Anàlisi de correspondència en matrius quadrades

Yan Hong Chen
Tutor: Josep Lluís Solé

Febrer 2016

Índex

1 Introducció

De taules de freqüències quadrades en podem trobar en diferents àmbits. Habitualment són taules a on les files i les columnes tenen les mateixes categories en circumstàncies diferents. Per exemple, en l'àmbit demogràfic les taules de moviments migratoris recullen el nombre d'habitants en una zona abans i després d'algun succés determinat. En sociologia trobem les taules de mobilitat social; en investigació de mercats, taules de fidelitat a la marca; en psicologia, les matrius de confusió. Sovint aquestes taules són dominades per la seva diagonal amb valors molt elevats i no permeten veure la relació existent dels valors situats fora de la diagonal.

L'anàlisi d'aquestes matrius quadrades és l'objectiu d'aquest treball. Volem aplicar una anàlisi de correspondència **(AC)** sobre aquestes matrius dominades per la seva diagonal per poder veure què ens diuen els valors que es situen fora de la diagonal.

Presentem, en primer lloc, els conceptes bàsics de l'AC: l'**espai de perfils** (tema 2), les **distàncies** χ^2 **i la inèrcia** (tema 3). Expressarem la inèrcia en funció dels perfils i de la massa (equació 8). S'aplicaran les definicions sobre unes dades reals extretes de l'INE, el nombre d'habitants de Catalunya en funció del lloc de naixement per províncies. (Padró continu de l'1 de gener 2014) (Taula 1).

Es defineix la **matriu d'inèrcia** (equació 11) i es tracta **l'algoritme AC** (tema 4) basat en la **descomposició de valors singulars (SVD)** (tema 3.5) aplicats a les dades **"smoke"** (Taula 6) del paquet "ca" de R (tema 4.1), amb la intenció de poder representar la informació a través de gràfics bidimensionals (tema 7). El cost de poder resumir la informació en un espai més petit es valorarà a través de les inèrcies i la seva descomposició (tema 6). De fet l'AC és una **anàlisi de components principals** (ACP) ad hoc a les taules de contingències (tema 3.6).

Arribats a aquest punt, ja tenim el material a punt per fer front a l'**AC de les taules quadrades** amb gran pes a la diagonal (tema 8). Hi apliquem l'algoritme AC basat en SVD, creat a partir de la teoria AC amb les dades clàssiques de **Karl Pearson** sobre mobilitat social (professions de pares i fills) (Taula 7)), el primer cas històric que es va estudiar de l'AC en matrius quadrades amb un pes important a la diagonal.

L'estratègia que s'intenta aplicar per evitar aquest excessiu pes a la diagonal és dividir l'anàlisi de correspondència (AC) en dues parts: **anàlisi de la part simètrica** de la taula (amb la diagonal inclosa), que conté la majoria de la informació de la taula, i l'**anàlisi de la part antisimètrica** amb la resta dels valors. Aquesta segona anàlisi ens hauria de permetre visualitzar la magnitud i el flux existents entre les files i les columnes.

En un primer intent de l'estratègia, **descomposem la matriu de dades** en dues matrius, una simètrica i una altra antisimètrica. La matriu simètrica ens permet l'estudi de la simetria de les dades, però amb la matriu antisimètrica veiem que en tant que conté valors negatius, moltes de les definicions bàsiques pròpies de l'AC no permeten treballar amb aquests valors cosa que comporta un canvi en l'estratègia.

La idea és que en comptes de descomposar, es **composa una nova matriu** (tema 8.2) a partir de la matriu de dades, que ens permet fer l'AC de la part simètrica i l'antisimètrica alhora. Sobre la nova matriu s'aplica l'algoritme AC basat en SVD i també es s'ofereix una modificació de l'algoritme AC basat en la **descomposició de valors singulars Generalitzat** (GSVD).

Finalment s'aplica de nou l'algoritme AC basat en el SVD (tema 9) a unes dades del cens del 1996 referents als moviments migratoris obligats pel treball entre les comarques de Catalunya (taula en l'apèndix C.1).

L'apèndix, que consta de tres apartats (A, B i C), inclou els codis de R. En l'apartat A (apèndix A) hi ha els codis essencials, és a dir, les funcions de R creades que permeten executar l'algoritme AC sense fer servir el paquet "ca" de R. En l'apartat B (apèndix B) hi ha els codis R utilitzats pels diferents exemples i finalment l'apartat C (apèndix C) són les dades del cas pràctic (Moviments migratoris obligats pel treball) i els codis R.

2 Perfils i espai de perfils

En l'anàlisi de correspondències (AC) Els percentatges respecte al total són les **freqüències relatives**.
Els percentatges files o percentatges columnes s'anomenen **perfils fila** i **perfils columna** respectivament.

Considerem la taula contingències de la **població** per lloc de **naixement i províncies**. Les **files** indiquen el **lloc de naixement** i les **columnes** indiquen el nombre de **població** de cada província. (Padró continu a 1 de gener del 2014) (Taula 1)

2.1 Taula de freqüències

Taula de freqüències (Taula 1)
És el recompte dels casos.

	Bcn	Gir	Lld	Trg	Total
Barcelona	3394	89	32	89	3604
Girona	36	396	2	2	436
Lleida	49	4	269	10	332
Tarragona	38	3	4	418	463
Total	3517	492	307	519	4835

Taula 1: Població per lloc naixement i províncies. Padró continu 1 gener 2014. Unitats: Milers de persones. Font: INE

2.2 Taula de freqüències relatives o Matriu de Correspondències

Cada valor és dividit pel total del nombre de casos. (Taula 2)
Sigui n_{ij} els elements de la taula de freqüències.
n el nombre total d'elements, n=4835
Els **elements de la matriu de correspondències** o les freqüències relatives s'obtenen com:

$$p_{ij} = \frac{n_{ij}}{n} \tag{1}$$

2.3 Perfils fila

Taula de perfils fila (Taula 3)
Cada valor és dividit pel total de la seva respectiva fila. Els totals de les columnes dividit pel total dels casos l'anomenem **Perfil fila mitjà o Massa de les columnes**. La característica dels perfils

	Bcn	Gir	Lld	Trg	Total
Barcelona	0.702	0.018	0.007	0.018	0.745
Girona	0.007	0.082	0.000	0.000	0.090
Lleida	0.010	0.001	0.056	0.002	0.069
Tarragona	0.008	0.001	0.001	0.086	0.096
Total	0.727	0.102	0.063	0.107	1.000

Taula 2: Freqüències relatives.

fila és que la suma de les files és 1 i es poden considerar com una ponderació de les columnes. Fixem-nos que la **columna total són tots 1**.

$$p_{i.} = \frac{n_{ij}}{n_{i.}}; \quad n_{i.} \text{ és el marginal de la fila } i \tag{2}$$

	Bcn	Gir	Lld	Trg	Total
Barcelona	0.942	0.025	0.009	0.025	1.000
Girona	0.083	0.908	0.005	0.005	1.000
Lleida	0.148	0.012	0.810	0.030	1.000
Tarragona	0.082	0.006	0.009	0.903	1.000
Perfil fila mitjà o Massa columna	0.727	0.102	0.063	0.107	1.000

Taula 3: Perfils fila.

Perfil fila mitjà o Massa de la columna
És una mitjana ponderada dels cincs perfils en AC s'anomena **centroide**.
Per exemple:

$$\textbf{Perfil fila mitjà}{=}\frac{3517}{4835}*\text{P.Bcn} + \frac{492}{4835}*\text{P.Gir} + \frac{307}{4835}*\text{PLld} + \frac{519}{4835}*\text{P.Trg}$$
$$\textbf{Perfil fila mitjà}{=}0.727*P.Bcn + 0.102*P.Gir + 0.063*P.Lld + 0.107*\text{P.Trg}$$

2.4 Perfils columna

Taula de perfils columna (Taula 4)
Cada valor és dividit pel total de la seva respectiva columna. Els totals de les files dividit pel total dels casos l'anomenem **Perfil columna mitjà o Massa de les files**. Al igual que en el cas anterior, la suma de les columnes és 1 i es poden considerar com una ponderació de les files. Fixem-nos que en la **fila del total tots són 1**.

$$p_{.j} = \frac{n_{ij}}{n_{.j}}; \quad n_{.j} \text{ és el marginal de la columna } j \tag{3}$$

Coodenades perfils columna
Cada perfil columna es pot escriure com una combinació lineal de les files. Les coordenades dels perfils columna són una mitjana ponderada de les files i.e **Massa de les files**.
Per exemple:

$$\text{P.Bcn}{=}\frac{3394}{3517}*\text{Barcelona} + \frac{36}{3517}*\text{Girona} + \frac{49}{3517}*\text{Lleida} + \frac{38}{3517}*\text{Tarragona}$$
$$\text{Bcn}{=}0.965*\text{Barcelona} + 0.01*\text{Girona} + 0.014*\text{Lleida} + 0.011*\text{Tarragona}$$

	Bcn	Gir	Lld	Trg	Perfil columna mitjà o Massa fila
Barcelona	0.965	0.181	0.104	0.171	0.745
Girona	0.010	0.805	0.007	0.004	0.090
Lleida	0.014	0.008	0.876	0.019	0.069
Tarragona	0.011	0.006	0.013	0.805	0.096
Total	1.000	1.000	1.000	1.000	1.000

Taula 4: Perfils columna.

3 Distància χ^2 i inèrcia

En l'AC, la variabilitat de la taula freqüències es mesura amb la **inèrcia**, un concepte molt relacionat amb la distància χ^2.
Es definirà primer la **inèrcia** a partir de l'estadístic χ^2 i l'escriurem en funció del perfil fila, perfil fila mitjà (massa columna) i del perfil columna mitjà (massa fila). Després aplicarem la formulació a les dades de la taula de **Població** (Taula 1). Finalment es definirà la **distància** χ^2

3.1 Inèrcia total

Definim **Inèrcia o Inèrcia total**: $\Phi^2 = \dfrac{\chi^2}{n}$
És una mesura de la variància total de la taula de contingència.
Recordem que **l'estadístic** χ^2 es defineix com:

$$\chi^2 = \sum \frac{(Observat - esperat)^2}{esperat} \tag{4}$$

Els valors esperats de cada cel·la:

$$n_{\text{esperat}} = n_{i.} \frac{n_{.j}}{n}$$

Per tant tenim que χ^2 és:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - n_{i.} \dfrac{n_{.j}}{n}\right)^2}{n_{i.} \dfrac{n_{.j}}{n}} \tag{5}$$

Expressem l'estadístic χ^2 en funció de **perfils fila** i **perfils fila mitjà (massa columna)** simplement **dividint** el numerador i denominador pel **quadrat del marginal fila** corresponent.

Perfil fila

$$p_{i.} = \frac{n_{ij}}{n_{i.}}$$

Perfil fila mitjà

$$c_j = \frac{n_{.j}}{n}\text{; n és el nombre total de casos.}$$

Per tant **dividint** numerador i denominador pel **quadrat del marginal fila** tenim:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(\dfrac{n_{ij}}{n_{i.}} - \dfrac{n_{i.}\dfrac{n_{.j}}{n}}{n_{i.}}\right)^2}{\dfrac{n_{i.}\dfrac{n_{.j}}{n}}{n_{i.}^2}} = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{i.}\frac{\left(\dfrac{n_{ij}}{n_{i.}} - \dfrac{n_{.j}}{n}\right)^2}{\dfrac{n_{.j}}{n}}$$

Per últim **dividim per** n a cada banda i expressem la inèrcia en funció de la **massa de les files**.

Massa de les files

$$r_i = \frac{n_{i.}}{n}$$

$$\frac{\chi^2}{n} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{n_{i.}}{n}\frac{\left(\dfrac{n_{ij}}{n_{i.}} - \dfrac{n_{.j}}{n}\right)^2}{\dfrac{n_{.j}}{n}} = \sum_{i=1}^{I}\sum_{j=1}^{J} r_i\frac{(p_{i.} - c_j)^2}{c_j} = \Phi^2 \tag{6}$$

3.2 Càlcul de la inèrcia total amb les dades

Calculem l'estadístic χ^2 per les dades de la taula 1 Si no hi hagués diferència entre els llocs de naixement, els perfils de totes les files haurien de ser similars a:

$$\textbf{perfil fila mitjà} = c_j = \frac{n_{.j}}{n} = \frac{3517}{4835}, \frac{492}{4835}, \frac{307}{4835}, \frac{519}{4835} = (0.727,\ 0.102,\ 0.063,\ 0.107)$$

Així doncs els **valors esperats de cada cel·la** (suposant que no hi ha diferències pel lloc de naixement) és el **total** de cada **fila** ($n_{i.}$) multiplicat **pel seu perfil fila mitjà** ($\frac{n_{.j}}{n}$). (Taula 5)

$$n_{\text{esperat}} = n_{i.}\frac{n_{.j}}{n}$$

Exemple, els valors esperats pels nascuts a Barcelona serien:

(3604*0.727, 89*0.102, 32*0.063, 89*0.107)=(2468.81, 9.056, 2.032, 9.553)

Valors esperats

	Bcn	Gir	Lld	Trg
Barcelona	2468.810	9.056	2.032	9.553
Girona	26.187	40.296	0.127	0.215
Lleida	35.643	0.407	17.080	1.073
Tarragona	27.641	0.305	0.254	44.869

Taula 5: Valors esperats

Ara ja podem **calcular l'estadístic** χ^2:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(n_{ij} - n_{i.}\dfrac{n_{.j}}{n}\right)^2}{n_{i.}\dfrac{n_{.j}}{n}}$$

$$\chi^2 = \frac{(3394-2468.810)^2}{2486.810} + ... + \frac{(89-9.553)^2}{9.553} + ...$$
$$+\frac{(38-27.641)^2}{27.641} + ... + \frac{(418-44.869)^2}{44.869}$$

L'estadístic χ^2 ho podem expressar també **en funció dels perfils fila** observats i dels **perfils fila mitjà (massa columna)** on cada terme queda multiplicat pel marginal fila. Només cal **dividir** el numerador i el denominador pel **quadrat del marginal fila** corresponent.

L'estadístic χ^2 queda:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(\dfrac{n_{ij}}{n_{i.}} - \dfrac{n_{i.}\dfrac{n_{.j}}{n}}{n_{i.}}\right)^2}{\dfrac{n_{i.}\dfrac{n_{.j}}{n}}{n_{i.}^2}} = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{i.}\frac{\left(\dfrac{n_{ij}}{n_{i.}} - \dfrac{n_{.j}}{n}\right)^2}{\dfrac{n_{.j}}{n}} \tag{7}$$

Recordem que els totals de les files són: (3604, 436, 332, 463)

$$\chi^2 = \frac{(3394/3604-2468.810/3604)^2}{2486.810/3604^2} + ... + \frac{(89/3604-9.553/3604)^2}{9.553/3604^2} + ...$$
$$+\frac{(38/463-27.641/463)^2}{27.641/463^2} + ... + \frac{(418/463-44.869/463)^2}{44.869/463^2}$$

Recordem que els Pefils fila mitjà són: (0.727, 0.102, 0.063, 0.107)
I tenim doncs:

$$\chi^2 = 3604\frac{(3394/3604-0.727)^2}{0.727} + ... + 3604\frac{(89/3604-0.107)^2}{0.107} + ...$$
$$+463\frac{(38/463-0.727)^2}{0.727} + ... + 463\frac{(418/463-0.107)^2}{0.107}$$

Per últim si **dividim per** nombre total de casos **(n)** a cada banda de l'equació tenim **la massa de les files**

$$\frac{\chi^2}{n} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{n_{i.}}{n}\frac{\left(\dfrac{n_{ij}}{n_{i.}} - \dfrac{n_{.j}}{n}\right)^2}{\dfrac{n_{.j}}{n}} = \sum_{i=1}^{I}\sum_{j=1}^{J} r_i\frac{(p_{i.}-c_j)^2}{c_j} = \Phi^2 \tag{8}$$

Recordem que la massa de les files són (0.745, 0.09, 0.069, 0.096)

$$
\begin{aligned}
\frac{\chi^2}{4835} &= \frac{3604}{4835}\frac{(3394/3604-0.727)^2}{0.727}+...+\frac{3604}{4835}\frac{(89/3604-0.107)^2}{0.107}+...\\
&+\frac{463}{4835}\frac{(38/463-0.727)^2}{0.727}+...+\frac{463}{4835}\frac{(418/463-0.107)^2}{0.107}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\chi^2}{4835} &= 0.745\frac{(3394/3604-0.727)^2}{0.727}+...+0.745\frac{(89/3604-0.107)^2}{0.107}+...\\
&+0.096\frac{(38/463-0.727)^2}{0.727}+...+0.096\frac{(418/463-0.107)^2}{0.107}
\end{aligned}
$$

Tenim doncs que la **Inèrcia o Inércia total** és: $\Phi^2 = \dfrac{\chi^2}{n}$
(Codi R pels càlculs de l'estadístic χ^2 i de la inèrcia total a l'apèndix A.1)

Podem calcular la inèrcia total per les dades de la població per províncies i lloc de naixement (Taula 1)

```
## Inèrcia total per les dades de la població per províncies i lloc de naixement
## [1] 2.091
```

La inèrcia serà alta quan els perfils fila presenten grans desviacions respecte la seva mitjana i serà baixa quan es trobin a prop.

