

La práctica del análisis de correspondencias

MICHAEL GREENACRE

Catedrático de Estadística en la Universidad Pompeu Fabra

Separata del apéndice A

Teoría del análisis de correspondencias

Primera edición: julio 2008
ISBN: 978-84-96515-71-0

Traducción: Jordi Comas Angelet
Revisión: Carles M. Cuadras Avellana

© **Michael Greenacre, 2008**
© de la edición en español, **Fundación BBVA, 2008**

www.fbbva.es

Teoría del análisis de correspondencias

El análisis de correspondencias se basa en resultados directos de la teoría de matrices. Utiliza especialmente la descomposición de una matriz en valores singulares (DVS), el fundamento de muchos métodos multivariantes, como el análisis de componentes principales, el análisis de correlaciones canónicas, todas las variantes de los biplots lineales, el análisis discriminante y el escalado métrico multidimensional. En este apéndice resumimos la teoría del análisis de correspondencias, así como la de los métodos relacionados que hemos visto en este libro. Hemos preferido utilizar la notación matricial porque es más concisa, y también porque se halla más próxima a la implementación del método en el lenguaje de cálculo R.

Contenido

Matriz de correspondencias y notación preliminar	266
Algoritmo básico de cálculo	266
Un apunte sobre la descomposición en valores singulares (DVS)	267
El modelo bilineal del AC	268
Ecuaciones de transición entre filas y columnas	268
Puntos adicionales	269
Inercia total y distancias χ^2	269
Contribuciones de filas y columnas a las inercias principales	270
Contribuciones de los ejes principales a las inercias de los puntos (correlaciones al cuadrado)	270
Agrupación de Ward de perfiles fila o de perfiles columna	270
Tablas concatenadas	271
AC múltiple	271
AC conjunto	272
Porcentaje de inercia explicado por el ACCo	272
Contribuciones en ACC	273
Inercias ajustadas en ACM	274
AC y ACM de subgrupos	274
Análisis de tablas asimétricas cuadradas	274
AC canónico	275
Tablas para contrastar agrupaciones o dimensiones significativas	276

Sea \mathbf{N} una matriz de datos $I \times J$, con sumas positivas de filas y columnas (casi siempre \mathbf{N} consta de valores no negativos, sin embargo, existen algunas excepciones como la que describimos al final del capítulo 23). Para simplificar la notación, en primer lugar, transformamos la matriz \mathbf{N} en la *matriz de correspondencias* \mathbf{P} , dividiendo \mathbf{N} por la suma total de sus elementos $n = \sum_i \sum_j n_{ij} = \mathbf{1}^T \mathbf{N} \mathbf{1}$ (utilizamos $\mathbf{1}$ para simbolizar un vector de unos, con una longitud adecuada a su uso. El primer $\mathbf{1}$ es $I \times 1$, el segundo es $J \times 1$ para coincidir con el número de filas y de columnas de \mathbf{N}).

Matriz de correspondencias:

$$\mathbf{P} = \frac{1}{n} \mathbf{N} \quad (\text{A.1})$$

Hemos utilizado la siguiente notación:

Masas de filas y columnas:

$$\begin{aligned} r_i &= \sum_{j=1}^J p_{ij} & c_j &= \sum_{i=1}^I p_{ij} \\ \text{es decir, } \mathbf{r} &= \mathbf{P} \mathbf{1} & \mathbf{c} &= \mathbf{P}^T \mathbf{1} \end{aligned} \quad (\text{A.2})$$

Matrices diagonal de masas de filas y de columnas:

$$\mathbf{D}_r = \text{diag}(\mathbf{r}) \quad \text{y} \quad \mathbf{D}_c = \text{diag}(\mathbf{c}) \quad (\text{A.3})$$

A partir de ahora, expresaremos todas las definiciones y todos los resultados en términos de estos valores relativos $\mathbf{P} = \{p_{ij}\}$, $\mathbf{r} = \{r_i\}$ y $\mathbf{c} = \{c_j\}$, cuyos elementos, en todos los casos, suman 1. Multiplicando por n recuperamos los elementos de la matriz original \mathbf{N} : $np_{ij} = n_{ij}$, $nr_i =$ suma de la i -ésima fila de \mathbf{N} , $nc_j =$ suma de la j -ésima columna de \mathbf{N} .

