



Revista Espiga

ISSN: 1409-4002

revistaespiga@uned.ac.cr

Universidad Estatal a Distancia

Costa Rica

Rojas Valenciano, Ligia

Análisis de niveles múltiples: una herramienta eficaz para realizar estudios sobre factores asociados en Educación y Ciencias Sociales

Revista Espiga, núm. 11, enero-junio, 2005, pp. 101-126

Universidad Estatal a Distancia

San Pedro de Montes de Oca, Costa Rica

Disponible en: <http://www.redalyc.org/articulo.oa?id=467846084008>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

# **Análisis de niveles múltiples: una herramienta eficaz para realizar estudios sobre factores asociados en Educación y Ciencias Sociales**

*Ligia Rojas Valenciano*

## **RESUMEN**

*En los estudios socioeducativos con frecuencia surgen inquietudes metodológicas en términos de cómo investigar las relaciones entre los individuos y la sociedad. Esto se debe a que los individuos interactúan con el contexto social al cual pertenecen y como resultado se ven influidos por el medio en que viven. Los individuos y los diversos grupos sociales a que pertenecen se encuentran agrupados en un sistema jerárquico definido y separado por niveles.*

\* Profesora Adjunta de la Universidad de Costa Rica, Doctora en Educación, Máster en Evaluación Educativa y Licenciada en Salud de la Mujer y Perinatología.

*Rec. 15-06-2005. Aprob. 20-02-2006.*

*Por tal motivo, los investigadores que intentan explicar comportamientos individuales, han visto la necesidad de estudiar conjuntamente las variables que describen tanto a los individuos como a su contexto; este tipo de investigación es conocida como investigación multinivel o de niveles múltiples. Es un modelo de análisis estadístico alternativo que puede ser aplicado a estudios que pretenden explicar factores asociados tanto en el área de educación como en las ciencias sociales.*

## **DESCRIPTORES**

*Regresión múltiple, análisis multinivel, regresión multinivel, investigación cuantitativa, investigación multinivel.*

## **ABSTRACT**

*In socio-educative studies, methodological concerns as to how to investigate the relationships between the individuals and society frequently arise. This is due to the fact that individuals interact with the social context to which they belong to and as result they are influenced by the environment in which they live. The individuals and the diverse social groups to which they belong to are grouped in a hierarchical system defined and separated by levels.*

*Therefore, researchers trying to explain individual behaviours have felt the need of studying the variables that describe both the individuals and their context. This type of investigation is known as multi-level or of multiple levels. It is an alternative model of statistic analysis that can be applied to studies that try to explain factors associated to education and social sciences.*

## KEY WORDS

*Multiple regression, Multilevel analysis, Multilevel regression, Quantitative investigation, Multilevel investigation.*

## INTRODUCCIÓN

En las ciencias sociales y educativas es común buscar explicaciones para los fenómenos del comportamiento. Cuando se habla de “explicar” el comportamiento, se hace referencia a que existe un planteamiento causal subyacente que permite identificar factores que influyen sobre el fenómeno que se estudia.

Uno de los enfoques cuantitativos más utilizados para estudiar esta temática es el análisis de regresión múltiple. Por ejemplo, en el campo de la educación, interesa explicar dimensiones de la enseñanza y el aprendizaje tales como, rendimiento académico, repitencia y deserción, entre otros. En ciencias de la salud interesa explicar entre otras, dimensiones epidemiológicas; tales como la incidencia y prevalencia de ciertas enfermedades, comportamientos de autocuidado y preventivos en las personas.

Desde el punto de vista metodológico, tradicionalmente ha sido difícil establecer evidencias empíricas para la causalidad. Así, los llamados “estudios de factores asociados” se refieren a un enfoque que utilizando metodologías cuantita-

tivas, principalmente modelos de regresión múltiple, permite evidenciar empíricamente asociaciones entre los factores y la variable que se va a explicar.

En los modelos clásicos de regresión múltiple no se puede analizar simultáneamente el efecto de variables independientes en diferentes niveles de agregación. Por ejemplo, con respecto al rendimiento académico, se podría hipotetizar que el desempeño del estudiante puede depender de factores propios del individuo, como es el caso de la condición socioeconómica, motivación para el estudio y nivel educativo de los padres o encargados. Pero también su rendimiento puede depender de variables contextuales referidas al centro educativo; tales como, infraestructura, tamaño de los grupos y estrategias pedagógicas de los y las docentes. Inclusive, podría pensarse que su rendimiento está también determinado por variables del contexto de la comunidad en que viven, como condición urbana-rural, acceso a los servicios básicos, aspectos de seguridad, entre otros.

Con el instrumental clásico de regresión, si se quieren estudiar los factores antes mencionados como determinantes del rendimiento se tendría que estimar separadamente tres ecuaciones de regresión: una en donde se incluyen variables

individuales, otra con las variables contextuales referidas al centro educativo y otra con las contextuales referidas a la comunidad.

Por el contrario, en los modelos de niveles múltiples, se permite en un mismo modelo de análisis incorporar no sólo variables individuales, sino también variables del contexto asociadas a la comunidad y al centro educativo. Por tal motivo, este modelo es más poderoso que la regresión clásica para el estudio de los factores asociados.

La aplicación de estos modelos, involucra procedimientos matemáticos y estadísticos de cierta complejidad, que requieren de conocimiento sustantivo y metodológico por parte del investigador e investigadora. Sin embargo, actualmente existen a disposición de ellos y ellas, diversos paquetes informáticos que han facilitado la aplicación del modelo, y por ende, contribuyen a su divulgación.

Este artículo tiene un carácter introductorio sobre los aspectos teóricos, metodológicos y de aplicación práctica del modelo de análisis de niveles múltiples.

## Justificación

Las investigaciones en las Ciencias Sociales y educativas generalmente se refieren al estudio de problemas

en donde existe una relación entre los individuos y la sociedad. Con esto se quiere decir que los individuos interactúan con el contexto social; por lo tanto, las personas (individuos) están influenciadas por los grupos sociales a los cuales pertenecen. Los individuos (unidades del nivel 1) viven y se desarrollan en diversos grupos organizados jerárquicamente (unidades de nivel 2 y niveles superiores) (Goldstein, 1987:1). Una ilustración de lo anterior es la siguiente: en una estructura de dos niveles, puede ser que las unidades del nivel 1 sean estudiantes, quienes están agrupados dentro de las escuelas, las cuales conforman las unidades del nivel 2.

Como consecuencia de lo expuesto anteriormente, las y los investigadores se vieron interesados en explicar o analizar fenómenos en donde simultáneamente se incluyen variables de los individuos (nivel 1), con variables descritas para los grupos sociales (niveles superiores). A este tipo de investigación se le llama "*estudios de niveles múltiples*".

El término "*niveles múltiples*", es conocido también como análisis de "*multinivel*"; es decir, se refiere a una estructura de datos agrupados. El término es usado genéricamente para todos los modelos que presentan la característica antes mencionada.

