

Métodos no lineales de escalado óptimo: una aplicación al análisis del empleo en la compañía ferroviaria MZA

Fabiola Portillo[†], Cecilio Mar[‡] y Tomás Martínez^{*}

[†]*Universidad de La Rioja*, [‡]*University of Kent, UK*,

^{*}*Universidad Complutense de Madrid*

March 15, 2007

Abstract

En este estudio se aplican los métodos no lineales de escalado óptimo con el fin de analizar la información contenida en una amplia base de datos sobre los trabajadores de la empresa ferroviaria MZA, una de las mayores concesionarias de este sector en España durante la segunda mitad del siglo XIX y primera del XX. Dado que esta base de datos incluye tanto variables cualitativas como cuantitativas, el procedimiento seguido consiste en la transformación de los valores cualitativos en cuantitativos mediante los métodos de escalado óptimo. Una vez cuantificadas las variables cualitativas, es posible aplicar las técnicas estándar del análisis multivariante. En este trabajo, se ha utilizado el análisis de Componentes Principales, ya que permite la visualización de los resultados, facilitando su comprensión a los no especialistas. Los resultados obtenidos muestran que los trabajadores de la compañía de nacionalidad española tendían a ser empleados en trabajos menos cualificados y, consecuentemente, remunerados con menores salarios que los trabajadores de otros países.

1 Introducción

En las últimas décadas, numerosos trabajos han tratado de evaluar la importancia de las contribuciones de R.A. Fisher a la estadística (Box, 1978;

Efron, 1998; Hald, 1998; Welsh y Robinson, 2004). Un rasgo característico de este autor generalmente destacado en dichos trabajos es la importancia que atribuye a la inferencia así como sus contribuciones en este ámbito, con novedosas propuestas de métodos prácticos para abordarla. En esta amplia literatura, no obstante, los métodos no lineales de Escalado Óptimo (EO) —o, en términos de Fisher, «*the appropriate scoring technique*»— han sido uno de los aspectos menos estudiados y, asimismo, las aplicaciones en economía a que han dado lugar han sido menos numerosas que las generadas por otros procedimientos de inferencia desarrollados por Fisher (Gower, 1990). A este respecto, Savage (1976) apunta como principal motivo de la situación descrita, el elevado grado de dificultad que presenta la comprensión los métodos de EO tal y como los formula el autor, insertos en un complicado contexto de análisis multivariante de datos.

La idea de asignar valores cuantitativos a conceptos cualitativos mediante los métodos no lineales de EO fue inicialmente planteada por Fisher (1938) —en la séptima edición de *Statistical Methods for Research Workers*—, con el fin de posibilitar la aplicación de las técnicas estadísticas estándar del análisis multivariante a datos cualitativos. Tras esta aportación inicial, los trabajos de Hayashi (1952), Kruskal (1964), De Leeuw *et al.* (1976), Takane *et al.* (1980), De Leeuw (2004), Meulman *et al.* (2004) y Neal y Roberts (2006), principalmente, han dotado a esta propuesta metodológica de las herramientas necesarias para su aplicación sistemática a numerosos ámbitos de la investigación científica, siendo de especial interés, por su relevancia para el desarrollo de este enfoque, el método de estimación de Mínimos Cuadrados Alternantes (MCA) (Young, 1972).

En este artículo se aplican los métodos de EO para analizar la información contenida en una amplia base de datos sobre los trabajadores de la empresa ferroviaria «Compañía de los Caminos de Hierro de Madrid a Zaragoza y Alicante» (MZA), una de las mayores concesionarias de este sector en España durante la segunda mitad del siglo XIX y primera del XX. La compañía MZA se estableció en Madrid en 1856, tras la aprobación de la Ley General de Ferrocarriles de 1855, y continuó desarrollando sus actividades hasta 1941, año en el que fue nacionalizada y pasó a formar parte de la compañía ferroviaria nacional española, RENFE. A comienzos del siglo XX, MZA controlaba más de un tercio del tendido ferroviario nacional y, junto con la «Compañía de los Caminos de Hierro del Norte de España», ambas representaban aproximadamente el 75% del empleo total del sector.

Para cada uno de los empleados de la compañía MZA, esta base de datos

contiene información sobre la edad en que fueron contratados, el número de años que permanecieron en la empresa y el salario inicial, todas ellas variables de tipo cuantitativo. Además incluye otras características de los empleados, como son, estado civil, tipo de trabajo desarrollado, motivos por los que cesó su relación con la compañía y lugar de nacimiento, todas ellas variables cualitativas. Así pues, para explotar la información que contienen estos datos de naturaleza multivariante, es necesario trabajar conjuntamente con variables cualitativas y cuantitativas, y por tanto medidas en distintos tipos de escalas, lo que condiciona los métodos estadísticos aplicables (Stevens, 1951). El procedimiento seguido consiste en la transformación de los valores cualitativos en cuantitativos mediante los métodos de EO. Una vez cuantificadas las variables cualitativas, es posible aplicar las técnicas estándar del análisis multivariante del modo habitual. En este trabajo se ha utilizado el análisis de Componentes Principales (CP), ya que permite la visualización de los resultados y, de este modo, resultan más fácilmente comprensibles.

