



<Artículo>

Modelos lineales multinivel en SPSS y su aplicación en investigación educativa

Tatiana Iñiguez-Berrozpe, Francesco Marcaletti

Enviado: 02/05/2017
 Aceptado: 20/06/2017
 Publicado en prensa: 15/09/2017
 Publicado: 08/01/2018

//Resumen

En el presente artículo abordamos el análisis de datos que se encuentran agrupados o anidados en unidades más amplias, como ocurre en muchas ocasiones en el ámbito educativo. Para ello, mediante la base de datos que proporciona la prueba PISA (OCDE), donde el alumnado (nivel 1) se agrupa en centros (nivel 2), se ejemplifica el análisis que más se adecúa a este tipo de datos estructurados siguiendo una jerarquía, los modelos lineales multinivel (MLM). A través del programa SPSS se lleva a cabo el MLM que permite, teniendo en cuenta dichos agrupamientos, analizar la influencia de las variables independientes de nivel 1 (personales y familiares) y las de nivel 2 (todas las relacionadas con el centro) en los resultados relacionados con el rendimiento del alumnado en la Prueba PISA 2015 (variable dependiente). En el ejemplo presentado, se demuestra no solo que las variables tanto de nivel 1 como de 2 tenían un efecto relevante en el resultado del alumnado, sino también que el análisis mediante el MLM permite tener en cuenta los efectos inter-niveles, en este caso el agrupamiento en centros, en el resultado.

//Palabras clave

Modelos Lineales Multinivel; Modelos Jerárquicos Multinivel; PISA; Análisis Cuantitativo; SPSS.

//Referencia recomendada

Iñiguez-Berrozpe, T., y Marcaletti, F. (2018). Modelos lineales multinivel en SPSS y su aplicación en investigación educativa. *REIRE Revista d'Innovació i Recerca en Educació*, 11(1), 26-40. <http://doi.org/10.1344/reire2018.11.118984>

//Datos de las autoras

Tatiana Iñiguez-Berrozpe. Departamento de Psicología y Sociología. Facultad de Educación, Universidad de Zaragoza. tatianai@unizar.es, ORCID: <http://orcid.org/0000-0003-4530-9645>

Francesco Marcaletti. Instituto de Estudios Superiores de la Familia Universitat Internacional de Catalunya. fmarcaletti@uic.es, ORCID: <http://orcid.org/0000-0003-3273-5311>

//Títol

Models lineals multinivell en SPSS i la seva aplicació en investigació educativa

//Resum

En aquest article abordem l'anàlisi de dades que es troben agrupades o niades en unitats més àmplies, com passa molt sovint en l'àmbit educatiu. Per fer-ho, mitjançant la base de dades que proporciona la prova PISA (OCDE), en què l'alumnat (nivell 1) s'agrupa en centres (nivell 2), s'exemplifica l'anàlisi que més s'adequa a aquest tipus de dades estructurades seguint una jerarquia, els models lineals multinivell (MLM). A través del programa SPSS es porta a terme l'MLM que permet, tenint en compte aquests agrupaments, analitzar la influència de les variables independents de nivell 1 (personals i familiars) i les de nivell 2 (totes les relacionades amb el centre) en els resultats relacionats amb el rendiment de l'alumnat en la prova PISA 2015 (variable dependent). En l'exemple presentat, es demostra no només que les variables tant de nivell 1 com de nivell 2 tenen un efecte rellevant en el resultat de l'alumnat, sinó que, a més, l'anàlisi mitjançant l'MLM permet tenir en compte els efectes internivells, en aquest cas l'agrupament en centres, en el resultat.

//Paraules clau

Models lineals multinivell; Models jeràrquics multinivell; PISA; Anàlisi quantitativa; SPSS.

//Title

Multilevel linear models in SPSS in educational research

//Abstract

In this article, we analyse data that are grouped or nested in larger units, as is often the case in education. To do this, we use the database that the PISA test (OECD) provides, in which students (level 1) are grouped in centres (level 2). The aim is to illustrate the analysis that best suits this type of hierarchy structured data following multilevel linear models (MLM). MLM are determined using the SPSS program and considering the groupings. Thus, we analyse how the independent variables of level 1 (personal and familiar) and those of level 2 (all related to the centre) influence results related to students' performance in the PISA 2015 test (dependent variable). In the example, we demonstrate not only that the variables of both levels 1 and 2 had a considerable impact on students' results, but also that the analysis through MLM considers inter-level effects, in this case grouping in centres, in the result.

//Keywords

Multilevel linear models; multilevel hierarchical models; PISA; quantitative analysis; SPSS.

1. Introducción

Dentro de la investigación educativa, el rendimiento académico del alumnado es uno de los indicadores que más se utilizan para evaluar la eficiencia de los sistemas educativos. Para estandarizar la evaluación de dicho rendimiento, el mecanismo más empleado es la prueba PISA (OCDE, 2016), que se realiza cada 3 años al alumnado de 15 años de los países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE). Mediante esa prueba se evalúa con mayor profundidad la competencia científica, matemática o lectora en cada una de sus ediciones. La última prueba PISA se llevó a cabo en 2015, y se centró en la competencia científica. Las dos encuestas principales se realizan a alumnado (incluida la evaluación de competencias, además de otras características sociopersonales y familiares) y a personal de dirección de centro (que debe responder a cuestiones relacionadas sobre las características del mismo). En esta última edición, los microdatos proporcionados por la OCDE ofrecen información sobre $n = 6736$ alumnos/as de 15 años, y $n = 201$ escuelas para el caso español.¹

