Análisis de correspondencias

Eliseo Martínez H.

1. Elecciones en París

Hemos decidido presentar un legendario ejemplo para explicar el objetivo del Análisis de Correspondencia. Este ejemplo se encuentra en el *libro Introducction a L'Analyse des Données* realizado bajo la dirección de Georges Morlat et al., y curiosamente editado por la SOCIETE de MATHEMATIQUES APPLIQUES et de SCIENCES HUMAINES, 1976

Los resultados en París de una elección presidencial se pueden resumir en una "tabla de contingencia" con p filas y q columnas.

		$\longleftarrow$	distritos	$\longrightarrow$	
			j		
$\uparrow$					
candidatos	i		$m_{i\,j}$		
$\downarrow$					

El número entero $m_{i\,j}$ denota el número de votos obtenido en el distrito "j" por el candidato "i".

Se considera aquí:

- tres conjuntos: el conjunto I de los candidatos; el conjunto J de los distritos; el conjunto K de los votantes
- dos características cualitativas: el carácter "candidato" denotado por x; el carácter "distrito" denotado por y, donde los conjuntos de las modalidades son respectivamente I y J.

La tabla de contingencia describe los efectos de las clases de la partición inducida sobre K por la aplicación

$$\begin{aligned} K &\longrightarrow I \times J \\ k &\longrightarrow (x(k), y(k)) = (i, j) \end{aligned}$$

en que a todo votante se asocia un candidato y un distrito.

Nos interrogamos aquí:

- sobre el electorado de los diferentes candidatos
- sobre el modo de votar en los diferentes distritos de París.

¿Los candidatos " i " y "i' " son ellos electoralmente comparables?

¿Los distritos " j " y " j' " están "próximos" en cuanto a la destinación de la papeleta del voto?

Como podemos deducir en este ejemplo, el análisis de correspondencia tiene como objetivo dar respuesta a las dos interrogantes emanadas de este ejemplo, basada en la idea geométrica de "proximidad" ya sea entre los candidatos o los distritos. Y las herramientas matemáticas que elaboraremos se generarán de la tabla de contingencia, y estableceremos una equivalencia con las componentes principales y coordenadas principales para variables cualitativas. Por otro lado, una tabla de contingencia está asociada a las frecuencias conjuntas de dos variables cuantitativas, sin embargo, como lo veremos en el próximo ejemplo, la técnica del análisis de correspondencia se puede utilizar en otro tipo de tablas bidimensionales.

2. Presupuestos de los países

Los presupuestos diseñados para diferentes países de la Europa pueden ser descritos en una tabla de p filas y q columnas.

		$\longleftarrow$	sectores	$\longrightarrow$	
			j		
$\uparrow$					
países	i		$m_{i\,j}$		
$\downarrow$					

El número $m_{i\,j}$ es la cantidad, expresada en millones de francos[1], asignada al sector " j " por el país " i " como presupuesto sectorial.

Nos interrogamos aquí:

- sobre la manera de establecer el presupuesto en los diferentes países,
- sobre la contribución de los países a los diferentes presupuestos sectoriales.

¿Dos estados " i " y "i' " son politicamente equivalentes en cuanto a la asignación de su presupuesto?

¿Dos sectores están "próximos" en cuanto a las sumas asignadas por los diferentes estados?

Nota: los números $m_{i\,j}$ no son en general enteros, pero haciendo un cambio adecuado de escala, se pueden considerar como cantidades enteras.

Se considera aquí:

- tres conjuntos: el conjunto de los países de Europa; el conjunto de los sectores de prsupuesto para el estado; el conjunto K de francos.
- dos características cualitativas: el carácter "país" denotado por x; el carácter "sector

[1] Si queremos actualizar el ejemplo, obtenido del mismo libro citado, deberíamos expresar el presupuesto en euros.

presupuestario" denotado por y, donde los conjuntos de las modalidades son respéctivamente I y J.

Los valores de $m_{i\,j}$, expresados en francos, son los efectos de las clases de partidción inducidas sobre K por la aplicación (x, y).