Los resultados obtenidos mediante la aplicación del procedimiento descrito muestran que los trabajadores de la compañía de nacionalidad no española tendían a ser empleados en trabajos más cualificados y, consecuentemente, remunerados con mayores salarios que los trabajadores españoles. También es posible observar que la mayoría de los trabajadores que interrumpieron su relación con la compañía por motivos extraordinarios — dimisión o despido disciplinario — lo hicieron poco tiempo después de entrar a trabajar en la empresa, no observándose relación entre este hecho y el salario inicial, ni tampoco con la categoría profesional en el momento de ser contratados.

El resto del trabajo se ha estructurado del siguiente modo. En primer lugar, se describen los métodos estadísticos que han servido de base para la investigación, así como los datos empleados en su aplicación. A continuación se muestran los resultados obtenidos, para pasar seguidamente a su análisis y discusión. Finalmente, se resumen las principales conclusiones del estudio.

2 Escalado Óptimo y Mínimos Cuadrados Alternantes

En las últimas décadas, diversos autores han reinterpretado los métodos no lineales de EO introducidos por Fisher (1938), presentando sus ideas en un

marco más moderno y unificado, como es el que proporciona el Análisis de Correspondencias Múltiples (Gower, 1990; Welsh y Robinson, 2004). Este marco de análisis ha permitido, por un lado, la profundización en los fundamentos metodológicos del EO, estableciendo sus relaciones con los Métodos de Medias Recíprocas propuestos por Fisher y Mackenzie (1923) para modelos multiplicativos, así como con el Análisis de Correspondencias de Hirschfeld (1935) y el Análisis de Correlación Canónica de Hotelling (1936); por otro lado, ha posibilitado su evolución hasta alcanzar el grado de desarrollo necesario para su aplicación sistemática en numerosos ámbitos de la investigación científica.

El principio de los métodos no lineales de EO implica la consideración de las observaciones como categóricas —bien porque estas posean dicha naturaleza o bien porque sean contempladas como el resultado de un proceso de medida finito—, representando las distintas categorías mediante parámetros, cuya optimización conduce a los denominados parámetros de escalado óptimo. Dichos parámetros son el resultado de aplicar una serie de transformaciones a las observaciones y están sujetos a las restricciones implícitas en las características de medida de la variable —por ejemplo, restricciones de orden en variables ordinales— además de las que requiere su normalización para la eliminación de ambigüedades. Así pues, desde este punto de vista, los datos consisten en observaciones que se sitúan en una más o menos amplia variedad de categorías, considerándose empíricamente equivalentes todas las observaciones incluidas en una misma categoría.

La estimación de los parámetros basada en el principio de MCA parte de su clasificación en dos tipos o grupos exhaustivos y mutuamente excluyentes: por un lado, los parámetros del modelo y, por otro, los parámetros de los datos o parámetros de escalado óptimo. Para abordar su estimación es necesario definir una función de pérdida y optimizar alternativamente con respecto a uno y otro tipo de parámetros, en un proceso iterativo que continua hasta alcanzar el óptimo —para lo cual existen actualmente algoritmos que presentan escasos problemas de mínimos locales—. El criterio de optimización, o función de pérdida, puede consistir en la suma del cuadrado de los residuos —lo que conduce al análisis de la varianza— o en otras medidas de disparidad entre las observaciones y los parámetros estimados, siendo la más empleada el índice de stress de Kruskal (1964).

Así pues, la optimización se lleva a cabo en un proceso iterativo que se inicia con la obtención de las estimaciones mínimo cuadráticas de los parámetros de un tipo, asumiendo que los restantes permanecen constantes, lo que cons-

tituye un método de estimación mínimo-cuadrático condicional. El principio mínimo-cuadrático está condicionado por los valores de los parámetros de los otros grupos, que sucesivamente se van alternando como resultado de las distintas iteraciones desarrolladas hasta alcanzar la transformación óptima de las observaciones.

Formalmente, los métodos de EO se definen como un problema de transformación, t , del vector de n observaciones, o , en parámetros óptimamente escalados, z^* , esto es,

$$t[o] = [z^*] \quad (1)$$

donde cada uno de los elementos de este último, z_i^* , representan de forma cuantitativa las correspondientes observaciones o_i . Los tipos de transformaciones que, en cada caso, han de aplicarse a los datos para su cuantificación dependen básicamente de la naturaleza de las variables —bien sean continuas o discretas, nominales u ordinales—, y se basan en la estimación mínimo-cuadrática del modelo, sujeta al estricto cumplimiento de las restricciones de medida que presentan las observaciones (ver Young, 1981). En este sentido, el valor numérico asignado a z_i^* es el valor óptimo del parámetro para la observación o_i . La función de pérdida a minimizar más comúnmente empleada para obtener dichos valores óptimos es el índice de stress de Kruskal (1964), Φ ,

$$\Phi = \frac{\|\hat{z} - z^*\|}{\|z^*\|} \quad (2)$$

que constituye un índice de la suma del cuadrado de los residuos normalizados de la estimación (ver De Leeuw, 1984; De Leeuw *et al.*, 1984). En la expresión (2), siguiendo la notación habitual, $\|z^*\|$ denota la norma euclidiana del vector z^* .

