



Hachuel, Leticia
Boggio, Gabriela
Wojdyla, Daniel

Instituto de Investigaciones Teóricas y Aplicadas, de la Escuela de Estadística

MODELOS LOGIT MIXTOS: UNA APLICACIÓN EN EL ÁREA DE LA SALUD*

1. INTRODUCCIÓN

En la actualidad la información estadística proveniente de diversas áreas tales como las ciencias sociales, médicas y económicas, tiene por lo general una estructura jerárquica compleja o de niveles múltiples. Dichas estructuras a veces tienen un origen natural mientras que en otros casos puede ser la resultante de diseños de investigación. Más aún, puede ocurrir que los datos no estén obviamente anidados pero considerarlos así suele conducir a análisis más eficientes. Ejemplos de este tipo de estructura se reconocen en estudios poblacionales donde es posible identificar a individuos, que integran familias, a su vez reunidas en barrios, los cuales conforman ciudades. O bien, alumnos agrupados en cursos de diferentes escuelas.

Este agrupamiento, cualquiera sea su origen, provoca observaciones que tienden a estar correlacionadas dentro de los grupos. Es decir las respuestas de individuos pertenecientes a un mismo grupo tienden a ser más similares que las de diferentes grupos. Ignorar esta relación haciendo uso de las técnicas estadísticas tradicionales puede sobredimensionar los efectos del grupo y por lo tanto distorsionar las conclusiones. En esencia lo que ocurre es que los individuos agrupados proveen menos información que la que hubiesen proporcionado si no tuvieran esa característica en común: pertenecer a algún grupo determinado. Por lo tanto para estudiar si una covariable influye sobre las respuestas de individuos agrupados según alguna característica en común, el empleo de modelos de regresión clásicos, ignorando el agrupamiento, podría conducir a resultados erróneos.

La modelización en niveles múltiples es una técnica estadística apropiada para el análisis de este tipo de datos. En términos generales, este enfoque incluye en el modelo efectos aleatorios relacionados con los diferentes niveles de información (Goldstein, 2003).

Si bien el uso de efectos aleatorios en modelos lineales para variables respuestas con distribución normal está bien estudiado, su consideración para el caso de variables respuestas categóricas es de desarrollo reciente. Los modelos que abarcan variables respuesta no normales y efectos fijos y aleatorios en el predictor, constituyen la clase denominada modelos lineales generalizados mixtos –MLGM- (Agresti, 2002; Agresti et al, 2000).

* Este trabajo forma parte de un proyecto de intercambio entre la Secretaría de Salud Pública de la Municipalidad de Rosario, el Instituto de la Salud Juan Lazarte y la Escuela de Estadística de la Facultad de Ciencias Económicas y Estadística de la UNR.



Este trabajo tiene por objeto estudiar y poner a prueba este tipo de modelos en el área de la salud en el marco del proyecto "Modelos de niveles múltiples para respuesta categórica"^{*}. La posibilidad de plantear este tipo de modelos constituye una manera de integrar la dimensión poblacional o de sociedad junto con la dimensión individual o micro en la investigación en el campo de la epidemiología y la salud pública (Diez-Roux, 2004).

En particular, se ensaya el ajuste de modelos lineales generalizados mixtos en el análisis del problema del bajo peso del recién nacido tomando como punto de referencia los partos acontecidos en una maternidad pública de la ciudad de Rosario.

2. MODELOS DE NIVELES MÚLTIPLES

Los modelos de regresión logística son los que tradicionalmente se utilizan para el estudio de respuestas binarias. El término binario se refiere a los dos posibles resultados de la variable respuesta, ya sea por su naturaleza categórica o bien por haber sido dicotomizada una variable continua de interés. La regresión logística permite expresar linealmente la relación entre las variables explicativas y el logaritmo del odds de respuesta positiva.

En la explicación de ciertos fenómenos es pertinente considerar factores relacionados a diferentes niveles de una estructura jerárquica. Para el caso más simple de dos niveles, implicaría manejar dos tipos de covariables: las "entre grupo", cuyo valor puede variar de grupo a grupo pero es el mismo para todas las unidades de un mismo grupo, y las "intra-grupo", cuyos valores pueden variar entre las unidades del mismo grupo.

