

TÉCNICAS DE ESCALADOS MULTIDIMENSIONALES APLICADAS AL FRACASO DEL ALUMNO EN LA ASIGNATURA DE MATEMÁTICAS EMPRESARIALES.

Autores: Camacho Peñalosa, M.E.; Fernández Geniz, P., Gómez Domínguez, D.; Masero Moreno, I.; Vázquez Cueto, M.J. y Zapata Reina, A.

Facultad de Ciencias Económicas y Empresariales
Departamento de Economía Aplicada III
Avda. Ramón y Cajal, nº 1
41018- SEVILLA
Tfno: 954557553-954551673- 954557556
Fax: 954551667
e-mail: pepi@us.es,
imasero@us.es

Resumen

El propósito de este trabajo es realizar un análisis de las características personales y de formación previa que pueden influir en el fracaso del alumno en la asignatura de Matemáticas Empresariales. Para la realización del mismo, y dado que disponemos de un número importante de variables, trataremos de sintetizar esta información mediante la construcción de un mapa que sintetice la información, utilizando la técnica de los Escalados Multidimensionales.

Palabras clave: Escalados multidimensionales, Análisis Profit, Matemáticas Empresariales, Regresión.

Abstract

In this paper, we try to measure the effect of some variables about personal data and previous learning of students, in the probability to pass the first level of mathematics for Administration and Business. The multidimensional scaling approach is used in this paper to reduce the dimensionality of the observations.

Keywords: Multidimensional Scaling, Profit, Mathematics, Regression .

1.- INTRODUCCIÓN.

Desde hace varios años venimos interesándonos por los posibles condicionantes de las deficiencias en conocimientos matemáticos observadas en los alumnos que se incorporan en los últimos años a los estudios universitarios de tipo económico-empresarial¹. Esto se traduce, en líneas generales, en peores resultados en las calificaciones obtenidas en las pruebas de evaluación² y en el aumento de los no presentados a dichas pruebas.

Una de las hipótesis de partida ha sido que el bachillerato de procedencia condiciona los conocimientos previos del alumno y, por tanto, el aprendizaje de las asignaturas de matemáticas de la Licenciatura de Administración y Dirección de Empresas³, sus calificaciones en dichas asignaturas y su decisión de abandonar las mismas, sobre todo el primer año en el que se cursan.

Sin embargo, este no es el único aspecto que puede influir sino que también hay que considerar ciertos cambios que afectan al estudiante cuando se incorpora a la vida universitaria, por ejemplo, existe un menor control por parte de la institución educativa⁴ al que también se une, en algunos casos, un menor control por parte del entorno familiar, que suele ser aún mayor cuando el alumno abandona la residencia familiar al iniciar sus estudios universitarios.

Otro aspecto a considerar es que algunos alumnos recurren a algún tipo de apoyo extra para la preparación de las asignaturas, ya sea acudiendo a algún centro no universitario “academia” que las imparte, o recurriendo a un profesor particular.

Todo lo anterior ha sido tenido en cuenta en la realización del estudio que presentamos, en el que tratamos de analizar hasta qué punto cada uno de estos aspectos influye en los resultados obtenidos por los alumnos en las disciplinas de matemáticas. Para ello, hemos estructurado este trabajo en los siguientes apartados. En el apartado dos abordamos la descripción de las variables consideradas, en el apartado tres se recoge la metodología que se va a utilizar, en el cuarto se exponen los resultados obtenidos y se cierra el mismo exponiendo algunas reflexiones finales.

2.- VARIABLES A CONSIDERAR Y FUENTE DE DATOS.

Teniendo en cuenta las consideraciones recogidas en el apartado introductorio, definimos las siguientes cinco variables.

VARIABLE 1: Bachillerato cursado. Se han considerado cinco modalidades a las que se les ha asignado los siguientes valores. Bachillerato de Arte (1), Bachillerato en Ciencias de la Naturaleza (2), Bachillerato en Humanidades y Ciencias Sociales (3), Bachillerato en Tecnología (4) y otros (5).

