

Fantasías Lineales: Una Revisión del Análisis de Correspondencias



Jorge Alagón y Felipe González
Millward Brown México

Hablar de fenómenos “no lineales” es absurdo: es como hablar de los “no elefantes” en un zoológico.

Stanislaw Ulam

1. Hipótesis Engañosas

El análisis de correspondencias es una herramienta popular para hacer análisis exploratorio de datos categóricos, por ejemplo, de los niveles de respaldo para diferentes marcas a lo largo de un grupo determinado de atributos. La idea es obtener un mapa perceptual donde se representan gráficamente las distintas relaciones que hay entre las marcas y los atributos. Las ventajas comúnmente citadas de este análisis son: que no hace supuestos importantes acerca de los datos (basta que los datos sean de tipo categórico), y que el mapa perceptual que se obtiene es de fácil interpretación. Una exposición de este método se puede encontrar en el número anterior de esta misma publicación ([VE]).

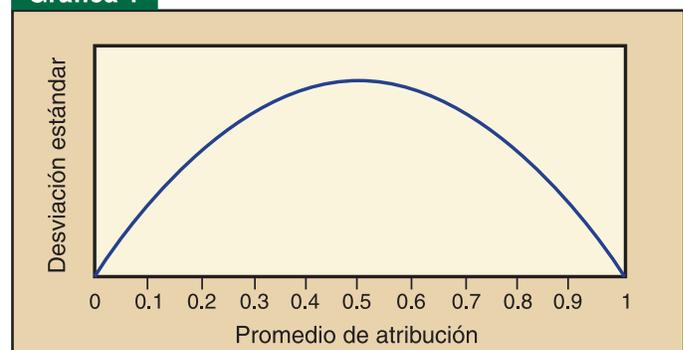
En cuanto a la primera ventaja, no sobra señalar que justamente cuando es difícil encontrar los supuestos, es cuando debemos ser más cuidadosos: esto sólo quiere decir que las hipótesis importantes están escondidas detrás del aparato técnico, o peor aún, que de tanto usar las mismas hipótesis hemos aprendido a dejar de verlas. La segunda ventaja es relativa y al final depende de que podamos resolver satisfactoriamente la cuestión de las hipótesis.

Quizá el supuesto más importante del análisis de correspondencias es la *estructura lineal de los datos*. El análisis de correspondencias es, a fin de cuentas, *un método lineal basado en la proyección lineal de vectores en un espacio lineal*. Muchas de las aplicaciones del análisis de correspondencias en investigación de mercados se realizan sobre tablas de imagen con porcentajes (casos), pero de entrada sabemos que el comportamiento de los porcentajes (casos) *no* es lineal. El supuesto de linealidad tiene los siguientes efectos indeseables:

Primero, para calcular el centroide de la tabla, a partir del cual se hace el análisis de correspondencias, se promedian los niveles de atribución a lo largo de marcas y atributos. Esto no siempre es correcto. Como ejemplo, supongamos que en cierto atributo dos marcas X y Y, ambas bien conocidas, obtienen un respaldo del 30% y del 2% respectivamente, del total de entrevistados. El promedio de atribución es entonces 16%. Ahora preguntamos al lector, ¿qué parece más probable, que X baje del 30% al 16% o que Y suba del 2% al 16%? Parecería que con el promedio estamos sobreestimando un “justo medio” entre los dos niveles de atribución.

El segundo aspecto indeseable del supuesto de linealidad está en la estructura de varianza que asume. En el análisis de correspondencias, al calcular la inercia, esperamos variaciones chicas para porcentajes chicos y variaciones grandes para porcentajes grandes. Esto se puede ver directamente en la fórmula de la inercia total. Pero porcentajes muy chicos (cercanos a 0%) o muy grandes (cercanos al 100%) tienen poco espacio para variar, mientras que porcentajes alrededor del 50% tienen más espacio. Es difícil justificar el tratamiento asimétrico de los extremos de la escala en correspondencias (ver gráfica 1).