El algoritmo de cálculo para obtener las coordenadas de los perfiles fila y de los perfiles columna en relación con los ejes principales, utilizando la descomposición en valores singulares (DVS), es el siguiente:

Paso 1 del AC. Cálculo de la matriz \mathbf{S} de residuos estandarizados:

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} (\mathbf{P} - \mathbf{r} \mathbf{c}^T) \mathbf{D}_c^{-\frac{1}{2}} \quad (\text{A.4})$$

Paso 2 del AC. Cálculo de la DVS de \mathbf{S} :

$$\mathbf{S} = \mathbf{U} \mathbf{D}_\alpha \mathbf{V}^T \quad \text{donde} \quad \mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I} \quad (\text{A.5})$$

donde \mathbf{D}_α es la matriz diagonal de valores singulares (positivos) en orden descendente: $\alpha_1 \geq \alpha_2 \geq \dots$

Paso 3 del AC. Coordenadas estándares de las filas Φ :

$$\Phi = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U} \quad (\text{A.6})$$

Paso 4 del AC. Coordenadas estándares de las columnas Γ :

$$\Gamma = \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V} \quad (\text{A.7})$$

Paso 5 del AC. Coordenadas principales de las filas \mathbf{F} :

$$\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U} \mathbf{D}_\alpha = \Phi \mathbf{D}_\alpha \quad (\text{A.8})$$

Paso 6 del AC. Coordenadas principales de las columnas \mathbf{G} :

$$\mathbf{G} = \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V} \mathbf{D}_\alpha = \Gamma \mathbf{D}_\alpha \quad (\text{A.9})$$

Paso 7 del AC. Inercias principales λ_k :

$$\lambda_k = \alpha_k^2, \quad k=1,2,\dots,K \quad \text{donde} \quad K = \min\{I-1, J-1\} \quad (\text{A.10})$$

Las filas de las matrices de coordenadas en (A.6)–(A.9) hacen referencia a las filas o a las columnas, según cada caso, de la tabla original. Mientras que las columnas de estas matrices hacen referencia a los ejes principales, o dimensiones, de las cuales hay $\min\{I-1, J-1\}$. Es decir, uno menos el número de filas o de columnas, el que sea menor. Fijémonos en el cálculo de las escalas de las coordenadas principales y estándares:

$$\mathbf{F} \mathbf{D}_r \mathbf{F}^\top = \mathbf{G} \mathbf{D}_c \mathbf{G}^\top = \mathbf{D}_\lambda \quad (\text{A.11})$$

$$\Phi \mathbf{D}_r \Phi^\top = \Gamma \mathbf{D}_c \Gamma^\top = \mathbf{I} \quad (\text{A.12})$$

Es decir, la suma de cuadrados ponderada de las coordenadas principales en la k -ésima dimensión (es decir, su inercia en la dirección de esta dimensión) es igual a la inercia principal (o valor propio) $\lambda_k = \alpha_k^2$, el cuadrado del k -ésimo valor singular. Mientras que la suma de cuadrados ponderada de las coordenadas estandarizadas es igual a 1. Todas las matrices de coordenadas tienen columnas ortogonales, en las que siempre utilizamos las masas para el cálculo de los productos escalares (ponderados).

La DVS es un resultado matemático fundamental del AC, de la misma manera que lo es para otras técnicas de reducción de la dimensión, como el análisis de componentes principales, el análisis de correlaciones canónicas, y el análisis discriminante lineal. Esta descomposición expresa cualquier matriz rectangular como el producto de tres matrices de estructura simple, como vimos anteriormente en (A.5): $\mathbf{S} = \mathbf{U} \mathbf{D}_c \mathbf{V}^\top$. Las columnas de las matrices \mathbf{U} y \mathbf{V} son, respectiva-

Un apunte sobre la descomposición en valores singulares (DVS)

mente, los *vectores singulares izquierdo y derecho*. Los valores positivos α_k de la diagonal de \mathbf{D}_α son, en orden descendente, los *valores singulares*. Si $\mathbf{S}\mathbf{S}^\top$ y $\mathbf{S}^\top\mathbf{S}$ son matrices cuadradas simétricas, la DVS está relacionada con la descomposición en *valores y vectores propios* de una matriz cuadrada simétrica de la siguiente manera $\mathbf{S}\mathbf{S}^\top = \mathbf{U}\mathbf{D}_\alpha^2\mathbf{U}^\top$ y $\mathbf{S}^\top\mathbf{S} = \mathbf{V}\mathbf{D}_\alpha^2\mathbf{V}^\top$. Por tanto, los vectores singulares son también vectores propios de las respectivas matrices, y los valores singulares son las raíces cuadradas de sus valores propios. La utilidad práctica de la DVS es que podemos construir otra matriz $I \times J$ $\mathbf{S}_{(m)}$ a partir de las primeras m columnas $\mathbf{U}_{(m)}$ y $\mathbf{V}_{(m)}$, y los primeros m valores singulares $\mathbf{D}_{\alpha(m)}$ como: $\mathbf{S}_{(m)} = \mathbf{U}_{(m)}\mathbf{D}_{\alpha(m)}\mathbf{V}_{(m)}^\top$. Por lo tanto, $\mathbf{S}_{(m)}$ es la aproximación mínimos cuadrados de rango m a \mathbf{S} (*teorema de Eckart-Young*). Dado que el objetivo de hallar los subespacios de pocas dimensiones que mejor se ajusten coincide con el objetivo de hallar matrices de rango pequeño por mínimos cuadrados, la DVS resuelve de forma satisfactoria y muy compacta nuestro problema. La única adaptación necesaria es incorporar en la DVS la ponderación de filas y de columnas con las masas de manera que obtengamos aproximaciones por mínimos cuadrados ponderados. Si definimos una forma generalizada de la DVS, con valores singulares normalizados y ponderados con las masas, entonces podremos obtener directamente el resultado del AC. Por ejemplo, la DVS generalizada de los *cocientes de contingencia* $p_{ij}/(r_i c_j)$, los elementos de la matriz $\mathbf{D}_r^{-1}\mathbf{P}\mathbf{D}_c^{-1}$, centrada en el valor constante 1, nos conduce directamente a las coordenadas estándares de filas y de columnas:

