

ANÀLISI DE MATRIUS QUADRADES NO SIMÈTRIQUES: UN ENFOCAMENT INTEGRAL USANT ANÀLISI DE CORRESPONDÈNCIES

J. DAUNIS I ESTADELLA*
T. ALUJA BANET**
S. THIÓ FDEZ. DE HENESTROSA*

En aquest article s'introdueix una proposta integrada d'anàlisi de matrius quadrades no simètriques mitjançant anàlisi de correspondències. En ella s'han integrat les dues principals famílies de metodologies que proposen solucions a la problemàtica inherent d'aquest tipus de taules: les diagonals sobrecarregades i els zeros estructurals. S'aplica a l'exemple de l'estudi dels moviments de població deguts al treball entre les 41 comarques catalanes.

Analysis of square skew-symmetric tables: An integral vision by correspondence analysis

Paraules clau: Matrius quadrades, diagonal sobrecarregada, anàlisi de correspondències, zeros estructurals, descomposició en valors singulars, dades mancants

Classificació AMS (MSC 2000): 62H17, 62H25

* Departament d'Informàtica i Matemàtica Aplicada. Universitat de Girona. Av. Lluís Santaló s/n. 17071 GIRONA. Email: josep.daunis@udg.es santiago.thio@udg.es.

** Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya. C/Pau Gargallo, núm. 5. 08028 BARCELONA. Email: tomas.aluja@upc.es.

– Rebut al desembre de 2001.
– Acceptat a l'octubre de 2002.

1. INTRODUCCIÓ

En diferents àmbits científics, ens trobem davant la presència de taules quadrades, on les fileres i les columnes es refereixen al mateix conjunt d'objectes o categories. Aquesta és una tipologia de taula d'ús molt freqüent i en les que ens centrarem en el cas de taules de tipus no simètric, de més ampli ús, on no hi ha d'haver una relació simètrica entre fileres i columnes. Anem a il·lustrar la problemàtica amb les dades objecte del nostre estudi: els moviments migratoris obligats pel treball, entre les comarques de Catalunya corresponents al cens de 1996 (Institut d'Estadística de Catalunya, 2001).

Troblem a la Taula 1 les dades, on les columnes són els orígens dels treballadors i les fileres les destinacions, dels moviments entre les 41 comarques catalanes deguts al treball. Clarament ens trobem amb una matriu quadrada no simètrica. Per a ajudar a interpretar més bé els moviments, ho referirem al mapa adjunt de les 41 comarques catalanes (Figura 1).



Figura 1. Les 41 comarques catalanes. (Origen Web de l'Idescat. Disseny propi)

Habitualment, aquestes taules són analitzades mitjançant anàlisi de correspondències (AC), però en l'anàlisi d'aquests tipus de taules, la principal problemàtica sorgeix en la inèrcia de la diagonal. En aquestes diagonals, els efectius habitualment poden ser o molt

grans amb relació amb la resta d'efectius (diagonals carregades), com podem trobar en l'exemple de mobilitat donat per Kazmierczak (1978), o nuls per l'estructura de la taula (zeros estructurals). Aquesta darrera problemàtica, els zeros estructurals, també es pot extreure, però no tan freqüentment, fora de la diagonal.

L'explicació d'aquesta problemàtica es troba en l'anàlisi de la inèrcia i l'estudi de la descomposició de la inèrcia en aquests tipus de taules per a cadascuna de les caselles. En l'AC la *inèrcia total*, mesura de la variació de la taula \mathbf{P} de terme general p_{ij} és:

$$I(T) = \sum_i \sum_j \frac{(p_{ij} - r_i c_j)^2}{r_i c_j}$$

on p_{ij} és la freqüència relativa de cada casella i r_i i c_j les marginals de fileres i columnes respectivament, essent r i c els vectors marginals. La inèrcia deguda a cada casella (i, j) és $I_{ij} = (p_{ij} - r_i c_j)^2 / (r_i c_j)$.

Es pot descomposar la inèrcia total de la taula en la inèrcia de la diagonal $I(D)$ i en la de fora de la diagonal $I(\overline{D})$, com a la suma de totes les inèrcies de les caselles pertanyents a aquesta o aquella. En la nostra taula, la importància de la inèrcia de la diagonal és del 97.8% sobre una inèrcia total de 28.65. Per tant la representació gràfica factorial obtinguda és bàsicament el moviment dels treballadors que tenen l'origen i el destí dins la pròpia zona. Aquest percentatge tan elevat és un tret característic de les matrius quadrades no simètriques objecte d'estudi.

L'anàlisi d'aquestes taules és problemàtica, per la qual cosa existeixen principalment dues famílies de mètodes de tractament. El nostre objectiu serà analitzar les metodologies existents i proposar una metodologia d'anàlisi integradora que superi la influència d'aquestes caselles productores d'inèrcies estructurals per a una millor anàlisi de les relacions globals entre les modalitats.

2. ANTECEDENTS: METODOLOGIES EXISTENTS DE SOLUCIÓ

La gran influència de la diagonal, en aquestes taules objecte d'estudi, ha portat a la necessitat d'estudiar propostes de modificació de les anàlisis en el sentit de treure la influència d'aquestes caselles. Aquestes propostes de modificació es basen principalment en reconstituir els elements de les diagonals, amb metodologies diferents, atenent a diferents propostes de reconstitució, però també s'ha abordat el problema des d'una metodologia d'intentar descomposar l'anàlisi en referència a la simetria o no simetria de les dades. A les següents seccions tractarem les propostes que diferents autors han realitzat, les compararem breument i finalment proposarem la nostra metodologia de resolució.