Gráfica 1





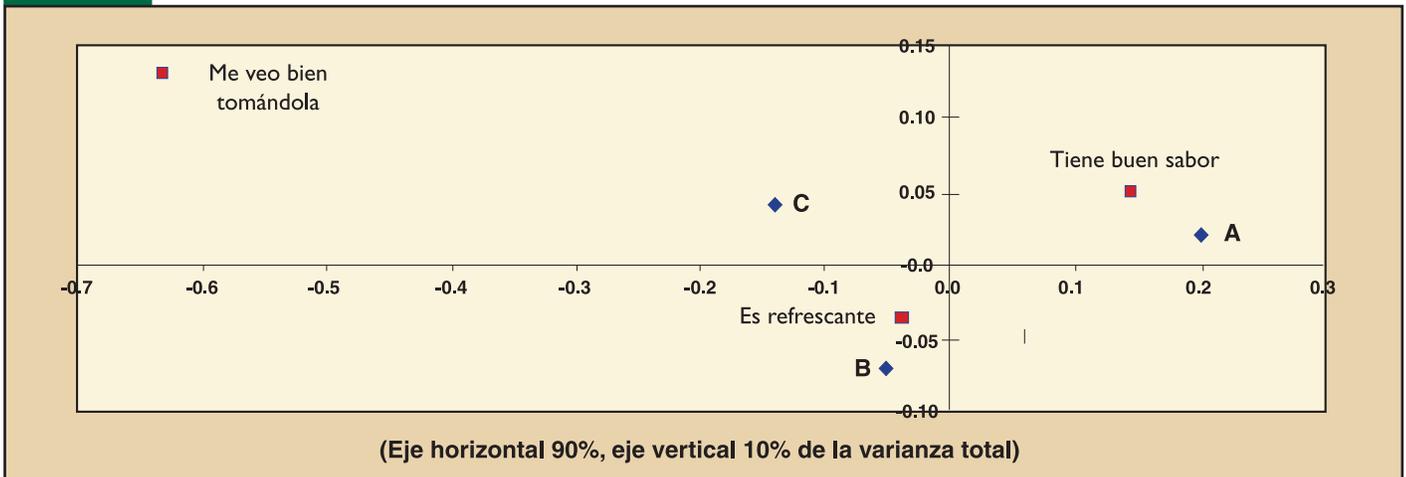
2. Dos Casos Concretos

Considérese la tabla 1 ficticia de atribuciones, donde A, B y C son marcas de bebidas dadas:

	A	B	C
Es refrescante	80	90	95
Tiene buen sabor	55	40	45
Me veo bien tomándola	1	5	10

Las celdas de la tabla contienen los porcentajes de respaldo en cada atributo para cada marca sobre la muestra total. La gráfica 2 es el mapa de correspondencias simétrico ([Gr]) obtenido para la tabla 1.

Gráfica 2

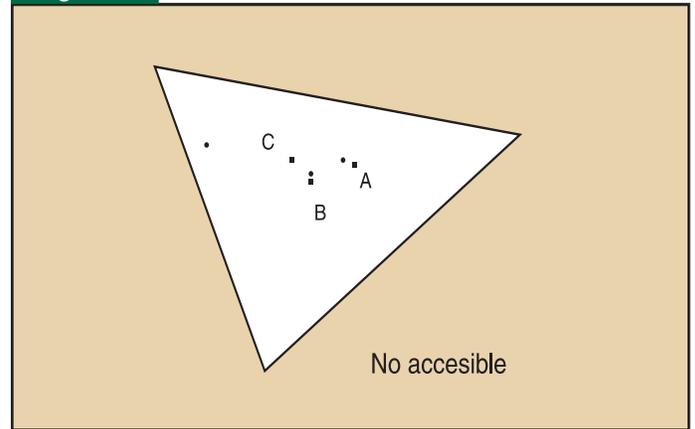


El atributo más importante de la categoría, “Es refrescante”, explica muy poca varianza en los datos (6% de la inercia total), mientras otro que tiene muy poco respaldo, “Me veo bien tomándola”, explica mucha (64%). Así que para empezar, probablemente estamos dando demasiado peso al atributo menos mencionado. Y peor aún, no estamos reconociendo que esos 5 puntos porcentuales que separan a C de las demás marcas en “Es refrescante” indican que C es líder en este atributo.

