



II Congreso de la Asociación Latinoamericana de Población

Guadalajara, México, 3 – 5 de Septiembre de 2006

La demografía latinoamericana del siglo XXI Desafíos, oportunidades y prioridades

La regresión logística y los factores determinantes de las migraciones internas para zonas rurales de Cuba

Cristina López-Calleja Hiort-Lorenzen

Centro de Estudios Demográficos. Universidad de La Habana
cristina@cedem.uh.cu

PONENCIA PARA EL EVENTO INTERNACIONAL
II Congreso de la Asociación Latino Americana de Población:
“La demografía latinoamericana del siglo XXI: desafíos, oportunidades y prioridades”
Guadalajara, Méjico, 3 al 5 de septiembre, 2006

Temática específica: **Movilidad y migraciones**

Título: **LA REGRESIÓN LOGÍSTICA Y LOS FACTORES DETERMINANTES DE LAS MIGRACIONES INTERNAS PARA ZONAS RURALES DE CUBA.**

Autora: **Cristina López-Calleja Hiort-Lorenzen¹**
Centro de Estudios Demográficos, Universidad de La Habana

E-mail: cristina@cedem.uh.cu, clopez@comuh.uh.cu

Resumen: Las migraciones internas se han convertido en los últimos quince años en un problema de fundamental importancia para muchos países. Algunos autores plantean que entre los factores que inciden en los migrantes se encuentran el sexo, la edad, el color de la piel, el nivel de instrucción y de educación, la ocupación y la calificación. Hay diferencias factoriales por países en cuanto a las personas que migran según grupos de edad, sexo, escolaridad y la ocupación. Hemos construido tablas cruzadas de movimientos migratorios a partir de los datos de la Encuesta Nacional de Migraciones Internas (ENMI, Cuba, 1995) para estudiar la dinámica de las migraciones internas que tienen su origen en zonas rurales del país, y realizan dos movimientos desde ese origen: hacia el medio rural (RRR), hacia el medio urbano (RUU), hacia el medio rural y luego hacia el medio urbano (RRU), y otro hacia el medio urbano y luego hacia el medio rural (RUR). El análisis, para cada una de estas rutas migratorias, se hace por medio de modelos de regresión logística con sus conocidas razones de disparidad (odds ratio, en inglés). Se construyen los modelos que más se ajustan a los datos, y se analizan sus bondades de ajuste y diagnósticos.

Como resultados de este trabajo se obtuvo que el movimiento migratorio se realiza en familia, ya que para las cuatro rutas migratorias las personas que tienen vínculo conyugal migran más que las que no lo tienen, lo cual ratifica lo obtenido mediante tablas cruzadas en trabajos basados en la ENMI. En las rutas donde aparece como destino el medio urbano las mujeres migran más que los hombres y en la ruta RRR la migración es similar para los dos sexos. Además migran más los jóvenes, también migran más las personas con vínculo laboral en la ruta RUR, y las de mayor calificación en la ruta RUU.

En la década de 1980-90, el llamado **Modelo de Regresión Logística** se ha convertido en un método estadístico de gran aplicación en diversas investigaciones, cuando se tiene una situación donde se requiere analizar la relación existente entre una variable dependiente discreta, (variable respuesta con dos o más valores posibles) y una o más variables independientes o explicativas, que a menudo son llamadas covariables.

En este trabajo se utiliza la regresión logística porque mediante este método se puede analizar la dependencia de una variable binaria que da la condición de las personas de ser migrante o no serlo, de otras variables que en este caso representan las características de esas personas, consideradas éstas en conjunto, es decir simultáneamente.

La variable dependiente Y, solo puede tomar valores 0 ó 1, y se define de la siguiente manera en este trabajo:

$$Y = \begin{cases} 1 & \text{Si la persona migra} \\ 0 & \text{Si la persona no migra} \end{cases}$$

El modelo de regresión logística cuando hay p variables independientes se expresa por²:

$$\ln \frac{\Pi(X)}{1-\Pi(X)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = g(X); \quad \text{siendo} \quad \Pi(X) = \frac{1}{1 + e^{-\left(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p\right)}}$$

Esta expresión lineal en las variables X_i permite estimar la probabilidad de que ocurra el suceso en cuestión cuando las variables independientes toman determinados valores, por supuesto es necesario conocer antes las estimaciones de los llamados **coeficientes logísticos** β_i

Si se tienen variables aleatorias discretas el modelo es más complejo; es necesario definir otras variables llamadas ficticias o del diseño. Por ejemplo si de las p variables X_2 tiene tres categorías y X_3 tiene dos categorías el modelo sería:

$$g(X) = \beta_0 + \beta_1 X_1 + \beta_{21} D_{21} + \beta_{22} D_{22} + \beta_{31} D_{31} + \beta_4 X_4 + \dots + \beta_p X_p$$

¹ Dra. en Ciencias Económicas, Master en Ciencias Matemáticas con mención en Probabilidades y Estadística y Master en Estudios de Población, Profesora e Investigadora Titular del Centro de Estudios Demográficos (CEDEM).