El enfoque de análisis más simple e "ingenuo" de este tipo de información es ignorar dicha estructura jerárquica, es decir ignorar la correlación existente entre las respuestas de individuos pertenecientes a un mismo grupo, y aplicar un modelo de regresión logística convencional. El costo de esta simplicidad es que una correlación positiva entre las observaciones provoca una sobredispersión que exagera la significación de los tests estadísticos (Hosmer & Lemeshow; 2000).

A diferencia de este primer enfoque, los modelos estadísticos de niveles múltiples consideran la influencia de los diferentes niveles de la jerarquía sobre la respuesta a través de la incorporación en el modelo de efectos aleatorios asociados a características medidas en esos distintos niveles. De esta manera se obtiene una estructura de asociación entre las respuestas de los individuos que refleja la tendencia a respuestas más similares entre los pertenecientes a un mismo agrupamiento. Se trata de un modelo que tiene en cuenta los datos correlacionados a través de la consideración de los niveles de un factor como seleccionados de una población de niveles con el fin de realizar inferencias sobre esa población.

Los modelos con efectos aleatorios han sido bien estudiados para el caso de respuestas normales. En contraste, sólo recientemente, han sido aplicados para respuestas categóricas. En particular para datos binarios el modelo básico, denominado "modelo logístico-normal", tiene un enlace logit con un predictor lineal que contiene efectos aleatorios distribuidos normalmente con variancias desconocidas. El mismo constituye un ejemplo de la extensión de los modelos lineales generalizados denominada "modelos lineales generalizados mixtos".

En términos formales los modelos de regresión logística mixtos o también denominados de tipo grupo- específico (GE) se explicitan en su forma general (Aerts et al, 2002):

^{*} Proyecto de Investigación de la Secretaría de Ciencia y Tecnología de la UNR – PID ECO29 - 2004.

$$\text{logit}\{P(Y_{ij} = 1 / X_{ij}, \alpha_i, \beta_{GE})\} = \alpha_i + X'_{ij}\beta_{GE} \quad \text{donde:} \quad (1)$$

Y_{ij} representa la respuesta aleatoria binaria en la j -ésima unidad $j=1, \dots, n_i$, del i -ésimo grupo, $i=1, \dots, m$;

$\text{logit}\{P(Y_{ij} = 1)\} = \log \frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)}$, siendo este último cociente el odds de respuesta positiva;

α_i es el término intercepto aleatorio para el i -ésimo grupo;

X_{ij} es un vector de covariables de interés para el j -ésimo individuo en el i -ésimo grupo y

β_{GE} es el correspondiente vector de parámetros.

El vector X_{ij} puede contener tanto componentes a nivel de grupo, $X_{(f)ij}$ como también componentes que varían dentro del grupo $X_{(v)ij}$. En correspondencia con estas componentes, el vector de parámetros β_{GE} se clasifica en $\beta_{GE(f)}$ y $\beta_{GE(v)}$. Con esta notación el modelo (1) se escribe:

$$\text{logit}\{P(Y_{ij} = 1 / X_{ij}, \alpha_i, \beta_{GE})\} = \alpha_i + X'_{(f)ij} \beta_{GE}^{(f)} + X'_{(v)ij} \beta_{GE}^{(v)}. \quad (2)$$

La verosimilitud bajo el modelo de regresión logística de efectos mixtos se obtiene integrando la ecuación (2) bajo el supuesto que α_i se distribuye según G . El modelo "logístico-normal" supone que G es la distribución Normal, es decir $\alpha_i \rightarrow N(0, \sigma^2)$ (Stiratelli et al., 1984).

El modelo es por lo tanto, un modelo logístico mixto con efectos fijos para las covariables, tanto las de nivel grupal como individual, más un efecto aleatorio asociado a grupo.

La inclusión del término α_i en la ecuación (2) provoca el aumento de la correlación entre respuestas dentro de un grupo en relación a la correlación entre grupos. La idea básica es que las probabilidades logísticas para observaciones de la respuesta en un grupo tengan un valor común α_i . Por lo tanto, si los α_i son muy diferentes, las respuestas de un mismo grupo están más altamente correlacionadas que el conjunto total de respuestas. Es decir, cuanto mayor es la diferencia entre los α_i , mayor es la correlación intra que entre grupos. La heterogeneidad en los α_i es simplemente una función de su variancia, por lo que la correlación intra-grupo aumenta cuando aumenta σ^2 .

