# ANÁLISIS MULTINIVEL DE MOVIMIENTOS MIGRATORIOS: CONSIDERACIONES Y ESTRATEGIA.

Minerva Montero[*], Ana Boquet[**], Adriana Martínez[***].
*Instituto de Cibernética, Matemática y Física
**Instituto de Planificación Física
***Oficina Nacional de Estadística

**ABSTRACT**
This paper considers the use of multilevel models in a demography research. Several recommendations related to the applications and interpretations of the multilevel models are discussed, also describes the main characteristics as well as the assumptions in each model. Attention focuses on the potential of the multilevel approach to explain the possible contextual effects in relation to individual factors and its flexibility to incorporate more integrals strategies. It is concluded that multilevel models are a powerful tool to the analysis of demographic data, particularly studies of migratory movements.

**KEY WORDS:** multilevel models, hierarchical models, contextual analysis, demography studies.

**MSC:** 62P25

**RESUMEN**
En este trabajo se considera el uso de los modelos multinivel en una investigación demográfica. Se discuten algunas recomendaciones relacionadas con la aplicación e interpretación de los modelos multinivel y se describen las características y los supuestos esenciales de cada modelo propuesto. La atención se centra en el potencial del enfoque multinivel para explicar los posibles efectos contextuales en relación con los factores individuales y su flexibilidad para incorporar estrategias más generales. Se concluye que los modelos multinivel ofrecen una poderosa herramienta para el análisis de datos demográficos, en particular para el estudio de movimientos migratorios.

## 1. INTRODUCCIÓN

En las investigaciones sociales más recientes se le concede gran importancia a la relación individuo-contexto (Chaix y Chauvin, 2002; Lumme y col., 2008). La idea fundamental es que las interacciones entre los individuos y sus contextos influyen en el comportamiento individual, a la vez que conforman características y propiedades de grupos (Macintyre, 1986; Duncan y col., 1998). Los individuos y los grupos sociales son percibidos como un sistema jerárquico de individuos y grupos, definidos en niveles separados. Es bien conocido que cuando se analizan datos jerárquicos, ignorar los grupos, o sea, tratar los datos como una muestra aleatoria proveniente de una única población, puede provocar serios problemas inferenciales (Snijders y Bosker, 1999).

La necesidad de investigar la interacción entre las variables que describen a los individuos y las que describen el contexto, considerando la falta de independencia de las respuestas dentro de los contextos, conduce entonces a un tipo de investigación en la que es recomendable la aplicación de los modelos multinivel o modelos jerárquicos (Goldstein, 1995; Bryk. y Randenbush, 1992). En este artículo se considera el caso particular del modelo logit multinivel (Wong y Mason, 1985).

El objetivo de este trabajo es mostrar algunas ideas para el diseño e implementación de una estrategia de análisis de datos jerárquicos en una investigación contextual. El problema que se presenta es parte de un estudio de migraciones internas del Instituto de Planificación Física. La discusión se orienta al análisis de aquellas cuestiones que surgen durante la aplicación de la modelación multinivel para estudiar las relaciones entre los individuos que migran a Ciudad de La Habana y el contexto en el cual ellos viven.

En la sección 2 se describe el problema y el conjunto de datos disponible; así como la teoría básica referida a los modelos multinivel utilizados. En la sección 3 se muestra cómo los componentes de la varianza no sólo pueden usarse como indicadores de la variación existente entre los grupos, sino también como guía para encontrar características interesantes en los datos. Se discute además el caso en que para alcanzar un buen ajuste del modelo, es necesario controlar diferentes factores de confusión. Los

resultados se exponen en la sección 4. En la sección 5 se discute, en base a los resultados obtenidos, la flexibilidad de la modelación multinivel para investigar el efecto de las variables consideradas en el estudio. Finalmente, se hacen algunas recomendaciones generales.

## 2. DATOS Y METODOLOGÍA

### 2.1. Movimientos migratorios internos

Los movimientos de la población constituyen un tema de especial interés para muchos especialistas de diversas esferas; en particular, para quienes se ocupan del Ordenamiento Territorial, ya que las migraciones representan una respuesta territorial de la población a las transformaciones socioeconómicas (Boquet, 1997).

A pesar de ser la migración interna, un tema que tiene una considerable base informativa, y sobre la cual se han realizado numerosos estudios e investigaciones, aún no se cuenta con suficientes elementos para llegar a conclusiones sobre la conveniencia de los flujos que se producen, y consecuentemente, si deben ser estimulados o no, en una política de distribución espacial de la población (CEDEM, 1997; Álvarez, 1998; Montes y López-Callejas, 1999). La incorporación de métodos estadísticos a los estudios de ordenamiento territorial, le ofrece al investigador una herramienta muy útil para analizar los enormes volúmenes de información que se generan hoy día.

