

Análisis multinivel en la investigación sociológica cuantitativa: implicancias y conveniencias

Andrea Alvarado Urbina¹ y Camila Jara Ibarra²

Fecha de recepción: 17 de junio de 2013

Fecha de aprobación: 27 de septiembre de 2013

Resumen:

La complejidad que la realidad social presenta dificulta su estudio pues muchas veces nos encontramos frente a fenómenos con estructuras jerárquicas o con distintos niveles de análisis. La emergencia del análisis multinivel busca aumentar la precisión analítica al incluir explícitamente esta complejidad, modelando las relaciones jerárquicas, considerando la dependencia entre observaciones y proporcionando estimaciones sobre la variabilidad y replicabilidad de los parámetros en distintos contextos sociales.

Este trabajo busca estimular la reflexión en cuanto al uso, el aporte y la conveniencia del Análisis Multinivel en el ámbito de la investigación social cuantitativa. Para ello, se realiza una descripción de la técnica con sus principales conceptos y aplicaciones, para luego exponer, mediante los resultados de una investigación multinivel, sus conveniencias y usos en términos prácticos.

Palabras clave: Análisis Multinivel, Regresión, Regresión Logística Multinivel, Investigación Cuantitativa.

Abstract:

The complexity of social reality makes it difficult to study as we many times face phenomena with hierarchical structures or multiple levels of analysis. The multilevel analysis intends to increase analytical accuracy by explicitly including such complexity, modelling hierarchical relations, taking into account the dependency of observations, and providing estimations on the variability and the reproducibility of parameters from different social contexts. This work intends to encourage reflections on the use, contribution and advantage of multilevel analysis to quantitative

1 Socióloga, Universidad de Chile, Estudiante de Master, Columbia University, Dirección postal: José Domingo Cañas 1640, depto 1405, Ñuñoa, Santiago, Chile. Dirección electrónica: <andrea.alvarado.urbina@gmail.com

2 Estudiante de Doctorado, Universidad de Leiden, Dirección postal: Johan Huizingagebouw, Doelensteeg 16, 2311 VL Leiden, Holanda. Dirección electrónica: <c.jara.ibarra@hum.leidenuniv.nl>

social research. It theoretically explores the technique as well as its main concepts and applications, and it presents its practical advantages and uses through the results of a multilevel research.

Key Words: multilevel analysis; regression; multilevel logistic regression; quantitative research

Introducción

Siendo el objetivo de la sociología la explicación de los fenómenos sociales, el fin de la metodología de la investigación es concretar este objetivo con la mayor rigurosidad y precisión posible.

El muestreo aleatorio simple no es utilizado frecuentemente, ya que no permite obtener una representación precisa cuando la población desde la que se extrae la muestra presenta mayores grados de complejidad. Así, las muestras estratificadas dan un paso adelante, y muchas veces logran retratar con bastante exactitud una población determinada.

Como consecuencia, nos encontraremos la mayoría de las veces frente a bases de datos que, de forma espontánea o producto de la técnica de muestreo, se pueden agrupar por niveles o, al contrario, desagregar de acuerdo a sus estructuras. Ejemplos de esto serán personas en hogares, alumnos en escuelas, etc. De esta forma, las observaciones en un nivel están 'anidadas' en niveles superiores, lo que podría generar dependencia o una mayor correlación entre las observaciones pertenecientes a un mismo grupo. En otras palabras, individuos que pertenecen a un mismo contexto podrían tender a ser más homogéneos en sus comportamientos, opiniones y actitudes que aquellos insertos en diferentes realidades, generándose así una estructura de correlación intracontextual o la también llamada Correlación Intra-Clase.

Este hecho tiene directas consecuencias sobre el análisis de la información, ya que justamente uno de los supuestos de los modelos de regresión tradicional es la independencia entre las

observaciones. De existir grupos con mayor homogeneidad dentro de una muestra, este supuesto sería violado, derivando en errores en los procedimientos inferenciales, sesgos en la estimación del error típico o en la aceptación de relaciones que no existen.

El Análisis Multinivel surge como respuesta a este problema al incluir explícitamente esta complejidad, modelando las relaciones jerárquicas, considerando la dependencia entre observaciones y proporcionando estimaciones sobre la variabilidad y replicabilidad de los parámetros en distintos contextos sociales.

La popularidad de esta técnica ha ido en aumento desde sus primeras apariciones en el campo de la investigación en educación en los años ochenta, hasta ser hoy reconocida como una herramienta fundamental para el quehacer de las ciencias sociales.

Este trabajo busca estimular la reflexión en cuanto al uso, el aporte y la conveniencia del Análisis Multinivel en el ámbito de la investigación social cuantitativa. Para ello, la exposición se realizará en dos secciones; primero realizaremos un abordaje teórico de la técnica con sus principales conceptos e implicancias, y luego presentaremos un extracto del estudio 'Determinantes de la Opinión sobre la Integración Económica y Política de Latinoamérica' como ejemplo práctico de una investigación multinivel, para conocer con ello su conveniencia y uso, centrándonos en la capacidad del modelo de integrar paralelamente características individuales y variables asociadas con un segundo nivel –o países, en el caso de esta investigación.

Cabe destacar, finalmente, que el foco de este estudio no se encuentra en la manera en que los latinoamericanos evalúan la integración política y económica del continente, sino en cómo la técnica de Análisis Multinivel abre posibilidades de investigación social que de otra forma resultan muy complejas de abordar. Por ello, este texto no incluye el detalle del procesamiento estadístico, sino sólo los datos que permiten ejemplificar los alcances de la técnica desarrollada.

Enfoque y Principales Conceptos

La investigación multinivel surge a mediados de los años ochenta en el campo de la educación. Al analizar el desempeño de los estudiantes respecto a diferentes materias, los investigadores notaron que los resultados no eran independientes entre los alumnos; se observó que los estudiantes formaban *clusters* en relación a sus clases o sus profesores, lo que se definió entonces como una estructura de dos niveles. El fenómeno se identificó también al momento de analizar a las clases o a los profesores ya que igualmente estos estaban agrupados en torno a los colegios a los cuales pertenecían, estableciéndose con ello un tercer nivel para el análisis (Goldstein, 2011; Leeuw y Meijer, 2007; Aparicio y Morera, 2007; Twisk, 2006).

