

<Artículo>

Cómo aplicar un *cluster* jerárquico en SPSS

Ruth Vilà-Baños, María-José Rubio-Hurtado, Vanesa Berlanga-Silvente, Mercedes Torrado-Fonseca

Fecha de presentación: 30/08/2013

Fecha de publicación: 10/01/2014

//Resumen

El análisis de conglomerados o *cluster* es una técnica multivariante que busca agrupar elementos o variables tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre ellos, mediante una estructura jerarquizada para poder decidir qué nivel jerárquico es el más apropiado para establecer la clasificación.

El programa SPSS dispone de tres tipos de análisis de conglomerados: el análisis de conglomerados jerárquico, bietápico y de K medias. Aplicaremos el método jerárquico como el más idóneo para determinar el número óptimo de conglomerados existente en los datos y su contenido para nuestro caso práctico.

//Palabras clave

Clasificación, conglomerados, *cluster* jerárquico y aglomerativo.

// Referencia recomendada

Vilà-Baños, R. Rubio-Hurtado, M. J., Berlanga-Silvente, V., y Torrado-Fonseca, M. (2014). Cómo aplicar un *cluster* jerárquico en SPSS. [En línea] *REIRE, Revista d'Innovació i Recerca en Educació*, 7 (1), 113-127. Accesible en: <http://www.ub.edu/ice/reire.htm>

// Datos de los autores

Ruth Vilà-Baños. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). ruth_vila@ub.edu

María-José Rubio-Hurtado. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). mjrubio@ub.edu

Vanesa Berlanga-Silvente Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). berlanga.silvente@ub.edu

Mercedes Torrado-Fonseca. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). mercedestorrado@ub.edu



1. Introducción

El cluster es una técnica de clasificación que sirve para poder detectar y describir subgrupos de sujetos o variables homogéneas en función de los valores observados dentro de un conjunto aparentemente heterogéneo. Se fundamenta en el estudio de las distancias entre ellos, permitiendo en el análisis, cuantificar el grado de similitud, en el caso de las proximidades, y el grado de diferencia, en el caso de las distancias. Como resultado aparecen agrupaciones o clusters homogéneos.

A diferencia de otras pruebas de clasificación, como por ejemplo el análisis discriminante (véase Torrado y Berlanga, 2013), las agrupaciones o conglomerados que se establecen se configuran a *posteriori*. En este caso el investigador no tiene conocimiento de la existencia de los subgrupos o conglomerados, ni del número resultante, ni mucho menos de las características que los definen. Es una técnica, por lo tanto, eminentemente exploratoria y descriptiva sin variables dependientes.

El objetivo fundamental de esta técnica es la configuración de grupos similares y homogéneos para poder entender y estudiar mejor los fenómenos sociales y educativos. A modo de ejemplo, si nuestro objetivo es conocer las características que identifican las escuelas públicas que agrupan a los estudiantes con un mayor éxito académico necesitaremos conocer las características básicas que definen y describen los diferentes grupos de escuelas. El análisis de *clusters* nos permitirá dar respuesta a interrogantes del tipo: ¿Las escuelas públicas se pueden agrupar? y ¿cuáles son las características comunes de las escuelas agrupadas?

En el proceso de aplicación de esta técnica el investigador tiene que tomar una serie de decisiones: en primer lugar, debe seleccionar las variables relevantes para identificar los grupos (en el caso del ejemplo, líneas de la escuela, número de profesorado fijo, rendimiento académico, etc.); en segundo lugar, ha de establecer las medidas de similitud para controlar las medidas de proximidad entre las unidades y, finalmente, debe fijar qué procedimiento o método de agrupación se utilizará para configurar los conglomerados.

En cuanto al procedimiento, el análisis de *cluster* permite hacer una agrupación jerarquizada o no. El procedimiento jerárquico es más adecuado para muestras pequeñas en las que los grupos se configuran por agrupaciones sucesivas de individuo a individuo, o de individuo a grupo configurando una estructura arborescente con niveles, que desemboca en una jerarquización de *clusters*. El procedimiento no jerárquico (K-medias) parte, de entrada, de la determinación de un número de grupos y asigna los casos a grupos diferenciados sin que unos dependan de otros (*clusters* no jerárquicos).

El método de la técnica puede ser aglomerativo o divisivo en función del punto de partida. Si partimos de tantos grupos como objetos tengamos y se van obteniendo las agrupaciones pertinentes, nos situamos en el método aglomerativo. Sin embargo, si partimos de un único grupo y este se va subdividiendo, optamos por el método divisivo. El método más utilizado es el aglomerativo.