Otro punto débil del mapa perceptual basado en el análisis de correspondencias es que no todas sus partes corresponden a cifras realizables en la tabla de imágenes: hay partes que no forman parte del paisaje perceptual y que no tienen ningún significado. Típicamente la parte que tiene sentido en el mapa perceptual consiste en una figura convexa que contiene a todos los puntos incluídos en el análisis (ver

figura 1, donde el área no accesible está señalada en beige).

Figura 1



Un ejemplo con datos reales es el siguiente. La tabla 2 muestra los porcentajes de respaldo (*top two boxes*, en una escala de cinco puntos) para cada marca en cada atributo:

Con el método de correspondencias, por ejemplo, el atributo “prestigio” explica el 3% de la inercia, mientras 25% es explicada por “barata”. En este punto nos preguntamos: ¿Es correcto que “barata” discrimine más que “prestigio”, aún cuando evidentemente el atributo “prestigio” divide a las marcas en dos grupos claros, de manera mucho más dramática que el atributo “barata”?

Más aún, resulta ser que el atributo “prestigio” es muy importante para la categoría en cuanto a consideración de compra. Correspondencias podría sugerir que ninguna marca se apropia del todo de este atributo y que quizá hasta podría ser un área de opor-

Tabla 2

	A	B	C	D	E	F	G	H	I	J	K
Prestigio	59	41	93	94	61	68	62	94	92	82	48
Calidad	56	45	88	80	59	50	50	88	78	76	37
Diferente	42	38	68	79	51	47	37	76	57	54	37
Gustaría vieran consumir	30	39	67	55	52	38	36	61	56	49	28
Barata	46	73	58	61	55	69	41	44	60	65	66
Tradición	39	23	82	82	53	54	42	82	81	67	26
Disfruta consumir	47	42	72	87	61	50	51	82	81	66	57
Natural	25	24	54	21	49	26	20	66	48	49	32
Nutritiva	43	7	69	13	53	27	33	80	15	37	15
Contribuye a la salud	32	24	57	24	48	25	28	70	27	40	17
Inspira confianza	53	34	83	57	62	51	40	79	64	58	35
Buen sabor	55	39	66	74	65	56	49	69	61	78	47

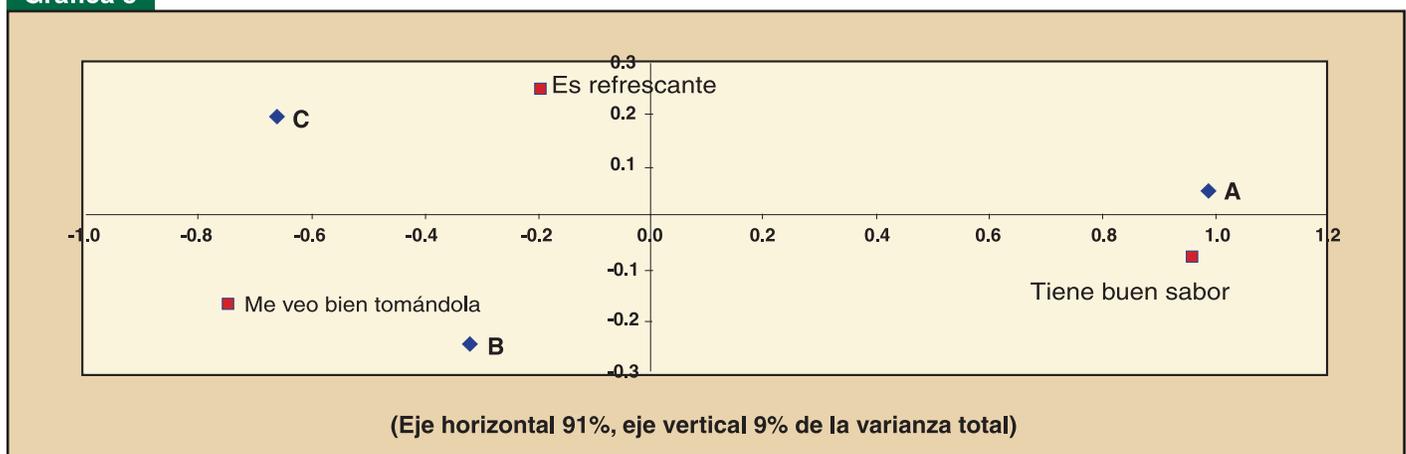
tunidad. El lector se dará cuenta fácilmente de cómo nos estamos acercando peligrosamente hacia una conclusión errónea: la realidad es que “prestigio” es importante en consideración de compra, discrimina bien entre las marcas, y debe ser reconocida como propiedad de las marcas C, D, H e I.