Estos resultados de mayor homogeneidad entre *clusters* cuestionaron las técnicas de regresión tradicionales; al violarse el supuesto de independencia entre las observaciones, las estimaciones y los resultados podrían estar sesgados o ser incorrectos. El análisis multinivel da cuenta de esta dependencia y la modela abiertamente. Con ello, define y analiza micro y macro interacciones o la relación mutua entre los individuos y su contexto. Ejemplos de estructuras multiniveles serán estudiantes en las clases y en las escuelas, pacientes en clínicas, votantes en distritos, personas en hogares en barrios y ciudades, etc. (Snijders y Bosker, 1999).

Una vez definidas las unidades de análisis, es posible identificar dos tipos de estructuras: jerárquica y no jerárquica. Como su nombre lo indica, una estructura jerárquica supone un orden en el que un primer nivel se encuentra anidado única y exclusivamente en una unidad de un nivel superior. Las estructuras no jerárquicas son aquellas donde el primer nivel se encuentra anidado en dos o más niveles superiores al mismo tiempo. En la práctica, se suele observar un tipo de estructura o ambas al mismo tiempo³.

3 Ejemplos de los primeros serían los ya nombrados estudiantes (unidades en el nivel 1) anidados en un único profesor (unidades en el nivel 2) que a su vez se

Para poder definir las estructuras, es necesario en este punto distinguir adecuadamente los niveles de las variables involucradas en la investigación o, lo mismo, aprender a diferenciar los efectos aleatorios de los fijos. En un diseño de investigación, un factor es aleatorio si sus niveles consisten en una selección al azar de una población de niveles posibles. Un factor es considerado fijo, si sus niveles son escogidos premeditadamente y por un procedimiento no aleatorio. Un factor aleatorio será entonces aquel cuyas unidades remiten a una muestra aleatoria de una población mayor mientras los efectos fijos son un número menor y fijo de clasificaciones (Milliken y Johnson, 1993)⁴.

Cuando una estructura jerárquica o niveles de análisis han sido apropiadamente identificados en un diseño de investigación, es el momento de optar por la utilización de un Análisis Multinivel como técnica para el manejo de los datos. Con ello, se opta por una herramienta que resuelve el problema de la correlación entre los individuos que afecta a la estimación mediante los mínimos cuadrados ordinarios. De igual forma, con la utilización de un análisis multinivel evitamos la falacia atomística o el error de centrar el estudio sólo en el nivel micro sin considerar la influencia que el contexto pueda tener sobre las observaciones. Por otra parte, un análisis concentrado sólo en los niveles superiores incurre en la llamada falacia ecológica, al no dar cuenta de

encuentran agrupados en escuelas (unidades en el nivel 3). Ejemplos de estructuras no jerárquicas serían personas (unidades en el nivel 1) agrupadas simultáneamente en vecindarios (unidades en el nivel 2) y en lugares de trabajo (unidades en el nivel 2).
 4 Tal como lo explica Murillo (2008), en los modelos de regresión clásicos los parámetros que se estiman son el intercepto (o punto de corte) y las pendientes. Estos coeficientes se asumen como fijos, esto es, comunes a todos los sujetos y son estimados a partir de los datos mientras los coeficientes aleatorios son variables y se distribuyen según una función de probabilidad. En el caso de una estructura multinivel, en tanto, los coeficientes del primer nivel son tratados como aleatorios en el segundo nivel. Esto es, los modelos multinivel están compuestos por dos partes: una general común a todos los contextos que es la llamada parte fija, y otra que representa la especificidad de cada contexto, que se estima y que varía a través de la varianza en los distintos niveles.

diferencias y complejidades a nivel individual (Hox, 2002; Hill y Rowe, 1996; Martín et al., 2009).

Estimaciones más precisas y lecturas más realistas de un determinado fenómeno social, la magnitud de las varianzas que operan en los diferentes niveles y, en definitiva, la relación de los individuos y los medios donde se desenvuelven son alguna de las implicancias del uso de esta técnica. Volveremos a este punto más adelante.

Principios básicos de Análisis Multinivel

El análisis multinivel es considerado una extensión del análisis de regresión tradicional. Por medio de este último se estudia la relación lineal entre dos o más variables, de forma que el comportamiento de la variable dependiente es influenciado por una combinación de predictores. La forma típica es la ecuación:

$$y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

donde la intersección será el valor de cuando el valor de la variable X sea igual a 0. Representa el cambio cuando X varía en una unidad. Supongamos ahora que los datos o la información referente ha sido recolectada en distintos ambientes físicos y sociales. Un modelo adecuado entonces sería la ecuación:

$$y_{ij} = \beta_0 + \beta_1 X_{1ij} + u_j + \varepsilon_{ij}$$

donde hemos agregado el subíndice j para representar estos distintos ambientes o contextos que han sido aleatoriamente seleccionados. Incluimos además el efecto aleatorio que representa las diferencias o la variación entre las medias de los distintos escenarios. De esta manera, el valor de y y su nivel de significancia nos indicará la magnitud del efecto de un determinado contexto para el desempeño de la variable estudiada.

Una alternativa a este enfoque sería el realizar diferentes análisis de regresión tradicionales o Ancovas, controlando o corrigiendo cada vez por cada uno de los ambientes mencionados.

Esto podría hacerse, por ejemplo, mediante la inclusión de variables *dummy* en el modelo, representando cada una de las unidades de este *segundo nivel*. Como resultado, tendríamos tantas ecuaciones y rectas como contextos o unidades en el segundo nivel posea un diseño determinado. El problema de esta solución, sin embargo, radica en que muchas veces estos contextos no son de especial interés por sí mismos, es decir, no es el comportamiento específico en cada uno de estos ambientes el foco del estudio, sino más bien lo que nos interesa es dar cuenta de que efectivamente existen diferencias generales entre las medias de la variable para cada uno de los ambientes estudiados. Esto es, existe un efecto asociado al contexto en que las observaciones se encuentran insertas.