En los dos ejemplos considerados, los elementos (i, j) del producto cartesiano $I \times J$ están asociados a la "masa" $m_{i\,j}$; en efecto, si A es un subconjunto de $I \times J$ parece lógico, tanto para el primero como el segundo ejemplo, asociar la masa:

$$m(A) = \sum \{m_{i\,l} \ / \ (i, j) \in A\}$$

donde $m(A)$ representa por ejemplo la cantidad en millones de francos asignadas por Francia y Polonia a la investigación y a la construcción.

En ambos casos, se ha asociado al conjunto finito $I \times J$ de $p \times q$ elementos una medida positiva; en ambas situaciones el análisis (factorial) de correspondencia es una técnica excelente:

- para precisar simultáneamente la proximidad entre elementos de I y entre elementos de J.
- para precisar la estructura de dependencia entre los caracteres cualitativos x e y.

3. Construcción de una tabla de contingencia

Supongamos que queremos medir a una cantidad de individuos o unidades muestrales dos caracterìsticas cualitativas, de tal menara que la primera característica puede asumir I valores, y la segunda característica puede asumir J valores, de tal manera que una unidad eventualmente puede tomar el par de valores dentro del conjunto $I \times J$. Para fijar ideas consideremos una tradicional tabla de contingencia presentada por Sir Ronald Fisher en 1940, que presenta la clasificación de 5387 escolares escoceses según su color de pelo y color de ojos. En este ejemplo los posibles valores para el color de los ojos son {claros, azules, castaños, oscuros}de manera que $I = 4$, y para el color del pelo tenemos los siguientes resultados {rubio, pelirrojo, castaño, oscuro, negro} y en este caso $J = 5$. La tabla de contingencia elaborada por Fisher se entrega en la Tabla 1.

Para definir el color de los ojos utilizaremos variables binarias con la siguiente codificación

$$\begin{aligned} claros &= (1, 0, 0, 0) \\ azules &= (0, 1, 0, 0) \\ castaños &= (0, 0, 1, 0) \\ oscuros &= (0, 0, 0, 1) \end{aligned}$$

y con esto, entonces, podemos construir una matriz binaria de dimensión 5387×4, que llamaremos $\mathbf{X}_a$ y cuya fila indicará un determinado escolar y mediante sus columnas de

manera unívoca determinará el color de los ojos de tal escolar, esto es a modo de ejemplo

$$\mathbf{X}_a = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

indicando con esto que el primer escolar tiene los ojos azules, el segundo escolar los tiene claros, el penúltimo escolar tiene los ojos castaños, y el último los tiene oscuros.

De manera similar codificamos el color del pelo mediante variables binarias como sigue,

$$\begin{aligned} rubio &= (1,0,0,0,0) \\ pelirrojo &= (0,1,0,0,0) \\ castaño &= (0,0,1,0,0) \\ oscuro &= (0,0,0,1,0) \\ negro &= (0,0,0,0,1) \end{aligned}$$

y así podemos generar una matriz $\mathbf{X}_b$de 5387×5 que entregará la información de manera unívoca sobre el color del pelo de cada uno de los escolares escoceses. Ahora si realizamos el producto $\mathbf{X}_a^t \mathbf{X}_b$ obtenemos esencialmente la Tabla 1, puesto que dicho producto nos entregará la suma de todas las personas que tienen un determinado par de caracterìsticas.

C. ojos	Color del pelo					
	rubio	pelirrojo	castaño	oscuro	negro	total
claros	688	116	584	188	4	1580
azules	326	38	241	110	3	718
castaños	343	84	909	412	26	1774
oscuros	98	48	403	681	85	1315
total	1455	286	2137	1391	118	5387

