zicem, variabilele dihotomice A și B, aproape toate măsurătorile pe care le facem (de exemplu, cunoscutul Pearson χ^2) indică doar cantitatea de neindependență dintre acestea. Nu putem însă cunoaște cu certitudine dacă relația măsurată dintre acestea este una autentică, sau dacă, de fapt, am măsurat simultan relațiile celor două variabile cu o a treia, care există și pe care, atunci când nu este observată direct, o numim *latentă*.

În cercetarea asistată de instrumentele statisticii trebuie avut în vedere că lucrurile nu sunt întotdeauna ce par a fi. Paradoxul Simpson este un exemplu elocvent în acest sens. Analiza de clase latente, ca oricare altă metodă statistică, este o încercare de apropiere către mijlocul unui adevăr presupus. Poate din acest motiv în cartea lor de referință *Latent Structure Analysis*, Paul F. Lazarsfeld and Neil W. Henry au apelat la un încântător și inspirat citat din poemul *The Secret Sits* scris de Robert Frost:

„Dansăm rotund într-un inel și presupunem,
dar Secretul stă în mijloc și cunoaște”³.

Pe lângă produsul *Latent Gold* pe care l-am prezentat, aflat în continuă dezvoltare, în prezent mai există câteva produse *software* care pot fi de interes pentru cercetătorii interesați de analiza structurilor latente. LEM este un program gratuit realizat de Jeroen Vermunt, mai puțin complex și cu o interfață utilizator mai puțin „prietenoasă”. WINMIRA 2001 este un produs specializat în clase latente și modele sau mixturi *Rasch*, cu o interfață ușor de utilizat, compatibil SPSS pentru date și care produce prezentări grafice color. Are, de asemenea, posibilitatea de bootstrap pentru mărirea acurateții rezultatelor. *Mplus*, un program disponibil pe <http://www.statmodel.com>, este destinat modelării cu variabile latente atât discrete cât și de tip continuu.

BIBLIOGRAFIE

1. Abraham, D., Bădescu, I., Chelcea, S., *Interethnic Relations in Romania*, Cluj-Napoca, Editura Carpatica, 1995.
2. Agresti, A., *Categorical data analysis*, New York, Wiley, 1990.
3. Bogardus, E.S.A., *Social Distance Scale*, în “Sociology and Social Research”, nr. 1, 1925.
4. Chelcea, S., Atitudinile etnice ale românilor, în S. Chelcea, *Personalitate și societate în tranziție*, București, Editura Știință și Tehnică S.A., 1994, pp. 227–240.

³ „We dance round in a ring and suppose,
but the Secret sits in the middle and knows”.