Aunque lo más común es que nuestros alumnos procedan de la opción de Bachillerato en Ciencias Sociales y Humanidades, existen alumnos que proceden de otras opciones. Además, debemos tener en cuenta que no sólo va a influir el bachillerato cursado sino también el itinerario curricular elegido por el alumno, ya que sus conocimientos

dependerán de las asignaturas de matemáticas que haya cursado. Introducimos, por tanto, una segunda variable que tendrá en cuenta este condicionante.

VARIABLE 2: Asignaturas de matemáticas cursadas en los dos años de bachillerato asignando los siguientes valores. Ninguna (0), sólo han cursado la de primer curso (1), sólo han cursado la de segundo curso (2), han cursado matemáticas los dos años de bachillerato (3).

Hemos de considerar que algunos de los cambios que se producen en el alumno o su entorno, cuando inicia los estudios universitarios, no son fáciles de recoger en un estudio de este tipo. Nosotros sólo vamos a incidir en la mayor autonomía o “menor control” de los padres, de aquellos alumnos que abandonan la residencia familiar al iniciar sus estudios universitarios. Tenemos en cuenta, por tanto, una tercera variable.

VARIABLE 3: El alumno vive en el domicilio familiar (1), no vive en el domicilio familiar (0).

La cuarta variable que tendremos en cuenta considera si el alumno dispone o no de apoyo extra en la preparación de las asignaturas de matemáticas.

VARIABLE 4: El alumno recurre a una “academia” o profesor particular para la preparación de las matemáticas (1), no cuenta con este apoyo (0).

Consideramos una quinta variable, para analizar si el género puede condicionar los resultados del alumno.

VARIABLE 5: Mujer (0) y hombre (1).

La variable dependiente que tratamos de explicar es haber superado o no la asignatura de Matemáticas I⁵.

Para la recogida de datos, en el presente curso (2002/03) se les facilitó a los alumnos de la asignatura Matemáticas II de la Licenciatura de Administración y Dirección de Empresas una ficha en la que, además de sus datos personales, se les solicitaba información sobre el bachillerato de procedencia, asignaturas de matemáticas cursadas en el mismo, apoyo en la preparación de las asignaturas de matemáticas de la licenciatura, acceso a internet en el domicilio, calificación obtenida en la asignatura Matemáticas I y además un breve cuestionario para evaluar sus conocimientos previos.

Tras la recogida de las fichas se procedió a la creación de una base de datos sobre la que hemos trabajado para extraer la muestra que se utilizará en este estudio. El procedimiento de extracción ha sido el de muestreo aleatorio simple.

3.- METODOLOGÍA.

En primer lugar tratamos de averiguar si las variables independientes pueden separar a los individuos en dos grupos claramente diferenciados, los que han superado la asignatura de Matemáticas I y aquellos que no la han superado⁶.

Para cada uno de los individuos de la muestra se dispone de una información 5-dimensional que trataremos de representar en un espacio de dimensión menor (2 o a lo sumo tres dimensiones). Para ello haremos uso de la técnica de los Escalados Multidimensionales que brevemente describimos a continuación.

El escalado multidimensional es un conjunto de técnicas o métodos matemáticos utilizados en el campo de las Ciencias Sociales que permite representar las proximidades entre un conjunto de objetos o estímulos como distancias en un espacio de baja dimensionalidad. Mediante estas técnicas, se transforman las distancias o disimilaridades entre observaciones en un espacio de dimensión menor, de tal forma que se conserva la estructura subyacente a los datos. Al mismo tiempo se persigue, mediante la aplicación de la técnica, obtener una representación geométrica de los datos en un espacio de mínima dimensionalidad, de forma que sea interpretable por simple inspección visual. Es esta última característica del escalado multidimensional la que lo hace especialmente atractivo.

