

Cómo se lleva a cabo un análisis de las preferencias mediante la ordenación de perfiles

Escribe Vicente Manzano-Arrondo

1. El sentido de la tarea

En una investigación sobre la feminización de la pobreza a partir de la contratación de personal, se diseñó un cuestionario que recogía información de quienes tienen la responsabilidad de contratar personal en un conjunto amplio de empresas en Andalucía Occidental. El instrumento contaba con ítems tipo Likert de grado de acuerdo en el se encontraban afirmaciones sobre la relevancia del género en el perfil de la persona que busca empleo. Para observar la importancia relativa del género, se incluyeron variables de perfil diverso, como tener o no hijos, ser inmigrante o autóctono o poseer una cualificación alta o baja. El cuestionario incluía afirmaciones como "Prefiero contratar a una persona autóctona antes que a alguien inmigrante". Los resultados mostraron que el género no influye en las decisiones de la persona responsable de contratar.

Además del cuestionario se diseñaron tarjetas. Cada tarjeta definía un hipótico solicitante de empleo: hombre o mujer, inmigrante o autóctono, con alta o baja cualificación, etc. Las características de perfil se combinaron en un conjunto de tarjetas y se pidió a quienes contratan que ordenaran las tarjetas en función de su preferencia: las primeras son las que contrataría antes. He aquí un ejemplo de tarjeta: mujer, sin hijos, autóctona, alta cualificación, disponible a media jornada.

Los resultados fueron muy ilustrativos. El cuestionario no suministró diferencias de preferencia para el género. Pero las tarjetas mostraron sin lugar a dudas, que las mujeres eran preferidas a los hombres cuando el puesto de trabajo es de baja calidad (inestable, horas parciales, sueldo bajo), mientras que los hombres eran preferidos para los mejores puestos, con independencia de la cualificación u otras características.

¿Por qué ocurrió esto? La persona que responde al cuestionario tiene ante sí una tarea sencilla. Puede concentrarse en cada ítem y responder según su criterio. Ese criterio puede coincidir o no con lo que se le pide. Puede ocurrir que las respuestas se vean afectadas por la deseabilidad social (se responde en función de lo que se considera socialmente más aceptado) y la aquiescencia (se responde en función de lo que se considera que satisface a quienes han ideado el cuestionario, tendiendo a estar de acuerdo con los enunciados). De este modo, quienes tienen la responsabilidad de contratar procuraron mostrar que no realizan su labor de manera sesgada, de tal forma que no penalizan a las mujeres, la inmigración, tener hijos, etc. Y pueden controlar esta imagen, consciente o inconscientemente, mediante sus respuestas a cada ítem.

No obstante, esta tarea es inviable en una ordenación de perfiles. Cada una de ellas se ha diseñado de tal modo que no se pueden controlar todas las variables al mismo tiempo. Hay que optar. Es una tarea compleja. En su origen, suele hablarse de *análisis conjunto*, pero es una denominación poco ilustrativa, tomada literalmente del inglés, en cuya versión también es poco acertada. Es muy frecuente utilizar tarjetas, por lo que se podría denominar *análisis de tarjetas*. Pero en muchas ocasiones el estímulo complejo puede ser una colección de botellas de cerveza (que varían en altura, color del vidrio, anchura, posición de la etiqueta...) o cepillos de dientes (que varían en color, forma, material, etc.). Por eso me gusta llamar a esto *análisis de perfiles*.

2. Como se diseñan los perfiles

Como en cualquier investigación, lo más importante es el objetivo. Hay que saber qué es lo que se pretende medir con los perfiles. Sigamos otro ejemplo. Queremos estudiar la conducta de publicación en revistas por parte del profesorado universitario. Tras un estudio de la literatura y una recogida de datos previa, se identifican las variables y valores que puedes encontrar en la tabla 1.

Tabla 1. Ejemplo de variables y categorías para análisis de perfiles.

#	Variable	Categoría 1	Categoría 2
1	Idioma de la publicación	Inglés	Español
2	Revista indexada	En Q1* ¹	Fuera de JCR* ²
3	Redacción realizada por	Mí	Otras personas
4	Donde mi orden de firma es	Primero de tres	Quinto de seis
5	Se paga por publicar	Sí	No
6	El artículo se leerá	Mucho	Muy poco
7	Y contará con una utilidad social	Clara	Ninguna

1: Cuartil 1 del listado JCR

2: *Journal Citation Reports* de la empresa Clarivate

Tabla 2. Definición de tarjetas (perfiles) mediante combinaciones ortogonales.

Tabla 2a. Combinaciones ortogonales.

