

AGRUPAMIENTO DE FILAS Y COLUMNAS HOMOGÉNEAS EN MODELOS DE CORRESPONDENCIA

MARINELLI, CLAUDIA* NÉLIDA WINZER†

Recibido/Received: 7 Sep 2002, Versión revisada: 13 Jun 2003

Resumen

Goodman (1981) propone criterios de homogeneidad y estructura para Modelos de Asociación que permiten determinar si pueden agruparse ciertas filas o columnas en una tabla de contingencia. Luego muestra las relaciones entre scores canónicos y los scores correspondientes a Modelos de Asociación. Gilula (1986) desarrolla resultados de agrupamiento sugeridos por los scores canónicos asociados en una tabla de contingencia bajo un modelo de Correlación Canónica RC. Por otra parte, el Análisis de Correspondencia se puede ver como una reparametrización del Modelo de Correlación Canónica en virtud de la propuesta de Goodman (1986) y van der Heijden *et al.* (1994).

Palabras clave: Criterio de Homogeneidad, Correlación Canónica, Inercia, Modelos de Asociación.

Abstract

Goodman(1981) proposed homogeneity and structures criterias in Associaton Models which allow to determine if certain rows or columns in a contingency table should be grouped. In later works, he showed the relations between canonicals scores and the one corresponding to an Association Models. Gilula (1986) developed grouping results suggested by the canonical scores in a contingency table under a RC Canonical Correlation Model. On the other hand, Correspondence Analysis can be seen like a reparametrization of Canonical Correlation Model by virtue of Goodman (1986) and Van der Heijden *et al.* (1994) result.

Keywords: Homogeneity Criterion, Canonical correlation, Inertia, Association Models.

Mathematics Subject Classification: 62H25.

*NUCOMPA, Universidad Nacional del Centro de la Provincia de Buenos Aires. Tandil, Argentina.
Tel: +54-2293-447104 Int. 412. E-Mail: cmarine@exa.unicen.edu.ar

†Universidad Nacional del Sur, Bahía Blanca, Argentina.

1 Introducción

Los agrupamientos apropiados de filas y columnas de una tabla de contingencia a dos vías o a vías múltiples podrían simplificar el Análisis de Asociación entre dos o más variables aleatorias categóricas, puesto que se reduce entre otras cosas, el número de parámetros del modelo.

Otro motivo para generar agrupamientos es que un modelo puede no ajustar bien una tabla original pero si hacerlo sobre una subtabla formada por los subconjuntos de filas y/o columnas. Por otra parte, los test χ^2 no son recomendables cuando las celdas tienen valores esperados bajos, lo que puede ser solucionado utilizando agrupamientos.

Leo A. Goodman [5] y más tarde Zvi Gilula ([2] y [3]), discuten la relación entre scores canónicos y scores correspondientes a Modelos de Asociación. Goodman especifica bajo qué circunstancias los scores canónicos estimados podrían ser similares, en magnitud, a los correspondientes scores de asociación estimados bajo un modelo estructural específico. Esto les permite probar que los agrupamientos, en un Modelo de Asociación, bajo el criterio de homogeneidad, podrían ser sugeridos por los scores canónicos estimados.

Peter van der Heijden [12] propone una fórmula de reconstitución de la matriz de probabilidad en un Análisis de Correspondencia. Esta fórmula permite ver que este análisis descompone la matriz de probabilidad en dos componentes: uno debido a la independencia y otro a los residuos.

Utilizando esta descomposición, Goodman [7] y van der Heijden *et al* [13], muestran al Análisis de Correspondencia como una reparametrización del Modelo Canónico de Fisher, al cual denominan Modelo de Correspondencia RC o saturado.

En este artículo se presenta en primer lugar el Modelo de Correspondencia mostrando su relación con un Modelo Canónico. Luego a partir de la identidad entre ambos modelos se establecen los criterios de agrupamiento bajo un Modelo de Correspondencia RC o saturado.

