

# La práctica del análisis de correspondencias

**MICHAEL GREENACRE**

Catedrático de Estadística en la Universidad Pompeu Fabra

---

Separata del capítulo 2

## Perfiles y espacio de perfiles

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](http://www.fbbva.es)



## Perfiles y espacio de perfiles

El concepto de *perfil*, un conjunto de frecuencias relativas, es fundamental para el análisis de correspondencias (AC de aquí en adelante). Estos conjuntos de frecuencias relativas, o *vectores*, tienen características geométricas especiales debido a que la suma de sus elementos es 1 (o el 100%). Cuando analicemos una tabla de frecuencias nos podemos fijar en las frecuencias relativas de las filas o en las frecuencias relativas de las columnas, las llamaremos *perfiles fila* y *perfiles columna*, respectivamente. En este capítulo representaremos los perfiles como puntos en un espacio de perfiles. En particular, lo estudiaremos para el caso especial de perfiles con sólo tres elementos.

### Contenido

Perfiles .....	25
Perfil medio .....	26
Perfiles fila y perfiles columna .....	26
Tratamiento simétrico de las filas y de las columnas .....	27
Consideración asimétrica de una tabla de datos .....	27
Representación de los perfiles en el espacio de perfiles .....	28
Los vértices definen los extremos del espacio de perfiles .....	29
El sistema de coordenadas triangular (o ternario) .....	29
Situación de los puntos en un sistema de coordenadas triangular .....	30
Geometría de los perfiles con más de tres elementos .....	30
Datos en una escala de razón .....	31
Datos en una escala común .....	32
RESUMEN: Perfiles y espacio de perfiles .....	32

Veamos una vez más los datos de la imagen 1.3, una tabla de frecuencias con cuatro filas (los países) y tres columnas (los tipos de día). En el AC, el concepto de *perfil*, un conjunto de frecuencias divididas por su total, es el primer concepto y más fundamental. En la imagen 2.1 se muestran los perfiles de las filas, que llamaremos perfiles fila, de estos datos. Por ejemplo, el perfil de Noruega es

**Imagen 2.1:**  
Perfiles fila (país):  
frecuencias relativas de los  
tipos de día de cada viaje, y  
perfil fila medio que  
muestra las frecuencias  
relativas de todos los viajes  
conjuntamente

País	Festivos	Medias jornadas	Jornadas completas
Noruega	0,33	0,06	0,61
Canadá	0,07	0,20	0,73
Grecia	0,14	0,86	0,00
Francia/Alemania	0,08	0,08	0,83
Media	0,15	0,36	0,49

[0,33 0,06 0,61], donde  $0,33 = 6/18$ ,  $0,06 = 1/18$ ,  $0,61 = 11/18$ . Decimos que se trata del «perfil de Noruega respecto al tipo de día». También podemos expresar el perfil como porcentaje; por ejemplo, en este caso es [33% 6% 61%], como vimos en la imagen 1.5. De forma similar, el perfil de Canadá respecto al tipo de día es [0,07 0,20 0,73], igual que ocurre con Noruega, la mayor concentración se produce en la categoría *jornadas completas*. En cambio, Grecia tiene un perfil de [0,14 0,86 0,00], la mayor concentración se produce en la categoría *medias jornadas*, y así sucesivamente. Estos son los valores que hemos representado en la imagen 1.4(b) de la página 20.

#### Perfil medio

En la imagen 2.1, además de los perfiles de los cuatro países, tenemos una fila adicional que hemos llamado *media*. Se trata del perfil de la fila final [13 31 42] de la imagen 1.3, que contiene la suma de las columnas de la tabla; es el perfil de todos los viajes considerados conjuntamente. En el capítulo 3 se explica más específicamente porqué lo llamamos *perfil medio*. Por el momento, basta con darnos cuenta de que de los 86 días de viaje, sin tener en cuenta el país visitado, el 15% fueron *festivos*, el 36% *medias jornadas* y el 49% *jornadas completas* de trabajo. Cuando comparemos perfiles, podemos comparar el perfil de un país con el de otro, pero también el perfil de un país con el perfil medio. Por ejemplo, en los datos de la imagen 2.1, vemos que los países con perfiles más parecidos son los de Canadá y de Francia/Alemania. Podemos ver que, comparados con el perfil medio, ambos tienen un mayor porcentaje de *jornadas completas* y, en cambio, están por debajo de la media por lo que respecta a *festivos* y *medias jornadas*.