Si se tiene una variable aleatoria con k categorías se definen k-1 variables ficticias.

Las variables ficticias se pueden definir de varias formas, por ejemplo para la variable raza de este trabajo, se definen dos variables D₁ y D₂ por tener dicha variable tres categorías:

		D ₁	D ₂
Blanca	(1)	0	0
Negra	(2)	1	0
Mestiza	(3)	0	1

Todos los coeficientes logísticos beta se estiman en un procesador mediante un programa de Estadística. En el caso de tener el modelo una sola variable independiente, se define el cociente de las llamadas razones logísticas (odds) como **Razón de Disparidad** (odd ratio), mediante la expresión:

$$\frac{\frac{\Pi(1)}{1-\Pi(1)}}{\frac{\Pi(0)}{1-\Pi(0)}} = \Psi = \Psi(1,0)$$

que es la razón de disparidad de migrar respecto a no migrar y se llega a:

$$\Psi = e^{\beta}$$

La razón de disparidad es una medida de asociación que ha encontrado un amplio uso, y ofrece una aproximación de cuanto más frecuente (positivo o negativo) es para la variable respuesta o variable dependiente Y, que haya presencia u ocurrencia de un suceso (X=1), o que haya ausencia o no ocurrencia de dicho suceso (X=0).

Si la razón de disparidad se extrae de una tabla de contingencia de 2x2 se le denomina **razón de disparidad cruda**, también esta razón de disparidad cruda se puede obtener mediante las expresiones de la regresión logística las cuales se obtienen de un programa estadístico computarizado.

Si la variable discreta tiene tres categorías el análisis es más complejo y las razones de disparidad se expresan por:

$$\frac{\frac{\Pi(3)}{1-\Pi(3)}}{\frac{\Pi(1)}{1-\Pi(1)}} = \Psi(3,1) \qquad \frac{\frac{\Pi(2)}{1-\Pi(2)}}{\frac{\Pi(1)}{1-\Pi(1)}} = \Psi(2,1)$$

En cuanto a **la estrategia en la construcción de modelos en la Regresión Logística**, se conoce que una modelación exitosa de una situación compleja basada en un conjunto de datos, tiene que tener en cuenta la especificidad de la rama del saber de la que se trata, de los métodos estadísticos a aplicar y de la experiencia y sentido común de los investigadores.

El criterio para incluir variables en un modelo puede variar de un problema a otro y de una disciplina científica a otra, pero en general se debe seleccionar un mínimo de variables en el modelo de manera que éste resulte numéricamente estable y más fácil de generalizar, dichas variables se deben seleccionar por criterios de expertos y/o según los procedimientos estadísticos. También a partir de las variables seleccionadas se analizan las posibles interacciones, sobre la base de consideraciones estadísticas y prácticas para lograr los mejores modelos según el ajuste y el diagnóstico realizado.

Varias son las **medidas para la bondad de ajuste** y ellas dan una indicación general sobre el ajuste del modelo.

Una **prueba de diagnóstico** muy utilizada es la de *Hosmer-Lemeshow* (1980-82), En el sentido de tener una mejor adherencia a la distribución $\chi^2_{(g-2)}$ es preferible utilizar un agrupamiento basado en los percentiles de las probabilidades estimadas, usualmente con g = 10 grupos. Estos grupos usualmente se denominan **deciles de riesgo**.

Dinámica de los movimientos migratorios con origen rural: Rural-Rural-Rural, Rural-Urbano-Rural, Rural-Rural-Urbano y Rural-Urbano-Urbano.

En este trabajo se ha querido profundizar en los movimientos migratorios, y es por ello, que dada la posibilidad de información que dio la ENMI, se analiza también el movimiento que el migrante realiza antes de llegar al lugar de residencia actual, es decir se analiza, la migración de aquellas personas cuyo lugar de nacimiento está en una zona

² Hosmer, D.W.& Lemeshow, S., (1984), Applied Logistic Regression. Ed. Hohn Wiley & Sons, Inc., Canada.

rural y que ha migrado al menos dos veces, es decir se contempla un movimiento migratorio anterior a su establecimiento en el lugar de residencia actual, y diferente al de origen.