Los alumnos y alumnas que participan en la prueba no pueden ser considerados sujetos independientes (atendiendo solo a sus distintos condicionantes sociopersonales y familiares), sino que se encuentran agrupados en centros, que, debido a sus diversas características y formas de gestión, pueden condicionar de manera distinta su rendimiento académico. Esta particularidad de los datos provoca que los análisis cuantitativos “tradicionales” (modelos de regresión logística binaria, modelos de regresión multivariante...) no sean los más apropiados para este tipo de estudios, ya que obvian este agrupamiento de las unidades muestrales. Para ello, el tipo de análisis que más se ajusta a los datos anidados o jerárquicos es el basado en modelos lineales multinivel (en adelante MLM), también llamados modelos jerárquicos lineales o modelos mixtos lineales. Aunque existen diferentes tipos de modelos y programas informáticos para su análisis, el presente artículo se centra en el que más se ajusta al ámbito educativo, tomando como ejemplo los datos del último informe PISA publicado (OCDE, 2016) y el programa de análisis estadístico SPSS. Su propósito no es entrar en profundas disquisiciones metodológicas, sino ser lo más claro y didáctico posible para guiar a la persona investigadora en el análisis.

2. Justificación del empleo de MLM en investigación educativa

En la mayoría de los tipos de análisis estadístico los datos se tratan como si estuvieran organizados en un mismo nivel. Sin embargo, en muchas ocasiones, los datos están estructurados de manera jerárquica (ver Fig. 1). En otras palabras, que algunas de las variables de análisis están anidadas o agrupadas en otras de mayor nivel. El ejemplo más característico dentro de la investigación educativa es la distinción entre el nivel 1, que corresponde al grupo de estudiantes, y el nivel 2, que puede corresponder al aula o al centro donde se agrupa dicho alumnado (Kreft y DeLeeuw, 1994). Si, como ocurre en la evaluación mediante la prueba PISA (OCDE, 2016), se evalúa el rendimiento del alumnado respecto a una competencia instrumental específica, y se tienen en cuenta los elementos contextuales tanto del propio chico o chica (personales, familiares, sociales), como los del centro (titularidad, autonomía, forma de agrupamiento del alumnado...), se está hablando de datos jerárquicos de nivel 1 y 2

¹ Se ha tomado como referencia el indicador “Spain” para simplificar el ejemplo, ya que en estos microdatos también se encuentra el indicador “Spain: regions”, con un mayor número de estudiantes y centros.

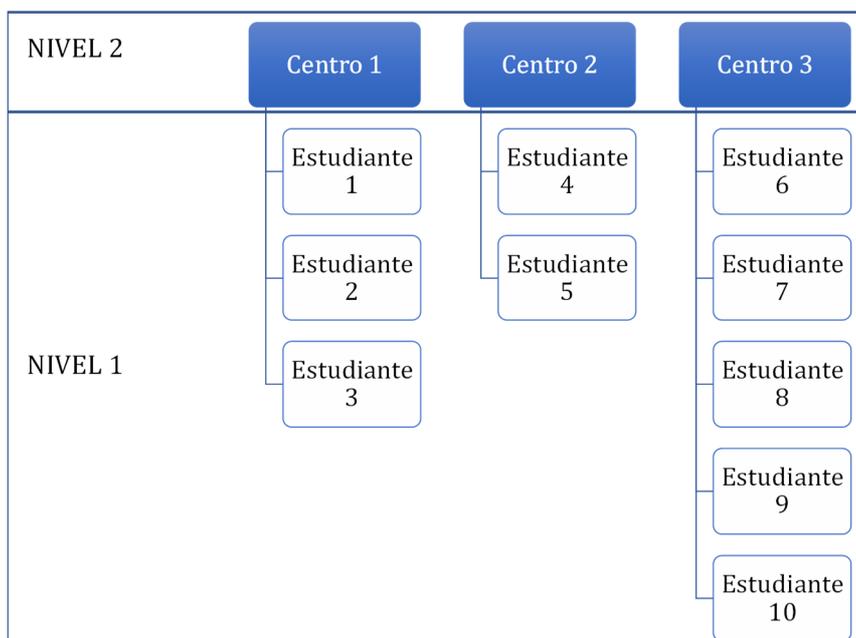


respectivamente. Esta condición se cumple en el caso de PISA, ya que en la realización de la prueba se escoge aleatoriamente a los y las estudiantes, pero previamente se han seleccionado los centros a los que pertenecen.

Tradicionalmente, los datos del informe PISA se analizaban mediante análisis basados en regresiones lineales múltiples o, en menor medida, regresiones logísticas, y se obviaban los efectos interniveles (Blanco Blanco, López Martín, Ruiz de Miguel, 2014). Esto provocaba que se ignoraran variables de contexto que pueden ser significativas para la evaluación del rendimiento del alumnado, y no se tenía en cuenta que estos datos no cumplen la condición de independencia, ya que, normalmente, cuando se trabaja con datos agrupados, las observaciones en cada grupo pueden presentar características similares.

A diferencia de estos, los MLM, al tener en cuenta tanto las variables asociadas a cada individuo como las contextuales que afectan al grupo, permiten comprender la varianza que produce cada nivel de agregación. De esta forma, puede analizarse qué efecto tiene la situación individual de cada alumno o alumna en su rendimiento académico y luego sumar el efecto de variables contextuales dependientes del centro, ya que los datos van introduciéndose en niveles. Así, los MLM permiten establecer la medida en que las correlaciones dependen del nivel individual o del nivel grupal.