$$\mathbf{D}_r^{-1}\mathbf{P}\mathbf{D}_c^{-1} - \mathbf{1}\mathbf{1}^\top = \Phi\mathbf{D}_\alpha\Gamma^\top \quad \text{donde} \quad \Phi^\top\mathbf{D}_r\Phi = \Gamma^\top\mathbf{D}_c\Gamma = \mathbf{I} \quad (\text{A.13})$$

El modelo bilineal del AC

Los pasos 1 a 4 del algoritmo básico nos permiten escribir los datos de \mathbf{P} de la siguiente manera [véanse también (13.4), pág. 139, y (14.9), pág. 150]:

$$p_{ij} = r_i c_j \left(1 + \sum_{k=1}^K \sqrt{\lambda_k} \phi_{ik} \gamma_{jk} \right) \quad (\text{A.14})$$

(*fórmula de reconstitución*). En notación matricial:

$$\mathbf{P} = \mathbf{D}_r(\mathbf{1}\mathbf{1}^\top + \Phi\mathbf{D}_\alpha^{1/2}\Gamma^\top)\mathbf{D}_c \quad (\text{A.15})$$

Dadas las relaciones simples (A.8) y (A.9), entre las coordenadas principales y las coordenadas estándares, podemos escribir este modelo bilineal de distintas maneras [véanse también (14.10) y (14.11) en la pág. 150].

Ecuaciones de transición entre filas y columnas

Los vectores singulares izquierdo y derecho están relacionados linealmente. Por ejemplo, multiplicando la DVS de la derecha por \mathbf{V} : $\mathbf{S}\mathbf{V} = \mathbf{U}\mathbf{D}_\alpha$. Las variaciones consistentes en expresar estas relaciones en términos de coordenadas principales y de coordenadas estándares proporcionan las ecuaciones de transición [véanse (14.1) y (14.2), así como (14.5) y (14.6) para las ecuaciones escalares equivalentes]:

Coordenadas principales en función de las coordenadas estándares (relaciones baricéntricas):

$$\mathbf{F} = \mathbf{D}_r^{-1} \mathbf{P} \mathbf{\Gamma} \quad \mathbf{G} = \mathbf{D}_c^{-1} \mathbf{P}^T \mathbf{\Phi} \quad (\text{A.16})$$

Coordenadas principales en función de las coordenadas principales:

$$\mathbf{F} = \mathbf{D}_r^{-1} \mathbf{P} \mathbf{G} \mathbf{D}_\lambda^{-1/2} \quad \mathbf{G} = \mathbf{D}_c^{-1} \mathbf{P}^T \mathbf{F} \mathbf{D}_\lambda^{-1/2} \quad (\text{A.17})$$

Las ecuaciones (A.16) son las que mencionamos en el capítulo 3. Expresan los perfiles como medias ponderadas de los vértices, en las que los pesos son los elementos del perfil. Son las ecuaciones que configuran los *mapas asimétricos*. Las ecuaciones (A.17) muestran que los dos conjuntos de coordenadas principales, que configuran los *mapas simétricos*, también están relacionados por una relación baricéntrica (media ponderada), pero con factores de escala diferentes en cada eje (las inversas de las raíces cuadradas de las inercias principales).

Utilizamos las ecuaciones de transición para situar puntos adicionales en el mapa. Por ejemplo, dada una columna adicional con valores en \mathbf{h} ($I \times 1$), la dividimos por su total $\mathbf{1}^T \mathbf{h}$ para obtener el perfil columna $\tilde{\mathbf{h}} = (\mathbf{1} / \mathbf{1}^T \mathbf{h}) \mathbf{h}$. A continuación transponemos el perfil para obtener el vector fila en la segunda ecuación de (A.16). Por ejemplo, para calcular las coordenadas \mathbf{g} de la columna adicional:

$$\mathbf{g} = \tilde{\mathbf{h}}^T \mathbf{\Phi} \quad (\text{A.18})$$

La inercia total de la matriz de datos es igual a la suma de los cuadrados de la matriz \mathbf{S} de (A.4):

$$\text{inercia} = \text{traza}(\mathbf{S} \mathbf{S}^T) = \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} \quad (\text{A.19})$$

La inercia también es la suma de los cuadrados de los valores singulares, es decir, la suma de los valores propios:

$$\text{inercia} = \sum_{k=1}^K \alpha_k^2 = \sum_{k=1}^K \lambda_k \quad (\text{A.20})$$