2.1. Reemplaçament dels elements de la diagonal sota hipòtesi d'independència

Aquesta proposta, així com les següents, està sustentada en base als *mètodes d'anàlisi factorial en referència a un model* proposats per Escofier (1984), ja que la taula objecte d'estudi és la taula original menys el model d'independència imposat als elements diagonals.

Per anul·lar la component inercial dels elements de la diagonal De Leeuw i van der Heijden (1988) suggereixen redefinir els elements de la diagonal com a producte dels marges, és a dir, sota un model d'independència: $\hat{p}_{ii} = r_i c_i$. Aquesta reconstitució de la matriu de dades sota el supòsit d'independència comporta que les caselles reconstituïdes tinguin, en principi, una inèrcia nul·la: $I(D) = 0$, però no es conserven els marges de les taules i, a més, la reconstitució està influïda pels propis valors de les caselles influents.

L'anàlisi de correspondències de la taula reconstituïda sota la hipòtesi d'independència ens dona que encara es conserva la gran relació d'un mateix lloc com a origen i com a destinació. Aquesta relació és deguda al fort paper que tenen les caselles de la diagonal en les marginals corresponents.

2.2. Tractament com a *dada mancant* de les caselles diagonals

La tècnica usada en aquest apartat s'inclou en les tècniques per al tractament de taules incompletes proposades per Nora (1974), també tractades per Benzécri (1973), així com per Greenacre (1984) i de Leeuw i van der Heijden (1988) entre d'altres, essent ara aplicada al tractament de dades influents.

La influència de les caselles de la diagonal en les marginals porta a una proposta de solució basada en el tractament de les dades de la diagonal com a *dades mancants* (*missing values*), és a dir, considerarem a tots els efectes que aquestes dades són mancants i per tant la seva reconstitució a partir de les marginals no serà influïda per aquestes. Es redefeix cada element de la diagonal com:

$$\hat{p}_{ii} = \frac{r_i^0 c_i^0}{1 - (r_i^0 + c_i^0)}$$

on el superíndex zero significa que l'element de la diagonal ha estat exclòs dels respectius marginals.

Es pot veure que, aplicant aquesta metodologia, es perd l'estructura deguda a la presència de grans quantitats en les caselles de la diagonal i només resten les quantitats degudes a les caselles de fora de la diagonal.

Aquesta reconstitució de la matriu per la imputació d'efectius com a dades mancants comporta que les cel·les reconstituïdes tenen una inèrcia gairebé nul·la, no es conserven

les marginals de la taula però la reconstitució ara no està influïda pels valors diagonals influents. Aquesta mateixa tècnica pot ser usada per al tractament de zeros estructurals fora de la diagonal o de les subtaules diagonals en anàlisi de correspondències múltiples (Greenacre, 1988).

2.3. Reemplaçament dels elements de la diagonal mitjançant la fórmula de reconstitució

Basada en el tractament tipus mancant aplicat a la diagonal i proposat per Mutombo (1973) i Nora (1974), per a dades mancants, hi ha el tractament k-EM. Aquest es basa en un algorisme tipus EM, Dempster (1976), però amb la introducció d'un nou paràmetre, l'ordre k de reconstitució inherent a les anàlisis factorials.

Aquest algorisme està basat en fer una primera reconstitució tipus mancant, on aquest serà el valor inicial de l'algorisme. Llavors s'aplica l'algorisme EM, on les etapes d'estimació (E) i maximització (M) es basen en la imputació dels valors obtinguts en l'etapa anterior i la reconstitució d'ordre k d'aquests valors mitjançant la fórmula inherent a les anàlisis factorials. En aquesta metodologia es conserven les marginals havent-ne exclòs d'elles els elements de la diagonal.

Altres propostes en aquesta mateixa línia, d'imposar un model sobre les dades, han estat proposades, vegi's per exemple Foucart (1985), però també hi ha metodologies alternatives com la de la següent secció.

2.4. Descomposició de la matriu com l'addició d'una matriu simètrica i d'una d'antisimètrica

En un altre vessant totalment diferent, trobem la proposta de Constantine i Gower (1978), revisada per Greenacre (2000), de solució basada en la descomposició de la matriu com a suma d'una matriu simètrica i d'una altra d'antisimètrica. La matriu objecte d'estudi $\mathbf{P}-\mathbf{rc}^T$ pot ésser descomposada en $\mathbf{Q}+\mathbf{R}$ on \mathbf{Q} és una matriu simètrica i \mathbf{R} una matriu antisimètrica. Si anomenem $\mathbf{A}=\mathbf{P}-\mathbf{rc}^T$, llavors podem trobar \mathbf{Q} i \mathbf{R} de la següent manera:

$$\mathbf{Q} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T) \quad \text{i} \quad \mathbf{R} = \frac{1}{2}(\mathbf{A} - \mathbf{A}^T)$$

Utilitzant les propietats de les matrius \mathbf{Q} -amb marginals \mathbf{w} - i \mathbf{R} , i com que:

$$\mathbf{P} - \mathbf{w}\mathbf{w}^T = \mathbf{Q} - \mathbf{w}\mathbf{w}^T + \mathbf{R}$$

Això ens permet fer una descomposició de la inèrcia en la part deguda a la simetria i a la no-simetria. L'anàlisi de $\mathbf{Q}-\mathbf{w}\mathbf{w}^T$ reflexa la inèrcia de la simetria (valors diagonals i mo-

viments recíprocs) i l'anàlisi de \mathbf{R} reflexa la no-simetria (els balanços dels moviments recíprocs).