El estudio de los flujos migratorios desde las Ciudades de Interés Nacional[1] (CIN) hacia Ciudad de La Habana es uno de los problemas en que es necesario manejar una gran cantidad de datos para lograr llegar a conclusiones sobre las características de los migrantes y el papel que juegan las circunstancias particulares de cada una de las ciudades. Estos elementos se incorporan a los diagnósticos, tanto de los estudios nacionales de ordenamiento territorial como al planeamiento provincial y urbano, que deben establecer las acciones y medidas a tomar en aquellos casos donde se identifiquen desequilibrios importantes.

Históricamente, en la isla se han producido migraciones hacia la capital, cuya procedencia se conoce en términos de grandes regiones o provincias. En este trabajo se busca identificar con un mayor nivel de detalle, una lógica migratoria que explique no sólo cómo se mueven los cubanos, sino quiénes lo hacen.

### 2.2. Estructura de los datos

El análisis de la relación entre el contexto y los individuos se realizó siguiendo un enfoque multinivel, en el cual, los individuos y las CIN se consideran como dos niveles separados de un sistema jerárquico (Hox y Kreft, 1994). Los individuos representan las unidades de nivel-1 y las ciudades representan las unidades de nivel-2.

Cuando el problema comprende dos niveles, se diferencian dos tipos de covariables: las "contextuales" o de nivel-2, cuyos valores en este caso, varían de ciudad a ciudad (siendo el mismo para todos los individuos de una misma ciudad), y las "individuales" o de nivel-1 (cuyos valores varían entre los individuos de la misma ciudad).

Las variables individuales (y sus categorías) seleccionadas para el estudio se muestran en la Tabla 1. Las variables contextuales (y sus valores máximos y mínimos) se muestran en la Tabla 2.

La metodología propuesta se aplica a datos del año 2000. Para la selección de la muestra de datos se consideró una estrategia de muestreo en dos etapas, donde se selecciona una muestra simple aleatoria de CIN y dentro de cada ciudad, se selecciona una muestra simple aleatoria de individuos. Los datos de cada migrante se obtuvieron a través de la estadística continua que ofrece la Oficina Nacional de Estadística (ONE). Para la etapa estudiada se seleccionó una muestra de 27 CIN y 12087 individuos. Por los resultados satisfactorios que se obtuvieron, esta metodología continuará aplicándose en el Instituto de Planificación Física para contrastar la evolución de las condiciones del entorno.

---

[1] Las Ciudades de Interés Nacional (CIN) son asentamientos de gran importancia para el ordenamiento de ámbito nacional por el papel que juegan como centros territoriales (Bermúdez, E., 2003).

| | | |
|---|---|---|
| Variable Respuesta | Y | 1 = Emigran hacia Ciudad Habana<br>0 = Emigran hacia otra provincia |
| Sexo | SEXO | 1 = Femenino<br>0 = Masculino |
| Edad | Edad1 | 1 = 15-24 años<br>0 = Otros |
| | Edad2 | 1 = 25-34 años<br>0 = Otros |
| | Edad3 | 1 = 35-44 años<br>0 = Otros |
| | Edad4 | 1 = 45-59 años<br>0 = Otros |
| Ocupación | Ocup1 | 1 = Trabajador agropecuario<br>0 = Otros |
| | Ocup2 | 1 = Profesionales, técnicos y dirigentes<br>0 = Otros |
| | Ocup3 | 1 = Trabajadores administrativos y de servicios, obrero no agropecuario.<br>0 = Otros |
| | Ocup4 | 1 = Busca empleo<br>0 = Otros |
| | Ocup5 | 1 = Ama de casa<br>0 = Otros |
| | Ocup6 | 1 = Estudiantes<br>0 = Otros |

Tabla 1 Variables individuales para los migrantes de las CIN.

### 2.3. Análisis multinivel

La estructura impuesta por el diseño jerárquico, sugiere la siguiente organización de los datos: Se tiene un conjunto de $J = 27$ CIN (unidades de nivel-2). Cada una de las CIN está subdividida en $I$ subpoblaciones, determinadas por las clases de una única variable nominal (Sexo, Edad u Ocupación), o el producto de una clasificación cruzada de varias de estas variables. De cada subpoblación se selecciona una muestra de $n_{ij}$ ($i$=1,⋯,$I$; $j$=1,⋯,27) individuos. Todos los individuos con la misma combinación de valores se tratan como un subgrupo en los datos. Por tanto, cada unidad de nivel-2, incluye $I$ subgrupos (unidades de nivel-1). Se supone que los individuos de cada subgrupo se clasifican en una de las dos categorías de la variable respuesta dicotómica $Y$ (Emigración). La información se resume en 27 tablas de contingencia $I$×2.