Las **rutas migratorias**³ que se analizan son:

- Personas que nacieron en una zona rural y residieron en una zona urbana antes de migrar de nuevo a otra zona rural (RUR).
- Personas que nacieron en una zona rural y residieron en una zona rural antes de migrar de nuevo a otra zona rural (RRR).
- Personas que nacieron en una zona rural y residieron en una zona rural antes de migrar a una zona urbana (RRU).
- Personas que nacieron en una zona rural y residieron en una zona urbana antes de migrar de nuevo a otra zona urbana (RUU).

El **gráfico 1**, muestra la intensidad de esos movimientos en cinco diferentes períodos de tiempo, establecidos bajo distintos criterios de alcance temporal que responden a intervalos, algunos iguales, y otros se identifican con períodos de alguna significación económica.

Aplicación del Análisis de Regresión Logística

Se considera *el período 1990-1995*, debido a que para los demás períodos considerados en la encuesta, ésta no proporcionaba los datos necesarios según los requerimientos del método estadístico a aplicar.

De los 72 050 individuos de la ENMI se seleccionaron los individuos de 15 a 64 años que hacen un total de 46706 individuos y de éstos se seleccionaron aquellos individuos que cumplen los requerimientos de pertenecer a cualquiera de las cuatro rutas de migración en estudio, los cuales hicieron un total de 11 193 individuos para toda la nación; De esta cantidad se tomaron los 3939 migrantes del período 1990-1995.

Para el análisis de los datos se utilizaron cuatro muestras que corresponden a cada uno de los cuatro grupos de migrantes definidos, incluyéndose en cada grupo, las personas que nacieron en las zonas rurales y no migraron, que fueron 4926 individuos.

La cantidad de migrantes, según las rutas de migración seleccionadas aparece en la **tabla 2**.

El universo en este trabajo para las migraciones internas con origen rural lo constituye toda la población de Cuba que corresponde a las cuatro rutas migratorias especificadas anteriormente incluyéndose las personas que no migraron, que tienen residencia en las zonas rurales, de esta población, se tomó la muestra constituida por los individuos de la ENMI que clasificaron en esas cuatro rutas migratorias, y por los individuos que no migraron del medio rural.

Las variables que aparecen en los modelos son las variables ficticias definidas a partir de las siguientes variables: Sexo, Raza, Edad en grupos, Situación conyugal, Nivel educacional, y Situación laboral.

Al calcular el cociente de las razones de migrar y no migrar se obtienen las razones de disparidad crudas, que al compararlas con las razones de disparidad que se obtienen de los modelos de regresión logística, se tiene que estas últimas proporcionan una mayor precisión acerca del comportamiento de las variables determinantes en las rutas migratorias, ya que **en el análisis de regresión logística se consideran todas las variables simultáneamente**, incluyendo la condición de migrar como variable dependiente.

El análisis de regresión logística incluye la **selección de los modelos**, así como el **ajuste** de éstos, y su **diagnóstico**.

En este trabajo, **las variables se seleccionaron por criterios de expertos, utilizándose en especial para una nueva selección de las variables el método de selección paso a paso** (hacia delante y hacia detrás) en el cual las variables son incluidas o excluidas del modelo en forma secuencial sobre la base de criterios estadísticos.

Este método permite el examen de un gran conjunto de modelos que de otra manera no podrían ser analizados.

Se analizaron también a partir de las variables seleccionadas las posibles interacciones, sobre la base de consideraciones estadísticas y prácticas, incluyéndose aquellas interacciones que mejoraban el modelo.