#### Perfiles fila y perfiles columna

Hasta ahora nos hemos fijado en los perfiles fila con el objetivo de comparar los diferentes países entre sí. Sin embargo, también podemos considerar la imagen 1.3 como un conjunto de columnas y comparar cómo se distribuyen los diferentes tipos de día con relación a los países. En la imagen 2.2 mostramos los perfiles de las columnas, que llamaremos perfiles columna, así como el perfil columna medio. Por ejemplo, de los 13 *festivos*, el 46% fueron en Noruega, el 8% en Canadá, el 31% en Grecia y el 15% en Francia/Alemania. Podemos comparar los valores

PAÍS	<i>Festivos</i>	<i>Medias jornadas</i>	<i>Jornadas completas</i>	<i>Media</i>
Noruega	0,46	0,03	0,26	0,21
Canadá	0,08	0,10	0,26	0,17
Grecia	0,31	0,81	0,00	0,34
Francia/Alemania	0,15	0,06	0,48	0,28

**Imagen 2.2:**

*Los perfiles de los tipos de día con relación a los países, y el perfil columna medio*

de los perfiles de los tipos de día con los valores del perfil columna medio, para ver si sus valores están por encima o por debajo de los de la media. Así por ejemplo, el 46% de los *festivos* los pasé en Noruega, no obstante, el número de días que pasé en Noruega fue sólo el 21% del total de los 86 días de viaje, un gran número de *festivos* comparado con la media.

Veamos otra vez la proporción de 0,46 (6/13) de *festivos* que pasé en Noruega (imagen 2.2) y comparémosla con la proporción 0,21 (18/86) de todos los días que pasé en este país. Podríamos calcular el cociente  $0,46/0,21 = 2,2$ , y llegar a la conclusión de que *festivos* en Noruega está ligeramente por encima de dos veces la media. Llegamos exactamente a la misma conclusión si hacemos un cálculo similar con los perfiles fila. Así, en la imagen 2.1, podemos ver que la proporción de *festivos* en *Noruega* era de 0,33 (6/18) mientras que, considerando todos los países conjuntamente, la proporción de *festivos* era de 0,15 (13/86). Vemos que  $0,33/0,15$  es 2,2 veces mayor que la media. La misma proporción que hemos obtenido a partir de los perfiles columna (llamamos a esta proporción *cociente de contingencia*, lo veremos de nuevo en los próximos capítulos). Tanto si razonamos a partir de los perfiles fila como a partir de los perfiles columna, llegamos a la misma conclusión. En el capítulo 8 mostraremos que el AC analiza de forma similar las filas y las columnas de una tabla de contingencia; también podríamos decir que trata filas y columnas de forma *simétrica*.

Tratamiento simétrico de las filas y de las columnas

A pesar de ello, en la práctica, a menudo vemos y consideramos las tablas de datos de forma no simétrica, o *asimétrica*, ya sea como un conjunto de filas o un conjunto de columnas. Por ejemplo, dado que cada fila de la imagen 1.3 constituye un viaje distinto, podría ser más natural ver la tabla como formada por filas, como en la imagen 2.1. La decisión sobre cuál es la forma más adecuada de analizar una tabla depende de la naturaleza de los datos y de los objetivos de investigación; a menudo no es una decisión consciente. Para conocer la decisión adoptada por un determinado investigador, tenemos que fijarnos en cómo éste interpreta los datos, si se refiere a porcentajes en filas o a porcentajes en columnas. Sin embargo, cualquiera que sea la decisión, los resultados del AC no dependen de esta elección.