Básicamente el análisis de escalado multidimensional se basa en lo siguiente, dada n observaciones de 3^p , se define una medida de proximidad que cuantifica el grado de similitud o disimilitud entre éstas. A partir de esta medida se construye la matriz de similaridad o disimilaridad⁷. Como medida de proximidad se puede utilizar la correlación.

La técnica parte de una matriz de distancias o disimilaridades. Existen multitud de variantes de los escalados multidimensionales con funciones de costes ligeramente diferentes y algoritmos de optimización. Los algoritmos diseñados para analizar la matriz de disimilaridad, y que pueden ser utilizados para reducir la dimensionalidad del conjunto de datos, se dividen en dos tipos básicos: métricos, cuando se parte de una matriz de distancia⁸, y no métricos, cuando se parte de una matriz de similitudes o similaridades⁹.

En el supuesto del escalado métrico, la disimilaridad δ_{ij} entre dos observaciones i y j vendrá representada en un espacio multidimensional por una distancia d_{ij} , de forma que:

$$\delta_{ij} = f(d_{ij})$$

si d_{ij} es la distancia euclídea¹⁰:

$$d_{ij} = \left[\sum_{s=1}^p (x_{is} - x_{js})^2 \right]^{1/2}$$

f es una función lineal con pendiente positiva¹¹.

En el escalado no métrico (Shepard, 1962) se plantea que la función que relaciona las disimilaridades con las distancias es una función desconocida a la que sólo se le impone que sea una función monótona¹². El objetivo es encontrar unas coordenadas que sean capaces de reproducir las distancias a partir de la única condición de monotonía. Ello obliga a definir :

1.- Un criterio de bondad de ajuste que sea invariante ante transformaciones monótonas de los datos.

2.- Un algoritmo para obtener las coordenadas, optimizando el criterio establecido.

Habitualmente se utiliza como criterio de ajuste la denominada función de STRESS definida como:

$$S = \frac{\sum_{ij} (d_{ij} - \hat{d}_{ij})^2}{\sum_{ij} d_{ij}^2}$$

siendo \hat{d}_{ij} las denominadas “disparidades”, que son unos valores ajustados a las distancias y que están en un orden lo más similar posible a los datos.

En realidad, el S-stress es una medida de maldad de ajuste que será mejor cuanto más próximo a cero sea su valor. Además del Stress, existe otro índice de ajuste del modelo a los datos, denominado RSQ. Este último es una correlación cuadrática entre las disparidades derivadas a partir de los datos originales, y las distancias derivadas por el modelo de escalamiento, de modo que puede ser interpretado como la proporción de varianza en las disparidades que es explicada por las distancias (Real Deus, 2001, p.55). El RSQ es un valor entre cero y uno, y el ajuste será mejor cuanto más próximo a uno sea su valor.

El procedimiento de cálculo se hace en los siguientes pasos:

- Generar una configuración inicial de puntos en un espacio de dimensionalidad prefijada.
- Normalizar la configuración.
- Calcular las distancias entre cada par de puntos.
- Ajustar las distancias mediante una función monótona, obteniéndose las disparidades.
- Calcular el valor de la función de STRESS¹³. Si es aceptable, la configuración de puntos de partida es la solución final, en caso contrario, hay que mover la configuración de puntos en alguna dirección y se comienza de nuevo.

Tres son las cuestiones que hay que tener en cuenta al utilizar este método:

1.- La obtención de mínimos locales, es decir, al minimizar la función de STRESS el programa puede caer en un mínimo local y no continuar iterando, dando ésta como solución final. Esto no ocurre con la mayoría de los programas informáticos que se utilizan para este análisis.

2.- La decisión acerca del número de dimensiones de la solución final. Se puede optar por un método de prueba y error, es decir, utilizando distintas dimensiones y computando las funciones de STRESS hasta obtener una configuración que sea fácilmente interpretable y con un STRESS aceptablemente bajo. La segunda opción es basarse en un análisis previo del número de dimensiones mediante el Análisis de Componentes Principales.