5. Clogg, C.C., Latent Class Models, în Arminger, G., Clogg, C.C., Sobel, M.E. (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York, Plenum, 1995, pp. 311–359.
6. Dayton, C.M., *Latent Class Scaling Analysis*, Sage University Papers Series, Quantitative Applications in the Social Sciences, series no. 07–126, Thousand Oaks, CA: Sage, 1998.
7. Dayton, C.M., Macready, G.B., Use of categorical and continuous covariates in latent class analysis, în Hagenaars, J.A., McCutcheon, A.L., (eds.), *Applied Latent Class Analysis*, Cambridge, UK, Cambridge University, 2002, pp. 213–233.
8. Formann, A.K., Kohlmann, T., *Latent class analysis in medical research*, în “Statistical Methods in Medical Research”, Vol. 5, No. 2, 1996, pp. 179–211.
9. Galindo-Garre, F., Vermunt, J.K., *Testing log-linear Models with inequality constraints: A comparison of asymptotic, bootstrap, and posterior predictive p values*, în “Statistica Neerlandica”, Vol. 59, No. 1, 2005, pp. 82–94.
10. Goodman, L.A., Latent Class Analysis. The Empirical Study of Latent Types, Latent Variables, and Latent Structures, în Hagenaars, J.A., McCutcheon, A.L. (eds.), *Applied Latent Class Analysis*, Cambridge, Cambridge University, 2002, pp. 3–55.
11. Henry, W. Neil, *Latent Structure Analysis at Fifty*, paper presented at the “1999 Joint Statistical Meetings”, Baltimore, August 11, 1999.
12. Lazarsfeld, F.P., Henry, W.N., *Latent Structure Analysis*, Boston, Houghton Mifflin, 1968.
13. Magidson, J., Vermunt, J.K., Comparing latent class factor analysis with the traditional approach in data mining, în Bozdogan, H. (ed.), *Statistical Data Mining and Knowledge Discovery*, Boca Raton, Chapman & Hall/CRC, 2003, pp. 373–383.
14. Magidson, J., Vermunt, J.K., *Latent class factor and cluster models, bi-plots and related graphical displays*, în “Sociological Methodology”, Vol. 31, 2001, pp. 223–264.
15. Mihăilescu, I., în Zamfir, C., Vlăsceanu, L., (coord.), *Dicționar de sociologie românească*, București, Editura Babel, 1998, p. 177.
16. Herseni, T., *Sociologia distanței*, în „Societatea de mâine”, an VIII, nr. 3–4, 1931, p. 77.
17. Herseni, T., *Sociologia vecinătății*, în „Societatea de mâine”, an VIII, nr. 6–7, 1931, p. 141.
18. Herseni, T., *Sociologia spațiului*, în „Societatea de mâine”, an VIII, nr. 5, 1931, p. 114.
19. Herseni, T., *Alte distanțe sociale*, în „Societatea de mâine”, an VIII, nr. 16–17, 1931, p. 314.
20. Magidson J., Vermunt J.K., Latent class models, în Kaplan, D. (ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, Thousand Oaks, Sage Publications, chapter 10, 2004, pp. 175–198.
21. Stăvărache, Fl., *Traian Herseni. Câteva considerații asupra distanței sociale*, în „Studii și cercetări din domeniul științelor socioumane”, Cluj Napoca, vol. 10, 2002.
22. Van der Ark, L.A., Van der Heijden, P.G.M., Graphical display of latent budget and latent class analysis, în Blasius J., Greenacre, M. (eds.), *Visualization of categorical data*, Boston, Academic Press, 1998, pp. 489–509.
23. Vermunt, J.K., Magidson, J., Factor Analysis with categorical indicators: A comparison between traditional and latent class approaches, în Van der Ark, A., Croon, M.A., Sijtsma, K., (eds.), *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences*, Mahwah, Erlbaum, 2005, pp. 41–62.
24. Vermunt, J.K., Magidson, J., *Latent GOLD 4.0 and IRT modeling*, 2006, disponibil online la <http://www.statisticalinnovations.com/products/LGIRT.pdf>.
25. Vermunt, J.K., Magidson, J., *Latent GOLD 4.0 User's Guide*, Belmont, Massachusetts, Statistical Innovations Inc., 2005, disponibil online la <http://www.statisticalinnovations.com>.
26. Vermunt, J.K., Magidson, J., *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Belmont Massachusetts, Statistical Innovations Inc., 2005, disponibil online la <http://www.statisticalinnovations.com>.
27. Vlăsceanu, L., în Zamfir, C., Vlăsceanu, L., (coord.), *Dicționar de sociologie românească*, București, Editura Babel, 1998, pp. 177–178.
28. *** *Barometrul de Opinie Publică*, Fundația Soroș România, 2007, disponibil online la http://www.osf.ro/ro/program_articol.php?articol=107.

The goal of this paper is to bring solid arguments for the latent class analysis use as a research tool in the social and behaviour sciences. I started with a technical presentation of the latent class analysis method, limited to what is strictly necessary for the basic understanding and subsequently, continuing with a practical research example. Using the Latent Gold software aimed to identify patterns of the social distance expressed by the wish not to have as neighbours 11 socially stigmatized person categories as drug addicts, immigrants, Roma people, ex convicts, etc. The analyzed data have been provided by Soros Foundation Romania and are part of the Public Opinion Barometer, October 2007. The initial 11 items set was split up into two subsets, using a special form of factorial analysis technique with continuous latent components, which have been later explored using latent classes cluster analysis. For the first variables subset, a four latent classes model was identified as the best fit and for the second variables subset, a three latent classes model was identified as the best fit. In the intolerant and distant clusters prevail women and people from rural areas. However, the women are more likely to accept HIV infected neighbors and the people from rural areas are more tolerant with Roma people. In the tolerant and tolerant with Roma people, the men prevail. In the urban areas, the tolerance for HIV infected neighbors is higher. The people with high educational level are predominant in the cluster HIV-tolerant, but they are distant in respect to drugs or alcohol addicted people.

Keywords: latent classes, factorial analysis, social distance.

Primit: 20. 06. 2010

Acceptat: 14. 12. 2010

Redactor: Ioan Mărginean