Figura 1
Ejemplo de una estructura de datos jerárquica en dos niveles: estudiantes (nivel 1) agrupados en centros (nivel 2)



Fuente: Elaboración propia.

Además de su adecuación a este tipo de datos, los MLM tienen otras ventajas importantes a la hora de ser empleados para el análisis estadístico: por un lado, no es necesario tener en cuenta las asunciones de homogeneidad de las pendientes de regresión, como ocurre en otros análisis de covarianza; a diferencia de las regresiones múltiples, los datos no deben cumplir la asunción

de independencia; los datos perdidos no suponen un problema para los MLM, de nuevo, a diferencia de los distintos tipos de regresiones, test ANOVA, ANCOVA, etc. (Field, 2013).

Finalmente, aunque los datos jerárquicos se relacionan habitualmente con el análisis de los datos PISA y en el presente artículo van a emplearse como caso de ejemplo, esta situación se da de manera muy habitual en un gran número de investigaciones educativas (normalmente, en las distintas investigaciones educativas, los cuestionarios que realizan profesorado y alumnado se distribuyen y recogen por centro), por lo que los MLM pueden extenderse mucho más allá del análisis de PISA.

3. Condiciones de aplicación del análisis mediante MLM

Como acaba de verse, los MLM permiten análisis que otros modelos de regresión impiden, ya que se puede trabajar con muestras cuyos datos no son independientes. No obstante, al ser una extensión de una regresión múltiple, las premisas que se asumen deben ser, en general, similares a este tipo de análisis. Esto es, siguiendo a Field (2013): una variable dependiente de tipo cuantitativa continua; varias variables predictoras que deben ser cuantitativas o categóricas; las varianzas de los predictores no deben ser cero; debe existir una multicolinealidad no perfecta entre dos o más predictores; los predictores no deben estar correlacionados con variables externas (no introducidas en el modelo).

Específicamente, dada la naturaleza de los MLM, los datos que hay que analizar seguirán una estructura jerárquica similar a la expresada en la Figura 1. Esta jerarquía puede ser simple (de dos niveles, como la que se va a emplear en el presente trabajo) o más compleja (de 3 o más niveles). Lo importante es que la persona investigadora conozca la estructura de dicha jerarquía, ya que a la hora de introducir los datos en SPSS (o cualquier otro *software* empleado para estos tipos de modelos), es esencial conocer el tipo de agrupamiento de los datos para que el análisis sea correcto (en este caso, alumnado agrupado en centros). Esto se debe a que las variables contextuales en la jerarquía especificada provocan dependencia en los datos. En otras palabras, los residuos correlacionarán, ya que los datos no son independientes (como se asume en otros análisis cuantitativos), lo cual se denomina "correlación intraclase" (CIC) (Field, 2013, p. 729). En este ejemplo de PISA con alumnado y centros, esta CIC representa la proporción de variabilidad total que puede atribuirse a la pertenencia a un centro.

Por otro lado, hay que tener en cuenta que el proceso de modelización puede realizarse de distintas formas, ya que pueden escogerse distintas estructuras de covarianza, y se tiene la posibilidad de que los parámetros no sean fijos o constantes, sino que puedan variar. Es decir, hay que diferenciar si los efectos de las variables predictoras son fijos en los distintos niveles de análisis, además de la existencia de un término aleatorio, que recoge las diferencias intergrupos (Andreu Abela, 2011), tal y como se especifica en el siguiente apartado.

Por último, para el análisis de los datos jerárquicos empleando MLM se precisa también de un *software* específico. Algunos de los programas que sirven para este tipo de modelización son: HLM: Hierarchical linear and nonlinear modeling (HLM); SAS (PROC MIXED function); MLwiN; aML; lme4; R. Sin embargo, en este caso, se apela a la familiaridad con el programa y se emplea



el paquete de IBM SPSS Statistics (versión 22, aunque esta función se incluye desde la versión 15).

4. Proceso de modelización multinivel

Para explicar la modelización multinivel pueden resumirse de manera práctica los pasos que hay que tener en cuenta. En el siguiente epígrafe, se ejemplifica mediante un caso concreto en SPSS. Para reducir la complejidad de la explicación se ha simplificado la definición matemática a solo las ecuaciones esenciales.

En este caso, para comprender mejor el proceso se ha tomado como ejemplo un tipo de investigación tradicional con los datos PISA: los factores asociados al rendimiento de los y las estudiantes. Así, la variable dependiente será la media del rendimiento en ciencias de la prueba PISA 2015 del alumnado español (PV1),² variable de tipo numérico continuo, y se analizará qué factores de nivel 1 (asociados al alumno o alumna) y qué elementos del nivel 2 (asociados a los centros) inciden en mayor o menor medida en dicho rendimiento. Estas variables son o bien categóricas o bien continuas.

Tipos de datos

Como ya se ha especificado anteriormente, los datos con los que se va a trabajar deben seguir una estructura jerárquica conocida por la persona investigadora. Se trabajará con una variable dependiente cuantitativa continua y varios factores o variables independientes que correspondan a los dos o más niveles jerárquicos en los que están estructurados los datos.

Modelo nulo

Pese a que la modelización en este tipo de análisis puede resultar más compleja que en una regresión múltiple, se procede, básicamente, como en un modelo univariado. Para ello, en primer lugar, se especifica el modelo nulo, sin covarianzas, expresado mediante la siguiente ecuación:

[1]

$$Y_{ij} = \beta_{0j} + e_{ij}$$

Este modelo nulo, que más tarde tendrá efectos aleatorios en los dos niveles que se especifiquen, solo consiste en la variable dependiente o de respuesta (en este caso el rendimiento académico del alumno), representada por Y_{ij} , el coeficiente (β_{0j}), que representa la media del rendimiento de su centro y el error (e_{ij}) o residuos en torno a esa media, sin incluir, por el momento, las variables explicativas.