Llavors en comptes de l'anàlisi estàndard, això és la descomposició en valors singulars generalitzada (DVSG) de la tripleta $(\mathbf{P} - \mathbf{r}\mathbf{c}^T, \mathbf{D}_r^{-1}, \mathbf{D}_c^{-1})$, les anàlisis proposades són les DVSG:

$$(\mathbf{Q} - \mathbf{w}\mathbf{w}^T, \mathbf{D}_w^{-1}, \mathbf{D}_w^{-1})$$

$$(\mathbf{R}, \mathbf{D}_w^{-1}, \mathbf{D}_w^{-1})$$

És a dir se centra en referència a les marginals de \mathbf{Q} i es prenen també les mètriques de la part simètrica.

3. PROPOSTA D'ANÀLISI DE MÀTRIXS QUADRADES NO SIMÈTRIQUES

Atenent a la problemàtica descrita i a les diferents propostes realitzades pels diferents investigadors, l'anàlisi global que suggerim agafarà part d'aquestes propostes (la descomposició en part simètrica i no-simètrica, la reconstitució EM), però també el càlcul de la importància de l'efectiu de la diagonal en relació al global i la problemàtica de la reconstitució influïda per la diagonal. L'anàlisi la dividirem en tres etapes que, per la seva relació de similitud amb els moviments migratoris, seran anomenades:

1. Anàlisi migració intra
2. Anàlisi migració entre i recíproca
3. Anàlisi saldos migratoris

Passarem ara a descriure cadascuna d'aquestes tres parts i posteriorment farem la seva aplicació sobre les nostres dades.

3.1. Anàlisi migració intra

La primera etapa és l'anomenada *anàlisi de la migració intra* o l'anàlisi dels moviments dins de la mateixa zona, és a dir l'estudi de la importància dels valors de la diagonal en referència als seus respectius marges, això ens donarà quins percentatges d'individus romanen dins la mateixa categoria, o fent el símil, quins individus romanen dins la pròpia zona en els seus desplaçaments laborals.

Per a aquesta primera etapa el que farem serà calcular els percentatges del moviment intern de la zona en referència al moviment global vers la mateixa zona, aquest ja serà un primer indicador de l'anàlisi, ja que tota la part no recollida en l'etapa Intra serà susceptible de ser recollida en les etapes posteriors. És una primera part d'anàlisi purament descriptiva.

3.2. Anàlisi migració entre i recíproca

La segona etapa és l'anomenada *anàlisi de la migració entre i recíproca* o l'anàlisi dels moviments simètrics entre cadascuna de les parelles de categories, és a dir, l'estudi de la component simètrica, definida ja en apartats anteriors, però a la qual li traurem la influència dels valors de la diagonal, ja que aquests influeixen en la part simètrica, i la seva importància ja ha quedat expressada en l'apartat anterior.

Per a treure la influència de la diagonal, en l'estudi de la part simètrica el que farem és aplicar l'algorisme de la reconstitució tipus k-EM a la matriu $\mathbf{Q}-\mathbf{ww}^T$, matriu de la component simètrica, reconstitució ja exposada en l'apartat 2.3.

Les etapes de l'algorisme són:

- S'assignen els valors inicials de \hat{q}_{ii} .
- Es realitza una AC de la matriu amb els valors reconstituïts tipus *mancant* (etapa M, o de maximització, de l'algorisme EM), per a trobar els eixos i les coordenades factorials.
- S'utilitza la fórmula de reconstitució de les anàlisis factorials, expressada en termes dels valors propis λ i de les coordenades estàndards ψ i ϕ , d'ordre $k = 1$ del valor \hat{q}_{ii} (etapa E, o d'estimació, de l'algorisme EM).
- Es reiteren els passos (b) i (c) fins a la convergència dels valors reconstituïts.

Així:

$$\hat{q}_{ii}^{(1,m)} = r_i^{(1,m-1)} c_i^{(1,m-1)} \left\{ 1 + \sqrt{\lambda_\alpha^{(1,m-1)}} \psi_{\alpha i}^{(1,m-1)} \phi_{\alpha i}^{(1,m-1)} \right\}$$

on els superíndexs $(1,m)$ representen l'ordre de reconstitució i l'ordre d'iteració respectivament.

- Una vegada obtinguda aquesta convergència, s'incrementa l'ordre de reconstitució, essent ara $k = 2$, prenent com a valor inicial l'obtingut en l'etapa (d) del procés anterior. Es reiteren les etapes de maximització i d'estimació fins a assolir de nou la convergència, amb ordre de reconstitució $k = 2$.

Es finalitza el procés augmentant l'ordre de reconstitució, utilitzant a l'inici de cada nova etapa com a valor inicial el valor obtingut al final del procés d'iteració d'ordre inferior. La fórmula general de l'etapa d'estimació és:

$$\hat{q}_{ii}^{(k,m)} = r_i^{(k,m-1)} c_i^{(k,m-1)} \left\{ 1 + \sum_{\alpha=1}^k \sqrt{\lambda_\alpha^{(k,m-1)}} \psi_{\alpha i}^{(k,m-1)} \phi_{\alpha i}^{(k,m-1)} \right\}$$

S'acaba el procés d'acord, per exemple, amb els valors propis significatius en l'anàlisi de la taula, indicador habitual de la qualitat de representació de cada eix.

3.3. Anàlisi saldos migratoris

Finalment, la darrera etapa és l'anomenada *anàlisi dels saldos migratoris* o l'anàlisi de la diferència de moviments no recíprocs entre categories, és a dir, l'estudi de la component no simètrica dels moviments entre objectes.