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

Pérez (2009) propone la siguiente clasificación de métodos de análisis de conglomerados: aglomerativos-divisivos, jerárquicos-no jerárquicos, solapados-exclusivos, secuenciales-simultáneos, monotéticos-politéticos, directos-iterativos, ponderados-no ponderados y métodos adaptativos-no adaptativos.

A continuación centraremos la explicación en el análisis de *cluster* jerárquico y aglomerativo.

Definición de la técnica: *cluster* jerárquico y aglomerativo

Esta técnica parte de la premisa de que todo fenómeno debe ser ordenado para ser entendible. Sokal y Sneath (1963) son los autores que más han influido en el desarrollo de esta técnica todavía vigente. Se trata de un conjunto de técnicas (fundamentalmente algoritmos) y métodos estadísticos multivariantes de clasificación automática de datos. Concretamente, a partir de una tabla se trata de situar todos los casos en grupos homogéneos (conglomerados o *clusters*) no conocidos previamente pero que los propios datos sugieren (los individuos parecidos serán asignados a un mismo conglomerado, mientras que los diferentes estarán en conglomerados distintos). Es una técnica que se puede utilizar para agrupar individuos o variables. El procedimiento e interpretación de los resultados siguen el mismo patrón. La única diferencia es el contexto donde se ubica la clasificación.

Es habitual utilizar otras técnicas que complementan o confirman la formación de grupos homogéneos de conglomerados, a saber, el análisis discriminante (técnica *ad hoc*) permite verificar la existencia de relaciones causales entre la pertenencia a un conglomerado determinado y los valores de las variables (Sneath y Sokal, 1973). Los resultados de un *cluster* pueden ampliarse también con regresión logística, añadiendo siempre nuevas variables independientes y constituyendo los conglomerados agrupando individuos y no solo variables.

Los principios básicos del análisis de conglomerados recogidos por Pérez (2009) son:

- Es un método estadístico multivariante de clasificación automática de datos.
- Tiene la finalidad de revelar concentraciones en los datos para un agrupamiento eficiente en conglomerados según su homogeneidad.
- El agrupamiento puede ser para casos o variables (cualitativas o cuantitativas).
- Es esencial un uso adecuado del concepto de distancia, ya que los grupos se realizan según la proximidad o lejanía de unos con otros.
- Es fundamental que dentro de un conglomerado los elementos sean homogéneos, y lo más diferentes a los contenidos del resto.
- Es una técnica de clasificación *post hoc*: el número de conglomerados se determina en función de los datos, y puede no ser definido de antemano.



2. Condiciones de aplicación del análisis de conglomerados

El análisis de conglomerados suele iniciarse estimando las similitudes entre los individuos mediante la correlación de las distintas variables (cualitativas o cuantitativas). A continuación se procede a comparar los grupos según las similitudes, y finalmente se decide cuántos grupos se construyen. El objetivo será formar el mínimo número de grupos posible, lo más homogéneos dentro de cada grupo, y lo más heterogéneos entre grupos.

De entrada, para aplicar esta técnica no se debe cumplir ningún supuesto paramétrico, sin embargo sí se deben considerar una serie de axiomas previos:

- Si las variables están en escalas muy diferentes será necesario *estandarizar las variables* (o trabajar con las desviaciones respecto de la media). También puede hacerse un análisis factorial previo y trabajar con puntuaciones factoriales.
- Observar *valores perdidos y atípicos*, ya que los valores atípicos deforman las distancias y producen *clusters* unitarios.
- Análisis previo de *multicolinealidad*, ya que las variables correlacionadas son nocivas para el análisis de conglomerados.
- El número de observaciones en cada conglomerado debe ser relevante, ya que pueden haber valores atípicos que difuminen las agrupaciones.
- El resultado debe tener sentido conceptual.
- Se pueden realizar otros análisis de forma complementaria: discriminante, regresión logística, etc.

3. Procedimiento

Como venimos diciendo, el objetivo del análisis de conglomerados consiste en conseguir una o más particiones de un conjunto de individuos a partir de determinadas características de los constituyentes. Se podrá decir que los individuos son similares si pertenecen a la misma clase, grupo o conglomerado de tal forma que todos los individuos que están en el mismo conglomerado se parecen entre sí, y son diferentes a los individuos de otro conglomerado. Los miembros de un conglomerado tienen características genéricas comunes, que difícilmente pueden resumirse en una única variable. A partir de la identificación de grupos o *clusters* se podrá trabajar de forma segmentada sin necesidad de seguir trabajando con toda la muestra.

El procedimiento básico es el siguiente:

- Selección de los individuos objeto de estudio.** Se debe prestar especial atención a los casos atípicos que pueden distorsionar la agrupación de grupos homogéneos.