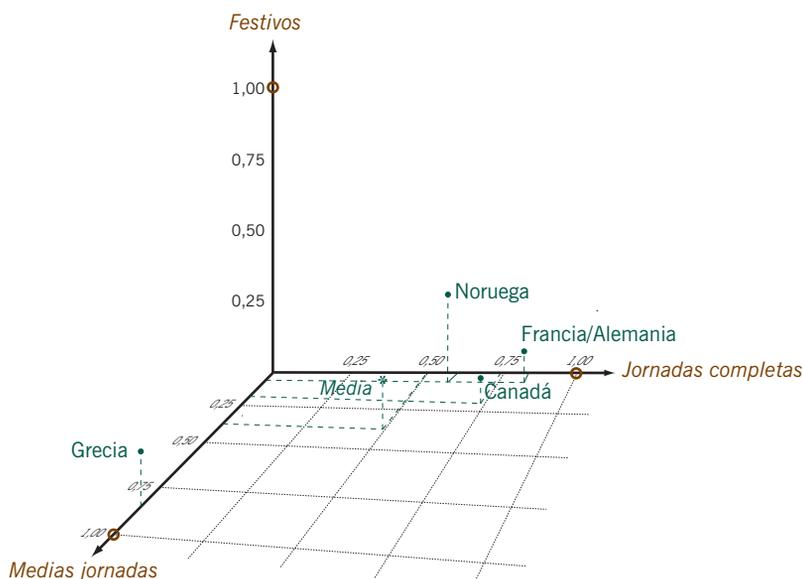
Consideración asimétrica de una tabla de datos

### Representación de los perfiles en el espacio de perfiles

Consideremos ahora una forma completamente distinta de representar los cuatro perfiles fila y el perfil fila medio de la imagen 2.1. A diferencia de la imagen 1.4(b), en la que el eje horizontal identificaba el tipo de día y el eje vertical representaba los porcentajes de días, ahora proponemos utilizar tres ejes, que correspondan a los tres tipos de día, a modo de diagrama de dispersión tridimensional. Imaginar tres ejes perpendiculares no es difícil: basta con mirar el suelo de la habitación en la que trabajamos y buscar una esquina despejada, veremos tres ejes como los que mostramos en la imagen 2.3. En cada uno de estos tres ejes de la habitación podríamos situar a uno de los tres elementos del perfil. Podemos considerar estos tres elementos como las coordenadas de un punto que represente todo el perfil —situación bastante distinta de la del gráfico de la imagen 1.4(b), en la que teníamos un punto distinto para cada uno de los elementos del perfil—. Etiquetamos los tres ejes como *festivos*, *medias jornadas* y *jornadas completas* y los calibramos de 0 a 1. Ahora, representar los cuatro perfiles es un ejercicio sencillo. El perfil de Noruega [0,33 0,06 0,61] (imagen 2.1), se halla a 0,33 unidades en el eje *festivos*, a 0,06 unidades en el eje *medias jornadas* y a 0,61 unidades en el eje *jornadas completas*. Otro ejemplo es el perfil de Grecia [0,14 0,86 0,00], que toma el valor cero en la dirección *jornadas completas*, por tanto su posición se halla en la «pared» de la izquierda, definida por las coordenadas 0,14 y 0,86 en los ejes *festivos* y *medias jornadas*, respectivamente, que delimitan la «pared». A los restantes perfiles los podemos representar de la misma manera en el espacio tridimensional, incluyendo el perfil fila medio [0,15 0,36 0,49].

#### Imagen 2.3:

Posiciones de los cuatro perfiles fila (●) de la imagen 2.1 así como la del perfil fila medio (\*) en un espacio tridimensional, que hemos representado como la esquina de una habitación con baldosas en el suelo. Así, por ejemplo, Noruega toma el valor 0,06 en el eje *medias jornadas*, 0,61 en el eje *jornadas completas* y 0,33 en la dirección vertical correspondiente al eje *festivos*. En cada eje hemos representado los puntos correspondientes a la unidad (vértices) por círculos huecos (○)

