

## **LOS MODELOS JERÁRQUICOS LINEALES Y SU APLICACIÓN A LA INVESTIGACIÓN EDUCATIVA**

*Arturo de la Orden Hoz*  
Universidad Complutense de Madrid

No resulta excesivamente arriesgado afirmar que el principal avance metodológico de los últimos años en investigación en Ciencias Sociales son los Modelos Jerárquicos Lineales (HLM), también llamados Modelos Multinivel. Desde su nacimiento como una generalización de los modelos lineales tradicionales se han desarrollado de forma tan vertiginosa que modelos tan novedosos como los Causales se han convertido en un caso particular de los HLM. Constituyen un marco de análisis natural para el tratamiento de datos con una estructura multinivel o anidada, situación muy común dentro de la investigación educativa.

En efecto, una de las características de los sistemas educativos que debe ser considerada en un modelo adecuadamente especificado es la estructura organizativa multinivel de la acción escolar y el hecho relacionado de que los datos generados por la investigación educativa institucional muy frecuentemente proceden de esquemas muestrales por conglomerados. Por ejemplo, en el estudio sobre Los resultados escolares, dentro del Diagnóstico del Sistema educativo Español. La Escuela Secundaria Obligatoria, 1997 (INCE. 1998), los alumnos se muestrean en clases que, a su vez, se muestrean de escuelas que, a su vez, se muestrean de provincias o de Comunidades Autónomas. Los extraordinarios avances en la investigación de la acción y efectos de la escuela en los años recientes están vinculados al uso de métodos de regresión lineal multinivel como respuesta a las exigencias analíticas de datos jerárquicamente organizados o anidados, resultados de la aplicación de esquemas muestrales por conglomerados. Estos métodos multinivel o jerárquicos lineales permiten el análisis de modelos explicativos que contienen variables a diferentes niveles (alumno, clase, escuela, distrito, región, estados u otros).

La noción central es que modelos lineales diferentes para el primer nivel o micro-nivel deben ajustarse para cada contexto (macro-nivel). Habitualmente cada grupo tiene las mismas variables explicativas y la misma variable de respuesta, pero con diferentes coeficientes de regresión. Todos los modelos del primer nivel están relacionados por un modelo de segundo nivel en el que los coeficientes de regresión del micro-nivel se regresan en un segundo nivel de variables explicativas.

La naturaleza del segundo nivel determina la naturaleza del modelo para el conjunto de los datos. Y aquí caben muy diversas posibilidades, incluso que no haya ninguna variable de 2º nivel que relacione los diversos contextos y se establezca una ecuación de regresión para cada contexto. Sin embargo, esto no añade nada nuevo desde el punto de vista estadístico, se trata de un análisis que emplea los coeficientes de regresión del primer nivel como variables de respuesta en un segundo nivel. En este análisis (llamado **PENDIENTES COMO RESULTADOS**, Burstein et al, 1978) las regresiones intra y entre grupos no están conectadas unas con otras. Hay de hecho análisis separados. La cuestión central es que en este tipo de modelos los coeficientes son fijos no aleatorios. Si se analizan los datos completos (en los dos niveles) estaríamos ante un modelo en el que varían los coeficientes, donde cada grupo es analizado separadamente y tiene sus propios coeficientes de regresión. Estamos ante un modelo poco parsimonioso, que se ocupa solo de las diferencias y no de los puntos en común.

La idea de un análisis de regresión separado para cada grupo, seguido por la regresión de los coeficientes del primer nivel en el segundo no es suficiente para especificar un modelo multinivel. Es necesario destacar que los modelos multinivel implican la **INTEGRACIÓN ESTADÍSTICA** de los distintos modelos especificados para cada nivel de interés. La integración más simple se da en el caso del **MODELO DE COEFICIENTES ALEATORIOS** (también conocidos como modelos de componentes de varianza, modelo jerárquico lineal o modelo de componentes de error, Kreft y de Leeuw, 1994). Donde los *coeficientes de regresión del primer nivel son tratados como variables aleatorias en el segundo nivel*. Esto significa que los coeficientes del primer nivel son entendidos como los originadores de una distribución de probabilidad. Los parámetros más importantes de esta distribución (media y varianza) están entre el conjunto de parámetros que deben ser estimados en el modelo multi-nivel. Al añadir un segundo nivel de variables explicativas el modelo de coeficientes aleatorios se hace más general y, al mismo tiempo, más útil. Esto es a lo que comúnmente llamamos modelos multinivel.