Un aspecto importante en la aplicación de este método lo constituye el establecimiento de las restricciones de medida que definen la estructura de los datos, y que se agrupan en tres tipos, como son, las restricciones de proceso, las restricciones de nivel y las restricciones de condicionalidad. Dentro del primer tipo, se encuentran las restricciones de medida correspondientes a procesos de naturaleza discreta o continua. En cuanto a las restricciones de nivel, básicamente permiten diferenciar entre variables nominales, ordinales y numéricas, bien sean de intervalo o de razón. Por último, las restricciones de condicionalidad tienen que ver con las relaciones existentes entre las ob-

servaciones dentro de las distintas categorías en que se agrupan los datos, y que pueden estar limitadas o condicionadas por algún aspecto inherente a la medición de los datos. Cuando este sea el caso, las limitaciones existentes deberán condicionar la comparabilidad de las observaciones. En este ámbito, son comunes las restricciones de tipo matriz condicionada —por ejemplo, en estudios de opinión, cuando no puede confiarse en que las escalas de respuesta sean utilizadas idénticamente por los distintos individuos de la muestra— así como las de tipo columna condicionada, característico del análisis multivariante, donde las columnas contienen variables de distinta naturaleza —tales como lugar de nacimiento, tipo de trabajo desarrollado en la empresa y salario inicial percibido por el trabajador—.

Actualmente, esta metodología se ha generalizado de manera que contempla la transformación de cualquier tipo de variables cualitativas en otras de naturaleza cuantitativa y, asimismo, su aplicación a una amplia variedad de técnicas de análisis multivariante de datos —como son, la regresión múltiple, el análisis de correlaciones canónicas, el análisis discriminante y el análisis de componentes principales— o al análisis de correspondencias múltiples. Entre las aplicaciones del escalado óptimo al ámbito de la economía cabe destacar Didow *et al.* (1985), Serrano-Cinca *et al.* (2003), Serrano-Cinca *et al.* (2004), Portillo *et al.* (2006).

3 Datos

Los datos empleados proceden de los archivos de la compañía MZA, de los que se ha extraído información sobre los 940 trabajadores que fueron contratados durante el periodo 1882 a 1885, en plena expansión de la compañía. En 1878 MZA había adquirido la empresa ferroviaria «Madrid-Badajoz» y, posteriormente, en 1898, adquiriría la compañía «Tarragona-Barcelona-Francia». De esta muestra de trabajadores, únicamente 533 han servido de base al estudio por no disponer para el resto de una ficha personal suficientemente completa. Una vez depurados los datos, el número total de trabajadores se redujo a 527.

Junto con el nombre y apellidos, la base de datos contiene información sobre la fecha y lugar de nacimiento, empleo para el que fue contratado el trabajador en el momento de incorporarse a la empresa, salario inicial, estado civil, fechas de comienzo y de extinción del contrato de trabajador con la empresa y sus causas —incluidas jubilación, muerte y enfermedad—. Esta información se resume en las ocho variables siguientes: edad de incorpo-

ración del trabajador en la compañía, número de años de permanencia en la misma, salario inicial, lugar de nacimiento, estado civil, primer trabajo desempeñado, sección en la que desarrolla su actividad y motivos para la extinción del contrato con la empresa. Mientras que las tres primeras son variables cuantitativas, las cinco últimas tienen naturaleza cualitativa. En las *tablas 1 y 2* se resumen los estadísticos descriptivos y las distribuciones de frecuencias de las variables cuantitativas y cualitativas analizadas, respectivamente.

Variable	Media		Mediana	Mínimo	Máximo
	Estadístico	Error típico			
Edad de incorporación (años)	27,91	0,36	26	13	53
Experiencia en la compañía (años)	9,30	0,53	3	0	53
Sueldo inicial (reales)	2.747,00	37,65	2.500	500	10.000

TABLA 1
Estadísticos descriptivos de las variables cuantitativas

Dada la procedencia de los trabajadores de la compañía MZA, la variable lugar de nacimiento se ha clasificado en cuatro categorías: nacidos en Madrid, nacidos en el resto de España, nacidos en Francia y otros. Principalmente son trabajadores de nacionalidad española y la gran mayoría de ellos procedentes de Madrid. Diferenciando Madrid del resto de España es posible además contrastar si estos últimos ocupaban diferentes empleos a los del resto de españoles, tal vez como consecuencia de la localización efectiva de la sede central de la compañía en esta localidad. MZA fue fundada con capital francés y 16 trabajadores en la muestra tenían esta nacionalidad, el mayor contingente de empleados de la compañía no españoles. Este hecho podría obedecer a una mayor cualificación de estos trabajadores, traídos a España por la compañía para aplicar sus especiales conocimientos en este ámbito, hipótesis que será posteriormente comprobada en el análisis. “Nacidos en otros lugares” abarca al resto de trabajadores procedentes de una amplia variedad de países, desde Cuba hasta Gran Bretaña.