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

b. Selección de las variables o datos que describen y caracterizan la muestra. Se deben incluir transformaciones a partir de las variables que los definen (tipificación de variables, etc.). La selección de las variables relevantes a los objetivos del estudio es decisiva para que los resultados tengan sentido. La elección de las variables debe ser coherente con la investigación que se lleva a cabo. La inclusión de variables no relevantes aumenta la posibilidad de tener casos atípicos. De entrada, pues, optaremos por un mismo tipo de variables; en caso de no ser así se deberá proceder a su estandarización.

c. Elección de la medida de proximidad entre los individuos. El conocimiento de las distancias ayudará a interpretar las agrupaciones resultantes y a determinar cuál es el punto de corte más adecuado. Las medidas de similitud/distancia definen la proximidad y no covariación, y vienen determinada por la escala de medida de las variables (ordinal o de intervalo-razón). El resultado de la técnica depende de la medida de asociación-similitud-distancia utilizada; así distintas medidas de proximidad pueden ofrecer resultados distintos. Con variables dicotómicas deberá utilizarse alguna de las distancias que SPSS integra dentro de la opción binaria; con variables cualitativas, se tendrá que seleccionar la media dentro de la opción *frecuencias*; y para variables cuantitativas, SPSS ofrece la opción *intervalo*.

<i>Variables cuantitativas</i>	<i>Variables cualitativas</i>	<i>Datos dicotómicos</i>
- Distancia euclídea	- Chi cuadrado	- Jaccard
- Distancia euclídea al cuadrado	- Phi cuadrado	- Russel y Rao
- Coseno de vectores		- Sokal y Sneath
- Correlación de Pearson		- Rogers y Tanimoto
- Distancia métrica de Chebynev		
- Bloque, Manhattan o City-block		
- Distancia de Minkowski		

Figura 1. Clasificación de las principales medidas de distancia, según el tipo de variable (Visauta, 2002:282).

d. Elección del criterio para agrupar los individuos en conglomerados y la ejecución del algoritmo. No existe un criterio único para seleccionar el mejor algoritmo y, por tanto, la decisión es subjetiva en función del objetivo pretendido. Se puede obtener una agrupación progresiva (jerárquico). Dentro de esta tipología, podemos identificar diferentes métodos. De entre todos el más utilizado es el método promedio entre grupos, el resto de métodos requieren trabajar con la distancia euclídea D2 como criterio.

Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

Tipo de métodos	Descripción del método
Distancia mínima o vecino más próximo	Los grupos se unen considerando la menor de las distancias existentes entre los miembros más cercanos de distintos grupos. Crea grupos más homogéneos. Ayuda a detectar <i>outliers</i> , pero no es útil para resumir datos. Los <i>clusters</i> son demasiado grandes. Es el método más sencillo.
Distancia máxima o vecino más lejano	Los grupos se unen considerando la menor de las distancias existentes entre los miembros más lejanos de distintos grupos. Los grupos resultantes son más heterogéneos. Es útil para detectar <i>outliers</i> y los <i>clusters</i> son pequeños y compactos.
Media o promedio entre grupos	La distancia entre los grupos se obtiene calculando la distancia promedio entre todos los pares de observaciones independientemente de que estén próximos o alejados. Agrupa los conglomerados con un tamaño óptimo y fusiona <i>cluster</i> con varianzas pequeñas. Es uno de los métodos más utilizados.
Vinculación intra grupos	Es una variante de la anterior aunque en este caso se combinan los grupos buscando que la distancia promedio dentro de cada conglomerado sea la menor posible.
Ward o Método de varianza mínima	La distancia entre dos <i>clusters</i> se calcula como la suma de cuadrados entre grupos en el ANOVA. Se persigue la minimización de la varianza intragrupal y maximiza la homogeneidad dentro de los grupos. Suele ser muy adecuado aunque los <i>clusters</i> que genera suelen ser pequeños y muy compactos. Es especialmente sensible a los <i>outliers</i> .
Centroide	La distancia entre dos <i>clusters</i> se calcula como la distancia entre sus centroides. Se trabaja con los valores originales. Las variables deben ser cuantitativas de intervalo. Este método es sensible si los tamaños de los grupos son muy diferentes
Agrupación de medianas	Variante del método anterior en la que no se considera el número de casos que forman cada uno de los agrupamientos, sino solo el número de conglomerados.