| Variables contextuales | Min. | Max |
|---|---|---|
| Envejecimiento | 3.73 | 26.57 |
| Densidad | 21.5 | 168.9 |
| Área | 273.7 | 6600 |
| Residencia Formal | 7.2 | 931.3 |
| Centros | 0.4 | 40.93 |
| % Zona Industrial | 2 | 51.84 |
| % Mejores viviendas | 39.3 | 94 |
| % Viviendas en buen estado | 42 | 87.9 |
| % Viviendas con agua | 20.8 | 100 |
| % Viviendas alcantarillado | 8.57 | 100 |
| % Vivienda Teléfono | 2.2 | 95 |
| % Empleo Femenino | 29 | 70 |
| % Construcción de viviendas vías formales | 21 | 100 |
| % Facultades | 0 | 36 |
| % Hospitales especializados | 0 | 9 |
| % Policlínicos | 0 | 9 |

Tabla 2: Variables contextuales para las CIN.

Sea $p_{ij}$ la proporción de individuos con la respuesta 1 en el subgrupo $i$ de la CIN $j$, y $\pi_{ij}$ su valor esperado. Para modelar la variación de las proporciones se utilizó la función logit. En la siguiente sección se ofrecen detalles de los modelos utilizados, siguiendo la formulación de Snijders y Bosker , (1999).

### 2.3.1. Modelo "nulo"

Si no se toman en cuenta las variables explicativas, entonces la probabilidad de éxito $\pi_{ij}$ (emigrar hacia Ciudad de La Habana) es constante en cada CIN, luego, la probabilidad de éxito en la CIN $j$ se denota por $\pi_j$. Aquí se considera que las CIN se toman de una población de CIN y las probabilidades de éxito en ellas se reconocen como variables aleatorias definidas en esa población. El modelo nulo para una variable respuesta dicotómica se refiere a la población de CIN y especifica la distribución de probabilidad para las $\pi_j$. Esto se expresa para una función general $f(\pi_j)$ por la fórmula:

$$logit(\pi_j) = \gamma_0 + u_{0j} ,$$

donde, $\gamma_0$ es el promedio poblacional de las probabilidades transformadas y $u_{0j}$, la desviación aleatoria de este promedio con respecto al promedio para la CIN $j$. Si $f$ es la función logit, entonces $f(\pi_j)$ es el log-odds para la CIN $j$, luego, para la función logit, el log-odds tiene una distribución normal en la población de CIN. Las desviaciones $u_{0j}$ se suponen errores aleatorios independientes con distribución normal y varianza $\sigma_0^2$.

Se denota por $\pi_0$ la proporción correspondiente al valor promedio $\gamma_0$, definido por:

$$f(\pi_0) = \gamma_0 .$$

Para la función logit, esto significa que $\pi_0$ es la llamada transformación logística de $\gamma_0$, definida por:

$$\pi_0 = logistic(\gamma_0) = \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}.$$

Aquí $\exp(\gamma_0) = e^{\gamma_0}$ denota la función exponencial, donde $e$ es la base del logaritmo natural. Las funciones logística y logit son cada una inversa de la otra.

### 2.3.2. Modelo "intercepto y/o pendiente aleatoria"

Cuando las subpoblaciones son el resultado de cruzar, por ejemplo, dos factores dicotómicos potencialmente explicativos de las proporciones observadas, entonces se podría ajustar un modelo de "efectos principales" en la parte fija, que estaría constituido por un intercepto y un coeficiente indicador de cada factor. Para una representación multinivel de esta situación, el modelo de nivel-1 se escribe como:

$$logit(\pi_{ij}) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} , \quad (5)$$

donde el coeficiente intercepto podría ser aleatorio en el nivel-2, así:

$$\beta_{0j} = \gamma_0 + u_{0j} . \quad (6)$$

Las ecuaciones (5) y (6) se pueden generalizar si se incluyen múltiples predictores de nivel-1. Considerando $t$ predictores dicotómicos de nivel-1, $X_k\ (k = 1,2,\cdots,t)$, entonces el modelo de nivel-1 está dado por:

$$logit(\pi_{ij}) = \beta_{oj} + \sum_{k=1}^{t} \beta_{kj}x_{kij} \quad . \quad (7)$$