Per a realitzar aquesta darrera part realitzarem l'estudi ja presentat de la matriu **R**, matriu corresponent a la no-simetria, on realitzarem sobre aquestes dades la descomposició en valors singulars generalitzada de la tripleta:

$$(\mathbf{R}, \mathbf{D}_w^{-1}, \mathbf{D}_w^{-1})$$

És a dir ponderem en referència a les marginals i prenem també les mètriques de la part simètrica **Q**.

4. EXEMPLE D'APLICACIÓ

En aquesta secció anem a aplicar la proposta d'anàlisi efectuada a l'exemple dels moviments deguts al treball entre les comarques catalanes, el qual el tractarem en base als tres ítems presentats: (1) anàlisi migració intra, (2) anàlisi migració entre i recíproca i (3) anàlisi saldos migratoris.

4.1. Anàlisi migració intra

Començarem amb l'estudi de la migració intra, és a dir la importància dels moviments de treballadors de la pròpia zona sobre el total d'efectius que arriben a la zona. En la Figura 2 trobem una representació gràfica d'aquests moviments. Podem observar que, del total de moviments sobre cadascuna de les zones, trobem que com a mínim un 61.2% d'aquests moviments pertanyen a gent de la pròpia zona (Baix Llobregat) mentre que arriba a uns màxims de més d'un 90% a la Val d'Aran, Segrià, Osona, Alt Urgell, Alt Empordà i Garrotxa. És a dir, molt més de la meitat de les migracions sobre una determinada zona són internes. Es dona la circumstància que les zones amb una migració Intra més altes es corresponen a zones més perifèriques i a zones nord, mentre que aquelles on la migració intra és més baixa són les que estan situades en zona d'influència de grans capitals, sobretot de Barcelona. Ho podem visualitzar globalment en la Figura 3.

4.2. Anàlisi migració entre i recíproca

L'anàlisi de la migració recíproca entre zones, comporta fer l'anàlisi de la matriu de moviments simètrics, però sense la influència dels valors de la diagonal, per la qual cosa es realitzarà una anàlisi tipus k-EM de la matriu de simetria.

MIGRACIÓ INTRA

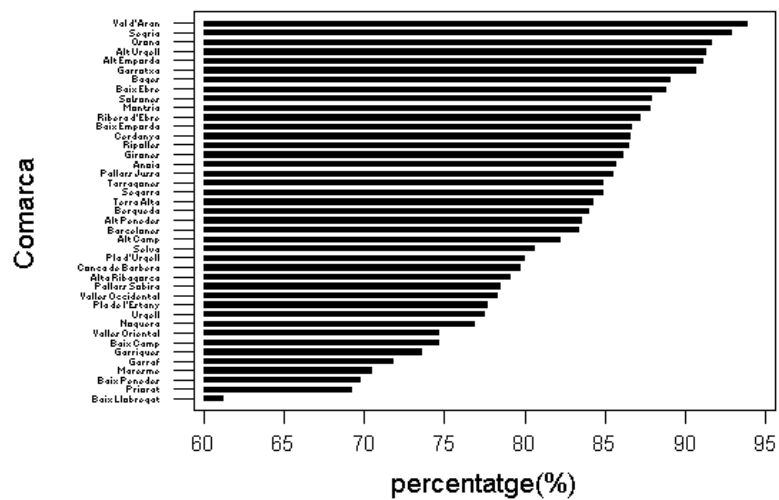


Figura 2. Anàlisi de la migració intra.

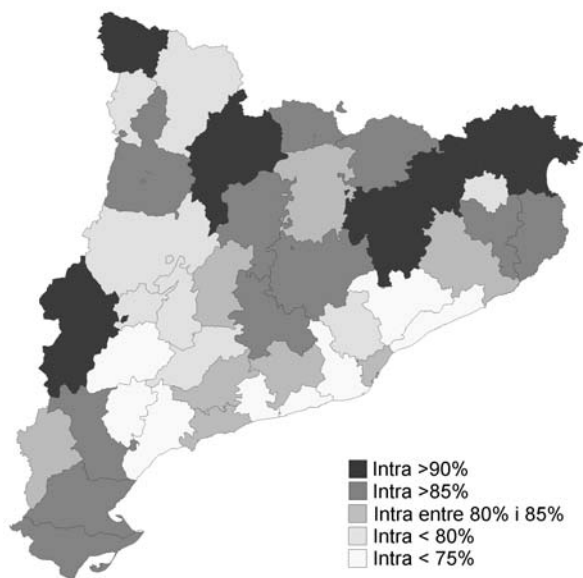


Figura 3. Mapa de la migració intra.

En l'anàlisi dels dos primers eixos factorials (qualitat = 33.7%) podem trobar en el primer eix les zones on els intercanvis són importants situades totes elles a les comarques meridionals amb dues comarques de transició. Queden configurades a l'extrem de l'eix pel Baix Camp i el Tarragonès, diferenciades del Montsià i Baix Ebre, per una banda, i més cap a l'origen de l'eix Alt Camp, Priorat, Ribera d'Ebre i Terra Alta. Les comarques de transició són la Conca de Barberà i el Baix Penedès. En l'origen de l'eix hi ha un gran cúmul de punts.

Mentre que en el segon eix trobem un altre grup, format per les comarques de la regió occidental amb la mateixa estructura de dues zones i comarques de transició cap a l'origen. Garrigues, Noguera, Pla d'Urgell, Segarra, Segrià, i Urgell formarien la primera zona, i més propera a l'origen, una segona zona formada per Alt Urgell, Alta Ribagorça, els dos Pallars i la Val d'Aran, i de transició cap al gran centre el Solsonès i l'Anoia.