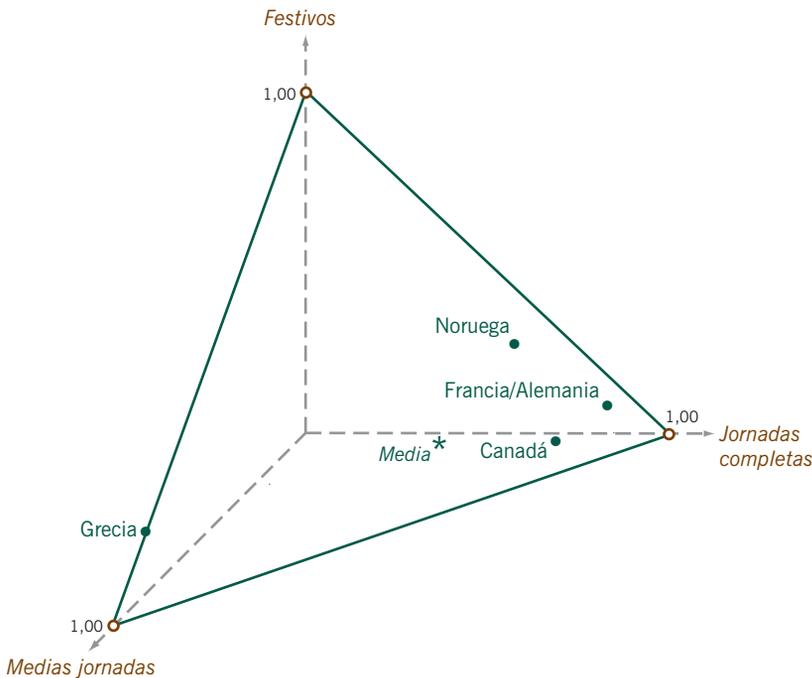


Con un poco de imaginación podríamos ver, como hemos representado gráficamente en la imagen 2.4, que los perfiles de la imagen 2.3 se hallan todos exactamente en un plano definido por un triángulo equilátero que une los tres *puntos unidad* [1 0 0], [0 1 0] y [0 0 1], situados en sus respectivos ejes. Llamaremos vértices a los tres extremos de este triángulo equilátero. Los vértices coinciden con los perfiles extremos, es decir perfiles totalmente concentrados en un solo tipo de día. Por ejemplo, el vértice [1 0 0] corresponde a un viaje con sólo *festivos* (desafortunadamente, ficticio en mi caso). De la misma manera, el vértice [0 0 1] corresponde a un viaje con sólo *jornadas completas* de trabajo.

Visto que, en realidad, en el espacio tridimensional, todos los perfiles se hallan en un triángulo (bidimensional), podemos situar este triángulo en un plano, como en la imagen 2.5. ¡Mirar los perfiles en un plano es más cómodo que intentar imaginar sus posiciones tridimensionales en la esquina de una habitación! A este tipo especial de representaciones las llamaremos *sistemas de coordenadas triangular* (o *ternario*). Las podemos utilizar siempre que tengamos datos compuestos de tres elementos que sumados den 1, como es el caso de los perfiles fila de nuestro ejemplo. Este tipo de datos es frecuente, por ejemplo, en geología y en química, disciplinas en las que se obtienen muestras que se caracterizan por la pro-

Los vértices definen los extremos del espacio de perfiles

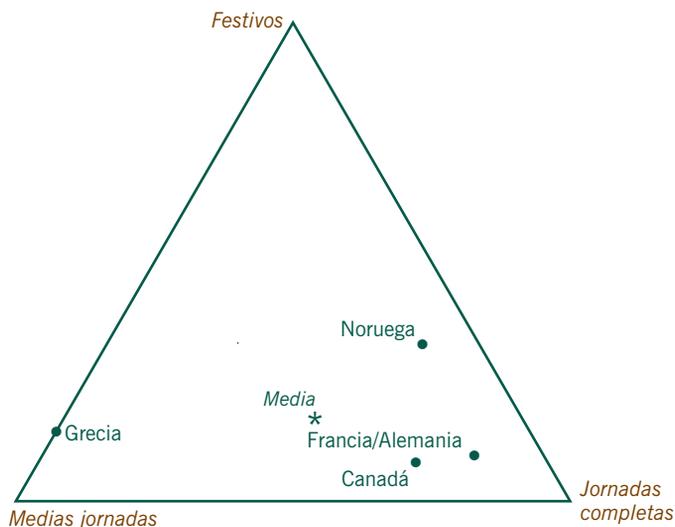
El sistema de coordenadas triangular (o ternario)



