Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales

por María CASTRO
Universidad Complutense de Madrid
y Luis LIZASOAIN
Universidad del País Vasco-Euskal Herriko Unibertsitatea

En este mes de Junio de 2011, cuando se entrega este artículo, se está celebrando el primer centenario de la empresa norteamericana IBM. Esta celebración conmemora la mejora de los procedimientos para el procesamiento del censo en Estados Unidos. En 1880 se tardaron siete años en procesar el censo de la población. En 1890 la población de Estados Unidos ascendía a 62.622.250 personas y se estimaban necesarios 10 años para completar el censo. Y es cuando aparece la maquina tabuladora de Hollerith que funcionaba actualizando unos contadores a partir de los datos contenidos en una tarjeta perforada. Se emplearon solamente seis semanas en procesar el censo de 1890. Eso sí, el coste fue un 198% superior al del censo anterior.

La generalización y disponibilidad de los ordenadores personales junto con la impresionante capacidad de los procesadores actuales hace que los hechos relevantes hace tan sólo cien años nos hagan sonreír en la actualidad.

Las técnicas de investigación estadísticas utilizadas para profundizar en el conocimiento de los fenómenos educativos son hijas de su tiempo y del paradigma de investigación que manejan los investigadores en cada periodo. La evolución en la investigación educativa viene de la mano de dos revoluciones. Por un lado los investigadores avanzan en sus planteamientos teóricos planteando modelos integrados que quieren estudiar las relaciones entre constructos no directa-

mente observables insertos en contextos complejos. Y por otro, el desarrollo técnico-científico del software para el análisis de datos que ha permitido someter a prueba estos modelos complejos en entornos computacionales razonables en términos de tiempo y esfuerzo. El impacto de las mejoras computacionales en el desarrollo de los modelos cuantitativos de investigación es tan evidente que a día de hoy no nos acordamos de que hace tan sólo 30 años era realmente complejo procesar y estimar algunos de los modelos que hoy nos parecen habituales.

Y si bien tanto las complejas relaciones existentes entre las variables implicadas en los fenómenos educativos, como las técnicas y procedimientos estadísticos que normalmente se utilizan se conocen bien desde el siglo XIX, no es sino hasta las últimas décadas del siglo XX cuando se ha producido una revolución tecnocientífica en el desarrollo del software para el análisis de datos que ha ido transformando las prácticas científicas en el ámbito de la educación.

El software para el tratamiento estadístico de datos muestra desde hace ya dos décadas una gran velocidad en el procesamiento de grandes cantidades de datos y variables, ha mejorado su usabilidad, permitiendo a los usuarios de herramientas de estadística aplicada el desarrollo de modelos complejos con interfaces fácilmente manejables, eso sin considerar la generalización de la disponibilidad de herramientas de computación. Se puede decir que, si bien todo es susceptible de mejora, no hay límite tecnológico evidente a nivel de usuario a la hora de realizar investigaciones cuantitativas, pues ha habido una gran evolución en los recursos y en la relaciones con esos recursos. Entonces, cabe preguntarse ¿dónde están los límites?

Modelización: de la variable aislada al modelo latente

El enfoque clásico del análisis de datos consiste, en general, en aplicar técnicas de contraste de hipótesis y también en determinar la significación de las medias de asociación comunes. Sin embargo, las alternativas a esta orientación surgen en la década de los 60 y se basan en la estimación de parámetros y en el ajuste y comparación de modelos de probabilidad a los datos empíricos (Ato y López García, 1996). El eje central de este enfoque es el denominado *modelado estadístico*. Consiste en la aplicación de un conjunto de procesos con el objeto de conseguir una explicación apropiada de una variable de respuesta (en nuestro caso tendrán niveles de medición de intervalos) a partir de una función ponderada de una o más variable explicativas. Al no ser la explicación perfecta, se precisa de la inclusión de la diferencia entre los datos y el modelo, denominado error o residual (Judd y McClelland, 1989; Estes, 1991; Lunneborg, 1994; Judd, McClelland y Culhane, 1995).

El modelado estadístico busca el modelo más simple que sea capaz de explicar los datos con un mínimo error posible. Siguiendo la metáfora de Dobson (1990), hay un cierto paralelismo entre los datos científicos y la información implicada en un mensaje. Un mensaje implica una señal que está distorsionada por un ruido. La 'señal' se puede entender como una

descripción matemática de las principales características de los datos y el 'ruido' como aquellas características no explicadas por el modelo. Del mismo modo, los datos científicos pueden articularse en forma de modelos matemáticos que incorporan ambos componentes, unos sistemáticos o determinísticos y otros aleatorios. Otra forma de conceptualizar los modelos estadísticos, es como una cuantificación y partición de la varianza atribuible tanto a la 'señal' como al 'ruido'. El objetivo del modelado estadístico es extraer de los datos tanta información como sea posible sobre la 'señal'.

Tal y como señala Gaviria (2000) las técnicas cuantitativas de investigación evolucionan vinculadas no sólo a la tecnología sino fundamentalmente al planteamiento epistemológico que asumen los investigadores en cada momento. Así señala que desde mediados de los 70 se evoluciona desde planteamientos exploratorios y empiristas hacia la búsqueda de una lógica causal en las relaciones dentro del los fenómenos educativos.

En el extremo más exploratorio, se encuentran las técnicas de minería de datos (*data mining*) que tienen como objetivo la extracción de información relevante de grandes bases de datos empleando para ello algoritmos o técnicas que tratan de localizar información no trivial.

En el extremo más confirmatorio, aparece la idea de variable latente, no observable directamente y no manipulable, naciendo así los *modelos de ecuaciones estructurales*. A mediados de los 80, el contexto cobra una nueva dimensión e irrumpe con fuerza en el panorama de las técnicas de investigación los modelos multinivel o *modelos jerárquicos lineales*.

En este artículo asumimos la tendencia creciente en investigación educativa al planteamiento y prueba de modelos comprehensivos sobre los fenómenos educativos. De manera más específica nos centraremos en los métodos estadísticos de segmentación, en los modelos de ecuaciones estructurales y en los modelos multinivel, que vienen dominando el panorama de la investigación educativa desde mediados de los 70 y que gracias al desarrollo tecnológico se han podido generalizar recientemente. Así, no vamos a relacionar aquí un catalogo de técnicas para variables de intervalos. Se presentan sólo aquellas en las que tradicionalmente se ha utilizado variables de intervalos como variable de respuesta, aunque haya adaptaciones y posibilidad de utilizar estos modelos con otras distribuciones distintas a la normal.