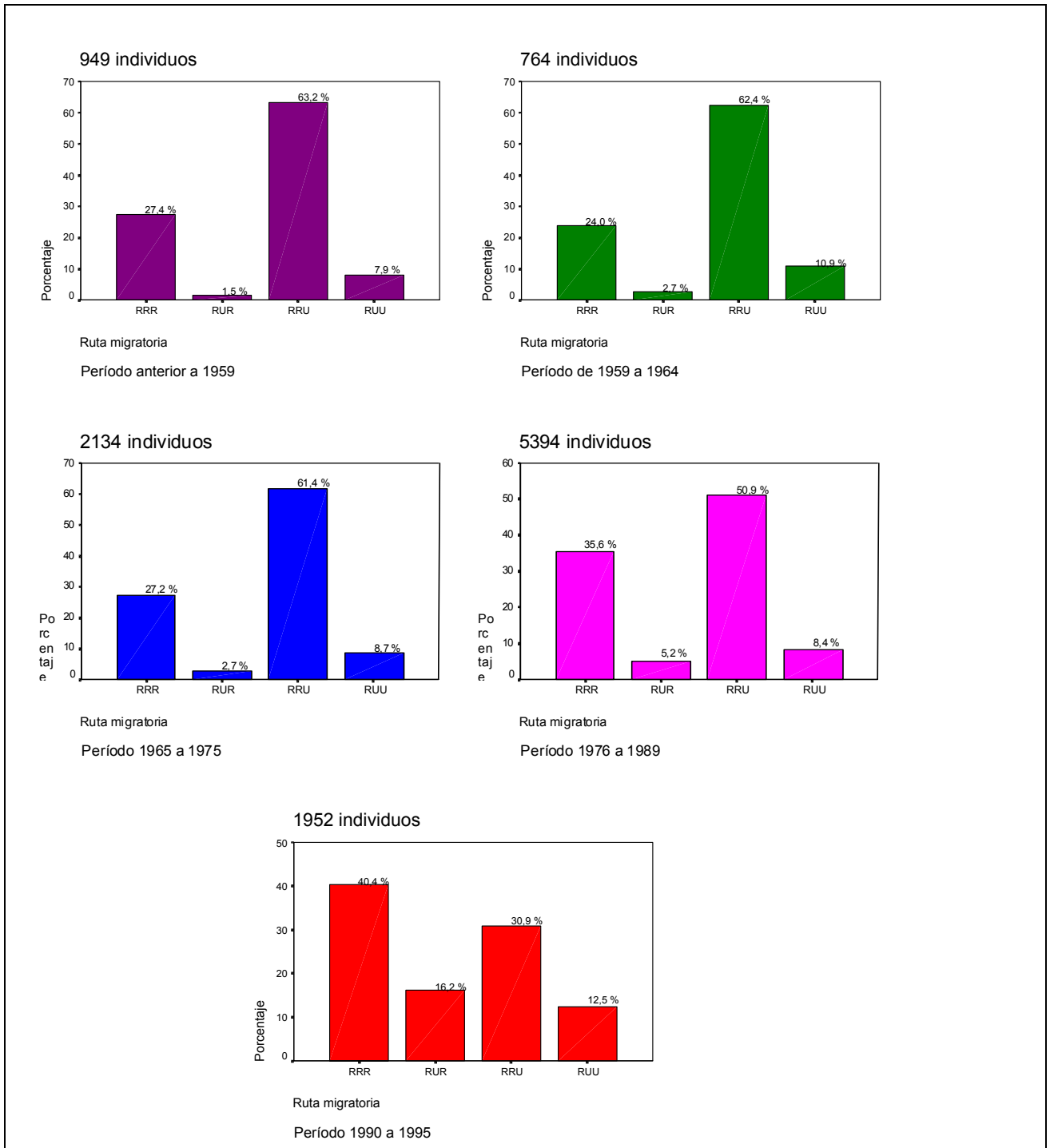
Luego de obtener el modelo de regresión logística **se probó el ajuste completo del modelo**. Para ello se utilizaron una serie de procedimientos estadísticos como son: *el estadígrafo Sumario de las desviaciones residuales (Deviance)*, el estadígrafo R^2 de Cox, Snell y Nagelkerke⁴, y **para el diagnóstico se utiliza la Prueba de Hosmer-Lemeshow de deciles de riesgo**⁵.

³ Se utilizará la denominación de **ruta migratoria** para designar a la corriente o movimiento que realiza la persona que abandona su lugar de nacimiento para establecerse en otro lugar, el cual también abandona para establecerse en el lugar de su residencia en el momento que es entrevistado para la Encuesta Nacional de Migraciones Internas.

⁴ Norusis, M.J., (1997), Manual del módulo SPSS Professional Statistics del paquete de programas del SPSS para Windows, Ed. SPSS Inc.

⁵ Hosmer, D.W. & Lemeshow, S., (1984), Applied Logistic Regression. Ed. John Wiley & Sons, Inc., Canada.

Gráfico 1: Migrantes de origen rural en cinco períodos de tiempo y según los diferentes movimientos o rutas migratorias: Rural-Rural-Rural, Rural-Urbano-Rural, Rural-Rural-Urbano y Rural-Urbano-Urbano.