El modelo general de análisis quedaría representado por la siguiente ecuación:

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + (\epsilon_{ij} + \mu_{0j} + \mu_{1j} x_{1ij})$$

Donde la parte fija está representada por  $\beta_0$  y  $\beta_1$  y la parte aleatoria por los términos que aparecen entre paréntesis en la ecuación. La varianza del primer nivel ( $\sigma_{\epsilon_{ij}}^2$ ) representa la variación entre individuos, estudiantes dentro de cada centro. La varianza del segundo nivel ( $\sigma_{\mu_{0j}}^2$ ,  $\sigma_{\mu_{1j}}^2$  y  $\sigma_{\mu_{01}}^2$ ) representan las variaciones entre los centros. Se ve claramente como no estamos hablando de dos modelos diferentes, sino de un *único modelo general como el presentado en la ecuación*. La principal innovación se encuentra en

el término de error, que refleja tanto la variación en el primer nivel como en el segundo nivel. Esta forma de descomponer la varianza es la diferencia fundamental con los tradicionales modelos de regresión, ya que ante esta situación, no se cumpliría el supuesto de homocedasticidad (para cada punto la varianza es constante,  $v(y | x) = cte$ ).

Hoy existen varios programas estadísticos disponibles para estimar modelos multinivel, siendo el más conocido y usado el HLM (Bryk, Raundebush, Seltzer y Congdon 1988; y Bryk, Raundebush y Congdon 1996). El enfoque utilizado en este programa resuelve las dificultades más comúnmente encontradas en el análisis de datos multinivel: el problema de la unidad de análisis, los sesgos de agregación, la dependencia de los individuos dentro de los grupos que dificulta la correcta estimación de los errores típicos, la heterogeneidad de regresión (las diferentes magnitudes del efecto para los diferentes grupos) y los diferentes significados y efectos de una variable, resultantes de utilizar distintos niveles de agregación.

En este simposio se pretende ofrecer el trabajo de investigación sobre la aplicación de los Modelos Jerárquicos Lineales a la investigación educativa desarrollado en cooperación entre el departamento MIDE de la Complutense y el Centro de Investigación y Documentación Educativa (CIDE). Concretamente se abordará su aplicación en la evaluación de sistemas educativos, en la investigación sobre eficacia escolar y en las investigaciones meta-analíticas.

## REFERENCIAS

- Bryk, A.S.; Raundebush, S.W.; Seltzer, M. y Congdon, R.T. (1988). *An introduction to HLM: Computer program and user's guide*. Chicago, University of Chicago.
- Bryk, A.S.; Raundebush, S.W. y Congdon, R.T. (1996). *HLM: Hiererachical linear and non linear modeling with the HLM/2L and HLM/3L programs*. Chicago, Scientific Software International.
- Burstein, L.; Linn, R.L. and Capell, F.J. (1978). *Analyzing multilevel data in presence of heterogeneous within-class regressions*. *Journal of Educational Statistics* 3, 347-383.
- De la Orden, A. y otros (1998). *Los resultados escolares. Diagnóstico del Sistema Educativo. La Escuela Secundaria Obligatoria*. 1997. Vol. 2, Madrid, MEC-INCE.
- Kreft, I.G.G. y Leeuw, J. (1994). *The gender gap in earnings: a two-way nested multiple regression analysis with random effects*. *Sociological Methods and Research* 22 (3), 319-341.