El modelo de nivel-2 para el intercepto está dado por la ecuación (6) y el modelo de nivel- 2 para las pendientes está dado por:

$$\beta_{kj} = \gamma_k , \quad \text{para } k = 1,2,\ldots,t. \quad (8)$$

Sustituyendo las ecuaciones (6) y (8) en la ecuación (7) se obtiene el "modelo de intercepto aleatorio":

$$logit(\pi_{ij}) = \gamma_0 + \sum_{k=1}^{t} \gamma_k x_{kij} + u_{0j} \quad . \quad (9)$$

Por tanto, un nivel de diferencia entre las categorías de cada variable indicadora $X_k$ , está asociado con una diferencia de $\gamma_k$ en el log-odds o, equivalentemente, un cociente de $exp(\gamma_k)$ en sus odds. Se supone que las desviaciones $u_{0j}$ tienen media 0 (dados los valores de todas las variables explicativas) y una varianza de $\sigma^2_{u_0}$.

Los modelos en los cuales los coeficientes de regresión también varían aleatoriamente entre las unidades de nivel-2 se llaman modelos de pendientes aleatorias.

### 2.3.3. Modelo "general"

Los modelos presentados hasta ahora se pueden generalizar incluyendo múltiples variables contextuales. Al suponer que hay $q$ predictores de nivel-2 $z_{sj}$ (s=1,2,⋯,q), entonces, el modelo de nivel-2 para el intercepto está dado por:

$$\beta_{oj} = \gamma_{00} + \sum_{s=1}^{q} \gamma_{0s} z_{sj} + u_{oj} \,, \tag{11}$$

y el modelo de nivel-2 para las pendientes aleatorias está dado por:

$$\beta_{kj} = \gamma_{k0} + \sum_{s=1}^{q} \gamma_{ks} z_{qs} + u_{kj} \,. \tag{12}$$

Sustituyendo (12) y (11) en (7) se obtiene el "modelo logit multinivel general".

$$logit(\pi_{ij}) = \gamma_{00} + \sum_{k=1}^{t} \gamma_{k0} x_{ki} + \sum_{s=1}^{q} \gamma_{0s} z_{sj} + \sum_{s=1}^{q} \sum_{k=1}^{t} \gamma_{ks} z_{sj} x_{ki} + u_{oj} + \sum_{k=1}^{t} u_{kj} x_{ki} \,.$$

Se supone que los errores $u_{kj}$ (k=0,1,⋯,t) tienen esperanza cero y varianza $\sigma^2_{u_k}$. La covarianza de $u_{k_j}$ y $u_{k_j'}$ (k≠k') se denota por $\sigma_{u_{kk'}}$.

### 3. ALGUNAS CONSIDERACIONES

**Análisis de la varianza contextual:** Las estimaciones de los componentes de la varianza se usan principalmente como indicadores de la variación existente entre las CIN. En cada modelo se reportan sus valores estimados para mostrar cuánta varianza residual queda como un potencial para ser "explicado" por variables contextuales. En este artículo, la estimación de los componentes de la varianza se usa además para detectar, y como consecuencia, explorar, diferencias entre contextos o grupos, tales como el caso de las CIN, cuando una varianza grande puede estar indicando que algunas CIN se desvían considerablemente de la curva de regresión media.

**Control de los factores de confusión:** En ocasiones, después de realizar un análisis multinivel, las estimaciones de los parámetros asociados a las variables explicativas en el nivel contextual no dicen mucho sobre las causas de la heterogeneidad entre los grupos. El interés de la investigación no debe entonces centrarse sólo en el nivel de agregación, porque es posible que además de los factores que se consideraron en el diseño del estudio, existan otros, asociados tanto a las variables explicativas, como a las de respuesta, a éstos se les conoce como factores de confusión y también deben ser considerados.

Cuando se ha detectado la presencia de algún factor de confusión[2], uno de los métodos más utilizados para su control en la etapa de análisis, consiste en agrupar la muestra en conjuntos que son internamente homogéneos respecto al factor de confusión. A los conjuntos así formados se les llamará estratos. Si se supone que los estratos son diferentes, entonces es necesario ajustar un modelo global (Montero y col., 1999) que describa cómo varían los parámetros de acuerdo a los cambios en los niveles del factor de confusión (o de cualquier otro factor que interactúe y sea de interés para el estudio).