$Tabla 1$

4. La matriz de frecuencias condicionadas por filas $\mathbf{R}$

La matriz $\mathbf{X}_a^t \mathbf{X}_b$,o en nuestro caso la matriz de la Tabla 2 se divide por el número n de casos observados, y de esta forma obtener la matriz de frecuencias relativas $\mathbf{F} = (f_{i\,j})_{I \times J}$, donde $f_{i\,j}$ es la frecuencia relativa asociada a la $i-$ésima fila y $j-$ésima columna.. Entonces

$$\sum_{i=1}^{I} \sum_{j=1}^{J} f_{i\,j} = 1$$

Desde el punto de vista de la teorìa de la probabilidad, la matriz $\mathbf{F}$ define una probabilidad sobre el espacio $I \times J$ de la manera siguiente

$$P_{I\,J}\{(i,\,j)\} = f_{i\,j} \tag{1}$$

cuya interpretación es elemental, por ejemplo de la distribución de los escolares escoceses según el color de los ojos y color del pelo, la probabilidad de encontrar, entre esos 5387 escolares escoceses, un escolar con los ojos azules y el pelo negro es de

$$P_{IJ}\{(2,5)\} = \frac{3}{5387} = 0.0005569$$

Y además la probabilidad en (1) nos permite definir dos tipos de probabilidades, a saber

$$P_J\{i\} = \sum_{j=1}^{J} f_{i\,j} = f_{i\,\cdot} \ ; \ \forall \ i \in I \tag{2}$$

$$P_I\{j\} = \sum_{i=1}^{I} f_{i\,j} = f_{\cdot\,j} \ ; \ \forall \ j \in J \tag{3}$$

Interpretando estas probabilidades como: P_J es la probabilidad sobre el conjunto finito I (probabilidad marginal de I); y P_I es la probabilidad sobre el conjunto J (probabilidad marginal de J).

En el caso de nuestro ejemplo, sobre los escolares escoceses , la matriz **F** es obtenida de la siguiente tabla:

C. ojos	Color del pelo				
	rubio	pelirrojo	castaño	oscuro	negro
claros	0.124	0.021533	0.10841	0.034899	0.00074253
azules	0.060516	0.07054	$0.0\,4737$	$0.0\,2042$	0.0005569
castaños	0.063672	0.015593	0.16874	$0.0\,7648$	0.0048264
oscuros	0.018192	0.0089103	0.07481	0.12642	0.015779

$Tabla\ 2$

Haremos el análisis que sigue para las filas I de la matriz **F**, no obstante el mismo análisis se puede efctuar para las columnas de **F**, será simétrico y equivalente toda vez que es arbitraria la eleción de fila y columna asociada a las dos variables cualitativas.

Las filas de la matriz **F** se deben considerar como I puntos en el espacio $\mathbb{R}^J$. Y al igual que el análisis en componentes principales vamos intentar representar estos I puntos en un espacio de dimensión inferior que nos permita apreciar sus distancias relativas, es decir se dará respuesta a la primera interrogante anunciada en los dos ejemplos precedentes.

Debemos considerar dos tipos de inconvenientes en las filas de la matriz **F**, que son:

- Las filas de la matriz **F**, como elementos de $\mathbb{R}^J$, no (siempre) tienen el "mismo peso". A modo de ejemplo, podemos observar que las filas de las frecuencias de la Tabla 2 para los ojos claros y azules han sido generadas de universos sobre datos muy dispares donde los datos para los escolares de ojos claros supera largamente a los escolares escoceses de ojos azules.
- La distancia euclídea como una medida de proximidad no siempre es la más adecuada, por lo que presentaremos otro tipo de distancia.

Cada fila de la matriz F, como lo establace la ecuación en (2), tiene asignada una medida o frecuencia relativa, esta es $f_{i\,\cdot} = \sum_j f_{i\,j}$, que se puede obtener matricialmente como

$$\mathbf{f} = \mathbf{F}^{\,t}\,\mathbf{1}$$

donde las entradas del vector $\mathbf{f}$ son precisamente los valores $f_{i\,\cdot}$. Luego podemos dar a cada fila un peso proporcional a su frecuencia relativa, de otra forma considerar sobre cada fila la distribución condicionada. Analíticamente, crearemos una matriz $\mathbf{R} = (r_{i\,j})$ que tendrá las siguientes entradas:

$$r_{i\,j} = \frac{f_{i\,j}}{f_{i\,\cdot}} \tag{4}$$

Y esta nueva matriz si que tendrá el mismo peso puesto que

$$\sum_{j=1}^{J} r_{i\,j} = \sum_{j=1}^{J} \frac{f_{i\,j}}{f_{i\,\cdot}} = 1 \;;\; i = 1, \ldots, I$$

Matricialmente $\mathbf{R}$ se forma como sigue:

$$\mathbf{R} = \mathbf{D}_f^{-\,1}\mathbf{F}$$

donde $\mathbf{D}_f = diag\{f_{1\,\cdot}, \ldots, f_{I\,\cdot}\}$ es la matriz diagonal de orden $I \times I$ cuyos elementos sobre la diagonal son los elementos del vector $\mathbf{f}$.

