



<Artículo>

Análisis de clasificación con variable criterio en SPAD

Angelina Sánchez-Martí, Antoni Ruiz-Bueno

Enviado: 09/07/2017

Aceptado: 18/09/2017

Publicado en prensa: 04/10/2017

Publicado: 08/01/2018

//Resumen

En este artículo se presentan las características, el procedimiento y la utilidad de la técnica de análisis de clasificación con variable criterio a partir de un conjunto grande de datos, principalmente con variables categóricas. Este análisis se inscribe entre las técnicas comúnmente conocidas como *minería de datos* o *data mining*, encargadas de analizar las relaciones o asociaciones entre todas las variables de una base de datos. El trabajo describe paso a paso cómo aplicar esta técnica estadística con el apoyo del *software* SPAD —un paquete estadístico empleado para realizar análisis de tipo multivariante— y aporta un ejemplo de cómo se aplica. Se trata de una técnica de la escuela de estadística francesa que, pese a ser poco conocida, proporciona un análisis de clasificación altamente interesante cuando se trabaja con grandes cantidades de datos —algo cada vez más recurrente en investigación educativa y más típico de los análisis secundarios que se llevan a cabo en nuestro campo.

//Palabras clave

Análisis de clasificación; Técnica de caracterización; Minería de datos; Análisis multivariante; Técnica de investigación.

//Referencia recomendada

Sánchez-Martí, A., y Ruiz-Bueno, A. (2018). Análisis de clasificación con variable criterio en SPAD. *REIRE Revista d'Innovació i Recerca en Educació*, 11(1), 41-53. <http://doi.org/10.1344/reire2018.11.119354>

//Datos de los autores

Angelina Sánchez-Martí. Universidad Autónoma de Barcelona. Departamento de Pedagogía Sistemática y Social, España. angelina.sanchez@uab.cat, ORCID: <http://orcid.org/0000-0002-4719-4688>

Antoni Ruiz-Bueno. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación, España. antoniruizbueno@ub.edu, ORCID: <http://orcid.org/0000-0001-9651-3633>



//Títol

Anàlisi de classificació amb variable «criteri» a SPAD

//Resum

Aquest article presenta les característiques, el procediment i la utilitat de la tècnica d'«anàlisi de classificació» amb variable «criteri» a partir d'un conjunt gran de dades, principalment amb variables categòriques. Aquesta anàlisi es troba entre les tècniques conegudes habitualment com a «mineria de dades» o *data mining*, encarregades d'analitzar les relacions o associacions entre totes les variables d'una base de dades. El treball descriu pas a pas com s'aplica aquesta tècnica estadística amb el suport del programari SPAD —un paquet estadístic utilitzat per fer anàlisis de tipus multivariant— i n'aporta un exemple d'aplicació. Es tracta d'una tècnica de l'escola d'estadística francesa que, malgrat ser poc coneguda, proporciona una anàlisi de classificació molt interessant quan es treballa amb grans quantitats de dades, fet cada vegada més recurrent en investigació educativa i més típic de les anàlisis secundàries que es porten a terme en el nostre camp.

//Paraules clau

Anàlisi de classificació; Tècnica de caracterització; Minería de dades; Anàlisi multivariant; Tècnica d'investigació.

//Title

Classification analysis with criterion variable using SPAD

//Abstract

This article presents the characteristics, procedure and utility of the technique of classification analysis with criterion variable in a large set of data, using mainly categorical variables. Classification analysis forms part of the techniques commonly known as data mining that analyse relationships or associations between variables. The paper describes step by step how to apply this statistical technique with the support of SPAD software, a statistical package for multivariate analysis, and provides an example of its application. The technique is drawn from the French school of statistics. Despite being little known, it is a very useful classification analysis for working with large amounts of data, a situation that is increasingly common in educational research, and more typical of the secondary analyses that are carried out in our field.

//Keywords

Classification analysis; Characterization technique; Data mining; Multivariate analysis; Research technique.



1. La minería de datos: ¿qué es y para qué puede utilizarse?

En los últimos años, el acceso y la monitorización de grandes cantidades de información y datos, sumado a la mejora del “software” y de los equipos informáticos para el análisis, ha dado lugar a la fusión de procedimientos propios del análisis exploratorio con la inteligencia artificial y la estadística: lo que hoy día se conoce como “data mining” o “minería de datos”. Este término se ha vinculado al uso de métodos específicos para explorar grandes matrices de datos, que no son otra cosa que algoritmos más o menos sofisticados que se aplican para descubrir relaciones nuevas y significativas, patrones, perfiles y tendencias.