Si se tienen L estratos, sean $W_1, W_2, \ldots, W_{L-1}$, las variables indicadoras que permiten definir la pertenencia a los L estratos.

$$W_l = \begin{cases} 1 & si\ y_{ij}\ es\ una\ obsevación\ obtenida\ en\ el\ estrato\ l-ésimo. \\ 0 & en\ cualquier\ otro\ caso. \end{cases}$$

donde $y_{ij}$ es el valor de la variable respuesta Y para la i-ésima subpoblación de la j-ésima tabla. Si se consideran $X_1, X_2, \ldots, X_t$, las variables explicativas medidas al nivel de las unidades del primer nivel, se tendrá el siguiente modelo general:

$$log\,it(\pi_{ij}) = \beta_{0j} + \sum_{k=1}^{t} \beta_{kj} x_{ki} + \sum_{I}^{L-1} \alpha_l \omega_{lij}, \quad i = 1,2,\ldots,I, \quad j = 1,2,\ldots,J.$$

[2] Se llamará confusión al sesgo que aparece en la estimación de los parámetros cuando la relación entre las variables explicativas (nivel 1 y 2) y la respuesta está distorsionada por los efectos de algún factor.

donde $\omega_{lij}$ es el valor de la variable $W_l$ para la i-ésima subpoblación de la j-ésima tabla. Nótese que los coeficientes $\alpha_l$, $l$=1,2,…,$L$-1, son fijos; o sea, no cambian de una tabla a otra, como es el caso de los $\beta_j = (\beta_{1j}, \beta_{2j}, \ldots, \beta_{tj})$, los cuales son llamados coeficientes aleatorios. Las ecuaciones en el segundo nivel se definen como en (11) y (12), que constituyen $t$ ecuaciones en las variables Z, cuyas observaciones son tomadas al nivel de grupo. Al aplicar este modelo queda eliminado el "ruido" correspondiente a las variables que representan al factor de confusión cuyo efecto en sí mismo no es de interés para el estudio.

## 4. RESULTADOS

Los detalles de cada modelo utilizado pueden encontrarse en Martínez (2004). El primer modelo que se propone es un modelo nulo, cuya varianza de nivel-1 estimada (0.918) es cercana a la unidad. Esto significa que no existe evidencia de variación extra-binomial (Goldstein, 1995; Jacob, 2000). En lo adelante, la varianza en el nivel-1 se fija a 1.0, lo cual es equivalente a suponer que los datos siguen la distribución Binomial, por tanto, $\sigma_e^2$ no tiene aquí interpretación útil.

En la Tabla 3 se presentan las estimaciones para cinco modelos diferentes. El modelo A es del tipo "modelo nulo" (Hox, 1995). La varianza estimada, indicadora de la variación existente entre las CIN, es estadísticamente significativa, lo que sugiere que la proporción de personas que emigran hacia Ciudad de La Habana varía entre las CIN. El parámetro fijo estimado (-1.18) se refiere a la distribución subyacente definida por la función logit y no a las proporciones. Para determinar la proporción esperada, se debe usar la transformación inversa a la función logit.

Por otro lado, las relaciones encontradas entre la variable respuesta y las variables explicativas individuales no parecen cambiar de una ciudad a otra, las diferencias entre las CIN se expresan sólo en términos de los interceptos. El modelo B incluye la variable individual Sexo, suponiendo fijas las pendientes de la regresión; el intercepto representa la condición para la cual la variable explicativa Sexo es cero, por tanto, el valor -1.33 para el intercepto, estima la respuesta esperada para los hombres. El signo del coeficiente estimado, asociado a la variable Sexo, confirma lo que se esperaba, o sea, ser mujer contribuye a aumentar la probabilidad de emigrar hacía Ciudad de La Habana, dado el potencial de empleo terciario en la ciudad.

El modelo C adiciona las variables indicadoras Edad del emigrante, en este caso el intercepto representa la condición en la cual ambas variables explicativas (Sexo y Edad) son cero. Los coeficientes estimados para cada una de las variables relativas a Edad, indican que la probabilidad de emigrar hacia Ciudad de La Habana es mayor entre los miembros del grupo Edad 4 (45-59 años), seguido por el grupo Edad 3 (35-44 años), el grupo Edad 2 (25-34 años) y el grupo Edad 1 (15-24 años).

| Parámetros | Modelo A | Modelo B | Modelo C | Modelo D | Modelo E |
|---|---|---|---|---|---|
| Fijos | | | | | |
| *Constante* | -1.18 (0.10) | -1.33 (0.11) | -1.55 (0.11) | -3.48 (0.27) | -4.13 (0.38) |
| *Variables individuales* | | | | | |
| Sexo | | 0.272 (0.04) | 0.30 (0.04) | -0.06 (0.06) | |
| Edad 2 | | | 0.11 (0.05) | 0.09 (0.06) | |
| Edad 3 | | | 0.37 (0.06) | 0.28 (0.07) | 0.22 (0.05) |
| Edad 4 | | | 0.55 (0.07) | 0.47 (0.08) | 0.41 (0.06) |
| Ocup 2 | | | | 3.06 (0.26) | 3.04 (0.26) |
| Ocup 3 | | | | 2.60 (0.26) | 2.59 (0.26) |
| Ocup 4 | | | | 1.04 (0.26) | 1.03 (0.26) |
| Ocup 5 | | | | 2.20 (0.26) | 2.14 (0.25) |
| Ocup 6 | | | | 2.43 (0.27) | 2.34 (0.26) |
| *Variables contextuales* | | | | | |
| Envejecimiento | | | | | 0.05 (0.02) |
| Aleatorio | | | | | |
| Nivel CIN ($\sigma_u^2$) | 0.28 (0.08) | 0.28 (0.08) | 0.27 (0.08) | 0.25 (0.07) | 0.20 (0.06) |
| -2 log (verosimilitud) | | | | | |