Para nuestro ejemplo de los escolares escoceses la nueva tabla quedaría como sigue:

C. ojos	Color del pelo					
	rubio	pelirrojo	castaño	oscuro	negro	total
claros	0.435	0.073	0.369	0.119	0.003	1
azules	0.454	0.053	0.336	0.153	0.004	1
castaños	0.193	0.047	0.512	0.232	0.015	1
oscuros	0.075	0.037	0.307	0.518	0.065	1

$Tabla\ 3$

En lo que sigue se construirá sobre la matriz $\mathbf{R}$ una distancia adecuada sobre sus filas.

5. La matriz de frecuencias estandarizada por columna $\mathbf{Y}$

Cada fila $\mathbf{r}_i^t$ de la matriz $\mathbf{R}$ es un elemento del espacion $\mathbb{R}^J$, y puesto que los elementos de esta fila suman 1, se tiene que el espacio generador de los vectores $\{\mathbf{r}_i^t\,; i = 1, \ldots, I\}$ es de dimensión $J - 1$. Ahora bien, nuestro objetivo central es proyectar estos puntos en un espacio de dimensión menor de tal forma que estas proyecciones "denuncien" la estructura de proximidad entre los puntos (las filas), es decir aquellos puntos que estén próximos se note geométricamente dicha proximidad en el espacio de proyección, y los puntos que estén alejados muestren su este alejamiento en sus proyecciones. Para este objetivo debemos definir adecuadamente una distancia entre los elementos filas $\mathbf{r}_i^t$ y $\mathbf{r}_k^t$.

Definamos la siguiente distancia cuadrática sobre las filas de la matriz $\mathbf{R}$

$$D^2(\mathbf{r}_i, \mathbf{r}_k) = \sum_{j=1}^{J} \frac{(r_{i\,j} - r_{k\,j})^2}{f_{\cdot\,j}} \tag{5}$$

Podemos notar que la distancia definida en (5) es esencialmente la distancia cuadrática euclídea, donde cada diferencia entre las filas es ponderada por las frecuencias marginales de cada columna. Esta ponderación también obedece al hecho de considerar el peso que puede otorgar el "otro" atributo que participa a través de las columnas. De otra forma, la diferencia entre los atributos filas será ponderada adecuadamente por cada atributo columna, puesto que $\sum_j f_{\cdot\,j} = 1$. A la distancia definida en (5) se le llama distancia χ^2 (ji-cuadrado), y se calcula matricialmente como

$$D^2(\mathbf{r}_i, \mathbf{r}_k) = (\mathbf{r}_i - \mathbf{r}_k)^t \mathbf{D}_c^{-1} (\mathbf{r}_i - \mathbf{r}_k)$$

donde $\mathbf{D}_c$ es la matriz diagonal de $J \times J$ cuyos elementos en la diagonal son los términos $f_{\cdot\,j}$.

Notemos lo siguiente,

$$\begin{aligned} D^2(\mathbf{r}_i, \mathbf{r}_k) &= \sum_{j=1}^{J} \frac{(r_{i\,j} - r_{k\,j})^2}{f_{\cdot\,j}} = \sum_{j=1}^{J} \left(\frac{r_{i\,j}}{\sqrt{f_{\cdot\,j}}} - \frac{r_{k\,j}}{\sqrt{f_{\cdot\,j}}} \right)^2 \\ &= \sum_{j=1}^{J} \left(\frac{f_{i\,j}}{f_{i\cdot}\sqrt{f_{\cdot\,j}}} - \frac{f_{k\,j}}{f_{i\cdot}\sqrt{f_{\cdot\,j}}} \right)^2 \end{aligned}$$