Fuente: Elaborada por la autora según base de datos de la ENMI, CEDEM-ONE-IPF, 1995

Resultados de la aplicación del método de la Regresión Logística

Al analizar las tablas de la condición migratoria con cada una de las variables seleccionadas, se hallaron las razones de disparidad crudas de dichas tablas, o sea el cociente de las razones de migrar y no migrar, con la categorización que fue necesario introducir para la regresión logística. Más adelante se analizan estas razones de disparidad comparándolas con las obtenidas mediante la regresión logística.

Tabla 2: Posibles rutas migratorias en Cuba, según zonas rural o urbana. Período 1990-1995.

				Zona residencia anterior				Total
				Urbana		Rural		
				Zona de residencia actual		Zona de residencia actual		
Migrante	Zona de nacimiento	Urbano	Cantidad	Urbana	Rural	Urbana	Rural	1987
			% fila	64,9%	23,6%	5,4%	6,1%	
	% columna	84,1%	59,7%	15,0%	13,4%	50,4%		
	% total	32,7%	11,9%	2,7%	3,1%	50,4%		
Total	Rural	Cantidad	244	316	604	788	1952	
	% fila	12,5%	16,2%	30,9%	40,4%	100,0%		
	% columna	15,9%	40,3%	85,0%	86,6%	49,6%		
	% total	6,2%	8,0%	15,3%	20,0%	49,6%		
Total		Cantidad	1533	785	711	910	3939	
		% fila	38,9%	19,9%	18,1%	23,1%	100,0%	
		% columna	100,0%	100,0%	100,0%	100,0%	100,0%	
		% total	38,9%	19,9%	18,1%	23,1%	100,0%	

Fuente: Elaborada por la autora según base de datos de la ENMI, CEDEM-ONE-IPF, 1995

De la salida del programa de regresión logística se obtienen los resultados que aparecen en las **tablas 5, 7, 9 y 11** (Sólo se incluye la tabla 5 por la limitación establecida de seis cuartillas).

Como se puede observar en la primera columna aparecen el nombre de las variables y sus categorías exceptuando la categoría de referencia. Aquí se debe aclarar que *la categoría que se toma como referencia en cada variable, para cada ruta migratoria, es la que tiene valores de cero para todas las variables ficticias que se definen al inicio de la salida de cada programa de regresión logística.*

En la segunda columna aparecen los coeficientes β del modelo de regresión múltiple para las seis variables categóricas independientes o covariables, y para las variables ficticias que se definen.

Se puede observar, que en todos los modelos se logró que los coeficientes logísticos en cada modelo sean positivos, lo cual garantiza que las razones de disparidad sean números mayores que uno, lo cual permite una más clara interpretación de las mismas. Esto se logra cambiando la categoría de referencia de la última a la primera en la sintaxis del correspondiente programa.

También en el trabajo aparecen en las **tablas 4, 6, 8 y 10** los resultados de las pruebas de bondad de ajuste y las de diagnóstico de cada modelo: el estadígrafo Sumario de los residuos de Pearson y el estadígrafo Sumario de las desviaciones residuales, así como los estadígrafos R^2 de Cox y Snell y de Nagelkerke. Además se incluye en dichas tablas la prueba de Hosmer-Lemeshow y la correspondiente tabla con los deciles de riesgo donde aparecen las frecuencias observadas y el valor esperado de las frecuencias en cada uno de los deciles para los migrantes y para los no migrantes. (Sólo se incluye la tabla 4 por la limitación establecida de seis cuartillas).

También se obtuvo en la salida del programa de cada ruta migratoria la prueba de Wald y la correlación parcial para cada coeficiente logístico.

En las **tablas 5, 7, 9 y 11**, sobre algunos aspectos de la salida del programa de aplicación de la regresión logística para las cuatro rutas migratorias, se observa que:

- Los intervalos de confianza no contienen a la unidad y por tanto es coherente con la significación estadística. También el tamaño en los intervalos tiene rangos aceptables, lo cual no invalida la significación estadística.
- En algunos casos existen diferencias entre las razones de disparidad crudas y las razones de disparidad obtenidas mediante el modelo de regresión logística, lo cual indica que hay efectos confusores que se eliminan introduciendo, para el análisis, las demás variables en el modelo de la regresión logística. También se puede observar que todas las variables son muy significativas ($< 0,001$) excepto dos cuyas significaciones son 0,02 y 0,05.
- También se pudo observar que los valores de los estadígrafos de Wald y de Correlación Parcial apuntan hacia una buena significación de todos los coeficientes logísticos.