Para la interpretación de los coeficientes de las covariables es necesario tener en cuenta que el modelo es condicional sobre el efecto aleatorio para grupo, y dentro de ese grupo, el coeficiente de una covariable representa la magnitud del cambio en el logaritmo del odds de respuesta positiva que uno debería esperar ante un valor particular de la covariable versus otro valor de la misma. Debido a que el modelo especifica que ese coeficiente es el mismo para todos los grupos, se estima combinando la información de diferentes grupos, es decir promediando sobre todos los grupos de acuerdo a la distribución de ese efecto aleatorio para grupo.

Si la covariable en cuestión es del tipo intra-grupo, X_v , el coeficiente representa el logaritmo de la razón de odds común para la respuesta entre los que asumen un determinado valor versus otro valor diferente de la covariable a través de los grupos.

Si se supone que no hay cambio en el valor de la covariable entre los individuos de un grupo, es decir la covariable es del tipo X_f , el modelo GE persiste en medir el efecto de cambio en el status de la covariable en un grupo dado. Como este cambio no fue directamente medido, se dice que su estimación está enteramente basada en el modelo. En consecuencia, en este enfoque GE resulta difícil la interpretación de covariables que no varían



dentro del grupo y por lo tanto, el modelo grupo-específico (GE) es más útil cuando el objetivo es proveer inferencias acerca de los efectos de covariables medidas a nivel individual.

Otro enfoque usual para el tratamiento de medidas binarias correlacionadas es el ajuste de los denominados modelos promedio-poblacionales o marginales. Este tipo de modelos es más útil cuando se intenta inferir acerca de efectos de covariables a nivel grupal. En el modelo marginal el coeficiente asociado a una covariable representa la diferencia en el logaritmo de la razón de odds para la respuesta entre dos conjuntos de individuos caracterizados por valores diferentes de dicha covariable. Es decir, los parámetros de estos modelos tienen una interpretación tipo "promedio-poblacional" en el sentido que el efecto de las covariables se promedia a través de los grupos. Si bien este enfoque tiene en cuenta la asociación entre las respuestas de los individuos de un mismo grupo, consecuencia de la estructura jerárquica de los datos, no considera explícitamente la heterogeneidad de las respuestas entre los grupos.

3. APLICACIÓN

3.1 Los datos

Para la puesta a prueba de los modelos GE descriptos en la sección anterior se utiliza información provista por la Secretaría de Salud Pública de la Municipalidad de Rosario registrada de acuerdo al Sistema Informático Perinatal (SIP), el cual es una base de datos que contiene variables referidas a la madre y al hijo en aspectos básicamente biológicos. Uno de los indicadores perinatales obtenidos a partir del SIC es el porcentaje de niños con bajo peso al nacer en las maternidades municipales. Entre ellas se encuentra la Maternidad Martín, para la que se sabe que el porcentaje de niños con bajo peso al nacer históricamente ha sido del 7%. Resulta interesante estudiar las variaciones en este porcentaje de acuerdo a características propias de la madre y del embarazo.

Para ello se ajustan modelos tipo logit a los datos referidos a nacimientos ocurridos en la Maternidad Martín en el año 2003 dicotomizando la respuesta continua peso del recién nacido de la siguiente manera:

Variable Respuesta: Bajo peso al nacer: sí (entre 500 y 2500 g.) , no (>2500 g.).

Se excluyen abortos (peso menor que 500 g. o edad gestacional menor que 20 semanas), muertes intrauterinas, malformaciones y embarazos múltiples.