Variable	Categoría	Distribución	
		Frecuencia	Porcentaje
Lugar de nacimiento	Madrid	191	36,24
	Resto de España	309	58,63
	Francia	16	3,04
	Otros países	11	2,09
Estado civil	Soltero	270	51,23
	Casado	246	46,68
	Viudo	11	2,09
Primer empleo	Aprendiz, Peón o Ayudante	195	37,00
	Oficial	329	62,43
	Encargado	3	0,57
Sección	Calderería	63	11,95
	Fraguas	31	5,88
	Fundición	30	5,69
	Ajustaje	36	6,83
	Montaje	54	10,25
	Tornos	36	6,83
	Oficinas	3	0,57
	Instalaciones	8	1,52
	Pintura y carpintería	105	19,92
	Guarnecido	21	3,98
	Guardería	3	0,57
Carruajes	137	26,00	
Causas de baja	Dimisión	145	27,51
	Despido disciplinario	174	33,02
	Despido-reducción plantilla	10	1,90
	Jubilación	38	7,21
	Enfermedad	39	7,40
	Fallecimiento	84	15,94
	Traslado	37	7,02

TABLA 2
Distribución de frecuencias de las variables cualitativas

Por su parte, la variable estado civil contiene únicamente tres categorías: soltero, casado y viudo. No se observan datos perdidos en la muestra.

En cuanto a los diferentes tipos de trabajos desarrollados por los empleados en la compañía, cabe destacar que fueron cuidadosamente registrados. La amplia lista de ellos resulta poco operativa por su amplitud y, en algunos casos, pintoresca. Por ello, esta lista detallada de designaciones se ha resumido en tres categorías: aprendices o similares, oficiales u ocupaciones con una cualificación de nivel medio y, por último, encargados o supervisores.

Claramente esta variable es de tipo ordinal, dado que es mejor ser supervisor que ocupar un empleo de nivel medio y, asimismo, es mejor ser un trabajador medianamente cualificado que un aprendiz.

Por otra parte, es claro que no puede considerarse totalmente equivalente el empleo de un carpintero cualificado que el de un hábil mecánico, y estas diferencias se han tratado de captar mediante una variable que contiene las diversas secciones en las cuales la plantilla está dividida. Los datos especifican las doce secciones siguientes: ajustaje, calderería, carruajes, fraguas, fundición, guardería, guarnecido, instalaciones, montaje, oficinas, pintura y carpintería, y tornos.

Los motivos por los cuales se declara extinguida la relación de los trabajadores con la compañía constituyen un catálogo de las relaciones sociales entre ambos, que incluyen observaciones tales como “estar bebido y provocar desórdenes” o “declararse enfermo mientras está trabajando en otra empresa”. Todas ellas han sido clasificadas en las categorías: muerte, jubilación, enfermedad, dimisión, despido disciplinario, despido por reducción de plantilla y traslado. A este respecto, los datos no indican porqué el traslado tiene como consecuencia el cese del trabajador en la compañía. Tal vez este hecho implique que el trabajador pase a formar parte de la plantilla de otra empresa subsidiaria de MZA.

4 Resultados

En la *tabla 3* se resumen los parámetros de EO para las distintas categorías de las variables cualitativas contenidas en la base de datos MZA. Teniendo en cuenta que no se han introducido restricciones de nivel en el algoritmo, es interesante observar la ordenación que, de manera natural, han recibido las distintas categorías mediante la aplicación de este método.

Una vez cuantificadas las variables cualitativas, el análisis de CP pudo realizarse contemplando simultáneamente las ocho dimensiones asociadas a las ocho variables —cinco cualitativas y tres cuantitativas— que incluye esta base de datos. La *tabla 4* presenta los valores propios de las dimensiones relevantes obtenidos en la aplicación de dicho análisis, siguiendo el criterio de Joliffe (1972). Este criterio establece en 0,8 el límite del valor propio, por encima del cual puede considerarse relevante una dimensión —frente al criterio de Kaiser que considera tan solo las dimensiones con valores propios mayores que 1—.

Variable	Categoría	EO Parámetros
Lugar de nacimiento	Madrid	-0,42
	Resto de España	-0,02
	Francia	0,87
	Otros países	6,61
Estado civil	Soltero	-0,93
	Casado	0,89
	Viudo	2,97
Primer empleo	Aprendiz, Peón o Ayudante	-1,30
	Oficial	0,75
	Encargado	2,18
Sección	Calderería	-0,76
	Fraguas	-1,30
	Fundición	0,66
	Ajustaje	0,01
	Montaje	-1,42
	Tornos	-0,42
	Oficinas	-0,24
	Instalaciones	-0,88
	Pintura y carpintería	1,61
	Guarnecido	0,99
	Guardería	1,82
Carruajes	0,08	
Causas de baja	Dimisión	-0,65
	Despido disciplinario	-0,63
	Despido-reducción plantilla	-0,56
	Jubilación	2,52
	Enfermedad	1,18
	Fallecimiento	1,03
	Traslado	-0,51

TABLA 3

Parámetros de Escalado Óptimo (EO) que cuantifican las variables cualitativas

Como puede observarse en la *tabla 4*, cinco valores propios satisfacen la condición de Joliffe (1972) y explican el 90% de la variación en los datos. De estos resultados cabe deducir que un empleado de la compañía MZA podría ser descrito adecuadamente a partir de estas cinco características o dimensiones, que se analizan a continuación. Dado que las dos primeras características explican aproximadamente el 50% de la variabilidad en los datos, dichas características constituyen los dos componentes principales más relevantes.