Més explícit és el gràfic amb els tres primers eixos factorials (qualitat = 46.4%) (Figura 4) on es reproduïx l'estructura que inclou un tercer eix on les comarques nord-orientals ens repeteixen l'estructura zonal (una zona i comarques de transició) i es veu clarament que destaca la posició central del Barcelonès respecte els tres eixos independents de desplaçaments. L'anàlisi ens permet també elaborar un mapa de Catalunya classificat amb comarques amb gran relació entre i recíproca (Figura 5).

4.3. Anàlisi saldos migratoris

Finalment s'efectua l'estudi dels saldos migratoris, és a dir, les diferències entre moviments rebuts d'altres zones i moviments vers altres zones. La realització d'aquesta fase comporta l'estudi de la matriu corresponent a la no simetria. Trobarem en la Figura 6 la representació gràfica d'aquests moviments.

En aquest darrer gràfic dels dos primers eixos factorials (qualitat = 31.15%) podem veure les relacions de saldos migratoris, podent comprovar primerament que donat el caràcter antisimètric de la matriu analitzada, les representacions dels individus com a productors de saldos positius o negatius, és a dir com a receptors o com a productors de moviments de treballadors, està subjecte a una rotació de 90 graus.

Per altra banda, podem veure relacions de saldos entre les comarques del Baix Camp i el Tarragonès, amb un gran saldo a favor d'aquesta darrera comarca. Uns moviments amb origen al Priorat i destins a l'Alt i Baix Camp i al Tarragonès i finalment d'origen a la Terra Alta i destí a la Ribera d'Ebre. Es pot observar que aquestes zones amb grans diferències de saldos migratoris òbviament corresponen a zones que no tenen una gran migració intra. El tercer eix factorial ens aporta unes inter-relacions triangulars entre les Garrigues, la Noguera i el Pla d'Urgell. Aquestes relacions les podem trobar representades en la Figura 7.

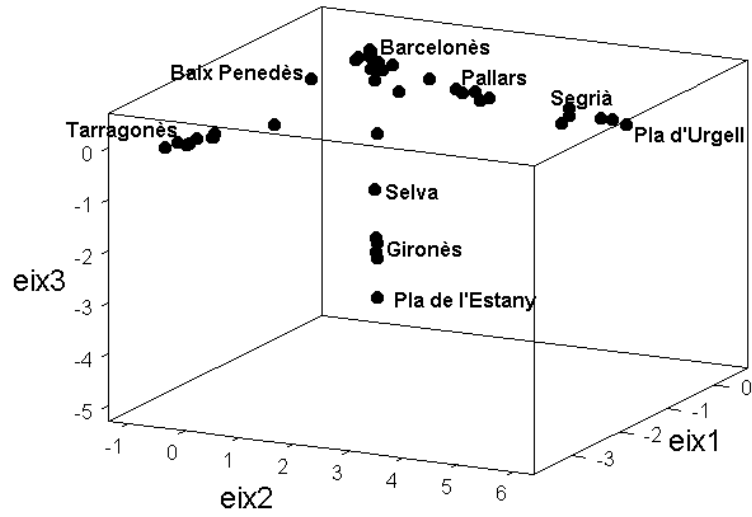


Figura 4. Gràfic factorial 3-EM de l'anàlisi de la migració entre i recíproca.

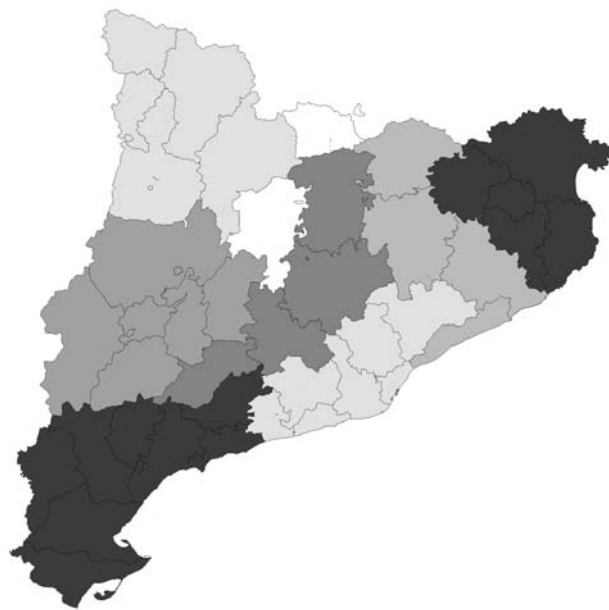


Figura 5. Mapa de les zones de migració entre i recíproca.

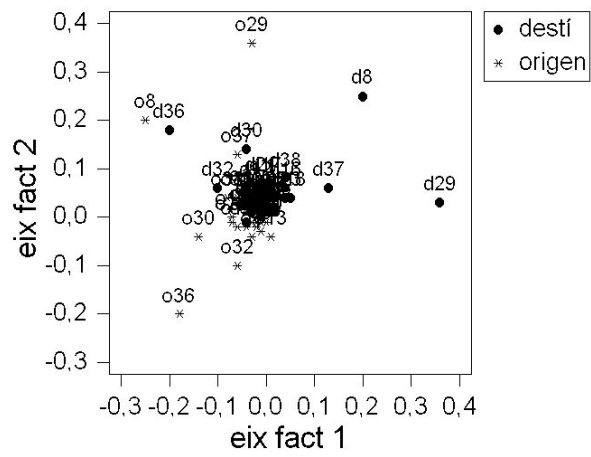


Figura 6. Gràfic factorial de l'anàlisi dels saldos migratoris.



Figura 7. Representació dels principals saldos migratoris.