Las distancias χ^2 entre los perfiles fila y los perfiles columna son:

$$\text{Las distancias } \chi^2 \text{ entre las filas } i \text{ e } i': \sum_{j=1}^J \left(\frac{p_{ij}}{r_i} - \frac{p_{i'j}}{r_{i'}} \right)^2 / c_j \quad (\text{A.21})$$

$$\text{Las distancias } \chi^2 \text{ entre las columnas } j \text{ y } j': \sum_{i=1}^I \left(\frac{p_{ij}}{c_j} - \frac{p_{i'j'}}{c_{j'}} \right)^2 / r_i \quad (\text{A.22})$$

Puntos adicionales

Inercia total y
distancias χ^2

Escribir el conjunto completo de distancias χ^2 en forma de matriz cuadrada simétrica exige un poco más de trabajo. En primer lugar, calculamos la matriz \mathbf{A} de «productos escalares χ^2 » entre los perfiles fila, por ejemplo como:

$$\text{Productos escalares } \chi^2 \text{ entre filas: } \mathbf{A} = \mathbf{D}_r^{-1} \mathbf{P} \mathbf{D}_c^{-1} \mathbf{P}^T \mathbf{D}_r^{-1} \quad (\text{A.23})$$

A continuación, definimos el vector \mathbf{a} como los elementos de la diagonal de esta matriz (es decir, los productos escalares de los perfiles fila por ellos mismos):

$$\mathbf{a} = \text{diag}(\mathbf{A}) \quad (\text{A.24})$$

Entonces, la matriz, $I \times I$, de distancias χ^2 al cuadrado es:

$$\text{Matriz de distancias } \chi^2 \text{ al cuadrado entre filas: } \mathbf{a} \mathbf{1}^T + \mathbf{1} \mathbf{a}^T - 2\mathbf{A} \quad (\text{A.25})$$

Para calcular la matriz $J \times J$ de distancias χ^2 al cuadrado entre perfiles columna, intercambiamos las filas por las columnas en (A.23). Definimos \mathbf{A} como $\mathbf{D}_c^{-1} \mathbf{P}^T \mathbf{D}_r^{-1} \mathbf{P} \mathbf{D}_c^{-1}$ y a continuación seguimos con (A.24) y (A.25).

Contribuciones de filas y
columnas a las inercias
principales

Las contribuciones de filas y de columnas a la inercia de la k -ésima dimensión son los componentes de la inercia:

$$\text{Para la fila } i: \frac{r_i f_{ik}^2}{\lambda_k} = r_i \phi_{ik}^2 \quad \text{para la columna } j: \frac{c_j g_{jk}^2}{\lambda_k} = c_j \gamma_{jk}^2 \quad (\text{A.26})$$

recordando la relación entre coordenadas principales y coordenadas estándares dadas en (A.8) y (A.9): $f_{ik} = \sqrt{\lambda_k} \phi_{ik}$, $g_{jk} = \sqrt{\lambda_k} \gamma_{jk}$. [Fijémonos en que las raíces cuadradas de los valores de (A.26) son exactamente las coordenadas propuestas para el biplot estándar de AC del capítulo 13, que muestra que las raíces cuadradas de estas coordenadas son las contribuciones a los ejes principales.]

Contribuciones de los
ejes principales a las
inercias de los puntos
(correlaciones al
cuadrado)

Las contribuciones de las dimensiones de la inercia a la i -ésima fila y a la j -ésima columna (es decir, los cosenos al cuadrado o las correlaciones al cuadrado) son:

$$\text{Para la fila } i: \frac{f_{ik}^2}{\sum_k f_{ik}^2} \quad \text{para la columna } j: \frac{g_{jk}^2}{\sum_k g_{jk}^2} \quad (\text{A.27})$$

Como vimos en el capítulo 11, los denominadores de (A.27) son los cuadrados de las distancias χ^2 entre los correspondientes perfiles y el perfil medio.

Agrupación de Ward de
perfiles fila o de perfiles
columna

Aquí describimos la agrupación del capítulo 15 en términos de filas. Podemos aplicar exactamente lo mismo a la agrupación de columnas. En cada paso del algoritmo, agrupamos las filas para minimizar la disminución del valor del estadístico χ^2 (de forma equivalente, para minimizar la disminución del valor de la

inercia, dado que la inercia $= \chi^2/n$, donde n es el total de la tabla). Este criterio de agrupación es equivalente a la agrupación de Ward, en la que ponderamos cada grupo con la masa total de sus miembros. Podemos demostrar que la medida de la diferencia entre filas es igual a la forma ponderada de la distancia ji-cuadrado entre perfiles. Supongamos que \mathbf{a}_i y r_i , con $i = 1, \dots, I$, simbolizan, respectivamente, los perfiles de la fila I de la matriz de datos y sus masas. A continuación identificamos el par que da la menor disminución de inercia, lo que es equivalente a buscar el par de filas (i, i') que minimiza la siguiente expresión:

$$\frac{r_i r_{i'}}{r_i + r_{i'}} \|\mathbf{a}_i - \mathbf{a}_{i'}\|_c^2 \quad (\text{A.28})$$

A continuación reunimos las dos filas sumando sus frecuencias, y recalculamos su perfil y su masa. En cada etapa de agrupación de los perfiles fila calculamos la misma medida de diferencia que vimos en (A.28) [véase (15.2) en la pág. 149 para una fórmula equivalente basada en perfiles de grupos], y reunimos los dos perfiles con la mínima diferencia. Por tanto, (A.28) es el nivel de agrupación en términos de disminución de inercia, o si multiplicamos por n , de disminución de χ^2 . En el caso de tablas de contingencia, podemos contrastar la significación del nivel de agrupación utilizando la tabla del final de este apéndice.