3.3 Distància χ^2 i transformació dels perfils

Definim **Distància de** χ^2 entre el "i" i el "i' éssim" perfil fila:

$$\|f_i - f_{i'}\|_c = \sqrt{\sum_{j=1} \frac{(p_{i.} - p_{i'.})^2}{c_j}} \tag{9}$$

On

$$f_{i.} = \frac{n_{ij}}{n_{i.}} \textbf{ Perfil fila i}$$

$$c_j = \frac{n_{.j}}{n} \text{ és } \textbf{Massa de la columna j}$$

Per fer representacions tridimensionals o sobre el pla, cal reduir la dimensionalitat dels punts. Aquesta reducció implica una pèrdua de la informació. Volem restringir aquesta pèrdua per mantenir la màxima informació. Voldríem poder visualitzar perfils en un espai de poques dimensions. Sobre aquest subespai projectarem els perfils fila i els vèrtex de l'espai de perfils.

L'AC és fer un anàlisi de components principals (ACP), on les distàncies de les projeccions són distàncies χ^2 amb els perfils ponderats per la inversa de l'arrel de la seva massa.

Volem projectar els perfils de manera que les distàncies χ^2 projectades sobre el subespai siguin màximes. Ens proposem doncs a **transformar els vectors perfils** perquè a l'hora de calcular les distàncies Euclídies de projeccions, les distàncies obtingudes siguin directament les distàncies χ^2. A més com en l'ACP centrerem les dades i les centrarem respecte el perfil mig.

Transformació dels vectors perfils

Si ponderem el perfil fila $f_{i.}$ amb w_j:

$$w_j = \frac{1}{\sqrt{c_j}} = \frac{1}{\sqrt{\frac{n.j}{n}}}$$

Quan calculem la distància de la manera habitual (Euclídia), la distància que obtenim serà la distància χ^2

Components de la matriu X. Vectors perfils transformats

$$x_{ij} = \frac{p_{ij} - (p_{i.}p_{.j})}{\sqrt{p_{i.}p_{.j}}} = \frac{\frac{n_{ij}}{n_{.j}} - (\frac{n_{ij}}{n_{i.}}\frac{n_{ij}}{n_{.j}})}{\sqrt{\frac{n_{ij}}{n_{i.}}\frac{n_{ij}}{n_{.j}}}} \tag{10}$$

3.4 Matriu d'inèrcia S

Un cop tenim els vectors dels perfils transformats, i.e la matriu X, la **matriu d'inèrcia S** (per les files) s'obté:

$$S = X^T X = D_r^{-1/2}(P - rc^T)D_c^{-1/2} = \tag{11}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{\frac{n_{1.}}{n}}} & ... & ... \\ ... & ... & ... \\ ... & ... & \frac{1}{\sqrt{\frac{n_{I.}}{n}}} \end{pmatrix} \left[\begin{pmatrix} \frac{n_{11}}{n} & ... & \frac{n_{1J}}{n} \\ ... & ... & ... \\ \frac{n_{I1}}{n} & ... & \frac{n_{IJ}}{n} \end{pmatrix} - \begin{pmatrix} \frac{n_{i.}}{n} \\ . \\ . \\ . \\ \frac{n_{I.}}{n} \end{pmatrix} \begin{pmatrix} \frac{n_{.1}}{n} & ... & \frac{n_{.J}}{n} \end{pmatrix} \right] \begin{pmatrix} \frac{1}{\sqrt{\frac{n_{.1}}{n}}} & ... & ... \\ ... & ... & ... \\ ... & ... & \frac{1}{\sqrt{\frac{n_{.J}}{n}}} \end{pmatrix}$$

Matriu de dades: N, dimensió (IxJ)

Matriu de correspondències: $P = \frac{1}{n}N$

Massa de files $\vec{r} = \left(\frac{n_{1.}}{n}, ..., \frac{n_{I.}}{n}\right)$

Massa de columnes $\vec{c} = \left(\frac{n_{.1}}{n}, ..., \frac{n_{.J}}{n}\right)$

Matriu diagonal de $\vec{r}$: $D_r = diag(\vec{r})$

Matriu diagonal de $\vec{c}$: $D_c = diag(\vec{c})$

A partir de la matriu S farem la **descompsicó dels valors singular** (SVD) i obtenim les inèrcies i les coordenades dels eixos principals.

3.5 Descomposició en valors singulars (SVD)

Sigui:
$\tilde{X}_{(n\times p)}$ matriu de rang k
$\tilde{X}_{(n\times p)} = U_{(n\times k)}\Lambda_{(k\times k)}\ T'_{(k\times p)}$ tal que $U'U = T'T = Id$

$$\Lambda_{(k\times k)} = \begin{pmatrix} \lambda_1^* & \dots \\ \dots & \lambda_k^* \end{pmatrix}$$

Llavors $\tilde{X}'_{(p\times n)}\tilde{X}_{(n\times p)}$, $\tilde{X}_{(n\times p)}\tilde{X}'_{(p\times n)}$ matrius simètriques i tenen els k mateixos valors singulars $(\lambda_1, ..., \lambda_k)$ diferents de zero i a més els valors singulars de $\Lambda_{(k\times k)}$ es poden expressar com:

$$\Lambda_{(k\times k)} = \begin{pmatrix} \lambda_1^* & \dots \\ \dots & \lambda_k^* \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} & \dots \\ \dots & \sqrt{\lambda_k} \end{pmatrix}$$

3.6 ACP i AC

L'AC és una adaptació de l'ACP per les taules de contingències. En l'ACP clàssic s'aplica el SVD sobre la matriu centrada de dades i en AC s'aplica sobre la taula de contingència modificada, i.e sobre la matriu d'inèrcia que hem anomenat S.
La modificació sobre la taula de contingència és la ponderació dels punts respecte la inversa de l'arrel de les masses files i columnes, perquè a l'hora de calcular les distàncies Euclídees amb els punts ponderats, tinguem la distància χ^2.

4 Algoritme AC

L'algoritme AC ens permetrà obtenir les **inèrcies principals i les coordenades principals i estàndards** de les files i les columnes de les taules de contingència. **A partir de la matriu d'inèrcia S, es realitza una SVD** que ens permetrà obtenir aquestes coordenades.

L'algoritme AC es pot resumir en **7 pasos** que es veuran després pas a pas amb les dades "smoke" del paquet de R "ca" (Taula 6)

Els 7 pasos de l'algoritme AC són:
Pas1. Càlcul de la **matiu S d'inèrcia**.

$$S = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$$

Pas2. Descomposició en valors singulars **(SVD) de la matriu S**.

$$S = UD_\alpha V^T \quad \text{on } U^TU = V^TV = I$$
D_α és la matriu diagonal dels valors singulars

$$D_\alpha = \begin{pmatrix} \sqrt{\lambda_1} & \dots \\ \dots & \sqrt{\lambda_k} \end{pmatrix}$$

A partir de les matriu ortogonals U i V s'obtenen les coordenades estàndards per les files (**rsc**) i les columnes (**csc**).

Pas3 del AC. Coordenades estàndards de les files **(rsc)**.

$$\text{rsc}=D_r^{-1/2}U$$

Pas4 del AC. Coordenades estàndards de les columnes **(csc)**.

$$\text{csc}=D_c^{-1/2}V$$

Amb les coordenades estàndards rsc i csc multiplicats per la matriu diagonal de valors singulars obtenim les coordenades principals de les files (**rpc**) i columnes (**cpc**).

Pas5 del AC. Coordenades principals de les files **(rpc)**.

$$\text{rpc}=D_r^{-1/2}UD_\alpha=\text{rsc}D_\alpha$$

Pas6 del AC. Coordenades principals de les columnes **(cpc)**.

$$\text{cpc}=D_c^{-1/2}VD_\alpha=\text{csc}D_\alpha$$

Pas7 del AC. **Inèrcies principals** (λ_k)
Les inèrcies principals són els valors λ_1,...,λ_k de la matriu D_α

$$\lambda_k \quad k = 1, 2...K \quad \text{on } K = min(I - 1, J - 1)$$

S'ha programat un **codi R ("MyAC")** que ens permetrà realitzar l'AC sobre una taula de contingència. La funció programada té com a arguments, la taula de contingència (format "data.frame" o "matrix") i en la seva sortida obtindrem un llistat amb totes les matrius necessàries per trobar la matriu S d'inèrcia, la seva descomposició de valors singulars, la descomposició de les inèrcies i les coordenades principals i estàndards de les files i columnes. (Codi R en l'apèndix A.2)

4.1 Algoritme AC amb les dades "smoke"

Considerem les dades "smoke" del paquet de R "ca" (taula 6) i apliquem pas a pas l'algoritme AC.
Calculem les matrius necessàries per l'algoritme AC:

```
## Matriu de correspondències P
##                        No fumador  Ocasional        Mig  Compulsiu
## DS: Directius Seniors 0.02072539 0.01036269 0.01554404 0.01036269
## DJ: Directius Joves   0.02072539 0.01554404 0.03626943 0.02072539
```

	No fumador	Ocasional	Mig	Compulsiu
DS: Directius Seniors	4	2	3	2
DJ: Directius Joves	4	3	7	4
ES: Empleats Seniors	25	10	12	4
EJ: Empleats Joves	18	24	33	13
SC: Secretàries	10	6	7	2

Taula 6: Dades smoke

```
## ES: Empleats Seniors  0.12953368 0.05181347 0.06217617 0.02072539
## EJ: Empleats Joves    0.09326425 0.12435233 0.17098446 0.06735751
## SC: Secretàries       0.05181347 0.03108808 0.03626943 0.01036269
##
##  Massa de les files r
## DS: Directius Seniors   DJ: Directius Joves  ES: Empleats Seniors
##            0.05699482            0.09326425            0.26424870
##    EJ: Empleats Joves       SC: Secretàries
##            0.45595855            0.12953368
##
##  Massa de les columnes c
## No fumador  Ocasional        Mig  Compulsiu
##  0.3160622  0.2331606  0.3212435  0.1295337
##
##  Matriu diagonal de massa files Dr
##            [,1]       [,2]      [,3]      [,4]      [,5]
## [1,] 0.05699482 0.00000000 0.0000000 0.0000000 0.0000000
## [2,] 0.00000000 0.09326425 0.0000000 0.0000000 0.0000000
## [3,] 0.00000000 0.00000000 0.2642487 0.0000000 0.0000000
## [4,] 0.00000000 0.00000000 0.0000000 0.4559585 0.0000000
## [5,] 0.00000000 0.00000000 0.0000000 0.0000000 0.1295337
##
## Matriu diagonal de massa columnes Dc
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.3160622 0.0000000 0.0000000 0.0000000
## [2,] 0.0000000 0.2331606 0.0000000 0.0000000
## [3,] 0.0000000 0.0000000 0.3212435 0.0000000
## [4,] 0.0000000 0.0000000 0.0000000 0.1295337
```

Pas1 del AC. Càlcul de la **matiu S d'inèrcia** (o de residus estandarditzats).

$$S = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$$

```
## Matriu S
##                        No fumador    Ocasional          Mig   Compulsiu
## DS: Directius Seniors  0.02020239 -0.025384382 -0.02043562  0.03468162
```

```
## DJ: Directius Joves   -0.05097522 -0.042054470  0.03644840  0.07864884
## ES: Empleats Seniors   0.15922216 -0.039477006 -0.07795287 -0.07298869
## EJ: Empleats Joves    -0.13394189  0.055330472  0.06404368  0.03413421
## SC: Secretàries        0.05373569  0.005097772 -0.02618966 -0.04953368
```

Pas2 del AC. Descomposició en valors singulars **(SVD) de la matriu S**.

$$S = UD_\alpha V^T \quad \text{on } U^TU = V^TV = I$$
$$D_\alpha \text{ és la matriu diagonal dels valors singulars}$$

```
## Descomposició en valors singulars de la matriu S
## $d
## [1] 2.734211e-01 1.000859e-01 2.033652e-02 5.600256e-17
##
## $u
##             [,1]        [,2]       [,3]        [,4]
## [1,] -0.05742524 -0.46212293  0.8332653 -0.08273461
## [2,]  0.28923816 -0.74239515 -0.5061482 -0.16781906
## [3,] -0.71554563 -0.05475038 -0.1303234 -0.68230206
## [4,]  0.57530335  0.38957951  0.1097504 -0.67236949
## [5,] -0.26469630  0.28376408 -0.1430158  0.21765779
##
## $v
##            [,1]        [,2]       [,3]      [,4]
## [1,] -0.8087001 -0.17127755 -0.0246170 0.5621941
## [2,]  0.1756411  0.68056865  0.5223178 0.4828671
## [3,]  0.4069601  0.04167443 -0.7151246 0.5667835
## [4,]  0.3867013 -0.71116353  0.4638695 0.3599079
```

Pas3 del AC. Coordenades estàndards de les files **(rsc)**.

$$\text{rsc}{=}D_r^{-1/2}U$$

```
## Coordenades estàndards de les files. (rsc)
##                             Eix1       Eix2       Eix3       Eix4
## DS: Directius Seniors -0.2405388 -1.9357079  3.4903231 -0.3465529
## DJ: Directius Joves    0.9471047 -2.4309584 -1.6573725 -0.5495202
## ES: Empleats Seniors  -1.3919733 -0.1065076 -0.2535221 -1.3273035
## EJ: Empleats Joves     0.8519895  0.5769437  0.1625337 -0.9957385
## SC: Secretàries       -0.7354557  0.7884353 -0.3973677  0.6047597
```

Pas4 del AC. Coordenades estàndards de les columnes **(csc)**.

$$\text{csc}{=}D_c^{-1/2}V$$

```
## Coordenades estàndards de les columnes. (csc)
##                  Eix1        Eix2        Eix3 Eix4
## No fumador -1.4384714 -0.30465911 -0.04378737    1
## Ocasional   0.3637463  1.40943267  1.08170100    1
## Mig         0.7180168  0.07352795 -1.26172451    1
## Compulsiu   1.0744451 -1.97595989  1.28885615    1
```

Pas5 del AC. Coordenades principals de les files **(rpc)**.

$$\text{rpc}=D_r^{-1/2}UD_\alpha=\text{rsc}D_\alpha$$

```
## Coordenades principals de les files (rpc)
##                              Eix1        Eix2         Eix3          Eix4
## DS: Directius Seniors -0.06576838 -0.19373700  0.070981028 -1.940785e-17
## DJ: Directius Joves    0.25895842 -0.24330457 -0.033705190 -3.077454e-17
## ES: Empleats Seniors  -0.38059489 -0.01065991 -0.005155757 -7.433239e-17
## EJ: Empleats Joves     0.23295191  0.05774391  0.003305371 -5.576390e-17
## SC: Secretàries       -0.20108912  0.07891123 -0.008081076  3.386809e-17
```

Pas6 del AC. Coordenades principals de les columnes **(cpc)**.

$$\text{cpc}=D_c^{-1/2}VD_\alpha=\text{csc}D_\alpha$$

```
## Coordenades principals de les columnes (cpc)
##                   Eix1         Eix2          Eix3         Eix4
## No fumador -0.39330845 -0.030492071 -0.0008904827 5.600256e-17
## Ocasional   0.09945592  0.141064289  0.0219980349 5.600256e-17
## Mig         0.19632096  0.007359109 -0.0256590867 5.600256e-17
## Compulsiu   0.29377599 -0.197765656  0.0262108499 5.600256e-17
```

Pas7 del AC. **Inèrcies principals λ_k**

$$\lambda_k \quad k=1,2...K \quad \text{on } K=min(I-1,J-1)$$

```
## Inèrcies principals explicades
##  Dim1  Dim2  Dim3
## 0.075 0.010 0.000
```

Del capítol 3, sabem que la inèrcia total és:

```
## Inèrcies Total
## [1] 0.085
```

La inèrcia total és la suma dels valors singulars de la matriu S, i.e la suma de les inèrcies principals explicades. Podem saber doncs quin és el percentatge d'inèrcia total explicada per cada dimensió.

```
## Percentatge d'inèrcia explicat
##   Dim1   Dim2   Dim3
## 87.756 11.759  0.485
```

5 Simetria entre l'anàlisi de files i de columnes

L'anànilis de correspondència realitza simultàneament un anàlisi per fila com per columna. La inèrcia total dels perfils fila i columna és la mateixa.
Existeix un factor que permet passar de les coordenades dels vèrtex d'un anàlisi a les coordenades dels perfils de l'altre anàlisi: (Codi R en l'apèndix B.1)

$$\text{coordenada del perfil} = \text{coordenades vèrtex} * \sqrt{\text{inèrcia principal(eix principal)}} \tag{12}$$

Així si agafem la primera columna de la matriu de les coordenades principals dels perfils fila i la primera columna de la matriu de les coordenades estàndard, vèrtex fila:

```
## Primera columna de la matriu de les coordenades principals dels perfils fila (rpc)
##                                Eix1
## DS: Directius Seniors -0.06576838
## DJ: Directius Joves    0.25895842
## ES: Empleats Seniors  -0.38059489
## EJ: Empleats Joves     0.23295191
## SC: Secretàries       -0.20108912
##
##  Coordenades estàndard vèrtex fila (rsc1).
##                               Eix1
## DS: Directius Seniors -0.2405388
## DJ: Directius Joves    0.9471047
## ES: Empleats Seniors  -1.3919733
## EJ: Empleats Joves     0.8519895
## SC: Secretàries       -0.7354557
```

Comprovem que efectivament es compleix la relació. (Equació 12).