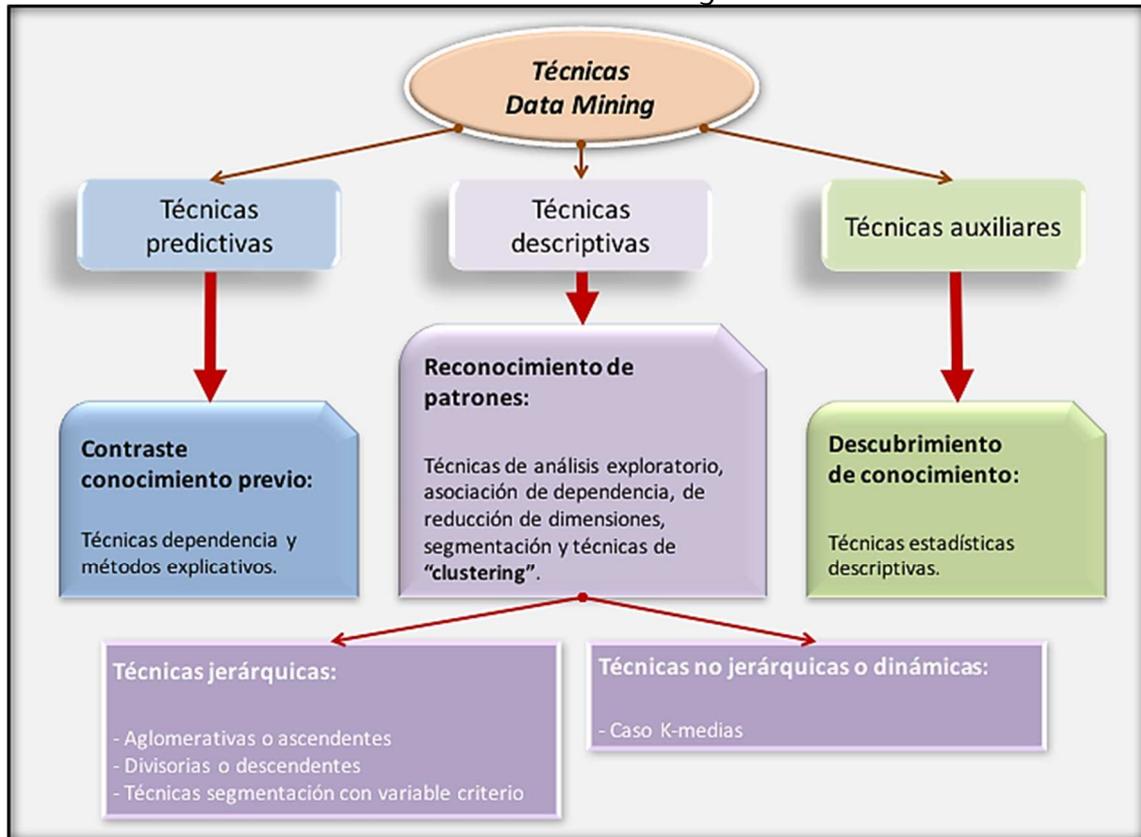
Estas técnicas son de gran utilidad para tomar decisiones, porque permiten confrontar numerosas informaciones de modo que resultan ser una estrategia analítica mucho más rica que mediante análisis separados. Para Bécue y Valls (2005) el hecho de poder disponer de representaciones simplificadas de grandes tablas de datos deviene un instrumento de síntesis altamente relevante, debido a su capacidad para reducir la dimensionalidad. De este modo, permiten extraer cuáles son las tendencias más destacadas, jerarquizarlas y, a su vez, eliminar los efectos marginales que perturben la percepción global de los hechos recogidos en los datos.

Dentro de las técnicas del “data mining” (Figura 1), se distingue entre las técnicas predictivas, las descriptivas y las auxiliares. Según Pérez y Santín (2008), las predictivas especifican el modelo para los datos en base a un conocimiento teórico previo, que debe contrastarse tras el proceso de minería de datos antes de aceptarlo como válido. Las variables de este tipo de técnicas se clasifican inicialmente en dependientes e independientes —de manera similar a las técnicas del análisis de la dependencia o los métodos explicativos del análisis multivariante—. Por su parte, las técnicas descriptivas parten del hecho de que todas tienen el mismo estatus y es a partir del reconocimiento de patrones que permiten la creación de modelos. En este grupo se incluyen las técnicas de “clustering” y segmentación, las de asociación y dependencia, las de análisis exploratorio de datos y las de reducción de la dimensión —factorial de componentes principales, correspondencias y de escalamiento multidimensional—. Por último, las técnicas auxiliares, a diferencia de las anteriores, son herramientas de apoyo más limitadas y superficiales y están enfocadas a descubrir el conocimiento a partir de los datos. En este caso, estas técnicas responden a nuevos métodos basados en técnicas estadísticas descriptivas, a partir de consultas e informes, enfocadas en general hacia la verificación.

De entre todas las técnicas mencionadas hasta hora, el análisis que nos ocupa en este artículo queda enmarcado dentro de las técnicas descriptivas y podríamos decir que se ubica en los denominados “análisis de formación de conglomerados, clases, grupos o clústers”. En este paraguas se agrupan, como decíamos, tanto las técnicas de formación de grupos —las técnicas que son, en cierta manera, exploratorias y automáticas, como es el caso de las técnicas de “clustering”—, como las de segmentación. Dentro de las técnicas automáticas se encuentran las jerárquicas, que pueden ser, a su vez, de dos tipos: las aglomerativas o ascendentes y las divisorias o descendentes. La característica de estas técnicas es que cuando el objeto (o en su defecto las variables) es clasificado en una etapa determinada, este no puede clasificarse de nuevo. Por el contrario, cuando en el procedimiento de cálculo pueden reasignarse los elementos a los grupos en distintas etapas, a estos procedimientos se les denomina “no jerárquicos o

dinámicos" (es el caso del K-medias) (Cea D'Ancona, 2004; González López Valcárcel, 1991; Lebart, Morineau y Fénelon, 1985).

Figura 1
Técnicas "data mining"



Fuente: Elaboración propia.