Dimensión	Varianza explicada	
	Valor propio	% Varianza total
1	2,103	26,29
2	1,883	49,82
3	1,394	67,25
4	1,042	80,28
5	0,812	90,43

TABLA 4

Valores propios y varianza total explicada mediante componentes principales

Para abordar la interpretación de las dimensiones es necesario basarse en el análisis de las saturaciones de las variables, que se resumen en la *tabla 5*. Como ya es conocido, las variables con mayor ponderación en los componentes son las que los explican en mayor medida. En el primer componente, las variables con mayores ponderaciones son el salario inicial, la edad de incorporación del trabajador en la empresa, el estado civil, y el primer empleo. Por tanto, un trabajador casado, de edad avanzada, que se incorporase a la compañía como supervisor alcanzaría un valor elevado en este primer componente. Por el contrario, tendrían valores resucidos los jóvenes, solteros, que se unieron a la empresa como aprendices. Así pues, este componente puede calificarse como “edad y categoría” previa a la incorporación de los trabajadores en la compañía.

Variable	Dimensión				
	1	2	3	4	5
Estado civil	,626	-,010	,647	-,148	-,122
Lugar de nacimiento	,104	-,191	-,099	,858	-,452
Edad de incorporación	,724	,109	,523	-,020	-,175
Primer empleo	,616	-,515	-,457	-,052	,091
Sección	,334	-,432	-,484	-,445	-,457
Salario inicial	,653	-,255	-,141	,289	,583
Causa de baja	,412	,795	-,315	,017	-,067
Experiencia	,299	,828	-,361	,007	-,020

TABLA 5

Tabla de saturaciones de los cinco componentes principales

Por su parte, las variables asociadas al segundo componente son, principalmente, la experiencia y la causa de baja, por lo que representarían la “experiencia en la empresa”. El tercer componente está asociado con el estado civil, la edad de incorporación y la sección de trabajo. Así, los casados que se incorporaron a la empresa a una edad avanzada y trabajaron en las secciones de pintura, tapicería o guardería, tendrían valores elevados en este componente. Por último, el cuarto componente está claramente relacionado con el lugar de nacimiento, y el hecho de que esta variable sea un componente en sí mismo indica, además, que no está correlacionada con las restantes variables ni con los otros componentes. Por ello, puede concluirse que el lugar de nacimiento constituye una variable singular que permite diferenciar distintos grupos de trabajadores en la empresa MZA, cuestión que se analiza en el siguiente epígrafe.

5 Interpretación gráfica

Por un lado, la *figura 1* representa gráficamente a los trabajadores de MZA en las dimensiones que determinan los dos primeros componentes principales, anteriormente analizados. Dado que cada punto representa un trabajador de la compañía, se ha indicado asimismo en la figura el salario y el lugar de nacimiento de cada uno de ellos.

Como puede observarse, el salario aumenta de izquierda a derecha, indicando que la edad y la categoría profesional al incorporarse a la compañía —primera dimensión— están claramente relacionadas con el salario que percibían los trabajadores, algo que era de esperar. También es notorio en la figura que los trabajadores nacidos fuera de España se encuentran muy próximos entre sí, lo que indica que comparten características comunes. De su posición en la figura se deduce que los trabajadores no españoles recibieron salarios superiores a la media.

Por otro lado, la *figura 2* muestra nuevamente los datos de MZA en la dimensión de los dos primeros componentes principales, pero en este caso en relación con la causa de cese de los trabajadores en la compañía y el salario inicial.

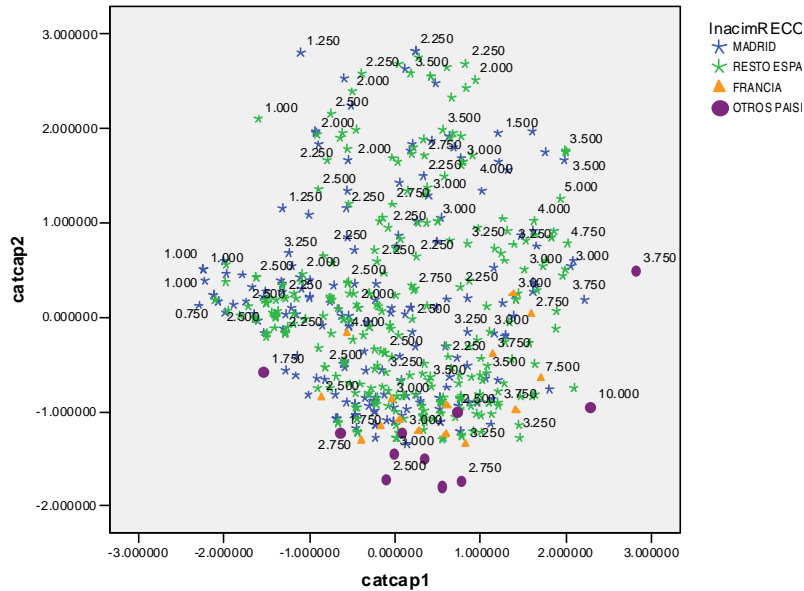


FIGURA 1

Representación gráfica de dos primeros componentes principales, con indicación del salario inicial y el lugar de nacimiento de los trabajadores