**Imagen 2.4:**  
 Los perfiles de la imagen 2.3 se hallan en un triángulo equilátero formado uniendo los vértices del espacio de perfiles. Por tanto, los perfiles tridimensionales son, en realidad, bidimensionales. El perfil de Grecia se halla en el borde del triángulo debido que su valor para jornadas completas es cero

**Imagen 2.5:**

El triángulo de la imagen 2.4 con los perfiles fila (países). Las tres esquinas, o vértices, del triángulo representan las columnas (tipo de día)



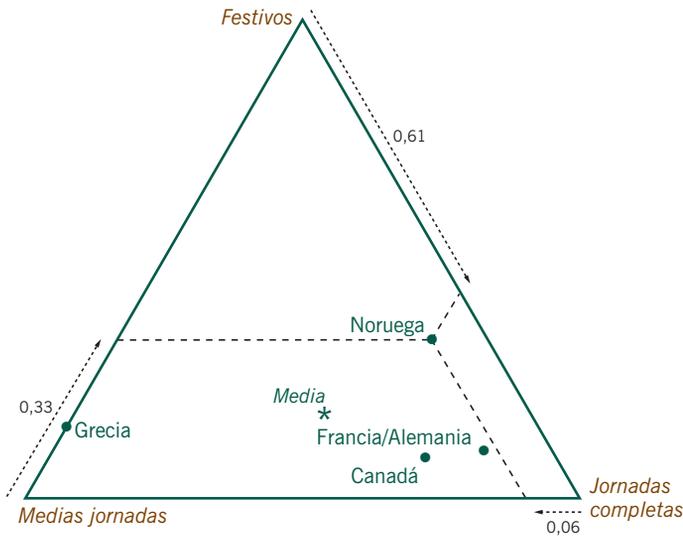
porción de tres elementos, en peso o volumen, y que por tanto podemos representar gráficamente como un solo punto en un sistema de coordenadas triangular (o ternario).

#### Situación de los puntos en un sistema de coordenadas triangular

Supongamos que tenemos un triángulo equilátero en blanco y los valores de los perfiles, ¿cómo podemos situar los perfiles sin tener que pasar por el espacio tridimensional que mostramos en las imágenes 2.3 y 2.4? En el sistema de coordenadas triangular, los lados del triángulo definen tres ejes. Suponemos que cada lado tiene una longitud igual a 1 y que los calibramos de forma lineal de 0 a 1. A continuación situamos los valores de los perfiles en sus respectivos ejes. Por ejemplo, como vemos en la imagen 2.5, para situar a Noruega tomamos el valor 0,33 en el eje *festivos*, 0,06 en el eje *medias jornadas* y 0,61 en el eje *jornadas completas*. A partir de estos valores de los perfiles, podemos trazar líneas paralelas a los lados del triángulo. Estas líneas paralelas confluyen en un punto que define la posición de Noruega (imagen 2.6). En realidad, es suficiente cualquier combinación de dos de las tres coordenadas de los perfiles, para situar los perfiles de la manera indicada. La tercera coordenada es, en realidad, innecesaria. Es otra manera de demostrar que los perfiles son intrínsecamente bidimensionales.

#### Geometría de los perfiles con más de tres elementos

Sólo podemos utilizar el sistema de coordenadas triangular para perfiles con tres elementos. Sin embargo, como veremos en el próximo capítulo, podemos generalizar fácilmente esta idea a perfiles con cualquier número de elementos. En tal caso llamamos al sistema de coordenadas *sistema de coordenadas baricéntri-*



**Imagen 2.6:**  
Noruega  
[0,33 0,06 0,61]

co («baricentro» es sinónimo de «media ponderada»). La dimensionalidad de este sistema de coordenadas es siempre igual al número de elementos de los perfiles menos uno. Así, acabamos de ver que los perfiles con tres elementos se hallan exactamente en un espacio de perfiles triangular de dos dimensiones. La dimensionalidad de los perfiles con cuatro elementos es tres: los perfiles se hallan en un tetraedro de cuatro extremos situado en un espacio tridimensional. El triángulo bidimensional y el tetraedro tridimensional son ejemplos de lo que en matemáticas se llama un *simplex regular*. Con el fin de poder percibir un espacio de perfiles tridimensional, en el apéndice de cálculo, indicamos la codificación R que permite visualizar un ejemplo tridimensional. Para perfiles de dimensión superior a tres, necesitaríamos para ser capaces de «ver» el espacio de perfiles con mayores dosis de imaginación. Afortunadamente, como veremos más adelante, el AC nos ayudará mucho a visualizar este tipo de perfiles multidimensionales.