² Aunque PISA proporciona 10 valores plausibles (PV) que puntúan dicho rendimiento, se ha tomado solo el primer valor (PV1) para simplificar el ejemplo. Si la investigación se hiciera realmente, debería replicarse este análisis para los 10 PV y calcular la media.



Tatiana Iñiguez-Berrozpe, Francesco Marcaletti. *Modelos lineales multinivel en SPSS y su aplicación en investigación educativa*

La intercepción de cada centro (β_{0j}) se establece entonces igual a una gran media o media general (γ_{00}) y un error aleatorio (u_{0j}):

[2]

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

De tal forma que, sustituyendo [2] en [1] se obtiene:

[3]

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

Introducción de variables explicativas o contextuales

Una vez se ha definido el modelo nulo, que tendrá efectos aleatorios, hay que introducir las variables explicativas o contextuales por niveles. Así el modelo de nivel 1 vendría expresado por:

[4]

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$

En este punto, entendiendo que la relación entre la variable Y_{ij} y la variable x_{ij} se verán afectadas por la variable contextual z_j , se introduce el modelo de nivel 2 definido por los siguientes coeficientes:

[5]

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_j + u_{1j}$$

Así, el modelo combinado de ambos niveles (1 y 2), que tiene ya en cuenta la jerarquía de los datos, podría expresarse de manera simplificada de la siguiente forma:

[6]

$$Y_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}x_{ij}z_j + (u_{0j} + u_{1j}x_{ij} + e_{ij})$$

Como puede apreciarse, este modelo incluye tanto efectos fijos (los que están fuera del paréntesis), como aleatorios (los que están dentro del paréntesis); por tanto, es un modelo mixto.

5. Caso práctico en SPSS³

Entrada de datos

Aunque varios autores afirman que SPSS no es el mejor programa para el análisis de datos cuantitativos mediante MLM (Tabachnick y Fidell, 2001; Twisk, 2006), la entrada de datos depende un poco del tipo de modelo multinivel que se desea ejecutar, ya que, si se desea un tipo de análisis longitudinal, el proceso varía. No obstante, dado el ámbito de estudio y la utilización extendida de SPSS en investigación educativa, se asume el tipo de análisis más habitual (sincrónico) y a través de este programa. En cuanto a los datos en sí, sus características deben ser similares al resto de los análisis que se efectúan mediante SPSS, esto es, una fila por caso y cada caso respondiendo a una serie de variables categóricas y cuantitativas numéricas.

Tabla 1
Variables empleadas para el MLM con datos PISA 2015 en SPSS⁴

Variabls independientes de nivel 1	Variabls independientes de nivel 2	Variable dependiente
Número de recursos culturales en casa (CULTPOSS)	Titularidad del centro (SC013Q01TA)	Rendimiento del alumno/a en ciencias (PV1SCIE)
Índice social, económico y cultural familiar (ESCS)	Proporción de ordenadores por estudiante (SC004Q01TA)	
Nivel ocupacional de los padres (hisei)	Tamaño del centro (SCHSIZE)	

En el caso de PISA, los microdatos que proporciona la OCDE se encuentran en tres bloques distintos: (1) encuesta del alumnado, (2) encuesta a las escuelas (contestada por los directores de centro) y (3) encuesta del profesorado. Por ello, en este caso hay que fusionar, en primer lugar, las encuestas 1 y 2 en un mismo archivo de datos SPSS⁵.

³ Los microdatos empleados están disponibles en acceso libre en: <http://www.oecd.org/pisa/data/2015database/>. Para que el seguimiento del caso sea más sencillo, se han respetado los nombres de las variables.

⁴ Nuevamente, para simplificar el ejemplo, no se toman todos los factores contextuales que recomienda la literatura para introducir en el modelo, porque son excesivamente numerosos para una explicación metodológica. Se escogen solo algunos que han demostrado en pasadas ediciones tener un efecto relevante en el rendimiento del alumnado. En el caso del análisis, se ha tomado el análisis más empleado en la literatura sobre datos PISA, es decir, aquel que toma la variable centro como aleatoria y los factores contextuales como fijos.

⁵ Para realizar la fusión de las encuestas deben descargarse del enlace especificado en la nota a pie de página n.º 3 los siguientes archivos: "Student questionnaire data file" (CY6_MS_CMB_STU_QQQ.sav) y "School questionnaire data file" (CY6_MS_CMB_SCH_QQQ.sav). Una vez descargados, deben abrirse ambos archivos y ordenarse por una variable que compartan, por ejemplo, la primera "CNTRYID" (*Datos > Ordenar casos*) y se guardan ambos archivos. Seguidamente se realiza la fusión de archivos: en la encuesta de estudiantes (CY6_MS_CMB_STU_QQQ.sav) *Datos > Fusionar archivos > Añadir variables*. En ese cuadro de diálogo se selecciona la opción "Un conjunto de datos abierto" dentro de la encuesta de escuelas (CY6_MS_CMB_SCH_QQQ.sav). En el siguiente cuadro se selecciona "Asignar casos en variables clave", "Los datos están ordenados por orden de variables clave en ambos archivos" y "El conjunto de datos activo es una tabla de claves", y se coloca como variable clave en el cuadro "CNTRYID". Se acepta y se comprueba que en la base de datos resultante se encuentran tanto las variables de alumnado como las de las escuelas. El archivo ya fusionado se puede descargar directamente desde el apartado "Herramientas del artículo > Archivos complementarios" que ofrece la página web de este artículo.