Minería de datos: árboles estadísticos de decisión

El incremento exponencial tanto de la capacidad de almacenamiento como de la potencia computacional de los sistemas informáticos ha posibilitado el desarrollo de un conjunto de técnicas en las que confluyen la inteligencia artificial y la estadística y que se engloban bajo la denominación general de minería de datos (*data mining*).

Tal y como el nombre apunta, el objetivo de estas técnicas es el de extraer información relevante (algunos dirán que conocimiento) de grandes bancos o bases

de datos empleando para ello algoritmos o técnicas que tratan de localizar información no trivial, entendiendo por tal diferencias, patrones, relaciones significativas, efectos de interacción, etc. que se supone se encuentran escondidos en dichas grandes masas de datos. En definitiva, se trata de aplicar el principio de "separar la mena de la ganga" tan empleado en minería y metalurgia.

Son muchos los términos que se asocian a este campo. Entre otros, citemos las redes neuronales, las técnicas de clasificación (supervisada o no), los algoritmos genéticos, el reconocimiento de patrones, la visualización, la minería de textos, la minería de la web, el aprendizaje automático, etc.

De la misma manera, son muchas y diversas las técnicas que se emplean en este campo. Por razones evidentes, vamos a centrarnos en las técnicas estadísticas y más específicamente en una de ellas que son los *árboles de decisión*, técnica también conocida como *métodos estadísticos de segmentación*.

Los árboles de decisión o clasificación son un conjunto de técnicas que permiten definir y validar modelos de forma que se pueda determinar qué variables (predictoras) explican los cambios de una variable dependiente. Son técnicas estadísticas explicativas de la familia de la regresión o el análisis discriminante, pero tienen la ventaja de que tanto la variable criterio como las predictoras pueden ser de cualquier tipo (cuantitativas o categoriales) lo que en el contexto de la investigación educativa es siempre una cuestión a tener en cuenta.

El principio básico del modo de operar de los árboles de decisión consiste en dividir progresivamente un conjunto en clases disjuntas. El proceso se inicia tomando en consideración el total de casos de la muestra y todas las variables incluidas en el modelo.

Sobre este conjunto inicial –denominado nodo raíz– se efectúa una partición del grupo original en 2 ó más subgrupos atendiendo a los valores de la variable predictora que más se asocie a la variable dependiente.

Una vez efectuada esta primera segmentación, el proceso se re-inicia en cada uno de los subgrupos establecidos en el paso anterior de forma que estos subgrupos se siguen subdiviendo hasta que el proceso de segmentación finaliza cuando se alcanza alguno de los criterios de parada establecidos a priori. El resultado se plasma en un árbol de decisión que muestra la estructura y relaciones entre las variables para cada uno de los segmentos o subgrupos (nodos).

Para ilustrar brevemente el procedimiento, pensemos en el caso de un típico banco de datos relativos a una evaluación de un sistema educativo o de un conjunto de centros, por ejemplo los resultados de una evaluación diagnóstica en una Comunidad Autónoma cualquiera. En el mismo se dispondrá al menos de una variable de respuesta o dependiente que en nuestro caso suele ser habitualmente la puntuación en una prueba que mida el desempeño de los estudiantes en alguna competencia básica, por ejemplo la competencia numérica y matemática. Junto a esta va-

riable dependiente hay un amplio conjunto de variables relativas a diferentes aspectos del estudiante, de su familia, los docentes, el centro, el clima escolar, etc. En principio todas ellas conforman el conjunto de variables explicativas o predictoras. Como ha quedado dicho, todas estas variables pueden ser de cualquier tipo, cuantitativas o categoriales.

Si con este hipotético banco de datos realizamos un análisis de segmentación, el proceso se inicia analizando todas las variables predictoras con el objetivo de determinar cuál de ellas –y qué partición de la misma– es la que maximiza la diferencia entre los grupos con respecto a la variable criterio. Y una vez seleccionada la variable y establecida la partición se generan los subgrupos –nodos– en función de la misma, de forma que ya se ha obtenido el primer nivel del árbol. Como antes decíamos, ahora el proceso se reiniciaría de la misma manera para cada uno de los subgrupos obtenidos.

Queremos insistir, pues es característico de las técnicas de minería de datos, en el hecho de que el algoritmo debe buscar no sólo la variable que más se asocie a la dependiente, sino la partición que maximice la información, la diferencia, con la variable dependiente. Es por tanto un procedimiento que implica una enorme cantidad de cómputo.

Supongamos el caso más sencillo en que todas las variables (la dependiente y las explicativas fuesen dicotómicas) y por simplicidad pensemos en que hay sólo 3 variables explicativas. En tal caso, el algoritmo se limitaría a buscar cuál de las tres es la que más se asocia a la variable dependiente, cuál maximiza la diferencia. Y para ello bastaría con realizar los tres contrastes de independencia en tablas de contingencia 2x2 y seleccionar la variable explicativa que cruzada con la dependiente arrojase un mayor valor de chi cuadrado.

Pero la cosa se complica si las variables no son dicotómicas. Sigamos con la variable dependiente dicotómica y pensemos, por ejemplo, en que las 3 variables explicativas son categoriales con 4 valores (A, B, C, D). En tal caso, el algoritmo no puede limitarse sólo a buscar cuál es la variable explicativa que más se asocia a la dependiente sino que además, para ello debe buscar para cada una de ellas cuál es la agrupación de valores que lleva asociado un mayor valor de chi cuadrado. Y una vez hecho esto finalmente escoger la variable y partición que maximice el valor del estadístico.

En este caso, y para cada una de las explicativas, debe examinar las siguientes 14 posibles agrupaciones de valores:

1 A-B-C-D	2 A-BCD	3 B-ACD	4 C-ABD	5 ABC-D	6 AB-CD	7 AC-BD
8 AD-BC	9 AB-C-D	10 A-BC-D	11 A-B-CD	12 AC-B-D	13 AD-B-C	14 BD-A-C

Es decir, para cada una de las tres variables explicativas se han de calcular 14 chi cuadrados, por tanto en nuestro ejemplo 42, y seleccionar de entre esas 42 particiones la de mayor valor del estadístico.