A partir de les "coordenades vèrtex", obtenim les "coordenades del perfil":

```
## Coordenades del perfil a partr de les coordenades del vèrtex
##                                Eix1
## DS: Directius Seniors -0.06576838
## DJ: Directius Joves    0.25895842
## ES: Empleats Seniors  -0.38059489
## EJ: Empleats Joves     0.23295191
## SC: Secretàries       -0.20108912
```

A partir de les "coordenades del perfil", obtenim les "coordenades vèrtex":

```
## Coordenades vèrtex a partir de les coordenades del perfil
##                              Eix1
## DS: Directius Seniors -0.2405388
## DJ: Directius Joves    0.9471047
## ES: Empleats Seniors  -1.3919733
## EJ: Empleats Joves     0.8519895
## SC: Secretàries       -0.7354557
```

6 Contribucions a la inèrcia

La **inèrcia total** es pot descomposar respecte cada cel·la de la taula de contingències, respecte les files, les columnes i dels eixos principals.
Recordem que la **inèrcia total és de 0.085**

6.1 Descomposició de la inèrcia respecte a cada cel·la

(Codi R a l'apèndix B.2)
La matriu S, escrita en funció de la matriu de freqüències relatives amb la ponderació de la inversa de l'arrel de les masses de files i columnes és:

$$S = D_r^{-1/2}(P - rc^T)D_c^{-1/2}$$

La matriu S de les dades és:

```
## Matriu d'inèrcia S
##                        No fumador    Ocasional         Mig   Compulsiu
## DS: Directius Seniors  0.02020239 -0.025384382 -0.02043562  0.03468162
## DJ: Directius Joves   -0.05097522 -0.042054470  0.03644840  0.07864884
## ES: Empleats Seniors   0.15922216 -0.039477006 -0.07795287 -0.07298869
## EJ: Empleats Joves    -0.13394189  0.055330472  0.06404368  0.03413421
## SC: Secretàries        0.05373569  0.005097772 -0.02618966 -0.04953368
```

La **contribució** de cada **cel·la** a la inèrcia total s'obté de S^2

```
## Contribució de cada cel·la a la inèrcia
##                         No fumador    Ocasional          Mig   Compulsiu
## DS: Directius Seniors 0.0004081367 6.443668e-04 0.0004176144 0.001202814
## DJ: Directius Joves   0.0025984732 1.768578e-03 0.0013284858 0.006185640
## ES: Empleats Seniors  0.0253516953 1.558434e-03 0.0060766503 0.005327349
## EJ: Empleats Joves    0.0179404290 3.061461e-03 0.0041015931 0.001165144
## SC: Secretàries       0.0028875239 2.598727e-05 0.0006858985 0.002453585
```

Els **percentatges** de contribució de cada cel·la:

```
## Percentatge de contribució de cada cel·la
##                        No fumador  Ocasional       Mig Compulsiu
## DS: Directius Seniors  0.4790907 0.75638912 0.4902161  1.411922
## DJ: Directius Joves    3.0502142 2.07604334 1.5594412  7.261004
## ES: Empleats Seniors  29.7590525 1.82936562 7.1330675  6.253502
## EJ: Empleats Joves    21.0593478 3.59369185 4.8146494  1.367703
## SC: Secretàries        3.3895159 0.03050513 0.8051410  2.880138
```

La cel·la que més contribueix són els Empleats Seniors, No fumadors, la seva contribució és del 29.76%

Fixem-nos que les sumes de files o columnes són justament les contribucions a les inèrcies de files i columnes.

6.2 Descomposició de la inèrcia total respecte de les files

(Codi R en l'apèndix: B.3)
La **contribució de les files** a la inèrcia total es calcula sumant les files de S^2

$$\sum_i s_{ij}^2; \text{ on } s_{ij} \text{ són els elements de la matriu } S \tag{13}$$

La **contribució de les files** a la inèrcia total i els seus **percentatges** de contribució:

```
## Contribució de les files a la inèrcia total
##                           Inèrcia % Inèrcia
## DS: Directius Seniors 0.002672932  3.137618
## DJ: Directius Joves   0.011881177 13.946703
## ES: Empleats Seniors  0.038314129 44.974987
## EJ: Empleats Joves    0.026268627 30.835392
## SC: Secretàries       0.006052995  7.105300
```

El perfil fila que més contribueix amb la inèrcia són ES: Empleats Seniors.
Si fem la suma de les contribucions de les files a la inèrcia recuperem la **inèrcia total 0.08519**

6.3 Descomposició de la inèrcia total respecte de les columnes

(Codi R en l'apèndix B.4)
La **contribució de les columnes** a la inèrcia total es calcula sumant les columnes de S^2

$$\sum_j s_{ij}^2; \text{ on } s_{ij} \text{ són els elements de la matriu } S \tag{14}$$

La **contribució de les columnes** a la inèrcia total i els seus **percentatges** de contribució:

```
## Contribució de les columnes a la inèrcia total
##             No fumador   Ocasional         Mig   Compulsiu
## Inèrcia     0.04918626 0.007058828  0.01261024  0.01633453
## % Inèrcia 57.73722102 8.285995064 14.80251519 19.17426872
```

La columna que més contribueix a la inèrcia són: No fumador.
Si fem la suma de les contribucions de les columnes a la inèrcia recuperem la inèrcia total 0.08519

6.4 Descomposició de la inèrcia respecte els eixos principals

(Codi R a l'apèndix B.5)
És la **contribució de cada fila a la inèrcia de l'eix principal**. Es calcula a partir de les masses de files i les coordenades principals de les files.
Recordem del pas 3 de l'algoritme de AC tenim que les coordenades principals de les files (rpc):

$$\text{rpc}{=}D_r^{(-1/2)}U$$

La **descomposició** de la inèrcia de l'eix principal **es calcula**:

$$\text{InerciaEix} = D_r * (rpc)^2 \tag{15}$$

Les masses de les files:

```
## Masses de les files
## DS: Directius Seniors   DJ: Directius Joves  ES: Empleats Seniors
##            0.05699482            0.09326425            0.26424870
##    EJ: Empleats Joves       SC: Secretàries
##            0.45595855            0.12953368
```

Les coordenades principals de les files (rpc):

```
## Coordenades principals de les files (rpc)
##                              Eix1        Eix2         Eix3          Eix4
## DS: Directius Seniors -0.06576838 -0.19373700  0.070981028 -1.940785e-17
## DJ: Directius Joves    0.25895842 -0.24330457 -0.033705190 -3.077454e-17
## ES: Empleats Seniors  -0.38059489 -0.01065991 -0.005155757 -7.433239e-17
## EJ: Empleats Joves     0.23295191  0.05774391  0.003305371 -5.576390e-17
## SC: Secretàries       -0.20108912  0.07891123 -0.008081076  3.386809e-17
```

La **descomposició de la inèrcia dels eixos principals** és:

```
## Descomposició de la inèrcia de l'eix principal:
##                              Eix1         Eix2         Eix3
## DS: Directius Seniors 0.000246530 2.139245e-03 2.871574e-04
## DJ: Directius Joves   0.006254251 5.520975e-03 1.059519e-04
## ES: Empleats Seniors  0.038277077 3.002754e-05 7.024216e-06
## EJ: Empleats Joves    0.024743316 1.520329e-03 4.981565e-06
## SC: Secretàries       0.005237932 8.066039e-04 8.459041e-06
```

El **percentatge de contribució** de les files als eixos principals:

```
## Percentatges de contribució a la inèrcia de l'eix principal
##                             Eix1       Eix2        Eix3
## DS: Directius Seniors  0.2893889 2.51114983 0.337079269
## DJ: Directius Joves    7.3415433 6.48078836 0.124371491
## ES: Empleats Seniors  44.9314940 0.03524778 0.008245366
## EJ: Empleats Joves    29.0449077 1.78463661 0.005847603
## SC: Secretàries        6.1485392 0.94683091 0.009929633
```

Al primer eix, la fila que més contribueix són els Empleats Seniors seguits dels Empleats Joves.

Si es realitza la **suma per columnes de les contribucions**, obtenim la **inèrcia explicada** de cada eix (pas 7 de l'algoritme AC):

```
## Suma de les contribucions de les files en cada eix i.e la inèrcia explicada
##         Eix1         Eix2         Eix3
## 0.0747591059 0.0100171805 0.0004135741
```

7 Representació bidimensional

La majoria de vegades, les representacions d'AC seran bidimensionals, degut a que estem més acostumats a treballar en dues dimensions. Habitualment el primer eix principal serà l'eix de les abscisses x i el segon eix principal serà l'eix de les ordenades y.

7.1 Mapa asimètric

Representem el **mapa asimètric** del AC (Gràfic 1) (Codi R en l'apèndix B.6 tant a partir del package "ca" de R com un codi propi implementat.) Les **files** es representen en **coordenades principals** i les **columnes en coordenades estàndards**.

El mapa asimètric és una **representació conjunta dels perfils i dels vèrtex**. Es representen les **files en coordenades principals** i les **columnes en coordenades estàndards** o viceversa, i.e. si estiguéssim interessats en l'anàlisi de les columnes, representaríem les columnes en coordenades principals i les files en coordenades estàndards.

La **inèrcia del primer eix** és de: 0.07476
El **percentatge d'inèrcia explicada** és de: 87.756%

La **inèrcia del segon eix** és de: 0.01002
El **percentatge d'inèrcia explicada** és de: 11.759%

La **inèrcia total** és de: 0.08519

Per tant, el percentatge de la inèrcia acumulada en la representació bidimensional (les sumes de les inèrcies dels dos eixos) és de: 99.515% de la inèrcia total. Només s'ha perdut 0.485% de la

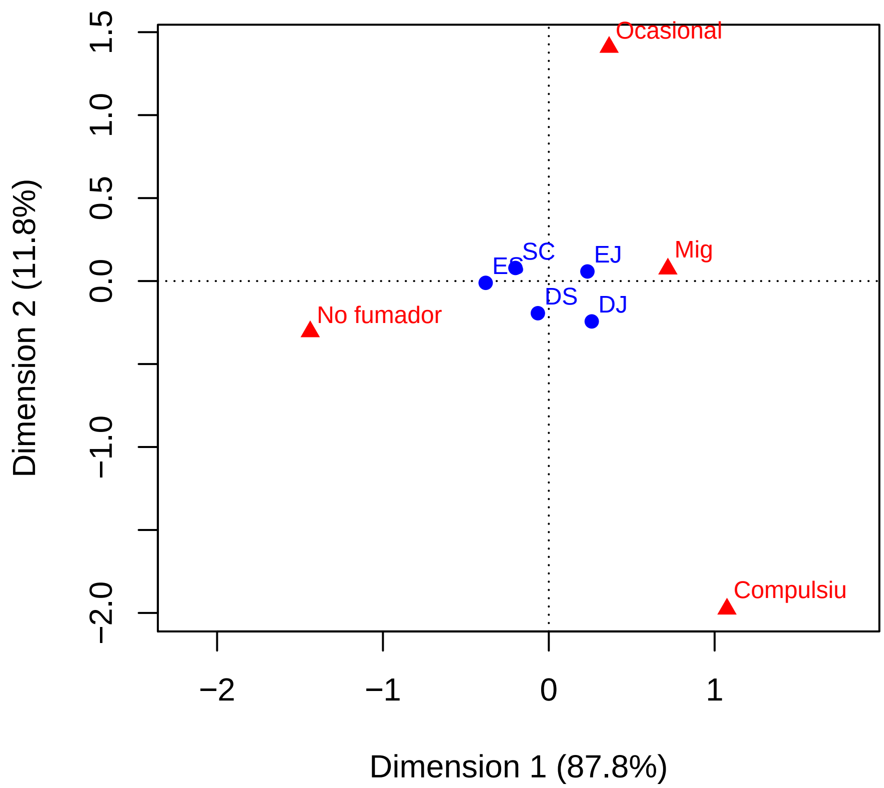

Figura 1: Mapa asimètric

inèrcia dels perfils.

Centre del gràfic

El centre del gràfic és el perfil mig, podem considerar les desviacions dels diferents empleats en diferents direccions a partir del perfil mig.

Eix abscisses

Les desviacions més grans es produeixen en l'eix de les x, primer eix principal (d'esquerra a dreta) entre els que sí són fumadors i dels que no ho són.

Eix ordenades

Si mirem l'eix vertical, separa els tres grups de fumadors. Com indica el percentatge d'**inèrcia** explicada per l'**eix vertical, molt inferior a la inèrcia de l'eix horitzontal**, els perfils no es diferencien tant, no hi ha tanta desviació. Es pot veure que el perfil "EJ" té més fumadors "ocasionals" que compulsius en comparació amb els del perfil "DJ", ja que es troba més a prop del vèrtex "ocasionals".

Distàncies

Distàncies entre els perfils i els vèrtex En els mapes asimètrics, podem interpretar les distàncies dels perfils respecte als vèrtex com que estan més relacionats els perfils amb els vèrtex més propers. Així per exemple, els perfil fila "EJ" i "DJ" respecte el vèrtex fumadors ocasionals, els "EJ" que són més a prop tenen més fumadors ocasionals que "DJ".

Eixos anidats

Els mapes en els AC tenen els eixos **anidats**, i.e la representació óptima en una determinada dimensió conté totes les representacions óptimes de menor dimensió.

7.2 Mapa simètric

Representem el **mapa simètric** del AC (Gràfic 2) (Codi R en l'apèndix B.7) El mapa simètric, representa **en el mateix mapa els perfils fila i columna en coordenades principals**. És el solapament de dos mapes diferents! Les distàncies entre les files són aproximadament distàncies χ^2 i les distàncies entre les columnes també són distàncies χ^2. En aquest tipus de mapes no està definit les distàncies entre files i columnes. **No és possible** deduir a partir de la proximitat de perfil - vèrtex la seva associació.