Modelo nulo

Una vez se tienen claros los datos que hay que analizar y su estructura, se pasa a ejecutar el análisis, en primer lugar, sin tener en cuenta dicha estructura. Para ejecutar el MLM se usa el comando *Modelos mixtos > Lineales* (Fig. 2), donde aparece el cuadro de diálogo que puede verse en la Fig. 3. Este cuadro de diálogo especifica la naturaleza jerárquica de los datos, si bien como por el momento se está ignorando dicha jerarquía, se ignora y se pulsa *Continuar*.

Figura 2
Cuadro de diálogo en SPSS para llevar a cabo un MLM

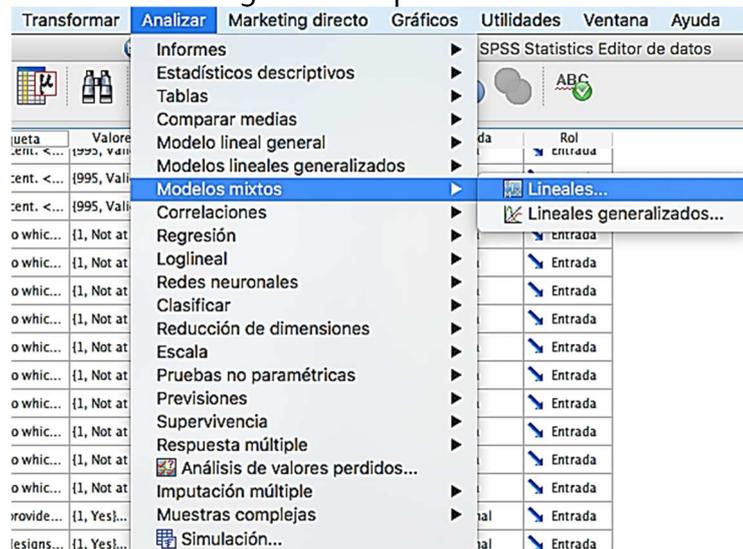
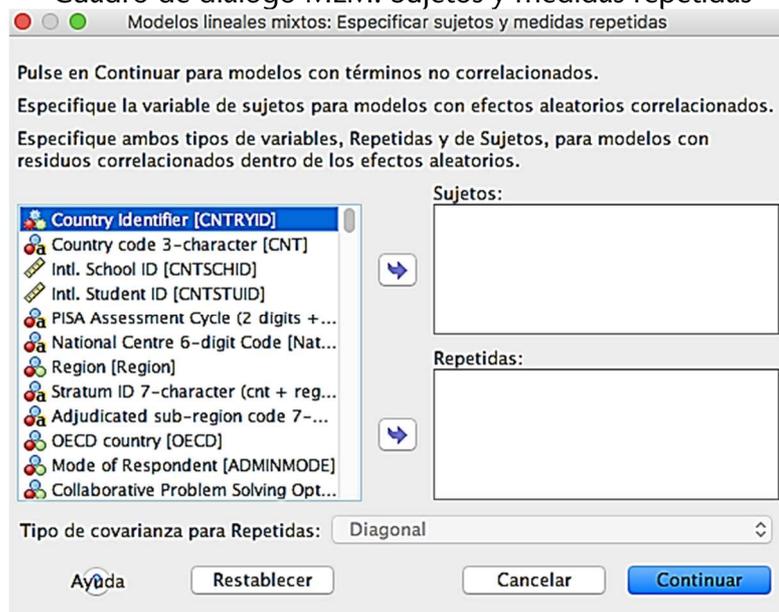


Figura 3
Cuadro de diálogo MLM. Sujetos y medidas repetidas



En el siguiente cuadro de diálogo que aparece (Fig. 4) hay que colocar la variable dependiente "Plausible Value 1 in Sciences" en la casilla "Variable dependiente" y en



Tatiana Iñiguez-Berrozpe, Francesco Marcaletti. *Modelos lineales multinivel en SPSS y su aplicación en investigación educativa*

“Factores” hay que incluir la identificación del centro (“Intl. School ID”). A continuación hay que pulsar la opción *Aleatoria...* y en el siguiente cuadro de diálogo, seleccionar “CNTSCHID” y pulsar *Añadir*. A continuación hay que trasladarlo a la casilla de *Modelo* (Fig. 5). Pulsando *Continuar* se vuelve al cuadro de diálogo de la Fig. 4, donde hay que seleccionar *Estadísticos...* y allí marcar las opciones *Estadísticos descriptivos*, *Estimaciones de los parámetros* y *Contrastes sobre los parámetros de covarianza*.