Supongamos que concatenamos horizontal o verticalmente las tablas \mathbf{N}_{qs} , con $q = 1, \dots, Q$, $s = 1, \dots, S$, para formar una matriz compuesta \mathbf{N} . Si las frecuencias marginales son las mismas en todas las filas y en todas las columnas (caso que sucede cuando cruzamos de forma independiente los mismos individuos en varias tablas), entonces, la inercia de \mathbf{N} es la media de las inercias de las tablas \mathbf{N}_{qs} :

$$\text{inercia}(\mathbf{N}) = \frac{1}{QS} \sum_{q=1}^Q \sum_{s=1}^S \text{inercia}(\mathbf{N}_{qs}) \quad (\text{A.29})$$

Supongamos que la matriz original de datos categóricos es $N \times Q$, es decir, N casos y Q variables. El AC múltiple clásico (ACM) tiene dos versiones. En la primera transformamos los casos clasificados por variables en una matriz binaria \mathbf{Z} en la que recodificamos los datos categóricos como variables binarias. Si la variable q -ésima tiene J_q categorías, esta matriz binaria tendrá $J = \sum_q J_q$ columnas (mostramos un ejemplo en la tabla de la imagen 18.1). La versión binaria del ACM no es más que la aplicación del algoritmo clásico de AC a la matriz \mathbf{Z} . Así obtenemos las coordenadas de los N casos y de las J categorías. La segunda versión del ACM, consiste en calcular la matriz de Burt $\mathbf{B} = \mathbf{Z}^T \mathbf{Z}$ de todos los cruces de las Q variables (mostramos un ejemplo en la tabla de la imagen 18.4). La versión Burt del ACM consiste en aplicar el algoritmo básico del AC a la matriz \mathbf{B} . Obtenemos así las coordenadas de las J categorías (\mathbf{B} es una matriz simétrica). Las

Tablas concatenadas

AC múltiple

coordenadas estándares de las categorías son idénticas en los dos versiones del ACM, y las inercias principales de la versión Burt son los cuadrados de los de la versión binaria.

AC conjunto

En el AC conjunto (ACCo) ajustamos las tablas situadas fuera de la diagonal de la matriz de Burt, ignorando las matrices situadas en la diagonal. El algoritmo que utilizamos es un procedimiento iterativo de mínimos cuadrados alternados. Aplicamos el AC sobre la matriz de Burt, que vamos modificando sucesivamente sustituyendo las matrices de la diagonal por valores estimados en el AC de la iteración previa. En el apéndice de cálculo explicamos más detalladamente este algoritmo. Cuando en el algoritmo del ACCo llegamos a la convergencia, llevamos a cabo el AC en la última matriz de Burt modificada, $\tilde{\mathbf{B}}$, que tiene en su diagonal matrices perfectamente ajustadas por construcción. Así, por ejemplo, si suponemos que la solución que buscamos es bidimensional, entonces las matrices de la diagonal modificadas cumplen exactamente (A.14), en la que utilizamos sólo dos términos del modelo del AC bilineal (o fórmula de reconstitución).

Porcentaje de inercia explicado por el ACCo

De la misma forma que la inercia total de $\tilde{\mathbf{B}}$ incluye la inercia Δ atribuible a las matrices de la diagonal, las dos inercias principales, $\tilde{\lambda}_1$ y $\tilde{\lambda}_2$, también la incluyen. Para hallar el porcentaje de inercia explicado por la solución bidimensional tenemos que descontar la inercia Δ tanto de la inercia total como de las dos inercias principales. Podemos obtener Δ a partir de la diferencia entre la inercia de la matriz de Burt original \mathbf{B} (de la que conocemos las inercias de la diagonal) y la de la matriz de Burt modificada $\tilde{\mathbf{B}}$ de la siguiente manera [aquí utilizamos el resultado (A.29) que podemos aplicar a las matrices \mathbf{B}_{qs} de \mathbf{B} , y a las de $\tilde{\mathbf{B}}$, que tienen las mismas matrices fuera de la diagonal]:

$$\begin{aligned} \text{inercia}(\mathbf{B}) &= \frac{1}{Q^2} \left(\sum \sum_{q \neq s} \text{inercia}(\mathbf{B}_{qs}) + \sum_q \text{inercia}(\mathbf{B}_{qq}) \right) \\ &= \frac{1}{Q^2} \left(\sum \sum_{q \neq s} \text{inercia}(\mathbf{B}_{qs}) + (J - Q) \right) \\ \text{inercia}(\tilde{\mathbf{B}}) &= \frac{1}{Q^2} \left(\sum \sum_{q \neq s} \text{inercia}(\mathbf{B}_{qs}) \right) + \Delta \end{aligned}$$