Al analizar las causas del cese de los trabajadores se observa que en la parte superior de la figura se sitúan aquellos que dejaron la compañía por causas ordinarias —jubilación, enfermedad, y muerte—, mientras que en la parte inferior se sitúan los que cesaron por razones extraordinarias —reducción de plantilla, despido disciplinario, dimisión y traslado—. Esto sugiere que la conflictividad laboral se concentra en la parte inferior de la figura y, en este sentido, cabe cuestionarse con qué variables estaría asociada tal conflictividad laboral. Para explorar esta cuestión, se aplicará el análisis de Ajuste de Propiedades —o *Property Fitting* (Pro-Fit)—.

Inicialmente, Schiffman *et al.* (1981) desarrollan esta herramienta para variables cuantitativas y, más recientemente, Mar Molinero y Mingers (2007) amplian su ámbito de aplicación, demostrando sus buenas propiedades también para el caso de dicotomías.

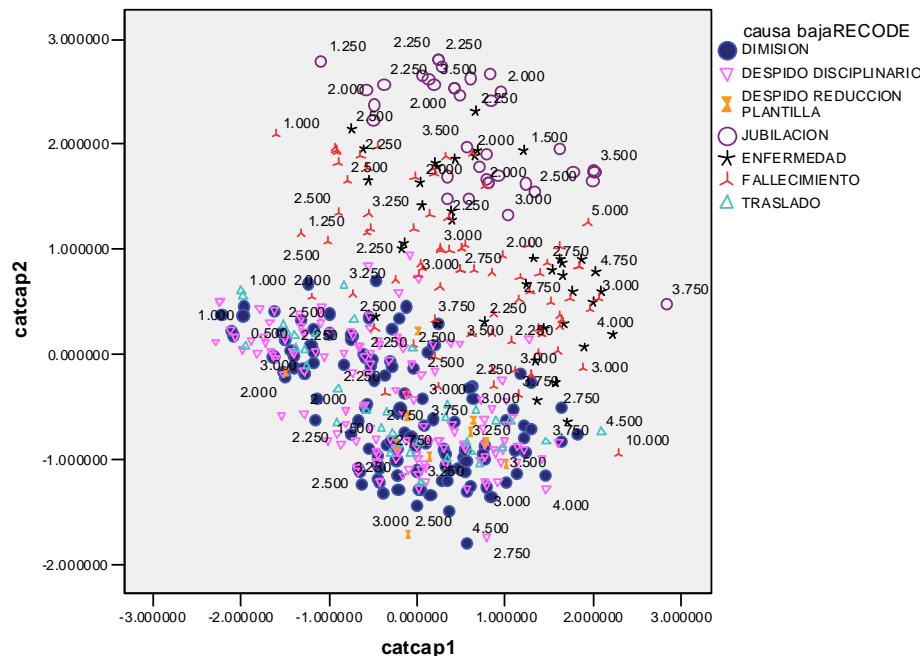


FIGURA 2

Representación gráfica de dos primeros componentes principales, con indicación del salario inicial y la causa del cese de los trabajadores

La aplicación de Pro-Fit requiere estimar una regresión en la que cada individuo es una observación. La variable dependiente es la característica de los datos que nos interesa relacionar con la posición en la figura; un ejemplo sería la edad de los trabajadores cuando se incorporan. Las variables explicativas son los valores de los componentes principales para cada individuo. Cuando la variable dependiente es una dicotomía- por ejemplo, en el caso de cese normal versus cese anormal-, la regresión es logística. Al estar basada en la regresión, todas las medidas de bondad de ajuste nos ayudan a evaluar la calidad de la explicación. Pero la gran ventaja de la técnica Pro-Fit es que los resultados se pueden representar de un modo gráfico por medio de unas líneas orientadas. Es conveniente normalizar los vectores asociados en las líneas Pro-Fit para dejar bien claro con qué características de la representación están asociados.

Las características siguientes han sido tratadas como propiedades: edad

de incorporación, salario inicial, experiencia en la empresa, razón de cese (normal versus anormal), estado civil (soltero/no soltero), lugar de nacimiento (España/no España), y tipo de trabajo (aprendiz/otro). Las últimas cuatro propiedades son dicotomías. Los resultados técnicos se muestran en la *tabla 6* y los vectores asociados con las propiedades se pueden ver en la *figura 3*. Los coeficientes de determinación fueron altos en todos los casos, indicando que es adecuado el interpretar las figuras con la ayuda de Pro-Fit.