A continuación, en el trabajo se presenta la aplicación del método de la regresión logística para las diferentes “rutas migratorias” R-R-R, R-R-U, R-U-R y R-U-U.

Tabla 4 de los deciles de riesgo y pruebas de bondad de ajuste y de diagnóstico del modelo para la ruta RRR.

Deciles de riesgo para la Prueba de bondad de ajuste de Hosmer y Lemeshow					
No Migrante			Migrante		
Frecuencias			Frecuencias		
Grupo	Observada	Esperada	Observada	Esperada	Total
1	332.000	332.580	15.000	14.42	347.000
2	446.000	448.857	35.000	32.143	481.000
3	478.000	480.866	42.000	39.134	520.000
4	623.000	619.196	62.000	65.804	685.000
5	87.000	85.194	8.000	9.806	95.000
6	893.000	899.989	124.000	117.011	1017.000
7	470.000	462.053	76.000	83.947	546.000
8	682.000	664.031	167.000	184.969	849.000
9	469.000	487.204	218.000	199.796	687.000
Ji-cuadrada gl Signif.					
Prueba de bondad de ajuste de Hosmer-Lemeshow:			7.0692	7	.4217
-2 Log Verosimilitud (-2LV _{modelo}) o Sumario de los cuadrados de las desviaciones residuales (Deviance): 4038.610					
Bondad de ajuste o Sumario de los cuadrados de los residuos de Pearson: 5283.825					
Ji-cuadrada gl Signif.					
Prueba de la razón de verosimilitud o del Modelo [-2LV _{modelo} -(-2LV _{con solo constante})]: 249.770 5 .0000					
Estadígrafo de Cox y Snell (R ²): .047					
Estadígrafo de Nagelkerke (^R ²): .047					

En este trabajo, debido a la limitación de paginas, de los 4 análisis para las cuatro rutas, sólo se presenta el correspondiente al:

Análisis de la construcción del modelo para la Ruta RRR.

Se analizó el modelo de regresión logística con todas las variables, recodificadas o no, **seleccionándose** mediante los métodos de paso a paso hacia atrás y hacia adelante, **el modelo con las variables** correspondientes a la **edad**, a la **situación conyugal** y a la **situación laboral**.

Las variables correspondientes al color de la piel, a la educación y al sexo fueron eliminadas por el programa en ese orden.

Luego de todos estos análisis, se decidió tomar el modelo que se presenta, al cual se le aplicaron la bondad de ajuste y el diagnóstico correspondiente, obteniéndose que **el modelo se ajusta razonablemente bien a los datos**.

Análisis de las razones de disparidad obtenidas del modelo de regresión logística para la Ruta RRR:

Al observar las razones de disparidad en la **tabla 5**, se puede decir que para un determinado estatus del individuo en las otras variables a las que no se hace referencia:

- Las personas de 30 a 49 años migran casi el duplo (1,61) de lo que lo hacen las personas de 50 años ó más.
- Las personas de 15 a 29 años migran más del triplo (3,63) de lo que lo hacen las personas de 50 años ó más.
- Las personas divorciadas, separadas, o viudas migran más del duplo (1 / 0,3852 =2,60) de lo que lo hacen las personas solteras.
- Las personas unidas o casadas migran más (1,58) que las personas divorciadas, separadas, o viudas.
- Las personas que no tienen vínculo laboral migran algo más (1,40) que las que tienen vínculo laboral (trabajan o están disponibles).

Análisis comparativo de las razones de disparidad cruda con las de la regresión logística para la Ruta RRR:

Al observar la **tabla 5** se puede decir que:

- La razón de disparidad de las personas sin vínculo laboral respecto a las personas con vínculo laboral disminuyó (de 1,52 a 1,40) lo que indica que el efecto de las otras variables disminuye la magnitud de su efecto.

- Por ejemplo se observa que el grupo de edades 30 a 49 años y el grupo de edades de 15 a 29 años aumentaron su razón de disparidad (1,48 a 1,61 y 2,65 a 3,63 respectivamente) así como también aumentó la razón de disparidad correspondiente al grupo de solteros ($1 / 0,61 = 1,64$ a $1 / 0,38 = 2,60$), queriendo esto decir que las personas sin vínculo laboral migraron en mayor medida por ser solteras y estar comprendidas en esos grupos de edades (49 años o menos), que por no tener vínculo laboral.
- El mismo razonamiento se sigue con la otra razón de disparidad cruda que disminuyó de 1,74 a 1,58, que es la correspondiente a la categoría de estar unido o casado respecto a estar divorciado o viudo o separado.
- Las personas unidas o casadas migraron también por estar comprendidas en los grupos de edades mencionados en el punto 1.