Tarjetas	Variables						
	1	2	3	4	5	6	7
1	1	2	1	2	1	1	2
2	2	1	1	1	1	2	2
3	2	1	1	2	2	1	1
4	2	2	2	1	2	1	2
5	1	2	1	1	2	2	1
6	1	1	2	1	1	1	1
7	1	1	2	2	2	2	2
8	2	2	2	2	1	2	1

Tabla 2b. Comprobación de la ortogonalidad

		2. Indexación	
		1	2
1. Idioma	1	6, 7	1, 5
	2	2, 3	4, 8

En las celdas se muestra la identificación numérica de las tarjetas que cuentan con los valores marginales.

Una vez contamos con las variables y sus valores, hay que generar las combinaciones. En el ejemplo, 7 variables con 2 valores implican $2^7 = 128$ tarjetas. No hay humano que pueda trabajar con eso. Así que utilizamos un conjunto de combinaciones que llamamos *ortogonales*. Son muy pocas, pero suficientes como para que ninguna variable se confunda

con el resto. Observa la tabla 2a, donde se muestra cómo se definen cada una de las ocho tarjetas ortogonales resultantes a partir de las siete variables. Cada uno de los dos valores de cada variable están presentes cuatro veces en el conjunto de las tarjetas. Si se toman dos variables cualesquiera, observarás que hay dos tarjetas para cada combinación de valores de las dos variables. Por ejemplo, tomemos las variables 1 (idioma) y 2 (indexación). Observa que hay dos tarjetas para cada una de las cuatro combinaciones entre ambas variables (tabla 2b). Eso mismo ocurre tomando cualesquiera dos pares. Por ello, los efectos principales de las variables no están confundidos y podremos seguir con el diseño.

Como resultado, se generan ocho tarjetas. En la figura 1 tienes un ejemplo concreto para una de ellas.

Figura 1. Tarjeta de perfil.

Publicado en:	Inglés
En una revista:	Fuera de JCR
Pago por publicar:	Sí
Se leerá:	Muy poco
Utilidad social:	Clara
Orden de firma:	Primero de tres
Escrito por:	Otras personas

¿Cómo se generan combinaciones ortogonales? Técnicamente no se explica en dos párrafos. Si diseñas perfiles basados en variables dicotómicas, la tabla 3 puede serte de gran ayuda. Con una variable necesitarás dos perfiles. Con 2 variables, 4 perfiles. De 3 a 7 variables, 8 perfiles (así son la mayoría de las aplicaciones que llevo a cabo). Y de 8 a 11 variables, 12 perfiles, que todavía resultan manejables.

Tabla 3. Diseño de perfiles ortogonales con variables dicotómicas.

Tar	Número de variables dicotómicas										
	1	2	3	4	5	6	7	8	9	10	11
1	0	1	001	0011	10100	010111	0110011	00000000	111010101	1111100001	11000110011
2	1	00	100	0110	00000	111010	1111000	01101100	111101000	1001001101	00000000000
3		10	111	1011	01110	101101	1010101	11001101	010110011	1100111000	01100011110
4		11	000	1000	10111	000000	1001011	10101010	100101111	0000000000	11001001101
5			110	1110	11010	100110	0101101	11010110	001100101	0100100111	10110010101
6			011	0000	00011	001011	0000000	00111111	110000110	1001110110	00010101111
7			101	1101	11001	011100	0011110	11100011	100011001	0011101010	10011011010
8			010	0101	01101	110001	1100110	10110100	001011110	1110001110	01111000011
9								01110001	101110010	0010111101	01011110100
10								01011010	011001011	1010010011	11110101000
11								00000111	000000000	0101011011	00101111001
12								10011001	010111100	0111010100	10101100110

3. Cómo se recogen los datos

A cada participante se le pide que ordene los perfiles según un criterio concreto: a qué persona contrataría antes, a qué destino turístico le gustaría acudir, cuál de estas personas cree que ha triunfado más claramente, elija la publicación que más le gustaría realizar... Básicamente hay dos estilos para pedir que se ordene: o bien se indica el criterio general, o bien se le pide que escoja un perfil del conjunto total, después que escoja el siguiente de entre los que quedan y así hasta finalizar la ordenación.

La presentación de las tarjetas depende del medio de realización de la tarea. Si es presencialmente, las tarjetas se imprimen, se suministran barajadas y cada participante las ordena siguiendo su propio procedimiento. Si es por ordenador, quien participa no goza de tanta libertad, el modo de programar la utilidad condiciona el método de ordenación. En algunos casos, las tarjetas se presentan por pares, mediante un algoritmo concreto, hasta ordenarlas internamente. En otro, se presentan como ocurre con el juego del solitario. En otros, se van visualizando las tarjetas al pulsar sobre ellas en una fila donde se superponen. Etc. Si el vehículo es el teléfono celular o móvil, el asunto se restringe bastante, pues el tamaño de la pantalla no da para hacer maravillas.