En el caso de las técnicas de segmentación, podemos encontrarnos con el análisis de clúster con variable criterio, que es el análisis que desarrollaremos en este artículo. No se trata propiamente de formas de clasificación automática, sino que más bien se parecen a las técnicas de regresión múltiple paso a paso (Lebart et al., 1985), ya que no se trata de extraer o poner de manifiesto los grupos o clases de un conjunto de variables, sino de encontrar el conjunto de individuos más "explicativos" de las modalidades de una variable cualitativa considerada de antemano (cabe apuntar que también es posible el análisis con una variable cuantitativa). El principio de cálculo es buscar la dicotomía inducida por una variable más relacionada con la variable que hay que segmentar o tomada como criterio de clasificación.

Tomando todos estos elementos en consideración, este análisis es interesante por dos motivos principales: por un lado, facilita que la persona investigadora pueda tomar decisiones óptimas en el contexto en el que se encuentra a la luz de toda la información y datos de los que dispone. Por el otro, permite estudiar de forma integrada un conjunto muy grande de datos que, a partir de un análisis estadístico de tipo unidimensional o bidimensional, sería muy costoso. Por tanto, responde a un tipo de análisis para la exploración de los datos que no pretende buscar relaciones entre las variables analizadas —lo que daría lugar a un análisis exploratorio tradicional—, sino centrar la atención en la generación de soluciones que conduzcan a predicciones útiles que

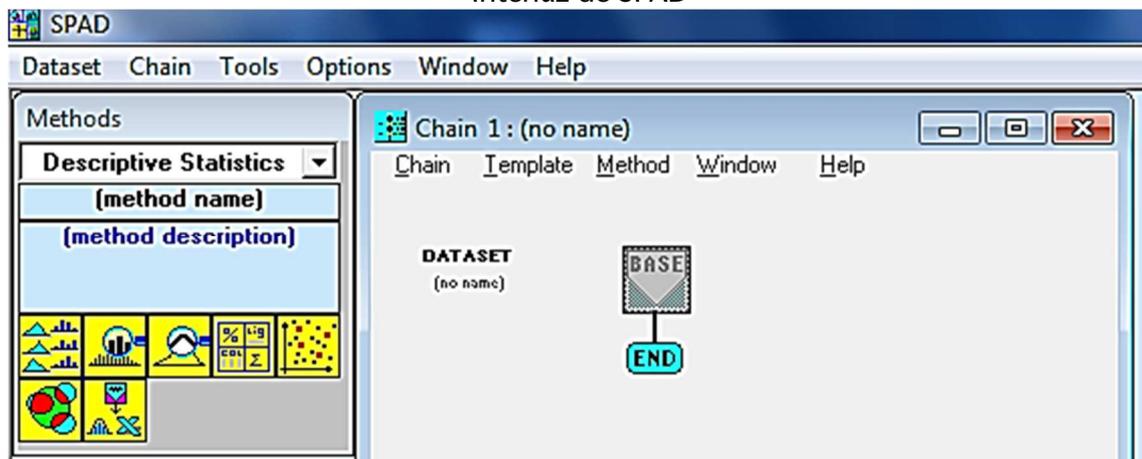


puedan ser analizadas bajo una óptica clasificatoria y de ordenación de todas las variables o categorías de mayor a menor asociación entre ellas.

2. Introducción al programa estadístico SPAD

Para este tipo de análisis, usamos y aconsejamos el programa estadístico SPAD (Système Portable pour l'Analyse de Donne), en su versión 5.6. La siguiente figura (Figura 2) muestra la ventana principal del programa. En ella se encuentran todos los métodos o técnicas disponibles (iconos en amarillo) agrupados según el tipo. Un aspecto interesante de la interfaz del programa es que en la ventana central se visualiza, en forma de iconos, la parametrización de los análisis, es decir, la secuencia llevada a cabo de los análisis y sus ficheros de resultados. Por último, como en todo "software", el programa incluye una barra de herramientas con los comandos disponibles.

Figura 2
Interfaz de SPAD



Fuente: Elaboración propia.

A continuación, y antes de presentar la técnica de análisis que nos interesa, planteamos cómo manejar con eficacia una base de datos con SPAD¹.

Organización y tratamiento previo de los datos

En la gran mayoría de las investigaciones con corpus empírico, la exploración inicial de los datos suele ser el primer paso del análisis de los resultados. Aunque el propósito de la recogida de datos sea uno u otro —la verificación de hipótesis, el descubrimiento de una teoría, la confección de un modelo, etc.—, antes de aplicar cualquier otra técnica analítica es aconsejable hacer un estudio exploratorio de los datos.

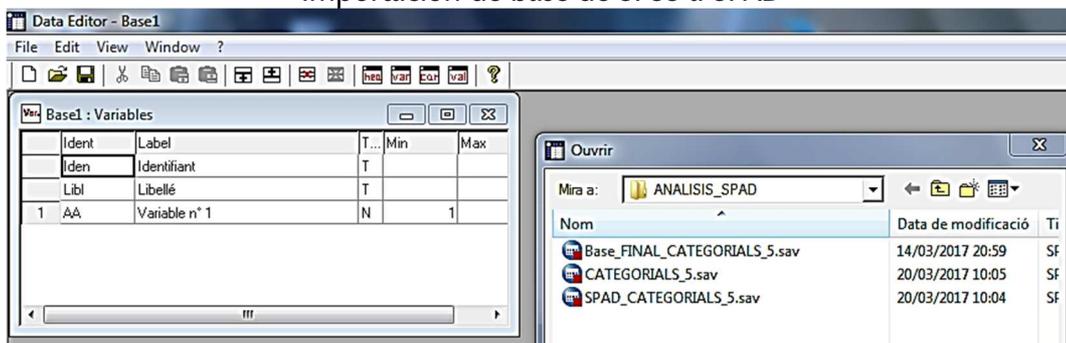
En el caso de investigaciones con diseños metodológicos de tipo cuantitativo, las primeras exploraciones tienen que ver con el análisis de cada una de las variables a partir de procesos de estadística descriptiva univariante respetando la naturaleza de las mismas, lo cual suele implicar la recodificación y la creación de nuevas variables a lo largo del proceso. En el campo de las ciencias sociales, hoy en día es prácticamente inevitable no gestionar,