5. CONCLUSIONS

En aquest treball es presenta una proposta d'anàlisi de matrius quadrades no simètriques. En ella s'han integrat les dues principals metodologies de tractament, la reconstitució de caselles i la descomposició en simetria-antisimetria, aprofitant els avantatges de cadascun dels dos mètodes. Aquesta integració pretén donar una nova visió, més clarificadora, del tractament d'aquests tipus de matrius. S'ha integrat en la metodologia un primer apartat de l'anàlisi dels moviments dins la pròpia zona, ja que és un dels principals indicadors de moviment. En segon lloc, l'anàlisi dels moviments simètrics sense ser influïts pels moviments interns, ens permet detectar regions en el mapa de forta relació mútua i finalment els moviments no simètrics ens permeten detectar els diferents pols d'atracció sobre el territori.

REFERÈNCIES

- Benzécri, J. P. (1973). *L'analyse des Données*. Dunod, Paris.
- Constantine, A. G. i Gower, J. C. (1978). «Graphical representation of asymmetry». *Applied Statistics*, 27, 297-304.
- Daunis-i-Estadella, J. i Aluja-Banet, T. (1997). «Análisis de correspondencias de matrices cuadradas no simétricas. Problemática y tratamientos». *Actas XXIII Congreso Nacional de Estadística e Investigación Operativa*, 40.3-40.4.
- Daunis-i-Estadella, J., Aluja-Banet, T. i Thió-Henestrosa, S. (1997). «Reconstitución de datos influyentes en análisis de correspondencias». *NGUS'97 IV International Meeting of Multidimensional Data Analysis*, 108-111.
- Dempster, A. P., Laird, N. M. i Rubin, D. R. (1976). «Maximum likelihood from incomplete data via EM algorithm». *Journal of the Royal Statistical Society-Series B*, 93, 1-38.
- Escofier, B. (1984). «Analyse factorielle en reference a un modèle. Application a l'analyse de tableaux d'échanges». *Revue de Statistique Appliquée*, 32 25-36.
- Foucart, T. (1985). «Tableaux symétriques et tableaux d'échanges». *Revue de Statistique Appliquée*, 33, 37-54.
- Greenacre, M. J. (1984). *Theory and applications of Correspondence Analysis*. Academic Press, London.
- Greenacre, M. J. (1988). «Correspondence analysis of multivariate categorical data by weighted least-squares». *Biometrika*, 75, 457-467.
- Greenacre, M. J. (2000). «Correspondence analysis of square asymmetric matrices». *Applied Statistics*, 49, 297-310.

- Institut d'Estadística de Catalunya (2001). «Estadística de població 1996». *Vol 12. Fluxos de mobilitat obligada pel treball o estudi. Dades Comarcals i municipals*, 21-25.
- Kazmierczak, J. B. (1978). «Migrations interurbaines dans la banlieue sud de Paris». *Cahiers de l'Analyse des Données*, 3, 203-218.
- De Leeuw, J. i van der Heijden, P. G. M. (1988). «Correspondence analysis of incomplete contingency tables». *Psychometrika*, 53, 223-233.
- Mutombo, F. K. (1973). «Traîtement des données manquantes et rationalisation d'un réseau de stations de mesures». *PhD thesis, Université Paul et Marie Curie, Paris VI*.
- Nora, C. (1974). «Une méthode de reconstitution et d'analyse de données incomplètes». *PhD thesis, Université Paul et Marie Curie, Paris VI*.

Taula 1. Moviment de treballadors a les 41 comarques catalanes.