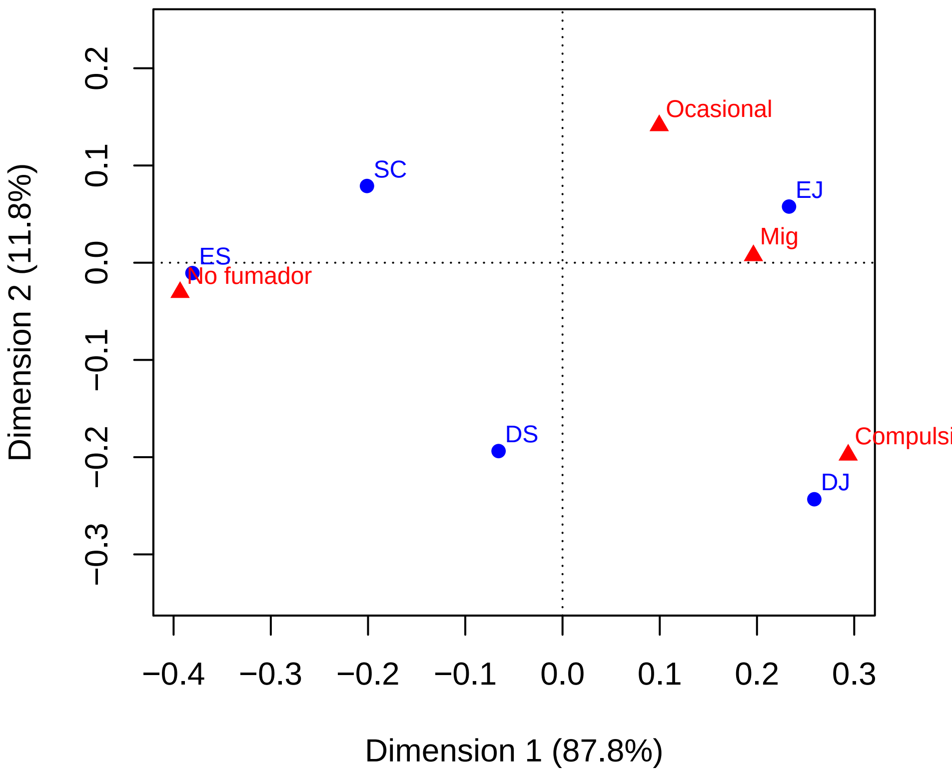

Figura 2: Mapa simètric

8 Anàlisi de taules quadrades

L'anàlisi realitzat fins ara era aplicable a qualsevol dimensió de taules de freqüències. A partir d'ara, estudiem amb detall l'AC de les taules quadrades. Considerem les taules de freqüències quadrades, que generalment les files i les columnes fan referència als mateixos objectes en dues circumstàncies diferents. Habitualment es caracteritzen perquè tenen valors elevats a la diagonal, indicant una forta associació que fàcilment emmascara les associacions més sutils fora de la diagonal que no queden reflectides en els eixos principals.
Per poder aplicar **AC a les taules quadrades**:

1. **Anàlisi de la part simètrica de la taula (amb la diagonal)**. És la part que conté la major part de la inèrcia.

2. **Anàlisi de la part antisimètrica**. Aquesta part és la que mostra la magnitud i el sentit de flux entre files i columnes i viceversa.

8.1 Descomposició de la taula quadrada de freqüències (N)

Podem descomposar la taula de dades com:
1. **Part simètrica**: Flux mitjà entre files i columnes (S)
2. **Part antisimètrica**: Flux diferencial (T)

$$S = \frac{1}{2}(N + N^T)$$
$$T = \frac{1}{2}(N - N^T)$$

$$N = \frac{1}{2}(N + N^T) + \frac{1}{2}(N - N^T) = S + T \tag{16}$$

Exemple Toy

$$\begin{pmatrix} 1 & 3 \\ 7 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 5 & 2 \end{pmatrix} + \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}$$

La **matriu simètrica (S)** tindrà la mateixa diagonal que la matriu original (N) i la **matriu antisimètrica (T)** tindrà zeros en la diagonal i els seus elements oposats situats fora de la diagonal tindran els signes canviats.
Atenció! Aquí la matriu simètrica l'anomenem S, no confondre amb la matriu d'inèrcia!

Considerem les dades de **Karl Pearson** referent a la mobilitat social (Taula 7). Són unes dades de fa més de 100 anys que van donar peu a l'AC de les taules quadrades amb **gran pes en la diagonal**. Es tracta de les **professions dels pares i dels fills**. En les files tenim 14 professions dels pares i a les columnes hi ha les professions dels seus fills. Donat que molts fills segueixen la professió dels pares, els valors de la diagonal són elevats, tot i que hi ha asimetries destacables. En podem veure per exemple en el nombre de pares "militars" (50) mentre que el nombre de fills militars és de 84. El **flux** de fills cap a la professió "militar" provenen majoritàriament de pares "propietaris rurals" (17) i de pares "comerciants" (12). També podem veure el flux de sortida cap a altres professions els fills de pares "comerciants" (106) que han escollit professions com "artistes", "teòlegs" i "Docents i científics".

	Mil	Artist	Mes	Art	Teo	Agr	Prurals	Adv	Esc	Com	Met	Mar	Pol	Docs	Total
Militars	28	0	4	0	0	0	1	3	3	0	3	1	5	2	50
Artistes	2	51	1	1	2	0	0	1	2	0	0	0	1	1	62
Mestres	6	5	7	0	9	1	3	6	4	2	1	1	2	7	54
Artesans	0	12	0	6	5	0	0	1	7	1	2	0	0	10	44
Teolegs	5	5	2	1	54	0	0	6	9	4	12	3	1	13	115
Agricultors	0	2	3	0	3	0	0	1	4	1	4	2	1	5	26
Propietaris rurals	17	1	4	0	14	0	6	11	4	1	3	3	17	7	88
Adv	3	5	6	0	6	0	2	18	13	1	1	1	8	5	69
Escriptors	0	1	1	0	4	0	0	1	4	0	2	1	1	4	19
Comerciants	12	16	4	1	15	0	0	5	13	11	6	1	7	15	106
Metges	0	4	2	0	1	0	0	0	3	0	20	0	5	6	41
Marins	1	3	1	0	0	0	1	0	1	1	1	6	2	1	18
Polítics	5	0	2	0	3	0	1	8	1	2	2	3	23	1	51
Docents i cientifics	5	3	0	2	6	0	1	3	1	0	0	1	1	9	32
Total	84	108	37	11	122	1	15	64	69	24	57	23	74	86	775

Taula 7: Files són professions dels pares i columnes són professions dels fills

Estudiem les **inèrcies** de la matriu sense descomposar. Volem comprovar que efectivament la diagonal de la matriu de dades porta un percentatge alt de la inèrcia total.
La **inèrcia total** de les dades és:

```
## Inèrcia total
## [1] 1.297
```

Mirem ara la **inèrcia** a la qual contribueix cadascuna de les **cel·les** i en particular mirem què sumen les inèrcies de les cel·les que estan en la diagonal:

```
## Els 7 primers casos de files i columnes
##                       Mil Artist    Mes    Art    Teo    Agr Prurals
## Militars           0.1214 0.0090 0.0014 0.0009 0.0102 0.0001  0.0000
## Artistes           0.0043 0.2680 0.0017 0.0000 0.0080 0.0001  0.0015
## Mestres            0.0000 0.0011 0.0098 0.0010 0.0000 0.0160  0.0047
## Artesans           0.0062 0.0072 0.0027 0.0597 0.0007 0.0001  0.0011
## Teolegs            0.0058 0.0098 0.0029 0.0003 0.0918 0.0002  0.0029
## Agricultors        0.0036 0.0009 0.0032 0.0005 0.0004 0.0000  0.0006
## Propietaris rurals 0.0075 0.0133 0.0000 0.0016 0.0000 0.0001  0.0140
##
##
##  La resta de casos de files i columnes
##                         Adv    Esc    Com    Met    Mar    Pol   Docs
## Adv                  0.0343 0.0099 0.0008 0.0042 0.0007 0.0004 0.0012
```

```
## Escriptors             0.0003 0.0041 0.0008 0.0003 0.0004 0.0005 0.0022
## Comerciants            0.0021 0.0017 0.0234 0.0005 0.0019 0.0012 0.0011
## Metges                 0.0044 0.0001 0.0016 0.1234 0.0016 0.0004 0.0006
## Marins                 0.0019 0.0003 0.0005 0.0001 0.0722 0.0001 0.0006
## Polítics               0.0044 0.0036 0.0001 0.0011 0.0019 0.0871 0.0049
## Docents i cientifics   0.0001 0.0015 0.0013 0.0030 0.0000 0.0018 0.0108
```

Les **inèrcies de la diagonal** sumen:

```
## [1] 0.92
```

És a dir que **la diagonal explica un 70.93% de la inèrcia total**

La inèrcia la podem descompondre de la següent manera:

$$\textbf{Inèrcia total} = \textbf{Inèrcia diagonal} + \textbf{Inèrcia fora de la diagonal} \tag{17}$$

Per tant tenim:

1.297=0.92+0.377
100%=70.911%+29.089%

Efectivament la diagonal de la nostra matriu quadrada porta molta càrrega d'informació.

Descomponem ara les nostres dades en dues matrius segons l'equació 16, la **part simètrica (matriu S)**, i la **part antisimètrica matriu T)** i realitzarem l'AC per cada matriu. (Codi R en l'apèndix B.8)

$$N = \frac{1}{2}(N + N^T) + \frac{1}{2}(N - N^T) = S + T$$

$$\begin{pmatrix} 28 & 0 & 4 & 0 & \dots \\ 2 & 51 & 1 & 1 & \dots \\ 6 & 5 & 7 & 0 & \dots \\ 0 & 12 & 0 & 6 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} 28 & 1 & 5 & 0 & \dots \\ 1 & 51 & 3 & 6.5 & \dots \\ 5 & 3 & 7 & 0 & \dots \\ 0 & 6.5 & 0 & 6 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} + \begin{pmatrix} 0 & -1 & -1 & 0 & \dots \\ 1 & 0 & -2 & -5.5 & \dots \\ 1 & 2 & 0 & 0 & \dots \\ 0 & 5.5 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Matriu S

Els primers casos (taula 8)

	Mil	Artist	Mes	Art	Teo	Agr	Prurals	Adv	Esc	Com	Met	Mar	Pol	Docs
Militars	28.00	1.00	5.00	0.00	2.50	0.00	9.00	3.00	1.50	6.00	1.50	1.00	5.00	3.50
Artistes	1.00	51.00	3.00	6.50	3.50	1.00	0.50	3.00	1.50	8.00	2.00	1.50	0.50	2.00
Mestres	5.00	3.00	7.00	0.00	5.50	2.00	3.50	6.00	2.50	3.00	1.50	1.00	2.00	3.50
Artesans	0.00	6.50	0.00	6.00	3.00	0.00	0.00	0.50	3.50	1.00	1.00	0.00	0.00	6.00
Teolegs	2.50	3.50	5.50	3.00	54.00	1.50	7.00	6.00	6.50	9.50	6.50	1.50	2.00	9.50
Agricultors	0.00	1.00	2.00	0.00	1.50	0.00	0.00	0.50	2.00	0.50	2.00	1.00	0.50	2.50

Taula 8: Primers casos de la matriu S a partir de la descomposició de la matriu quadrada N

Matriu T

Els primers casos (taula 9)

	Mil	Artist	Mes	Art	Teo	Agr	Prurals	Adv	Esc	Com	Met	Mar	Pol	Docs
Militars	0.00	-1.00	-1.00	0.00	-2.50	0.00	-8.00	0.00	1.50	-6.00	1.50	0.00	0.00	-1.50
Artistes	1.00	0.00	-2.00	-5.50	-1.50	-1.00	-0.50	-2.00	0.50	-8.00	-2.00	-1.50	0.50	-1.00
Mestres	1.00	2.00	0.00	0.00	3.50	-1.00	-0.50	0.00	1.50	-1.00	-0.50	0.00	0.00	3.50
Artesans	0.00	5.50	0.00	0.00	2.00	0.00	0.00	0.50	3.50	0.00	1.00	0.00	0.00	4.00
Teolegs	2.50	1.50	-3.50	-2.00	0.00	-1.50	-7.00	0.00	2.50	-5.50	5.50	1.50	-1.00	3.50
Agricultors	0.00	1.00	1.00	0.00	1.50	0.00	0.00	0.50	2.00	0.50	2.00	1.00	0.50	2.50

Taula 9: Primers casos de la matriu T a partir de la descomposició de la matriu quadrada N

AC de la matriu simètrica S

(codi R en l'apèndix B.9)

La **inèrcia total de la part simètrica (S)** és 1.1485, un **88.55% de la inèrcia total** (que és de 1.297)

Els **percentatges d'inèrcia explicada pels dos primers eixos** són: 33.84% + 33.84%= 54.04%

Presentem el **mapa simètric** i.e. **files i columnes en coordenades principals**. (Figura 3)

Les coordenades de les files i columnes són les mateixes. El mapa mostra aquesta simetria, els fills tenen la mateixa professió que els seus pares. Sense fluxos en aquesta part de la matriu descomposta.

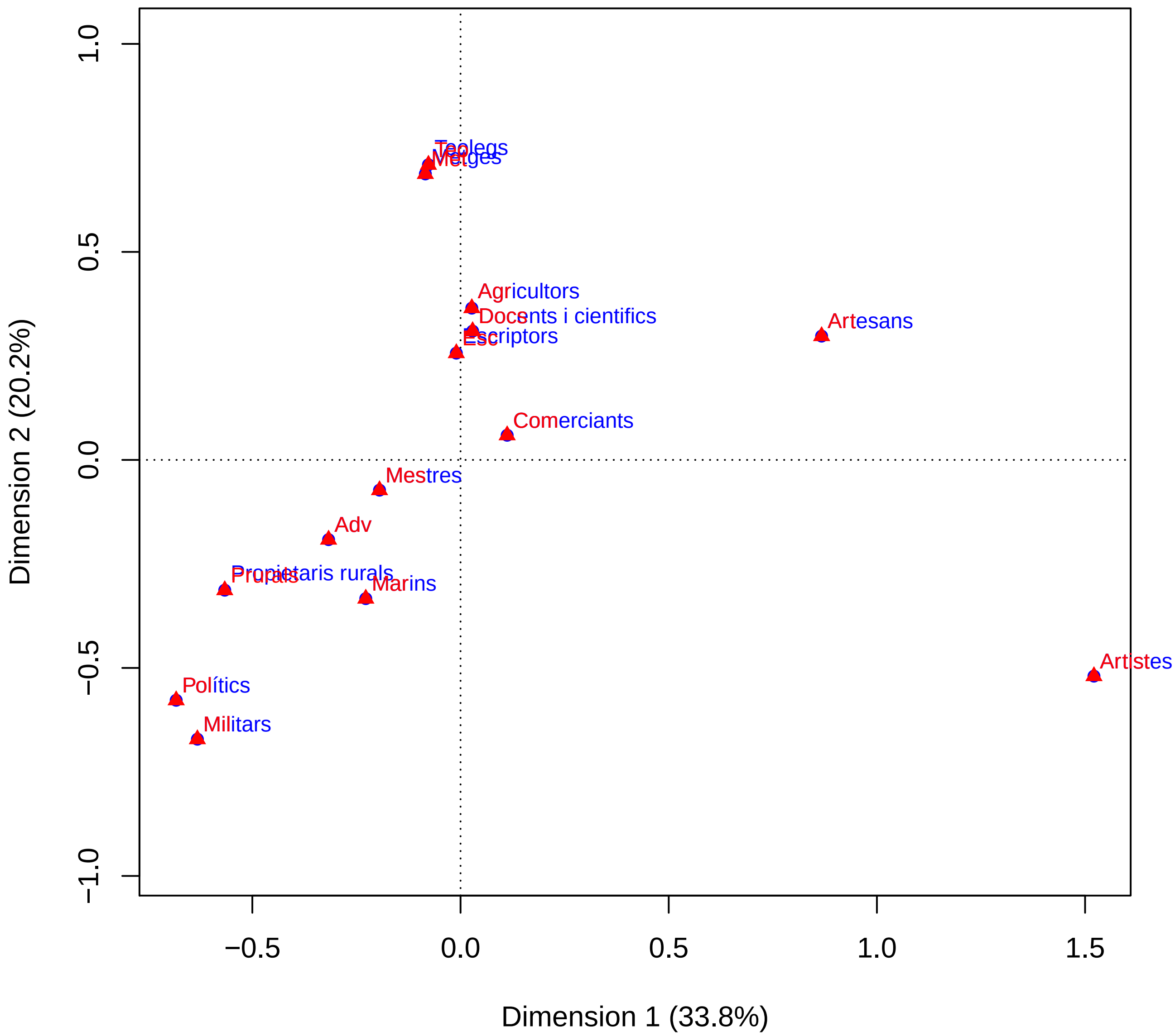

Figura 3: Mapa simètric AC de la matriu simètrica S Coordenades files=Coordenades columnes

AC a la matriu antisimètrica T

Atenció!