¹ El archivo de datos utilizado como ejemplo se puede descargar directamente desde el apartado "Herramientas del artículo > Archivos complementarios" que ofrece la página web de este artículo.

organizar y tratar los datos a través de procesos estadísticos, con el apoyo de programas informáticos como el SPSS-Win (Statistical Package for the Social Sciences) (Vilà, 2006). Por este motivo, previamente a cualquier otro tratamiento, se aconseja optar por la creación de lo que llamamos una “base de datos de trabajo” o “matriz de gestión” en la que integrar todos los datos recogidos mediante escalas, cuestionarios u otros instrumentos que formen parte del estudio. Esta matriz contendrá todas aquellas variables que interesan para la finalidad de la investigación. Una vez organizados los datos, se sugiere llevar a cabo análisis estadísticos de tipo exploratorio con el apoyo de procedimientos gráficos de representación de matrices de datos que permitan conocer o aproximarse a las variables que van a ser tratadas “a posteriori” con el apoyo del “software” SPAD.

Tras la configuración de la base de datos (SPSS, Excel...) y de su exploración inicial, es necesario también preparar los ficheros que van a utilizarse en SPAD para asegurarnos de que este “software” es capaz de interpretar adecuadamente las variables que conformarán su propia base interna de datos. El programa permite crear su propia base de datos, con extensión .sba. También permite importar una base de datos de SPSS, de extensión .sav. Por los motivos antes mencionados, sugerimos optar por este procedimiento. Dado que el archivo predefinido de trabajo de SPAD en el que se almacenan los datos es el archivo BASE (.sba), convertiremos la base de datos .sav de SPSS a .sba. Para ello, accedemos a la pestaña “Dataset” de la barra de herramientas y seleccionamos “New dataset”. Una vez dentro, seleccionamos “File” y abrimos el archivo SPSS que almacena los datos (la matriz de gestión antes mencionada, cuya extensión es precisamente .sav) (Figura 3).

Figura 3
Importación de base de SPSS a SPAD



Fuente: Elaboración propia.

La base, una vez abierta en SPAD, mantiene todas las etiquetas tanto de las variables como de los valores de cada variable de nuestra matriz de datos inicial. SPAD reconoce el archivo .sav de SPSS, pero no permite trabajar con él. Por lo tanto, debe salvarse con la extensión que exige este programa (.sba). Es importante recordar que la conversión a dicha extensión puede sobrescribir el archivo, de modo que sugerimos salvar la nueva Base.sba con un nombre diferente.

Para el tipo de análisis que se plantea en esta ficha es necesario tener en cuenta que el análisis trabaja con las modalidades o categorías de todas las variables de tipo cualitativo (llamadas “categoriales”). Por lo tanto, como persona investigadora cabe preguntarse si interesa convertir algunas de las variables que formen parte de la base de datos inicial de



SPSS antes de trasladar la base a SPAD, de modo que este programa las trate y reconozca como categoriales y no como variables continuas. En otras palabras, si la base de datos SPSS contiene variables de tipo escalar, por ejemplo, estas podrían convertirse a categoriales (recodificarlas). Un procedimiento simple para hacerlo consiste en convertir cada variable escalar en las categorías que nos proporcionan los cuartiles. De este modo, cada categoría de la variable va a ser considerada por el programa como una unidad. Este procedimiento debe hacerse antes de convertir la base de datos SPSS a SPAD: Si no el programa entenderá que estas variables son continuas.

3. Análisis multivariado de datos con el "software" SPAD: el análisis de clasificación con variable criterio

El análisis propuesto con el apoyo del programa estadístico SPAD es, según Bécue y Valls (2005), una técnica original y poco conocida para caracterizar una variable respuesta a partir de un conjunto de variables explicativas. Este método se diferencia respecto a cuando la variable respuesta es categórica o continua. En cualquier caso, se aplica un algoritmo que realiza comparaciones múltiples entre medias y/o proporciones para encontrar las variables explicativas relacionadas.

Esta técnica de clasificación de la escuela estadística francesa permite, por ejemplo, saber qué características diferencian a los estudiantes que transitan de una etapa formativa a otra de los estudiantes que no lo hacen, tomando como variables explicativas la edad, el género, el nivel de estudios, etc. En el caso de que la variable sea continua, nos permitiría encontrar las diferencias entre individuos con valores altos y valores bajos de la variable respuesta. Por lo tanto, cuando la variable es categórica, la categoría de la variable respuesta define implícitamente un subgrupo de individuos del conjunto global de participantes. El algoritmo realiza pruebas estadísticas para comparar las proporciones de individuos que presentan las categorías *a*, *b*, *c*, etc. dentro del subgrupo respecto a las proporciones en la muestra global. En cambio, en el caso de que se desee caracterizar el subgrupo de individuos a partir de una variable explicativa continua se realiza un test de comparación de medias entre la media del subgrupo respecto a la media global (Bécue y Valls, 2005). Eso nos permitirá decidir si nos interesa recodificar previamente alguna de las variables continuas a variables categoriales.