3.- La interpretación de la solución. En muchos casos no es fácil y se debe tener en cuenta que la colocación de los ejes es arbitraria, siendo válida cualquier rotación que

ayude a interpretar la solución siempre que se mantenga la distancia interpuntos. Normalmente la salida de los programas de ordenador es tal que el primer eje abarca la mayor variabilidad. Si las dimensiones no tienen una lectura clara, es preferible ofrecer una interpretación localizando conglomerados de puntos.

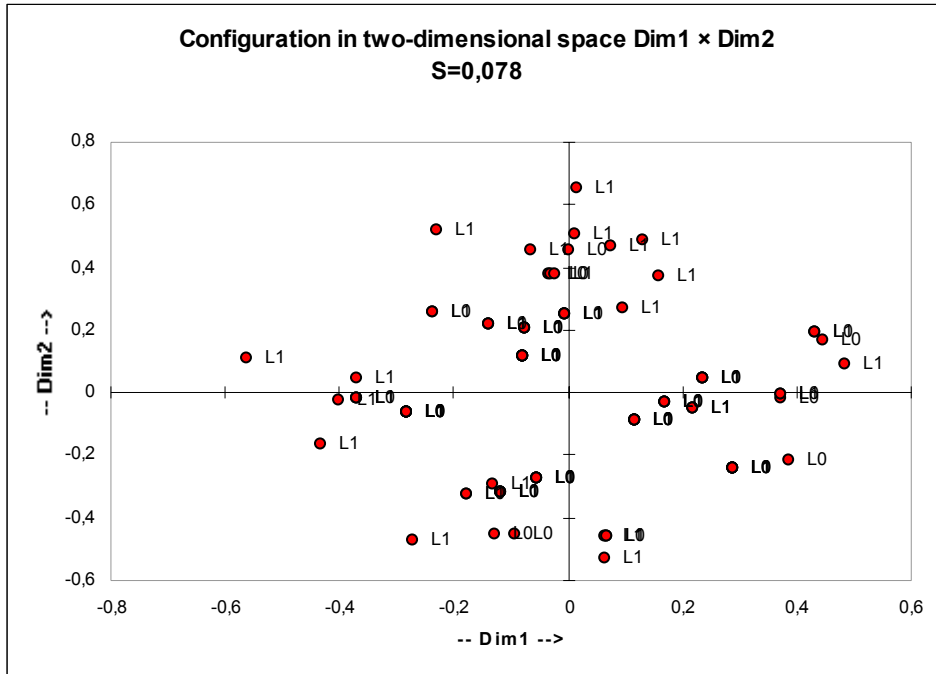
4.- RESULTADOS OBTENIDOS.

En este caso tenemos una matriz de 94 observaciones, correspondiente a 54 alumnos que han superado la asignatura de Matemáticas I de la Licenciatura de Administración y Dirección de Empresas¹⁴ y 40 alumnos que no la han superado. Para cada alumno hemos recogido un conjunto de cinco variables que se corresponden con las definidas en el apartado 2 de este trabajo, y que denominamos, bachillerato, matemáticas cursadas, familia, apoyo extra y sexo. Es decir, tenemos una matriz 94x5 donde por columnas están las variables y por filas las observaciones (alumnos).

La técnica del escalamiento se aplica mediante el paquete estadístico XLSTAT versión 6.0. Tras realizar diez iteraciones, la medida de error alcanzada (valor de función de STRESS) es de 0,078¹⁵.

El mapa bidimensional proporcionado es el que se muestra en la figura 1, donde se han situado las observaciones en función de sus coordenadas respecto a las dos dimensiones, L0 para los que no han superado la asignatura y L1 para los que si la han superado.

FIGURA 1.



En principio, en los cuatro cuadrantes aparecen ambos tipos de observaciones. Para poder distinguir mejor donde se sitúa cada grupo, los hemos representado por separado (figuras 2 y 3). Se puede observar que la dimensión dos es la que mejor distingue a los

grupos, situando a la mayor parte de los que no han superado la asignatura en la parte negativa de este eje y en la parte positiva a la mayor parte de los que la han superado.