En cualquier caso, la tarea de ordenación genera un listado de números que representan las posiciones de orden. Si contamos con 8 tarjetas y han respondido 6 personas, la apariencia del archivo de datos puede ser algo parecido a lo que se muestra en la tabla 4.

Tabla 4. 6 casos de ordenación de perfiles.

Caso	Posiciones							
	1	2	3	4	5	6	7	8
1	6	7	5	1	2	3	4	8
2	6	2	7	3	5	4	8	1
3	6	2	3	7	4	8	5	1
4	4	6	2	1	8	5	7	3
5	6	2	7	3	8	4	5	1
6	6	2	3	7	4	8	1	5

La participante representada como Caso 1 ha escogido en primer lugar el perfil 6, después el 7, el 5... hasta llegar al menos preferido, el perfil 8.

Obviamente el archivo de datos no solo incluye estas informaciones. Con seguridad habrá variables para medir otros aspectos relacionados con la tarea (como el tiempo empleado, por ejemplo), otros relativos a variables sociodemográficas (como la edad), otros más ligados a los objetivos del estudio, etc. Esas otras informaciones no afectan a lo que necesitamos saber para analizar la información que específicamente se refiere al modo de ordenar las tarjetas.

4. Simulación de respuesta

Podemos partir de que la participante ha realizado la ordenación atendiendo principalmente a pocas variables (quizá solo una) en primer lugar. Ha priorizado algún valor concreto de alguna variable concreta. Como hay varias repeticiones de ese valor, en segundo lugar ha utilizado otra variable para definir la ordenación dentro de la primera. Y así hasta conseguir el resultado final. En un conjunto de ocho perfiles ortogonales que surgen de 3 a 7 variables

dicotómicas, solo es posible realizar esa labor con 3 variables. El resto tendrán un orden aleatorio o accidental, mostrando una influencia nula en los criterios de la participante.

Para ver un ejemplo, vayamos a la tabla 1 sobre los criterios de publicación. Imagina que la participante considera que lo más importante es no pagar por publicar. Por ello pondrá en primer lugar las cuatro tarjetas que muestran el valor "No" en esa variable (variable 5, código de valor 2). Y en los cuatro últimos lugares pondrá las tarjetas con el valor "Sí" (código 1). El vector quedaría así:

Tarjetas →	3, 4, 5, 7				1, 2, 6, 8			
Paga	2	2	2	2	1	1	1	1

Ahora queda la tarea de ordenar dentro de cada uno de estos dos grupos de 4 tarjetas. Pues bien, en ambos podría priorizar que el idioma sea el español (variable 1, código 2). De ese modo:

Tarjetas →	3, 4		5, 7		2, 8		1, 6	
Paga	2	2	2	2	1	1	1	1
Idioma	2	2	1	1	2	2	1	1

Todavía tiene más juego para tomar decisiones. Imaginemos ahora que escoge el criterio de que el artículo tenga una clara utilidad (variable 7, código 1). De este modo:

Tarjetas →	3	4	5	7	8	2	6	1
Paga	2	2	2	2	1	1	1	1
Idioma	2	2	1	1	2	2	1	1
Utilidad	1	2	1	2	1	2	1	2

No siempre se puede hacer esto. Ordenar no suele consistir en algo tan sencillo como fijarse en una variable, después en otra (hasta aquí, siempre es posible) y después en otra, y otra. Ya desde la tercera podemos encontrarnos con valores *esclavos*, es decir situaciones donde no se puede escoger el orden deseado para la tercera variable porque coincide en las combinaciones de las otras dos. Eso es lo que tiene utilizar combinaciones ortogonales. Algún inconveniente tendría que haber en utilizar 8 tarjetas en lugar de $2^7 = 128$. Cuando la participante no puede ordenar según una variable porque está condicionada por el orden heredado de las previas, hay que aceptarlo o modificar la prioridad de las variables. En esto consiste la complejidad de la tarea. Hay que optar. En la práctica, no suelo observar que las personas hacen lo que acabo de describir por pasos, sino que leen el contenido completo de la tarjeta y van modificando los órdenes. Lo que siempre ocurre es que va a ser difícil encontrar la combinación perfecta y siempre hay que optar priorizando unas variables sobre otras. Así que, cuando la participante comprueba que controlar la tarea a la perfección es inviable, es mucho más fácil que no desee controlarla en absoluto y se deje llevar.