2 Modelo de correspondencia

Sea Π de dimensión $I \times J$ la matriz de elementos $\{\pi_{ij}\}$. La *inercia poblacional* [9] se define como;

$$In = In(I) = In(J) = \sum_{i=1}^I \sum_{j=1}^J \frac{(\pi_{ij} - \pi_{i.}\pi_{.j})^2}{\pi_{i.}\pi_{.j}} \quad (1)$$

donde $\pi_{i.}$ y $\pi_{.j}$ representan las probabilidades marginales fila y columna, respectivamente.

Por lo tanto, la inercia poblacional es la suma del cuadrado de los residuales q_{ij} :

$$q_{ij} = \frac{(\pi_{ij} - \pi_{i.}\pi_{.j})}{\sqrt{\pi_{i.}\pi_{.j}}}$$

Si r y c son los vectores de los marginales fila y columna, respectivamente, realizando la descomposición del valor singular generalizado para la matriz $(\Pi - rc')$, se tiene, $(\Pi - rc') = AD_\lambda B'$ donde $A'D_r^{-1}A = I = B'D_c^{-1}B$ con $D_r = \text{diag}\{r\}$ y $D_c = \text{diag}\{c\}$ [11].

La matriz de residuales $Q = \{q_{ij}\}$ puede construirse a partir de $(\Pi - rc')$ pre y postmultiplicando por $D_r^{-\frac{1}{2}}$ y $D_c^{-\frac{1}{2}}$ respectivamente. Por consiguiente, si E es la matriz cuyos elementos tienen la forma, $e_{ij} = \pi_i \cdot \pi_{.j}$,

$$Q = D_r^{-\frac{1}{2}}(\Pi - E)D_c^{-\frac{1}{2}} = D_r^{-\frac{1}{2}}(\Pi - rc')D_c^{-\frac{1}{2}} = D_r^{-\frac{1}{2}}AD_\lambda B'D_c^{-\frac{1}{2}} = UD_\lambda V', \quad (2)$$

con $U = D_r^{-\frac{1}{2}}A$ y $V = D_c^{-\frac{1}{2}}B$.

En esta descomposición U es de orden $I \times H$, V es de orden $J \times H$ y D_λ es no singular de orden $H \times H$, con $H \leq \min(I - 1, J - 1)$. Además $U'U = I = V'V$.

Las filas de $UD_\lambda^{-\frac{1}{2}}$, constituyen las coordenadas que permiten representar los perfiles filas y pueden ser vistas como ponderaciones de las columnas [1]. Análogamente, las columnas pueden representarse mediante las filas de $VD_\lambda^{-\frac{1}{2}}$. Si se definen $R^* = D_r^{-1}A$ y $C^* = D_c^{-1}B$, de la descomposición de Q se obtienen las fórmulas de transición, que muestran la posibilidad de expresar las ponderaciones de filas en función de columnas y viceversa, del siguiente modo [9]:

$$\begin{aligned} R^* &= D_r^{-1}\Pi C^* D_\lambda^{-1} \\ C^* &= D_c^{-1}\Pi' R^* D_\lambda^{-1} \end{aligned} \quad (3)$$

Reemplazando R^* y C^* en la descomposición del valor singular (2) resulta,

$$\Pi = E + \Pi C^* D_\lambda^{-1} R^* \Pi \quad (4)$$

Por otra parte, las coordenadas de los perfiles filas y columnas respecto a los ejes principales en un Análisis de Correspondencia están dadas por los elementos $\{x'_{ik}\}$ de $D_r^{-1}\Pi C^*$ y $\{y'_{jk}\}$ de $D_c^{-1}\Pi' R^*$ [8]. Entonces si m es el número de autovalores distintos de cero, cada elemento de Π puede expresarse del siguiente modo:

$$\pi_{ij} = e_{ij} \left(1 + \sum_{k=1}^m \frac{x'_{ij} y'_{ij}}{\lambda_k} \right) \quad (5)$$

Esto permite modelar cada celda en una tabla de Contingencia bajo un Análisis de Correspondencia.