Restando lo anterior nos lleva a:

$$\text{inercia}(\mathbf{B}) - \text{inercia}(\tilde{\mathbf{B}}) = \frac{J - Q}{Q^2} - \Delta \quad (\text{A.30})$$

Lo que nos da el valor Δ :

$$\Delta = \frac{J - Q}{Q^2} - (\text{inercia}(\mathbf{B}) - \text{inercia}(\tilde{\mathbf{B}})) \quad (\text{A.31})$$

Restando este valor del total y de, la suma de las inercias principales (suponiendo una solución bidimensional) nos proporciona el porcentaje de inercia explicado por la solución del ACCo:

$$100 \times \frac{\tilde{\lambda}_1 + \tilde{\lambda}_2 - \Delta}{\text{inercia}(\tilde{\mathbf{B}}) - \Delta} \quad (\text{A.32})$$

En la sección anterior vimos cómo eliminar la inercia extra de las matrices de la diagonal resultante del ACC de la matriz de Burt. A nivel de cada punto (fila o columna), se nos presenta la misma situación. Es decir, cada categoría j tiene un valor adicional de inercia, δ_j , atribuible a las matrices modificadas de la diagonal. Para la matriz de Burt original \mathbf{B} sabemos cómo calcular este valor, atribuible a las matrices de la diagonal: para el j -ésimo punto es $(1 - Qc_j)/Q^2$, donde c_j es la j -ésima masa (sumando estos valores para $j=1, \dots, J$, obtenemos $(J - Q)/Q^2$ que es la inercia extra total atribuible a las matrices diagonales de \mathbf{B}). Tal como vimos en el apartado anterior, podemos obtener las contribuciones de la solución bidimensional de las inercias de los puntos de la manera siguiente:

Contribuciones en ACC

$$\text{inercia}(j\text{-ésima categoría de } \mathbf{B}) = \text{componentes de fuera de la diagonal} + \frac{1 - Qc_j}{Q^2}$$

$$\text{inercia}(j\text{-ésima categoría de } \tilde{\mathbf{B}}) = \text{componentes de fuera de la diagonal} + \delta_j$$

Restando lo anterior (los «componentes fuera de la diagonal» son los mismos) nos lleva a:

$$\text{inercia}(j\text{-ésima categoría de } \mathbf{B}) - \text{inercia}(j\text{-ésima categoría de } \tilde{\mathbf{B}}) = \frac{1 - Qc_j}{Q^2} - \delta_j$$

lo que nos permite calcular δ_j :

$$\delta_j = \frac{1 - Qc_j}{Q^2} - (\text{inercia}(j\text{-ésima categoría de } \mathbf{B}) - \text{inercia}(j\text{-ésima categoría de } \tilde{\mathbf{B}})) \quad (\text{A.33})$$

Descontando este valor de inercia de la j -ésima categoría y, de forma similar, de la suma de los componentes de la inercia de las dos dimensiones, obtenemos las contribuciones relativas (calidades) en relación al resultado bidimensional del ACCo:

$$\frac{c_j \tilde{g}_{j1}^2 + c_j \tilde{g}_{j2}^2 - \delta_j}{(\sum_k c_j \tilde{g}_{jk}^2) - \delta_j} \quad (\text{A.34})$$

donde \tilde{g}_{jk} es la coordenada principal de la j -ésima categoría del eje k del AC de $\tilde{\mathbf{B}}$ (solución del ACCo). La suma en el denominador es para todas las dimensiones. Fijémonos en que $\sum_j \delta_j = \Delta$; es decir, sumando (A.33) obtenemos (A.31).

Inercias ajustadas en ACM

Podemos ajustar la solución del ACM para optimizar el ajuste de las matrices situadas fuera de la diagonal (lo podríamos denominar ACC *condicionado al* resultado del ACM). Según vimos en el capítulo 19, podemos obtener los ajustes óptimos por regresión de mínimos cuadrados ponderada. Sin embargo, debido a que la solución no está anidada, preferimos introducir una ligera modificación —fácil de calcular a partir del resultado del ACM de la matriz de Burt— que sí nos permite obtener una solución anidada aunque sea subóptima. El mencionado ajuste lo realizamos de la siguiente manera (podemos consultar las págs. 198-200 del cap. 19):

Inercia total ajustada de la matriz de Burt:

$$\text{inercia total ajustada} = \frac{Q}{Q-1} \times \left(\text{inercia de } \mathbf{B} - \frac{J-Q}{Q^2} \right) \quad (\text{A.35})$$

Las *inercias principales ajustadas* (*valores propios*) de la matriz de Burt:

$$\lambda_k^{\text{adj}} \left(\frac{Q}{Q-1} \right)^2 \times \left(\sqrt{\lambda_k} - \frac{1}{Q} \right)^2, \quad k = 1, 2, \dots \quad (\text{A.36})$$

Aquí λ_k indica la k -ésima inercia principal de la matriz de Burt. Por tanto $\sqrt{\lambda_k}$ es la k -ésima inercia principal de la matriz binaria. Hacemos los ajustes sólo en las dimensiones para las cuales $\sqrt{\lambda_k} > \frac{1}{Q}$, y no utilizamos más dimensiones; por tanto, los porcentajes de inercia no suman el 100%. Podríamos demostrar que estos porcentajes son estimaciones que están por debajo de los porcentajes que obtenemos en el ACCo, sin embargo, en la práctica son valores muy similares.