En base a los resultados hallados en la caracterización socioepidemiológica de madres atendidas en maternidades municipales en el año 1999 (Informe estadístico de la gestión de Salud Pública en la ciudad de Rosario - Secretaría de Salud Pública, 2001) se seleccionaron variables para intentar describir algunos aspectos considerados importantes en el estudio del bajo peso al nacer. Ellas son:

Variables explicativas:

- Convivencia: con pareja (casada o unión consensual), sin pareja (soltera, separada o viuda).
- Nivel educacional máximo alcanzado: ninguno, primaria, secundaria, universitario
- Total de controles durante el embarazo: varía de 0 a 9.
- Edad: medida en años



- Condición de primípara: no, si.
- Forma de Terminación del Parto: espontáneo, no espontáneo (fórceps, cesárea, otro).

Se dispone también de información sobre variables indicadoras de áreas geográficas que permiten agrupar a las mujeres de acuerdo a su lugar de residencia. Ellas son:

- Distrito municipal: centro, norte, noroeste, oeste, sudoeste, sur.
- Seccional Policial: incluye las 22 seccionales y 6 subcomisarías del municipio.

3.2 Ajuste de modelos

En las indagaciones epidemiológicas dirigidas a estudiar un determinado problema de salud-enfermedad a nivel poblacional es frecuente reconocer una estructura jerárquica en la información, ya que los individuos pertenecen a grupos familiares, con determinadas relaciones entre sus miembros, a su vez estas familias están integradas a un grupo social, con particulares condiciones de vida y de salud, de acceso y de utilización de los servicios de atención y dichos grupos comparten un espacio social en un territorio determinado.

Si bien la posición socioeconómica continúa siendo el más consistente predictor de bienestar, es probable que los procesos políticos, culturales e institucionales a nivel barrial también contribuyan en forma importante a los resultados en salud (O'Campo, 2003). Además la supuesta heterogeneidad entre barrios responde a que los mismos no constituyen una mera delimitación geográfica sino que son producto de dinámicas sociales particulares. El supuesto teórico es que dentro de cada grupo –barrio-, las respuestas tenderán a estar correlacionadas, por tratarse de expresión de condiciones, atributos de individuos, familias, próximos entre sí, que comparten condiciones de vida y salud semejantes (Luppi et al, 2004).

Los datos disponibles permiten el reconocimiento de una estructura jerárquica que responde a un agrupamiento de las mujeres según el lugar de residencia, ya sea distrito municipal o seccional policial. Si bien no se dispone del nivel de desagregación deseado, en barrios, se decide utilizar estos indicadores más agregados en un análisis multinivel. Más aún, se elabora ad-hoc un agrupamiento en áreas que no sólo tiene en cuenta la clasificación del municipio en seccionales policiales sino que también utiliza información complementaria indicadora de la salud de la población. Para su construcción se utiliza la ubicación geográfica de las seccionales policiales y el nivel de prevalencia de talla baja o muy baja en niños en edad escolar según consta en un trabajo sobre Situación Nutricional de Escolares realizado en el ámbito de la Secretaría de Salud Pública de la Municipalidad de Rosario (Aronna et al, 2003).

El enfoque adecuado para tener en cuenta la estructura jerárquica de la información recién descripta es el correspondiente a los modelos grupo- específico o modelos mixtos formalizados según la ecuación (2). Sin embargo las variables disponibles reducen esta formulación a la consideración de variables sólo del tipo denominado $X_{(V)}$, es decir las que varían de individuo a individuo dentro de cada grupo. El nivel de agregación mayor de la jerarquía traducido en la pertenencia a, por ejemplo un distrito municipal, queda reducido a un coeficiente aleatorio α_i , ya que no se dispone de características propias de ese nivel.

Previamente se ajusta el modelo logístico convencional mencionado como el correspondiente al enfoque ingenuo en la sección anterior a fin de evaluar las posibles distorsiones en la validez de las inferencias.

Para el ajuste de los modelos se utilizan los procedimientos LOGISTIC y NLMIXED de



SAS (SAS Institute, 1999).

Enfoque ingenuo

Bajo el supuesto de independencia de las observaciones, se ajustaron modelos de regresión logística convencional considerando posibles diferencias en la probabilidad de bajo peso del recién nacido según la edad de la madre, el nivel educacional, el tipo de convivencia, el número de controles prenatales, la condición de primípara y la forma de terminación del parto. En la Tabla 1 se presentan los resultados hallados después de excluir del modelo aquellas variables que resultaron no significativas ($p < 0.10$). Se elige incluir en el modelo la variable edad con una componente lineal y otra cuadrática y la interacción entre la edad y la condición de primípara.