Tabla 5 : Algunos aspectos de la salida del programa de aplicación de la regresión logística para la ruta de migración Rural-Rural-Rural, donde el origen y destino de todos los desplazamientos fueron las propias zonas rurales

Variables en la ecuación	B _i	Sig	Exp(B _i) Razón de Disparidad	Intervalo de confianza		Razón de Disparidad Cruda
				Límite Inferior	Límite superior	
De 30-49 años Referencia: 50 años o más	,4754	,0012	1,6086	1,2070	2,1438	1,48
De 15 a 29 años Referencia: 50 años o más	1,2895	,0000	3,6309	2,7576	4,7807	2,65
Soltero Ref: divorciado-viudo-Separado	-,9541	,0000	,3852	,2666	,5564	0,61
Unido o casado Ref: divorciado-viudo-Separado	,4564	,0015	1,5784	1,1909	2,0919	1,74
Sin vínculo laboral Referencia: con vínculo laboral	,3346	,0000	1,3974	1,1903	1,6406	1,52

Fuente: Elaborada por la autora con la salida del programa de regresión logística del SPSS.

Conclusiones

La aplicación del método estadístico del Análisis de Regresión Logística que se presenta en este trabajo, muestra sin lugar a dudas, resultados muy interesantes que contribuyen a ampliar el conocimiento en el tema de las migraciones, ofreciendo una interpretación multivariada de fenómenos sociales que son muy complejos y tienen múltiples y variadas relaciones.

La aplicación del método estadístico de regresión logística se ajustó adecuadamente a las exigencias de la investigación, por cuanto brindó la posibilidad de ponderar el efecto neto de cada una de las variables, una vez que se elimina el efecto en su conjunto de las demás variables sobre la probabilidad de migrar, y relaciona las características de los migrantes, que proporcionan esas variables, con las características de los no migrantes.

- De acuerdo con los resultados obtenidos se puede decir que parece que para las rutas analizadas la migración es selectiva atendiendo a diferentes atributos sociodemográficos de los individuos que participan: sexo, edad, situación conyugal, nivel de instrucción y situación laboral. En el análisis, el color de la piel no resultó importante para ninguna de las rutas estudiadas.
- Se obtuvo que en algunos casos existen diferencias entre, las razones de disparidad crudas obtenidas de tablas cruzadas y las razones de disparidad obtenidas mediante el modelo de regresión logística, lo cual indica que hay efectos confusores, por lo cual las razones de disparidad obtenidas mediante la regresión logística son las válidas, ya que al analizar el efecto simultáneo de todas las variables el resultado es diferente al que se hace al considerar sólo dos variables a la vez, de lo cual se infiere que en esta investigación los resultados obtenidos no se podrían haber alcanzado sólo con métodos tradicionales de tablas cruzadas.
- El análisis de *la aplicación del modelo de regresión logística permitió no solo revelar a aquellas variables, según tipos de migración, que devienen significativas para determinar la probabilidad de migrar en las rutas analizadas, sino probar algunas hipótesis formuladas acerca de la importancia de determinados atributos sociodemográficos en el acto de la migración*, como por ejemplo, *la importancia de la familia en el acto de migrar*.

En este sentido probablemente lo más llamativo resulte el hecho de que la variable sexo no tenga influencia alguna para la ruta migratoria RRR entre las zonas rurales. La participación femenina se hace notoria en las rutas RUR, RRU y RUU entre las personas migrantes. Aquí se puede observar, que cuando el destino del migrante es al menos una zona urbana hay diferencia, en cuanto al sexo, con la ruta donde el destino es siempre una zona rural, ratificándose de

cierta manera la cuarta ley de Ravenstein⁶. De igual manera, en los movimientos migratorios, con excepción de la ruta RUU, el atributo edad privilegia a la cohorte de los más jóvenes.