3. Mejores Mapas Perceptuales

Por lo menos en el área de la regresión, durante los últimos 30 años se han establecido métodos para tratar con el tipo de fenómenos mencionados arriba (por ejemplo, la regresión logística y la probit [A]), y hoy pocos nos empeñaríamos en usar métodos puramente lineales para estudiar las dependencias de un conjunto de datos categóricos.

Ya existen varias alternativas mejores para construir mapas perceptuales. El trabajo de Aitchison sobre bimapas (*biplots*) de variaciones relativas (ver [AG]) y los ratio maps de Greenacre, basados en el trabajo de Aitchison ([G]), apuntan en esta dirección. Otra posibilidad que presentamos aquí es la de hacer mapas de *logits*. En primer lugar, transformamos la tabla 1, por ejemplo, en valores *logit* (logaritmo de los nomios correspondientes). Con esta nueva tabla hacemos un bimapa simétrico (ver [K] p. 128-34, donde también se estudian otras variaciones útiles de los bimapas). Se pueden dar muchas razones a favor de pensar en las unidades de los *logits* como cantidades lineales, tanto teóricas (en la mecánica estadística, por ejemplo, son de uso común para pasar de distribuciones a energías que son aditivas, [Go])

Gráfica 3



como empíricas. Para la tabla 1, el mapa que resulta es el que se muestra en la gráfica 3.

La interpretación de este bimapa (ver [K]) sugiere, en contraste con el mapa de correspondencias, que B tiene una fortaleza en el atributo “Me veo bien tomándola”, y que en realidad no es especialmente fuerte en el atributo “Es refrescante”. Además, C es claramente líder en todos los atributos menos “Tiene buen sabor”. Alguien podrá argumentar que estas son sutilezas. Y esa persona tendría razón. Pero también hay que recordar que muchas veces *los hallazgos importantes se esconden detrás de esas sutilezas*.

Otra ventaja de este mapa es que cada punto en él corresponde a configuraciones posibles de atribuciones. Dicho de otra manera, si alguien señala un punto en el mapa y nos pregunta, ¿qué atribuciones necesitaría para que una marca X estuviera aquí?, le podemos responder fácilmente. En el mapa de correspondencias, como señalamos arriba, esto es en ocasiones imposible. Para el ejemplo de la tabla 2 tenemos el siguiente bimapa simétrico (gráfica 4), donde se resuelven los problemas señalados arriba: aquí, prestigio, nutritiva y barata son los atributos que discriminan más.