En el caso de que una variable explicativa sea ordinal el asunto se simplifica un tanto pues no todas las combinaciones posibles son admisibles dado que hay que respetar el orden subyacente de forma que, en nuestro ejemplo, dada una variable ordinal de 4 valores, sólo serían admisibles las particiones que respeten el orden de las cuatro categorías. En este caso, la mitad (1, 2, 5, 6, 9, 10 y 11).

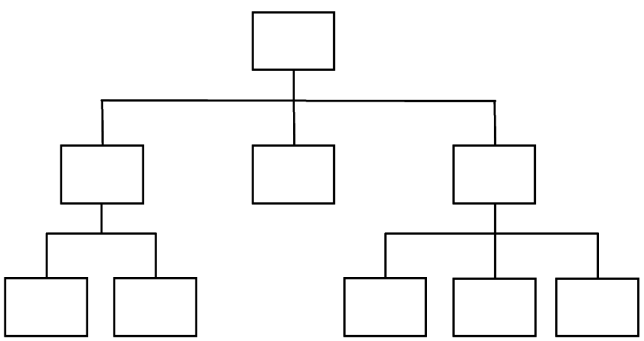
En cambio, si las variables explicativas son cuantitativas, la cosa toma otro cariz. Desde un punto de vista teórico, si la variable es continua, las posibilidades son infinitas. Como en la práctica cualquier variable se hace discreta en función del nivel de precisión de la medida, hay un número finito de posibles particiones, pero es evidente que en este caso la masa de cálculos a realizar se dispara. Para encarar este problema, los algoritmos fijan unos parámetros del número máximo de particiones a realizar.

Con estas líneas, pensamos que queda clara la complejidad computacional de este tipo de técnicas y justificada la afirmación inicial de que son sólo viables gracias a la capacidad de almacenamiento y a la potencia de cómputo de los sistemas informáticos.

En función de las diversas combinaciones posibles de los niveles de medida de las variables, los algoritmos emplean diferentes estadísticos para evaluar el grado de asociación de las mismas. Los más habituales son chi cuadrado, la F y el índice de impureza de Gini.

Como antes hemos señalado, en aplicación de este modo de operar, el procedimiento va realizando segmentaciones sucesivas a partir del nodo raíz y de los subsiguientes nodos que en cada nivel se van generando hasta que se satisface alguno de los criterios de parada previamente establecidos. Los más habituales se refieren al nivel alfa de significación de los estadísticos de asociación; al número de efectivos en los nodos, bien en los nodos-padres o bien en los nodos-hijo; o al número de niveles de profundidad del árbol.

FIGURA 1: *Árbol de decisión*



La Figura 1 muestra la estructura de un hipotético árbol con 2 niveles y 8 nodos, de los cuales 6 son terminales. Siguiendo con nuestro ejemplo, la primera variable de segmentación bien pudiera ser el nivel socioeconómico y cultural medio de las familias de un centro escolar, de forma que esta variable (ISECM, cuantitativa u ordinal) –de entre todas las variables predictoras– es la que genera una mayor diferencia entre los grupos con respecto a la puntuación en matemáticas (variable cuantitativa). Es la que más se asocia. Ninguna otra agrupación atendiendo a los valores de otra variable predictora es capaz de generar una diferencia mayor.

El segundo nivel se genera a partir de los tres subgrupos del nivel socioeconómico medio. Por ejemplo, el nodo 1 agrupará a los centros con menor ISECM y a su vez podría subdividirse, por ejemplo, atendiendo a la titularidad pública o privada (variable cualitativa) de dichos centros. El nodo 2 es un nodo terminal y no hay ninguna variable que con respecto a la cuál se generen diferencias significativas en matemáticas entre los sujetos de dicho nodo. Y, por último el nodo 3 es el de los centros con mayor ISECM y se subdivide en tres subgrupos atendiendo por ejemplo a 3 niveles del clima escolar (variable ordinal).

Este sencillo ejemplo nos muestra las capacidades analíticas de esta técnica. Mediante la misma disponemos de un modelo que nos explica la varianza en la puntuación en matemáticas en función del nivel socioeconómico medio de las escuelas en primer lugar. Pero luego, para las diferentes agrupaciones de centros, otras variables con fuerte capacidad explicativa serían la titularidad o el clima escolar.

Vemos, pues, lo que al principio se apuntaba. Son técnicas múltiples que operan dentro de un modelo de dependencia al modo de las técnicas de regresión. Pero además de sus capacidades analíticas y de –como acabamos de ver– poder operar con cualquier tipo de variables, otra de las principales ventajas que estas técnicas aportan estriba en que sus resultados se presentan de forma gráfica siendo de muy sencilla interpretación.

Existen varios algoritmos o procedimientos de generación de árboles de decisión. Los dos más habituales son el CHAID *(CHi-squared Automatic Interaction Detector),* desarrollado por Kass (1980), y el CART *(Classification and Regression Trees)* desarrollado por Breiman y colaboradores (1984). La diferencia básica entre ambos radica en que CHAID genera modelos no binarios mientras que CART siempre subdivide un nodo padre en dos nodos hijos. Cada uno tiene sus ventajas y limitaciones. Los árboles no binarios tienden a ser más extensos y menos profundos mientras que los binarios suelen tener más niveles y habitualmente las soluciones binarias son más sencillas de interpretar. Por el contrario, CHAID, como más adelante veremos, puede emplearse en tareas de clasificación y generación de subgrupos precisamente en razón de que no se limita a soluciones binarias.

Por último, una importante ventaja de CART es que ofrece la opción de "po-

dar" el árbol para evitar el sobreajuste que es uno de los principales problemas de estas técnicas. La poda consiste en una optimización del modelo propuesto mediante un algoritmo de coste-complejidad (Kim, 1991) que elimina las ramas y nodos que incrementan la complejidad del modelo sin aportar excesiva información.

Junto con los procedimientos de generación del árbol, estas técnicas ofrecen también recursos para evaluar el grado de ajuste de los modelos como tablas de clasificación errónea, estadísticos de error o porcentaje de varianza explicada. Adicionalmente, pueden generar reglas de clasificación de casos en función de las especificidades del modelo.

Como antes se ha señalado, estas técnicas pueden emplearse para la generación y validación de modelos predictivo-explicativos al estilo de la regresión múltiple, el análisis discriminante o la regresión logística. Pero además tienen otras posibilidades de las que estos carecen.