Hasta ahora, hemos visto el concepto de perfil en un contexto de datos de frecuencias, el principal tipo de datos que utilizamos en el AC. Sin embargo, el AC lo podemos aplicar a un abanico mucho más amplio de tipos de datos. De hecho lo podemos aplicar siempre que tenga sentido expresar los datos como valores relativos, es decir, que podamos expresar los datos en una *escala de razón*. Por ejemplo, supongamos que tenemos datos sobre los importes de dinero invertidos por diferentes países en distintas áreas de investigación —los valores relativos de interés podrían ser, por ejemplo, los porcentajes invertidos en medio ambiente, en biomedicina, etc.—. Otro ejemplo podrían ser las medidas

Datos en una escala de razón

morfométricas de organismos vivos, por ejemplo peces, como la longitud total, la anchura, la longitud de las aletas, etc., en centímetros, que podríamos expresar con relación a la suma total. Dicha suma total sería una manera de expresar el tamaño del pez, de modo que podríamos analizar y comparar a partir de los perfiles de los diferentes peces, y no de las medidas originales que constituyen su morfología.

#### Datos en una escala común

Una condición necesaria para los datos en el AC es que expresemos todas las observaciones en la misma escala. Por ejemplo, los recuentos de individuos en una tabla de frecuencias, una determinada unidad monetaria en una tabla sobre investigación de inversiones, o las medidas en centímetros en un estudio morfométrico. En el AC no tendría sentido que analizáramos datos expresados en distintas escalas de medida. A no ser que hagamos una transformación previa que nos permita homogeneizar las distintas escalas de la tabla. La mayor parte de los datos de este libro son datos de frecuencias, sin embargo, en el capítulo 23 se analiza una amplia variedad de otro tipo de datos. Veremos también cómo recodificarlos para que sean apropiados para el AC.

#### RESUMEN: Perfiles y espacio de perfiles

1. Un *perfil* es un conjunto de frecuencias (u otros valores positivos o ceros) divididas por su total, es decir es un conjunto de frecuencias relativas.
2. En las tablas de contingencia, las filas y las columnas definen conjuntos de frecuencias, que podemos reexpresar con relación a sus respectivos totales para obtener así perfiles fila o perfiles columna.
3. También podemos expresar las frecuencias marginales de las tablas de contingencia con relación a sus totales (el total de la tabla) para obtener así el perfil fila medio y el perfil columna medio.
4. Comparando los perfiles fila con su media, llegamos a las mismas conclusiones que comparando los perfiles columna con su media.
5. Podemos representar los perfiles con  $m$  elementos como puntos en un espacio  $m$ -dimensional. Sin embargo, debido a que la suma de estos  $m$  elementos es 1, los perfiles ocupan, en realidad, una región restringida de este espacio de dimensionalidad  $(m - 1)$  que llamaremos *símplex*. Las líneas de unión de todos los pares de los  $m$  puntos unidad de los  $m$  ejes perpendiculares delimitan el símplex. A estos puntos unidad les llamamos *vértices* del símplex o del espacio de perfiles. Llamamos *sistema de coordenadas baricéntrico* al sistema de coordenadas definido por el símplex.
6. Es fácil visualizar los perfiles con tres elementos. El símplex formado por la unión de los tres vértices es simplemente un triángulo equilátero. A este caso particular de sistema de coordenadas baricéntrico lo llamamos *sistema de coordenadas triangular* (o *ternario*).

7. Podemos ampliar la idea de perfil a datos expresados en *escalas de razón*, es decir, a datos expresados como valores relativos. En dicho caso, es necesario haber obtenido todos los datos originales en la misma escala de medida.