En resumen, al determinar las variables sociodemográficas de mayor impacto para la probabilidad de migrar mediante la aplicación de la regresión logística, **en las diferentes rutas migratorias analizadas, con origen rural, se obtuvo como resultado lo siguiente:**

1. **Es posible caracterizar a las personas que se desplazan desde zonas rurales a zonas rurales también, como aquellas cuyas edades están entre 15 y 49 años, o sea, los jóvenes y los no tan jóvenes, que no tienen vínculo laboral, con una situación conyugal donde predominan los unidos o casados y en menor medida los divorciados, viudos o separados.**
2. **Es posible caracterizar a las personas que se desplazan desde zonas rurales a las urbanas y luego de nuevo a las rurales, como aquellas mujeres cuyas edades están entre 15 y 49 años, o sea, las más jóvenes y las no tan jóvenes, con una situación conyugal donde predominan los unidos o casados, y de escolaridad media o universitaria.**
3. **Es posible caracterizar a las personas que se desplazan desde zonas rurales a las rurales y luego a zonas urbanas, como las mujeres unidas o casadas, con edades entre 15 a 49 años, y que trabajan o están disponibles.**
4. **Es posible caracterizar a las personas que se desplazan desde zonas rurales a las urbanas y luego de nuevo a zonas urbanas, como las mujeres unidas o casadas, con edades de 30 años o más, y que tienen nivel medio básico o medio superior o nivel universitario.**

Bibliografía

- Agresti, A.,(1984), Analysis of Ordinal Categorical Data. Ed. Hohn Wiley & Sons,Inc.,Canada.
- Arguello, O., (1981), Migraciones internas: universo teórico y objetos de investigaciones, Notas de Población, CELADE, Santiago de Chile, Año IX, No. 25 .
- Bishop, Y. M., Fienberg, S. E., Holland, P. W., (1978), Discrete Multivariate Analysis, Theory and Practice, Ed. The Alpine Press, The Massachusetts Institute of Technology, EEUU.
- Bouquet, A., (1997), Estudio Descriptivo de las migraciones internas de Cuba en los años 1989-1995, s/f, s/p.
- De Jong, G. and Fawcett, J. (1981), Motivations for migration: an assessment and a value-expectancy research model, New York, Pergamon Press.
- Fernández Arias, M.,(1975), Las migraciones internas: algunos lineamientos teóricos desde la perspectiva del materialismo histórico, Facultad de Ciencias Sociales, Universidad de Costa Rica.
- Grupo de Trabajo Nacional de la Encuesta Nacional de Migraciones Internas, (1997), Las Migraciones internas en Cuba, Una exploración por niveles del Sistema de Asentamientos Poblacionales, Centro de estudios Demográficos, Universidad de la Habana.
- Hosmer, D.W.& Lemeshow, S., (1984), Applied Logistic Regression. Ed. Hohn Wiley & Sons, Inc., Canada.
- León, E., (1996), Encuesta Nacional de Migraciones Internas (ENMI). Elementos de su diseño, Grupo Ejecutivo Nacional.
- Lee Everett,(1975) Una Teoría de las migraciones. En Elizaga, Juan C. Y John Macisco, Migraciones Internas. Teoría. Método y Factores Sociológicos, CELADE, Stgo de Chile.
- López-Calleja, C. y Segura, T., (1999), Variables sociodemográficas determinantes para la construcción de un modelo de migraciones en las zonas rurales en Cuba, Ponencia presentada, en el IX Encuentro de Estadística Cuba-México en la Conferencia Internacional de Ciencia y Tecnología para el Desarrollo (CIMAF 99), C. de la Habana y Ponencia presentada en el Encuentro de Cientistas Sociales de las Ciudades de Córdoba, Buenos Aires y La Habana en febrero de 1999, Aceptado para publicación en 1999, Ciudad de la Habana.
- Morejón Seijas, B., (1994), Marcos Teóricos para la Interpretación de la Migración, CEDEM, Universidad de la Habana, Cuba.
- Mulder C. (1991): Dynamics in short and long distances migration: age –specific motives. Paper prepared for the International Seminar on Population Geography, Netherlands, April 7-10. Amsterdam. University of Amsterdam.
- Norusis, M.J., (1997), Manual del modulo SPSS Professional Statistics del paquete de programas del SPSS para Windows, Ed. SPSS Inc.
- Pregibon, D., Logistic Regression Diagnostic, (1981), Princeton University, The annals of Statistics, Vol. 9, N° 4, 705-724.
- Shawn Malia Kanaiauapuni., (1999), Reframing the Migration Question: An Empirical Analysis of Men, Women, and Gender in Mexico, CDE Working Paper #99-15, Selección de Internet.

• ⁶ Ravenstein, E.G., (1989) Historiador económico inglés, que al publicar en 1989 su trabajo sobre las leyes de la migración, dejó establecida una contribución teórica que aún hoy sustentan muchas de las investigaciones que se realizan en el campo socialista.