Figura 2. Clasificación de los métodos de agrupamiento.

e. **Identificar las agrupaciones resultantes.** Mediante una representación gráfica de los conglomerados obtenidos para visualizar los resultados, mediante un dendograma o árbol lógico. Este gráfico resume el proceso de agrupación: en el eje de abscisas se sitúan los sujetos y en el eje de ordenadas aparecen las distancias utilizadas para agrupar *clusters*. Los sujetos o variables similares se conectan mediante enlaces a partir del método aglomerativo. La posición del enlace determina el nivel de similitud entre los objetos.

f. **Interpretación de los resultados obtenidos.** La decisión e interpretación final sobre el número adecuado de agrupaciones es totalmente subjetiva. En la decisión final tiene que primar un equilibrio entre un número reducido de *clusters*, con lo que obtendremos agrupaciones heterogéneas, y un número excesivo de agrupaciones con la dificultad de



interpretación que lleva asociada. Generalmente se asocia la distancia óptima como punto de corte cuando en el nivel de agrupación se producen saltos bruscos.

A partir de la utilización del SPSS como recurso informático, obtenemos dos tipos de información, una numérica y, por lo tanto, la cuantificación de las distancias entre grupos; y otra gráfica que permite visualmente e intuitivamente identificar los grupos o conglomerados que se han formado.

4. Análisis de *cluster* jerárquico mediante SPSS

Siguiendo con el ejemplo anterior, si nuestro objetivo es conocer las características que identifican las escuelas públicas de una comarca catalana que agrupan a los estudiantes con un mayor éxito académico desarrollaremos un análisis de *clusters*¹ que nos permitirá dar respuesta a los interrogantes: ¿Las escuelas públicas se pueden agrupar? y ¿cuáles son las características comunes de las escuelas agrupadas?

Para ello, partimos de una base de datos de 30 escuelas públicas con las siguientes variables: líneas de la escuela, proyectos de innovación del centro, porcentaje de plantilla fija de profesorado, la media de notas del alumnado en el curso 2012-13, porcentaje de abandono escolar del alumnado y de tránsito a la postobligatoria, y el número de ordenadores de los que dispone el centro.

Considerando el tamaño reducido de la muestra y el objetivo del estudio que no dispone de un número determinado de agrupaciones, el procedimiento más adecuado es un *cluster* de tipo jerárquico, con procedimiento aglomerativo. Concretamente, nos interesa clasificar los casos (las escuelas) en *clusters*. A continuación proponemos los siguientes pasos:

Paso 1. Revisión de las condiciones de aplicación de la técnica

Dado que las variables están en escalas muy diferentes será necesario *estandarizar las variables*. Para ello, con SPSS se ejecuta el procedimiento *descriptivo* marcando la opción de *variables tipificadas* y se crean las nuevas variables en la matriz que se utilizarán en el estudio (figura 3).

¹ Los datos utilizados en este caso son ficticios y responden a una finalidad pedagógica.



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

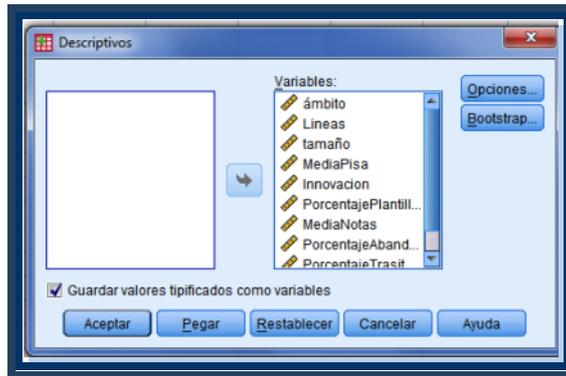


Figura 3. Tipificación de las variables seleccionadas para el estudio.

Paso 2. El procedimiento *cluster*

Para realizar un *análisis cluster jerárquico* elegiremos el menú *analizar*, opción *clasificar conglomerados jerárquicos* (figura 4), y seleccionaremos las variables relevantes en nuestra investigación.