Variable	β_1	β_2	β_3	β_4	β_5	β_6	R^2
Edad de incorporación	0.74	0.10	0.53	-0.01	-0.17	0.37	0.93
Salario inicial	0.68	-0.26	-0.11	0.40	0.55	0.06	0.82
Experiencia	0.27	0.87	-0.40	0.01	-0.02	-0.11	0.95
Causa de baja: normal/extraordinaria	0.42	0.84	-0.32	0.02	-0.09	-0.03	0.96 (Nagelkerke)
Esto civil: soltero/no soltero	0.50	0.00	0.77	-0.12	-0.15	-0.35	1.00 (Nagelkerke)
Lugar de nacimiento: España/no España	-0.07	0.19	0.11	-0.85	0.47	0.04	1.00 (Nagelkerke)
Primer empleo: Aprendiz/resto	0.48	-0.68	-0.51	-0.08	0.05	-0.22	0.99 (Nagelkerke)

TABLA 6

Resultados del análisis de ajuste de Propiedades: cosenos direccionales y coeficientes de determinación (en el caso de logit bivalente, el R^2 Nagelkerke)

Los vectores que forman un ángulo agudo en la *figura 3* corresponden a variables que están positivamente correlacionadas. Vemos que las variables EF (Experiencia en la empresa) y RL (motivo de cese) están altamente asociadas, indicando que quienes pasan muchos años en la empresa la dejan de modo “normal”. Dicho de otro modo, a los trabajadores con muchos años de estancia en la empresa no los despiden. El reverso también es cierto. Los trabajadores que dejan la empresa de un modo anormal suelen ser aquellos que han permanecido en ella por poco tiempo.

También hay asociación positiva entre la edad en el momento de la incorporación (JA) y el estado civil (MS), confirmando que los trabajadores que están casados son de más edad que los solteros. Tanto JA como MS tienen asociaciones positivas con el salario inicial (IS). Esto indica que los trabajadores de más edad suelen recibir salarios más altos al incorporarse.

Es interesante ver que el salario inicial y la experiencia en la empresa no están asociados (los dos vectores son ortogonales), como tampoco lo está el salario inicial con el motivo de cese. El nivel de empleo en el primer trabajo está relacionado con el salario inicial, tal como sería de esperar. El nivel de empleo en el primer trabajo está negativamente asociado con el haber nacido en España, lo que se deduce del hecho de que las variables FJ y PB están casi sobre la misma línea pero en direcciones opuestas.

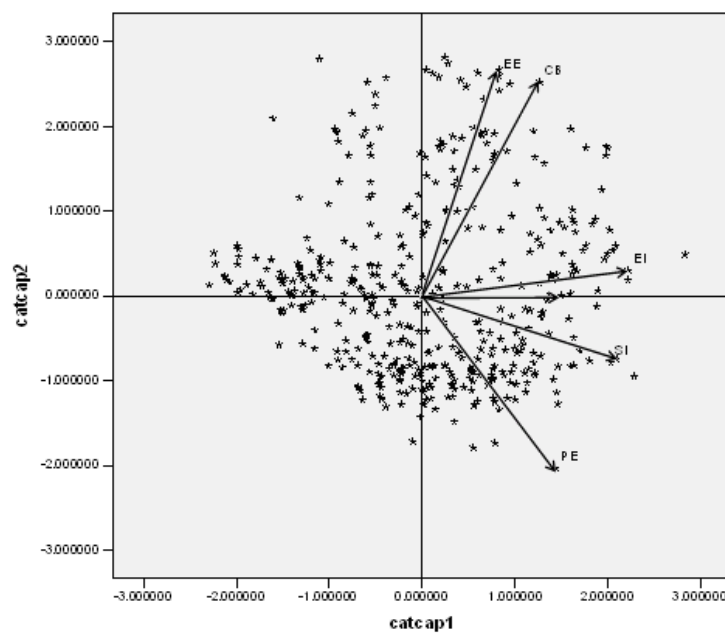


FIGURE 3

Resultados del análisis de Property Fitting. Lugar de nacimiento (LN), experiencia (EE), causa de cese (CB), edad de incorporación (EI), estado civil (MS), salario inicial (SI), primer empleo (PE).

Sería posible continuar la exploración de los datos teniendo en cuenta el tercer componente categórico, pero basta con dar un somero repaso a los cosenos direccionales para darse cuenta de que la tercera dimensión contribuye poco a las explicaciones que hemos dado. La verdadera historia que nos transmiten los datos se halla en la representación en el primer y segundo componentes principales.

6 Conclusiones

La combinación de variables cualitativas y cuantitativas complica el análisis de muchos conjuntos de datos. En este trabajo se ha hecho ver que la técnica del Escalado Óptimo al asignar valores numéricos a categorías cualitativas hace posible la aplicación de los métodos tradicionales del análisis multivariante en casos en que no hubiera sido posible hacerlo.

Nuestros datos contenían información sobre 527 trabajadores de una de las mayores empresas ferroviarias españolas del siglo XIX. Muchos son los estudios de esta empresa desde el punto de vista de la gestión, pero nadie parece haberse preocupado de enfocar los estudios desde el punto de vista de los trabajadores que empleaba.