Por este motivo, se propone

...una estructura de análisis dentro de la cual se pueden reconocer los distintos niveles en que se articulan los datos, estando cada nivel representado por su propio modelo. Cada uno de estos modelos 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. (Murillo,1999:3).

A mediados de los años 80, algunos investigadores comenzaron a introducir aproximaciones sistemáticas a los modelos estadísticos para analizar los datos estructurados jerárquicamente. Los trabajos de Aitkin y otros en el año 1981, sobre los datos en los estilos de enseñanza y trabajos subsecuentes con Longford en 1986, dieron paso en los años 90 al desarrollo de técnicas establecidas, experiencias y paquetes informáticos para aplicar diferentes modelos de niveles múltiples en áreas como educación, epidemiología, geografía, salud, entre otros.

Los modelos de niveles múltiples, han sido desarrollados como una respuesta a las necesidades metodológicas en investigación para

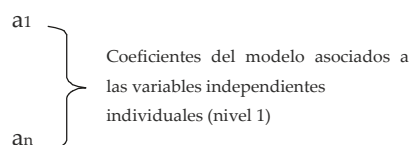
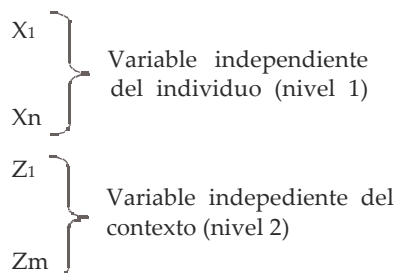
lograr analizar integralmente fenómenos del comportamiento humano.

Estos modelos permiten distinguir con precisión los efectos de las variables representadas en un nivel, y ofrecen, a diferencia de los modelos clásicos de regresión, la posibilidad de recolectar la estructura anidada de los datos en sus niveles.

Entonces, un modelo de niveles múltiples puede ser expresado como:

$$Y = (a_1X_1 + a_2X_2 + \dots + a_nX_n) + (b_1Z_1 + b_2Z_2 + \dots + b_mZ_m); \text{ en donde:}$$

Y = Variable dependiente



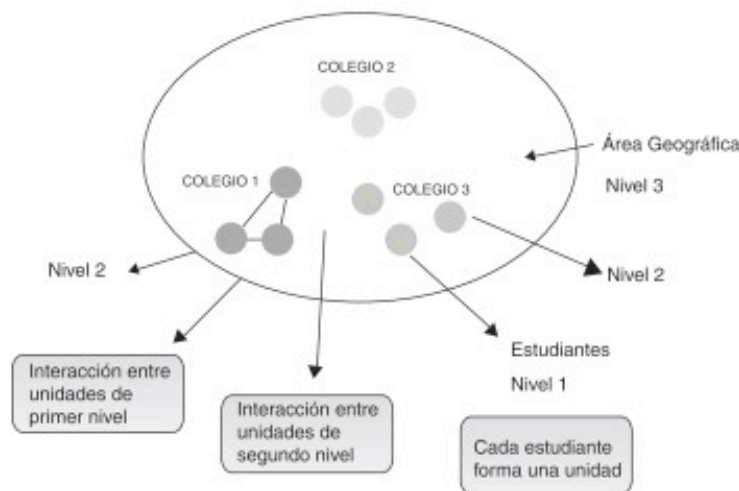
$b_1$   
 $b_m$  } Coeficientes del modelo asociados a las variables del contexto (nivel 2)

Los coeficientes  $a_1 \dots a_n$  y  $b_1 \dots b_m$ , representan los pesos o ponderaciones de las correspondientes variables independientes  $X_1 \dots X_n$  y  $Z_1 \dots Z_m$ . El criterio fundamental que guía el procedimiento de estimación de estos valores en una ecuación de regresión específica es la minimización de los residuos o errores de estimación, definiendo éstos como la diferencia entre los valores observados de la variable dependiente y los valores estimados con base en el modelo de regresión. Cuando  $a_i$  es positivo existe una relación directa entre  $X_i$  y  $Y$  (cuando aumenta una, aumen-

ta la otra), manteniendo constantes o controlando las otras variables independientes en el modelo. Por el contrario, si el coeficiente de regresión  $a_i$  es negativo la relación entre  $X_i$  y  $Y$  es inversa, es decir, al aumentar una disminuye la otra (Montero, 2001:sp).

Por ejemplo, en educación las áreas geográficas forman un nivel (nivel 3); los colegios en cada área corresponden al nivel 2; y los estudiantes en cada colegio conforman las unidades de primer nivel 1 (individual); obsérvese el Esquema 1.

Otro ejemplo de una estructura jerárquica es cuando se desea estudiar aspectos de salud en individuos que asisten a centros de salud dentro de comunidades. La



ESQUEMA: Ilustración hipotética de un modelo de tres niveles

comunidad, en este caso, corresponde a un tercer nivel, en donde se van a medir diversas variables como pueden ser condición urbana o rural, densidad de población, disponibilidad de los servicios, entre otros. Los centros de salud se ubican en un segundo nivel de jerarquización, en donde se van a medir variables tales como: recursos materiales, recursos humanos, nivel de capacitación del personal, entre otros. Finalmente, un primer

nivel lo conforman las personas que asisten a dicho centro de salud, donde se pueden estudiar variables sociodemográficas tales como: sexo, edad, condición socioeconómica, y estado de salud, entre otras. Véase una aplicación real de estos modelos en el Cuadro 1.

En el Cuadro 1, se analiza un modelo de dos niveles en donde se presentan las siguientes variables individuales (nivel 1)

CUADRO 1

MODELO FINAL PARA EL ANÁLISIS DE REGRESIÓN ( $R^2=0,1513$ )

satisind	Coef.	Robust Std. Err.	t	P> t	95% Conf.	Interval
avail	0.414239	0.215906	1.919	0.058	-0.01476	0.843239
user	0.249008	0.104712	2.378	0.02	0.040947	0.457069
knowstd	0.139884	0.100122	1.397	0.166	-0.05906	0.338823
ed1	-0.56417	0.340508	-1.657	0.101	-1.24076	0.112409
ed2	-0.25165	0.356383	-0.706	0.482	-0.95978	0.456472
ed3	-0.42418	0.336999	-1.259	0.211	-1.09379	0.245428
ed4	-0.5204	0.365628	-1.423	0.158	-1.24689	0.206099
ed5	-0.32197	0.350991	-0.917	0.361	-1.01938	0.375439
age1	0.031628	0.287841	0.11	0.913	-0.54031	0.603561
age2	-0.08044	0.276076	-0.291	0.771	-0.629	0.468119
age3	-0.09101	0.275734	-0.33	0.742	-0.63889	0.456867
age4	-0.03482	0.260002	-0.134	0.894	-0.55144	0.481795
age5	0.134588	0.286684	0.469	0.64	-0.43505	0.704223
age6	-0.21517	0.287421	-0.749	0.456	-0.78627	0.355932
ntype1	-1.13944	0.284544	-4.004	0	-1.70482	-0.57405
ntype4	-1.0963	0.225305	-4.866	0	-1.54397	-0.64862
ntype2	-1.20996	0.212581	-5.692	0	-1.63235	-0.78756
ntype5	-0.88278	0.293329	-3.01	0.003	-1.46562	-0.29994
staff3	0.11787	0.12125	0.972	0.334	-0.12305	0.358791
natgui	0.257662	0.174642	1.475	0.144	-0.08935	0.604671
broch	0.268125	0.117727	2.278	0.025	0.034203	0.502046
plmij	-0.03574	0.107517	-0.332	0.74	-0.24937	0.177898
condom	-0.37222	0.131263	-2.836	0.006	-0.63303	-0.1114

FUENTE: Elaboración de la autora con base en fuentes inéditas.