Tabla 3 Resultados de ajustar los modelos A, B, C, D y E.

El modelo D introduce además la variable indicadora Ocupación. Como en el modelo anterior, las categorías usadas para las variables indicadoras hacen necesaria la referencia a una condición ¨básica¨: ser

hombre entre 15-24 años, de ocupación obrero agropecuario. Note que de las tres variables contempladas en el modelo, la Ocupación es la que hace una contribución más significativa a la ecuación de regresión, sin embargo, ahora las variables Sexo y Edad 2 no aparecen como variables explicativas determinantes, ya que los coeficientes estimados asociados a estas variables no son significativos. De los parámetros estimados se puede inferir que los individuos en la segunda categoría ocupacional (profesionales, técnicos y dirigentes) tienen una probabilidad mayor de emigrar hacia Ciudad de La Habana que cualquier otra categoría ocupacional.

La variación del intercepto $\sigma_u^2$ entre las CIN sigue siendo estadísticamente significativa, y podría ser explicada por variables en el segundo nivel. En cada CIN se determinó el número de personas mayores de 60 años por cada 100 habitantes, este indicador de población fue el único entre los 15 considerados que tuvo una contribución significativa a la ecuación de regresión. Los parámetros estimados del modelo incluyendo la variable Envejecimiento se muestran en la última columna de la Tabla 3 (modelo E). El signo positivo del parámetro estimado, revela que la probabilidad de emigrar hacia Ciudad de La Habana es mayor si el migrante proviene de una ciudad envejecida. La varianza del intercepto ha disminuido pero continúa siendo significativa. La persistencia de variación significativa en el intercepto, después de la inclusión de las variables del nivel grupal, sugiere que otros factores de nivel grupal –posiblemente responsables de esa variación- necesiten ser explorados. Desafortunadamente, no se cuenta con más información acerca de las características de las CIN, por lo que es imposible definir nuevas variables al nivel-2, que puedan explicar la variación entre las CIN.

Teniendo en cuenta la dificultad para el planteamiento de nuevos modelos, estudios exploratorios más exhaustivos detectaron que la zona geográfica (Occidente, Centro y Oriente) resultó ser un posible factor de confusión. Para estudiar el efecto de la zona, que se considera fijo, se crean dos variables indicadoras. La Tabla 4 muestra las estimaciones del modelo F incluyendo estas variables. El signo negativo de los coeficientes estimados confirma que la probabilidad de emigrar hacia Ciudad de La Habana es mayor para los individuos procedentes de la zona Occidental.

| Parámetros | Modelo E | Modelo F |
|---|---|---|
| Fijos | | |
| *Constante* | -3.55 (0.33) | -3.63 (0.29) |
| *Variables individuales* | | |
| Sexo | | -0.10 (0.07) |
| Edad 2 | | 0.05 (0.07) |
| Edad 3 | 0.22 (0.05) | 0.24 (0.07) |
| Edad 4 | 0.41 (0.06) | 0.43 (0.08) |
| Ocup 2 | 3.04 (0.26) | 3.09 (0.28) |
| Ocup 3 | 2.59 (0.26) | 2.64 (0.28) |
| Ocup 4 | 1.03 (0.26) | 1.15 (0.28) |
| Ocup 5 | 2.14 (0.25) | 2.31 (0.29) |
| Ocup 6 | 2.34 (0.26) | 2.46 (0.29) |
| *Variables contextuales* | | |
| Envejecimiento | 0.04 (0.01) | |
| *Estratos* | | |
| Centro | -0.84 (0.15) | |
| Oriente | -0.58 (0.14) | |
| Aleatorio | | |
| Nivel CIN ($\sigma_u^2$) | 0.07 (0.02) | 0.045 (0.01) |
| -2 log (verosimilitud) | | |

Tabla 4 Resultados de ajustar los modelos E y F.

