Aplicación de modelos multinivel para variables binarias en estudios sobre logros académicos en escolares

Ramón Alvarez [1] **Sebastián Gadea** [1]
XI Congreso Latinoamericano de Sociedades de Estadística CLATSE 2014
[1] Unidad de Biometría, IESTA
Octubre, 20-23, 2014, La Serena, Chile

Resumen

En un estudio sobre las dificultades de aprendizaje llevado adelante por equipos de maestros, psiquiatras y psicomotricistas en escolares de contexto socio-económico bajo, se analizan los logros académicos. Cuando los individuos forman grupos o clusters, podríamos esperar que dos seleccionados de un mismo grupo tenderán a ser más parecidos que dos individuos seleccionados de entre los diferentes grupos. Por ejemplo, los niños aprenden en las clases, las condiciones de su grupo, tales como características de los maestros y la capacidad de otros niños en la clase, puede influir en el logro educativo de un niño. Por lo tanto, para evaluar tales dependencias se recurre a los modelos multinivel - también conocidos como modelos jerárquicos lineales, modelos mixtos, modelos de efectos aleatorios y los modelos de componentes de la varianza - para analizar los datos con una estructura jerárquica.
Para eso se toma en cuenta las variables contextuales relativas a escuela, grupo en la escuela y maestra, en 372 niños del departamento de Canelones de 1er grado, que forman parte de 7 escuelas públicas y 22 grupos. Se evalúa una escala de logro académico (ELA) que está conformada por 6 subescalas para medir logros en lectura de frases y palabras, adquisición de código escrito y dominio de repertorio numérico. El constructo ELA se dicotomiza tomando como categorías si logra la totalidad de las subescalas o no y sobre éste se aplica análisis multinivel .

Presentación del Problema

Con el objetivo de conocer el constructo ELA -rendimiento óptimo- teniendo en cuenta la estructura de la población, se realiza un estudio del mismo a partir de un modelo multinivel. Se presenta una nueva metodología de estimación del modelo, se expone la aplicación concreta, y se comenta brevemente una metodología alternativa.

Modelo multinivel

Considere la posibilidad de una estructura de dos niveles, donde un total de $n = 372$ niños (en el nivel 1) se anidan dentro de $J = 22$ grupos (en el nivel 2) con n_j individuos en el grupo j . Se denota por y_{ij} la respuesta para el individuo i en el grupo j.

$$y_{ij} = \pi_{ij} + \epsilon_{ij} \tag{1}$$

, donde $\pi_{ij} = E(y_{ij}|x_{ij}, u_j) = P(y_{ij} = 1)$. El modelo de intercepto aleatorios lineal generalizado para la dependencia de la respuesta probabilidad π_{ij} en x_{ij} se escribe [4]:

$$F^{-1}(\pi_{ij}) = \beta_0 + \beta_1 x_{ij} + u_j \tag{2}$$

donde F^{-1} es la función de enlace logit (monótona y diferenciable), $u_j \sim N(0, \sigma_u^2)$ efecto de los grupos, y x_{ij} variable explicativa del nivel individual.
Vemos como a pesar de que el lado izquierdo es una transformación no-lineal de π_{ij}, el lado derecho toma la misma forma que una ecuación para una variable continua . Los parámetros del modelo que deseamos estimar son β_0, β_1, y σ_u^2.

Modelo logit con intercepto aleatorio

y_{ij} se distribuye Bernoulli con media π_{ij} , la contribución del individuo i en el grupo j a la probabilidad es la función de densidad de probabilidad :

$$f(y_{ij}) = \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{(1-y_{ij})} \tag{3}$$

En un modelo logit $F^{-1}(\pi_{ij})$ es el log-odds de probabilidades de que y = 1. Si π_{ij} es la probabilidad de tener un óptimo logro académico, se tiene que:

$$\pi_{ij} = \frac{\exp^{\beta_0+\beta_1 x_{ij}+u_{ij}}}{1 + \exp^{\beta_0+\beta_1 x_{ij}+u_j}} \tag{4}$$

Luego, si aplicamos $\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$ tenemos que

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + u_j \tag{5}$$

Observemos que no existe relación lineal entre π_{ij} y los parámetros, sino que la relación lineal se da entre F^{-1} y los β. A éste modelo se denomina regresión logística.