3 Inercia y Correlación

Con el objeto de explicitar la relación entre la *inercia* y la *correlación canónica* se realizará una reparametrización del Modelo Canónico:

Si $m \leq \min(I - 1, J - 1)$ el Modelo de Correlación Canónica C_m para una tabla de contingencia $I \times J$, puede expresarse como [4]:

$$\pi_{ij} = \pi_i \cdot \pi_{.j} \left(1 + \sum_{k=1}^m \rho_k x_{ik} y_{jk} \right) \quad (6)$$

donde,

$$\begin{aligned} \sum_{i=1}^I x_{ik}\pi_i &= 0 & \sum_{i=1}^I x_{ik}^2\pi_i &= 1 \\ \sum_{j=1}^J y_{jk}\pi_{\cdot j} &= 0 & \sum_{j=1}^J y_{jk}^2\pi_{\cdot j} &= 0 \\ \sum_{i=1}^I x_{ik}x_{ik'}\pi_i &= 0 & \sum_{j=1}^J y_{jk}y_{jk'}\pi_{\cdot j} &= 0 \end{aligned} \quad (7)$$

con $k \neq k'$. Los parámetros ρ_k en la expresión (6) son la medida de correlación entre x_{ik} e y_{jk} , ya que de las fórmulas (6) y (7) resulta:

$$\sum_{i=1}^I \sum_{j=1}^J x_{ik}y_{jk}\pi_{ij} = \rho_k \quad (8)$$

y los x_{ik} e y_{jk} son los k -ésimos scores fila y columna para la celda (i, j) . El modelo (6) es el Modelo de Correlación Canónica Cm ($\rho_k \neq 0, \forall k \leq m$). Sin pérdida de generalidad, los ρ_k pueden ordenarse del siguiente modo $1 \geq \rho_1 \geq \rho_2 \geq \dots \geq \rho_m \geq 0$.

Bajo el Modelo de Correlación Canónica RC los parámetros x_{i1} e y_{i1} , estandarizados maximizan la correlación ρ_1 . Asimismo, los parámetros x_{i2} e y_{i2} , con la misma estandarización, maximizan la correlación ρ_2 y así hasta la correlación m -ésima. Los scores x_{ik} e y_{jk} están incorrelados con los $k-1$ anteriores.

De las expresiones (8) y (7) y teniendo en cuenta que la correlación entre dos scores consecutivos es cero se tiene la siguiente identidad:

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \frac{(\pi_{ij} - \pi_{i\cdot}\pi_{\cdot j})^2}{\pi_{i\cdot}\pi_{\cdot j}} &= \sum_{i=1}^I \sum_{j=1}^J \pi_{i\cdot}\pi_{\cdot j} \left(\sum_{k=1}^m \rho_k x_{ik}y_{jk} \right)^2 = \\ &= \sum_{k=1}^m \rho_k^2 \sum_{i=1}^I (x_{ik}^2\pi_i) \sum_{j=1}^J (y_{jk}^2\pi_{\cdot j}) = \sum_{k=1}^m \rho_k^2. \end{aligned}$$

Así, cada componente ρ_k^2 ($k = 1, \dots, m$) se divide a su vez en dos subcomponentes: una contribución relativa a las filas ($x_{ik}^2\pi_i$) y otra a las columnas ($y_{jk}^2\pi_{\cdot j}$).

P. van der Heijden [12], propone en primer lugar la fórmula de reconstitución (5) de la matriz de probabilidad en un Análisis de Correspondencia, la misma permite ver que este análisis descompone la matriz de probabilidad en dos componentes: uno debido a la independencia y otro a los residuos. Utilizando esta descomposición, Goodman [7] y van der Heijden *et al* [13] proponen el Modelo de Correspondencia RC o saturado:

$$\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j} \left(1 + \sum_{k=1}^m \frac{x'_{ik} y'_{jk}}{\rho_k} \right) \quad (9)$$

donde,

$$x'_{ik} = x_{ik}\rho_k \quad y'_{jk} = y_{jk}\rho_k \quad (10)$$

De este modo, en virtud de la expresión (5) y de las fórmulas de transición entre los pesos de filas y de columnas (3), el Análisis de Correspondencia puede verse como una reparametrización del Modelo de Correlación Canónica. Y a partir de esta relación es posible analizar propiedades del Modelo de Correspondencia a través de su relación con el de Correlación Canónica.