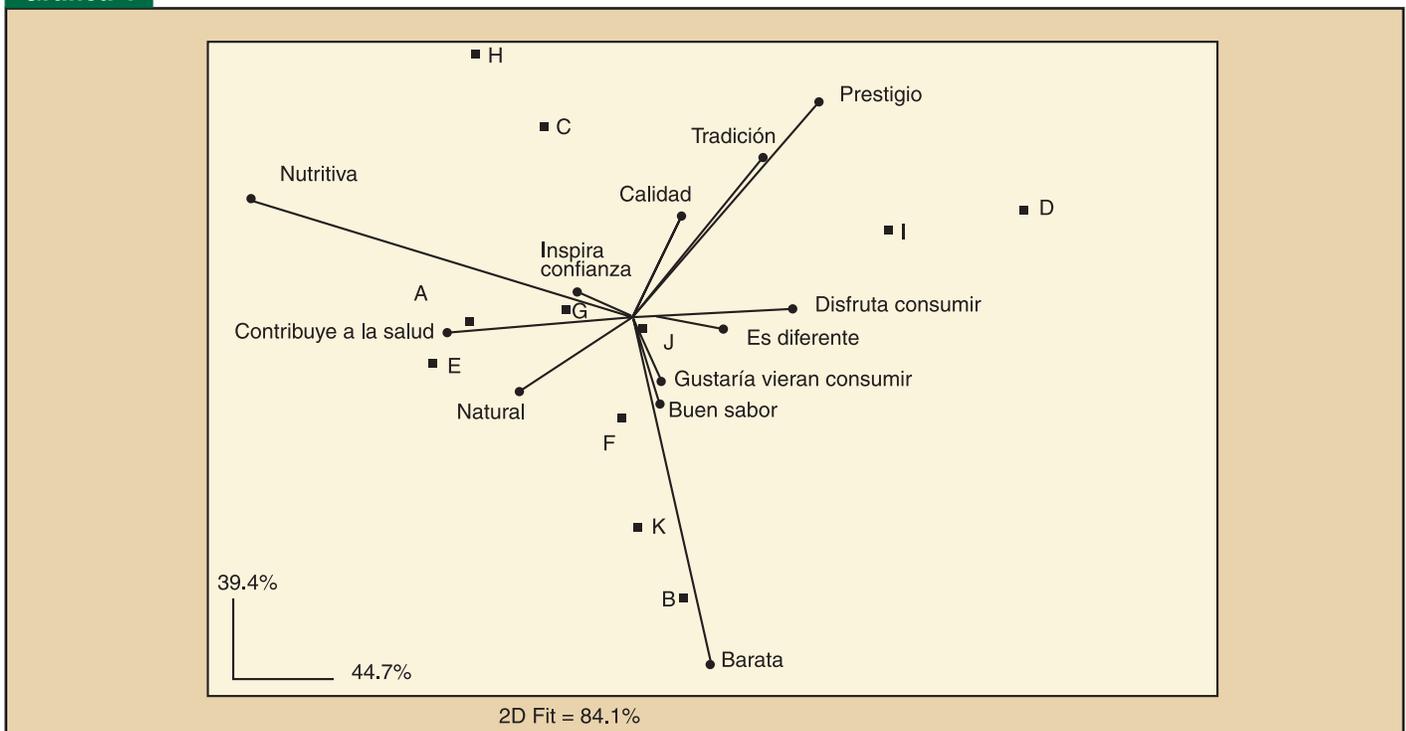
4. Conclusiones

El uso de mapas perceptuales (correspondencias o bimapas) siempre debe estar respaldado por otros métodos más cuidadosos y precisos de perfilación. Nosotros preferimos utilizar, por ejemplo, otras técnicas más refinadas como los *Brand Image Profiles* ([B]), que corrigen tanto los problemas mencionados arriba como otros que aparecen, por ejemplo, cuando hay distorsiones por el dominio claro de una marca sobre un atributo de su categoría o por atributos muy fuertemente correlacionados con el tamaño de marca.