1. No podem aplicar directament l'algoritme de AC, la **matriu T té valors positius i negatius**, de fet la suma de la matriu T és zero, no té sentit calcular la matriu de freqüències relatives.

2. La **suma de files i columnes** per obtenir les masses **tampoc té sentit** per aquesta matriu T.

Per tant caldria un algoritme especial per analitzar la matriu T. Per evitar això, **es fa una transformació de la matriu de dades** i es fa un **AC simultani de les parts simètriques i antisimètriques**.

8.2 Transformació de les matrius quadrades

Volem fer una transformació de la matriu de dades N perquè es pugui fer un AC simultani de les parts simètriques i antisimètriques. Considerem primer la descomposició en la part simètrica i antisimètrica de la matriu P i com centrem les matrius P i S. Després veurem la transformació de la matriu de dades N que ens permetrà fer l'AC simultani de les parts.

Sigui la matriu de dades N i la matriu de freqüències relatives P amb els marginals r i c:

$$N = \begin{pmatrix} n_{11} & \dots & n_{1I} \\ \dots & \dots & \dots \\ n_{I1} & \dots & n_{II} \end{pmatrix}, P = \begin{pmatrix} \frac{n_{11}}{n} & \dots & \frac{n_{1I}}{n} \\ \dots & \dots & \dots \\ \frac{n_{I1}}{n} & \dots & \frac{n_{II}}{n} \end{pmatrix}$$

Marginals fila de P: $r = \sum_i \frac{n_{i.}}{n}$
Marginals columna de P: $c = \sum_i \frac{n_{.i}}{n}$

A partir de la matriu de freqüències relatives P, fem la seva descomposició (equació 16) en matriu simètrica S i matriu antisimètrica T.

$$P = \frac{1}{2}(P + P^T) + \frac{1}{2}(P - P^T) = S + T$$

La matriu simètrica S tindrà els marginals files i columnes ($w = \frac{1}{2}(r + c)$) iguals a la **mitjana dels marginals de P**, mentre que la suma dels elements de la matriu T serà zero.
Podem doncs centrar la matriu P a partir de ww^T en comptes de rc^T com s'ha fet fins ara:

$$P - ww^T = S - ww^T + T \tag{18}$$

I la corresponent **descomposició de la inèrcia**:

$$\sum_i \sum_j \frac{(p_{ij} - w_i w_j)^2}{w_i w_j} = \sum_i \sum_j \frac{(s_{ij} - w_i w_j)^2}{w_i w_j} + \sum_i \sum_j \frac{t_{ij}^2}{w_i w_j} \tag{19}$$

La inèrcia calculada d'aquesta manera és superior a la calculada amb la matriu centrada amb rc^T. Com més diferents siguin r i c, més gran serà la inèrcia calculada.

Transformació matriu N

Per fer **l'AC simultani**, cal que es realitzi a partir de la matriu $\tilde{N}$:

$$\tilde{N} = \begin{pmatrix} N & N^T \\ N^T & N \end{pmatrix}$$

Matriu N és (I x I)
Nova matriu transformada és (2I x 2I)

Les dimensions corresponents a la matriu antisimètrica sempre són amb parells d'inèrcies principals iguals. Veiem-ho:

El nombre total de casos de la matriu transformada $\tilde{N}$ és $n + n + n + n = 4n$

La seva **matriu de correspondències** serà

$$\tilde{P} = \frac{1}{4n}\tilde{N} = \frac{1}{4n}\begin{pmatrix} N & N^T \\ N^T & N \end{pmatrix} = \frac{1}{4}\begin{pmatrix} P & P^T \\ P^T & P \end{pmatrix}$$

Les marginals fila i columna de la matriu $\tilde{P}$ coincideixen i valen:

$$\tilde{w} = \frac{1}{2}\begin{pmatrix} w \\ w \end{pmatrix}$$

A més la inèrcia total de $\tilde{P}$ és la mateixa que P:

$$\text{traça}[D_{\tilde{w}}^{-1}(\tilde{P} - \tilde{w}\tilde{w}^T)D_{\tilde{w}}^{-1}(\tilde{P} - \tilde{w}\tilde{w}^T)^T] = \text{traça}[D_w^{-1}(P - ww^T)D_w^{-1}(P - ww^T)^T]$$

A partir de l'equació 18 tenim:

$$\tilde{P} - \tilde{w}\tilde{w}^T = \frac{1}{4}\begin{pmatrix} P - ww^T & P^T - ww^T \\ P^T - ww^T & P - ww^T \end{pmatrix} = \frac{1}{4}\begin{pmatrix} S - ww^T & S - ww^T \\ S - ww^T & S - ww^T \end{pmatrix} + \frac{1}{4}\begin{pmatrix} T & T^T \\ T^T & T \end{pmatrix} \tag{20}$$

La descomposició en valors singulars generalitzada (GVSD) de $\tilde{P} - \tilde{w}\tilde{w}^T$ es pot fer per blocs a partir de les submatrius.

Nota: GSVD

Sigui la descomposició en valors singulars de la matriu $M_{(m \times n)}$

$$M = U\Sigma V' \text{ on}$$
$$U'W_uU = V'W_vV = Id$$

U i V són ortonormals donats les matrius de pesos W_u i W_v

Lema (Mostrem el lema de l'article [4])

Sigui T una matriu antisimètrica $(m \times m)$ amb valors propis (VAPs) $i\lambda_1, -i\lambda_1, ..., i\lambda_j, -i\lambda_j$, on $\lambda_p > 0$, $p = 1, ..., j$, $j = \frac{m}{2}$

Llavors la SVD de $T = UDV^T$ tindrà la matriu D com:

$$D = \begin{pmatrix} \lambda_1 & & & & & \\ & \lambda_1 & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & \lambda_j & \\ & & & & & \lambda_j \end{pmatrix}$$

Així tornant al nostre cas, tenim:

$$S - ww^T = XD_\lambda X^T; \text{ on } X^T D_w^{-1} X = I$$
$$T = YD_\mu JY^T; \text{ on } Y^T D_w^{-1} Y = I$$

J és un bloc de matrius 2x2

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{tal que } J^T = -J; \; J^T D_\mu = -D\mu J$$

Una matriu antisimètrica té els valors propis purament imaginaris i conjugats per parelles, però a l'AC utilitzem els valors singulars obtinguts amb la GSVD que són valors reals aparellats, així D_μ té els valors singulars aparellats. [1]

Seguint amb l'equació 20:

$$\tilde{P} - \tilde{w}\tilde{w}^T = \frac{1}{4}\begin{pmatrix} XD_\lambda X^T & XD_\lambda X^T \\ XD_\lambda X^T & XD_\lambda X^T \end{pmatrix} + \frac{1}{4}\begin{pmatrix} YD_\mu JY^T & -YD_\mu JY^T \\ -YD_\mu JY^T & YD_\mu JY^T \end{pmatrix} = \tag{21}$$

$$= \begin{pmatrix} \frac{1}{2}X & \frac{1}{2}Y \\ \frac{1}{2}X & -\frac{1}{2}Y \end{pmatrix} \begin{pmatrix} D_\lambda & 0 \\ 0 & D_\mu \end{pmatrix} \begin{pmatrix} \frac{1}{2}X & \frac{1}{2}YJ^T \\ \frac{1}{2}X & -\frac{1}{2}YJ^T \end{pmatrix} \tag{22}$$

Per tant veiem que la descomposició en valors singulars de $\tilde{P} - \tilde{w}\tilde{w}^T$ està formada a partir dels components simètrics i antisimètrics de $P - ww^T$ i tenim els valors singulars de la part simètrica i antisimètrica separats en respectivament D_λ i D_μ. A més els valors de D_μ són aparellats.

AC sobre la matriu transformada

Reprenem de nou les dades de karl Pearson referent a la mobilitat social (7) (Codi R en l'apèndix A.3)

I=14

[1]"Robust Late Fusion with Rank Minimization, Supplementary material" Guangnan Ye, Dong Liu, I-Hong Jhuo, Shih-Fu Chang

Matriu N és (I x I)=(14 x 14)
Inèrcia total és 1.59908

Mostrem les 2I-1=27 inèrcies principals explicades (els valors singulars) de la nova matriu són:

```
## Les inèrcies principals explicades són:
##    Dim1    Dim2    Dim3    Dim4    Dim5    Dim6    Dim7    Dim8    Dim9
## 0.38868 0.23204 0.15836 0.15836 0.14391 0.12376 0.08184 0.07074 0.04984
##   Dim10   Dim11   Dim12   Dim13   Dim14   Dim15   Dim16   Dim17   Dim18
## 0.04184 0.04184 0.02287 0.02205 0.01287 0.01287 0.01036 0.00759 0.00759
##   Dim19   Dim20   Dim21   Dim22   Dim23   Dim24   Dim25   Dim26   Dim27
## 0.00309 0.00309 0.00166 0.00115 0.00115 0.00062 0.00038 0.00038 0.00015
```

La **inèrcia total de la part simètrica** és la suma de les (I-1)=13 inèrcies principals (són els valors singulars que no estan repetits.)
Els parells d'inèrcia iguals corresponen a l'anàlisi antisimètric.

Per tant, les inèrcies de cada matriu i la seva inèrcia total és:

Les **inèrcies de la part simètrica** són:

```
##    Dim1    Dim2    Dim5    Dim6    Dim7    Dim8    Dim9   Dim12   Dim13
## 0.38868 0.23204 0.14391 0.12376 0.08184 0.07074 0.04984 0.02287 0.02205
##   Dim16   Dim21   Dim24   Dim27
## 0.01036 0.00166 0.00062 0.00015
```

La **inèrcia total de la matriu simètrica** és:

```
## [1] 1.14852
```

En la matriu simètrica hi ha el **71.82% de la inèrcia total**.

Les **inèrcies de la matriu antisimètrica** són:

```
##    Dim3    Dim4   Dim10   Dim11   Dim14   Dim15   Dim17   Dim18   Dim19
## 0.15836 0.15836 0.04184 0.04184 0.01287 0.01287 0.00759 0.00759 0.00309
##   Dim20   Dim22   Dim23   Dim25   Dim26
## 0.00309 0.00115 0.00115 0.00038 0.00038
```

La **inèrcia total de la matriu antisimètrica** és:

```
## [1] 0.45056
```

En la matriu antisimètrica hi ha el **28.18% de la inèrcia total** (de la matriu transformada). Les inèrcies que millor expliquen la matriu antisimètrica són les de dimensions 3 i 4, que són el **70.29% de la seva inèrcia**.

Representem el **mapa asimètric de la part antisimètrica de la matriu transformada**.(Figura 4).
(Codi R en l'apèndix B.10)
Representem les **files en coordenades principals (sobre els eixos de les dimensions 3 i 4)**. Els eixos tindran el **35.15% respecte la inèrcia de la part antisimètrica** i el 9.9% respecte de la inèrcia total de la matriu transformada.

Per la interpretació del mapa de la part antisimètrica de la matriu transformada, no fem servir les distàncies entre punts, **interpretem les àrees triangulars formats per parells de punts (pares - fills) i l'origen**. Les àrees indiquen la intensitat de les asimetries entre els parells de punts. Per exemple, el triangle format per pares "comerciants" - fills "docents i cientfícs" - origen, és un àrea gran, indica que hi ha un gran flux diferencial de pares "comerciants" amb fills "docents i científics". Les àrees petites indiquen que no hi ha un flux entre professions, però podria haver fluxos d'entrada.

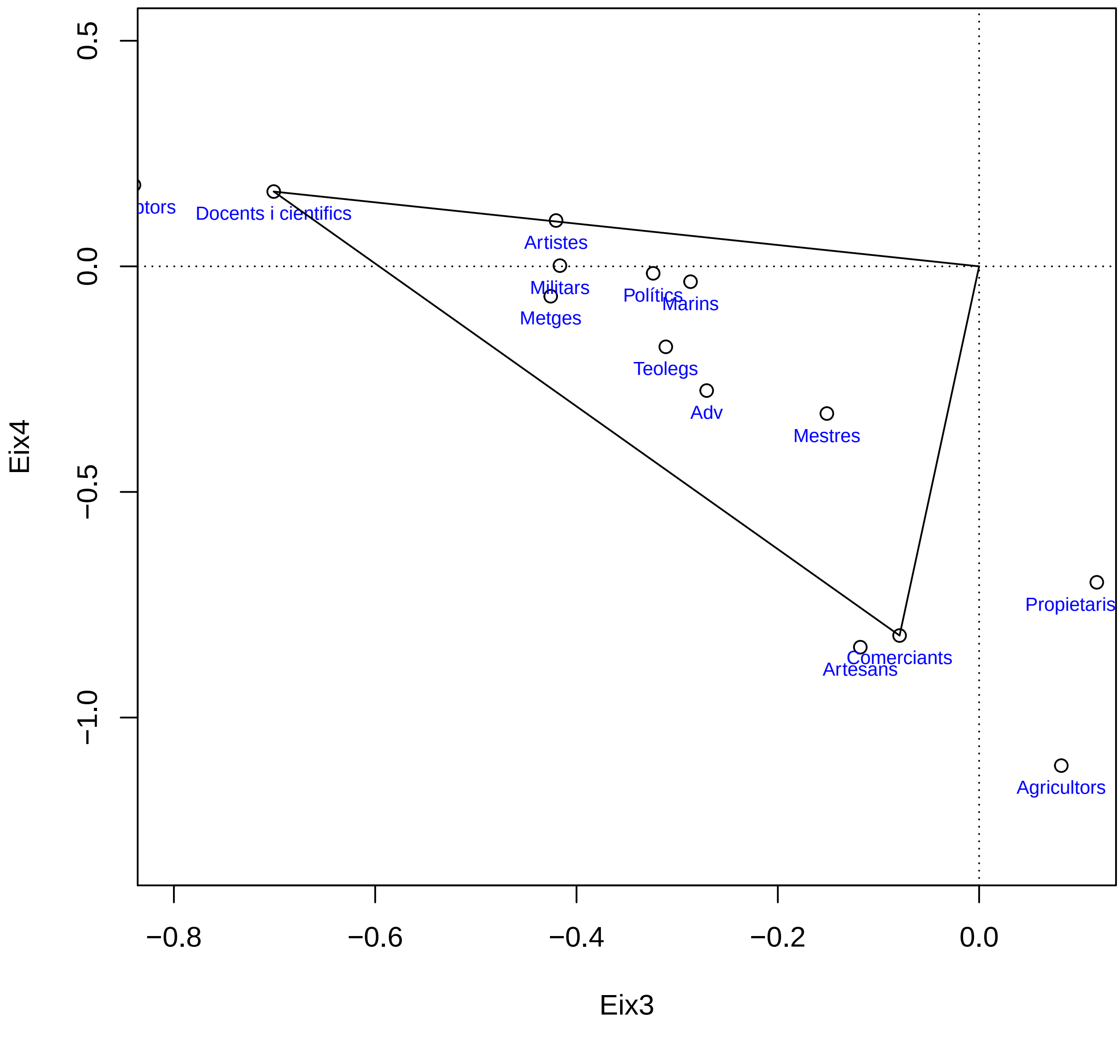

Figura 4: Mapa asimètric del AC de la part antisimètrica de la matriu transformada

9 Una segona aplicació de l'AC en Matrius Quadrades

Aplicarem ara el nostre codi R implementat sobre les dades dels **moviments migratoris obligats pel treball**, entre les comarques de Catalunya corresponents al cens del 1996 (Font: Institut d'Estadística de Catalunya (Idescat)) (Taula en l'apèndix C.1) La taula de contingència és doncs una taula de 41x41 on les files són les comarques d'origen dels treballadors i les columnes són les seves destinacions. La gran majoria dels treballadors treballen en la mateixa comarca d'origen, és només una minoria que han hagut de fer moviments migratoris. Aquesta part minoritària de la informació és sobre la que estem interessats. Volem saber quins han estat els seus moviments migratoris.