En el ejemplo que estamos siguiendo, una vez se han ordenado las tarjetas según los valores de las variables 1 (no pagar por publicar), 5 (en español) y 7 (con clara utilidad), las otras cinco variables están condenadas, sus valores son dependiente de la combinación entre las tres previas. El resultado entonces es:

Tarjetas →	3	4	5	7	8	2	6	1
Paga	2	2	2	2	1	1	1	1
Idioma	2	2	1	1	2	2	1	1
Utilidad	1	2	1	2	1	2	1	2
Indexada	1	2	2	1	2	1	1	2
Escrita por	1	2	1	2	2	1	2	1
Firma	2	1	1	2	2	1	1	2
Se lee	1	1	2	2	2	2	1	1

5. Análisis de datos

El análisis de la ordenación suele ser muy complejo, en el sentido de que se basa en técnicas de análisis estadístico que, a su vez, se basan en otras. Voy a utilizar otra perspectiva, cuya lógica puedes comprender paso a paso. Para comprenderla, utilizaremos el ejemplo simulado anterior.

El primer principio lógico es que si la participante ha priorizado un valor de una variable sobre todo lo demás, la primera mitad de tarjetas presentará ese valor. Si hay otras tarjetas y las variables son dicotómicas, preferir una variable sobre el resto puede generar uno de estos dos resultados:

1, 1, 1, 1, 2, 2, 2, 2
2, 2, 2, 2, 1, 1, 1, 1

La operación que hagamos implica otorgar la máxima puntuación a esa variable, pues ha sido la utilizada como criterio, con total claridad. La mínima puntuación debe ser obtenida por cualquier otra ordenación donde ningún valor sea sobresaliente. Con ocho tarjetas que surgen de variables dicotómicas, las ordenaciones que muestran indiferencia por la variable (pues los órdenes de sus valores son equivalentes) son estas seis:

1, 2, 2, 1, 1, 2, 2, 1
2, 1, 1, 2, 1, 2, 2, 1,
1, 2, 2, 1, 2, 1, 1, 2,
2, 1, 1, 2, 2, 1, 1, 2,
1, 1, 2, 2, 2, 2, 1, 1
2, 2, 1, 1, 1, 1, 2, 2

¿Cómo lo conseguimos? La estrategia va a consistir en sumar las posiciones que ocupan los valores. Dado que una posición más ventajosa se simboliza con un número más bajo, conforme más baja sea la suma, más importante ha sido ese valor. Vamos a calcularlo con la ordenación utilizada en el ejemplo. El orden de las tarjetas fue:

Orden →	1	2	3	4	5	6	7	8
Tarjetas →	3	4	5	7	8	2	6	1

Lo primero que hay que hacer es acompañar cada tarjeta por los valores que la definen. Eso ya está hecho en la simulación. Vamos a reproducirlo aquí.

Orden →	1	2	3	4	5	6	7	8
Tarjetas →	3	4	5	7	8	2	6	1
Paga	2	2	2	2	1	1	1	1
Idioma	2	2	1	1	2	2	1	1
Utilidad	1	2	1	2	1	2	1	2
Indexada	1	2	2	1	2	1	1	2
Escrita por	1	2	1	2	2	1	2	1
Firma	2	1	1	2	2	1	1	2
Se lee	1	1	2	2	2	2	1	1

Lo siguiente es conseguir un peso para cada uno de los dos valores de cada variable, en función de su posición. Eso se obtiene sumando sus posiciones. Vamos a añadir entonces dos columnas a la tabla. Una para la suma de posiciones del valor 1 y otra respecto al valor 2.

Orden →	1	2	3	4	5	6	7	8	Valores		
Tarjetas →	3	4	5	7	8	2	6	1	1	2	Σ
Paga	2	2	2	2	1	1	1	1	26	10	36
Idioma	2	2	1	1	2	2	1	1	22	14	36
Utilidad	1	2	1	2	1	2	1	2	16	20	36
Indexada	1	2	2	1	2	1	1	2	18	18	36
Escrita por	1	2	1	2	2	1	2	1	18	18	36
Firma	2	1	1	2	2	1	1	2	18	18	36
Se lee	1	1	2	2	2	2	1	1	18	18	36

Bien, vamos a echarle un tiempo a esta tabla.