esta última igualdad en virtud de (4). De tal forma que la distancia χ^2 no es más que la distancia cuadrática euclídea esta vez sobre entre las filas de la matriz $\mathbf{Y}$ de $I \times J$ definida por

$$\mathbf{Y} = \left(\frac{f_{i\,j}}{f_{i\cdot}\sqrt{f_{\cdot\,j}}} \right)$$

C. ojos	Color del pelo				
	rubio	pelirrojo	castaño	oscuro	negro
claros	0.837	0.326	0.587	0.235	0.015
azules	0.873	0228	0.536	0.301	0.029
castaños	0.374	0.205	0.815	0.455	0.095
oscuros	0.147	0.161	0.484	1.022	0.440

$Tabla\ 4$

La interpretación que tienen las entradas de esta matriz es como sigue. Las entradas de esta matriz representan las frecuencias relativas condicionadas por filas, $f_{i\,j}\ /\ f_{i\,\cdot}$, pero estandarizadas por su variabilidad, representada por la raíz cuadrada de la frecuencia relativa de cada columna, de esta forma las entradas de la matriz son comparables entre sí. En el ejemplo de los escolares escoceses, la matriz $\mathbf{Y}$ está dada en la Tabla 4.

Podemos entonces tratar a esta matriz como una matriz de datos estándar de la manera

habitual, eso es que representa I observaciones con J variables de preguntas (en columnas), y por lo tanto podemos intentar proyectar las observaciones en un espacio de dimensión menor de tal manera que se preserven las distancias existentes entre las observaciones (las filas). Esto significa encontrar una dirección unitaria $\mathbf{a}^t$ de tal forma que los puntos proyectados sobre esta dirección $\mathbf{Ya}$ tengan variabilidad máxima. Si definimos

$$y_p(\mathbf{a}) = \mathbf{Ya}$$

el vector dirección $\mathbf{a}$ se encuentra maximizando $y_p(\mathbf{a})^y \; y_p(\mathbf{a}) = \mathbf{a}^t\mathbf{Y}^t\,\mathbf{Ya}$ sujeto a la condición $\mathbf{a}^t\mathbf{a} = \mathbf{1}$, pero este problema es precisamente el cálculo del componente principal, de tal forma que el vector $\mathbf{a}$ es el autovalor unitario de la matriz $\mathbf{Y}^t\,\mathbf{Y}$ asociado al mayor autovalor de dicha matriz.

6. La matriz estandarizada Z

La función que estamos maximizando tiene su armazón en la matriz $\mathbf{Y}$, que hemos llamado de datos y construida e partir de la matriz de frecuencias $\mathbf{F}$. Sin embargo esta matriz $\mathbf{Y}$ le vamos a multiplicar sus entradas por el factor $f_{i\,\cdot}$ y así obtenemos una nueva matriz $\mathbf{Z}$ cuyas entradas son

$$z_{i\,j} = \frac{f_{i\,j}}{\sqrt{f_{i\,\cdot}\,f_{\cdot\,j}}}$$

y de esta forma vamos a encontrar una dirección $\mathbf{a}$ para proyectar los datos $\mathbf{Y}$ de tal mamnera que sea máxima la forma cuadrática

$$\mathbf{a}^t\mathbf{Z}^{\,t}\mathbf{Z}\,\mathbf{a} \tag{6}$$

con la condición $\mathbf{a}^t\mathbf{a} = \mathbf{1}$.

La maximización de (6) no es arbitraria puesto que

$$\mathbf{a}^t\mathbf{Z}^{\,t}\mathbf{Z}\,\mathbf{a} = \mathbf{a}^t\mathbf{D}_c^{-\,1/2}\mathbf{F}^t\mathbf{D}_f^{-1}\mathbf{F}\,\mathbf{D}_c^{-\,1/2}\mathbf{a} = \mathbf{a}^t\mathbf{Y}^t\,\mathbf{D}_f\;\mathbf{Ya}$$

De forma tal que el operador $\mathbf{D}_f$ entrega mayor peso justamente a las filas que tienen mayor frecuencia relativa que tienen poca frecuencia relativa en la matriz $\mathbf{Y}$. En definitiva la maximización de (6) entrega a cada fila un peso proporcional al número de datos que contiene.