Y una de las más reseñables es la capacidad de detectar efectos de interacción en subgrupos específicos de casos, en capas o niveles profundos del árbol o en algunas regiones del mismo. Y esto es un enfoque exploratorio clásico de la minería de datos. Puede darse incluso el caso de que a nivel superficial –para todo el gran grupo– no haya ninguna variable explicativa que genere diferencias significativas con respecto a la dependiente; pero eso no es óbice para que ello pueda ocurrir en algunos subgrupos específicos de sujetos. Y esto puede ser muy relevante y es difícil de encontrar con procedimientos clásicos habida cuenta del enorme número de posibilidades y combinaciones a examinar. En cambio, mediante los árboles de decisión basta con fijar el parámetro del nivel alfa en valores altos para impedir la parada del algoritmo y ejecutar el procedimiento de forma que –si se dispone de un número de casos suficientemente grande– el procedimiento será capaz de localizar estos hipotéticos subgrupos con efectos significativos de interacción.

En un trabajo desarrollado por Lizasoain y Joaristi (2010) con datos de una evaluación a gran escala en el estado de Baja California (México) se muestran dos de las principales potencialidades de este tipo de técnicas de minería de datos. Por una parte, la detección profunda de interacciones que se dan con variables cuya capacidad de segmentación opera exclusivamente en algunos subgrupos específicos de sujetos. En dicho trabajo se encontró que para grupos pequeños pero muy diferenciados con respecto a la puntuación en lengua española, aparecían variables predictoras muy distintas. En algunos nodos (compuestos por estudiantes de muy bajo rendimiento) aparecían variables ligadas con el cumplimiento formal del docente o del estudiante (puntualidad, falta de asistencia) o el estado físico de las aulas y escuelas.

La segunda cuestión hace referencia a lo que también antes ha sido apuntado referente al uso de estas técnicas como instrumento de clasificación. En el caso que nos ocupa, el modelo se diseñó con la puntuación en lengua española como variable

dependiente y con un amplio conjunto de variables explicativas incluyendo el identificador de cada centro. Y lo que ocurrió es que la primera variable de segmentación fue precisamente esta variable identificadora. Esto quiere decir que de entre todas las variables predictoras, ésta es la que genera una mayor diferencia entre los grupos con respecto al rendimiento en español. Esto supone que, para explicar las diferencias existentes en el rendimiento en lengua española, la escuela es clave.

Pero sin entrar ahora en esas consideraciones, lo que aquí queremos ilustrar es que, mediante el procedimiento CHAID, se generó un primer nivel del árbol compuesto por 9 nodos que agrupaban a los 71 centros escolares de la muestra. Dicho de otra manera, ordenar los 71 centros por su puntuación media en la variable dependiente es trivial. Lo que en cambio aquí se obtiene es la clasificación óptima de los mismos en función de la puntuación en lengua en una solución que maximiza la varianza entre grupos minimizando la existente dentro de cada grupo. Los grupos de escuelas así generados son, dentro de cada nodo, muy parecidos entre sí, y a la vez los nodos están lo más diferenciados posibles con respecto a la variable dependiente. Y una clasificación de estas características ya no es trivial y resulta de gran utilidad para usos y análisis posteriores.

Hasta ahora se han expuesto las principales ventajas y potencialidades de los árboles de decisión y de los dos algoritmos más habituales. De cualquier forma, el operar con un solo árbol tiene sus limitaciones por lo que, desde el enfoque exploratorio imperante, es recomendable emplear las dos estrategias y comparar los resultados de ambas.

Al margen de ello, un problema común de los árboles es que las soluciones plantean problemas de robustez pues un cambio de variable en los niveles iniciales origina modelos muy diferentes y en ocasiones es muy frecuente que haya variables explicativas con capacidad de discriminación muy similar. Dado que cada nodo del árbol es el resultante de las particiones previas, las subsiguientes en niveles inferiores están lógicamente predeterminadas por las anteriores. Esto hace que pequeños cambios en los datos generen soluciones muy distintas.

Para evitar estos problemas, Breiman (2001) propuso una metodología complementaria basada en el algoritmo CART que el mismo diseñó. Y la misma consiste en generar muchos árboles distintos (sin podar) a partir de muchos subconjuntos similares de los datos que se crean mediante remuestreo con reposisión (bootstrapping) de la muestra original. Igualmente, para cada nodo se seleccionan al azar un subconjunto de variables.

De esta forma se incrementa la variabilidad entre las diferentes soluciones y se reduce la dependencia con respecto a segmentaciones previas. Finalmente, con todas las soluciones generadas se calcula un promedio que permite obtener como resultado una ordenación de las variables en función de su importancia en el modelo de cara a la predicción de la variable dependiente.

Como no podía ser de otra manera, a esta técnica basada en multitud de árboles, Breiman la denominó bosques aleatorios (Random Forests). Vandamme, Meskens y Superby (2007) aplicaron este procedimiento en un estudio sobre la predicción del rendimiento académico. Strobl, Malley y Tutz (2009) hacen un recorrido por todo este conjunto de técnicas y su aplicación en la investigación en Psicología.

Con estas líneas confiamos en haber mostrado las principales aplicaciones y usos de este tipo de técnicas de minería de datos en la investigación educativa que, como se ha visto, ofrecen potencialidades dignas de ser tomadas en cuenta en ámbitos como la propia segmentación de poblaciones, el estudio y reducción de la dimensionalidad, la detección de la interacción y la validación de modelos predictivos. Dado el carácter marcadamente exploratorio de estas técnicas, lo más apropiado es emplearlas para diseñar y depurar modelos que puedan luego ser analizados, ajustados y evaluados mediantes técnicas confirmatorias. De cualquier forma, es preciso señalar que es requisito indispensable contar con un volumen de datos suficientemente grande como para que estos algoritmos puedan operar. De ahí que su aplicación resulte especialmente apropiada en, por ejemplo, evaluaciones a gran escala.

Con respecto a los programas informáticos para árboles de decisión, SPSS incorpora esta opción dentro del menú relativo a las técnicas de clasificación. Dependiendo de la versión del programa, este módulo va incorporado al módulo base o al de técnicas avanzadas.