Tabla 1: Estimaciones de los parámetros del modelo logit convencional

Parámetro	Estimación	Error Estándar	P- <i>asoc.</i>
Intercepto	-0.3227	1.6016	0.8403
Edad	-0.1019	0.1170	0.3839
Edad ²	0.0018	0.0020	0.3781
Sin pareja	-0.3371	0.1563	0.0310
Primípara	-4.8076	3.2467	0.1387
Controles prenatales	-0.3057	0.0316	<0.0001
Parto no espontáneo	0.9938	0.1606	<0.0001
Edad*Primípara	0.5104	0.2966	0.0853
Edad ² *Primípara	-0.0122	0.0068	0.0716

El test de bondad de ajuste de Hosmer Lemeshow para el modelo planteado da como resultado un valor de 8.6891 con 8 grados de libertad y probabilidad asociada igual a 0.3692, por lo que proporciona un ajuste adecuado a los datos.

De acuerdo a este modelo resultan altamente significativos los coeficientes asociados al tipo de convivencia (con o sin pareja), al número de controles prenatales y forma de terminación del parto (espontáneo o no). La probabilidad asociada a la interacción entre edad y condición de primípara se encuentra entre 0.05 y 0.10, lo que hace dudosa su importancia. Cabría preguntarse cuánto de esta significación estadística se alcanza por el hecho de ignorar la correlación entre las respuestas de individuos de un mismo distrito. Para responder a este interrogante se reanaliza la información bajo un enfoque grupo-específico a fin de controlar la posible asociación entre las respuestas de las mujeres residentes en un mismo distrito.

Enfoque grupo-específico

El objetivo principal del ajuste de modelos logísticos mixtos es realizar inferencias acerca de los efectos fijos. La inclusión de efectos aleatorios en el modelo es un mecanismo para caracterizar cómo la correlación positiva se presenta entre las observaciones dentro de un grupo. De esta manera el error estándar del intercepto aleatorio de un modelo constituye un resumen útil del grado de heterogeneidad de la población en estudio (Agresti, 2000).



Se ajusta entonces, un modelo de regresión logística mixto con efectos fijos para las mismas covariables recién consideradas y un efecto aleatorio asociado a cada distrito. Los resultados se muestran en la Tabla 2.

Tabla 2: Estimaciones de los parámetros del modelo logit con efecto aleatorio para cada distrito

Parámetro	Estimación	Error Estándar	P-asoc.
Intercepto	-0.1927	0.8048	0.8187
Edad	-0.1015	0.1175	0.4210
Edad ²	0.0018	0.0021	0.4155
Sin pareja	-0.3418	0.1568	0.0720
Primípara	-4.8640	3.2428	0.1843
Controles prenatales	-0.3067	0.0318	<0.0001
Parto no espontáneo	0.9915	0.1609	0.0008
Edad*Primípara	0.5133	0.2960	0.1336
Edad ² *Primípara	-0.0122	0.0067	0.1210
Variancia del efecto aleatorio	0.0281	0.0506	0.5993

El valor estimado de la variancia del efecto distrito, asumido aleatorio en el modelo, resulta igual a 0.0281 con alta probabilidad asociada ($p=0.5993$), por lo cual el grado de heterogeneidad no es importante ni resulta significativo.

Bajo este modelo pierden significación los coeficientes asociados a la interacción entre edad y condición de primípara y en menor medida el asociado a la variable tipo de convivencia. Estos resultados acuerdan con la posible reducción en la significación estadística de los efectos cuando se tiene en cuenta el agrupamiento de las respuestas según distrito, aún cuando la heterogeneidad entre ellos sea leve. Además el número reducido de grupos (6 distritos) puede justificar el hecho que las estimaciones de los coeficientes del modelo mixto no sean más grandes en valor absoluto que las del modelo convencional.

Los resultados hallados orientan al ajuste de un nuevo modelo más simple que sólo contiene los efectos principales de tipo de convivencia, condición de primípara, número de controles prenatales y forma de terminación del parto (Tabla 3).