Para detectar las categorías que caracterizan mejor se evalúa la desviación entre los porcentajes de la categoría en el grupo y el porcentaje en la muestra. Así, el programa SPAD calcula el estadístico de contraste y el p-valor correspondiente a la hipótesis nula de que no hay diferencias en la muestra y en el grupo para cada una de las categorías. Luego el resultado se transforma en el correspondiente valor-test (valor de la distribución normal estándar que corresponde a la misma probabilidad) (Grande y Abascal, 2005). Cuanto menor es el p-valor o mayor el valor-test mejor caracteriza la categoría.

Incorporación de la base de datos en el programa SPAD

El proceso del tratamiento de datos para aplicar la prueba estadística que nos interesa es el siguiente:

Angelina Sánchez-Martí, Antoni Ruiz-Bueno. *Análisis de clasificación con variable criterio en SPAD*

Para introducir el fichero de datos con el que vamos a trabajar, damos doble clic en el icono "Base". El fichero que debe incorporarse en el "software" pide la extensión .sba (precisamente la que hemos guardado anteriormente). Buscamos ese fichero en la carpeta en la que esté ubicado. Una vez introducida la base en su formato correspondiente, el icono de la "Base" se ve resaltado y aparece el nombre de la base introducida (Figura 4).

Figura 4
Incorporación de la matriz de datos al fichero BASE

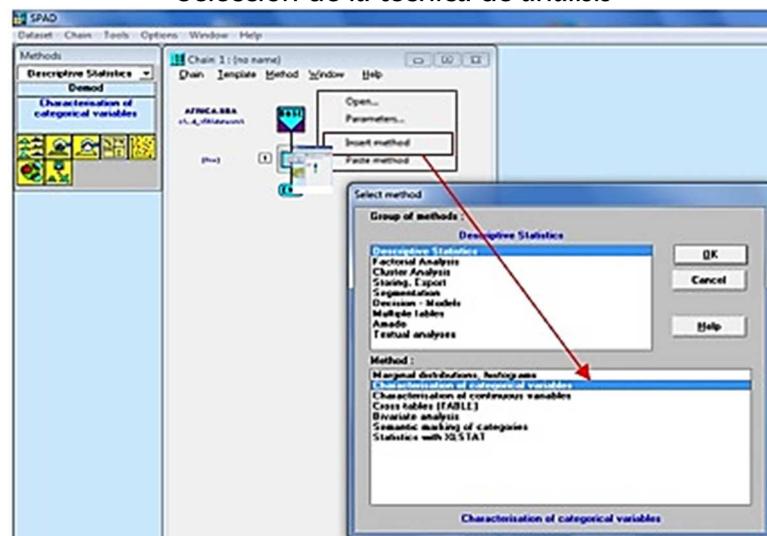


Fuente: Elaboración propia.

Ejecución de los análisis en el programa SPAD

Para poder iniciar un análisis, debemos situar el cursor nuevamente encima del icono "BASE" y pulsar el botón derecho del ratón para seleccionar la opción "Insert method". Nos aparece un nuevo icono. Pulsando dos veces encima de él nos aparecen las opciones correspondientes a los tipos de análisis. Nos interesa seleccionar la "técnica de clasificación con criterio" y, en este caso, de una variable categorial ("Characterization of categorical variables") (Figura 5). Si lo preferimos, también podemos introducir la técnica arrastrando su icono correspondiente (iconos en amarillo) en el recuadro del método ("Methods").

Figura 5
Selección de la técnica de análisis



Fuente: Elaboración propia.

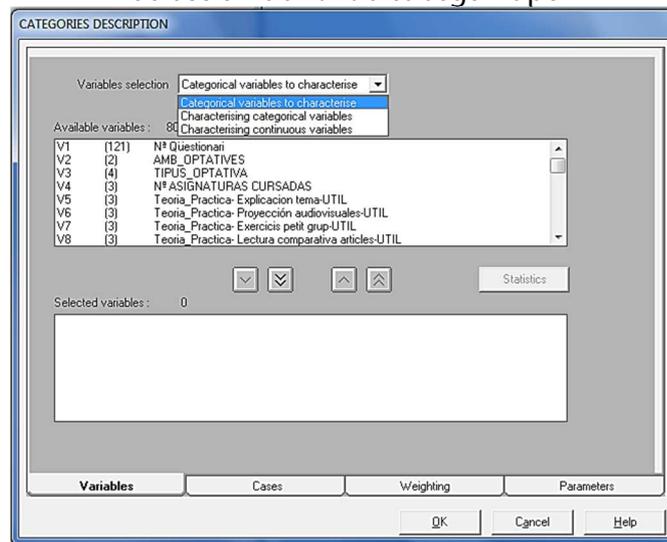


Angelina Sánchez-Martí, Antoni Ruiz-Bueno. *Análisis de clasificación con variable criterio en SPAD*

Posteriormente, pulsamos dos veces en el icono que visualiza el método aplicado ("Method: Demod"). Se abre una ventana en la que puede verse el listado completo de las variables de la base. La ventana contiene dos recuadros (Figura 6). El de arriba contiene las variables disponibles y en el de abajo irán apareciendo las variables que seleccionemos. La instrucción "Variables selection" nos permite tres opciones para seleccionar las variables de la base que nos interesen:

- "Categorical variables to characterise": indica las variables categóricas que seleccionemos para caracterizar, es decir, la o las variables que vamos a caracterizar a partir de todas las variables de nuestra base (categóricas y continuas). Si nos interesa caracterizar más de una variable, podemos seleccionar más de una.
- "Characterising categorical variables": contiene todas las variables categóricas que pueden caracterizar la variable antes escogida (la que deseamos diferenciar). Debemos decidir si nos interesa seleccionar todas las que tenemos disponibles o bien omitir alguna.
- "Characterising continuous variables": contiene todas las variables continuas que pueden caracterizar la variable antes escogida. Debemos decidir si nos interesa seleccionar todas las que tenemos disponibles o bien omitir alguna.