Con respecto a R, en el apartado específico dedicado al aprendizaje automático (http://cran.es.r-project.org/web/views/MachineLearning.html), es posible encontrar información muy detallada sobre los paquetes disponibles para este tipo de técnicas. Para obtener árboles CART el paquete más empleado y recomendado es *rpart* que viene incorporado al R básico. Otros paquetes de generación de árboles son *tree* y *party*. Para la técnica de bosques aleatorios, el paquete es *randomForest.*

Un buen manual de minería de datos en R es la obra de Torgo (2010). El autor tiene además una página web específica: *http://www.liaad.up.pt/~ltorgo/DataMiningWithR/*

Para facilitar el empleo de estas técnicas en R, recomendamos el programa Rattle que es un interfaz gráfico de usuario para minería de datos en R. Incluye árboles de decisión, bosques aleatorios y muchos más procedimientos proporcionando un entorno gráfico de trabajo al operar con los distintos paquetes de R. Es un programa de código abierto, libre y gratuito que está disponible en http://www.togaware.com/. Las obras de Williams (2009, 2011) constituyen excelentes referencias sobre Rattle.

Modelos de ecuaciones estructurales y modelos jerárquicos lineales

Bajo los modelos de ecuaciones estructurales (SEM a partir de ahora, respondiendo a *Structural Equation Models*) y bajo los modelos jerárquicos lineales (MLM, respondiendo a *Multilevel Models* o HLM, *Hierarchical Linear Models*) se

esconde una lógica confirmatoria de investigación, en la que el investigador diseña un modelo, que no es más que una hipótesis, que contrasta con los datos recogidos. Se diferencian así de los métodos estadísticos de segmentación que responden a una lógica exploratoria al buscar empíricamente las variables predictoras que más diferencias producen en la variable de respuesta.

Este planteamiento de modelización muestra también una gran diferencia frente a los estudios experimentales clásicos, que no está únicamente en la naturaleza experimental o no del diseño o en su base correlacional. La diferencia fundamental reside en la naturaleza de las variables. Los SEM buscan causas que están definidas como variables latentes, que provienen de la elaboración de un constructo teórico, mientras que en los diseños experimentales clásicos el origen de los cambios está en una o varias variables directamente observables. Además, los MLM introducen la consideración del contexto en esta lógica confirmatoria.

Los SEM son un método global para la cuantificación y prueba de teorías sustantivas. Metodológicamente, son una colección de técnicas estadísticas que permiten establecer un conjunto de relaciones entre una o más variables predictoras (que pueden ser continuas o discretas) con una o más variables de respuesta. Ambos tipos de variables pueden ser variables de medida o directamente observadas o variables latentes o factores. Por tanto, son modelos que incorporan tanto variables latentes y como variables medidas. Las variables latentes son constructos teóricos hipotéticos de especial relevancia de los que no hay un procedimiento operativo para ser observados directamente. Sin embargo, las manifestaciones de los constructos latentes pueden observarse registrando medidas específicas de otras variables directamente observables. Los SEM pueden ser utilizados tanto para cuantificar la plausibilidad de una hipótesis teórica compleja expresada a través de las potenciales relaciones entre constructos como para probar las relaciones con las medidas que lo definen. Así, los SEM tienen en cuenta el error de medida de las distintas variables consideradas en el ámbito socio-educativo, lo que representa una sustancial diferencia con los métodos estadísticos multivariados (Raykov y Marcoulides, 2006). Se suelen representar por medio de un diagrama de las relaciones entre las variables y un sistema de ecuaciones que las formaliza y expresa su función metodológica dentro del modelo.

Los MLM representan un enfoque global sobre cómo debe analizarse la información, pues reconoce una estructura jerárquica o anidada en la estructura de los datos, especialmente en los de naturaleza socio-educativa. La jerarquía no responde a una agrupación conceptual o teórica de las respuestas, sino a una estructura de agrupación que podríamos llamar "física" o natural de los datos. Por ejemplo, los estudiantes se agrupan en clases, las clases en escuelas, las escuelas a su vez en distritos, en ciudades, comunidades autónomas, etc. Esta pertenencia a un grupo o a un contexto real hace que los sujetos de una agrupación compartan cierto número de influencias y que a su

vez se diferencien de otros individuos que están "bajo la influencia" de otro contexto con características diferentes. Hay por tanto cierto nivel de homogeneidad entre los individuos de un contexto y cierta heterogeneidad entre contextos.

Los MLM se han desarrollado para tratar adecuadamente los datos de naturaleza jerárquica o anidada. Esto permite dejar de desarrollar una ecuación de regresión específica para cada contexto (por ejemplo, para cada escuela), desarrollando una ecuación para los microcontextos y una ecuación para cada uno de los macrocontextos o niveles incluidos en el modelo, dejando variar los coeficientes de los microniveles en los macroniveles. Así se puede tratar la homogeneidad dentro de cada contexto y la heterogeneidad entre contextos con un modelo único que aborda la dificultad de tratar de forma diferenciada la variabilidad correspondiente a cada nivel.

Tal y como señalan Gaviria y Castro (2005), estos modelos proponen una estructura de análisis dentro de la cual se pueden reconocer los distintos niveles en que se articulan los datos, estando cada subnivel representado por su propio modelo. Cada uno de estos submodelos expresa la relación entre las variables dentro de un determinado nivel y especifica cómo las variables de ese nivel influyen en las relaciones que se establecen en otros niveles. Es decir, constituyen una estrategia analítica que permite la formulación jerárquica de las fuentes de variación y con capacidad para dar cuenta de esta estructura. El análisis multinivel es una metodología para el análisis de datos con patrones complejos de variabilidad, enfocada a fuentes anidadas de variabilidad.

Claramente, los MLM son la técnica de moda en educación por la cantidad de artículos de investigación publicados utilizando estos modelos. La justificación es más profunda que un simple tema de frecuencia de uso. Las ventajas técnicas de los modelos jerárquicos, son muchas. Draper (1995) señala que los modelos jerárquicos proporcionan un entorno natural en el que expresar y comparar las teorías acerca de las relaciones estructurales entre variables de cada uno de los niveles en una jerarquía organizativa o de muestreo; además de tener en cuenta en sus calibraciones la autocorrelación presente en los datos.

Tal y como explica y demuestra detalladamente Gaviria (2000), los MLM no son más que un caso particular de los SEM. Además no hay un cambio en los conceptos básicos de causalidad y de la naturaleza de las variables causales, latentes u observables. Por tanto, no hay un cambio epistemológico, sino más bien una continuidad en los planteamientos y un momento de establecimiento operativo. De ahí que, aun siendo modelos aparentemente tan distintos, se traten conjuntamente en este artículo.