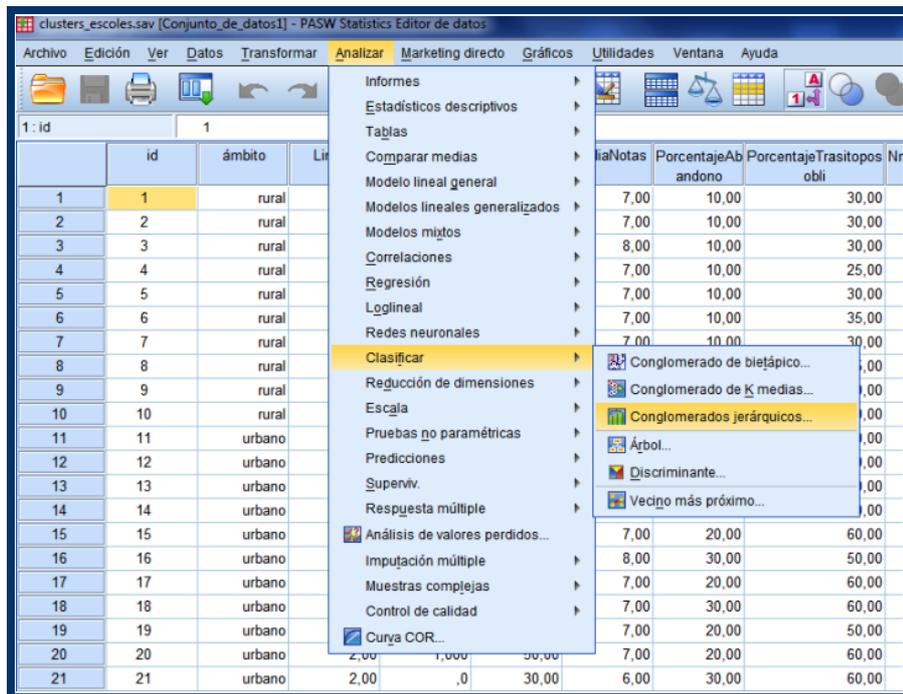


Figura 4. Cuadro de diálogo *Análisis de cluster jerárquico* con SPSS.

En el cuadro de diálogo de selección de las variables, encontramos las siguientes especificaciones:

- *Etiquetar los casos.* Si se desea el etiquetaje de los casos es necesario que el identificador esté definido como cadena en la matriz de datos. En nuestro ejemplo,



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

se ha añadido una variable tipo *cadena* con la denominación de escuela 1, escuela 2, etc.

- *Conglomerar casos o variables.* En nuestro ejemplo, conglomeramos las escuelas (casos).
- *Visualización de estadísticos y/o gráficos.* Los marcamos.

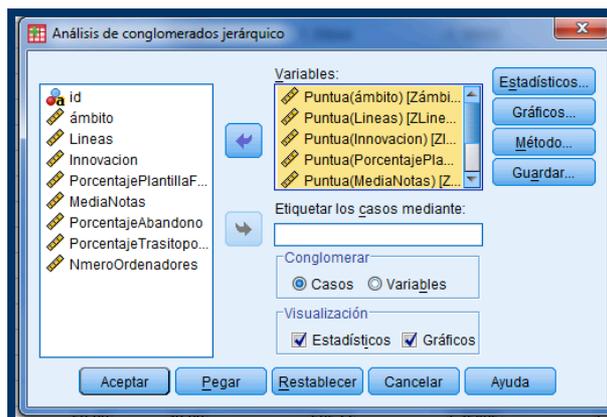


Figura 5. Cuadro de diálogo de selección de las variables.

En el mismo cuadro de diálogo de la figura 5, encontramos los siguientes cuatro botones:

1. *Estadísticos.* En este botón, dejamos por defecto la opción *historial de conglomeración*, que muestra los casos combinados en cada etapa, las distancias entre los casos y el último nivel del proceso en el que cada caso se unió a su conglomerado correspondiente. Otra opción es pedir la *matriz de distancias*, para mostrar la distancia de cada caso con el resto (figura 6).

Una opción interesante es el *conglomerado de pertenencia* cuando conviene especificar un número concreto de conglomerados o *clusters*, o bien, un rango específico. Si en nuestro caso necesitáramos, por ejemplo, establecer cuatro grupos de escuelas, marcaríamos la opción *solución única*; y si propusiéramos entre tres y cinco *clusters*, marcaríamos *rango de soluciones*.



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

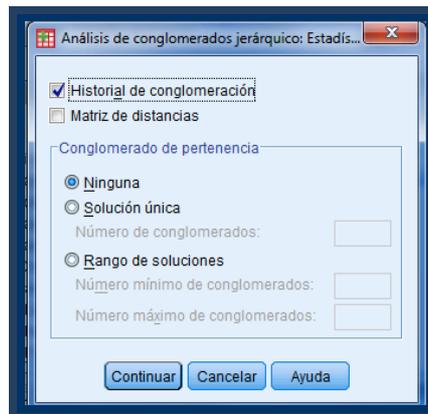


Figura 6. Cuadro de diálogo de la opción Estadísticos.

2. *Gráficos.* Es recomendable señalar la opción dendograma, que ofrece un gráfico muy visual de los pasos del *cluster* jerárquico, ofreciendo unas líneas verticales concertadas que designan combinaciones de casos. Por defecto, SPSS ofrece otra representación gráfica, el diagrama de témpanos al cual se le puede especificar el número de conglomerados y la orientación horizontal o vertical del gráfico (figura 7).