Figura 6
Selección de variables según tipo



Fuente: Elaboración propia.

Una vez seleccionadas todas las variables ("variables por caracterizar"² y "variables caracterizadoras"),³ como en todo programa podemos definir o modificar los parámetros de análisis. Por defecto, el programa trabaja con significaciones iguales o inferiores a 0,01. Como es conocido, el mínimo en ciencias sociales es el 0,05; por lo tanto, puede

² En este caso, si tomamos como prueba la matriz de ejemplo de este artículo, seleccionamos la variable V3 Transita a estudios PO Sí/No (generada).

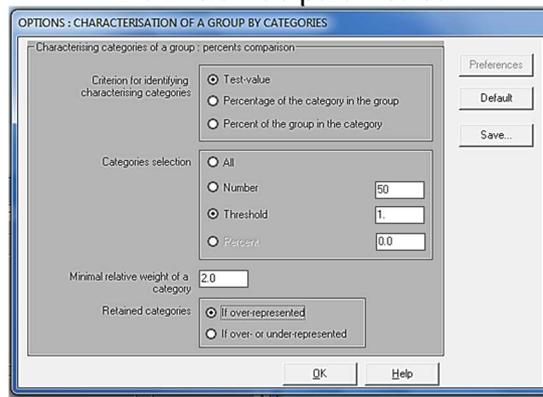
³ Siguiendo el caso del ejemplo, seleccionamos aquí el resto de las variables de la matriz de datos. En cada estudio deberemos seleccionar aquellas variables que nos sean de interés en función de nuestros objetivos.



Angelina Sánchez-Martí, Antoni Ruiz-Bueno. *Análisis de clasificación con variable criterio en SPAD*

cambiarse dicho parámetro (Parametres – Options⁴ – Threshold → De 1. a 5.). Además, sugerimos seleccionar en la opción “Retained categories” únicamente aquellas que sean representativas de las categorías de la variable (“If over-represented”) para que el listado discrimine solamente aquellas variables que aportan diferenciación estadística (Figura 7). Es aconsejable no tocar el resto de los parámetros que aparecen por defecto.

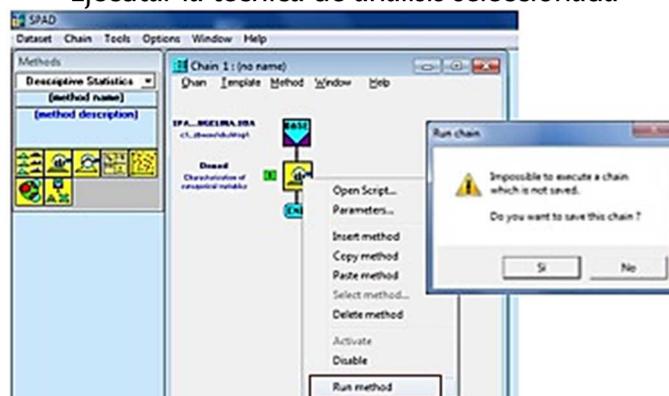
Figura 7
Definición de parámetros



Fuente: Elaboración propia.

Una vez realizados los pasos anteriores, ya podemos ejecutar el análisis. Pulsamos sobre el icono del método o técnica, ahora ya en amarillo, con el botón derecho del ratón para seleccionar la opción “Run method”⁵. SPAD nos devuelve una ventana que nos indica si queremos salvar el procedimiento que estamos llevando a cabo. Le decimos que sí. Se le debe asignar un nombre al fichero de análisis interno. Sugerimos guardarlo con una “a” (a.FIL). Cada vez que se haga un análisis nos pedirá lo mismo, de modo que el fichero “a” se irá reescribiendo (Figura 8).

Figura 8
Ejecutar la técnica de análisis seleccionada



Fuente: Elaboración propia.

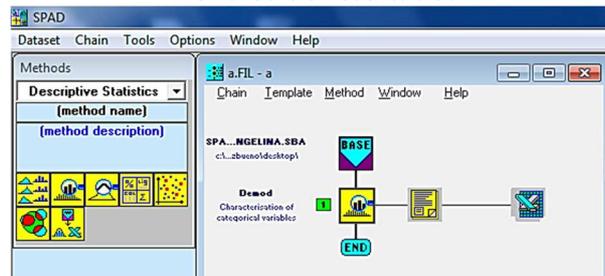
⁴ Debemos indicar esta opción únicamente en relación a la casilla “By categories”.

⁵ En caso de que aparezca un error con el mensaje “Des fichiers résultats sont en cours d'utilisation. Veuillez les fermer: fichier Excel,...” debemos cerrar el programa y volver a abrirlo pulsando el botón derecho y seleccionando “Ejecutar como administrador”.

Angelina Sánchez-Martí, Antoni Ruiz-Bueno. *Análisis de clasificación con variable criterio en SPAD*

Posteriormente nos aparecen en pantalla dos archivos de resultados en dos formatos distintos: uno en Excel y otro en .txt. El usuario puede decidir en qué formato prefiere visualizarlos. Sin embargo, la visualización de uno y otro no es exactamente la misma. Sugerimos optar por el formato .excel, ya que tiene la ventaja de que proporciona los resultados en una tabla que puede trasladarse a cualquier otro documento (Figura 9).