Nota: Ilustración de una aplicación de niveles múltiples. Variable dependiente: índice de satisfacción con los servicios de Planificación Familiar



**User** = usuaria, 1=Nueva, 0=antigua

**Knowstd** = conocimiento de las clientes acerca de los métodos.

**Ed 1, 2, 3, 4, 5** = variables dummy asociadas a las categorías de educación.

**Age 1, 2, 3, 4, 5, 6** = variables dummy asociadas a las categorías de edad.

**Pilinj, condom** = variables dummy asociadas al tipo de método de planificación familiar brindado a la cliente.

Las variables referentes al centro de salud (nivel 2) son las siguientes:

**Avail** = Disponibilidad de métodos: 1 = si el establecimiento provee al menos 3 métodos modernos de planificación familiar, 0 = otro caso.

**Ntype 1, 2, 4, 5** = variables dummy asociadas al tipo de centro de salud. (el grupo de comparación o grupo excluido son los hospitales privados)

**Staff3** = Disponibilidad de personal: 1 = si el establecimiento tiene al menos 3 personas asignadas para dar el servicio de planificación familiar y al menos 2 de ellos presentes en el momento de la consulta, 0 = otro caso.

**Natgui** = Existencia en el establecimiento de Normas de Atención en planificación familiar: 1= Si, 0 = No.

**Broch** = Disponibilidad de materiales educativos: 1= Si el establecimiento provee a sus clientes con al menos tres panfletos sobre métodos modernos, 0 = otro caso.

Por ejemplo, en la variable individual usuaria (user), se codificó con 1 a las usuarias nuevas y 0 a las antiguas; el coeficiente asociado de esta variable es igual a 0,249 con una probabilidad asociada de 0,02; lo cual indica que hay una relación significativa entre esta variable y la satisfacción con los servicios dados. Cuando el coeficiente es positivo se concluye que las usuarias nuevas tienen en promedio mayor satisfacción que las antiguas.

Dentro de las variables que corresponden al centro de salud, se encuentra el tipo de establecimiento (ntype 1, 2, 3, 4, 5); en donde, el 1 corresponde al hospital del Ministerio de Salud; el 2 a la Clínica de Planificación Familiar o Clínica de Salud Materno Infantil, adscrita también al Ministerio de Salud; el 3 es el Centro de Salud del Ministerio de Salud; el 4 hospital del Seguro Social y el 5 hospital privado. A partir de esta clasificación se formaron cuatro variables dicotómicas (0-1); en donde se comparó cada uno de los establecimientos públicos con los hospitales privados. Los coeficientes asociados a las variables presentan un valor negativo y estadísticamente significativo, dado que su probabilidad asociada es menor de 0,05. Por lo tanto, indica que en promedio las usuarias de estable-



cimientos públicos tienen menos satisfacción que las que asisten a hospitales privados.

En Costa Rica, por ejemplo, en educación, la aplicación de los modelos de análisis de niveles múltiples son prácticamente desconocidos y son pocos los estudios realizados utilizando las técnicas de análisis estadísticas tradicionales de regresión múltiple. Los análisis estadísticos más comunes son de naturaleza descriptiva y no es muy común que se apliquen procedimientos multivariados. Sin embargo, se debe mencionar que en el área de la salud, se cuenta con algo más de experiencia en el uso de técnicas multivariadas.

## Antecedentes

Existen diversos estudios que basan su metodología en modelos de niveles múltiples, que han sido desarrollados principalmente en Europa, Estados Unidos y Australia. No obstante, se encontró un estudio realizado en Colombia titulado: *Los insumos escolares en la educación secundaria y su efecto sobre el rendimiento académico de los estudiantes: un estudio en Colombia*, realizado por Piñeros Jiménez y Rodríguez Pinzón en 1998, en donde se aplica el modelo multinivel de dos niveles. En síntesis, el estudio pretendió medir el impac-

to de algunas características de las y los alumnos y de los establecimientos educativos que tienen sobre el rendimiento alcanzado por los estudiantes que terminaron el ciclo de educación media en diciembre de 1997.

Se han realizado otros estudios en diferentes campos; por ejemplo, la revista *Social Science of Medicine* (1993), publica un artículo escrito por Duncan, Jones y Moon en Inglaterra. Los autores argumentan que a través de un estudio de análisis de niveles múltiples se pudo determinar las diferencias regionales a escala nacional sobre la conducta del fumado y la bebida. La conducta ha sido entendida no solamente en términos de características individuales sino también relacionada con la cultura y factores del contexto.

En un análisis de niveles múltiples, aplicado a un estudio sobre los efectos del sistema de remuneración en el comportamiento y actitudes de practicantes generales de medicina, el cual compara las consultas médicas en instituciones de salud y las visitas médicas domiciliarias, realizado por Kristiansen y Holtedahl, publicado en 1993 por el *Journal of Epidemiology and Community Health*, se determinó que los médicos parecen estar influenciados por la remu-

neración monetaria. Aunque las visitas domiciliarias son parte de las características tradicionales, culturales y organizacionales, se concluyó que el incentivo monetario puede ser usado para cambiar la conducta y el valor de las visitas domiciliarias.

Finalmente, se presenta un estudio realizado por Jones, Moon y Clegg (1991) en el Politécnico de Portsmouth en Buckingham, titulado *Efectos ecológicos e individuales en la inmunización en niños(as): una aproximación de niveles múltiples*. Se analiza la inmunización realizada a escala individual tomando en cuenta el medio ambiente (ecológico), y los problemas que provienen desde distintas estrategias. La aproximación potencial del modelo toma en cuenta simultáneamente el proceso de ambos niveles (ecológico-individual).

Una cantidad considerable de artículos, investigaciones, entre ellos los editados por la Universidad de Southampton y la Universidad de Carolina del Norte en Chapel Hill, Estados Unidos en 1997; la revista *Multilevel Modelling Newsletter* a cargo del Instituto de Educación de la Universidad de Londres, Inglaterra en el 2001, investigaciones sobre diversos temas, utilizando el análisis de "multinivel", expuestos en la Tercera Conferencia

de Ámsterdam, en abril del 2001; entre otros, reseñan la aplicación de los modelos de análisis de niveles múltiples, y avalan esta sistematización estadística por su potencial para explicar la relación entre variables o unidades que conforman los diferentes niveles de jerarquía.