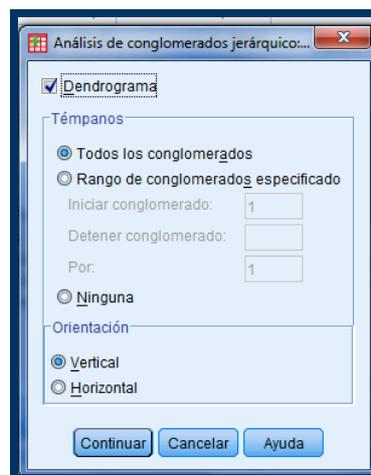


Figura 7. Cuadro de diálogo de la opción Gráficos.

3. *Método.* En primer lugar es importante elegir el tipo de distancia a considerar según la escala métrica de las variables (denominada *medida* en el cuadro de diálogo). En nuestro ejemplo, dada la escala de intervalo y razón de las variables incluidas, elegimos la *distancia euclídea al cuadrado* dentro de la opción *intervalo* del cuadro de diálogo. A continuación, elegimos el *método de conglomeración*, a elegir entre los que se proponen en la figura 2 vista anteriormente. En nuestro ejemplo, seleccionamos el *método de Ward*, dado que interesa minimizar la varianza intragrupal y maximizar la homogeneidad dentro de los grupos. Puesto que este método es muy sensible a los *outliers* (valores atípicos extremos), primero verificamos su inexistencia.



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

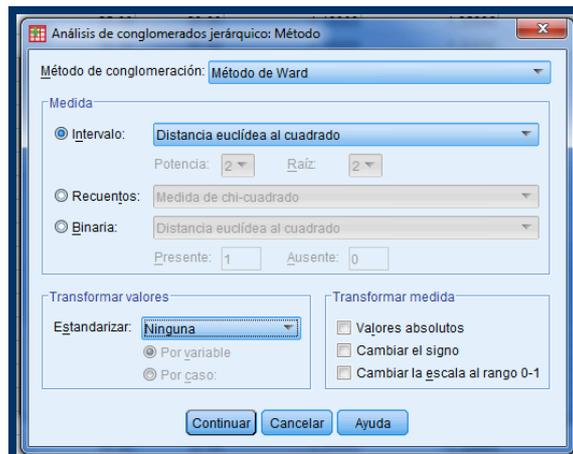


Figura 8. Cuadro de diálogo de la opción Método.

En el cuadro de diálogo de la figura 8, también se ofrece la opción de transformar los valores estandarizándolos. En el ejemplo se ha dejado por defecto *ninguna estandarización* dado que se utilizó anteriormente la opción *guardar valores tipificados como variables* (resumido en la figura 3).

4. *Guardar*. Finalmente seleccionaremos la opción *Guardar*, la cual crea una nueva variable que recoge el conglomerado al que pertenece cada caso, útil para hacer futuros análisis, ente ellos la descripción de los conglomerados, necesaria para interpretar los *clusters* resultantes. Desde la opción *Guardar* se ha de decidir entre un número de *clusters* concreto (*solución única*) o un *rango de soluciones*, con el que crear esta nueva variable que identifica los casos.

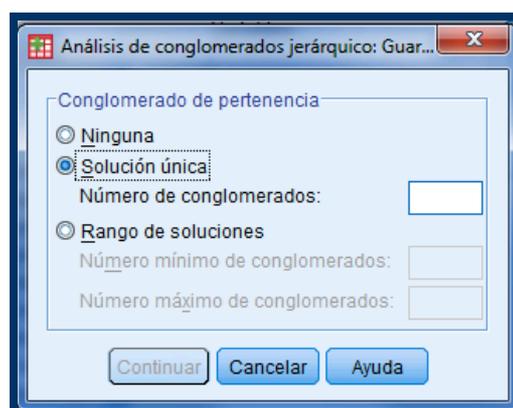


Figura 9. Cuadro de diálogo de la opción Guardar.

Paso 3. La interpretación de los *outputs*

El historial de conglomeración es la tabla principal de los *outputs* que detallan el análisis de *clusters* (figura 10). En la primera columna se enumeran las etapas del análisis, las columnas segunda y tercera indican los conglomerados unidos en cada etapa y, a continuación, la distancia en la que se produce el agrupamiento. La quinta y la sexta columna indican la



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

etapa en la que el conglomerado aparece por primera vez. Y finalmente, la última columna indica en qué etapa volverá a ser utilizado el agrupamiento que se acaba de formar.