Figura 9
Ficheros de resultados



Fuente: Elaboración propia.

El programa SPAD arroja los resultados por separado: por una parte, muestra las variables categóricas que caracterizan la variable criterio seleccionado; por la otra, muestra las variables continuas (que trata como unidades).

Ejemplo de análisis de caracterización de una variable cualitativa

Vamos a ver un ejemplo con la interpretación que puede derivar del mismo: la siguiente figura (Figura 10) muestra un extracto de las variables que definen el análisis hecho para caracterizar qué estudiantes transitan de estudios obligatorios a posobligatorios y qué estudiantes, por el contrario, no lo hacen. La figura muestra, en este caso, la salida de resultados en formato .txt.

Figura 10
Fragmento de resultados en formato .txt

```

CHARACTERISATION OF CATEGORICAL VARIABLES
DESCRIPTION OF: Transita a estudios PO Sí/No (generada)
CATEGORIES CHARACTERISATION BY CATEGORIES
CHARACTERISATION BY CATEGORIES OF CLUSTERS OR CATEGORIES
OF Transita a estudios PO Sí/No (generada)
Si
-----
T.VALUE  PROB.  ----  PERCENTAGES  ----  CHARACTERISTIC  IDEN  WEIGHT
GRP/CAT  CAT/GRP  GLOBAL  CATEGORIES      OF VARIABLES
-----
11.12  0.000  100.00  100.00  80.00  Si  Resolución momento de transicion (repite, abandona, gradua..  GC01  108
11.12  0.000  100.00  100.00  80.00  Si  Transita a estudios PO Sí/No (generada)  GC01  108
10.15  0.000  100.00  96.30  77.04  Transita (transita a  Detalle de la transición con los abandonos prematuros  GT05  104
5.77  0.000  100.00  57.41  45.93  Si  Transitan a otro instituto  GT01  62
4.71  0.000  100.00  44.44  35.56  Pendiente  Tipo estudios específico en el momento de transición  GT07  48
4.71  0.000  100.00  44.44  35.56  Pendiente  Tipo estudios elegido en el momento de transición  GC07  48
4.56  0.000  100.00  42.59  34.07  Bachillerato  Tipo estudios elegido en el momento de transición  GC01  46
3.72  0.000  91.25  67.59  59.26  Seguro  Aprobarás el curso actual Recodificado  VZ01  80
3.53  0.000  91.03  65.74  57.78  Si, seguro  Aprobarás el curso actual  GP02  78
3.11  0.001  91.18  57.41  50.37  Bien-Muy bien  Como me van los estudios Recodificado  VZ02  68
2.63  0.004  92.16  43.52  37.78  Bien  Cómo crees que te van los estudios  GP03  51
2.61  0.005  100.00  20.37  16.30  Bachillerato científ  Tipo estudios específico en el momento de transición  GT01  22
2.61  0.005  100.00  20.37  16.30  Bachillerato social-  Tipo estudios específico en el momento de transición  GT02  22
2.02  0.022  83.48  88.89  85.19  Si, seguro  Cuando aprueves 4º de ESO, ¿crees que continuarás estudiando  GP02  115
2.01  0.022  100.00  14.81  11.85  CFPM Gestió administ  Curso  GC09  16
1.95  0.025  87.69  52.78  48.15  INS 3  Instituto  GI03  65
1.80  0.036  89.36  38.89  34.81  Años primera vez escuela_Catalunya  GP02  47
1.79  0.037  100.00  12.96  10.37  1ºBATX  Curso  GC06  14
1.79  0.037  100.00  12.96  10.37  Ciclos Formativos  Tipo estudios elegido en el momento de transición  GC02  14
1.79  0.037  100.00  12.96  10.37  Ciclo formativo  Tipo estudios específico en el momento de transición  GT04  14
1.72  0.042  92.86  24.07  20.74  Más de 10  Número de amigos Recodificado  NU03  28
1.68  0.046  86.11  57.41  53.33  Primaria  Curso incorporac escuela_Catalunya  GP02  72
-----
No
-----
T.VALUE  PROB.  ----  PERCENTAGES  ----  CHARACTERISTIC  IDEN  WEIGHT

```

Continúa en la siguiente página

Angelina Sánchez-Martí, Antoni Ruiz-Bueno. *Análisis de clasificación con variable criterio en SPAD*