## Algunos aspectos sobre la teoría

La autora considera importante mencionar algunos conceptos básicos sobre la regresión múltiple, técnica que, adaptada a los modelos de niveles múltiples, se convierte en un aspecto teórico fundamental.

### Regresión múltiple

Este apartado está dedicado a la introducción de algunos conceptos básicos sobre la regresión múltiple. Puede ser que para el lector la connotación del nombre "*regresión múltiple*" sea impresionante; sin embargo, los investigadores sociales pueden llegar a una amplia comprensión de la regresión y su aplicabilidad sin penetrar profundamente en complejas ecuaciones matemáticas.

En un modelo de regresión múltiple se construye una ecuación en donde la variable dependiente o criterio (Y), aquella que interesa

predecir o estimar, se concibe como una función lineal de un conjunto de variables independientes<sup>2</sup> o predictoras ( $X_1, X_2, X_3, \dots, X_n$ ). Esta función es llamada ecuación de regresión y matemáticamente se puede expresar así:

$$Y = f(X_1, X_2, X_3, \dots, X_n) = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n$$

Se parte del supuesto de que el nivel o valor que observamos en la variable dependiente es el resultado de la influencia simultánea de varios factores. Las variables independientes son estos factores relevantes. Así, se busca que el conjunto de variables independientes permita predecir, con un grado razonable de precisión, el comportamiento de la variable dependiente (Montero, 2001:sp).

Por lo tanto, se pretende determinar cómo y en qué medida las variaciones provocadas en las variables independientes conducen a variaciones en la variable dependiente.

Otro propósito de la regresión múltiple es el de predecir, esto no implica necesariamente algu-

na deducción hacia el futuro, sino que se refiere al hecho de que la información acerca de una variable, puede ser utilizada para obtener información respecto a otra. Por consiguiente, la regresión múltiple pretende estimar el porcentaje de oscilación de la variable dependiente que puede ser explicado a partir de sus relaciones con las variables independientes en estudio.

Como ejemplo, se asume una variable dependiente "*logro académico*" y las variables predictoras o independientes "*condición socioeconómica, sexo y habilidad previa*";<sup>3</sup> por consiguiente, éstas últimas van a explicar o predecir la variable "*logro académico*".

Obsérvese el gráfico hipotético 1; el cual ilustra una relación de regresión lineal. En este caso corresponde a una regresión simple porque intervienen solamente dos variables. No obstante, los fundamentos para la regresión múltiple son similares, la diferencia estriba en que en este último se incluyen más de dos variables independientes.

El gráfico hipotético 1, corresponde al resultado del logro acadé-

1. Se refiere a la característica que se pretende explicar; también se le llama criterio.

2. Cuando se comparan varios grupos o situaciones respecto a una característica dada; estos grupos o situaciones constituyen la variable independiente (o experimental).

3. Se entiende por "habilidad previa", todas aquellas habilidades que posee un(a) estudiante antes de iniciar un aprendizaje determinado.

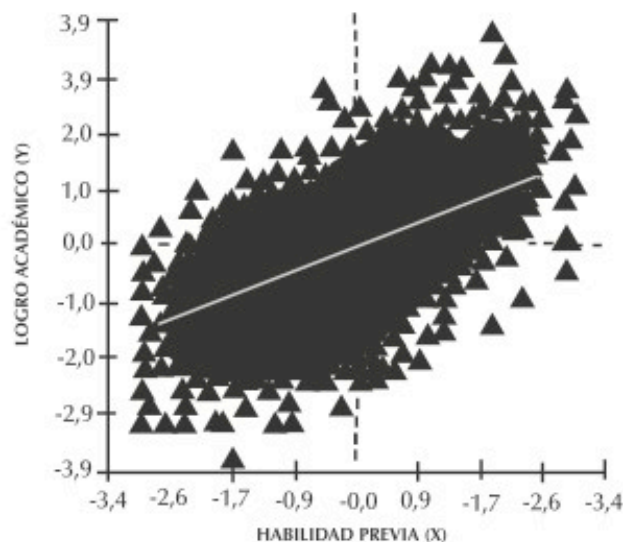


GRAFICO 1: Ilustración hipotética de una relación de regresión lineal simple.

mico y la habilidad previa de una muestra de estudiantes. La línea roja (recta de regresión) evidencia que el porcentaje del incremento en la habilidad está asociado con el incremento en el logro académico.

La recta de regresión representa el mejor fundamento para predecir valores de Y a partir de valores conocidos de X. Otro aspecto importante de señalar en el gráfico anterior, es que no todos los puntos que representan el valor observado para cada caso, se encuentran sobre la recta de regresión. A la diferencia entre el valor observado y el valor obtenido a partir de la

recta de regresión se le llama “*desviación*”; las desviaciones representan los errores de la predicción.

Las desviaciones con respecto a la recta representan los errores en la predicción. Si se suman las desviaciones de los valores respecto a la recta se tiene una base para calcular la varianza y la desviación estándar de los errores de predicción. A parte, a la varianza con respecto a la recta se le conoce como “*varianza residual*” y a la desviación estándar con respecto a la recta de regresión se le conoce como el “*error estándar de estimación*”.

De lo anterior se afirma que cuanto más pequeño sea el error estándar de la estimación, tanto menor será la dispersión de los puntos respecto a la recta de regresión.

Para cada uno de los conceptos antes mencionados se aplican fórmulas estadísticas específicas; sin embargo, no es el fin de este artículo de que el lector profundice en cada una de ellas; todo lo contrario, dar a conocer que existe una variedad de conceptos vinculantes en una regresión múltiple y ésta como fundamento teórico para los diferentes modelos de análisis de niveles múltiples.

Obsérvese en el gráfico hipotético 2 que muestra la representación en dos dimensiones del resultado de la aplicación de una regresión múltiple con dos variables independientes.

### Consideraciones sobre los modelos de niveles múltiples

A juicio de la autora y de acuerdo con la literatura consultada, parece existir un mayor desarrollo en lo referente a los paquetes computacionales que ayudan al procesamiento de los datos, comparado con la investigación en torno a la teoría subyacente en los modelos de niveles múltiples (J.J. Hox, 1995:120).

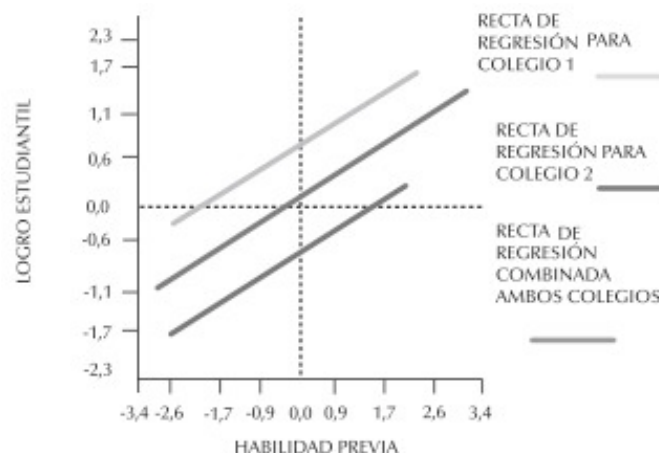


GRAFICO 2: Ilustración en dos dimensiones de una representación hipotética de los datos a partir de una regresión múltiple con dos variables independientes (habilidad previa y colegio de procedencia).