Al usar un análisis multinivel lo que se realiza no es una estimación para cada una de las rectas sino que una estimación de la variación entre ellas. Esto, pues se espera que las observaciones anidadas en cada uno de los ambientes tiendan a ser más homogéneas entre ellas (variación inter-grupos) generando una media en la variable relativa a este grupo y diferente a la media del grupo perteneciente a otro contexto (variación entre-grupos). Esta es la denominada 'Correlación Intraclase' y una regresión multinivel puede ser usada para su estimación. El modelo usado para este propósito no incluye predictores o variables explicativas reduciéndose a:

$$y_{ij} = \beta_0 + u_j + \varepsilon_{ij}$$

donde no se explica la varianza sino más bien la descompone en dos elementos independientes: y/o las varianzas de los errores del nivel más bajo y más alto respectivamente. Usando este modelo, podemos estimar la correlación intraclase ya que ésta es igual a:

$$y_{ij} = \beta_0 + u_j + \varepsilon_{ij}$$

Este valor nos permitirá entonces evaluar si es necesario optar por un análisis multinivel o continuar con la estimación por

medio de una regresión simple. Observando esta ecuación entonces, es posible deducir que mientras más pequeña sea la varianza dentro de los grupos, mayor será la correlación intra clases (ver: De la Cruz, 2008; Murillo, 2008).

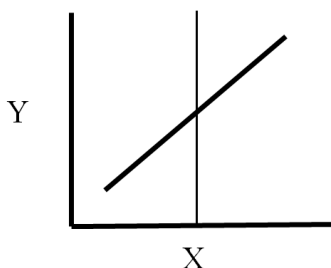
Este coeficiente es clave para la realización del análisis y supone una de las mayores innovaciones del mismo, dado que la regresión tradicional –asumiendo el supuesto de independencia– considera que cada nueva observación provee un 100 por ciento de nueva información. Cuando incorporamos el contexto y los *clusters* al análisis, la información provista por cada nueva observación correspondiente al mismo grupo no es 100 por ciento nueva, sino un porcentaje menor. La magnitud de la información proporcionada por cada nueva observación dependerá del tamaño del valor del coeficiente de correlación intra clase. Mientras más alto este valor, menor la proporción de nueva información en cada unidad de análisis y con ello mayor el error estándar del análisis multinivel comparado con el análisis tradicional (Twisk, 2006). Es aquí donde surge la posibilidad de aceptar un modelo que en la realidad no es significativo o donde los errores Tipo I podrían aparecer.

Volviendo a la variación entre grupos y a la lógica del análisis, se puede afirmar que la técnica multinivel asume como una variable aleatoria con media cero y variación constante. Este hecho presenta uno de los objetivos del modelo, esto es, analizar si la inclusión de predictores reduce o ‘explica’ la variación entre grupos o contextos (Paterson y Goldstein, 1991). De esta forma, la complejidad y la capacidad explicativa de los modelos multinivel aumenta en la medida que se permite a los efectos fijos variar a través de los diferentes niveles.

En la Figura 1 vemos un modelo de regresión tradicional o *de un nivel*, en el que por medio de la técnica de Mínimos Cuadrados se define una recta general para retratar la relación entre las variables X e Y. Al hacer esto, y tal como se ha mencionado, se generaliza la relación de ambas variables para todos los grupos

en que estas medidas están siendo analizadas, ignorando con ello las posibles diferencias en las variaciones entre grupos.

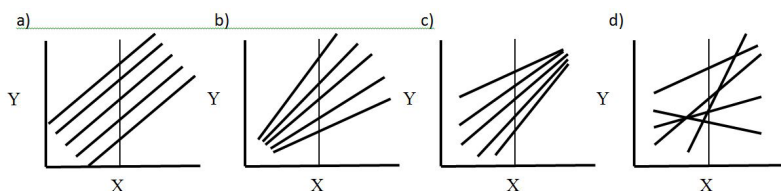
Figura 1. Modelo de regresión lineal tradicional



Al observar los gráficos presentados en la Figura 2, es posible reconocer en ellos las consecuencias de un análisis que considera y modela los niveles –dos niveles en estas gráficas– de un diseño. En el primer gráfico (a) se ha permitido que los interceptos varíen a través de las distintas unidades del segundo nivel (Modelo de Interceptos Aleatorios). Lo que se observa es una relación positiva entre X e Y al igual que en la Figura 1, pero esta vez se da cuenta del efecto de los contextos en los que la relación se establece. Este es el modelo multinivel más básico y que permite evaluar, junto al coeficiente de correlación Intra Clase, si la realización de un análisis por niveles comienza a ser justificado.

A continuación, nos encontramos con los modelos (b), (c) y (d), donde además de permitir que el intercepto varíe a través de los grupos, se incorpora la variación de la pendiente o *beta* a través de estos niveles (Modelo de Interceptos y Pendientes Aleatorias). En el segundo gráfico (b), vemos que los contextos generan una diferencia en la relación entre las variables, pero que esta vez se acentúa en la medida que X aumenta su valor. Lo contrario se observa en (c), donde para las unidades del segundo nivel son los menores valores de X los que definen una mayor diferencia. En la gráfica (d) finalmente, los grupos o contextos producen una diferencia en la relación entre las variables aunque mucho más compleja y difícil de interpretar.

Figura 2. Modelos de regresión lineal para dos niveles



En el cálculo de los resultados de cada uno de los tres modelos revisados (Tradicional, Intercepto Aleatorio y Pendiente e Intercepto Aleatorio) el número de parámetros estimados es diferente. En el modelo de intercepto aleatorio un parámetro adicional es estimado, esto es, la variación de los interceptos a través de los grupos (y no un intercepto para cada uno de ellos). El modelo de Pendiente e Intercepto Aleatorio genera 3 parámetros adicionales ya que estimaremos la varianza de los interceptos, la varianza de las pendientes y además un tercer parámetro: la correlación o covarianza entre ambos. Tal como observamos en la Figura 2, el hacer variar los interceptos y las pendientes genera distintos patrones de comportamiento entre las variables analizadas. En (b) vemos que la relación entre interceptos y pendientes es positiva, en (c) identificamos una relación negativa y en (d) independencia entre ambas. Tal como se observa a esta altura, lo que se ha retratado en este último párrafo es la existencia de una relación latente o implícita en la información, escondida en la regresión tradicional y evidenciada por la ejecución de un análisis jerárquico.