De las expresiones (7) y (10) se obtienen las restricciones sobre x'_{ik} e y'_{jk} que se denominarán a partir de ahora “*Scores de Correspondencia*”:

$$\begin{aligned} \sum_{i=1}^I x'_{ik} \pi_{i \cdot} &= 0 & \sum_{i=1}^I x'_{ik} x'_{ik'} \pi_{i \cdot} &= \rho_k^2 \delta_{kk'} \\ \sum_{j=1}^J y'_{jk} \pi_{\cdot j} &= 0 & \sum_{j=1}^J y'_{jk} y'_{jk'} \pi_{\cdot j} &= \rho_k^2 \delta_{kk'} \end{aligned} \quad (11)$$

A partir de (8) y (11), ρ_k en (9) es una medida de correlación entre x'_{ik} e y'_{jk} . Además de (9) y (11) se tiene que,

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J \frac{(\pi_{ij} - \pi_{i \cdot} \pi_{\cdot j})^2}{\pi_{i \cdot} \pi_{\cdot j}} &= \sum_{i=1}^I \sum_{j=1}^J \pi_{i \cdot} \pi_{\cdot j} \left(\sum_{k=1}^m \frac{x'_{ik} y'_{jk}}{\rho_k} \right)^2 = \\ &= \sum_{k=1}^m \frac{1}{\rho_k^2} \sum_{i=1}^I (x'_{ik} \pi_{i \cdot}) \sum_{j=1}^J (y'_{jk} \pi_{\cdot j}) = \sum_{k=1}^m \rho_k^2. \end{aligned}$$

De este modo, esta medida de alejamiento de la independencia, *inercia*, puede partitionarse en m componentes ρ_k^2 , para $k = 1, \dots, m$ y cada componente ρ_k^2 puede a su vez partitionarse en dos contribuciones: $x'_{ik} \pi_{i \cdot}$ de la i -ésima fila ($i = 1, \dots, I$) e $y'_{jk} \pi_{\cdot j}$ de la j -ésima columna ($j = 1, \dots, J$).

Dado que en un Análisis de Correspondencia la medida de alejamiento de la independencia es la suma de las inercias principales, de las descomposiciones anteriores, la *correlación canónica* al cuadrado puede ser vista como una *inercia principal*.

4 Agrupamiento en tablas de contingencia a dos vías bajo un Modelo de Correspondencia

Dada la matriz de probabilidad con elementos $\Pi = (\pi_{ij})$ de dimensión $(I \times J)$, dos columnas j y h de Π se dicen *homogéneas* si $\frac{\pi_{ij}}{\pi_{\cdot j}} = \frac{\pi_{ih}}{\pi_{\cdot h}}$. Análogamente se define la homogeneidad entre filas.

Se han definido dos criterios para determinar cuándo pueden agruparse las filas o columnas en una tabla de contingencia: el de *homogeneidad* [6] y el de *estructura* [4].

El criterio de homogeneidad establece que 2 ó más filas (columnas) pueden combinarse si dichas filas (columnas) son homogéneas. Zvi Gilula [4] demuestra que, bajo Modelos Canónicos, 2 o más filas (columnas) son homogéneas si los correspondientes conjuntos de scores son iguales.

Por el principio de equivalencia distribucional [9], el agrupamiento de columnas (filas) homogéneas no cambia la geometría de las filas(columnas) en el sentido que se conservan las distancias χ^2 entre ellas.

En base a lo expuesto por Gilula y teniendo en cuenta la reparametrización de van der Heijden, se desarrollará aquí una extensión del método de agrupamiento cuando se supone válido un Modelo de Correspondencia. Aunque se expondrán los resultados sólo para las columnas su aplicación a las filas se hace en forma totalmente análoga.