Sin embargo, la teoría señala que uno de los aspectos más importantes para utilizar este tipo de modelos es tener un criterio de agrupación lo más claro posible, de tal manera que las variables puedan ser asignadas, adecuadamente, a los niveles correspondientes.

No obstante, se puede enfrentar a un agrupamiento en donde los límites algunas veces no son tan claros y pueden tener una connotación de arbitrarios; esto quiere decir que la asignación de variables no siempre va a ser obvia ni simple. Por lo tanto, la decisión de ubicar una variable en un nivel y no en otro va a depender de las suposiciones teoréticas y del objeto de estudio. Por ejemplo, si se desea estudiar áreas geográficas para conocer la calidad y cantidad de colegios que en ellas se encuentran, el área geográfica correspondería a un segundo nivel (nivel alto), y los colegios a un primer nivel (nivel bajo). Pero si el interés investigativo reside más bien en predecir el efecto de la calidad educativa que brindan esos colegios en población estudiantil de un área geográfica; los colegios vienen a conformar un segundo nivel (nivel alto), y la población estudiantil corresponderá a un primer nivel de jerarquía (nivel bajo).

Cuando existe un gran número de variables en diferentes niveles de jerarquización, se encuentran, de la misma manera, numerosas posibilidades de interacción en el cruce de los niveles. De esta forma, la teoría especificará cuáles variables pertenecen a cuál nivel y los efectos esperados en las interacciones de éstos.

Un ejemplo de lo anterior puede mostrarse en una estructura jerárquica simple es las personas agrupadas en familias. Esta estructura tiene dos niveles: las personas corresponden a un nivel individual (nivel bajo), y ellas están agrupadas dentro de las familias en un segundo nivel (nivel alto). Las interacciones en el cruce de niveles pueden ser diversas y difieren en las familias, entre las cuales se encuentran características sociales, ambientales, predisposición genética, entre otros aspectos. Otros ejemplos son estudiantes en escuelas, pacientes en hospitales, entre otros.

Es importante resaltar que una de las fortalezas de los modelos de niveles múltiples estriba en que no requiere de un dato balanceado; o sea, con respecto al ejemplo antes mencionado, el número de personas en cada familia no tiene que ser el mismo. Obsérvese la Figura 1.

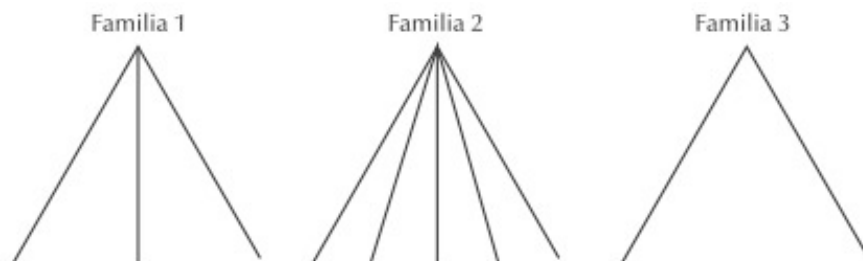


FIGURA 1: Diferencia en el número de personas que conforman diferentes familias en un estudio hipotético de niveles múltiples.

Cabe señalar, sin embargo, que algunos investigadores han descuidado ciertos aspectos teóricos muy importantes, que le pueden restar confiabilidad y validez al modelo; entre estos se pueden mencionar los relacionados con la influencia por parte de los individuos sobre el grupo conformado y la influencia de los individuos sobre la sociedad. Estos aspectos dejan de ser un riesgo cuando se organizan las variables en los niveles adecuados y se elige el modelo pertinente para el estudio.

Por otro lado, en las investigaciones socioeducativas a menudo existen grandes inconvenientes para relacionar individuos y sociedad; debido a que los individuos interactúan con el contexto social al que pertenecen, esto significa que las personas están influenciadas por el grupo social y las pro-

piedades de esos grupos sociales están a la vez influenciadas por los individuos quienes la conforman.

Para dar respuesta a lo anterior, los individuos y los grupos sociales se conceptualizan como un sistema jerárquico de individuos y grupos, definidos en niveles separados de este sistema jerárquico. Esto conduce a investigar la interacción entre variables que se desean explicar en los individuos y en los grupos sociales. De ahí que también se conoce al modelo de análisis de niveles múltiples, como análisis contextual.

Existen diferentes tipos de modelos de análisis de niveles múltiples, los cuales son aplicados según las necesidades metodológicas del estudio; a continuación se señalan algunos de ellos.



## Tipos de modelos de análisis de niveles múltiples

Existe una gran variedad de modelos de análisis de niveles múltiples, algunos de los cuales son el modelo de regresión de nivel múltiple, el modelo de nivel múltiple para estructuras de covarianza, entre los que se pueden citar: el factorial y el de trayectoria (J. J. Hox, 1995:45). Sin embargo, este artículo hará mención solamente al modelo de regresión de nivel múltiple, o regresión “multinivel”, debido a que es el tipo de modelo que ha utilizado la autora.

### Modelo de regresión de nivel múltiple

Es una versión de la familia de modelos de regresión múltiple. Es un modelo versátil. Si se utiliza el código “dummy”<sup>4</sup> para variables categóricas, éste puede ser usado para el análisis de varianza (ANOVA); además, puede ser aplicado en una amplia gama de problemas de investigación, que requieran análisis longitudinales y análisis de datos de encuestas.

4. Se comparan dos grupos; un grupo con valor 1 y otro con valor 0 (variables dicotómicas). Si el coeficiente es positivo, el grupo que tiene valor 1 presenta mayor promedio en la variable dependiente que el valor de 0.

En este modelo se asume que existe un conjunto de datos jerárquicos con una variable dependiente, la cual es medida en el nivel bajo y variables explicativas en todos los demás niveles. Conceptualmente, puede ser visto como un sistema jerárquico de ecuaciones de regresión.

Un ejemplo con el fin de clarificar el modelo sería el siguiente:

Se recolectan datos en colegios (J), estos datos se refieren a diferente número de estudiantes (Nj) en cada colegio. Como resultado, se tiene nivel 1 = “*estudiante*”, y dentro de éste se encuentra la variable dependiente: *rendimiento académico* (Y); además, de la variable explicativa: *condición socioeconómica* (X). En el nivel 2 = “*colegio*”, se tiene la variable explicativa: *tamaño del colegio* (Z).