Al ser el análisis multinivel una extensión de la regresión lineal, ciertos supuestos deben cumplirse. Si la variable resultado es continua, debe estar normalmente distribuida. Adicionalmente, los interceptos y las pendientes aleatorias también deben mostrar este comportamiento. Siempre se debe evaluar la presencia e influencia de outliers. El supuesto de independencia de los residuos, por otra parte y como ya se ha mencionado, no es un problema, es más bien la causa por la que se realiza un análisis multinivel en primer término (Twisk, 2006).

Una vez revisados los principales conceptos y la lógica tras la herramienta analítica multinivel, revisaremos un ejemplo de la aplicación del mismo, centrándonos específicamente en una de sus mayores ventajas y conveniencias: la estimación de un modelo que considere características individuales y variables asociadas con un segundo nivel, países en el caso de esta investigación. Así, el análisis de la interacción a través de dos niveles es el centro del estudio que presentamos a continuación.

Conveniencias del análisis multinivel: un ejemplo práctico

Tal como se ha revisado hasta este punto, el uso del análisis multinivel tiene múltiples implicancias y supone un nuevo enfoque en el tratamiento de la información cuantitativa en la medida que permite estimar la variabilidad de los fenómenos a través de distintos escenarios o contextos sociales. Otra ventaja importante de esta herramienta es que permite la inclusión de factores o variables explicativas de distintos niveles en un mismo modelo y, con ello, el análisis de los efectos de interacción tanto entre factores de un mismo como de diferentes niveles. Es sobre estas últimas conveniencias sobre la que específicamente se concentra esta sección.

En el estudio 'Apoyo público a la integración Latinoamericana: factores individuales y nacionales'⁵ se utilizó esta técnica estadística para evaluar el efecto de variables individuales⁶ y variables ecológicas o de contexto⁷ en la probabilidad de que los

5 La investigación corresponde a la tesis de máster en Análisis Cuantitativo en las Ciencias Sociales de la Universidad Católica de Lovaina realizado por Camila Jara durante el año 2011.

6 Se incluyen variables sociodemográficas (edad, género, nivel educacional, situación económica) y variables ideológicas (tendencia política, orgullo nacional).

7 Los años de estudio de la fuerza de trabajo se consideran como factor a nivel de país. Debido a que en América Latina no se observan sistemas de bienestar como en

latinoamericanos tuvieran una opinión favorable o desfavorable respecto a una eventual integración económica de la región. La decisión de incluir ambos niveles en el análisis responde tanto a motivaciones teóricas como a la búsqueda de la precisión metodológica.

Tal como se mencionó anteriormente, el enfocarse únicamente en un micro o en un macro nivel podría derivar en las falacias atomística o ecológica. La estrategia de mirar sólo un nivel de la estructura de la información puede producir incluso resultados contradictorios o dificultad para interpretar la causalidad o relación de los fenómenos. Más aún, en el contexto de este estudio y de sus antecedentes teóricos, la asimetría o variabilidad entre los países de Latinoamérica, especialmente en términos de sistemas políticos y modelos económicos, ha sido mencionada como un elemento inhibitor del desarrollo de una unidad regional (Rojas, 2006). Por tanto, el modelo del estudio incluye variables que miden la heterogeneidad entre países para evaluar así si dichas diferencias podrían influir también en la opinión de los latinoamericanos respecto al proyecto integracionista. El análisis multinivel permite evaluar entonces en qué medida el apoyo a la integración a nivel individual (opinión) es afectada por las diferencias entre países. Esto es, la locación de las personas es tomada en consideración dentro del modelo.

Asimismo, es posible argumentar que la opinión y las actitudes respecto a determinados temas son influenciados por el componente contextual. Hallazgos respecto a la evaluación ciudadana sobre la Unión Europea revelan que la diferencia entre la calidad y el funcionamiento de las instituciones públicas a través de los distintos países miembros es una variable explicativa importante para esta evaluación (Sánchez-Cuenca, 2000) in which support is the consequence of the interplay bet-

Europa, la desigualdad en la distribución del ingreso es abordada como una medida del bienestar e incorporada al presente análisis.

ween supranational and national politics: the higher citizens' opinion of the functioning of supranational institutions and the lower that of national institutions, the greater their support for integration. This is so because the worse the opinion of the national political system, the lower the opportunity cost of transferring sovereignty to Europe. In order to illustrate this hypothesis, I show that levels of national support for integration are higher in those countries that suffer greater corruption and have less-developed welfare states. Through an ordered logit analysis of one Eurobarometer survey, I test this political model at the individual level. Finally, I draw some conclusions from the empirical analysis for current discussions on the possibility of a European democracy." ,"DOI": "10.1177/1465116500001002001", "ISSN": "1465-1165, 1741-2757", "journalAbbreviation": "European Union Politics", "language": "en", "author": [{"family": "Sánchez-Cuenca", "given": "Ignacio"}], "issued": {"date-parts": [{"2000", 6, 1}]}, "accessed": {"date-parts": [{"2013", 10, 29]}]}, "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json". Para Brinegar y Jolly (2005) en su estudio respecto al apoyo a la comunidad europea, las actitudes de los ciudadanos están fuertemente influenciadas por la configuración política y económica de sus países. Para los autores, las variables a nivel individual tienen un efecto más débil en el apoyo a la integración que el que comúnmente se piensa. Para ello, proveen evidencia que sugiere que las personas hacen evaluaciones sociotrópicas para emitir sus juicios respecto a la integración europea, y señalan que las características individuales deben ser entendidas en el marco de los factores contextuales.

Mediante el estudio que se presenta a continuación, se intenta mostrar un ejemplo representativo de un problema de estructura jerárquica que es analizado a través de la técnica multinivel. Dos niveles son incorporados en un mismo modelo y, con ello, factores explicativos en ambos niveles y sus efectos de interacción. Se incorporan así las sugerencias de los autores antes señalados y se

avanza en una mirada más profunda respecto al fenómeno del apoyo a la integración.