Realitzem primer un **breu descriptiu de la taula de dades**. (Codi R en l'apèndix C.2)
Hi ha un total de 2192976 d'individus.
Si observem la diagonal dels perfils fila, veurem quines comarques són les que han tingut més moviments de migració i quines menys. Així doncs la comarca amb menys moviments migratoris és: "Terra.Alta..37." és a dir, qui té el valor de la diagonal més alt. La comarca amb més moviments migratoris serà qui té el valor més baix, en aquest cas és "Baix.Llobregat..11."
El perfil fila de la comarca "Terra.Alta..37." té un 95.81% del marginal fila
El perfil fila de la comarca "Baix.Llobregat..11." té un 67.46% del marginal fila
Si calculem la mitjana de la diagonal dels perfils fila: 0.87 i.e. un 87% i la seva variància 0.00489, observem que la diagonal té molt pes i els moviments migratoris obligats pel treball és un percentatge molt baix en general.

Si calculem la inèrcia que aporta les cel·les de la diagonal, veurem que el percentatge d'inèrcia respecte la inèrcia total és elevada.

```
## Inèrcia total de les dades és:
## [1] 28.65979
##
## Inèrcia que aporta les cel·les de la diagonal és:
## [1] 28.02075
```

La diagonal aporta **97.77% de la inèrcia total**.
Si no féssim cap transformació a la matriu de dades, conclouríem que la matriu és simètrica.

9.1 AC sobre la matriu transformada de les dades

(Codi R a l'apèndix C.3)
Sigui N les dades originals de dimensions (I x I)=(41 x 41) La **inèrcia total de les dades transformades** és:

```
## [1] 28.58762
```

Les inèrcies principals explicades són:

```
##          Dim1          Dim2          Dim3          Dim4          Dim5
## 9.424781e-01 9.333200e-01 9.206392e-01 8.996228e-01 8.772883e-01
##          Dim6          Dim7          Dim8          Dim9         Dim10
## 8.751765e-01 8.647459e-01 8.482826e-01 8.401162e-01 8.355590e-01
##         Dim11         Dim12         Dim13         Dim14         Dim15
## 8.134015e-01 8.088789e-01 7.953715e-01 7.936682e-01 7.787898e-01
##         Dim16         Dim17         Dim18         Dim19         Dim20
## 7.746473e-01 7.598475e-01 7.515592e-01 7.508077e-01 7.335168e-01
##         Dim21         Dim22         Dim23         Dim24         Dim25
## 7.274093e-01 7.202878e-01 7.139828e-01 7.107557e-01 6.862025e-01
##         Dim26         Dim27         Dim28         Dim29         Dim30
## 6.722843e-01 6.580715e-01 6.433607e-01 6.160241e-01 6.145129e-01
##         Dim31         Dim32         Dim33         Dim34         Dim35
## 6.017460e-01 5.992311e-01 5.916369e-01 5.543172e-01 5.436614e-01
##         Dim36         Dim37         Dim38         Dim39         Dim40
## 5.371976e-01 5.316285e-01 4.881586e-01 4.665980e-01 2.935694e-01
##         Dim41         Dim42         Dim43         Dim44         Dim45
## 3.000222e-03 3.000222e-03 2.363323e-03 2.363323e-03 2.001110e-03
##         Dim46         Dim47         Dim48         Dim49         Dim50
## 2.001110e-03 6.261223e-04 6.261223e-04 5.272317e-04 5.272317e-04
##         Dim51         Dim52         Dim53         Dim54         Dim55
## 4.824663e-04 4.824663e-04 1.672725e-04 1.672725e-04 1.314850e-04
##         Dim56         Dim57         Dim58         Dim59         Dim60
## 1.314850e-04 9.874765e-05 9.874765e-05 6.711067e-05 6.711067e-05
##         Dim61         Dim62         Dim63         Dim64         Dim65
## 5.934619e-05 5.934619e-05 5.195186e-05 5.195186e-05 2.877629e-05
##         Dim66         Dim67         Dim68         Dim69         Dim70
## 2.877629e-05 1.555544e-05 1.555544e-05 5.465121e-06 5.465121e-06
##         Dim71         Dim72         Dim73         Dim74         Dim75
## 3.158992e-06 3.158992e-06 1.682997e-06 1.682997e-06 5.842672e-07
##         Dim76         Dim77         Dim78         Dim79         Dim80
## 5.842672e-07 1.558025e-07 1.558025e-07 2.566714e-09 2.566714e-09
##         Dim81
## 1.597117e-31
```

Fixem-nos en que la última inèrcia és zero.

Estudiem les inèrcies de la part simètrica i de la part antisimètrica de la matriu de dades.

9.2 Part simètrica

Les inèrcies de la part simètrica són aquelles que no estan repetides en la matriu transformada, són:

```
##      Dim1      Dim2      Dim3      Dim4      Dim5      Dim6      Dim7
## 0.9424781 0.9333200 0.9206392 0.8996228 0.8772883 0.8751765 0.8647459
##      Dim8      Dim9     Dim10     Dim11     Dim12     Dim13     Dim14
## 0.8482826 0.8401162 0.8355590 0.8134015 0.8088789 0.7953715 0.7936682
##     Dim15     Dim16     Dim17     Dim18     Dim19     Dim20     Dim21
## 0.7787898 0.7746473 0.7598475 0.7515592 0.7508077 0.7335168 0.7274093
##     Dim22     Dim23     Dim24     Dim25     Dim26     Dim27     Dim28
## 0.7202878 0.7139828 0.7107557 0.6862025 0.6722843 0.6580715 0.6433607
##     Dim29     Dim30     Dim31     Dim32     Dim33     Dim34     Dim35
## 0.6160241 0.6145129 0.6017460 0.5992311 0.5916369 0.5543172 0.5436614
##     Dim36     Dim37     Dim38     Dim39     Dim40
## 0.5371976 0.5316285 0.4881586 0.4665980 0.2935694
```

La **inèrcia total de la part simètrica** és la suma de les (I-1)=40 inèrcies principals (són els valors singulars que no estan repetits)

```
## [1] 28.56835
```

Si comparem la inèrcia total de la matriu transformada amb la inèrcia de la part simètrica de la matriu transformada, efectivament veiem que la part simètica té un **99.93% de la inèrcia total** de les dades transformades.

9.3 Part antisimètrica

Les inèrcies de la matriu antisimètrica són:

```
##        Dim41        Dim42        Dim43        Dim44        Dim45
## 3.000222e-03 3.000222e-03 2.363323e-03 2.363323e-03 2.001110e-03
##        Dim46        Dim47        Dim48        Dim49        Dim50
## 2.001110e-03 6.261223e-04 6.261223e-04 5.272317e-04 5.272317e-04
##        Dim51        Dim52        Dim53        Dim54        Dim55
## 4.824663e-04 4.824663e-04 1.672725e-04 1.672725e-04 1.314850e-04
##        Dim56        Dim57        Dim58        Dim59        Dim60
## 1.314850e-04 9.874765e-05 9.874765e-05 6.711067e-05 6.711067e-05
##        Dim61        Dim62        Dim63        Dim64        Dim65
## 5.934619e-05 5.934619e-05 5.195186e-05 5.195186e-05 2.877629e-05
##        Dim66        Dim67        Dim68        Dim69        Dim70
## 2.877629e-05 1.555544e-05 1.555544e-05 5.465121e-06 5.465121e-06
##        Dim71        Dim72        Dim73        Dim74        Dim75
## 3.158992e-06 3.158992e-06 1.682997e-06 1.682997e-06 5.842672e-07
##        Dim76        Dim77        Dim78        Dim79        Dim80
## 5.842672e-07 1.558025e-07 1.558025e-07 2.566714e-09 2.566714e-09
```

La **inèrcia total de la matriu antisimètrica** és 0.0193, i.e és un 0.07% de la inèrcia total de les dades transformades.
La inèrcia de la part antisimètrica és gairebé inapreciable, i per aquest motiu, les inèrcies s'han acumulat justament en els últims valors de la descomposició de valors singulars.

9.4 Mapa asimètric

Representem el mapa asimètric de la part antisimètrica de les dades. (Figura 5) Les dimensions 41 i 42 són les que tenen més inèrcia, contribueixen un 31.149% de la seva inèrcia total (0.0193). (Codi R a l'apèndix C.4)

No és estrany que justament siguin les dimensions 41 i 42 les dimensions amb més inèrcia perquè la inèrcia de la part antisimètrica de les dades era molt petita en comparació amb la inèrcia total. Els valors singulars de la part antisimètrica s'han concentrat justament en les últimes dimensiones, i per això les dimensions que més contribueixen a la part antisimètrica són la 41 i 42.

En el mapa (Figura 5) que explica el 31.149% de la inèrcia de la matriu antisimètrica podem veure els fluxos dels moviments migratoris interpretant les àrees dels triangles formats pels parells de punts i l'origen. Els fluxos s'interpreten en el sentit de les agulles del rellotge amb el triangle format per parells de punts i l'origen de les coordenades.

Si mirem per exemple la comarca de l'Urgell i la Segarra, veiem que forma un triangle molt gran, això indica que habitants d'Urgell que s'han desplaçat a un altre comarca, hi ha un gran flux de moviment migratori cap a la Segarra. Són 198 d'origen Urgell desplaçats cap a la Segarra obligats per motius de feina.

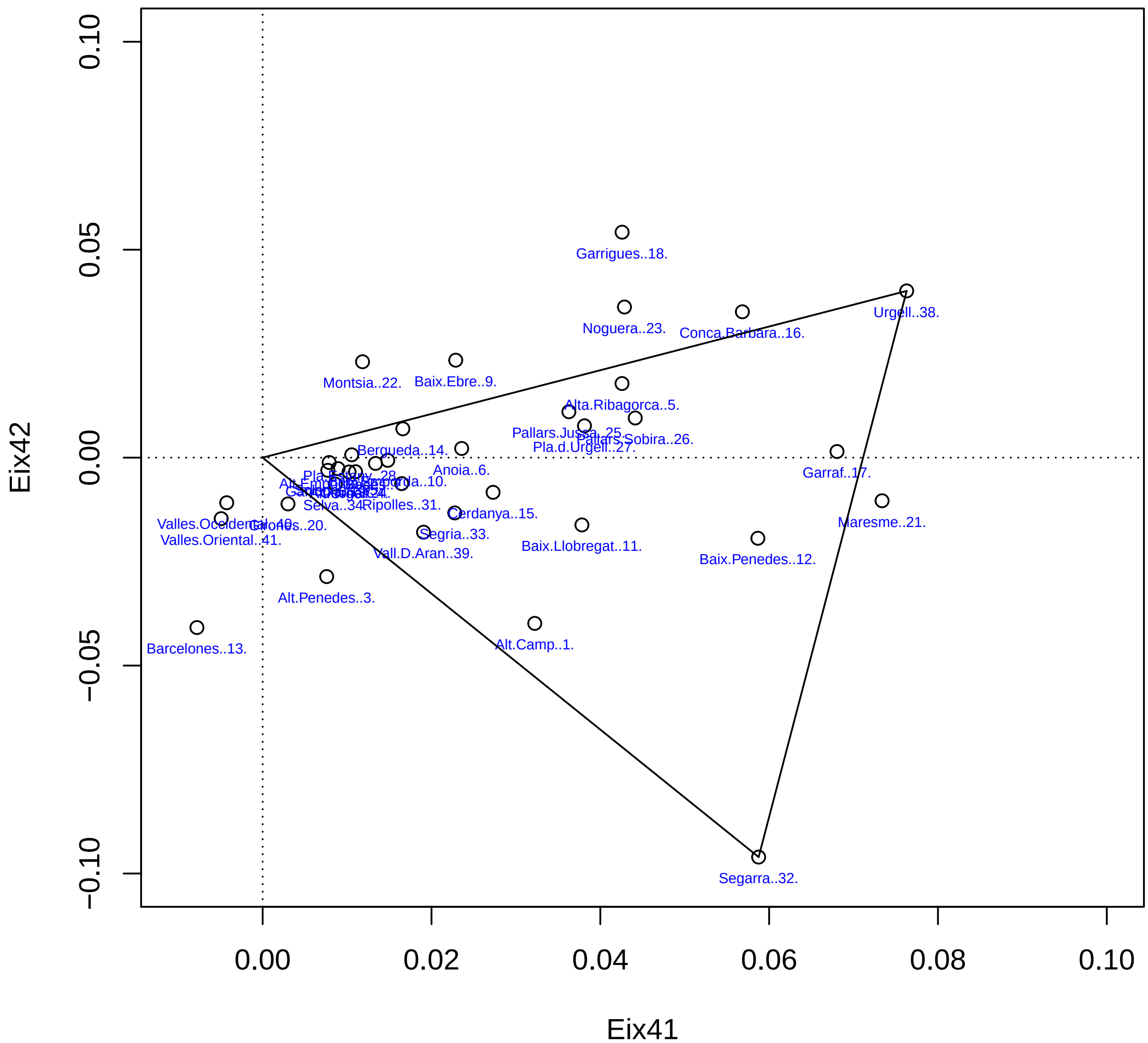

Figura 5: "Mapa asimètric de la part antisimètrica de la matriu transformada"

10 Conclusions

L'anàlisi de correspondències (AC) ens permet estudiar les relacions estadístiques existents (no implica relacions de causa efecte) entre files i columnes, així com fer una descripció més detallada de les taules de contingència. És una adaptació de l'anàlisi de components principals (ACP); la gran diferència respecte a l'ACP, però, és la matriu sobre la qual s'obtenen els valors singulars: mentre que en l'ACP els valors singulars resulten d'aplicar la SVD sobre la matriu de dades centrada, en l'AC la SVD s'aplica sobre la **matriu d'inèrcies**.

La matriu d'inèrcies s'obté a partir de la matriu de freqüències relatives, centrada respecte a la massa de files i columnes i ponderada per la inversa de l'arrel de la massa de les files i columnes. Dit d'una altra manera, ponderem els punts originals i calculem la distància euclídia de les projeccions dels perfils sobre el subespai. La distància d'aquestes projeccions són les distàncies χ^2. Recordem que l'estadístic χ^2 ens permet decidir si dues variables són independents o no ho són. A partir de la matriu d'inèrcies s'estudien les contribucions a la inèrcia de cada cel·la, de les files, de les columnes. La inèrcia de cada eix també es pot descomposar per saber com en contribueix cada perfil.

En taules quadrades on tota la inèrcia és carregada sobre la diagonal hi ha informació que queda oculta. Per veure la part oculta cal fer una transformació de la matriu de dades que ens permeti fer una anàlisi AC i obtenir alhora les inèrcies i les coordenades de la part simètrica i antisimètrica. Per provar-ho s'ha aplicat un **AC amb SVD** sobre les dades clàssiques de **Karl Pearson**, unes dades quadrades amb cert pes a la diagonal. Els valors singulars obtinguts de la part simètrica i de la part antisimètrica estaven barrejats, tot i que els valors singulars de la part antisimètrica sempre són valors aparellats. En canvi, en una segona aplicació on el pes de la diagonal ha estat sumament més gran que la resta, els valors singulars obtinguts han aparegut ordenats: primer els valors de la part simètrica i després els de la part antisimètrica. Això és així a causa, justament, del gran pes de la diagonal en comparació amb la resta dels seus valors.

Pel que fa a la representació de les coordenades i la seva interpretació, la transformació de la matriu de dades provoca que la representació i la interpretació dels mapes sigui diferent de l'habitual. En el mapa es representen les files en coordenades principals i el que s'estudia són els fluxos que es produeixen d'un perfil a l'altre. Aquests fluxos es representen amb triangles formats pels parells de punts i l'origen. Una àrea gran del triangle indica que hi ha molt flux d'un perfil a l'altre.

Els codis de R propis ("FunctionInercia", "MyAC" i "ACMatriuQuad") han estat contrastats amb els exemples proposats en el llibre "La práctica del análisis de correspondencias" de Michael Greenacre ([1]).