AC y ACM de subgrupos

El AC de subgrupos es simplemente una aplicación del algoritmo del AC a una parte seleccionada de la matriz estandarizada residual \mathbf{S} que vimos en (A.4) (*no* es una parte de la matriz original). Sin embargo, para todos los cálculos utilizamos las masas de la matriz completa. Supongamos que trabajamos en una parte de la matriz original que contiene todas las filas y sólo una parte de las columnas. En consecuencia, las filas estarán centradas (es decir, sus medias ponderadas se hallan en el origen del mapa), mientras que no ocurrirá lo mismo con las columnas. Llevamos a cabo el ACM de subgrupos realizando el AC de un subgrupo de la matriz binaria o de la matriz de Burt. En el caso de la matriz de Burt, trabajar con una selección de categorías implica que tenemos que especificar el subgrupo, tanto para filas como para columnas.

Análisis de tablas asimétricas cuadradas

Si la matriz de datos \mathbf{N} es una matriz cuadrada asimétrica, en la que tanto las filas como las columnas hacen referencia a los mismos objetos, podemos expresar \mathbf{N} como la suma de la parte simétrica y la parte asimétrica:

$$\begin{aligned} \mathbf{N} &= \frac{1}{2}(\mathbf{N} + \mathbf{N}^T) + \frac{1}{2}(\mathbf{N} - \mathbf{N}^T) \\ &= \text{simétrica} + \text{asimétrica} \end{aligned} \quad (\text{A.37})$$

Aplicamos el AC a cada una de estas matrices, con una ligera variación para la parte asimétrica. En la parte simétrica $\frac{1}{2}(\mathbf{N} + \mathbf{N}^T)$ procedemos con el AC usual, y así obtenemos un conjunto de coordenadas. Las masas son las medias de las masas de las filas y de las columnas correspondientes al mismo objeto: $w_i = \frac{1}{2}(r_i + c_i)$. El análisis de la parte asimétrica $\frac{1}{2}(\mathbf{N} - \mathbf{N}^T)$ consiste en la aplicación del algoritmo de AC, *sin centrar* y utilizando las mismas masas que las del análisis simétrico. Es decir, la matriz de «residuos estandarizados» de (A.4) es la matriz de «diferencias estandarizadas»:

$$\mathbf{S} = \mathbf{D}_w^{-\frac{1}{2}} \left[\frac{1}{2}(\mathbf{P} - \mathbf{P}^T) \right] \mathbf{D}_w^{-\frac{1}{2}} \quad (\text{A.38})$$

donde \mathbf{P} es la matriz de correspondencias y \mathbf{D}_w es la matriz diagonal de las masas w_i . Tal como vimos en el capítulo 22, podemos sustituir estos dos análisis por un AC habitual de la matriz compuesta:

$$\begin{bmatrix} \mathbf{N} & \mathbf{N}^T \\ \mathbf{N}^T & \mathbf{N} \end{bmatrix} \quad (\text{A.39})$$

Al desarrollar el AC de la matriz compuesta —obtenida a partir de la matriz \mathbf{N} de $I \times I$ —, observamos que las inercias principales de las dimensiones simétricas son únicas, mientras que las inercias principales de las dimensiones asimétricas ocurren por pares. Por tanto, es fácil atribuir las $2I - 1$ dimensiones obtenidas del AC de la matriz compuesta. Por otra parte, observamos que en las coordenadas de las dimensiones simétricas aparecen repeticiones, mientras que para las dimensiones asimétricas aparecen repetidas pero con el signo cambiado (podemos ver un ejemplo en el capítulo 22).

En el AC canónico (ACC) disponemos de una matriz adicional de variables explicativas \mathbf{X} (independientes). En el análisis exigimos que las dimensiones del AC estén relacionadas linealmente con \mathbf{X} . Dividimos la inercia total en la inercia del *espacio restringido* directamente relacionada con las variables explicativas y la inercia correspondiente al *espacio no restringido*. Dado que \mathbf{X} son filas o columnas adicionales, el ACC es un análisis «asimétrico». En ACC es habitual que las filas sean las unidades de muestreo y que \mathbf{X} sea un conjunto adicional de M columnas, es decir, que tengamos una matriz adicional de $I \times M$. En el ACC llevamos a cabo una regresión en la que calculamos la matriz restringida $I \times J$ cuyas columnas están linealmente relacionadas con \mathbf{X} . La diferencia entre \mathbf{P} y la matriz restringida es la matriz no restringida cuyas columnas no están linealmente relacionadas con \mathbf{X} . Por tanto el ACC consiste en aplicar el AC a una matriz restringida y (opcionalmente) a una matriz no restringida. En todos los cálculos, mantenemos las masas originales de filas y de columnas. Los resultados obtenidos —como coordenadas, inercias principales, contribuciones, fórmulas de reconstitución, etc.— son los

misimos que los del AC habitual. En el cálculo de medias y varianzas, damos por supuesto que las columnas de \mathbf{X} están estandarizadas y que utilizamos las masas de las filas como pesos. Si tenemos varias variables categóricas independientes, las codificamos como variables binarias de forma similar a como lo haríamos en un análisis de regresión que estandarizamos de la forma habitual.