Teorema 1 *Dada una tabla de contingencia con matriz de probabilidad $\Pi = (\pi_{ij})$ bajo un Modelo de Correspondencia RC o saturado, las columnas j y h son homogéneas si y sólo si los pesos y'_{jk} e y'_{hk} definidos en (10), son iguales $\forall k$, con $1 \leq k \leq m$.*

DEMOSTRACIÓN: Bajo un Modelo de Correspondencia RC se verifica que

$$\pi_{ij} = \pi_{i \cdot} \pi_{\cdot j} \left(1 + \sum_{k=1}^m \frac{x'_{ik} y'_{jk}}{\rho_k} \right)$$

Suponiendo que $\pi_{ij} > 0$ para todo i, j se tiene que:

$$\frac{\pi_{ij}}{\pi_{ih}} = \frac{\pi_{\cdot j} \left(1 + \sum_{k=1}^m \frac{x'_{ik} y'_{jk}}{\rho_k} \right)}{\pi_{\cdot h} \left(1 + \sum_{k=1}^m \frac{x'_{ik} y'_{hk}}{\rho_k} \right)} \quad (12)$$

Si las columnas j y h son homogéneas vale que $\frac{\pi_{ij}}{\pi_{\cdot j}} = \frac{\pi_{ih}}{\pi_{\cdot h}} \forall i$, con $1 \leq i \leq I$, por lo tanto,

$$\frac{\pi_{\cdot j} \left(1 + \sum_{k=1}^m \frac{x'_{ik} y'_{jk}}{\rho_k} \right)}{\pi_{\cdot h} \left(1 + \sum_{k=1}^m \frac{x'_{ik} y'_{hk}}{\rho_k} \right)} = 1$$

lo que implica que $\sum_{k=1}^m \frac{x'_{ik} (y'_{jk} - y'_{hk})}{\rho_k} = 0$ para $1 \leq i \leq I$.

Usando las identidades (10), esta expresión es equivalente a:

$$\sum_{k=1}^m x_{ik} (y_{jk} - y_{hk}) \rho_k \quad \text{para } 1 \leq i \leq I$$

Por definición de los scores canónicos y del modelo C_m (Gilula, [4] y [3]), los vectores $x_k = [x_{ik} : 1 \leq i \leq I]$ para $1 \leq k \leq m$ son linealmente independientes y $\rho_k > 0$ para todo k , entonces $\sum_{k=1}^m x_{ik} (y_{jk} - y_{hk}) \rho_k = 0$ si y sólo si $y_{jk} = y_{hk}$ para $1 \leq k \leq m$. Entonces $y_{jk} \rho_k = y_{hk} \rho_k$ para $1 \leq k \leq m$ o equivalentemente, $y'_{jk} = y'_{hk}$ con $1 \leq k \leq m$, que es lo que se quería mostrar.

La recíproca resulta trivial. ■

Sea $\Pi^{(1)}$ la matriz obtenida a partir de Π por agrupamiento de todas sus filas y columnas homogéneas bajo un Modelo de Correspondencia RC o saturado. De Gilula [4] y sabiendo que la forma canónica de una distribución bivariada es única por construcción (Lancaster, [10]) ambas matrices mantienen la misma forma canónica.

Si para una tabla de contingencia N , un subconjunto particular de scores de correspondencia son parecidos en todos los m conjuntos de scores, el teorema anterior sugiere que N podría agruparse sobre esas columnas y/o filas particulares obteniendo una tabla condensada $N^{(1)}$ de dimensión $I_1 \times J_1$. Si $L^2(N)$ es el estadístico de razón de verosimilitud para testear el alejamiento de la independencia en una tabla de contingencia N , Gilula [4] propone utilizar la diferencia $L_d^2(N) = L^2(N) - L^2(N^{(1)})$ como un test estadístico para probar la validez de un agrupamiento. Si este estadístico es pequeño indicaría que el agrupamiento es consistente con los datos.

Por la aditividad del estadístico L^2 , L_d tiene una distribución χ^2 con $(I - 1)(J - 1) - (I_1 - 1)(J_1 - 1)$ g.l.

Ejemplo

Se considera aquí una tabla de contingencia $N_{(6 \times 4)}$ (Tabla 1) referida a los niveles socioeconómicos y de salud mental en 1660 familias, la misma ha sido analizada en diferentes contextos por Goodman ([5] y [6]) y Gilula [4], entre otros (A, B, C, D, E y F en la Tabla 1 representan los niveles socioeconómicos).