A Codis R essencials

A.1 Càlcul Ji-quadrat i la inèrcia total

```
FunctionInercia<-function(DADES)
{
  ### Funcio que donat les dades, retorna el ji-quadrat i la seva inèrcia total ###

  DADES.rowsum <- apply(DADES, 1, sum)
  DADES.colsum <- apply(DADES, 2, sum)
  n<- sum(DADES)
  DADES.exp<- DADES.rowsum %o% DADES.colsum / n
  # Equivalentment
  DADES.exp<- as.matrix(DADES.rowsum) %*% as.matrix(t(DADES.colsum)) / n

  res<-list()
  ## ji-quadrat ###
  res$chi2<-sum((DADES - DADES.exp)^2 / DADES.exp)

  ### Inercia ###
  res$Inercia<-res$chi2/n

  return(res)
}
```

A.2 Codi algoritme AC

```
MyAC<-function(DADES)
{
  ###################
  ### Algoritme AC ###
  ###################

  # Matriu de correspondències P=(1/n)*N   (N matriu dades)
  # Centrem les dades respecte el nombre total de dades
  DADES.P<-DADES/sum(DADES)
  # sum(DADES.P)
  # 1/sum(DADES)*DADES

  # Massa de les files
  DADES.r<-apply(DADES.P,1,sum)
    # sum(DADES.r)
```

```r
# Massa de les columnes
DADES.c<-apply(DADES.P,2,sum)
# sum(DADES.c)

# Matriu diagonal de massa files
DADES.Dr<-diag(DADES.r)
# sum(DADES.Dr)

# Matriu diagonal de massa columnes
DADES.Dc<-diag(DADES.c)
# sum(DADES.Dc)

# Matriu Dr elevat a menys 1/2
DADES.Drmh<-diag(1/sqrt(DADES.r))

# Matriu Dc elevat a menys 1/2
DADES.Dcmh<-diag(1/sqrt(DADES.c))

DADES.P<-as.matrix(DADES.P)


### Matriu S ###
### Matriu de residus estandarditzats ###
### Matriu per diagonalitzar ###
DADES.S<-DADES.Drmh %*% (DADES.P - DADES.r %*% t(DADES.c)) %*% DADES.Dcmh
rownames(DADES.S)<-rownames(DADES)
colnames(DADES.S)<-colnames(DADES)

### S= U Lambda V'
DADES.svd<-svd(DADES.S)



### Càlcul de les coordenades principals (pc) i estàndards (sc) ###
# Coordenades estandard files (vèrtex files)
DADES.rsc<-DADES.Drmh %*% DADES.svd$u
rownames(DADES.rsc)<-rownames(DADES)
colnames(DADES.rsc)<-paste("Eix",1:(dim(DADES)[2]),sep="")

# Coordenades estandard columna
DADES.csc<-DADES.Dcmh %*% DADES.svd$v
rownames(DADES.csc)<-colnames(DADES)
colnames(DADES.csc)<-paste("Eix",1:(dim(DADES)[2]),sep="")
```

```r
# Coordenades principal files
DADES.rpc<-DADES.rsc %*% diag(DADES.svd$d)
rownames(DADES.rpc)<-rownames(DADES)
colnames(DADES.rpc)<-paste("Eix",1:(dim(DADES)[2]),sep="")

# Coordenades principal columna
DADES.cpc<-DADES.csc %*% diag(DADES.svd$d)
rownames(DADES.cpc)<-colnames(DADES)
colnames(DADES.cpc)<-paste("Eix",1:(dim(DADES)[2]),sep="")


res<-list()
res$P<-DADES.P
res$r<-DADES.r
res$c<-DADES.c
res$S<-DADES.S
res$svd<-DADES.svd
res$rpc<-DADES.rpc
res$cpc<-DADES.cpc
res$rsc<-DADES.rsc
res$csc<-DADES.csc

# Inèrcia total
res$IT<-FunctionInercia(DADES)$Inercia


# Inèrcia explicada per eix principal
res$InerciaExplicada<-(DADES.svd$d)^2
res$InerciaExplicada<-res$InerciaExplicada[-dim(DADES)[2]]
names(res$InerciaExplicada)<-paste("Dim",1:(dim(DADES)[2]-1),sep="")

# Percentatge de inèrcia explicada per eix principal
res$PctInerciaExplicada<-((DADES.svd$d)^2/FunctionInercia(DADES)$Inercia)*100
res$PctInerciaExplicada<-res$PctInerciaExplicada[-dim(DADES)[2]]
names(res$PctInerciaExplicada)<-paste("Dim",1:(dim(DADES)[2]-1),sep="")


#########################################################################
### Descomposició de les inèrcies respecte files i columnes        ###
### Inercia = sum(i-èssim massa)*(dist chi2 de i-èssim perfil mig) ###
```

```
  ### Pag 50                                                         ###
  ######################################################################
  # Contribució de les files a la Inèrcia (Inèrcia dels perfils fila)
  res$InerciaFila<-apply(res$S^2, 1, sum)

  # Contribució de les columnes a la Inèrcia total
  res$InerciaCol<-apply(res$S^2, 2, sum)

  # Contribucions de les cel·les a la inèrcia total
  res$InerciaCelda<-res$S^2

  # Percentatge de contribucions de les cel·les a la inèrcia total
  res$PctInerciaCelda<-100*res$S^2/sum(res$S^2)


  ####################################################
  ### Descomposició de les inèrcies respecte eixos ###
  ### (Descomposició de la inèrcia de l'eix)       ###
  ####################################################

  res$InerciaEix<-diag(DADES.r)%*%DADES.rpc^2
  res$InerciaEix<-res$InerciaEix[,-dim(DADES)[2]]
  rownames(res$InerciaEix)<-rownames(DADES)
  colnames(res$InerciaEix)<-paste("Eix",1:(dim(DADES)[2]-1),sep="")

  return(res)

}
```

A.3 AC sobre la matriu N transformada

```
ACMatriuQuad<-function(Dades)
{
### Transformació matriu original per AC simultani ###
Dades<-as.matrix(Dades)
Dades <- rbind(cbind(Dades,t(Dades)), cbind(t(Dades),Dades))
res<-list()
res$Matriu<-Dades
res$AC<-MyAC(Dades)
return(res)
}
```

B Codis R dels exemples

B.1 Simetria entre files i columnes

```
### Primera columna de la matriu de les ###
### coordenades principals dels perfils fila (rpc) ###
rpc1<-as.matrix(MyAC(Dades)$rpc[,1])
colnames(rpc1)<-c("Eix1")
cat("Primera columna de la matriu de les coordenades principals dels perfils fila (rpc)")
rpc1

rsc1<-as.matrix(MyAC(Dades)$rsc[,1])
colnames(rsc1)<-c("Eix1")
cat("\n Coordenades estàndard vèrtex fila (rsc1).")
rsc1

### Comprovem que efectivament es compleix la relaci\'o. \\
### A partir de les "coordenades v\`ertex", obtenim les "coordenades del perfil":

### Coordenades del perfil ###
CoordPerfil<-rsc1*sqrt(MyAC(Dades)$InerciaExplicada[1])
cat("Coordenades del perfil a partr de les coordenades del vèrtex")
CoordPerfil

### A partir de les "coordenades del perfil", obtenim les "coordenades v\`ertex":
cat("Coordenades vèrtex a partir de les coordenades del perfil")
CoordVertex<-rpc1*(1/sqrt(MyAC(Dades)$InerciaExplicada[1]))
CoordVertex
```

B.2 Càlcul de la contribució de cada cel·la a la inèrcia

```
# Contribucions de les cel·les a la inèrcia total
InerciaCelda<-MyAC(Dades)$S^2
# Percentatge de contribucions de les cel·les a la inèrcia total
PctInerciaCelda<-100*InerciaCelda/IT
```

B.3 Càlcul de la contribució a la inèrcia per fila

```
### Càlcul de les contribucions de les files a la inèrcia ###

# Contribució de les files a la Inèrcia (Inèrcia dels perfils fila)
```

```
cat("Contribució de les files a la inèrcia total")
InerciaFila<-as.matrix(apply(MyAC(Dades)$S^2, 1, sum),ncol=1)


# Percentatge de contribució de les files en la inèrcia
IT<-MyAC(Dades)$IT
PctInerciaFila<-(InerciaFila/IT)*100


IF<-cbind(InerciaFila,PctInerciaFila)
colnames(IF)<-c("Inèrcia", "% Inèrcia")
IF
```

B.4 Càlcul de la contribució a la inèrcia per columna

```
### Càlcul de les contribucions de les columnes en la inèrcia ###
# Contribució de les columnes a la Inèrcia total
cat("Contribució de les columnes a la inèrcia total")
InerciaCol<-apply(MyAC(Dades)$S^2, 2, sum)

# % de Contribucio de la columna en la inèrcia
PctInerciaCol<-(InerciaCol/IT)*100

IC<-rbind(InerciaCol,PctInerciaCol)
rownames(IC)<-c("Inèrcia", "% Inèrcia")
IC
```

B.5 Càlcul de la contribució de les files a la inèrcia de l'eix

```
  ####################################################
  ### Descomposició de les inèrcies respecte eixos ###
  ### (Descomposició de la inèrcia de l'eix)       ###
  ####################################################


  InerciaEix<-diag(Dades.r)%*%Dades.rpc^2
  InerciaEix<-InerciaEix[,-dim(Dades)[2]]
  rownames(InerciaEix)<-rownames(Dades)
  colnames(InerciaEix)<-paste("Eix",1:(dim(Dades)[2]-1),sep="")

  # Inèrcia total explicada
```

```
  # MyAC(Dades)£InerciaExplicada
  # colSums(InerciaEix)
```

B.6 Mapa asimètric del AC

```
rownames(smoke)<-c("DS", "DJ", "ES","EJ","SC")
### Mapa asimètric ###
plot(ca(smoke),map="rowprincipal")
#DistChi2(smoke)
#FunctionInercia(smoke)
  ############################################################################
  ### Mapa asimètric smoke. Codi propi ###
  ### Files en coordenades principals i columnes en coordenades estàndards ###
  ############################################################################

  rpc<-MyAC(smoke)$rpc[,1:2]
  csc<-MyAC(smoke)$csc[,1:2]
  plot(rbind(rpc,csc))
  # pch opcions dels punts en el plot
  text(rpc,rownames(rpc),cex=0.8,col="Blue",pos=4)
  text(csc,rownames(csc),cex=0.8,col="2",pos=4)
  abline(v=0, h=0,lty="dotted")
```

B.7 Mapa simètric del AC

B.8 Descomposició de la matriu quadrada de freqüències N

```
N<-mob
S<-1/2*(N+t(N))
T<-1/2*(N-t(N))
```

B.9 AC de la matriu simètrica S

```
### AC sobre la matriu simètrica S ###
InerciaS<-FunctionInercia(S)$Inercia
InerciaS<-round(InerciaS,4)

PctISExplicadaEix1<-round(MyAC(S)$PctInerciaExplicada[1],2)
PctISExplicadaEix2<-round(MyAC(S)$PctInerciaExplicada[2],2)
```

B.10 Mapa asimètric de la matriu transformada

```
### Mapa asimètric. Codi propi ###
# Files en coordenades principals i columnes en coordenades estàndards
# names(ACMatriuQuad(mob)£AC)
# ACMatriuQuad(mob)£AC£rpc[,3:4]

rpc<-ACMatriuQuad(mob)$AC$rpc[1:dim(mob)[1],3:4]
#rpc<-ACMatriuQuad(mob)£AC£rpc[(dim(mob)[1]+1):(2*dim(mob)[1]),3:4]
#csc<-ACMatriuQuad(mob)£AC£csc[1:dim(mob)[1],3:4]
#csc<-ACMatriuQuad(mob)£AC£csc[(dim(mob)[1]+1):(2*dim(mob)[1]),3:4]
# ?xlim
#?plot
#plot(rpc)
plot(rpc,xlim=c(-0.8,0.1),ylim=c(-1.3,0.5))
# plot(rbind(rpc,csc))
# plot(rbind(rpc,csc),xlim=c(-0.5,0.5),ylim=c(-2.2,0.5))

# pch opcions dels punts en el plot
text(rpc,rownames(rpc),cex=0.7,col="Blue",pos=1)
# text(csc,rownames(csc),cex=0.7,col="2",pos=1)
abline(v=0, h=0,lty="dotted")
### Triangle origen comerciants docents ###
#?segments
segments(x0=0,y0=0,x1=rpc[10,1],y1=rpc[10,2])
segments(x0=rpc[10,1],y0=rpc[10,2],x1=rpc[14,1],y1=rpc[14,2])
segments(x0=rpc[14,1],y0=rpc[14,2],x1=0,y1=0)
```

C Codis R Cas pràctic. AC en matrius quadrades

C.1 Dades dels moviments migratoris obligats pel treball

	A.Camp1	A.Emp2	A.Pen3	A.Urg4	A.Rib5	An6	Bag7	B.Camp8	B.Ebre9
Alt.Camp..1.	10119	0	17	0	0	17	0	579	34
Alt.Emporda..2.	4	32542	6	3	0	19	15	4	2
Alt.Penedes..3.	28	7	23368	2	0	411	54	41	8
Alt.Urgell..4.	1	0	1	5508	0	7	16	1	0
Alta.Ribagorca..5.	0	0	0	3	1029	0	0	1	0
Anoia..6.	7	16	262	3	5	27531	22	20	1
Bages..7.	6	9	31	13	1	180	48169	17	2
Baix.Camp..8.	417	2	15	0	1	5	15	37061	234
Baix.Ebre..9.	9	2	3	0	0	1	3	134	19487
Baix.Emporda..10.	1	453	10	2	4	14	7	7	1
Baix.Llobregat..11.	33	67	938	21	6	1115	804	130	31
Baix.Penedes..12.	95	4	487	1	0	13	15	170	23
Barcelones..13.	225	753	1879	162	91	1773	2130	1194	380
Bergueda..14.	0	4	42	7	0	12	436	9	1
Cerdanya..15.	0	3	1	70	0	1	20	0	1
Conca.Barbara..16.	289	0	4	0	0	100	4	95	4
Garraf..17.	10	3	343	3	1	36	17	33	9
Garrigues..18.	2	0	5	0	0	0	0	23	4
Garrotxa..19.	0	127	5	2	0	12	8	16	0
Girones..20.	2	1244	12	2	2	6	31	30	7
Maresme..21.	1	38	21	3	0	54	56	31	10
Montsia..22.	5	1	5	0	1	0	1	29	1180
Noguera..23.	1	1	2	22	2	5	4	7	0
Osona..24.	2	28	7	10	0	13	272	12	3
Pallars.Jussa..25.	0	0	0	9	23	1	1	1	4
Pallars.Sobira..26.	0	1	1	3	0	0	5	3	2
Pla.d.Urgell..27.	1	7	0	2	2	11	3	10	2
Pla.Estany..28.	0	139	14	0	0	1	7	0	1
Priorat..29.	6	0	0	0	0	1	1	50	4
Ribera.Ebre..30.	3	2	4	0	0	2	3	263	61
Ripolles..31.	0	17	0	1	0	2	10	2	4
Segarra..32.	13	7	2	11	1	331	40	2	7
Segria..33.	7	11	5	72	69	34	28	69	23
Selva..34.	1	107	7	2	0	6	19	6	6
Solsones..35.	0	0	4	42	1	9	159	4	0
Tarragones..36.	991	9	71	3	3	14	35	9449	331
Terra.Alta..37.	1	0	2	2	1	0	1	9	34
Urgell..38.	1	0	1	6	2	20	3	8	2
Vall.D.Aran..39.	2	5	1	2	47	2	1	9	0
Valles.Occidental..40.	24	84	279	32	6	300	1185	119	22
Valles.Oriental..41.	6	49	70	10	3	86	272	47	22