FIGURA 2

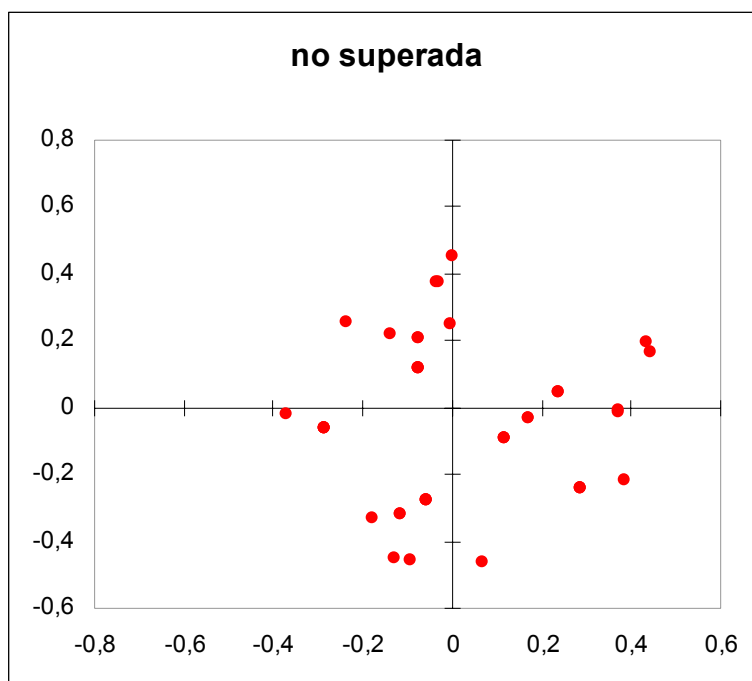
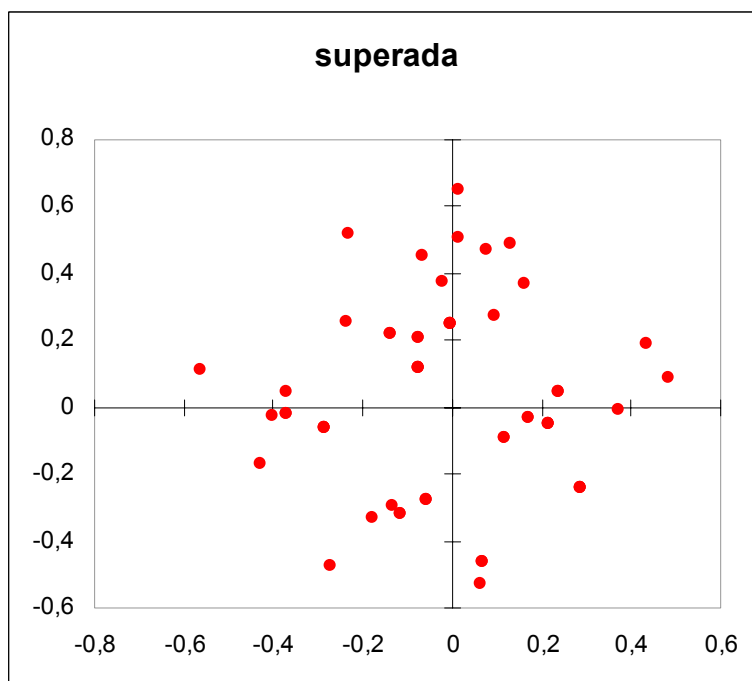