En primer lugar, observa que el total de las posiciones de los valores 1 y 2 siempre es 36. No es una casualidad, es la consecuencia de sumar los 8 primeros números naturales:

$$\sum_1^n i = \frac{n(n+1)}{2} \quad \sum_1^8 i = \frac{8 \cdot 9}{2} = 36$$

La mayor descompensación entre la suma de las posiciones de los valores ocurre cuando van correlativos, como ocurre con la variable *Paga*. En ese caso, uno de los valores ocupa las $n/2$ posiciones iniciales, por lo que suma:

$$\sum_1^{n/2} i = \frac{n/2 (n/2+1)}{2} \quad \sum_1^8 i = \frac{4 \cdot 5}{2} = 10$$

Mientras que el otro sumará $36 - 10 = 26$. Como ves, eso es lo que ocurre con la variable *Paga*. Las variables *Indexada*, *Escrita por*, *Firma* y *Se lee* muestran total indiferencia: la suma de las posiciones de ambos valores coinciden. Por último, *Idioma* y *Utilidad* cuentan con valores situados entre ambos extremos.

¿Qué miden esas sumas de posiciones? Dado que a mayor puntuación, menos preferencia se posee por ese valor, las puntuaciones miden importancia inversa. Eso lo corregiremos en breve. Antes vamos a dar un paso más en un recurso para conseguir la unicidad de interpretación: calcular el valor de cada categoría en función de su distancia al valor esperado si fueran categorías irrelevantes, es decir, a la mitad de la suma de posiciones, 18. Vamos a llamar g (de *graveco*, importancia en Esperanto) a esta versión del índice.

$$g_i = S_i - \frac{1}{2} \sum_1^n i = S_i - \frac{n(n+1)}{4} = S_i - 18$$

Añadiendo esto a la tabla:

	S_1	S_2	g_1	g_2
Paga	26	10	8	-8
Idioma	22	14	4	-4
Utilidad	16	20	-2	2
Indexada	18	18	0	0
Escrita por	18	18	0	0
Firma	18	18	0	0
Se lee	18	18	0	0

La importancia inversa de cada categoría está repetida, con signo contrario, porque se trata de variables dicotómicas. De hecho, no nos haría falta contar con dos columnas. Una de ellas, con ese mismo resultado en valores absolutos sería suficiente. Y el resultado ya puede ser interpretado como la importancia de la variable para entender la ordenación de las fichas.

Para facilitar la interpretación y la comparación con otras situaciones, una buena estrategia es transformar g en porcentaje. Y para completar la tabla solo nos falta incluir información sobre cuál ha sido la categoría (k , de *kategorio* en Esperanto) preferida.

Observa la tabla resultante. Tenemos información sobre la importancia (g) de cada variable y la categoría más relevante (k). Contando con esas dos columnas para todos los casos, es decir para todas las participantes, podemos interpretar qué ha hecho cada una de ellas: las variables que ha utilizado para ordenar, considerando preferencias concretas por sus valores.

	S_1	S_2	g	$g(\%)$	k
Paga	26	10	8	57	2
Idioma	22	14	4	29	1
Utilidad	16	20	2	14	2
Indexada	18	18	0	0	0
Escrita por	18	18	0	0	0
Firma	18	18	0	0	0
Se lee	18	18	0	0	0
		Σ	14	100	

Un ejemplo de análisis de datos de tarjetas

Variable	Valor 1	Valor 2
Ámbito	Ámbito internacional	Ámbito sevillano
Creatividad	Innovadora	Tradicional
Riqueza	Ganas dinero	No te da beneficios
Efectos en la gente	Es admirada	Deja indiferente
Efectos en empleo	Crea empleo de calidad	Crea empleo inestable

Tarjeta	Ámbito	Creatividad	Riqueza	E. gente	E. empleo
0	1	2	1	2	2
1	2	2	2	2	2
2	2	1	1	1	2
3	1	2	1	1	1
4	1	1	2	1	2
5	2	2	2	1	1
6	1	1	2	2	1
7	2	1	1	2	1

Participante 1: 7,2,4,3,0,1,5,6

Orden→	1	2	3	4	5	6	7	8	Valores					
Tarjetas→	7	2	4	3	0	1	5	6	1	2	Σ	g	g%	k
Ámbito	2	2	1	1	1	2	2	1	36					
Creatividad	1	1	1	2	2	2	2	1	36					
Riqueza	1	1	2	1	1	2	2	2	36					
E. gente	2	1	1	1	2	2	1	2	36					
E. empleo	1	2	2	1	2	2	1	1	36					

Participante 2: 7,0,6,3,2,5,4,1

Orden→	1	2	3	4	5	6	7	8	Valores					
Tarjetas→									1	2	Σ	g	g%	k
Ámbito									36					
Creatividad									36					
Riqueza									36					
E. gente									36					
E. empleo									36					