Con los datos anteriores, se puede realizar una ecuación de regresión separada para cada colegio, y predecir la variable dependiente *rendimiento académico* (Y), por la explicación de la variable explicativa *condición socioeconómica* (X). Obsérvese el gráfico 3 en donde se observa lo siguiente:

$Y_{ij} = B_{0j} + B_{1j} X_{ij} + e_{ij}$ , en donde:

$B_{0j}$  = intercepción

$B_{1j}$  = coeficiente de regresión (inclinación)

$e_{ij}$  = error residual

$j$  = colegios ( $j = 1 \dots j$ )

$i$  = estudiantes ( $i = 1 \dots N_j$ )

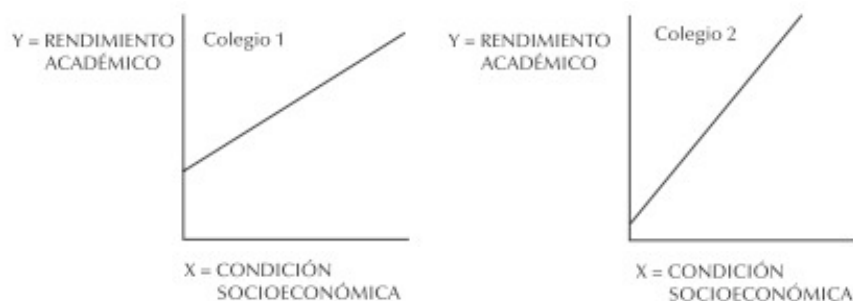


GRAFICO 3: Rendimiento académico según condición socioeconómica

La diferencia con el modelo de regresión usual, es que se asume que cada colegio está caracterizado por un coeficiente de intercepción diferente ( $B_0$ ) y por una inclinación diferente ( $B_1$ ). La varianza de error dentro de cada colegio es  $\sigma^2$ . Por otro lado, la mayoría de los modelos simples de análisis de niveles múltiples asumen que esa varianza de error es la misma para todos los colegios y se toma como la  $\sigma^2$ .

Los coeficientes para la intercepción y la inclinación varían a través de los diferentes colegios en estudio; por esta razón se dice que éste es un modelo de coeficientes aleatorios.

En el ejemplo presentado, cada colegio está caracterizado por su propio valor específico para los

coeficientes de intercepción ( $B_0$ ) y de inclinación ( $B_1$ ) para la variable *condición socioeconómica*. De esta manera, los estudiantes con un mismo valor en tales variables en un colegio con alto valor en la intercepción pueden predecir un alto grado de rendimiento académico, no así en un colegio con bajo valor para la intercepción. Además de las diferencias en los valores del coeficiente de inclinación para las mismas variables, se podría interpretar que la relación entre los aspectos sociales de las y los estudiantes y la predicción en su rendimiento no es la misma en todos los colegios.

A tomar ( $X$ ) como un valor específico en la variable socioeconómica, se puede observar el rendimiento académico para cada colegio. Obsérvese el Gráfico 4.

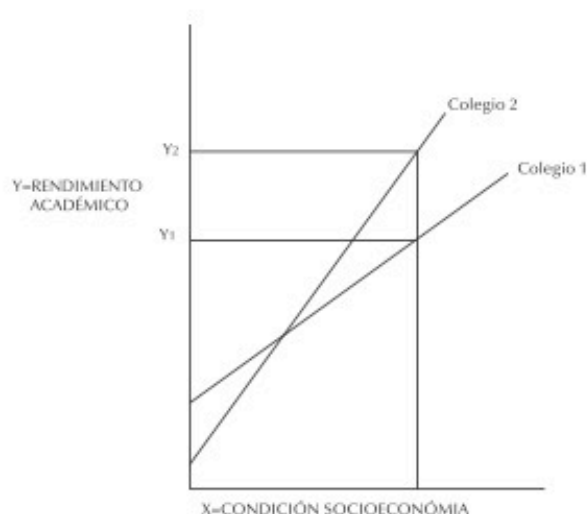


GRÁFICO 3: Ilustración hipotética rendimiento académico frente a condición socioeconómica según colegio.

La teoría estadística del modelo de regresión de niveles múltiples es compleja. Por tal motivo, se han creado una variedad de programas computacionales que ayudan al procesamiento de la información. No obstante, el óptimo uso de estos programas está sujeto a la convergencia en el modelo, ya que puede ser descalificado por errores de consistencia de los datos, muestras muy pequeñas, debido a que se requiere representatividad de las unidades tanto de mayor como de menor nivel (individual).

Entre los programas más utilizados para analizar un modelo de regresión de niveles múltiples están: HLM, VARCL, STATA y ML3/MLn.

El HLM (Bryk, Raudenbush, & Congdon, 1994), VARCL (Longford, 1990), y ML3/MLn (Prosser, Rasbash & Goldstein, 1991; Rasbash & Woodhouse, 1995). Las pequeñas diferencias se observan en la estimación producida por los programas; además, también pueden diferir en la cantidad de información que este brinda y en el diseño. No obstante, en teoría, cada programa debe conducir a las mismas conclusiones.

El HLM es un programa utilizado para el análisis de modelos de dos y tres niveles respectivamente; además, contiene un programa especial llamado VKHLM que puede ser usado para el llamado

*"modelo de meta-análisis"*. El HLM estima los parámetros, errores estándar y la covariancia para dos niveles. Para la mayoría de los parámetros estimados provee valores de "p", probabilidad asociada, como un indicador de su significancia estadística.

Este programa tiene una diferencia muy importante con respecto al VARCL, y es que este último usa la estimación máximo verosímil; por tal motivo, es muy utilizado para comparar la desviación de los diferentes modelos o inspeccionar los estimados y errores estándar de varios coeficientes en un modelo específico. El inconveniente que tiene este programa es que no contiene valores "p", de probabilidad asociada. Si el investigador o investigadora lo requiere, puede utilizarlo en conjunto con otro programa.

El programa MLn difiere de HLM y VARCL en que puede analizar datos con un número arbitrario de niveles (asumiendo que la computadora tenga suficiente capacidad de memoria); le ofrece al investigador o investigadora escoger entre la estimación máximo verosímil (llamada IGLS) y la estimación de

máxima verosimilitud restringida (llamada RIGLS). Posee comandos para manipulación de datos tales como variables estandarizadas, y no estandarizadas, además, de capacidades gráficas. Por otro lado, permite la estimación de efectos aleatorios en todos los niveles, y hay matrices de covarianzas correspondientes al número de niveles especificados para MLn. Además, requiere sólo un archivo de datos, el cual tiene que ser clasificado y contener variables que identifiquen las unidades en todos los niveles usados. Es un programa totalmente interactivo y muy poderoso.

El Stata (Stata Program Corporation), permite analizar sólo dos niveles, aunque si se tienen tres o más, se pueden anidar las unidades de segundo nivel en el tercero y analizar como si se tratara de dos niveles. Se obtienen estimaciones de los errores estándar para los coeficientes, valores de la probabilidad asociada y medidas de bondad de ajuste para el modelo.

En el siguiente apartado se presentan algunos aspectos generales que la o el investigador debe conocer con respecto a las variables.