FIGURA 3



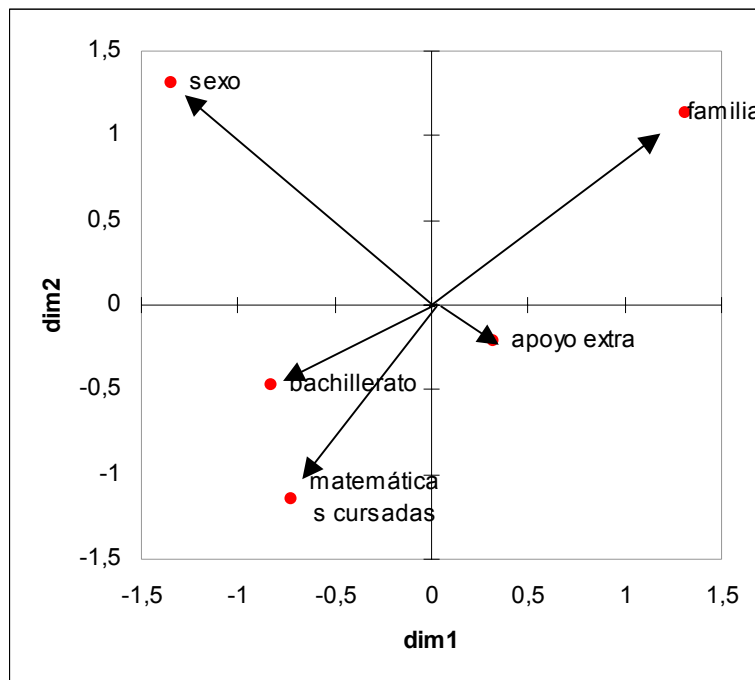
Una vez que hemos determinado el eje que mejor discrimina las observaciones, el problema que abordamos a continuación es la interpretación del mismo. Es decir, se trata de relacionar cada variable con las dos dimensiones que han servido para la construcción del mapa. Para ello utilizaremos el análisis proficit. Los resultados obtenidos se recogen en la tabla 1.

TABLA 1.

	dim1	dim2	R ²
Var1=bachillerato	-0.825	-0.475	0.061
Var2=mat. Cursadas	-0.721	-1.149	0.237
Var3=familia	1.310	1.134	0.780
Var4=apoyo extra	0.318	-0.218	0.055
Var5=sexo	-1.344	1.308	0.842

A continuación, representamos en el plano bidimensional (figura 4) las variables por medio de vectores que parten del origen y terminan en las coordenadas de la variable.

FIGURA 4



De la inspección visual del gráfico podemos extraer la siguiente información. Las variables que afectan de más a menos a la dimensión dos y por tanto que mejor discriminan son: apoyo extra, bachillerato de procedencia, entorno familiar, matemáticas cursadas y sexo. Aunque el bachillerato de procedencia y el número de matemáticas cursadas nos parecían a priori que eran las variables que más podían afectar a la superación o no de las asignaturas de matemáticas, no es ésta la lectura que puede extraerse del gráfico anterior. La variable apoyo extra es la que más condiciona, seguida de cerca por el bachillerato. Hemos de resaltar sin embargo, que cuando nos referimos al apoyo extra, en algunos casos no es tal, sino que la asistencia a clases en

“academia” se convierte en un sustituto de las clases en las aulas de la facultad. Esta sutil distinción no ha sido recogida en el estudio. Por último, hacemos notar, que la salida del domicilio familiar influye negativamente en nuestra variable dependiente. Esto quizás sea debido a que nos estamos refiriendo a alumnos de primer curso.

5.- REFLEXIONES FINALES.

En primer lugar queremos resaltar que el estudio realizado no nos conduce a una clara separación de los alumnos que superan las asignaturas de matemáticas de primer curso de la Licenciatura de Administración y Dirección de Empresas en función de las variables consideradas. Sin embargo, sí que se puede concluir una cierta separación entre alumnos que no superan la asignatura y alumnos que la superan, en función de lo que hemos denominado dimensión dos y que viene determinada de mayor a menor influencia por: apoyo extra, bachillerato de procedencia, entorno familiar, matemáticas cursadas y sexo.

En este trabajo se muestra además la utilidad de las técnicas de escalamiento multidimensional para obtener una representación multidimensional de datos multivariantes, especialmente útil si con dos dimensiones logramos sintetizar la información relevante.