Método y Principales Hallazgos

Mediante la presentación de este estudio se hace una invitación a evaluar la interacción entre micro y macro niveles en una investigación social y a reflexionar de qué forma la inclusión de ambos puede enriquecer los hallazgos y ampliar el alcance de su visión.

Dada la naturaleza dicotómica (0,1) de la variable resultado,⁸ en el estudio sobre el apoyo público a la integración latinoamericana se utilizó una regresión logística multinivel. El modelo considera que la variable dependiente Y_{ij} sigue una distribución *Bernoulli* $Y_{ij} \sim \text{Bernoulli}(1, \pi_{ij})$ con una varianza condicional $\text{var}(y_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})$ donde π_{ij} es la probabilidad para un individuo i de presentar la característica de interés o referencia. En este caso, $\pi_{ij} = \text{Pr}(y_{ij}=1)$ o declarar estar *a favor* de la integración económica de la región en el país j . Por tanto, considerando únicamente un nivel, el modelo se define como:

$$\text{Logit}[\pi_{ij}] = \log \left[\frac{\pi_{ij}}{1 - \pi_{ij}} \right] = \beta_{0j} + X'_{ij} \beta_1$$
$$\beta_{0j} = \beta_0 + u_j$$

8 En términos generales, en el estudio se utilizó un análisis secundario de dos fuentes diferentes: para las variables individuales se trabajó con el estudio Latinobarómetro 2009 con dos preguntas categóricas sobre el apoyo o rechazo a una eventual integración política y económica de la zona. Para las variables nacionales, por otra parte, se utilizó la base de datos del Statistical YearBook 2010 (CEPAL, 2010). Como técnica de análisis estadístico se optó por una regresión logística multinomial jerárquica o multinivel. El procesamiento de los datos se realizó con el software MLwin 2.2, integrando las características individuales y de los países en un único análisis con una estructura de dos niveles. Modelamos los *odds ratios* de una respuesta en función de las variables explicativas. El proceso de estimación fue secuencial, con 9 modelos para cada una de las dos variables dependientes, agregando cada vez un nuevo predictor desde el nivel individual hasta el nivel nacional.

donde X'_{ij} el vector designa a las variables explicativas. El coeficiente β_0 representa la constante del modelo y el vector $\beta_1=(\beta_{11}, \beta_{12}, \dots, \beta_{1k})$ incluye los coeficientes k de las covariables. La variación en el nivel individual no se incluye en la ecuación ya que ésta forma parte de la especificación del error de la distribución. Cuando el error de distribución es binomial, la varianza es una función de la proporción de la población y no se estima de manera separada (Hox, 2002). Al incluir el nivel país en el modelo, se permite que la constante varíe a través de los países analizados. El modelo logit multinivel:

$$\text{Logit}[\pi_{ij}] = \log \left[\frac{\pi_{ij}}{1 - \pi_{ij}} \right] = \beta_{0j} + X'_{ij}\beta_1$$

$$\beta_{0j} = \beta_0 + u_j$$

donde β_0 es la media de la constante para el total de la población. El efecto aleatorio del modelo está distribuido normalmente con una media igual a 0 y una varianza σ_u^2 y muestra la diferencia en la variación de la variable dependiente que puede ser atribuida al país.

Tal como se introdujo en la sección anterior, uno de los usos del análisis multinivel tiene que ver con su capacidad de evaluar en qué medida el fenómeno o variable estudiada varía a través de diferentes contextos (o unidades en niveles superiores). Esto se puede realizar mediante el modelo más simple en el que únicamente la variable de interés es estimada. Este es el llamado modelo vacío o de intercepto. En el caso del estudio realizado, el modelo vacío se usó para predecir la probabilidad de estar a favor de la integración económica de Latinoamérica y la forma en que esta opinión varía a través de los 18 países observados. Esto último, y tal como se mencionara, es relevante en la medida que en el caso de éste y otros estudios, el análisis del comportamiento de la variable dependiente en cada país (o colegio, hospital, etc.) no es de interés por sí mismo, sino más bien lo que interesa es estimar la forma en que el país ejerce influencia sobre las opiniones de los

individuos. Lo anterior se logra comparando el modelo vacío de primer nivel con el modelo de intercepto aleatorio.

En este caso, la estimación del modelo simple o de un nivel para el ratio de apoyo y rechazo a la integración es igual a $\exp(1.621) = 5.06$, implicando que los odds o probabilidades de apoyar la integración son 5 veces más altos que los de estar en contra⁹ de esta iniciativa (ver Tabla 1).

Entrando a las conveniencias del análisis multinivel, en la Tabla 1 se observan también los log-odds de apoyo a la integración cuando la estructura multinivel es tomada en cuenta. Cuando se permite que el intercepto varíe aleatoriamente a través de los países, un intercepto general para el país j es estimado: $1.721 + u_{0j}$, donde la varianza u_j —o la varianza *entre* países— es igual a 0.321 (SE=0.110) y significativa al 1 por ciento. Por tanto, hay diferencias significativas respecto a la opinión sobre la integración a través de los países latinoamericanos y esta variabilidad debe necesariamente ser considerada en el estudio. Así, cuando el efecto país es analizado, los log-odds del apoyo a la integración económica son 1.721 (en promedio a través de todos los países) equivalente a odds de $\exp(1.721) = 5.59$. Aplicando la transformación logística $\hat{\pi} = \frac{1}{1+e^{-(x)}}$ sería igual a una probabilidad estimada de $1/(1+\exp(-1.721)) = 0.85$. Esto es, existe un 85 por ciento de probabilidades de apoyar la integración latinoamericana en un país ‘promedio’ lo que es igual a la gran media o intercepto general de las probabilidades estimadas para los 18 países.

9 El nivel de significación del parámetro ($p < 0.01$) se obtuvo a partir de la prueba Wald.

Tabla 1. Parámetros y errores estándar del modelo vacío o de intercepto para predecir la probabilidad de “apoyar la integración económica en América Latina” (errores estándar entre paréntesis).