Volviendo al modelo D y calculando los residuos del intercepto en el nivel-2, se construye el gráfico que se muestra en la Figura 1, que muestra los 27 residuos de nivel-2 en orden ascendente, con su intervalo de confianza al 95%. Si se observan los intervalos, se aprecia que, hacía los extremos del eje horizontal, hay un grupo de 16 CIN donde éstos no contienen al cero. Esto significa que más de la mitad de las CIN difieren significativamente de los valores promedios predichos por los parámetros fijos. En particular se

destacan cuatro ciudades con los mayores residuos del intercepto (Pinar del Río, Güines, San Cristóbal y Artemisa) y una (Trinidad), cuyo residuo se encuentra muy por debajo del valor cero esperado.

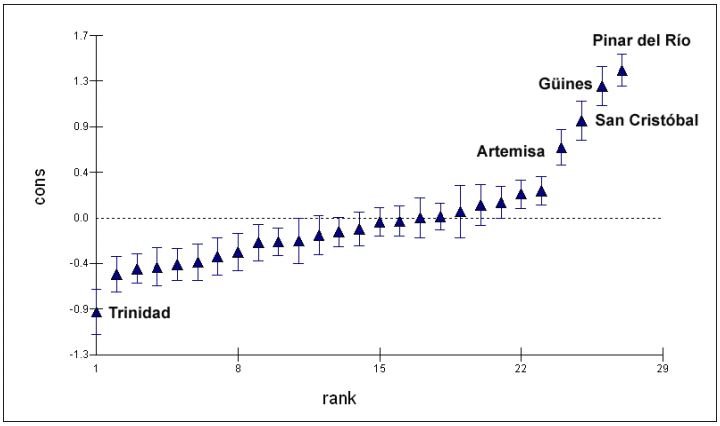


Figura 1: Residuos de las 27 CIN y sus intervalos de confianza

Las estimaciones del modelo F presentadas en la Tabla 4, se obtuvieron después de eliminar de la muestra las cuatro CIN antes mencionadas. Es notable como la varianza estimada pierde significación con respecto a la estimada en el modelo D, lo que corrobora que la variación más importante entre las CIN se debe principalmente a unas pocas ciudades atípicas. La variable Ocupación sigue teniendo un papel determinante al explicar las diferencias individuales en el comportamiento migratorio, aunque esta vez no alcanza a neutralizar el efecto de la variable Sexo. El resto de las variables mantienen las mismas relaciones con la respuesta.

## 5. DISCUSIÓN Y RECOMENDACIONES

Los modelos propuestos permiten explicar claramente el efecto de las variables consideradas en el estudio sobre la emigración de las CIN hacia la Ciudad de La Habana, además de que permiten una descripción de las diferencias entre las CIN.

Se comprobó que la variable explicativa más relevante es la Ocupación. De las personas que deciden migrar desde las ciudades más importantes del país, tienen mayores probabilidades de hacerlo hacia Ciudad de La Habana, los profesionales y técnicos, cuyas ocupaciones son afines con el perfil de la ciudad, como centro administrativo, investigativo y de servicios del primer nivel en el país (Boquet, 2003). La variable Edad del migrante, también resultó significativa. De los grupos de edades prefijados, tienen más probabilidades de dirigirse a Ciudad de La Habana los que se encuentran entre 45 y 59 años. Estas son las personas que cuando deciden migrar, tienen un destino muy preferencial, sin embargo, los jóvenes tienen mayor diversidad de preferencias para el destino de su migración.

Del conjunto de ciudades utilizadas en este estudio, los migrantes con mayores probabilidades de dirigirse a Ciudad de La Habana provienen de Pinar del Río, Artemisa, Güines y San Cristóbal. Los migrantes con menos probabilidades de dirigirse a Ciudad de La Habana fueron los de Trinidad.

Considerando como base la etapa exploratoria, se observan además, claras diferencias entre las zonas geográficas. De las personas que migran, tienen una probabilidad alta de hacerlo hacia Ciudad de La Habana, las que provienen de ciudades de la región occidental. Las probabilidades de las otras dos regiones son mucho más bajas y similares entre sí. Por otro lado, sólo el nivel de envejecimiento de la población ofreció algún elemento de diferenciación, es más probable migrar a Ciudad de La Habana cuando se proviene de las ciudades más envejecidas. Esto puede interpretarse como un indicador sintético del nivel de consolidación de las condiciones urbanas de una ciudad, que van más allá del físico de la ciudad para constituir un modo de vida. También es posible que otras características, no consideradas en el ejemplo, sean las responsables de la diferenciación entre las CIN, por lo que se recomienda indagar con otros indicadores para mayor certeza.