Otra cuestión importante a destacar de esta técnica es que la incorporación al análisis de una nueva observación no implica necesariamente tener que repetir el estudio completo con el nuevo conjunto de datos. Basta con incorporar al gráfico un nuevo punto. Las coordenadas de este nuevo punto pueden obtenerse regresando en sentido inverso a como se construyeron las líneas correspondientes a las variables.

Una vez que se realice la prueba de evaluación correspondiente a la asignatura Matemáticas II se procederá a realizar de nuevo el estudio incorporando información adicional que poseemos sobre los conocimientos previos de los alumnos

Finalmente, debemos señalar que no podemos generalizar los resultados ya que la información de la que de momento disponemos sólo corresponde a este curso académico.

BIBLIOGRAFIA.

Fernández, P. Et al. (2000): Algunas consideraciones sobre la influencia de los estudios previos en las calificaciones de Matemáticas para los alumnos de la Diplomatura de Empresariales, VIII Jornadas ASEPUMA, Sevilla, pp. 117-128.

Fernández, P. Et al. (2001): Reflexiones sobre la adecuación de los conocimientos previos de los alumnos al temario de Matemáticas de la Escuela de Empresariales, IX Jornadas ASEPUMA, Las Palmas de Gran Canaria.

Kruskal, J. B.(1964, a): Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika* , 29, pp. 1-27.

Kruskal, J. B. (1964, b): Nonmetric multidimensional scaling: a numerical method, *Psychometrika*, 29, pp. 115-130.

Mantha, V.(1998): Factor Analysis and Multidimensional Scaling, Seminar Papers, Department for Electrical and Computer Engineering, Universidad Estatal de Mississipi.
Peña, D.(2002): Análisis de Datos Multivariantes, McGraw-Hill, Madrid.
Real Deus J.E. (2001): Escalamiento Multidimensional, Cuadernos de Estadística, Ed. LA MURALLA, Madrid.
Shepard, R.N.(1962): The analysis of proximities: Multidimensional scaling with an unknown distance function, I y II, Psychometrika, 27, pp.125-140 y 219-246.

¹ En esta línea hemos presentado otros trabajos, Fernández et al. 2000 y 2001.

² Que suele ser por lo general el examen tradicional.

³ De igual forma se podría conducir el estudio para la Licenciatura de Economía y para la Diplomatura de Ciencias Empresariales.

⁴ Por ejemplo, no es habitual el control de la asistencia.

⁵ Esta asignatura se cursa en el primer semestre de la Licenciatura de Administración y Dirección de Empresas y sus contenidos son básicamente de Álgebra Lineal.

⁶ Este último grupo engloba a los alumnos suspensos y a los no presentados.

⁷ Se dirá de similaridad si a mayor semejanza le corresponde mayor puntuación y de disimilaridad si a mayor semejanza le corresponde menor puntuación.

⁸ Un ejemplo típico sería la matriz de distancias kilométricas entre un conjunto de ciudades dadas dos a dos y mediante esta técnica trataríamos de representar en un plano esas ciudades.

⁹ Una matriz de este tipo se obtendría de comparar por ejemplo una marca de tabaco con otras, o una marca de coche

¹⁰ Se podrían considerar otras distancias.

¹¹ Para ver cómo proceder a partir de la matriz de distancias puede consultarse Peña (2002).

¹² Creciente si es de disimilaridades y decreciente si es de similaridades.

¹³ Kruskal (1964 a y b) considera que un valor de la función de STRESS igual a 0,01 indica una solución excelente, entre 0,01 y 0,05 buena, entre 0,05 y 0,10 pasable, entre 0,10 y 0,15 regular y mayor de 0,15 mala.

¹⁴ Su calificación en la prueba de evaluación correspondiente es superior a cinco.

¹⁵ Se ha optado por dos dimensiones porque la función de Stress para una sólo dimensión es muy alta $S=0,259$, y aunque para tres dimensiones se mejora ostensiblemente el ajuste $S=0,031$ no compensa por la dificultad visual que entrañan las representaciones tridimensionales.