La estimación del error estándar del efecto aleatorio conduce a un valor de 0.2147 ($\hat{\sigma} = \sqrt{0.0461}$), el cual mide el grado de heterogeneidad entre los distritos en el logit de la probabilidad de bajo peso que no es atribuible a las covariables consideradas. Cabe señalar que el valor de esta estimación es levemente superior al encontrado en el modelo más parametrizado ($\hat{\sigma} = \sqrt{0.0281} = 0.1675$) precisamente debido a la exclusión de variables en el modelo. Sin embargo la variancia del efecto aleatorio sigue siendo estadísticamente no significativa.

Tabla 3: Estimaciones de los parámetros del modelo logit reducido con efecto aleatorio para cada distrito

Parámetro	Estimación	Error Estándar	P-asoc.
Intercepto	-0.8649	0.1059	0.0002
Sin pareja	-0.3127	0.1533	0.0875
Primípara	0.4149	0.1630	0.0437
Controles prenatales	-0.3169	0.0312	<0.0001
Parto no espontáneo	1.0070	0.1575	0.0007
Variancia del efecto aleatorio	0.0461	0.0625	0.4885

En este modelo reducido resultan significativas al 5% la condición de primípara, el número de consultas prenatales y la forma de terminación del parto. El aumento de la probabilidad asociada a la covariable tipo de convivencia hace que se desestime su influencia en el riesgo de bajo peso, por lo que se decide no interpretar su coeficiente.

Los coeficientes de las variables que mostraron un efecto significativo sobre la probabilidad de que el recién nacido sea de bajo peso se interpretan en términos de razones de odds condicionales. Éstas constituyen una medida aproximada del riesgo de que las mujeres residentes en un determinado distrito tengan un niño de bajo peso al nacer según sea el valor asumido por cada covariable en particular manteniendo constante el valor de las restantes.

Por lo tanto, para un mismo distrito la chance estimada de bajo peso en el recién nacido:

- es el 50% ($\exp(0.4149)=1.51$) mayor para las madres primíparas que para las que no lo son;
- es casi tres ($\exp(1.007)=2.74$) veces mayor cuando la terminación del parto no fue espontánea, en el sentido que requirió de cesárea, fórceps o alguna otra intervención;
- disminuye en la medida que la embarazada realice un número mayor de controles prenatales. Esta disminución es de un 30% ($\exp(-0.3169)=0.73$) por aumento unitario en el número de controles del embarazo.

Sin embargo, el intervalo de confianza correspondiente a la razón de odds entre bajo peso y condición de primípara ($IC_{95\%}$: 1.10; 2.08) muestra que dicha chance puede llegar a ser sólo un 10% mayor para las primíparas. En cambio con respecto a la forma de terminación del parto, el intervalo de confianza ($IC_{95\%}$: 2.01; 3.73) indica que el aumento en la chance de bajo peso al nacer es por lo menos del doble cuando el parto no termina en forma espontánea. Por último, en relación al número de consultas, el intervalo de confianza ($IC_{95\%}$: 0.68; 0.77) resulta lo suficientemente estrecho como para reafirmar el alcance de la estimación puntual.

La misma secuencia de análisis recién descrita se realiza teniendo en cuenta a seccional policial y área, en lugar de distrito, como formas aproximativas a la conceptualización de barrio, con resultados similares a los recién presentados. Resulta importante señalar que, y los resultados así lo corroboran, estas aproximaciones distan mucho de ser satisfactorias para representar a los barrios como grupos geográficos homogéneos en cuanto a la dinámi-



ca social.

4. DISCUSIÓN

Un concepto clave en epidemiología es que todos los determinantes y condicionantes de las enfermedades no pueden ser conceptualizados sólo como atributos a nivel individual, por lo que resulta necesario, cuando se quieren estudiar problemas de salud-enfermedad, considerar aspectos de los grupos a los cuales pertenecen los individuos.

El reconocimiento de esta diferenciación en niveles jerárquicos de los condicionantes del fenómeno en estudio provoca desafíos metodológicos. Los modelos de niveles múltiples constituyen un enfoque apropiado para el análisis de este tipo de datos.