Los pasos del ACC son los siguientes:

Paso 1 del ACC – Cálculo de la matriz \mathbf{S} de residuos estandarizados como en AC:

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} (\mathbf{P} - \mathbf{r}\mathbf{c}^\top) \mathbf{D}_c^{-\frac{1}{2}} \quad (\text{A.40})$$

Paso 2 del ACC – Cálculo de la matriz de proyección $I \times I$, de rango M , que proyecta sobre el espacio restringido:

$$\mathbf{Q} = \mathbf{D}_r^{\frac{1}{2}} \mathbf{X} (\mathbf{X}^\top \mathbf{D}_r \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{D}_r^{\frac{1}{2}} \quad (\text{A.41})$$

Paso 3 del ACC – Proyección de los residuos estandarizados para obtener la matriz restringida:

$$\mathbf{S}^* = \mathbf{Q}\mathbf{S} \quad (\text{A.42})$$

Paso 4 del ACC – Aplicación a \mathbf{S}^ de los pasos 1-6 del AC (página 252):*

Paso 5 del ACC – Inercias principales λ_k^ en el espacio restringido:*

$$\lambda_k^* = \alpha_k^2, \quad k=1,2,\dots,K \quad \text{donde} \quad K = \min\{I-1, J-1, M\} \quad (\text{A.43})$$

Paso 6 del ACC (opcional) – Proyección de los residuos estandarizados en el espacio no restringido:

$$\mathbf{S}^\perp = (\mathbf{I} - \mathbf{Q})\mathbf{S} = \mathbf{S} - \mathbf{S}^* \quad (\text{A.44})$$

Paso 7 del ACC (opcional) – Aplicación de los pasos 1-6 del AC a \mathbf{S}^\perp .

Tal como describimos en el capítulo 24, podemos expresar las inercias principales de (A.43) como porcentajes de la inercia total, o como porcentajes de la inercia restringida, que es la suma de los cuadrados de los elementos de \mathbf{S}^* , igual a $\sum_k \lambda_k^*$.

En el caso de tablas de contingencia obtenidas de muestras aleatorias, podemos contrastar la significación estadística de la primera inercia principal. Se trata de la misma prueba que utilizamos para la agrupación de Ward en el capítulo 15. En ese último caso podemos obtener el valor crítico del estadístico χ^2 en la tabla de la imagen A.1, de acuerdo con el tamaño de la tabla (en la página 161, podemos consultar el ejemplo sobre las tiendas de comida. Se trata de una tabla de 5×4 ,

	<i>J</i>									
<i>I</i>	3	4	5	6	7	8	9	10	11	
3	8,59									
4	10,74	13,11								
5	12,68	15,24	17,52							
6	14,49	17,21	19,63	21,85						
7	16,21	19,09	21,62	23,95	26,14					
8	17,88	20,88	23,53	25,96	28,23	30,40				
9	19,49	22,62	25,37	27,88	30,24	32,48	34,63			
10	21,06	24,31	27,15	29,75	32,18	34,50	36,70	38,84		
11	22,61	25,96	28,90	31,57	34,08	36,45	38,72	40,91	43,04	
12	24,12	27,58	30,60	33,35	35,93	38,36	40,69	42,93	45,10	
13	25,61	29,17	32,27	35,09	37,73	40,22	42,60	44,90	47,12	

Imagen A.1:
Valores críticos para la prueba de comparaciones múltiples en una tabla de contingencia de $I \times J$ (o $J \times I$). Podemos utilizar los mismos valores críticos para contrastar la significación de una inercia principal. El nivel de significación es del 5%

Fuente: Pearson, E.S. y Hartley, H.O. (1972). *Biometrika Tables for Statisticians*, Volumen 2: Tabla 51. Cambridge University Press, Gran Bretaña.

para la que el valor crítico de la tabla de la imagen A.1 es 15,24). Estos valores críticos son los mismos que utilizamos para contrastar la significación de la primera inercia principal. Así, para el mismo ejemplo sobre las tiendas de comida, que mostramos en la imagen 15.3, el valor de la primera inercia principal era de 0,02635, que expresada como χ^2 , es de $0,02635 \times 700 = 18,45$. Dado que 18,45 es mayor que el valor crítico 15,24, la primera inercia principal es estadísticamente significativa (a un nivel del 5%).