Parámetro	Estimación
<i>Efecto fijo del intercepto (en Nivel 1)</i>	1.621 (0.021)***
<i>Efecto fijo del intercepto (en Nivel 2)</i>	1.721 (0.136)***
<i>Efecto aleatorio σ_u^2</i>	0.321 (0.110)***

Notas: *** significativo al nivel 1%, ** significativo al 5% , * significativo al 10%

Si se asume que los log-odds para el apoyo a la integración siguen una distribución normal con una media 1.721 y varianza 0.321, los intervalos de confianza de 95 por ciento para β_{0j} son $1.721 \pm [(1.96 * \sqrt{(0.321)}) = 1.11] = (2.83, 0.61)$. Transformando los log-odds a probabilidades, este intervalo de confianza es (0.94, 0.65). Es decir, aun cuando las probabilidades de apoyo a la integración son iguales a 85 por ciento para todos los países de Latinoamérica, hay diferencias significativas en las probabilidades entre los países ya que el apoyo varía desde un 65 a un 94 por ciento cuando los efectos aleatorios son considerados y ningún factor o variable explicativa es incluida en el modelo.

Este primer paso o análisis de modelo vacío permite, por tanto, medir la variabilidad del fenómeno a través de las unidades de segundo nivel estudiadas y estimar la necesidad de incluir dicha variabilidad como un elemento explicativo. Además, gracias a la inclusión de la estructura multinivel en el análisis, no sólo la variación *entre* países puede ser medida, sino que también la variabilidad *dentro* de los países –o cualquier unidad de nivel mayor– puede ser estimada. Esta variabilidad se conoce como variación dentro del grupo o variación de cluster que, en el caso de este estudio, equivale a $1.721 \pm (1.96 * 0.136) = (1.99, 1.45)$. Es decir, la probabilidad promedio para el apoyo a la integración económica varía desde 81 a 88 por ciento *dentro* de los países de Latinoamérica.

Tal como se mencionara en la primera sección, otro elemento de gran relevancia proporcionado por el modelo nulo o vacío

para analizar el efecto del contexto es la correlación intra-clase o el coeficiente de partición de la varianza.¹⁰ Por tratarse de una variable dicotómica en el caso del apoyo a la integración, se asume que los efectos aleatorios están distribuidos normalmente y la correlación intra-clase estimada es, por tanto, igual a $\hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \pi^2 / 3)$. Lo anterior, puesto que en la formulación de la regresión logística se asume que los errores¹¹ siguen una distribución logística estándar con una media 0 y una varianza igual a Hedeker, 2007. Así, la correlación intra-clase para el modelo nulo de apoyo a la integración es igual a $\rho = \frac{0.321}{0.321+3.29} = 0.09$, lo que es equivalente a señalar que la correlación o dependencia entre individuos dentro de un país es igual a un 9 por ciento. En otras palabras, aproximadamente el 10 por ciento de la varianza residual del apoyo a la integración económica puede ser atribuida a las diferencias detectadas entre los países de la región.

De esta manera, es posible modelar y explicitar la influencia que ejerce el contexto en el que se enmarcan los sujetos sobre sus respuestas individuales. Para este estudio, por tanto, el país en que el individuo latinoamericano reside es uno de los factores que permite explicar una actitud positiva o negativa hacia la integración. Esta información habría sido omitida de no haber llevado a cabo un análisis que considerara la complejidad de niveles contenida en los datos.

Avanzando en la complejidad del modelo se integran los factores que, teóricamente, tendrían un impacto sobre la opinión de los latinoamericanos respecto a la integración. Considerando los efectos principales de cada uno de dichos factores, la constante

10 El VPC para un modelo de intercepto aleatorio de dos niveles es la proporción de la varianza residual total que es atribuible al nivel dos, es decir, interpretable como la correlación entre dos unidades del primer nivel en la misma unidad de nivel superior (Rasbash et al., 2009).

11 Si la distribución del error es binomial, la varianza es una función de la proporción de la población y no debe ser estimada de forma separada en la ecuación de regresión.

del modelo puede ser leída como la probabilidad promedio de apoyo a la integración a través de todos los países o unidades de segundo nivel para el perfil de referencia.¹² Esto es, para una mujer latinoamericana de 40 años de edad con una situación económica precaria, que sólo ha completado su educación primaria, que declara estar orgullosa de su país y tiene una tendencia política de derecha, la probabilidad de estar a favor de la integración es igual a 0.75. Asimismo, esta mujer vive en un país promedio tanto en términos del índice Gini (0.519) como de los años de educación de la fuerza de trabajo (8.4 años).

El efecto de todas las variables a nivel individual es significativo excepto por la variable edad. El detalle de estos efectos, cuya interpretación se asemeja a una regresión logística simple o de un nivel,¹³ escapa del objeto de este trabajo donde más bien se vuelve de especial relevancia exponer los hallazgos de las variables de segundo nivel y, más aún, los efectos de interacción entre niveles.

Las dos variables de segundo nivel incluidas en el modelo resultaron no ser significativas para explicar la opinión respecto a la integración económica de la región. Sin embargo, aun cuando las variables nacionales parecen no aportar significativamente a la explicación de esta opinión, los términos de interacción entre variables de ambos niveles resultaron ser no sólo efectos significativos sino que además presentan información relevante para la discusión y el alcance de los estudios de opinión. A través de la

12 Este perfil es definido por las categorías de referencia de cada una de las variables. La elección de éstas se basa en que individuos de dichas categorías tendrían, de acuerdo a los antecedentes teóricos, menos probabilidades de estar a favor de la integración latinoamericana lo que facilita la lectura e interpretación de los resultados de la regresión logística.

13 Aunque se pueden obtener conclusiones y tendencias similares cuando únicamente el nivel inferior es involucrado en el análisis, la regresión simple y su versión multinivel no son análisis equivalentes pues en el caso de esta última la estructura multinivel es incluida en el modelo y en el procesamiento de los datos en todo momento y, por tanto, los parámetros estimados debieran diferir entre ambas técnicas.