La puesta a prueba de esta clase de modelos en datos del campo de la salud ilustra cómo concebir las condiciones sociales y medio ambientales como una dimensión explicativa adicional en el estudio de un problema de salud a nivel individual. En particular para el estudio del bajo peso al nacer, la aplicación de los modelos de niveles múltiples, más concretamente un modelo logístico-normal, proviene de pensar al lugar de residencia de la embarazada como una posible fuente adicional de variabilidad. La inclusión de efectos aleatorios asociados a agrupamientos según el lugar de residencia condujo a resultados más confiables que los obtenidos con el enfoque convencional. La interpretación de los coeficientes de efectos fijos se puede realizar, como en toda regresión logística, en términos de razones de odds pero en estos modelos están referidas a cada grupo –distrito- en particular.

Los modelos mixtos utilizados tienen en cuenta la correlación producida por el agrupamiento de los individuos a través de la inclusión de efectos aleatorios en cada etapa de la jerarquía. Las variancias de estos efectos se estiman como parte del proceso de ajuste del modelo y miden la cantidad de variabilidad no explicada por los efectos fijos. En este enfoque grupo-específico, las estimaciones y la inferencia son condicionales sobre los valores de las covariables y de los efectos aleatorios. Si bien estos efectos aleatorios no son directamente observables se los piensa como que representan propensiones producto de las condiciones socio-ambientales compartidas. Por lo tanto, las estimaciones de este modelo cobran mayor relevancia en un contexto explicativo.

5. REFERENCIAS

- AERTS, M.; GEYS, H.; MOLENBERGHS, G.; RYAN, L. 2002. *Topics in Modelling of Clustered Data*. Chapman & Hall/CRC.
- ARONNA, A.; BALPARDA, L.; BISIO, S.; GÓMEZ, M.; LIBORIO, M.; MOYANO, C. SALARANO, D., 2003. Situación Nutricional de Escolares. Publicación de la Municipalidad de Rosario, Argentina.
- AGRESTI, A., 2002. *Categorical Data Analysis*, 2nd ed. John Wiley & Sons.
- AGRESTI, A.; BOOTH, J.; HOBERT, J.; CAFFO, B. 2000. Random-effects modeling of categorical response data. *Sociological Methodology*, 30: 27-81.
- CROUCHLEY, R., 1995. A random effects model for ordered categorical data. *Journal of American Statistical Association*, 90: 489-498.
- DIEZ-ROUX, A., 2004. The study of group-level factors in epidemiology: rethinking variables, study designs and analytical approaches. *Epidemiology reviews*, 26:104-111.



- DIGGLE, P. J.; HEAGERTY, P.; LIAGN K-Y.; ZEGER S. L. 2002. *Analysis of Longitudinal Data*, 2nd ed. Oxford University Press.
- GOLDSTEIN, H., 2003. *Multilevel Statistical Models*. 3rd. edition. Kendall's Library of Statistics. London.
- HOSMER, D. W.; LEMESHOW, S., 2000. *Applied Logistic Regression*, 2nd ed. John Wiley & Sons. New York.
- LUPPI, I.; BOGGIO, G., 2004. Modelos de niveles múltiples como estrategia analítica epidemiológica en estudios evaluativos. Actas 2° Jornadas Nacionales de Epidemiología y 3° Jornadas de Economía y Gestión de Salud. Rosario.
- O'CAMPO, P., 2003. Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health. *American Journal of Epidemiology*, 157:9-13.
- PREISSER, J.; ARCURY, T.; QUANDT, S. 2003. Detecting patterns of occupational illness clustering with alternating logistic regressions applied to longitudinal data. *American Journal of Epidemiology*, 158: 495-501.
- SAS Institute, Inc. 1999. SAS/STAT user's guide, version 8. Cary, NC.
- SECRETARÍA DE SALUD PÚBLICA. Municipalidad de Rosario, 2001. Informe estadístico de la gestión de Salud Pública en la ciudad de Rosario. *Boletín de Bioestadística*. Año 4 N° 1.
- STIRATELLI, R.; LAIRD, N.; WARE, J., 1984. Random effects models for serial observations with binary response. *Biometrics*, 40: 1025-1035.