Etapa	Conglomerado que se combina		Coeficientes	Etapa en la que el conglomerado aparece por primera vez		Próxima etapa
	Conglomerado 1	Conglomerado 2		Conglomerado 1	Conglomerado 2	
1	29	30	,000	0	0	2
2	26	29	,000	0	1	12
3	15	20	,000	0	0	13
4	12	18	,000	0	0	19
5	1	7	,000	0	0	17
6	2	6	,070	0	0	17
7	3	8	,140	0	0	20
8	23	28	,421	0	0	18
9	21	22	,701	0	0	26
10	5	10	1,032	0	0	24
11	11	17	1,366	0	0	19
12	25	26	1,787	0	2	18
13	14	15	2,233	0	3	22
14	4	9	2,747	0	0	20
15	13	19	3,358	0	0	21
16	24	27	4,122	0	0	23
17	1	2	5,045	5	6	24
18	23	25	5,973	8	12	23
19	11	12	7,028	11	4	27
20	3	4	8,117	7	14	25
21	13	16	9,344	15	0	22
22	13	14	11,635	21	13	27
23	23	24	14,326	18	16	26
24	1	5	17,221	17	10	25
25	1	3	21,963	24	20	29
26	21	23	28,039	9	23	28
27	11	13	34,727	19	22	28
28	11	21	88,564	27	26	29
29	1	11	203,000	25	28	0

Figura 10. Output Historial de conglomeración.

Se recomienda el análisis de los datos obtenidos a través de la representación gráfica del dendograma (figura 11). El dendograma está compuesto por una primera columna que identifica los casos (las escuelas en el ejemplo), y una primera fila que identifica las distancias.

En el ejemplo, hay agrupaciones muy tempranas en la primera distancia entre, por ejemplo, las escuelas 29, 30, 26, 25, 23, 28, 24 y 27; por este motivo la línea que las une está situada muy a la izquierda del gráfico. Todos estos agrupamientos están indicando una escasa distancia entre estas escuelas. El grupo anterior formado juntamente con las escuelas



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

21 y 22 se une en la segunda distancia, lo que indica que la agrupación se produce a una distancia mayor a la anterior.

Cuando las líneas de unión se sitúan más a la derecha, más distancia existe en las uniones. Por ejemplo, en la distancia doce, las escuelas anteriores se fusionan con otro grupo formado por varios niveles de agrupación previos (escuelas 12, 18, 11, 17, etc.). Finalmente se fusionan todas las escuelas en un conglomerado único en la distancia 25.

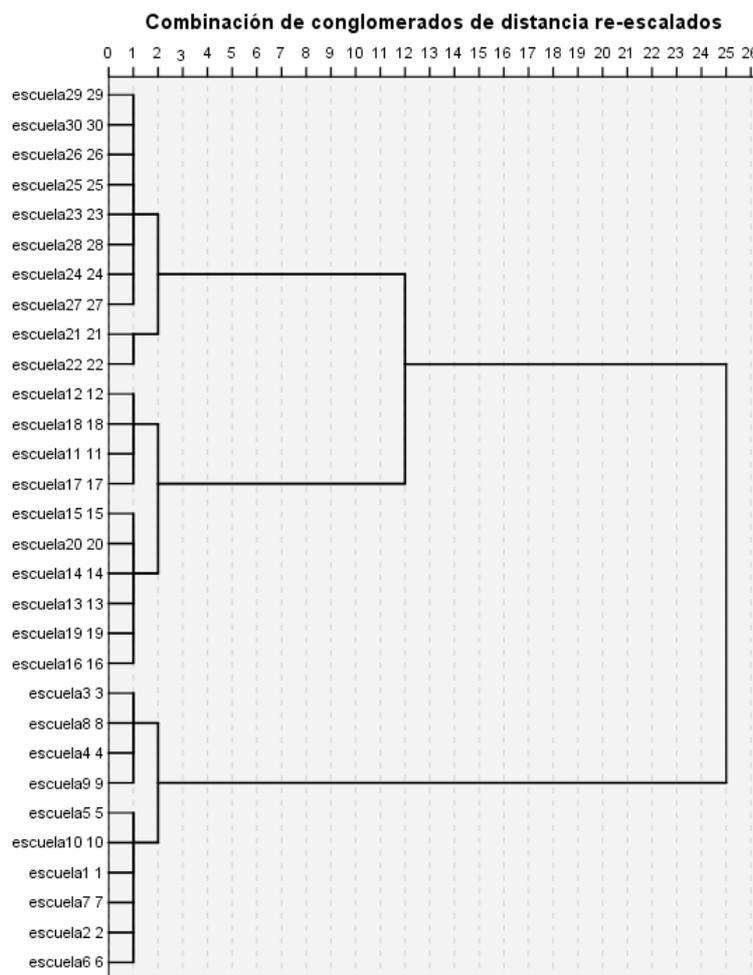


Figura 11. Output dendograma que utiliza una vinculación de Ward.