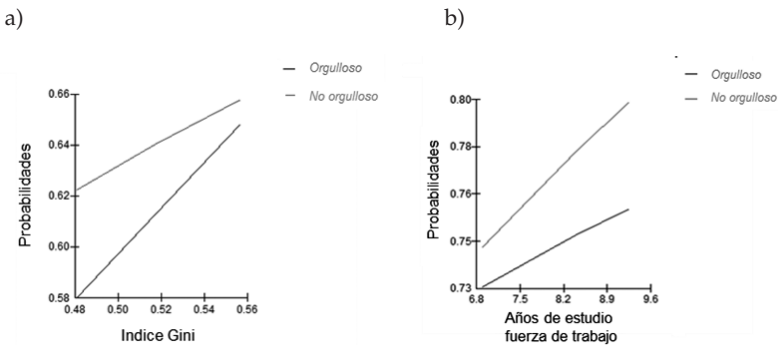
exposición de estos hallazgos se busca profundizar más aún en las características y potencialidades del análisis multinivel.

Para facilitar la interpretación de dichos efectos, tres valores característicos de las variables ecológicas fueron calculados y vinculados con las variables ideológicas y educacionales de nivel individual.¹⁴ El efecto multiplicativo del orgullo por la patria con ambas variables de nivel nacional es significativo. Para ambas categorías (orgullosos y no orgullosos) las probabilidades de apoyar la integración aumentan en la medida que los años de educación de la fuerza trabajadora del país son más altos; esta tendencia sin embargo es más pronunciada para la categoría 'no orgullosos'. Como se puede ver en el eje horizontal del gráfico de la Figura 3 (a), los valores característicos más altos y más bajos para los años de educación de un país son 6.890 y 9.258 respectivamente. Para un país con un universo de trabajadores con un mayor número de años de estudio (9.258), los odds para el apoyo a la integración económica son 1.29 veces más altos que para aquellos que no están orgullosos versus aquellos individuos que sí declaran sentirse orgullosos de su nación. En contraste, en un país con una fuerza de trabajo con menos años de estudio (6.890), el odds ratio es igual a 1.09, esto es, aquellas personas que declaran estar orgullosos y aquellos que declaran no estarlo presentan aproximadamente las mismas probabilidades de apoyar la integración de la región. Tal como se presentó en la primera sección, la representación gráfica de este término de interacción presenta una tendencia de 'abanico abierto', es decir, el rango de probabilidades de apoyar la integración es más estrecho (casi las mismas probabilidades) para las categorías de orgullo por la nación en aquellos países con una población activa de trabajadores con un menor promedio de años de estudio, rango que se amplía a medida que el número de años de estudio se expande.

14 Los efectos multiplicativos o de interacción fueron incluidos en el modelo de acuerdo a los antecedentes teóricos.

Una tendencia de ‘abanico cerrado’, por el contrario, se puede observar para el término de interacción entre las variables orgullo por la nación y la distribución del ingreso del país (ver Figura 3 (b)). Existe un efecto del orgullo por la nación en los odds de apoyar la integración que presenta una brecha más amplia entre sus categorías en aquellos países con un índice Gini menor o, lo mismo, países con una distribución más equitativa del ingreso. Los odds para ‘no orgulloso’ versus ‘orgulloso de la nación’ en países con una distribución de ingreso más equitativa son 31 por ciento más altos mientras que en países con distribución más desigual donde este ratio es igual a un 10 por ciento. Siendo el ‘orgullo por la nación’ un predictor eficiente a nivel individual, cuando su efecto es combinado con una variable ecológica o nacional su efecto se ve intensificado en aquellos países con una distribución del ingreso más simétrica y una población de trabajadores más educados.

Figura 3. Efecto de interacción de los años de estudio de la fuerza de trabajo y de la distribución del ingreso con el orgullo por la nación, en el apoyo a la integración económica

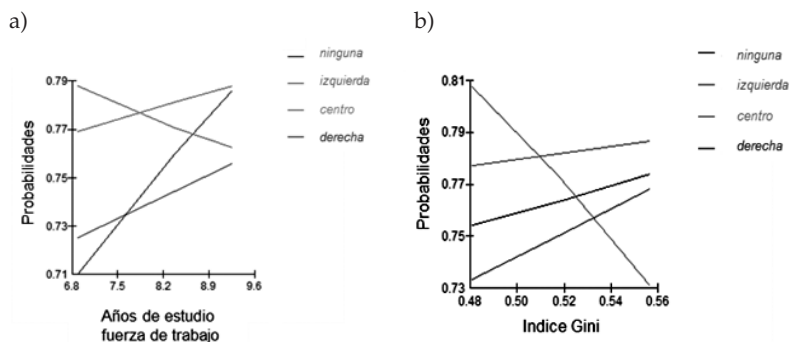


Al analizar el término de interacción entre la tendencia política y los años de escolaridad de la fuerza de trabajo presentada en la Figura 4 (a), se observa que las tendencias ‘ninguna’ e ‘izquierda’ presentan un comportamiento diferente cuando la variable nacional o de segundo nivel es observada. En países con una fuerza de trabajo más educada, los individuos con una ten-

dencia de izquierda tienen menores probabilidades de apoyar la integración que aquellos individuos de la misma preferencia política de países con trabajadores menos educados. Asimismo, los odds para el apoyo a la integración para personas que declaran ser de izquierda comparado con las personas que se declaran de derecha son prácticamente los mismos (odds=1.00) en aquellos países con una fuerza de trabajo más instruida, mientras los odds para este mismo ratio (personas de izquierda versus personas de derecha) son 1.38 veces más altas en países con trabajadores menos educados.