Figura 4
Cuadro de diálogo MLM. Efectos aleatorios

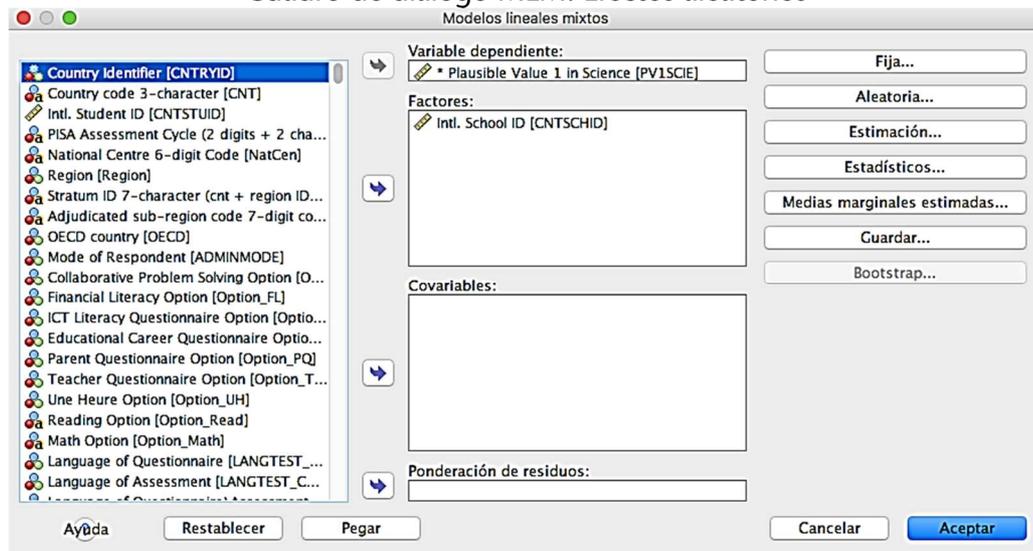
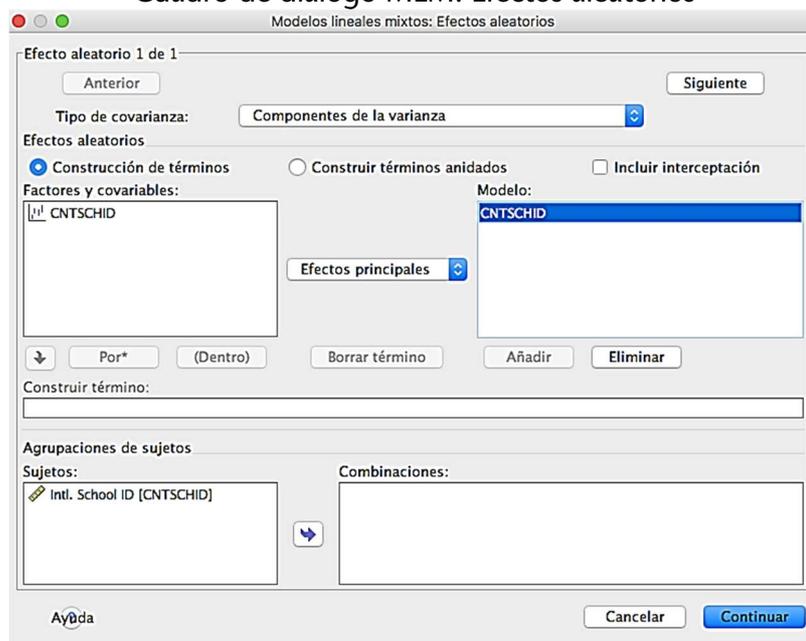


Figura 5
Cuadro de diálogo MLM. Efectos aleatorios



En el visor de resultados se mostrará varia información de interés, como, por ejemplo, en la primera tabla, el número de centros y de alumnos por centro, la media en la prueba de ciencias de cada centro, la variabilidad intracentro y la variabilidad intercentros. Sin



Tatiana Iñiguez-Berrozpe, Francesco Marcaletti. *Modelos lineales multinivel en SPSS y su aplicación en investigación educativa*

embargo, interesa también el *output* relacionado con las pruebas de efectos fijos. Como puede verse en la Fig. 6, con $p < 0,001$ (Sig. = ,000), indica que puede rechazarse la hipótesis de que el efecto del factor centro es nulo en el rendimiento del alumnado.

Figura 6
Pruebas de efectos fijos de tipo III^a

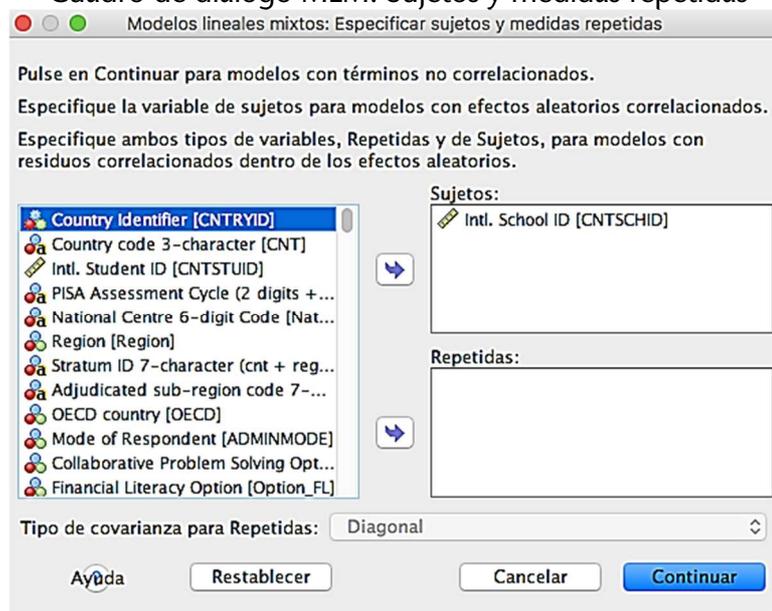
Origen	gl de numerador	gl de denominador	F	Sig.
Interceptación	1	199,024	38781,815	,000

a. Variable dependiente: * Plausible Value 1 in Science.

Modelización atendiendo a la jerarquía de los datos

A continuación, sí que va a interesar atender a la jerarquía de los datos, por lo que, en este caso sí que hay que atender al primer cuadro de diálogo que aparece en el comando *Modelos mixtos > Lineales*. En el cuadro denominado *Sujetos* hay que introducir la variable contextual que agrupa al alumnado, es decir, el centro ("Intl. School ID"), y pulsar *Continuar* (Fig. 7).