Ahora bien, sabemos que maximizar la expresión (6) significa encontrar los componentes principales de la matriz $\mathbf{Z}$. Luego buscamos entre los vectores propios, esto es

$$\mathbf{Z}^{\,t}\mathbf{Z}\,\mathbf{a} = \boldsymbol{\lambda}\mathbf{a}$$

con λ autovalor de $\mathbf{Z}^{\,t}\mathbf{Z}$.

En el caso de componentes principales para matriz de datos cuantitativos, el primer componente era el autovalor unitario asociado al mayor autovalor de la matriz obtenida por el producto en tre la traspuesta de la matriz de datos y la matriz de datos. Esta vez, esto no será posible dada que la estructura de $\mathbf{Z}$ conduce a que el mayor autovalor de $\mathbf{Z}^{\,t}\mathbf{Z}$ es el 1, y su autovector asociado $\mathbf{D}_c^{-\,1/2}$. Veamos la demostración.

Tenemos que

$$\mathbf{D}_c^{-1/2}\mathbf{F}^t\mathbf{D}_f^{-1}\mathbf{F}\,\mathbf{D}_c^{-1/2}\mathbf{a} = \boldsymbol{\lambda}\mathbf{a}$$

multiplicando por $\mathbf{D}_c^{-1/2}$ por la izquierda,

$$\mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1}\mathbf{F}\,\mathbf{D}_c^{-1/2}\mathbf{a} = \boldsymbol{\lambda}\,(\mathbf{D}_c^{-1/2}\mathbf{a})$$

Por otro lado las matrices $\mathbf{D}_f^{-1}\mathbf{F}$ y $\mathbf{D}_c^{-1}\mathbf{F}^t$ por construcción satisfacen lque

$$\mathbf{D}_f^{-1}\mathbf{F}\ \mathbf{1} = \mathbf{1}\,;\ \ \mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{1} = \mathbf{1}$$

y en consecuencia la matriz $\mathbf{D}_c^{-1}\mathbf{F}^t\mathbf{D}_f^{-1}\mathbf{F}$ tiene un autovalor 1 unido a un autovector $\mathbf{1}$. Luego haciendo $(\mathbf{D}_c^{-1/2}\mathbf{a}) = \mathbf{1}$ concluimos que la matriz $\mathbf{Z}^t\mathbf{Z}$ tiene un valor propio igual a uno, con vector propio $\mathbf{D}_c^{-1/2}$.

Entonces tenemos que obviar esta solución trivial que no nos entrega informaci{on sobre la estructura de las filas, por lo tanto se hace necesario tomar el mayor valor propio menos a la unidad de la matriz $\mathbf{Z}^t\mathbf{Z}$ para determinar el vector propio $\mathbf{a}$, y entonces proyectamos a la matriz $\mathbf{Y}$ sobre esta dirección, esto es

$$\mathbf{Ya} = \mathbf{D}_f^{-1}\mathbf{F}\,\mathbf{D}_c^{-1/2}\mathbf{a}$$

y el vector $\mathbf{Ya}$ es la mejor representación de las filas de la tabla de contingencia en una dimensión. Análogamente, si extraemos el próximo autovector asociado al siguiente mayor autovalor obtenemos una segunda componente, y así podemos representar las filas en un espacio de dimensión dos. Las coordenadas $\mathbf{C}_f$ de la representación de cada fila se obtienen de la siguiente forma

$$\mathbf{C}_f = \mathbf{Y}\,\mathbf{A}_2 = \mathbf{D}_f^{-1}\mathbf{F}\,\mathbf{D}_c^{-1/2}\mathbf{A}_2$$

donde $\mathbf{A}_2 = \{\mathbf{a}_1, \mathbf{a}_2\}$ contiene la columna de los dos vectores propios de $\mathbf{Z}^t\mathbf{Z}$.