Niveles de Salud Mental	A	B	C	D	E	F
Bueno (BU)	64	57	57	72	36	21
Síntoma Leve (SL)	94	94	105	141	97	71
Síntoma Moderado (SM)	58	54	65	77	54	54
Caso Severo (CS)	46	40	60	94	78	71

Tabla 1: Tabla de contingencia que cruza los niveles socioeconómicos y de salud mental en 1660 familias.

A través de este ejemplo se pretende mostrar la existencia de posibles agrupamientos, comparar los resultados obtenidos bajo un modelo de Correlación Canónica y los obtenidos bajo uno de correspondencia y de esta manera corroborar las analogías presentadas en apartados anteriores.

El modelo de correlación canónica $C_1 : (\rho_k \neq 0 : k \leq 1)$ es un buen ajuste a la Tabla 1, $L_2(N) = 2.77$ con 8 g.l. En la Tabla 2 se muestran las estimaciones de la correlación canónica y los scores fila y columna y en la Tabla 3 se dan los resultados de un análisis de correspondencia considerando sólo el primer autovalor. Relacionando la información contenida en las Tabla 2 y 3 se confirman las analogías entre los dos modelos: $\lambda = \rho^2$ y cada score de correspondencia es igual al producto de la correlación canónica ρ y el respectivo score canónico (10).

A partir de los scores de correspondencia en la Tabla 3, se observa que los scores correspondientes a las columnas A y B son aproximadamente iguales. Por el teorema del apartado 4 se puede concluir que pueden ser agrupadas. La Tabla 4 representa la tabla de contingencia condensada $N^{(1)}$, luego de agrupar las dos primeras columnas. Por otra parte se llevó a cabo un análisis de correspondencia con la tabla de contingencia condensada $N^{(1)}$ (Tabla 5). Si se realiza el test propuesto por Gilula, se confirma la validez del

Correlación canónica	Scores canónicos fila	Scores canónicos columna
0.1613	1.6087	1.1216
	0.1837	1.1467
	-0.0880	0.3658
	-1.4715	-0.0550
		-1.0252
		-1.7833

Tabla 2: Estimaciones de la correlación canónica y los scores fila y columna.

$\lambda = \rho^2$	Scores fila de correspondencia	Scores columna de correspondencia
0.026202	0.2595	0.1809
	0.0296	0.1850
	-0.0142	0.0590
	-0.2374	-0.0089
		-0.1654
		-0.2877

Tabla 3: Resultados del análisis de correspondencia para el primer autovalor.

agrupamiento: ($L^2(N) = 47.42$, $L^2(N^{(1)}) = 42.02$), con 15 y 12 g.l. respectivamente, por lo tanto $L_d^2(N) = 0.397$).

Niveles de Salud Mental	A+B	C	D	E	F
Bueno (BU)	121	57	72	36	21
Síntoma Leve (SL)	188	105	141	97	71
Síntoma Moderado (SM)	112	65	77	54	54
Caso Severo (CS)	86	60	94	78	71

Tabla 4: Tabla de contingencia condensada $N^{(1)}$ agrupando las dos primeras columnas.

Se graficaron las filas y las columnas en el plano de los dos primeros ejes de correspondencia de la tabla original (Gráfico 1) y de la tabla condensada (Gráfico 2). Los ejes han sido usados al sólo efecto de visualizar con más detalle la distribución de los puntos, aunque el segundo eje tiene un valor de inercia mucho menor que el primero ($\lambda_2 = 0.0014$).

En estos gráficos se ve claramente que las categorías A y B tienen un mismo comportamiento y que el agrupamiento $A + B$, no modifica el comportamiento de las restantes categorías socioeconómicas ni de los niveles de salud mental.

$\lambda = \rho^2$	Scores fila de correspondencia	Scores columna de correspondencia
0.029110	0.2634	0.1862
	0.0345	0.0629
	-0.0116	-0.0058
	-0.2476	-0.1616
		-0.3052

Tabla 5: Resultados del análisis de correspondencia para la tabla agrupada.