Estimación del modelo

Un método general para ajustar modelos de regresión es la estimación por máxima verosimilitud. La función de verosimilitud es el producto de todas las contribuciones individuales (a través de los n_j individuos en el grupo j y J grupos):

$$L\left(\beta_0, \beta_1, \sigma_u^2|u\right) = \prod_{j=1}^{J}\prod_{i=1}^{n_j} \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}} \tag{6}$$

La función de probabilidad dada por (6) se denota $L\left(\beta_0, \beta_1, \sigma_u^2|u\right)$ ya que su valor está condicionada al valor del azar del efecto u_j . Por esta razón, se conoce como la probabilidad condicional. Debido a que u_j es inobservable, trabajamos con la probabilidad marginal que no depende de u_j. La probabilidad marginal se obtiene promediando sobre los efectos aleatorios que, equivale a la integración sobre la distribución de los efectos aleatorios:

$$L\left(\beta_0, \beta_1, \sigma_u^2\right) = \int L\left(\beta_0, \beta_1, \sigma_u^2|u\right)\phi(u)du \tag{7}$$

donde $\phi(u)$ es la función de densidad de probabilidad de una distribución normal.
Los métodos más utilizados de la integración numérica son Gauss-Hermite cuadratura numérica y cuadratura adaptativa. Puede ser necesario un gran número de puntos de cuadratura para aproximar la distribución normal, que puede conducir a largos tiempos de estimación para grandes conjuntos de datos.

Cuasi-verosimilitud

Una alternativa para el método de máxima verosimilitud es el uso de una aproximación que reemplace (6) por un modelo lineal.
Procedimientos de cuasi-verosimilitud implican la aproximación π_{ij} en (6) por una expresión lineal, utilizando un desarrollo en serie de Taylor de $\pi_{ij} = F(\beta_0 + \beta_1 x_{ij} + u_j)$. Goldstein (1991)[1] propuso el uso del algoritmo iterativo de mínimos cuadrados reponderados (IRLS). Este método es una modificación del algoritmo IGLS (para ver descripción completa del método IGLS: Rodriguez y Goldman 1995)[2].

Algoritmo IRLS:

paso 0: Se obtienen valores de partida $\beta_0^{(0)}$, $\beta_1^{(0)}$ estimaciones de MCO y $\sigma_u^2 = 0$ (estimaciones de un modelo de un solo nivel).
paso 1: Obtenemos $\beta_0^{(m)}$, $\beta_1^{(m)}$ (estimaciones de β_0, β_1 obtenidas en la iteración m) a partir de $\beta_0^{(m-1)}$, $\beta_1^{(m-1)}$ y $\sigma_u^{2(m-1)}$
paso 2: Obtenemos $\sigma_u^{2(m)}$ a partir de $\beta_0^{(m)}$, $\beta_1^{(m)}$ y $\sigma_u^{2(m-1)}$.
paso 3: Calculamos $\pi^{(m)}$ probabilidad de respuesta evaluada en $\beta_0^{(m)}$, $\beta_1^{(m)}$ y $w^{(m)} = \pi^{(m)}(1 - \pi^{(m)})$.
Los pasos 1, 2 y 3 se repiten, en cada iteración m, hasta que una iteración adicional del procedimiento conduce a un cambio relativo "pequeño" en las estimaciones de los parámetros, por lo que se consigue el punto de convergencia. El criterio de convergencia se conoce comúnmente como *tolerancia*.
La aproximación de primer orden para π_{ij} (omitiendo los subíndices ij) es:

$$\pi^{(m+1)} \approx \pi^{(m)} + w^{(m)}[(\beta_0 - \beta_0^{(m)}) + (\beta_1 - \beta_1^{(m)})x] + w^{(m)}u \tag{8}$$

Después de sustituir (8) en $y_{ij} = \pi_{ij} + \epsilon_{ij}$ y reordenando, obtenemos un modelo lineal:

$$\tilde{y}_{ij} = \tilde{\beta}_0 + \tilde{\beta}_1\tilde{x}_{ij} + \tilde{u}_j + \epsilon_{ij}$$

, donde los $\tilde{y}_{ij}$ y $\tilde{x}_{ij}$ son versiones de la variable dependiente y la variable explicativa (en función de las variables originales Y, x y los valores de los $\beta_0^{(m)}$,$\beta_1^{(m)}$ y $\pi^{(m)}$) y $\tilde{u}_j = w^{(m)}u_j$.

Intrepretación de los parámetros

1. β_0 intercepto global ; $\beta_0 + u_j$ intercepto para el grupo j.
2. σ_u^2 varianza entre grupos ajustada para x, o simplemente varianza residual del nivel 2.
3. El Odds-ratio $\exp^{\beta}$ = OR = $\frac{Odds|_{x+1}}{Odds|_x}$ cuantifica la magnitud de la relación entre la variable respuesta y el cambio en una unidad de la variable de entrada de interés.

Aplicación

Se evalúa una escala de logro académico (ELA) que está conformada por 6 subescalas para medir logros en lectura de frases y palabras, adquisición de código escrito y dominio de repertorio numérico. El constructo ELA se dicotomiza tomando como categorías si logra la totalidad de las subescalas o no y sobre éste se aplica análisis multinivel. Las variables utilizadas en el modelo refieren, además del sexo y la edad, a la escucha, la atención, la memoria, la comprensión de consignas, la autonomía y el permanecer sentado (con ellas se conforma el constructo *Autorregulación*), que son habilidades necesarias e imprescindibles para todo proceso de aprendizaje.

Media por grupo

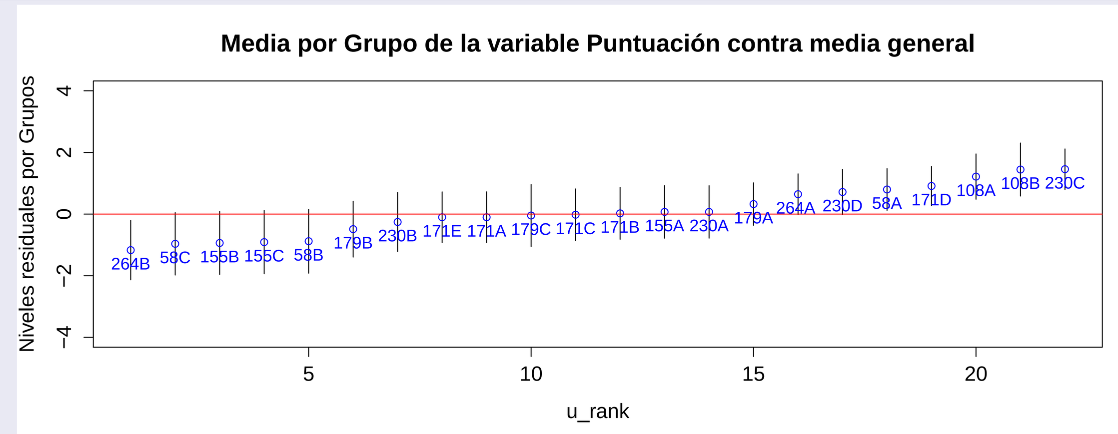

Figura: Se ordenan los grupos en forma creciente a partir de los valores estimados de los u_j.

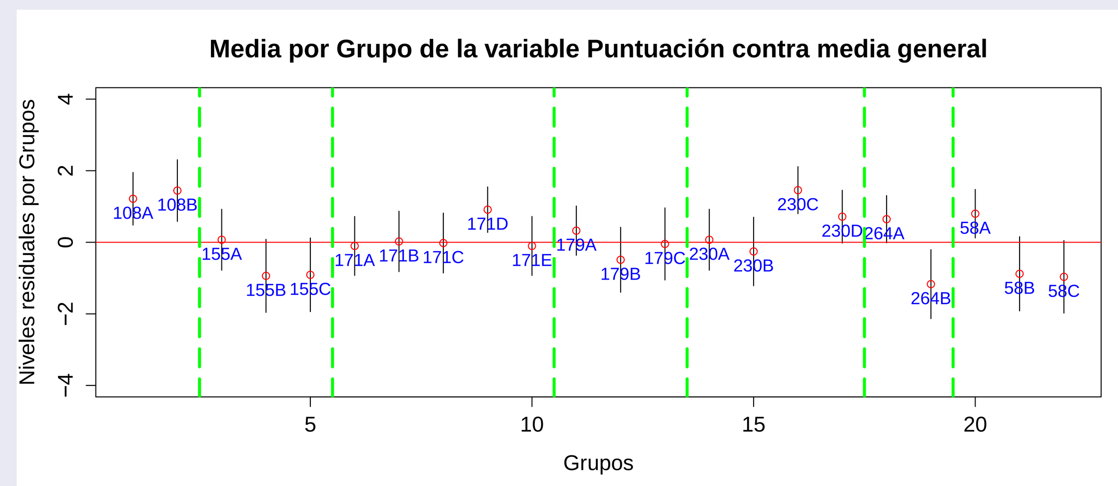

Figura: Se ordenan los grupos por escuela.

Efecto de los grupos

Para discutir la existencia del efecto de los grupos, comenzamos estimando un modelo de dos niveles, modelo con sólo el efecto del intercepto y el de los grupos, y lo comparamos mediante la *desvianza* con un modelo lineal nulo.[3]

Modelo de dos niveles:	**Modelo nulo:**
$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + u_{0j}$	$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0$
$AIC : 301{,}6$	$AIC : 313{,}39$
$LogLik : -148{,}78$	$LogLik : -155{,}69$

De las estimaciones del modelo (utilizando la aproximación de Laplace), podemos decir que el log-odds de probabilidades de lograr la totalidad de las subescalas para un grupo "promedio"(uno con $u_{0j} = 0$) se estima con un $\beta_0 = -1{,}98$. La intersección del grupo j es -1.98 + u_{0j}, donde la varianza de u_{0j} estimada es $\hat{\sigma_u^2} = 1{,}019$.
Desvianza La desvianza D se define como 2 veces la diferencia entre el logaritmo de la verosimilitud de los modelos a testear, independientes entre sí.
$D = 2 * (-148{,}78 - (-155{,}69)) = 13{,}82^a$ $D \sim \chi_1^2$, con *1-p-valor* de 0.00857. Con lo cual hay evidencia de que la varianza entre grupos es diferente de cero.

[a] La prueba estadística no tiene una distribución de muestreo estándar ya que la hipótesis nula de varianza cero está en el límite del espacio de parámetros. En este caso, el p-valor correcto es la mitad del obtenido a partir de las tablas de la distribución chi-cuadrado con 1 grado de libertad.

Variables explicativas

Se estima un modelo para el log-odds de probabilidades de lograr la totalidad de las subescalas:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = -8{,}481 - 0{,}035 * \text{E.Meses}_{ij} + 1{,}736 * \text{A.reg}_{ij} - 0{,}753 * \text{Sexo}_{ij} + u_j,$$

con $u_j \sim N(0; 1{,}103)$ AIC: 287.1

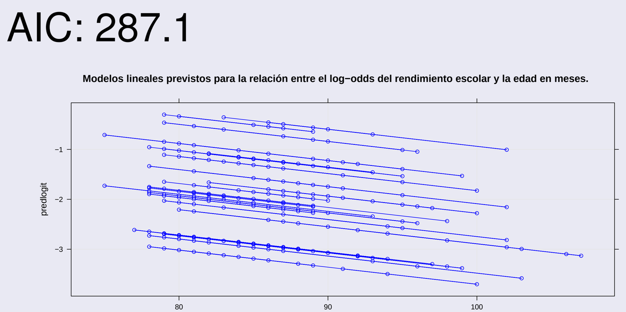

Figura: Modelos lineales previstos para la relación entre el log-odds del rendimiento escolar y la variable EDAD en meses.

Para un niño de 80 meses, el log-odds de probabilidades de lograr el óptimo escolar varía de aproximadamente -0.5 a -3 dependiendo del grupo al que asista. Esto se traduce en una gama de probabilidades de $\frac{\exp^{-0{,}25}}{[1+\exp^{-0{,}25}]} = 0{,}44$ a $\frac{\exp^{-3}}{[1+\exp^{-3}]} = 0{,}05$, por lo que hay fuertes efectos comunitarios. Para la variable *Autorregulación*, $\exp^{1{,}736} = 5{,}675$, vemos que el aumentar una categoría en la variable aumenta en 5.7 el ratio de tener un óptimo logro académico, en cambio para la variable *Sexo*, $\exp^{-0{,}753} = 0{,}471$, vemos que el ser varón lo disminuye en un 52.9 % [a]

[a] 1-hombre, 0-mujer

Alternativa metodológica a seguir

Modelo Umbral de intercepto aleatorios de dos niveles:
Se puede describir un modelo de umbral para una estructura de dos niveles, como una representación alternativa del modelo de intercepto aleatorios lineal generalizado dada por (2).
Se supone que subyace a la variable y_{ij} de respuesta binaria una variable latente continua y_{ij}^* que está relacionada con la observada:

$$y_{ij} = \begin{cases} 1 & \text{si } y_{ij}^* \leq 0 \\ 0 & \text{si } y_{ij}^* > 0 \end{cases} \tag{9}$$

El umbral o punto de corte es arbitrario porque y_{ij}^* es no observada. Podemos definir un modelo de intercepto aleatorios de dos niveles para y_{ij}^*, al igual que para cualquier variable continua:

$$y_{ij}^* = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}^* \tag{10}$$

en donde e_{ij}^* son los residuos del nivel 1 con media cero y varianza $\sigma_{e^*}^2$.
Debido a que y_{ij}^* es no observada, tenemos que establecer su escala, esto lo hacemos fijando $\sigma_{e^*}^2$. También tenemos que especificar la distribución de e_{ij}^*.
Suponiendo una distribución normal con $\sigma_{e^*}^2$= 1 conduce a un modelo probit de intercepto aleatorios, mientras que una distribución logística con $\sigma_{e^*}^2$= 3.29 lleva a un modelo logit de intercepto aleatorios. La forma logit del modelo se denomina a veces modelo logistic-normal porque el nivel residual 1 se supone que sigue una distribución logística, mientras que el nivel 2 residual se supone normal.

Referencias Bibliográficas

1. Goldstein H. (1991) Nonlinear multilevel models, with an application to discrete response data. Biometrika, 78, 45-51.
2. Rodríguez G. and Goldman N. (1995) An assessment of estimation procedures for multilevel models with binary responses. Journal of the Royal Statistical Society, Series A, 158, 73-89.
3. Snijders T.A.B. and Bosker R.J. (1999) Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Sage Publications Ltd, London.
4. Goldstein H. (2003) Multilevel Statistical Models. 3rd edn. Arnold, London.