10119	4	28	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
A. Camp-1	0	32542	7	0	0	16	9	2	2	453	67	4	753	4	3	0	289	10	2	0	2
A. Empordà-2	17	6	25368	1	0	262	31	15	3	10	983	487	1879	42	1	4	343	5	5	127	1244
A. Penedes-3	0	3	2	5508	3	3	13	0	0	2	21	1	162	7	70	0	3	0	2	2	21
A. Urgell-4	0	0	0	0	1029	5	1	1	0	4	6	0	91	0	0	0	1	0	0	0	3
A. Ribagorça-5	17	19	411	7	0	27531	180	5	1	14	1115	13	1773	12	1	100	36	0	12	6	
Anoia-6	0	15	54	16	0	252	48169	15	3	7	894	15	2130	436	20	4	17	0	8	31	
Bages-7	579	4	41	1	1	20	17	37061	134	7	130	170	1194	9	0	95	33	23	16	30	
B. Camp-8	34	2	8	0	0	1	2	234	19487	1	31	23	380	1	1	4	9	4	0	7	
B. Ebre-9	1	524	7	3	0	7	19	2	1	31730	122	9	1164	2	11	0	3	0	82	2179	
B. Empordà-10	14	36	968	3	3	693	464	47	16	28	142346	131	74642	19	9	10	646	1	17	100	
B. Llobregat-11	137	1	889	1	0	29	18	89	12	3	430	12050	2087	3	0	16	469	2	2	9	
B. Penedes-12	61	226	1080	32	5	586	806	206	73	186	52938	241	613094	85	45	24	1040	14	69	563	
Barcelona-13	1	4	18	9	0	15	760	5	1	7	62	4	501	10802	70	0	16	0	5	10	
Berguedà-14	0	2	3	71	0	1	9	0	0	3	25	0	358	10	4354	0	0	0	2	32	
Cerdanya-15	395	2	14	0	1	64	1	110	8	1	37	9	237	1	0	5101	0	17	0	1	
Conca Barba-16	17	5	855	3	2	44	33	39	5	0	1272	547	5563	6	6	11	23481	0	2	13	
Garrúf-17	14	0	7	5	1	10	11	20	0	4	28	8	173	0	0	71	4	4767	0	9	
Garrigues-18	0	148	3	1	0	1	5	3	0	53	22	0	262	1	4	0	0	0	0	9	
Garróxa-19	2	683	6	4	1	40	21	6	0	1065	150	6	1072	9	6	1	7	0	16914	656	
Gironès-20	5	52	59	0	0	51	65	11	10	69	1577	8	25345	10	7	2	43	1	25	319	
Maresme-21	19	2	3	1	0	3	10	74	1468	2	24	12	252	0	0	2	4	0	1	8	
Montsià-22	3	3	2	40	8	39	10	9	2	5	45	5	387	3	3	7	1	16	0	8	
Noguera-23	5	18	12	3	3	26	229	5	4	22	184	1	1672	61	26	0	15	0	27	45	
Osona-24	1	0	4	13	7	2	4	5	0	4	23	1	268	1	4	0	0	0	1	3	
Pallars J.-25	3	1	1	10	0	1	7	4	0	1	24	3	198	1	1	0	3	0	3	5	
Pallars S.-26	5	1	4	6	1	32	8	9	3	1	37	2	190	0	4	11	4	78	3	2	
Pla d'Urgell-27	0	179	0	0	0	1	0	0	0	67	8	1	131	0	1	0	2	1	300	1268	
Pla Espanyol-28	7	1	3	0	0	3	3	245	11	0	9	7	120	0	0	2	3	27	1	2	
Primitiu-29	7	0	7	1	0	0	1	175	102	5	10	8	163	0	2	3	7	3	0	4	
Ribera Ebre-30	0	30	8	0	1	2	8	2	1	25	40	0	382	9	52	0	0	0	159	129	
Ripollès-31	1	2	3	11	1	162	19	6	2	1	42	2	273	1	3	20	3	7	1	1	
Segarra-32	11	10	21	61	18	61	38	52	12	8	155	12	1345	4	7	25	11	245	4	36	
Segrià-33	0	114	8	1	0	9	15	5	4	241	105	10	1393	10	3	1	5	1	167	3331	
Selva-34	0	0	0	54	0	17	141	7	0	0	13	1	123	36	0	4	0	0	0	3	
Solsones-35	1000	8	109	5	0	29	19	3918	101	11	287	688	2205	1	1	140	101	19	2	21	
Tarragonès-36	2	0	1	1	1	0	2	1	30	67	0	15	9	94	0	1	0	2	0	0	
Terra Alta-37	7	4	6	4	1	56	9	9	1	0	43	2	270	0	1	30	3	23	0	6	
Urgell-38	0	2	0	3	7	0	1	0	0	0	9	0	77	0	1	0	2	0	0	1	
Vál d'Aran-39	16	54	292	12	1	199	677	25	16	63	6364	76	37886	25	4	7	109	7	23	132	
Valles Oc.-40	25	25	41	0	0	46	170	12	2	59	1417	16	15007	20	3	0	44	1	9	149	
Valles Or.-41																					790

A. Camp-1	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
A. Empordà-2	1	1	28	0	1	7	139	0	2	17	7	11	107	0	9	0	0	5	24	6
A. Penedes-3	5	2	7	0	1	0	14	0	4	0	2	5	7	4	71	2	1	1	84	49
A. Urgell-4	0	22	10	9	3	2	0	0	1	11	11	72	2	42	3	2	6	2	279	70
A. Ribagorça-5	1	2	0	23	0	2	0	0	0	0	1	69	0	1	3	1	2	47	6	3
Aonúa-6	0	5	13	1	0	11	1	1	2	2	331	34	6	9	14	0	20	2	300	86
Bages-7	1	4	272	1	5	3	7	1	3	10	40	28	19	159	35	1	3	1	1185	272
B. Camp-8	29	7	12	1	3	10	0	50	263	2	2	69	6	4	9449	9	8	9	119	47
B. Ebre-9	1180	0	3	4	2	2	1	4	61	4	7	23	6	0	331	34	2	0	22	22
B. Empordà-10	0	0	21	3	1	0	49	0	15	2	6	319	0	10	0	2	2	179	86	
B. Llobregat-11	13	5	126	5	5	6	10	0	12	6	20	56	81	3	210	0	16	6	9424	1912
B. Penedes-12	50	0	9	0	0	6	7	6	4	2	0	14	7	5	640	1	2	1	232	45
Barcelona-13	34	32	564	24	9	30	30	12	36	62	77	243	663	24	618	1	32	17	38554	16455
Berquesú-14	0	3	278	0	3	1	2	0	1	36	4	9	5	90	6	0	0	1	90	32
Cerdanya-15	0	5	10	0	0	1	1	0	0	22	3	5	20	6	1	0	0	0	63	10
Conca Barb.-16	4	2	4	0	1	2	0	3	1	0	17	29	3	0	262	0	30	2	30	10
Garraf-17	3	2	6	1	0	2	1	2	7	122	0	14	9	0	127	0	1	2	328	135
Garrigues-18	0	14	6	7	2	303	0	6	12	1	40	830	1	1	43	3	49	3	20	8
Garrotxa-19	0	0	24	0	0	0	208	0	0	92	0	1	186	0	4	0	1	0	33	21
Gironès-20	1	1	33	0	0	0	707	0	0	33	2	7	2620	0	19	0	1	2	180	101
Maresme-21	10	6	99	1	0	10	25	0	3	8	0	13	1908	1	61	0	4	4	2011	2505
Montsià-22	15782	3	5	2	0	1	0	0	34	0	0	16	6	1	184	21	1	1	30	7
Noguerà-23	0	9383	2	20	11	212	0	1	0	5	307	1265	8	22	26	0	277	14	36	19
Osona-24	2	1	45904	0	1	2	12	0	2	307	8	15	98	9	18	0	0	1	256	1065
Pallars J.-25	2	24	0	3890	42	6	0	0	2	0	19	131	4	3	16	0	7	5	38	8
Pallars S.-26	0	10	3	48	1839	5	0	0	1	0	3	94	2	0	3	0	3	19	37	7
Pla d'Urgell-27	1	170	2	5	0	0	7345	0	0	4	0	2	91	0	18	0	401	7	27	17
Pla d'Estany-28	0	0	8	0	1	0	0	0	0	0	1	13	1	0	224	2	0	0	23	13
Priorat-29	2	0	1	0	1	0	0	2095	217	0	1	18	2	0	181	59	1	0	19	1
Ribera d'Ebre-30	9	4	0	0	3	0	0	87	6113	0	2	18	2	0	151	0	1	1	26	8
Ripollès-31	0	0	362	1	0	0	13	0	0	8980	0	2	18	0	6	0	0	0	75	59
Segarra-32	0	49	3	1	0	35	0	0	0	1	5895	117	0	26	3	0	198	1	45	7
Segrià-33	3	451	18	62	24	656	3	5	41	14	157	55346	11	24	136	1	214	43	148	61
Selva-34	0	1	87	0	0	1	62	0	1	5	4	10	34022	3	11	0	0	3	284	700
Solsonès-35	0	8	3	5	3	4	1	0	0	2	22	35	5	5	3871	1	0	1	4	28
Tarragonès-36	44	2	13	0	3	1	1	25	115	3	6	73	9	2	52114	11	9	1	210	70
Terra Alta-37	11	2	3	1	0	0	0	4	280	0	5	2	1	0	95	3434	2	0	8	1
Urgell-38	1	220	6	10	3	380	1	0	0	0	907	466	5	4	26	0	8888	5	41	17
Val d'Aran-39	0	3	1	6	4	1	0	0	0	0	0	53	1	0	4	0	0	0	11	2
Valles Occ.-40	14	12	205	4	7	7	11	1	10	13	21	54	159	6	129	1	9	10	194624	6843
Valles Ori.-41	3	0	561	0	3	1	12	0	3	10	15	11	632	2	42	0	4	1	8989	82507