Para evaluar la bondad de ajuste del modelo se chequearon los residuos de nivel-1 y nivel-2 a través de gráficos de diagnóstico. Si las variables explicativas se van incorporando paso a paso, entonces se recomienda hacer un análisis de diagnóstico después de cada modelo propuesto. Este proceder iterativo

permite aumentar gradualmente la complejidad de los modelos. Otro aspecto muy importante a destacar es que los componentes de la varianza pueden utilizarse como guía para detectar CIN atípicas. Es especialmente esta característica, la que hace de los modelos multinivel una nueva herramienta para el desarrollo de teorías en los estudios demográficos.

En este trabajo, los efectos de las variables individuales se tratan como fijas, sin embargo, la relación entre las variables explicativas y la respuesta no tiene que ser la misma para todas las CIN. Para los nuevos estudios se recomienda probar con modelos de coeficientes aleatorios más complejos, incluyendo interacciones entre niveles, verificando siempre la estructura de los datos, en términos del número de variables y el tamaño de la muestra, ya que si no son adecuados, la sobreparametrización podría causar dificultades numéricas.

En general, se recomienda el uso de los modelos multinivel para el análisis de los movimientos migratorios de otros años, ajustando los contextos a las condiciones de los diferentes momentos, también resultará interesante ampliar las descripciones de los contextos para identificar exhaustivamente el factor discriminante entre las CIN. La estrategia propuesta no sólo permitirá el desarrollo de mejores modelos explicativos, sino que puede crear las bases para una mejor política de planeación.

**RECEIVED SEPTEMBER 2009**
**REVISED JUNE 2010**

## REFERENCIAS

[1] ÁLVAREZ, C., (1998): Elementos para la formulación de una política de distribución espacial de la población. IPF, La Habana, inédito.

[2] BERMÚDEZ, E., (2003): Sistema de asentamientos poblacionales En: Diagnóstico Preliminar del ENOT. IPF, La Habana, .

[3] BOQUET, A., (1997): Migraciones Internas: Estudio descriptivo de las migraciones internas en Cuba de 1989 a 1996. IPF, La Habana, 35 pp.

[4] BOQUET, A., (2003): Población y empleo En: Diagnóstico Preliminar del ENOT. IPF, La Habana.

[5] BRYK, A.S. y RANDENBUSH, S.W. (1992). **Hierarchical Linear Models**: **Applications and Data Analysis Methods**. Sage Pub. Newsbury Park.

[6] CEDEM, (1997): Resultados de la encuesta nacional de migraciones según niveles del sistema de asentamientos. CEDEM. La Habana. .

[7] CHAIX, B. y CHAUVIN, P., (2002): The contribution of multilevel models in contextual analysis in the field of social epidemiology: a review of literature. Chauvin **Rev. Epidemiol Santé Publique** 50, 489-99.

[8] DUNCAN, C., JONES K. y MOON G., (1998): Context, composition and heterogeneity: Using multilevel models in health research. **Social Science and Medicine,** 46, 97-117.

[9] GOLDSTEIN H., (1995): **Multilevel Statistical Models**. 2nd. Ed. Halsted Press, New York.

[10] HOX, J. y KREFT, I., (1994): Multilevel analysis methods. **Sociological methods and Research**, 22, 283-299.

[11] HOX, J.J., (1995): **Applied multilevel analysis**. TT- Publikaties. Amsterdam.

[12] JACOB M., (2000): Extra-binomial variation in logistic multilevel models—A simulation. **Multilevel Modelling Newsletter**, 12, 8-14.

[13] LUMME S., LEYLAND A.H. y KESKIMÄKI I., (2008): Multilevel modeling of regional variation in equity in health care. **Med Care**. 46, 976-83.

[14] MACINTYRE, S., (1986): The patterning of health by social position in contemporary Britain: direction for sociological research. **Social Science and Medicine,** 23, 393-415.

[15] MARTÍNEZ A., (2004): Análisis de datos demográficos: un enfoque multinivel. **Tesis de grado**. Facultad de Matemática y Computación, Universidad de la Habana.

[16] MONTERO, M., CASTELL, E. y DÍAZ, M., (1999): Un modelo multinivel multivariado con control de un factor de confusión. **Revista Multiciencia** 4 , 99-53.

[17] MONTES, N. y LÓPEZ CALLEJAS, C., (1999): Población y movilidad territorial en Cuba. CEDEM. La Habana.

[18] SNIJDERS T.A.B. y BOSKER R.J., (1999): **Introduction to basic and advanced multilevel modeling**, Sage, London.

WONG, G. Y., y MASON, W. M. (1985): The Hierarchical Logistic Regression. Model for Multilevel Analysis, **Journal of the American Statistical Association,** 80, 5 13-524.