Desde el punto de vista estadístico, los algoritmos y procedimientos de análisis tanto de los SEM como de los MLM se conocen desde principios del siglo XX. El primer antecedente de los SEM se remonta al año 1934 en el que Wright da a conocer los modelos de *path analysis* so-

bre las relaciones de tamaño de las mediciones óseas. El marco analítico de los modelos jerárquicos lineales está establecido desde que Fisher (1925) incorpora la posibilidad de analizar datos de naturaleza anidada en el marco del análisis de varianza.

En el contexto del Análisis Factorial, entendido como un antecedente de los SEM, se interpretaba la rotación como el intento de verificar empíricamente un constructo teórico. El impulso fundamental para el cambio de enfoque desde el exploratorio al confirmatorio vino con los trabajos de Jöreskog en el *Educational Testing Service* en 1967, al publicar un artículo que describía un algoritmo para la estimación por máxima verosimilitud de los parámetros del modelo de factores comunes y centró su atención en las *rotaciones procrustes*. Jöreskog (1969) incorporó al modelo de análisis factorial confirmatorio las ecuaciones lineales estructurales que se utilizaban ampliamente en economía. Esto dio lugar al modelo LISREL y al software asociado, que contribuyó decisivamente al uso intensivo y extensivo de estos modelos (Jöreskog, 1973, 1979).

En el caso de los modelos multinivel tienen que llegar Lindley y Smith, (1972) para formular de manera general el modelo jerárquico lineal. La estimación de estos modelos presentaba muy serias complicaciones, que no pudieron ser resueltas hasta que se introdujo el conocido *algoritmo EM* (algoritmo esperanza-maximización de Dempster, Laird y Rubin, 1977). Se usa en estadística para encontrar estimadores de máxima verosimilitud de parámetros que dependen de variables no observables. El algoritmo EM alterna pasos de esperanza (paso E), donde se computa la esperanza de la verosimilitud mediante la inclusión de variables latentes como si fueran observables, y un paso de maximización (paso M), donde se computan estimadores de máxima verosimilitud de los parámetros mediante la maximización de la verosimilitud esperada del paso E. Los parámetros que se encuentran en el paso M se usan para comenzar el paso E siguiente, y así el proceso se repite. Posteriormente se desarrollaron otros métodos de estimación para los MLM, como la Máxima verosimilitud Completa o Restringida usando el algoritmo Fisher scoring (Longford, 1987), los Mínimos Cuadrado Generalizados Iterativos (Goldstein, 1986) y otros basados en MCMC y Gibbs sampling (Smith y Roberts, 1993).

Debido a la complejidad matemática de la estimación y prueba de las relaciones mostradas en ambos modelos, la generalización del software estadístico ha sido un elemento fundamental para la difusión y uso de estas poderosas herramientas de investigación.

Hay numerosos programas para desarrollar los análisis necesarios para los SEM. Software como AMOS (Arbuckle & Wothke, 1999, recientemente incorporado en el paquete SPSS), EQS (Bentler, 2004), LISREL (Jöreskog & Sörbom, 1993a, 1993b, 1993c, 1999), Mplus (Muthén & Muthén, 2004), SAS PROC CALIS (SAS Institute, 1989), SEPATH (Statistica, 1998) y RAMONA (Browne & Mels, 2005) han contribuido grandemente a la

aplicación de estos modelos. Sin embargo, aunque estos programas tienen similares potencialidades son sobre todo LISREL y EQS los que históricamente han dominado el ámbito durante años (Marsh, Balla, & Hau, 1996). En los últimos años, Mplus también ha ganado popularidad entre los investigadores socio-educativos. Como pueden observar, desde la década de los 90 hay gran cantidad de software disponible para desarrollar y probar modelos de ecuaciones estructurales. LISREL dominó tanto el mercado que se ha llegado a confundir la metodología SEM con el paquete estadístico. En R, el paquete *sem* permite desarrollar este tipo de modelos. El trabajo de Fox (2006) constituye una buena introducción al respecto.

De igual manera, hay gran cantidad de software disponible para realizar MLM. En la actualidad la mayoría de los paquetes estadísticos generales (como SPSS o SAS) tienen su propio módulo para realizar este tipo de modelos, a través de los modelos mixtos. Entre estos paquetes generales, también destacaríamos el paquete WINBUGS *(http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml),* que permite ajustar un amplio conjunto de modelos a través de lenguaje de control estadístico. En el conjunto de paquetes para MLM es el paquete que permite desarrollar estimaciones bayesianas de los parámetros, que son de especial utilidad cuando se cuenta con poca información muestral.

Además una gran cantidad de paquetes estadísticos especializados, de los cuales destacan HLM *(www.ssicentral.com/hlm/)* desarrollado por Bryk, Raudenbush, y Congdon (1996, 2000, 2004) y el paquete MLwin *(http://www.bristol.ac.uk/cmm/software/mlwin/)* desarrollado inicialmente por Rabash, Goldstein, Browne, Yang y Woodhouse (2000). Ambos paquetes permiten tratar variables de respuesta continuas y discretas. HLM incorpora modelos de hasta tres niveles y estimaciones de máxima verosimilitud. MLwin también permite tratar modelos cruzados y modelos con cualquier número de niveles en la estructura. Los procedimientos de estimación son tanto máximo verosímiles como MCMC (*Markov Chain Monte Carlo methods*)

También los paquetes LISREL y MPLUS, dedicados inicialmente a los SEM, se utilizan para los MLM. Para una revisión completa del software para MLM recomendamos el capítulo 18 de Goldstein (2010). En R, los paquetes *multilevel* y *nlme* están dedicados a estos modelos.

Conclusiones

Y llega el momento de retomar la pregunta inicial, cuáles son las aportaciones de los últimos años en los procedimientos de modelización.

Desde el punto de vista metodológico, los procedimientos estadísticos de estimación para las modelos de segmentación, para los SEM y para los MLM están definidos desde hace tiempo atrás. Aunque el desarrollo de estos modelos supone la mejora de los procedimientos de estimación, la incorporación de relaciones no lineales entre las variables estudiadas o a la consideración de variables de respuesta con distribuciones distintas a la normal.