GRP/CAT	CAT/GRP	GLOBAL	CATEGORIES	OF VARIABLES				
11.12	0.000	100.00	100.00	20.00	No	GC02	27	
7.42	0.000	100.00	59.26	20.00	No	GC02	27	
6.15	0.000	100.00	44.44	11.85	No P.O.	GT06	16	
5.77	0.000	36.99	100.00	8.89	Repite (4º ESO)	GT02	12	
5.48	0.000	100.00	37.04	54.07	No	GT02	73	
5.48	0.000	100.00	37.04	7.41	no gradua y repite 4	Resolución momento de transición (repite, abandona, gradua..GM02	10	
5.48	0.000	100.00	37.04	7.41	repite ESO	Tipo estudios elegido en el momento de transición	GC04	10
5.12	0.000	100.00	33.33	6.67	Se desconoce	Tipo estudios elegido en el momento de transición	GC06	9
5.12	0.000	100.00	33.33	6.67	Se desconoce	Detalle de la transición con los abandonos prematuros	GT06	9
4.75	0.000	100.00	29.63	5.93	gradua y se desconoc	Resolución momento de transición (repite, abandona, gradua..GM05	8	
4.37	0.000	81.82	33.33	8.15	Desconocido	Tipo estudios específico en el momento de transición	GT08	11
3.95	0.000	100.00	22.22	4.44	Abandona	Tipo estudios elegido en el momento de transición	GC05	6
3.72	0.000	36.36	74.07	40.74	Duda	Aprobarás el curso actual Recodificado	V202	55
3.11	0.001	31.34	77.78	49.63	Mal-Regular	Como me van los estudios Recodificado	V201	67
2.48	0.007	39.29	40.74	20.74	Europa no-comunitari	País Padre	GP03	28
2.44	0.007	100.00	11.11	2.22	Fracasa (no aprueba	Detalle de la transición con los abandonos prematuros	GT01	3
2.44	0.007	100.00	11.11	2.22	no gradua y abandona	Resolución momento de transición (repite, abandona, gradua..GM04	3	
2.44	0.007	100.00	11.11	2.22	gradua y abandona	Resolución momento de transición (repite, abandona, gradua..GM03	3	
2.44	0.007	100.00	11.11	2.22	Abandona (aprueba 4º	Detalle de la transición con los abandonos prematuros	GT03	3
2.40	0.008	62.50	18.52	5.93	No, creo que repetir	Aprobarás el curso actual	GP03	8
2.38	0.009	30.00	66.67	44.44	Regular	Cómo crees que te van los estudios	GP02	60
2.15	0.016	55.56	18.52	6.67	4 ESO E	Curso	GC05	9
2.08	0.019	30.61	55.56	36.30	No lo sé	Aprobarás el curso actual	GP01	49
1.96	0.025	75.00	11.11	2.96	España	País Madre	GP01	4
1.79	0.037	32.35	40.74	25.19	INS 2	Instituto	GI02	34
1.76	0.039	38.89	25.93	13.33	No lo sé	Cuando aprueves 4º de ESO, ¿crees que continuarás estudiando?	GP01	18

Fuente: Elaboración propia.

Mediante los resultados obtenidos, podríamos concluir lo siguiente:

Los jóvenes que transitan de la etapa obligatoria a la posobligatoria en secundaria lo hacen tanto hacia bachillerato como hacia ciclos formativos, aunque la opción de bachillerato tiene una mayor influencia en la transición positiva. El hecho de continuar los estudios en el mismo instituto tiene influencia en cuanto a transitar. Presentan trayectorias educativas lineales, sin repetir cursos y, en el momento de transitar a la etapa posobligatoria, están convencidos de que continuarán estudiando. Creen que los estudios les van bien. Por el contrario, los alumnos o jóvenes que no transitan tienen tendencia a repetir el último curso de la educación obligatoria y no a abandonar. En su momento, dudaban de aprobar el último curso de ESO y son jóvenes que, en general, consideran que los estudios les van mal o regular.

Estas conclusiones dan cuenta del tipo de caracterización que puede hacerse. Las interpretaciones pueden acompañarse del detalle de los estadísticos arrojados. La columna peso ("Weight") recoge, por ejemplo, el número de personas encuestadas que ha elegido la categoría caracterizante, de manera que puede utilizarse junto al valor p (PROB.) y el valor del test (T.VALUE) para justificar las relaciones. Puede verse otro ejemplo de esta técnica de análisis aplicada en educación en Rubio, Ruiz Bueno y Martínez Olmo (2016).

<Referencias bibliográficas>

- Bécue, M., y Valls, J. (2005). *Manual de introducción a los métodos factoriales y clasificación con SPAD*. Bellaterra: Servei d'Estadística. Universidad Autònoma de Barcelona. Recuperado de <http://sct.uab.cat/estadistica/sites/sct.uab.cat/estadistica/files/manualSPAD.pdf>
- Cea D'Ancona, M. (2004). *Análisis multivariante. Teoría y práctica en la investigación social*. Madrid: Síntesis.
- González López-Valcárcel, B. (1991). *Análisis multivariante. Aplicación al ámbito sanitario*. Barcelona: SG Editores.
- Grande, I., y Abascal, E. (2005). *Análisis de encuestas*. Madrid: ESIC Editorial.



Angelina Sánchez-Martí, Antoni Ruiz-Bueno. *Análisis de clasificación con variable criterio en SPAD*

Lebart, I., Morineau, A., y Fénelon, J. (1985). *Tratamiento estadístico de datos. Métodos y programas*. Barcelona: Marcombo.

Pérez López, C., y Santín, D. (2008). *Minería de datos: técnicas y herramientas* (2ª ed.). Madrid: Thomson Ediciones Paraninfo.

Rubio, M. J., Ruiz-Bueno, A., y Martínez-Olmo, F. (2016). Percepción del alumnado sobre la utilidad de las actividades de aprendizaje para desarrollar competencias. *Revista de Investigación Educativa*, 34 (1), 221-240. <http://dx.doi.org/10.6018/rie.34.1.225131>

Vilà, R. (2006). *¿Cómo hacer un análisis cuantitativo de datos de tipo descriptivo con el paquete estadístico SPSS?* Butlletí LaRecerca, 6, 1-8. Recuperado de <http://www.ub.edu/ice/recerca/pdf/ficha6-cast.pdf>

Copyright © 2018. Esta obra está sujeta a una licencia de Creative Commons mediante la cual, cualquier explotación de ésta, deberá reconocer a sus autores, citados en la referencia recomendada que aparece al inicio de este documento.



Institut de Ciències de l'Educació
UNIVERSITAT DE BARCELONA