El análisis de los datos ha revelado varios aspectos de la compañía. Los trabajadores que no habían nacido en España solían recibir salarios más altos que los nacidos en España, aunque hubo españoles que recibieron, al incorporarse a la compañía, salarios tan altos como los que se pagaban a los extranjeros, y fueron contratados en niveles de la misma categoría.

El período que hemos estudiado se caracteriza por su conflictividad social. Esto se refleja en el modo en que dejaron la empresa los trabajadores. El estudio deja bien claro que muchos de los trabajadores que dejaron la empresa de modo anormal (reducción de plantilla, despido, etc), lo hicieron al poco de haberse incorporado. La razón de cese no está relacionada ni con el primer sueldo ni con el nivel de empleo en el momento de la incorporación.

References

- [1] Box, J.F. (1978) R.A. Fisher, the life of a scientist. Wiley, New York.
- [2] De Leeuw, J. (1984) Differentiability of Kruskal's stress at a local minimum. *Psychometrika*, 49(1), 111–112.
- [3] De Leeuw, J. (2004) Least Squares Optimal Scaling of Partially Observed Linear Systems, en K. van Montfort , J. Oud, y A. Satorra, (eds.), *Recent Developments on Structural Equation Models: Theory and Applications*. Kluwer, Norwell, MA.
- [4] De Leeuw, J. y Stoop, I. (1984) Upper bounds for Kruskal's stress. *Psychometrika*, 49(3), 289–437.

- [5] De Leeuw, J., Young, F.W. y Takane, Y. (1976) Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika*, 41(4), 471–503.
- [6] Efron, B. (1998) R.A. Fisher in the 21 st Century. *Statistical Science*, vol. 13, No. 2, 95-122. Comments by: Cox, D.R., R. Kass, O.E. Barndorff-Nielsen, D.V. Hinkley, D.A. S. Fraser and P. Dempster.
- [7] Fisher, R.A (1938, 1941) *Statistical methods for research workers* (ediciones 7^a y 8^a). Oliver and Boyd, Edinburgh, UK.
- [8] Fisher, R.A y Mackenzie, W.A. (1923) Studies sin crop variation: II. The manurial response of different potato varieties. *Journal of Agricultural Science*, 13, 311–320.
- [9] Gower, J.C. (1990) Fisher’s Optimal Scores and Multiple Correspondence Analysis. *Biometrics*, 46, 947–961.
- [10] Hald, A. (1998) *A. History of Mathematical Statistics from 1750 to 1930*. Wiley, New York.
- [11] Hayashi, C. (1952) On the predictions of the phenomena from qualitative data and quantifications of qualitative data from the mathematico–statistical point of view. *Annals of the Institute of Statistical Mathematics*, 3, 69–92.
- [12] Joliffe. I.T. (1972) Discarding variables in Principal Components Analysis. *Applied Statistics*, 21, 160–173.
- [13] Kruskal, J.B. (1964) Multidimensional Scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, 29(1), 1–27.
- [14] Kruskal, J.B. y Wish, M. (1978) *Multidimensional Scaling*. Sage. London. UK.
- [15] Mar Molinero, C.; y Mingers, J. (2006) Mapping MBA Programmes: an alternative analysis. Forthcoming in the *Journal of the Operational Research Society*.
- [16] Meulman J.J., van der Kooij A.J. y Heiser W.J. (2004) Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data, en D. Kaplan (ed.) *Handbook of quantitative methodology for the social sciences*, 49–70. Sage Publications, Inc, Thousand Oaks, CA.

- [17] Neal, P.J. y Roberts, G.O. (2006) Optimal Scaling for partially updating MCMC algorithms. *The Annals of Applied probability*, 16(2), 475–515.
- [18] Savage, L.J. (1976) On reading R.A. Fisher. *The Annals of Statistics*, 4, 441–500.
- [19] Schiffman, S.S., Reynolds, M.L. y Young, F.W. (1981) *Introduction to Multidimensional Scaling: Theory, Methods and Applications*. Academic Press, London
- [20] Serrano, C.; Mar Molinero, C.; y Chaparro, F. (2004) Spanish savings banks: a view on intangibles. *Knowledge Management Research & Practice*, 2, 103–117.
- [21] Stevens, S.S. (1951) Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley.
- [22] Welsh, A.H.; y Robinson, J. (2005) Fisher and inference for scores. *International Statistical Review*, 73, 131–150.
- [23] Takane, Y., Young, F.W. y De Leeuw, J. (1980) An individual differences additive model: An alternating least squares method with optimal scaling features. *Psychometrika*, 42, 7– 67.
- [24] Young, F.W. (1972) A model for polynomial conjoint analysis algorithms. In R.N. Shepard, A.K. Romney, and S. Nerlove (eds.), *Multidimensional scaling: Theory and applications in the behavior–sciences*. New York: Academic Press.
- [25] Young, F.W. (1981) Quantitative analysis of qualitative data. *Psychometrika*, 46(4), 357–388.