ENGLISH SUMMARY

ANALYSIS OF SQUARE NON-SYMMETRIC TABLES: AN INTEGRAL VISION BY CORRESPONDENCE ANALYSIS

J. DAUNIS I ESTADELLA*
T. ALUJA BANET**
S. THIÓ FDEZ. DE HENESTROSA*

In this paper a new integrated approach to the analysis of square non-symmetric tables is introduced by means of correspondence analysis. The application of correspondence analysis to such tables is not successful due to the strong role played by the diagonal values: overloaded diagonals and structural zeros. Two main families of methods of resolution are integrated in this paper. The resulting method is applied to a study of commuting between the 41 catalan counties.

Keywords: Square matrices, overloaded diagonals, correspondence analysis, structural zeros, singular value decomposition, missing data

AMS Classification (MSC 2000): 62H17, 62H25

* Departament d'Informàtica i Matemàtica Aplicada. Universitat de Girona. Av. Lluís Santaló s/n. 17071 GIRONA. Email: josep.daunis@udg.es santiago.thio@udg.es.

** Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya. C/Pau Gargallo, núm. 5. 08028 BARCELONA. Email: tomas.aluja@upc.es.

– Received December 2001.

– Accepted October 2002.

In many scientific areas we deal with square tables, where rows and columns refer to the same set of objects or categories. Such non-symmetric square tables are very common, where there is not a symmetric relation between rows and columns. We illustrate our problem in the context of migratory movements due to work (*commuting*) between the 41 catalan counties, using data from the 1996 census.

Often these tables are analyzed by means of correspondence analysis (CA), but the application of CA to such tables is not successful due to the strong role played by the diagonal values in the total inertia. On this diagonal the entries are often greater than the non-diagonal entries (overloaded diagonals) or empty due to the table structure (structural zeros).

The table's total inertia can be decomposed as the sum of diagonal inertia $I(D)$ and non-diagonal inertia $I(\overline{D})$.

The analysis of these tables is problematic and there are two families of methodologies to do it, based mainly on the reconstitution of diagonal elements, or trying to decompose the analysis referring to the symmetry and skew-symmetry of data. Our goal is to analyze both methodologies and to integrate them, avoiding the influence of those elements that produce structural inertia and analyzing the joint relationship existing between row and column categories separately from the polarities established among them.

These existing methodologies are:

- Replacement of diagonal elements under independence hypothesis
- Treatment of diagonal elements as *missing values*
- Replacement of diagonal elements using the reconstitution formula
- Decomposition of the matrix as the sum of a symmetric matrix and a skew-symmetric matrix

The proposed analysis can be divided into the following three steps, named by analogy with migratory movements:

1. Intra migration analysis
2. Between and reciprocal migration analysis
3. Migratory balance analysis

The first step, called intra migration analysis, studies the movements inside the same zone, that is it analyzes the importance of the diagonal in reference to the margins.

The second step, called between and reciprocal migration analysis, is the analysis of the symmetric movements between every two pairs of categories, that is, the study of the symmetric component without the diagonal influence using the k-EM reconstitution algorithm on the matrix that contains the symmetric component.

Finally, the last step, called migratory balance analysis, consists of the analysis of the difference of non-reciprocal movements between categories, that is the study of the skew-symmetric component of movements between the objects.

As a conclusion, this work presents a new proposal for the analysis of non-symmetric square matrices. Here we integrate the reconstitution methodology and the decomposition in both symmetric and skew-symmetric components, gaining the advantages of both methods. This integration tries to give a new vision, a clearer vision, of treatment of this kind of data matrix.