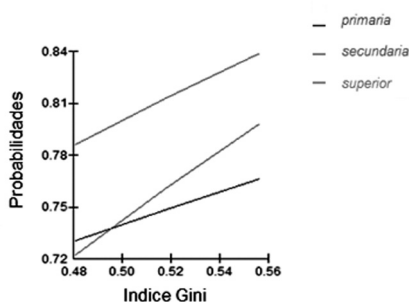
De la misma forma, cuando el efecto multiplicativo de la tendencia política con la distribución del ingreso es analizado, las probabilidades de apoyar la integración económica presentan una tendencia opuesta para individuos que se autodenominan de izquierda comparados con los de derecha y de centro, dependiendo de la distribución del ingreso del país. Como se puede ver en la Figura 4 (b), en aquellos países con una distribución del ingreso más equitativa, los odds para el apoyo a la integración para las personas de izquierda comparados con las personas de derecha son 1.59 veces más altos. En contraste, cuando el mismo odds ratio es analizado dentro de aquellos países con una peor distribución de ingreso, los odds para izquierdistas son 0.81 más altos (19 por ciento más bajos) que para derechistas.

Figura 4. Efecto de interacción de los años de estudio de la fuerza de trabajo y de la distribución del ingreso con la tendencia política, en el apoyo a la integración económica.



Finalmente, cuando el índice Gini es combinado con la variable individual nivel educacional (ver Figura 5) el efecto positivo de este término indica que mientras más desigual es un país en términos de distribución del ingreso, mayores son las probabilidades de apoyar la integración económica de la región para las tres categorías de educación analizadas. Sin embargo, la relación positiva entre educación secundaria y el índice Gini es más intensa en relación a las otras categorías de educación.

Figura 5. Efecto de interacción de la distribución del ingreso y el nivel educacional en el apoyo a la integración económica.



A partir de lo revisado, las correlaciones intra-clase y las variaciones de las tendencias de ciertas variables individuales cuando las variables nacionales son consideradas (efecto de interacción entre niveles), sugiere que la estructura multinivel debe ser considerada cuando los estudios comparativos sobre opinión pública son llevados a cabo. De otra manera, esta relevante información quedará solamente implícita en el análisis y podría generar, tal como se mencionara anteriormente, que relaciones espurias sean validadas o bien que las conclusiones del análisis pierdan robustez. El estudio presentado como ejemplo nos muestra que la opinión pública, así como tantos otros fenómenos sociales, es un problema complejo y multidimensional y que, por esa razón, las falacias atomistas o ecológicas deben ser evitadas o controladas a través de la inclusión de los múltiples niveles de la realidad dentro del análisis.

Conclusiones

La complejidad y las estructuras que encontramos en la realidad suponen un desafío para su abordaje desde la investigación social. Las ciencias sociales y sus métodos buscan dar un tratamiento riguroso de los fenómenos que observan; consecuencia de ello es la emergencia de los modelos multiniveles como respuesta al desafío de aumentar la precisión de los estudios e integrar esta complejidad en un único análisis.

La utilización de las técnicas de Análisis Multinivel tiene una serie de ventajas y múltiples implicancias para el análisis de datos cuantitativos. Uno de ellos es la posibilidad de modelar la variabilidad de acuerdo a los contextos, permitiéndose estimar la influencia que los distintos escenarios ejercen sobre un fenómeno estudiado. Otra ventaja es la integración de variables ecológicas e individuales en un único modelo, tal como se demostró con el estudio expuesto, así como observar los términos de interacción entre ambos tipos de variables.

La integración latinoamericana es solo parte de la enorme variedad de temáticas que pueden ser abordadas a partir de la aplicación del análisis multinivel. La incorporación de esta técnica dentro del set básico de instrumentos con los que cuentan los científicos sociales constituye un aporte significativo a la calidad, amplitud y relevancia de la investigación social. Esperamos que este trabajo haya logrado el objetivo de estimular la reflexión respecto al uso de esta técnica de investigación, su potencial y su conveniencia para el quehacer sociológico.

Referencias

- Aparicio, A.; Morera, M. 2007. La conveniencia del análisis multinivel para la investigación en salud: una aplicación para Costa Rica. *Poblacion y Salud en Mesoamerica*.
- Brinegar, A.P.; Jolly, S.K. 2005. Location, Location, Location National Contextual Factors and Public Support for European Integration. *European Union Politics* 6, 155–180.

- CEPAL, 2010. Statistical yearbook for Latin America and the Caribbean, 2010.
- De la Cruz, F., 2008. Modelos multinivel. *Revista Peruana de Epidemiología* 12.
- Goldstein, H., 2011. *Multilevel Statistical Models*. John Wiley & Sons.
- Hedeker, D., 2007. Multilevel Models for Ordinal and Nominal Variables, en: Leeuw, J. de, Meijer, E. (Eds.), *Handbook of Multilevel Analysis*. Springer, pp. 341–376.
- Hill, P.W.; Rowe, K.J., 1996. Multilevel Modelling in School Effectiveness Research. *School Effectiveness and School Improvement* 7, 1–34.
- Hox, J.J., 2002. *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates.
- Leeuw, J. de, Meijer, E. (Eds.), 2007. *Handbook of Multilevel Analysis*. Springer.
- Martín, J.J.M.; González, M. del P.L. del A.; Mochón, L.G.; Garrido, M.N.M.; Sánchez, F.J., 2009. Análisis multinivel de la influencia de características individuales, capital social y privación en el estado de salud percibida en España. 118 p.
- Milliken, G.A.; Johnson, D.E., 1993. *Analysis of Messy Data: Designed Experiments*. Taylor & Francis.
- Murillo, J., 2008. Los modelos multinivel como herramienta para la investigación educativa. *Magis, Revista Internacional de Investigación en Educación* 1, 45–62.
- Paterson, L.; Goldstein, H., 1991. New Statistical Methods for Analysing Social Structures: an introduction to multilevel models. *British Educational Research Journal* 17, 387–393.
- Rasbash, J.; Steele, F.; Browne, W.; Goldstein, H., 2009. *A User's Guide to MLwin*. Version 2.10. University of Bristol, Bristol.
- Rojas, F., 2006. *La Integración latinoamericana: visiones regionales y subregionales*. Editorial Juricentro.
- Sánchez-Cuenca, I., 2000. The Political Basis of Support for European Integration. *European Union Politics* 1, 147–171.
- Snijders, T.A.B.; Bosker, R.J., 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE.
- Twisk, J.W.R., 2006. *Applied Multilevel Analysis: A Practical Guide*. Cambridge University Press.