Figura 7
Cuadro de diálogo MLM. Sujetos y medidas repetidas

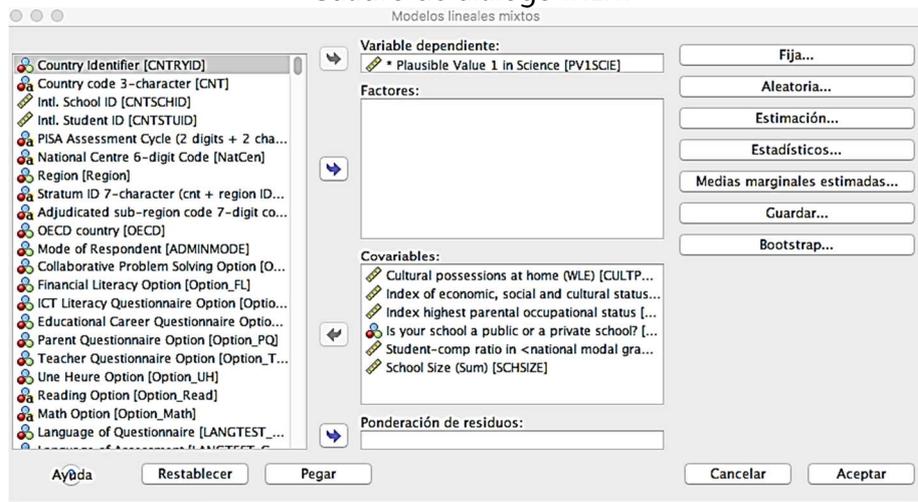


Seguidamente aparecerá el cuadro de diálogo con el que ya se ha trabajado (Fig. 4). Allí hay que introducir la variable dependiente en el cuadro destinado a tal efecto (Plausible Value 1 in Sciences), siendo ahora el momento de añadir las covariables (que van a hacer referencia a las variables independientes). Como ejemplo, se introducen las variables independientes señaladas previamente en la Tabla 1 (relacionadas con el alumno/a y con el centro) (Fig. 8), que van a considerarse como efectos fijos.



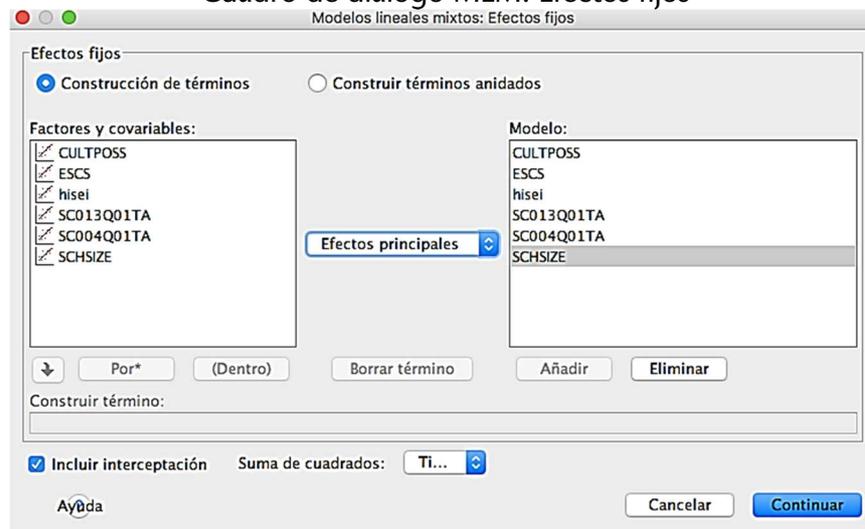
Tatiana Iñiguez-Berrozpe, Francesco Marcaletti. *Modelos lineales multinivel en SPSS y su aplicación en investigación educativa*

Figura 8
Cuadro de diálogo MLM



Así, en este mismo cuadro de diálogo hay que pulsar *Fija...* y en el siguiente hay que añadir las covariables al modelo, seleccionándolas y pulsando *Añadir* en el recuadro de *Modelo* (Fig. 9), elegir la opción *Efectos principales* en el desplegable central, y asegurarse de que la opción *Incluir interceptación* está marcada. Pulsar *Continuar* y regresar al cuadro de la Fig. 8.

Figura 9
Cuadro de diálogo MLM. Efectos fijos



A continuación, hay que añadir los efectos aleatorios. Para ello hay que seleccionar *Aleatoria...* y aparecerá el cuadro de diálogo de *Efectos aleatorios*. En este caso no hay que trasladar las covariables al modelo (ya que se ha indicado que son efectos fijos). En lugar de ello, en las agrupaciones de sujetos hay que añadir los centros ("Intl. School ID") al cuadro de *Combinaciones*. Hay que asegurarse nuevamente de que la casilla *Incluir interceptación* está marcada (Fig. 10). Pulsar *Continuar* y volver al cuadro de diálogo anterior. Con ello lo



Tatiana Iñiguez-Berrozpe, Francesco Marcaletti. *Modelos lineales multinivel en SPSS y su aplicación en investigación educativa*

que se indica al modelo es que los interceptos varían en función del centro al que pertenece el alumno o alumna, es decir, se tiene en cuenta el anidamiento de los sujetos, su estructura jerárquica.