Para la aplicación de todos estos modelos, ha sido fundamental la capacidad computacional desarrollada desde mediados de los 80 dada la cantidad de datos que precisan y la complejidad de los procedimientos de estimación. Esta capacidad tecnológica ha producido mejoras fundamentalmente en la *velocidad de procesamiento*, haciendo posible la estimación de los modelos, en la capacidad del software para *incorporar algoritmos de estimación* verdaderamente complejos y en la *usabilidad de los programas*, permitiendo interfaces más amigables incluso para los no expertos estadísticamente hablando. El *software* para el tratamiento estadístico de datos cuantitativos muestra desde hace ya más de dos décadas una gran velocidad en el procesamiento de grandes cantidades de datos y variables, ha mejorado las condiciones de acceso permitiendo a los usuarios de herramientas de estadística aplicada el desarrollo de modelos complejos con interfaces fácilmente manejables. Además de la creciente disponibilidad de la tecnología.

El efecto en la investigación educativa ha sido una generalización en el uso de modelos estadísticos sofisticados como nunca hasta el momento. Hay por ejemplo revistas científicas específicas para SEM, como *Structural Equation Modeling,* y los MLM capitalizan prácticamente revistas como *Educational Evaluation and Policy Analysis*. Nos parece claro que el límite tecnológico para los usuarios que realizan investigaciones cuantitativas en educación es imperceptible, pues ha habido una gran evolución en los recursos y en la relaciones con esos recursos. Así, la mejora del *software* facilita la ejecución de procedimientos hasta ahora alejados, los resultados se producen y se difunden con mayor velocidad.

¿Dónde pues están los límites? Desde luego no parece que en la tecnología ni en la metodología. Los modelos estadísticos presentados en estas páginas son procedimientos útiles, sofisticados, preparados para el tratamiento de grandes cantidades de datos y de distintos tipos de variables. Los límites están donde siempre se encuentran en el desarrollo de las ciencias, y es en el desarrollo de teorías pedagógicas fundamentadas que articulen y vertebren la investigación educativa a realizar. El desarrollo tecnológico no habilita el conocimiento ni de los fenómenos educativos objeto de estudio ni de los modelos estadísticos que se utilizan cada vez con más facilidad.

¿Cabe pensar que la mejora de la capacidad de computación está modificando el nuestros hábitos de investigación en Ciencias Sociales? ¿La evolución de algoritmos ya programados y su generalización puede suponer un uso más *naif* de las técnicas estadísticas, volviendo a patrones más exploratorios y menos fundados en teorías y de óptica más confirmatoria?

Dirección para la correspondencia: María Castro Morera. Departamento de Métodos de Investigación y Diagnóstico en Educación. Facultad de Educación. Universidad Complutense de Madrid. C/ Rector Royo Villanova S/N. 28040 Madrid. E-mail: maria.castro@edu.ucm.es.

Fecha de recepción de la versión definitiva de este artículo: 20.IX.2011

Bibliografía

ATO, M. y LÓPEZ GARCÍA, J. J. (1996) *Análisis estadístico para datos categóricos* (Madrid, Síntesis).

BENTLER, P. M. (2004) *EQS structural equations program manual* (Encino, CA, Multivariate Software, Inc.).

BREIMAN, L. (2001) Random Forests, *Machine Learning* 45:1, pp. 5-32.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A. y STONE, C. J. (1984) *Classification and regression trees* (Belmont, California, Wadsworth).

BROWNE, M. W. y MELS, G. (2005) Path Analysis (RAMONA), en SYSTAT 11 *Statistics III [computer software and manual]* (Richmond, CA, SYSTAT Software Inc.), pp. III-1-61.

BRYK, A. S.; RAUDENBUSH, S. W. y CONGDON, R. (1996) *HLM 4 for Windows [Computer software]* (Chicago, IL, Scientific Software International, Inc.).

DEMPSTER, A. P.; LAIRD, N. M. y RUBIN, D. B. (1977) Maximum Likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, Series B, 39, pp. 1-38.

DOBSON, A. (1990) *An introduction to generalized linear models* (London, Chapman y Hall).

DRAPER, D. (1995) Inference and Hierarchical Modeling in the Social Sciences, *Journal of Educational and Behavioral Statistics*, 20, pp. 115-147.

ESTES, W. K. (1991) *Statistical models in behavioral research* (Hillsdale, NJ, Elbaum).

FISHER, R. A. (1935) *The design of experiments* (London, Oliver and Boyd).

FOX, J. (2006) Structural Equation Modeling with the sem Package, *Structural Equation Modeling*, 13:3, pp. 465-486

GAVIRIA, J. L. (2000) *Cambios en las técnicas cuantitativas de investigación socio-educativa*, Actas del XII Congreso Nacional y II Iberoamericano de Pedagogía, Madrid, pp. 39-45.

GAVIRIA, J. L. y CASTRO, M. (2005) *Modelos jerárquicos lineales* (Madrid, La Muralla).

GOLDSTEIN, H. (1986) Multilevel mixed linear models analysis using iterative generalized least squares, *Biometrika*, 73, pp. 43-56.

GOLDSTEIN, H. (2010) *Multilevel Statistical Models* (UK, Wiley, 4th Edition).

JÖRESKOG, K. G, y SÖRBOM, D. (1993c) *LISREL8: The SIMPLIS command language* (Chicago, IL, Scientific Software Inc.).

JÖRESKOG, K. G. y SÖRBOM, D. (1993a) *LISREL8: User's reference guide* (Chicago, IL, Scientific Software Inc.).

JÖRESKOG, K. G. y SÖRBOM, D. (1993b) *PRELIS2: A Preprocessor for LISREL* (Chicago, IL, Scientific Software Inc.).

JÖRESKOG, K. G. (1970) A general method for analysis of covariance structures, *Biometrika*, 57, pp. 239-251.

JÖRESKOG, K. G. (1971) Statistical analysis of sets of congeneric tests, *Psychometrika*, 36, pp. 109-133.

JÖRESKOG, K. G. (1973) A general method for estimating a linear structural equation system, en Goldberger, A. S. y Duncan, O. D. (eds.) *Structural Equation Modelling for the Social Sciences* (New York, Seminar Press).

JÖRESKOG, K. G. (1979) Statistical Models and Methods for analysis of longitudinal data, en JÖRESKOG, K. G. y SÖRBOM, D. *Advances in factor analysis and structural equation models* (Cambridge, MA., Abt Books).

JÖRESKOG, K. G. y SÖRBOM, D. (1999) *LISREL8.30: User's reference guide* (Chicago, IL, Scientific Software Inc.).