Para interpretar estos resultados se recuerda que el objetivo es agrupar las escuelas considerando sus características para el éxito académico. Por lo tanto, tener un solo grupo no aporta información. Sabiendo que a menor distancia los conglomerados son más homogéneos, es conveniente detener el proceso de unión cuando las líneas horizontales sean muy largas: en el ejemplo se detiene el proceso en la distancia 12, tal como se dibuja en la figura 11 (líneas discontinuas paralelas); obteniendo tres *clusters* de 10 escuelas cada uno.



Ruth Vilà, María-José Rubio, Vanesa Berlanga, Mercedes Torrado. *Cómo aplicar un cluster jerárquico en SPSS.*

Un análisis en profundidad de las características de cada grupo de escuelas a partir de la variable creada con la opción *Guardar* (figura 5) ofrece más información sobre la naturaleza del agrupamiento. Por ejemplo, la opción *explorar* del SPSS puede ser muy útil para obtener una descripción básica de los conglomerados resultantes.

Tal como se resume en la figura 12:

- El *primer cluster* se caracteriza por agrupar las escuelas más pequeñas de una sola línea, que poseen profesorado estable, unas notas altas en el alumnado, un bajo porcentaje de abandono y baja transición del alumnado a la postobligatoria.
- El *segundo cluster* incluye centros un poco mayores de dos líneas, con menor proporción de profesorado estable, altas notas del alumnado y alta proporción de transición a la postobligatoria.
- Y finalmente, el *tercer cluster* incluye escuelas con un mayor número de líneas, con menos proyectos de innovación y menor equipación informática, profesorado inestable, notas más bajas en el alumnado y mayor índice de abandono.

Variables del análisis/ descriptivos		Conglomerados		
		1	2	3
Líneas	Media	1,00	2,00	2,80
	Desviación típica	,00	,00	,42
	Máximo	1,00	2,00	3,00
	Mínimo	1,00	2,00	2,00
Innovación	Media	1,400	1,400	,000
	Desviación típica	,516	,516	,000
	Máximo	2,000	2,000	,000
	Mínimo	1,000	1,000	,000
Porcentaje Plantilla Fija	Media	61,00	48,00	33,00
	Desviación típica	3,16	4,22	4,83
	Máximo	70,00	50,00	40,00
	Mínimo	60,00	40,00	30,00
Media Notas	Media	7,20	7,10	5,10
	Desviación típica	,42	,32	,57
	Máximo	8,00	8,00	6,00
	Mínimo	7,00	7,00	4,00
Porcentaje Abandono	Media	10,00	25,00	38,00
	Desviación típica	,00	5,27	4,22
	Máximo	10,00	30,00	40,00
	Mínimo	10,00	20,00	30,00
PorcentajeTránsitopostob	Media	29,50	58,00	49,00
	Desviación típica	2,84	4,22	8,76
	Máximo	35,00	60,00	60,00
	Mínimo	25,00	50,00	40,00
NúmeroOrdenadores	Media	20,50	29,50	30,00
	Desviación típica	3,69	3,69	,00
	Máximo	25,00	35,00	30,00
	Mínimo	15,00	25,00	30,00

Figura 12. Output de las características de cada conglomerado.



<Referencias bibliográficas>

- Baró, J. y Alemany, R. (2000). *Estadística II*. Ed. Fundació per a la Universitat Oberta de Catalunya. Barcelona.
- Bisquerra, R. (1989). *Introducción conceptual al Análisis Multivariable. Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD (Vol II)*. Barcelona: PPU.
- Peña Sánchez de Rivera, D. (1987). *Estadística. Modelos y Métodos*. Volumen 2. Alianza Editorial. Madrid.
- Pérez, C. (2009). *Técnicas estadísticas multivariantes con SPSS*. Madrid: Garceta grupo editorial.
- Pérez, C. (2008). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid: Pearson Prentice Hall.
- Sneath, P.H.A. y Sokal, R.R. (1973). *Numerical taxonomy*. San Francisco: Freeman.
- Sokal, R.R. y Sneath, P.H.A. (1963). *Principles of numerical taxonomy*. San Francisco: Freeman.
- Torrado, M. y Berlanga, V. (2013). Análisis Discriminante mediante SPSS. [En línea] *REIRE, Revista d'Innovació i Recerca en Educació*, 6 (2), 150-166.
- Visauta, B. (2002). *Técnicas de análisis multivariante para investigación social y comercial*. Madrid: Ra-Ma.

Copyright © 2014. 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.