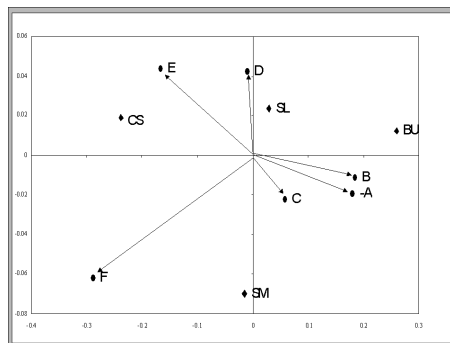


Gráfico 1: Primer plano principal del análisis de correspondencia sobre N .

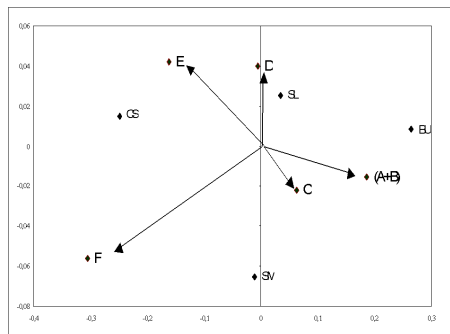


Gráfico 2: Primer plano principal del análisis de correspondencia sobre $N^{(1)}$.

5 Conclusiones

El método de agrupamiento planteado es un método exploratorio, ya que parte de la observación de los scores de correspondencia estimados. Si un conjunto particular de scores de dos filas o dos columnas, son cercanos en los m conjuntos de scores, a partir del resultado anterior puede modificarse la tabla de contingencia N agrupando tales filas o columnas obteniendo una nueva tabla $N^{(1)} (I_1 \times J_1)$, eligiendo m de tal forma que se recupere un porcentaje de inercia razonable.

Si se trabaja con los dos tipos de análisis se tiene la ventaja de tener, a partir de un Análisis de Correlación Canónica un test que permite precisar m y utilizar por otro lado la potencia gráfica de Análisis de Correspondencia.

Referencias

- [1] Gabriel, K. R. (1971) “The biplot graphic display of matrices with applications to principal component analysis”, *Biometrika* **58**(3): 453ss.
- [2] Gilula, Z.; Haberman S. (1986) “Canonical analysis of contingency tables by maximum likelihood”, *J.A.S.A.* **81**: 395ss.
- [3] Gilula, Z.; Haberman S. (1988) “The analysis of multivariate contingency tables by restricted canonical and restricted associations models”, *J.A.S.A.* **83**: 403ss.
- [4] Gilula, Z. (1986) “Grouping and associations in contingency tables”, *J.A.S.A.* **81**: 395ss.
- [5] Goodman, L. (1981) “Association models and canonical correlation in the analysis of cross-classifications having ordered categories”, *J.A.S.A.* **76**: 374ss.
- [6] Goodman, L. (1985) “The analysis of cross-classified data”, the 1983 Henry L. Rietz Memorial Lecture, *The Annals of Statistics* **13**(1): 10–69.
- [7] Goodman, L. (1986) “Some useful extensions of usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables”, *International Statistical Review* **54**(3): 243–309.
- [8] Greenacre, M. (1984) *Theory and Applications of Correspondence Analysis*. Academic Press, New York.
- [9] Greenacre, M.; Blasius, J. (1994) *Correspondence Analysis in the Social Sciences*. Academic Press, London.
- [10] Lancaster, A. (1979) *The Chi-Squared Distributions*. Wiley, New York.
- [11] Mardia, K.; Kent, J.; Bibby, J. (1979) *Multivariate Analysis*. Academic Press, London.
- [12] van der Heijden P.; Leeuw J. de (1985) “Correspondence analysis used complementary to loglinear analysis”, *Psychometrika* **50**: 429–447.
- [13] van der Heijden, P.; Mooijaart, A.; Takane, Y. (1994) “Correspondence analysis and contingency table models in correspondence analysis in the social sciences”, M. Greenacre & J. Blasius (Eds.) *Correspondence Analysis in the Social Sciences*, Academic Press, London: 370ss.