JUDD, C. M. y MCCLELLAND, G. H. (1989) *Data analysis: a model comparison approach* (San Diego, CA, Hartcourt, Brace and Jovanovich).

JUDD, C. M.; MCCLELLAND, G. H. y CULHANE, S. E. (1995) Data analysis: continuing issues in the everyday analysis of psychological data, *Annual Review of Psychology*, 46, pp. 433-465.

KASS, G. (1980) An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, 29: 2, pp. 119-127.

KIM, S. H. (1991) *An extension of CART's Pruning Algorithm*. Program Statistics Research Technical Report Nº 91-11 (Princeton, New Jersey. Educational Testing Service).

LINDLEY, D. V. y SMITH, A. F. M. (1972) Bayes estimates for the linear model, *Journal of the Royal Statistical Society*, Series B, 34, pp. 1-41.

LIZASOAIN, L. y JOARISTI, L. (2010) Estudio Diferencial del Rendimiento Académico en Lengua Española de Estudiantes de Educación Secundaria de Baja California (Mé-

xico), *Revista Iberoamericana de Evaluación Educativa*, 3:3, pp. 115-134. Ver http://www.rinace.net/riee/numeros/vol3-num3/art6.pdf (Consultado el 15.XII.2010).

LONGFORD, N. T. (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects, *Biometrika*, 74, pp. 817-827.

LUNNEBORG, C. E. (1994) *Modeling experimental and observational data* (Belmont, CA, Duxbury Press).

MARSH, H. W.; BALLA, J. R. y HAU, K.-T. (1996) An evaluation of incremental fit indices: A clarification of mathematical and empirical properties, en MARCOULIDES, G. A. y SCHUMACKER, R. E. (eds.) *Advanced structural equation modeling: Issues and technique* (Hillsdale, NJ, Lawrence Erlbaum Associates), pp. 315-353.

MUTHÉN, B.O. y MUTHÉN, L. (2004) *Mplus User's guide* (Los Angeles, CA, Muthén & Muthén).

RASBASH, J.; GOLDSTEIN, H.; BROWNE, W.; YANG, M. y WOODHOUSE, G. (2000) *The MLwiN Command Interface* (Bristol, University of Bristol, Centre for Multilevel Modeling).

RAUDENBUSH, S. W.; BRYK, A. S. y CONGDON, R. (2000) *HLM 5 for Windows [Computer software]* (Lincolnwood, IL, Scientific Software International, Inc.).

RAUDENBUSH, S. W.; BRYK, A. S. y CONGDON, R. (2004) *HLM 6 for Windows [Computer software]* (Lincolnwood, IL, Scientific Software International, Inc.).

RAYKOV, T. y MARCOULIDES, G. A. (2006) *A first course in structural equation modeling* (New Jersey, LEA).

SAS INSTITUTE (1989) *SAS PROC CALIS User's guide* (Cary, NC, Author).

SMITH, A. F. M. y ROBERTS, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov-chain Monte Carlo methods, *Journal of the Royal Statistical Society*, Series B, 55, pp. 3-23.

STATISTICA (1998) *User's guide* (Tulsa, OK, Statistica Inc).

STROBL, C.; MALLEY, J. y TUTZ, G. (2009) An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests, *Psychological Methods*, 14:4, pp. 323-348.

TORGO, L. (2010) *Data Mining with R: Learning with Case Studies* (London, Chapman & Hall).

VANDAMME, J.-P.; MESKENS, N. y SUPERBY, J. F. (2007) Predicting Academic Performance by Data Mining Methods, *Education Economics*, 15:4, pp. 405-419.

WILLIAMS, G. (2009) Rattle: A Data Mining GUI for R, *The R Journal*, 1:2, pp. 45-55.

WILLIAMS, G. (2011) *Data Mining with R and Rattle: The Art of Excavating Data for Knowledge Discovery* (Springer).

Resumen:
Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales

La investigación educativa ha avanzado planteamientos teóricos al definir y diseñar modelos integrados que quieren estudiar las relaciones entre constructos no directamente observables insertos en contextos complejos. De manera paralela, el *software* para el análisis de datos que ha permitido someter a prueba estos modelos complejos en entornos computacionales razonables en términos de tiempo y esfuerzo. Esta revolución tecno-científica en el desarrollo del *software* para el análisis de datos cuantitativos ha transformado las prácticas científicas en el ámbito de la educación.

En este artículo asumimos la tendencia creciente en investigación educativa al planteamiento y prueba de modelos comprehensivos sobre los fenómenos educativos. De manera más específica nos centraremos en los árboles estadísticos de decisión, en los modelos de ecuaciones estructurales y en los modelos multinivel, que vienen dominando el panorama de la

investigación educativa desde mediados de los 70 y que gracias al desarrollo tecnológico se han podido generalizar recientemente. El efecto en la investigación educativo ha sido una generalización en el uso de modelos estadísticos sofisticados como nunca hasta el momento. Hace falta una reflexión sobre cuáles son entonces los modelos fundamentados teóricamente que queremos probar con herramientas estadísticas tan potentes como las descritas en este artículo.

Descriptores: Modelización estadística/Minería de datos/Árboles de decisión/Modelos jerárquicos lineales/Modelos Multinivel/Modelos de ecuaciones estructurales.

Summary:
Statistical Modeling Techniques in Educational Research: Data Mining, Structural Equation Models and Hierarchical Linear Models

Educational research has advanced theoretical approaches by defining and designing integrated models oriented up to the study of relationships between constructs not directly observable embedded in complex contexts. In parallel, data analysis software has allowed testing these complex models in reasonable computing environments in terms of time and effort. This techno-scientific revolution in the developed software for quantitative data analysis has transformed scientific practice in educational research.

In this paper is assumed the growing trend in educational research on design and fit of comprehensive models of educational phenomena. More specifically we will focus on the statistical decision trees, in structural equation models and multilevel models, which come to dominate the landscape of educational research since the mid 70's and have recently been generalized thanks to technological development. The impact on educational research has been a widespread use of sophisticated statistical models than ever so far. It is necessary to think about what theoretically grounded models have to be tested with statistical tools as powerful as those being described in this paper.

Key Words: Statistical Modelling/Data Mining/Decision Trees/ Hierarchical Linear Models/HLM/ Multilevel Models/-MLM/Structural Equation Models/SEM.