# Aspectos generales relacionados con las variables en los modelos de niveles múltiples

En relación con las variables, éstas pueden ser definidas en cualquier nivel de jerarquía (alto o bajo) y algunas pueden ser medidas directamente en su nivel natural; por ejemplo: a nivel de centros educativos, pueden medirse variables como tamaño y denominación (pública-privada).

Otro punto que se toma en consideración es la movilización de las

variables de un nivel a otro ya sea “...por agregación, que significa que las variables de un nivel bajo pueden moverse a un nivel más alto de jerarquía; o por desagregación, que significa movilizar variables a un nivel más bajo” (Hox, 1995:2).

Swanborn (1981), citado por Hox (1995:2), conformó un esquema tipológico para adaptar las variables, el cual clarifica el nivel de medida al cual pertenecen las variables y la relación con respecto a la agregación y desagregación que puede haber entre ellas; se representa en la Figura 2.

NIVEL	1 (más bajo)	2	3	ETC
Tipo de variable	absoluta relacional contextual ⇐	analítica estructura I global relacional	analítica  estructura	
	(nivel conformado por individuos)	I estructural ⇐	global relacional contextual ⇐	

GRAFICO 3: Tipología para adaptar variables. Tomado de Hox, J. J. (1995).



⇒  
⇒



Aunque el nivel más bajo suele corresponder a las variables que se miden en individuos, no siempre sucede así. La definición de roles entre individuos, por ejemplo, puede colocarse en un nivel bajo de jerarquía, y en diseños longitudinales, se pueden definir como el nivel más bajo las medidas repetidas dentro de las variables individuales. (Goldstein, 1989, sp.) (Traducción de la autora).

A continuación se define los distintos tipos de variables.

1. *Variables globales o absolutas:* son las que se refieren únicamente al nivel en que están definidas, sin referencia a ninguna otra unidad o nivel de análisis; generalmente constituyen el nivel bajo. Por ejemplo, la inteligencia podría ser una variable global (se mide a nivel individual).
2. *Variables relacionales:* son las que conforman un nivel simple y describen las relaciones de una unidad a otra siempre dentro del mismo nivel. Por ejemplo, los índices de popularidad se consideran variables relacionales en el nivel individual.

3. *Variables analíticas y estructurales:* éstas se refieren a variables de niveles inferiores que se agregan en unidades mayores. Las analíticas se refieren a la agregación de una variable global o absoluta de un nivel más bajo. Las variables estructurales se refieren a la agregación de variables relacionales en el nivel más bajo. Por ejemplo, las características sociales de los individuos son variables analíticas cuando se agregan para toda una comunidad.
4. *Variables contextuales:* estas definen a las superunidades; todas las unidades en los niveles más bajos reciben el mismo valor en la variable que se mide en la superunidad a la cual pertenecen. Por ejemplo, el área geográfica, comunidad, etc.

## Errores en la interpretación

Existen varias falacias (errores en la interpretación) que pueden surgir cuando se aplica el modelo sin el conocimiento adecuado para ello. Hox (1995) se refiere a ellas de la siguiente manera:



Otorga incorrectamente características a los sujetos.

Diferencias en las variables de los sujetos a los contextos

1. *Falacia ecológica*: este error sucede cuando se interpretan datos agregados en un nivel individual, conocido como el efecto Robinson descrito por Robinson (1950:sp). Por ejemplo, Robinson describió datos agregados de la relación entre el porcentaje de negros e ilegales en Estados Unidos de Norteamérica en nueve regiones geográficas en 1930. La correlación ecológica entre las variables agregadas en el nivel región fue de 0,95, pero la correlación individual entre las variables absolutas, en el nivel individual, fue de 0,20. Por lo tanto, concluyó que en la práctica, una correlación ecológica no es cierta a su correspondiente correlación individual.
2. *Falacia atomística*: esta falacia se refiere a la inferencia que se realiza en un nivel alto desde análisis ejecutados en un nivel bajo. Una tipología de estas falacias es la Paradoja de Simpson, la cual se refiere al problema de conclusiones completamente erróneas si se agrupan los datos de poblaciones heterogéneas, colapsados y analizados como si fueran una simple población homogénea.

A continuación, se presenta, a manera de ilustración, algunos resultados obtenidos en una investigación realizada por la autora en donde se aplicó el modelo multinivel; el tipo utilizado es el que corresponde a la regresión múltiple.

### **Resultados obtenidos en la aplicación del modelo multinivel en una investigación**

Este apartado presenta a grandes rasgos una investigación realizada por la autora, en donde se aplicó un modelo de niveles múltiples de dos niveles. La misma tuvo como objetivo analizar los factores asociados a la repitencia de los y las estudiantes que cursan séptimo año en los colegios académicos, diurnos y públicos. La población objeto de estudio fueron 1442 estudiantes de séptimo año, 115 docentes que impartían clases de las asignaturas de Inglés, Español, Estudios Sociales, Ciencias y Matemáticas; y 14 directores y directoras de los centros educativos participantes en el estudio. Este estudio se enmarcó dentro del enfoque cuantitativo, no experimental, transversal y correlacional; además, se tomó en cuenta las pautas que brinda la operacionalización múltiple. El modelo de análisis de niveles múltiples permitió distin-

guir los efectos contextuales de los individuales. Esto quiere decir que, en un mismo modelo de análisis se incorporó no sólo variables individuales, sino también variables del contexto asociadas al centro educativo. Además, determinó qué parte de la variación de los indicadores se debe a, y es explicada por el contexto (los colegios) y qué parte por las características del individuo (los estudiantes repitentes). Para cada nivel en estudio se definen variables. Se utilizó el código "dummy" para variables categóricas, usado para el análisis de varianza (ANOVA).

Para el procesamiento de la información, se utilizó el programa Stata (Stata Program Corporation), el cual permitió analizar dos niveles y obtener estimaciones de los errores estándar para los coeficientes, valores de la probabilidad asociada y medidas de bondad de ajuste para el modelo.

En cuanto a los resultados propiamente del estudio; en términos generales se concluyó que la repitencia es un fenómeno multifactorial en donde intervienen aspectos propios del estudiante, de su contexto familiar, del centro educativo y de la dinámica pedagógica que desarrolla el docente. Más específicamente, y para resaltar el poder explicativo del modelo,

algunas conclusiones interesantes en la asignatura de matemáticas, la cual presentó el mayor porcentaje de repitencia en la muestra seleccionada, fueron las siguientes:

- Según la percepción de los y las estudiantes de séptimo año el sentirse contento en el centro educativo es un factor que influye en que se presente o no la condición de repitencia. Este estudio determinó que según la percepción de los y las estudiantes, quienes están contentos con el centro educativo tienen un 75% menos de posibilidad de ser repitentes en matemáticas; resultado estadísticamente significativo al 5%.
- Los y las estudiantes de séptimo año que realizan la mayor cantidad de tareas y asignaciones escolares tienen menos posibilidad de ser repitentes en matemáticas; resultado estadísticamente significativo al 5%.
- Los y las estudiantes de séptimo año que tienen alta autoestima tienen menos posibilidad de ser repitentes en matemáticas; resultado estadísticamente significativo al 5%.
- Los y las estudiantes de séptimo año que están motivados hacia el estudio tienen menos
- posibilidad de ser repitentes en matemáticas; resultado estadísticamente significativo al 5%.
- Según la percepción de los y las estudiantes de séptimo año, si el centro educativo está ubicado en una buena zona (puntaje alto), tienen menos posibilidad de ser repitentes en matemáticas; resultado estadísticamente significativo al 10%.
- Según la opinión de los y las estudiantes de séptimo año, si reciben ayuda para realizar sus tareas o deberes escolares tienen 51% menos de posibilidad de repetir en matemáticas; resultado estadísticamente significativo al 10%.
- Según la percepción de los y las estudiantes de séptimo año, cuanto es la frecuencia de la ayuda que reciben para realizar las tareas o deberes escolares, tienen más posibilidad de repetir en matemáticas, resultado estadísticamente significativo al 5%.
- Según la opinión del profesor y la profesora de séptimo año, si aplica adecuaciones curriculares a los y las estudiantes con necesidades educativas especiales, tienen menos posibili-

dad de repetir en matemáticas; resultado estadísticamente significativo al 10%.

- Según la opinión del profesor y la profesora de séptimo año, si existe una mala relación interpersonal entre profesor o profesora y estudiante, tienen más posibilidad de repetir en matemáticas, resultado estadísticamente significativo al 5%.

## CONCLUSIONES

Del presente artículo se concluye lo siguiente:

1. La teoría de los modelos de niveles múltiples parte del concepto de que las personas interactúan en diversos contextos sociales, de tal forma que las propiedades de esos grupos influyen el comportamiento de los individuos quienes conforman esos grupos y viceversa. Así, no se puede explicar integralmente un comportamiento individual, si no se analizan simultáneamente los factores del contexto.
2. La meta de los modelos de niveles múltiples es determinar el efecto directo de las variables individuales y del

contexto (en diferentes niveles de jerarquía), sobre variables de comportamiento humano.

3. El modelo de análisis de niveles múltiples en contraste con el análisis de regresión clásico, permite incorporar variables independientes de distintos niveles; así identifica qué parte de la varianza de la variable dependiente se debe al efecto de los factores individuales y qué parte se debe al contexto.
4. A pesar de que los modelos de análisis de niveles múltiples no son la panacea para todas las investigaciones; han sido la respuesta alternativa a los diversos problemas que se presentan cuando se aplican técnicas de análisis estadísticas tradicionales a estudios de factores asociados, o estudios que pretenden identificar causas en los fenómenos educativos y sociales.
5. El éxito en la aplicación de los modelos de análisis de niveles múltiples va a depender del conocimiento sustantivo y metodológico por parte del investigador o investigadora. Sustantivo en la medida en que se construya un adecuado y exhaustivo marco concep-

tual. Metodológico porque requiere de conocimientos en estadística multivariable.

## BIBLIOGRAFÍA

- Aitkin, M. y otros (1981) *Statistical Modelling of Data on Teaching Styles*. Amsterdam: J. Royal Statist. Soc.
- Aitkin, M.; Longford, N. (1986) *Statistical Modelling in School Effectiveness Studies*. Amsterdam: J. Royal Statist. Soc.
- Burch, A. (2001) "Introduction to Multilevel Modelling". Artículo de Internet. <http://multilevel.ioe.ac.uk>
- Craig, D. (1993) "Do Places Matter? A Multilevel Analysis of Regional Variations in Health related Behavioral in Britain". *Social Science of Medicine* 37 (6) 725-733. Internet. <http://tramss.data-archive.ac.uk>
- Goldstein, H. (1987) *Multilevel Models in Educational and Social Research*. United States. Oxford University. Redwood Burn Limited
- Goldstein, H. (1999) *Multilevel Statistical Models*. London: England University, Institute of Education. Kendalls Library of Statistics
- Hox, J. J. (1995) *"Applied Multilevel Analysis"*. Amsterdam. TT - Publikatie
- Jones, K. y otros (1991) "Ecological and Individual Effects in Childhood Immunisation. Uptake: A Multilevel Approach". *Social Science of Medicine*. 33 (4) 501-508. Internet. <http://tramss.data-archive.ac.uk>
- Kristiansen, I. y Høltedahl, K. (1993) "Effect of the Remuneration System on the General Practitioners choice between Surgery Consultations and Home Visits". *Journal of Epidemiology and Community Health*. 47: 481-484. Internet <http://tramss.data-archive.ac.uk>
- Little, Todd D. y otros (2000) *Modelling Longitudinal and Multilevel Data*. United States: Lawrence Erlbaum Associates
- Multilevel Model Project. (1989) *Multilevel Modelling Newsletter*. London: 1(1) Jan: 2-6
- Murillo, F. (1999) "Los Modelos Jerárquicos Lineales Aplicados a la Investigación sobre Eficacia Escolar". San José: *Revista de Investigación Educativa* 17 (2) 453-460
- Pedhazur, E. (1982) *Múltiple Regresión in Behavioral Research*. Explanation and Prediction. United States: Harcourt Brace College Publishers
- Piñeros, L. y Rodríguez, A. (1998). *Los Insumos en la Educación Secundaria y su Efecto sobre el Rendimiento Académico de los Estudiantes: un estudio en Colombia*. Bogotá: The World Bank, Latin America and the Caribbean Regional Office. Human Development Department
- Runyon, R. y Audrey, H. (1992) *Estadística para las Ciencias Sociales*. Estados Unidos: Addison- Wesley Iberoamericana
- Snijders, T. (2001) *The Statistical Evaluation of Social Network Dynamics*. Department of Statistics and Measurement Theory. Netherlands. University of Groningen.
- Statistical Modelling Transs (2001). "Multilevel Models". Artículo de Internet. <http://tramss.data-archive.ac.uk>

*¡Nuestros libros  
tienen la palabra!*



**EDITORIAL EUNED**

EDITORIAL UNIVERSIDAD ESTATAL A DISTANCIA

**LIBRERÍA  
MAGÓN**  
Mercedes  
de Montes de Oca,  
Carretera a Sabanailla.  
Tels: 253-9349  
253-8197

**LIBRERÍA  
FERNÁNDEZ GUARDIA**  
San José,  
Bajos del Teatro  
Melico Salazar.  
Tels: 223-9794  
Fax: 256-1765

**LIBRERÍA  
DOBLES SEGREDAS**  
Heredia,  
25 m N.  
Restaurante Fresas.  
Tels: 260-5159  
Fax: 261-2029

**LIBRERÍA  
CARLOS LUIS FALLAS**  
Alajuela,  
Frente a Plaza Ferias.  
Tel: 442-8640  
Fax: 443-6746

**LIBRERÍA  
MARIO SANCHO**  
Cartago, 200 m E.  
Tribunales  
de Justicia.  
Tel: 592-1017

Pedidos al por mayor al teléfono: 280-1451 / Fax: 280-1498