Métodos como el bimapa de logits han demostrado ser superiores tanto desde el punto de vista práctico como del teórico. Desde el punto de vista teórico, la superioridad es clara. Desde el punto de vista práctico, además de lo que hemos mostrado en este artículo, métodos similares como el de Aitchison ([AG]) ya han sido adoptados en el análisis composicional de muestras geológicas ([ABP], [AG]), donde la naturaleza muy precisa de las mediciones hace necesario abandonar el análisis de correspondencias clásico.

Aunque en la investigación de mercados el análisis de correspondencias sigue siendo útil y a veces hasta

Gráfica 4



convinciente, cada vez requerimos de técnicas y herramientas más precisas que nos permitan hacer hallazgos más sutiles, que estén respaldados por mayor rigor técnico. Lo presentado aquí, más que una crítica de las correspondencias, es una crítica del uso indiscriminado de los métodos lineales en la investigación de mercados, donde prácticamente no aparecen datos lineales de manera natural. El mundo simplemente no se comporta linealmente. Forzar nuestras predicciones y nuestro campo de atención a micromundos lineales tenía una importante razón de ser antes de que tuviéramos el poder de computación que tenemos hoy: los métodos no lineales requieren de mucho más esfuerzo computacional. Pero en estos días de computación barata, ya no es tan fácil justificar el sacrificio de la precisión y el rigor estadístico en el altar de la fantasía lineal.

Bibliografía

[A] Agresti A. *An Introduction to categorical data analysis*, Wiley Series in Probability and Statistics, Wiley-Interscience, 1998.

[ABP] Aitchison, J., Barceló-Vidal y Pawlowsky-Glahn *Some comments on compositional data analysis in archaeometry*, prepublicación aceptada en *Archaeometry*, 2002.

[AG] Aitchison, J. y Greenacre, M. *Biplots of compositional data*, *Journal of the Royal Stat. Society Series C (Applied Statistics)* 51 (4), p. 375-392, 2002.

[B] Brown. G., *Findings from brand attitude grids*, Millward Brown, 1986

[G] Greenacre, M., *Ratio maps and correspondence analysis*, prepublicación, 2002.

[Go] González, F. *Some integral transforms in the thermodynamic formalism*, Tesis de doctorado, Universidad de Warwick, 2002.

[Gr] Greenacre, M. *Theory and applications of correspondence analysis*, Academic Press, 1984.

[K] Krzanowski, W.J. *Principles of multivariate analysis* (Oxford Statistical Science Series v. 3), Oxford Science Publications, University Press, 1990.

[VE] Vargas, R. y Escobedo, N. *Análisis de correspondencias: más allá de un mapa*, Datos, Diagnósticos y Tendencias, Año 10, No. 36, Abril 2003.

¿ Cuántas llamadas necesitas hacer para preguntar
a todos ellos si les gusta lo que ofreces ?



Sólo una
llama a

 MERCADOTECNICO
FACTUM

investigación de mercados con sentido business

Tels: 52 86 64 00
52 86 51 18
52 11 66 67
Fax: 52 86 97 24

Acapulco 47
Col. Roma Norte
e-mail: factum99@prodigy.net.mx

No tome decisiones a oscuras

FACT BOOK® 2003
Tendencias del
Mercado Mexicano

Dirija su camino con la luz que
guiará sus decisiones de negocio.

Saber cómo se comporta el mercado mexicano,
resulta ser la guía ideal para conocer en dónde
se encuentran realmente, los retos y
oportunidades que enfrenta su negocio.

El desempeño en ventas de 99 categorías de
productos de consumo masivo, el perfil del
consumidor, así como los datos
macroeconómicos y demográficos más
relevantes de nuestro país, son sólo algunos
de los indicadores que ahora están a su alcance
a través de esta deslumbrante herramienta de
consulta bilingüe y exportable a Excel®.

¡Click!

El Fact Book® 2003 lo iluminará.

Llámenos ahora mismo y reciba una muestra
del producto TOTALMENTE GRATIS.

Ordene el Fact Book® vía Internet:

www.acnielsen.com/mx