Figura 10
Cuadro de diálogo MLM. Efectos aleatorios

Finalmente, en este cuadro hay que pulsar *Estadísticos...* y en las opciones que aparecen seleccionar *Estimaciones de los parámetros* y *Contrastes sobre parámetros de covarianza* (Fig. 11). Pulsar *Continuar* y *Aceptar* para que el programa efectúe el análisis.

Figura 11
Cuadro de diálogo MLM. Pestaña *Estadísticos*

Tatiana Iñiguez-Berrozpe, Francesco Marcaletti. *Modelos lineales multinivel en SPSS y su aplicación en investigación educativa*

Los resultados que aparecen son los siguientes:

Figura 12
Pruebas de efectos fijos de tipo III

Origen	gl de numerador	gl de denominador	F	Sig.
Interceptación	1	343,392	2239,658	,000
CULTPOSS	1	5739,137	79,860	,000
ESCS	1	5752,977	50,243	,000
hisei	1	5715,541	7,146	,008
SC013Q01TA	1	185,654	11,370	,001
SC004Q01TA	1	167,619	9,502	,002
SCHSIZE	1	168,356	4,512	,035

Variable dependiente: * Plausible Value 1 in Science.

Figura 13
Estimaciones de efectos fijos

Parámetro	Estimación	Error estándar	gl	t	Sig.	Intervalo de confianza al 95 %	
						Límite inferior	Límite superior
Interceptación	464,473393	9,814540	343,392	47,325	,000	445,169211	483,777576
CULTPOSS	11,273025	1,261466	5739,137	8,936	,000	8,800075	13,745974
ESCS	13,638477	1,924110	5752,977	7,088	,000	9,866497	17,410458
hisei	,225300	,084280	5715,541	2,673	,008	,060079	,390520
SC013Q01TA	16,064268	4,764139	185,654	3,372	,001	6,665459	25,463076
SC004Q01TA	,232950	,075572	167,619	3,082	,002	,083754	,382146
SCHSIZE	-,013210	,006218	168,356	-2,124	,035	-,025486	-,000933

Variable dependiente: * Plausible Value 1 in Science.

Figura 14
Estimaciones de parámetros de covarianza

Parámetro	Estimación	Error estándar	Wald Z	Sig.	Intervalo de confianza al 95 %	
					Límite inferior	Límite superior
Residuo	5877,998894	111,322151	52,802	,000	5663,811330	6100,286360
Interceptación Varianza [sujeto = CNTSCHID]	362,636966	59,426444	6,102	,000	263,016526	499,989758

a. Variable dependiente: * Plausible Value 1 in Science.

En estos *output* que proporciona SPSS, puede observarse que todos los factores del modelo tienen una influencia significativa en la variable dependiente, es decir, tanto las variables independientes de nivel 1 como las de nivel 2 contribuyen a la varianza del rendimiento del alumnado en ciencias ($p < 0,05$) (Fig. 12).

Para presentar los resultados, principalmente se reporta la estimación de los efectos fijos, el error estándar y la significación de cada variable (Fig. 13). En este caso, la *Titularidad del centro (SC013Q01TA)*, como variable de nivel 2, y el *Índice social, económico y cultural*



Tatiana Iñiguez-Berrozpe, Francesco Marcaletti. *Modelos lineales multinivel en SPSS y su aplicación en investigación educativa*

familiar (ESCS) y el *Número de recursos culturales en casa (CULTPOSS)*, como variables de nivel 1, una vez controladas el resto de las variables, son las que resultan más explicativas de la varianza del rendimiento en ciencias en el modelo simplificado que se ha empleado como ejemplo. Por otra parte, atendiendo a la Fig. 14, puede apreciarse la estimación del efecto del centro al que pertenece cada alumno o alumna. Al introducirlo como variable de agrupación y tratarlo como efecto aleatorio, se ha respetado la estructura jerárquica de los datos a la hora de evaluar los efectos fijos de las variables independientes antedichas de nivel 1 y nivel 2. Si se hubiera llevado a cabo el análisis sin introducir este anidamiento, por ejemplo, mediante una regresión lineal múltiple, los resultados habrían sido muy distintos, debido a que no se habría tenido en cuenta el efecto aleatorio que supone a los sujetos investigados su agrupamiento en el mismo centro (Blanco Blanco *et al.*, 2014; Andreu Abela, 2011; Field, 2013).

En definitiva, los MLM permiten introducir la variable de agrupamiento para estudios en los que los datos pueden no ser independientes entre sí, como en el caso presentado, en el que varios estudiantes encuestados pertenecen al mismo centro, atendiendo, por tanto, a los efectos internivel.

<Referencias bibliográficas>

- Andreu Abela, J. (2011). El análisis multinivel: una revisión actualizada en el ámbito sociológico. *Metodología de Encuestas*, 13, 161-176. Recuperado de <http://casus.usal.es/pkp/index.php/MdE/article/view/1017>
- Blanco Blanco, A., López Martín, E., y Ruiz de Miguel, C. (2014). Aportaciones de los modelos jerárquico-lineales multivariados a la investigación educativa sobre el rendimiento. Un ejemplo con datos del alumnado español en PISA 2009. *Revista de Educación*, 365, 122-149. <http://doi.org/10.4438/1988-592X-RE-2014-365-267>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll* (4.ª edición). Los Angeles: Sage.
- Kreft I. G. G., y De Leeuw, J. (1994). *Introducing multilevel modeling*. Londres: Sage.
- OCDE (2016). *Programme for International Student Assessment (PISA) 2015*. Recuperado de <http://www.oecd.org/pisa/>
- Tabachnick, B. G., y Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn & Bacon.
- Twisk, J. W. R. (2006). *Applied multilevel analysis: a practical guide*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511610806>

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