Taula 10: Moviments migratoris obligats pel treball (Taula 1/5) Font: Idescat

	B.Emp10	B.Llo11	B.Pen12	Bar13	Ber14	Cer15	C.Bar16	Gar17	Gar18
Alt.Camp..1.	1	14	137	61	1	0	395	17	14
Alt.Emporda..2.	524	36	1	226	4	2	2	5	0
Alt.Penedes..3.	7	968	889	1080	18	3	14	855	7
Alt.Urgell..4.	3	3	1	32	9	71	0	3	5
Alta.Ribagorca..5.	0	3	0	5	0	0	1	2	1
Anoia..6.	7	693	29	586	15	1	64	44	10
Bages..7.	19	464	18	806	760	9	1	33	11
Baix.Camp..8.	2	47	89	206	5	0	110	39	20
Baix.Ebre..9.	1	16	12	73	1	0	8	5	0
Baix.Emporda..10.	31730	28	3	186	7	3	1	0	4
Baix.Llobregat..11.	122	142346	430	52938	62	25	37	1272	28
Baix.Penedes..12.	9	131	12050	241	4	0	9	547	8
Barcelones..13.	1164	74642	2087	613094	501	358	237	5563	173
Bergueda..14.	2	19	3	85	10802	10	1	6	0
Cerdanya..15.	11	9	0	45	70	4354	0	6	0
Conca.Barbara..16.	0	10	16	24	0	0	5101	11	71
Garraf..17.	3	646	469	1040	16	0	0	23481	4
Garrigues..18.	0	1	2	14	0	0	17	0	4767
Garrotxa..19.	82	17	2	69	5	2	0	2	0
Girones..20.	2179	100	9	563	10	32	1	13	9
Maresme..21.	65	486	13	6693	21	13	4	51	3
Montsia..22.	0	13	50	34	0	0	4	3	0
Noguera..23.	0	5	0	32	3	5	2	2	14
Osona..24.	21	126	9	564	278	10	4	6	6
Pallars.Jussa..25.	3	5	0	24	0	0	0	1	7
Pallars.Sobira..26.	1	5	0	9	3	0	1	0	2
Pla.d.Urgell..27.	0	6	6	30	1	1	2	2	303
Pla.Estany..28.	49	10	7	30	2	1	0	1	0
Priorat..29.	0	0	6	12	0	0	3	2	6
Ribera.Ebre..30.	0	12	4	36	1	0	1	7	12
Ripolles..31.	15	6	2	62	36	22	0	122	1
Segarra..32.	2	20	0	77	4	3	17	0	40
Segria..33.	6	56	14	243	9	5	29	14	830
Selva..34.	319	81	7	663	5	20	3	9	1
Solsones..35.	0	3	5	24	90	6	0	0	1
Tarragones..36.	10	210	640	618	6	1	262	127	43
Terra.Alta..37.	0	0	1	1	0	0	0	0	3
Urgell..38.	2	16	2	32	0	0	30	1	49
Vall.D.Aran..39.	2	6	1	17	1	0	2	2	3
Valles.Occidental..40.	179	9424	232	38554	90	63	30	328	20
Valles.Oriental..41.	86	1912	45	16455	32	10	10	135	8

Taula 11: Moviments migratoris obligats pel treball (Taula 2/5) Font: Idescat

	Gar18	Garro19	Gir20	2016-03-21	Mont22	Nog23	Oso24	P.Jus25	P.Sob26	P.Urg27
Alt.Camp..1.	14	0	2	5	19	3	5	1	3	5
Alt.Emporda..2.	0	148	683	52	2	3	18	0	1	1
Alt.Penedes..3.	7	3	6	59	3	2	12	4	1	4
Alt.Urgell..4.	5	1	4	0	1	40	3	13	10	6
Alta.Ribagorca..5.	1	0	1	0	0	8	3	7	0	1
Anoia..6.	10	1	40	51	3	39	26	2	1	32
Bages..7.	11	5	21	65	10	10	229	4	7	8
Baix.Camp..8.	20	3	6	11	74	9	5	5	4	9
Baix.Ebre..9.	0	0	0	10	1468	2	4	0	0	3
Baix.Emporda..10.	4	53	1065	69	2	5	22	4	1	1
Baix.Llobregat..11.	28	22	150	1577	24	45	184	23	24	37
Baix.Penedes..12.	8	0	6	8	12	5	1	1	3	2
Barcelones..13.	173	262	1072	25345	252	387	1672	268	198	190
Bergueda..14.	0	1	9	10	0	3	61	1	1	0
Cerdanya..15.	0	4	4	7	0	3	26	4	1	4
Conca.Barbara..16.	71	0	1	2	2	7	0	0	0	11
Garraf..17.	4	0	7	43	4	1	15	0	3	4
Garrigues..18.	4767	0	0	1	0	16	0	0	0	78
Garrotxa..19.	0	16914	265	25	1	0	27	1	3	3
Girones..20.	9	656	44429	319	8	8	45	3	5	2
Maresme..21.	3	27	107	81561	4	8	48	14	5	6
Montsia..22.	0	0	1	10	15782	0	2	2	0	1
Noguera..23.	14	0	1	6	3	9383	1	24	10	170
Osona..24.	6	24	33	99	5	2	45904	0	3	2
Pallars.Jussa..25.	7	0	0	1	2	20	0	3890	48	5
Pallars.Sobira..26.	2	0	0	0	0	11	1	42	1839	0
Pla.d.Urgell..27.	303	0	0	10	1	212	2	6	5	8507
Pla.Estany..28.	0	208	707	25	0	0	12	0	0	1
Priorat..29.	6	0	0	0	0	1	0	0	0	0
Ribera.Ebre..30.	12	0	0	3	34	0	2	2	1	1
Ripolles..31.	1	92	33	8	0	5	307	0	0	4
Segarra..32.	40	0	2	0	0	307	8	19	3	208
Segria..33.	830	1	7	13	16	1265	15	131	94	866
Selva..34.	1	186	2620	1908	6	8	98	4	2	3
Solsones..35.	1	0	0	1	1	22	9	3	0	1
Tarragones..36.	43	4	19	61	184	26	18	16	3	18
Terra.Alta..37.	3	0	0	0	21	0	0	0	0	0
Urgell..38.	49	1	1	4	1	277	0	7	3	401
Vall.D.Aran..39.	3	0	2	4	1	14	1	5	19	7
Valles.Occidental..40.	20	33	180	2011	30	36	256	38	37	27
Valles.Oriental..41.	8	21	101	2505	7	19	1065	8	7	17

Taula 12: Moviments migratoris obligats pel treball (Taula 3/5) Font: Idescat

	P.Est28	Pri29	R.Ebre30	Rip31	Seg32	Segr33	Sel34	Sol35	Tar36
Alt.Camp..1.	0	7	7	0	1	11	0	0	1000
Alt.Emporda..2.	179	1	0	30	2	10	114	0	8
Alt.Penedes..3.	0	3	7	8	3	21	8	0	109
Alt.Urgell..4.	0	0	1	0	11	61	1	54	5
Alta.Ribagorca..5.	0	0	0	1	1	18	0	0	0
Anoia..6.	1	3	0	2	162	61	9	17	29
Bages..7.	0	3	1	8	19	38	15	141	19
Baix.Camp..8.	0	245	175	2	6	52	5	7	3918
Baix.Ebre..9.	0	11	102	1	2	12	4	0	101
Baix.Emporda..10.	67	0	5	25	1	8	241	0	11
Baix.Llobregat..11.	8	9	10	40	42	155	105	13	287
Baix.Penedes..12.	1	7	8	0	2	12	10	1	688
Barcelones..13.	131	120	163	382	273	1345	1393	123	2205
Bergueda..14.	0	0	0	9	1	4	10	36	1
Cerdanya..15.	1	0	2	52	3	7	3	0	1
Conca.Barbara..16.	0	2	3	0	20	25	1	4	140
Garraf..17.	2	3	7	0	3	11	5	0	101
Garrigues..18.	1	27	3	0	7	245	1	0	19
Garrotxa..19.	300	1	0	159	1	4	167	0	2
Girones..20.	1268	2	0	129	1	36	3331	3	21
Maresme..21.	11	5	4	24	9	22	1608	4	34
Montsia..22.	0	2	9	0	0	3	0	0	44
Noguera..23.	0	0	4	0	49	451	1	8	2
Osona..24.	8	1	0	362	3	18	87	3	13
Pallars.Jussa..25.	0	0	0	1	1	62	0	5	0
Pallars.Sobira..26.	1	1	3	0	0	24	0	3	3
Pla.d.Urgell..27.	0	0	0	0	35	656	1	4	1
Pla.Estany..28.	7345	0	0	13	0	3	62	1	1
Priorat..29.	0	2095	87	0	0	5	0	0	25
Ribera.Ebre..30.	0	217	6113	0	0	41	1	0	115
Ripolles..31.	4	0	0	8980	1	14	5	2	3
Segarra..32.	0	1	2	0	5895	157	4	22	6
Segria..33.	2	13	18	2	117	55346	10	35	73
Selva..34.	91	1	2	18	0	11	34022	5	9
Solsones..35.	0	0	0	0	26	24	3	3871	2
Tarragones..36.	2	224	181	6	3	136	11	1	52114
Terra.Alta..37.	0	2	59	0	0	1	0	0	11
Urgell..38.	0	1	1	0	198	214	0	1	9
Vall.D.Aran..39.	0	0	1	0	1	43	3	4	1
Valles.Occidental..40.	23	19	26	75	45	148	284	28	210
Valles.Oriental..41.	13	0	8	59	7	61	700	8	70

Taula 13: Moviments migratoris obligats pel treball (Taula 4/5) Font: Idescat

	T.Alta37	Urg38	V.Aran39	V.Occ40	V.Ori41
Alt.Camp..1.	2	7	0	16	25
Alt.Emporda..2.	0	4	2	54	25
Alt.Penedes..3.	1	6	0	292	41
Alt.Urgell..4.	1	4	3	12	0
Alta.Ribagorca..5.	0	1	7	1	0
Anoia..6.	2	56	0	199	46
Bages..7.	1	9	1	677	170
Baix.Camp..8.	30	9	0	25	12
Baix.Ebre..9.	67	1	0	16	2
Baix.Emporda..10.	0	0	0	63	59
Baix.Llobregat..11.	15	43	9	6364	1417
Baix.Penedes..12.	9	2	0	76	16
Barcelones..13.	94	270	77	37886	15007
Bergueda..14.	0	0	0	25	20
Cerdanya..15.	1	1	1	4	3
Conca.Barbara..16.	0	30	0	7	0
Garraf..17.	2	3	2	109	44
Garrigues..18.	0	23	0	7	1
Garrotxa..19.	0	0	0	23	9
Girones..20.	0	6	1	132	149
Maresme..21.	3	15	2	578	790
Montsia..22.	11	1	0	14	3
Noguera..23.	2	220	3	12	0
Osona..24.	3	6	1	205	561
Pallars.Jussa..25.	1	10	6	4	0
Pallars.Sobira..26.	0	3	4	7	3
Pla.d.Urgell..27.	0	380	1	7	1
Pla.Estany..28.	0	1	0	11	12
Priorat..29.	4	0	0	1	0
Ribera.Ebre..30.	280	0	0	10	3
Ripolles..31.	0	0	0	13	10
Segarra..32.	5	907	0	21	15
Segria..33.	2	466	53	54	11
Selva..34.	1	5	1	159	632
Solsones..35.	0	4	0	6	2
Tarragones..36.	95	26	4	129	42
Terra.Alta..37.	3434	0	0	1	0
Urgell..38.	2	8858	0	9	4
Vall.D.Aran..39.	0	5	2913	10	1
Valles.Occidental..40.	8	41	11	194624	8989
Valles.Oriental..41.	1	17	2	6843	82507

Taula 14: Moviments migratoris obligats pel treball (Taula 5/5) Font: Idescat

C.2 Descriptiu de les dades dels moviments migratoris obligats pel treball

```
Total<-sum(movmigra)
# movmigra[1:10,1:10]

PerfilsFila<-(movmigra/rowSums((movmigra)))
# PerfilsFila[1:5,1:5]
Mitjana<-mean(diag(PerfilsFila))
Mitjana<-round(Mitjana,2)

Variancia<-var(diag(PerfilsFila))
Variancia<-round(Variancia,5)

# Comarca amb menys moviment migratori
MenysMov<-which.max(diag(PerfilsFila))
rownames(movmigra)[MenysMov]
PctMenysMov<-round(diag(PerfilsFila)[MenysMov]*100,2)

# Comarca amb més moviment migratori
MesMov<-which.min(diag(PerfilsFila))
rownames(movmigra)[MesMov]
PctMesMov<-round(diag(PerfilsFila)[MesMov]*100,2)


# Massa fila
MassaFila<-rowSums(movmigra)/Total
# MassaFila[1:5]
# names(MassaFila)

# Perfil fila mig
colSums(movmigra)/Total
```

C.3 AC sobre la matriu transformada

```
names(ACMatriuQuad(movmigra))
names(ACMatriuQuad(movmigra)$AC)

# Dimensió de la matriu simètrica S
I<-dim(movmigra)[1]

# Inèrcia total
IT<-ACMatriuQuad(movmigra)$AC$IT
```

```
IT<-round(IT,5)

# Inèrcies (totals) principals exlicades
cat("Inercia explicada")
IE<-ACMatriuQuad(movmigra)$AC$InerciaExplicada

cat("Percentatge de la inèrcia explicada")
PctIE<-ACMatriuQuad(movmigra)$AC$PctInerciaExplicada

### Part simètrica ###
# Inèrcia de la part simètrica
IEsim<-IE[1:40]

# Inèrcia total de la matriu simètrica
sum(IEsim)

# Percentatge de la inèrcia de la part simètrica respecte
# la inèrcia total de la matriu transformada
PctIEsim<-round((sum(IEsim)/IT)*100,2)


### Part antisimètrica ###
# Inèrcia de la part antisimètrica
IEantisim<-IE[41:80]

# Percentatge de la inèrcia de la part antisimètrica respecte
# la inèrcia total de la matriu transformada
PctIEantisim<-round((sum(IEantisim)/IT)*100,2)

# Inèrcia explicada pels eixos de les dimensions 41 i 42
Ieix<-sum(IEantisim[1:2])
Ieix<-round((Ieix/sum(IEantisim))*100,3)
```

C.4 Mapa asimètric de la part antisimètrica de les dades transformades

```
### Mapa asimètric. Codi propi ###
# Cal escollir dues dimensions que tingui més inèrcia
# which.max(IEantisim)

# Files en coordenades principals i columnes en coordenades estàndards
# names(ACMatriuQuad(movmigra)£AC)
rpc<-ACMatriuQuad(movmigra)$AC$rpc[1:dim(movmigra)[1],41:42]
#csc<-ACMatriuQuad(movmigra)£AC£csc[1:dim(movmigra)[1],41:42]
```

```r
# plot(rbind(rpc,csc))
#plot(rbind(rpc,csc),xlim=c(-0.19,0.19),ylim=c(-0.19,0.19))
plot(rpc,xlim=c(-0.01,0.1),ylim=c(-0.1,0.1))
# pch opcions dels punts en el plot
text(rpc,rownames(rpc),cex=0.5,col="Blue",pos=1)
#text(csc,rownames(csc),cex=0.5,col="2",pos=1)
abline(v=0, h=0,lty="dotted")

### Moviment migratori 198 d'Urgell(38) cap a la Segarra(32) ###
segments(x0=0,y0=0,x1=rpc[38,1],y1=rpc[38,2])
segments(x0=rpc[38,1],y0=rpc[38,2],x1=rpc[32,1],y1=rpc[32,2])
segments(x0=rpc[32,1],y0=rpc[32,2],x1=0,y1=0)
```

Bibliografia

[1] ”La práctica del análisis de correspondencias”: Michael Greenacre
http://www.fbbva.es/TLFU/tlfu/ing/publicaciones/libros/fichalibro/index.jsp?codigo=300

[2] ”www.carme-n.org”
http://www.carme-n.org/?sec=data

[3] ”An Adaptation of Correspondence Analysis for Square Tables”: Michael Greenacre

[4] ”Robust Late Fusion with Rank Minimization, Supplementary material”: Guangnan Ye, Dong Liu, I-Hong Jhuo, Shih-Fu Chang

[5] ”Análisis de correspondencias: un ejemplo electoral”
http://erre-que-erre-paco.blogspot.com.es/2013/01/analisis-de-correspondencias-un-ejemplo.html

[6] ”Correspondence Analysis in R: The Ultimate Guide for the Analysis, the Visualization and the Interpretation - R software and data mining”
http://www.sthda.com/english/wiki/correspondence-analysis-in-r-the-ultimate-guide-for-the-analysis-the-visualization-and-the-interpretation-r-software-and-data-mining

[7] ”Análisis correspondencias simples y múltiples”: Santiago de la Fuente Fernández
http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf