



UNIVERSIDAD DE LA REPÚBLICA
Facultad de Ciencias Económicas y de Administración
Instituto de Estadística

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

Ramón Álvarez Sebastián Gadea
Diciembre 2014

Serie DT (14 / 05) - ISSN : 1688-6453

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

Ramón Álvarez ¹

Instituto de Estadística - Facultad de Ciencias Económicas y de Administración - UdelaR.

Sebastián Gadea ²

Instituto de Estadística - Facultad de Ciencias Económicas y de Administración - UdelaR.

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 individuos seleccionados de un mismo grupo tiendan a ser más parecidos que dos individuos seleccionados de entre los diferentes grupos. Por ejemplo, los niños aprenden en las clases nutriéndose de las condiciones de su grupo, tales como características de los maestros y la capacidad de otros niños en la clase, lo que 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 modelos de componentes de la varianza - para analizar los datos con una estructura jerárquica.

Se hace una breve introducción a los modelos multinivel y luego se presenta el caso particular de variables de respuesta binarias. Se presenta una nueva metodología de estimación del modelo para este tipo de variables y a continuación, se expone la aplicación concreta, donde se aplican modelos de intercepto aleatorio.

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), 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 el logro de la totalidad de las subescalas o no y, sobre éste se aplica análisis multinivel. Los resultados muestran que es necesario trabajar con modelos más complejos, manejando como alternativas futuro modelos de 3 niveles, con pendiente aleatoria o modelos de Umbral con interceptos aleatorios.

Palabras clave: Escalas, efectos fijos, efectos aleatorios, modelos multinivel, Variables binarias.

¹ramon@iesta.edu.uy

²sgadea@iesta.edu.uy

1. Introducción

Cuando se analiza la performance de estudiantes, hay que tener en cuenta que las observaciones que corresponden a individuos de la misma clase o grupo no pueden ser consideradas como independientes. Los estudiantes que pertenecen a una misma clase pueden verse como perteneciendo a una jerarquía, de manera tal que los individuos en esta situación siguen una especie de “clusterización”.

Si los individuos forman grupos o clusters, podríamos esperar que dos de ellos seleccionados de un mismo grupo tiendan 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, lo que puede influir en el logro educativo de un niño.

Esta situación se puede ver como una estructura de dos niveles de datos, donde el primer nivel son los estudiantes y el segundo nivel son las clases. Más aún, las clases que, en este caso están en el nivel 2 de la jerarquía, también forman parte de una estructura jerárquica más amplia, donde las escuelas pueden estar clusterizadas entre sí (Gelman and Hill, 2006).

Por lo tanto, para evaluar esas dependencias se recurre a los modelos multinivel - también conocidos como modelos jerárquicos lineales, modelos mixtos, modelos de efectos aleatorios y modelos de componentes de la varianza - para analizar los datos con una estructura jerárquica.

El análisis multinivel (AM) se desarrolló inicialmente para la investigación educativa (Goldstein, 1991b).

La misma situación de las escuelas se da en otras disciplinas, como por ejemplo en salud, cuando se lleva a cabo un estudio para investigar la relación entre el total de colesterol (CT) y edad.

$$CT = \beta_0 + \beta_1 * edad + \beta_2 * género + \varepsilon \quad (1)$$

Los pacientes en estudio en general pueden provenir de diferentes médicos consultantes, (es como habitualmente se pueden hacer los estudios) con lo cual se puede pensar que existe una estructura jerárquica que responde al médico del cual provienen los pacientes del estudio. Eso se puede resolver incorporando variables *dummies* que tengan en cuenta el médico tratante. Si hubiese 12 médicos involucrados, serían necesarias 11 variables *dummies* en el modelo, lo que implicaría una pérdida de potencia muy grande. Mas aún, si se desease comparar el modelo para un valor dado de edad en las mujeres por ejemplo, se tendría 11 modelos lineales univariados con diferentes interceptos, pero una misma pendiente. La idea que está detrás del (AM) es, en lugar de considerar los interceptos por separado, considerar la varianza del intercepto (Hox, 1995).

Si, además, se puede suponer que las observaciones están clusterizadas entre los médicos

el modelo convencional que incorpora interacción entre edad y tipo de médico es el de la ecuación ((2)).

$$\begin{aligned} \text{CT} = & \beta_0 + \beta_1 * \text{edad} + \beta_2 * \text{D1} + \dots \beta_m * \text{Dm-1} + \\ & \dots + \beta_{m+1} * \text{D1} * \text{edad} + \dots \beta_{2m-1} * \text{D11} * \text{edad} + \epsilon \end{aligned} \quad (2)$$

Si lo que interesa para el ejemplo planteado no es tener una pendiente diferente para cada médico sino un efecto global de interacción, con el (AM) se puede estimar la *varianza* de la pendiente, en lugar de considerar una pendiente diferente para cada combinación de edad y médico (Twisk, 2006).

Otro aspecto importante a considerar y que puede resultar una guía para saber si es necesario usar el enfoque multinivel, es estudiar el coeficiente de correlación intraclase (CCI) entre individuos que están clusterizados o anidados en niveles jerárquicos superiores (para el caso de los estudiantes - clase - escuela - distrito escolar o para el ejemplo del estudio sobre colesterol los -médicos-hospitales-ciudades) (Finch, 2014).

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2} \quad (3)$$

donde τ^2 representa la varianza entre clusters y σ^2 es la varianza dentro de los clusters. Un valor elevado de ρ estaría indicando que la variable de respuesta está muy asociada con la pertenencia de la observación a un cluster o, lo que es lo mismo, los individuos dentro del mismo grupo (por ejemplo, la escuela), son más parecidos en la variable medida de lo que son estudiantes de otro cluster.

Para evaluar el ICC (Finch, 2014) propone hacer un análisis de la varianza (ANOVA) debiendo estimar $\hat{\rho}$ usando para eso $\hat{\tau}^2$ y $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (n_j - 1) S_j^2}{N - C} \quad (4)$$

donde S_j^2 es la varianza intraclase

$$S_j^2 = \frac{\sum_{i=1}^n (y_{ij} - \bar{y}_j)^2}{n_j - 1} \quad (5)$$

n_j es el tamaño del cluster j , N es el tamaño total de la muestra, y C es el número total de clusters. En otras palabras, σ^2 es simplemente el promedio ponderado de la varianza intraclusters.

Para estimar $\hat{\tau}^2$ se debe calcular la varianza ponderada interclase

$$\hat{S}_B^2 = \frac{\sum_{i=1}^c (\bar{y}_j - \bar{y})^2}{\tilde{n}(C-1)} \quad (6)$$

donde \bar{y}_j es la media de la variable de respuesta en el cluster j y \bar{y} es la media global

$$\tilde{n} = \frac{1}{C-1} \left[N - \frac{\sum_{i=1}^c n_j^2}{N} \right] \quad (7)$$

Como no se puede usar directamente \hat{S}_B^2 como la estimación de τ^2 , ya que está impactada por la variación aleatoria de los individuos dentro de los clusters, se corrige mediante la siguiente relación

$$\hat{\tau}^2 = \hat{S}_B^2 - \frac{\hat{\sigma}^2}{\tilde{n}} \quad (8)$$

de donde puede calcularse

$$\rho_I = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{S}_B^2} \quad (9)$$

Antes de que el análisis multinivel fuera desarrollado, el problema de observaciones correlacionadas se abordaba de dos maneras: o bien ignorando el hecho de que estaban correlacionadas, o combinando las observaciones en un único valor. Por lo tanto, se debía calcular una especie de valor medio de las observaciones para cada grupo en primer término, para luego utilizar estos promedios en un análisis de regresión estándar. Este método se conoce como método de agregación.

Los 2 modelos de (AM) con intercepto aleatorio o con intercepto y pendiente aleatoria se estiman mediante máxima verosimilitud y coinciden con los modelos de regresión que, en la literatura de investigación se presentan con variedad de nombres tales como “modelo de coeficientes aleatorios” (de Leeuw y Kreft, 1986; Longford, 1993), “modelo de componentes de varianza” (Longford, 1986) o más recientemente, como un caso particular de los “modelos mixtos” que se presentan en (McCulloch, 2001).

En general se puede ver el (AM) como un caso de modelo mixto

$$\mathcal{Y} = \underbrace{\mathcal{X}\beta}_{\text{Coeficientes fijos}} + \underbrace{\mathcal{U}\delta}_{\text{Coeficientes aleatorios}} + \epsilon \quad (10)$$

que en forma escalar se puede representar como plantean (de Leeuw and Meijer, 2008)

$$y_i = \sum_{q=1}^{q=r} x_{qi} \beta_q + \sum_{s=1}^{s=p} u_{si} \delta_s + \epsilon_i \quad (11)$$

Una de las mayores ventajas del análisis multinivel es que puede ser utilizado para el análisis de otros tipos de variables de respuesta también, como el análisis multinivel logístico (AML), análisis multinivel de Poisson (AMP) para variables de conteo e incluso se puede desarrollar análisis de supervivencia de varios niveles.

También el (AM) puede ser usado para el modelado de datos longitudinales, donde los datos están correlacionados, al ser, por ejemplo, los pacientes los que se repiten y para los cuales se les mide una variable *tiempo-dependiente*. Este tipo de datos tiene una forma jerárquica que responde a individuos o unidades que se miden en más de una ocasión, como sucede en los estudios de crecimiento humano. Aquí las ocasiones se agrupan dentro de los individuos que representan las unidades de nivel 2 con las medidas en diferentes momentos que representan el nivel 1; en general hay mucha más variación entre los individuos que en las ocasiones dentro de los individuos (Hox, 1995).

En el ejemplo que se presenta en la sección 3 para el caso de un grupo de escolares interesa conocer el constructo ELA (escala de logro académico) teniendo en cuenta la estructura de la población, por lo cual el análisis efectuado incorpora los modelos multinivel. Se presenta en la sección una nueva metodología de estimación del modelo, en este caso de datos binarios, se expone la aplicación concreta, y se comenta brevemente metodologías alternativas a seguir a futuro.

2. Modelo multinivel

Retomando el ejemplo de los estudiantes que pertenecen a diferentes escuelas, es pertinente considerar la posibilidad de una estructura de dos niveles, donde un total de n niños (en el nivel 1) se anidan dentro de J grupos (en el nivel 2) con n_j individuos en el grupo j . y_{ij} representa la respuesta para el individuo i en el grupo j .

$$y_{ij} = \pi_{ij} + \epsilon_{ij} \quad (12)$$

donde $\pi_{ij} = E(y_{ij}|x_{ij}, u_j) = P(y_{ij} = 1)$. El modelo de intercepto aleatorio lineal generalizado para la dependencia de la respuesta (probabilidad π_{ij} en x_{ij}) se escribe tal como aparece en (Goldstein, 2003)

$$F^{-1}(\pi_{ij}) = \beta_0 + \beta_1 x_{ij} + u_j \quad (13)$$

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.

Se ve cómo 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 se desean estimar son β_0 , β_1 , y σ_u^2 , y que aparecen en la ecuación (18).

2.1. Modelo logit con intercepto aleatorio

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

$$f(y_{ij}) = \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{(1-y_{ij})} \quad (14)$$

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 logro académico óptimo, se tiene que:

$$\pi_{ij} = \frac{\exp^{\beta_0 + \beta_1 x_{ij} + u_{ij}}}{1 + \exp^{\beta_0 + \beta_1 x_{ij} + u_{ij}}} \quad (15)$$

Luego, si se aplica $\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$ (es decir $\log(\text{odds})$) se tiene que

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{ij} + u_{ij} \quad (16)$$

Observar 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 este modelo se denomina regresión logística.

2.2. 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):

$$L(\beta_0, \beta_1, \sigma_u^2 | u) = \prod_{j=1}^J \prod_{i=1}^{n_j} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \quad (17)$$

La función de probabilidad dada por (17) se denota $L(\beta_0, \beta_1, \sigma_u^2 | u)$ ya que su valor está condicionada al valor efecto u_j que es aleatorio, siendo en este caso, una probabilidad condicional. Debido a que u_j es inobservable, se trabaja 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:

$$L(\beta_0, \beta_1, \sigma_u^2) = \int L(\beta_0, \beta_1, \sigma_u^2 | u) \phi(u) du \quad (18)$$

donde $\phi(u)$ es la función de densidad de probabilidad de una distribución normal. Los métodos más utilizados de 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 conjuntos de datos voluminosos.

Para los parámetros es importante ver los siguientes aspectos:

1. En primer lugar se tiene β_0 que representa al intercepto global.
2. Por otra parte $\beta_0 + u_j$ representa al intercepto para el grupo j en la relación lineal entre el $\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$ y x .
3. σ_u^2 es la varianza entre grupos ajustada para x , o simplemente varianza residual del nivel 2.
4. El odds-ratio (OR) = $\frac{Odds|_{x+1}}{Odds|_x} = \exp^\beta$, sirve para cuantificar la magnitud de la relación entre la variable respuesta que se mide y el cambio en una unidad de la variable de entrada de interés.
5. β_1 es el efecto de un cambio de 1 unidad de x_1 en el log-odds de que $y = 1$.

3. Aplicación

Se trabaja con 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. Las variables de logro académico se componen a partir de tres bloques: *Lectura*, *Escritura espontánea* y *Numeración*. El bloque de *Lectura* lo componen las variables Lectura de palabra (Lect.Palabra) y Lectura de frase (Lect.Frases), éstas enumeran cuantas palabras o frases respectivamente fueron leídas por el niño. El recorrido de ambas variables es $\{0, 1, 2, 3, 4, 5, 6, 7\}$.

El bloque de *Escritura espontánea* lo componen la variables Etapa, Estructura y Escritura de nombre.

$$\text{Etapa} = \begin{cases} 0 & \text{Nada} \\ 1 & \text{HC-HV} \\ 2 & \text{Silábica} \\ 3 & \text{Silábica- Afabética} \\ 4 & \text{Afabética} \end{cases}$$

$$\text{Estructura} = \begin{cases} 0 & \text{Nada} \\ 1 & \text{Palabra} \\ 2 & \text{Enunciado} \\ 3 & \text{Texto} \end{cases}$$

$$\text{Escritura de nombre} = \begin{cases} 0 & \text{Nada} \\ 1 & \text{No logrado} \\ 2 & \text{Parcialmente logrado} \\ 3 & \text{Logrado} \end{cases}$$

El tercer bloque, *Numeración*, lo compone la variable Grilla 1, trata de evaluar si el niño es capaz de completar un cuadro con la numeración del 1 al 15 en los cuales hay números faltantes.

La variable ha sido recodificada:

$$\text{Grilla.1.rec} = \begin{cases} 0 & \text{Grilla 1} < 6 \\ 1 & 6 \leq \text{Grilla 1} \leq 12 \\ 2 & \text{Grilla 1} > 12 \end{cases}$$

El constructo ELA se dicotomiza tomando como categorías el logra de 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 en meses, son 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.

Para los diferentes modelos que se estiman se trabaja con el software *R* (R Core Team, 2014) usando diferentes librerías como

- *lme4* (Bates et al., 2014), (Bates et al., 2014)
- *psych*, (Fox and Weisberg, 2011)
- *lattice* (Sarkar, 2008) para el manejo de los gráficos.

Los resultados preliminares fueron presentados en la sesión de posters del XI Congreso de Sociedades Latinoamericanas de Estadística', CLATSE, La Serena, Chile (2014) (Alvarez and Gadea, 2014).

Se modifican los scripts desarrollados y presentados en el curso on line LEMMA (Learning environment for multilevel methodology and applications) del Centre for Multilevel Modelling de la Universidad de Bristol -(Szmaragd and Leckie, 2014).

3.1. Efecto de los grupos

Para discutir la existencia del efecto de los grupos, se comienza por estimar un modelo de dos niveles, es decir un modelo que sólo da cuenta del efecto del intercepto y el de los grupos, que se compara mediante la *desvianza* con un modelo lineal nulo, tal como plantean (Snijders and Bosker, 1999).

Modelo de dos niveles:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0 + u_j$$
$$AIC : 301,6$$
$$LogLik : -148,78$$

Modelo nulo:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_0$$
$$AIC : 313,39$$
$$LogLik : -155,69$$

De las estimaciones del modelo (utilizando la aproximación de Laplace), se puede decir que el logaritmo de los odds de lograr la totalidad de las subescalas en un grupo "promedio" (con $u_j = 0$) se estima con un $\hat{\beta}_0 = -1,98$. El intercepto del grupo j es $-1,98 + u_j$, donde la varianza de u_j estimada es $\hat{\sigma}_u^2 = 1,019$.

Para comparar ambos modelos se utiliza la desvianza D que se define como 2 veces la diferencia entre el logaritmo de la verosimilitud de los modelos a testear, independientes entre sí.

$$\mathcal{D} = -2 * (\mathcal{L}(\text{modelo}_1) - \mathcal{L}(\text{modelo}_0)) \quad (19)$$

$\mathcal{D} = 2 * (-148,78 - (-155,69)) = 13,82$ $\mathcal{D} \sim \chi_1^2$, con *p-valor* de 0.00857. Con lo cual hay evidencia de que la varianza entre grupos es diferente de cero.

La intersección $\hat{\beta}_0$ es compartida por todos los grupos, mientras que el efecto aleatorio u_j es específico del grupo j .

Se puede observar que los intervalos de confianza alrededor de las estimaciones residuales, varían mucho en su amplitud; los intervalos más amplios corresponden a grupos pequeños, disminuyendo el tamaño del intervalo de confianza, a medida que el tamaño del grupo aumenta.

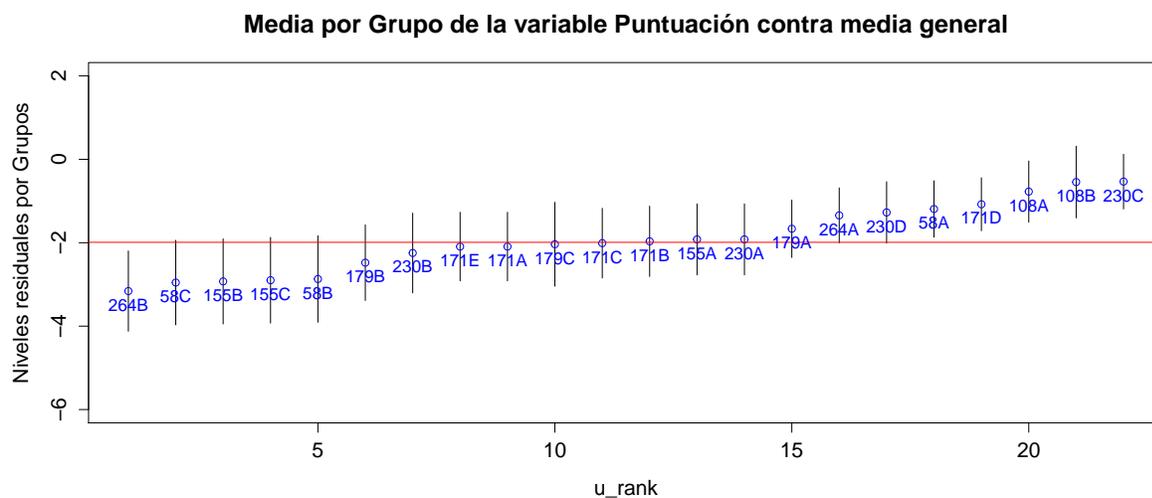


Figura 1: Grupos ordenados de forma creciente a partir de los valores estimados de los u_j .

3.2. Modelos con Variables explicativas

Se estima un modelo con la variable *edad en meses* (X_1):

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = -1,995 - 0,034 * X_{1ij} + u_j$$

con $u_j \sim N(0; 1,073)$ AIC: 302.0.

Este modelo puede descomponerse en 2 submodelos

$$y_{ij} = \beta_{0j} + \beta_j X_{1ij} + \epsilon_{ij} \tag{20}$$

$$\beta_{0j} = \beta_0 + u_j \tag{21}$$

Variable	Coefficiente	Std.	Error z	$Pr(> z)$
(Intercepto)	-1.995	0.0018014	-1103.4	$< 2e - 16 * **$
(X_1) Edad en meses	-0.0334	0.0017982	-18.6	$< 2e - 16 * **$

Cuadro 1: Modelo de Regresión con edad

El gráfico de la figura 3.2 muestra las diferentes regresiones, una para cada uno de los grupos.

Se observa que para un niño de 80 meses, el log-odds de lograr el óptimo escolar varía de aproximadamente -0.25 a -3 dependiendo del grupo al que asista. Esto se traduce en un rango 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 en la probabilidad de que un niño logre el óptimo escolar.

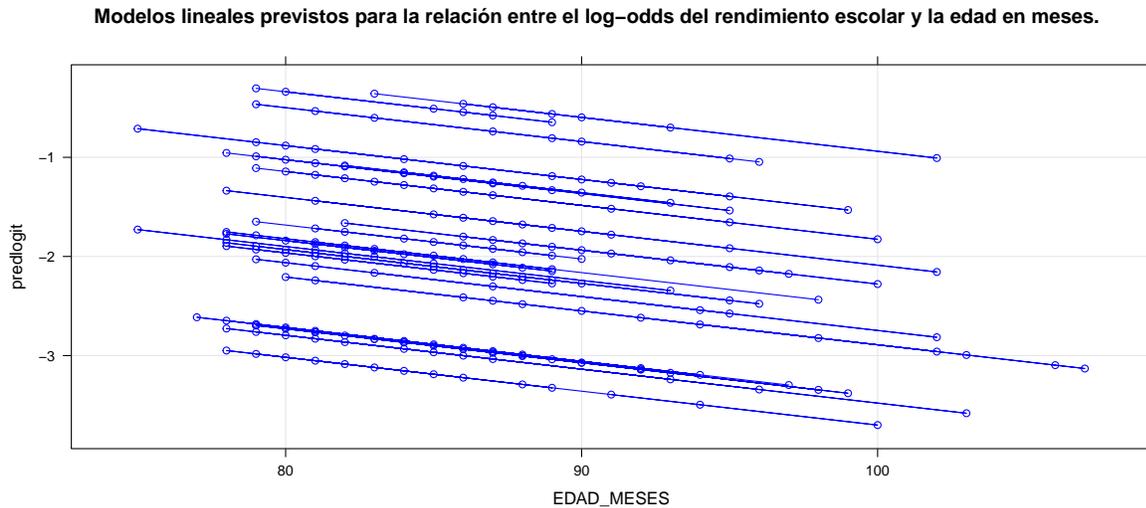


Figura 2: Modelos lineales log-odds del rendimiento escolar y la variable EDAD en meses para los diferentes grupos.

Se explora también una relación cuadrática entre la variable *edad en meses* y el óptimo logro académico:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = -1,987 - 0,033 * X_{1ij} - 0,0002 * X_{1ij}^2 + u_j$$

con $u_j \sim N(0; 1,073)$ AIC: 304.0

El término cuadrático no resulta significativa para el análisis.

Se considera un modelo ampliado para el log-odds de probabilidades de lograr la totalidad de las subescalas con las variables *Edad en meses* (X_1) *Autorregulación* (X_2), *Sexo* (X_3)

Variable	Coefficiente	Std.	Error z	$Pr(> z)$
(Intercepto)	-1.987	0.0018014	-1103.4	$< 2e - 16$ ***
EDAD en MESES X	-0.0334	0.0017982	-18.6	$< 2e - 16$ ***
EDAD en MESES X^2	-0.00021	0.0016175	-0.1	0.898

Cuadro 2: Modelo de Regresión con edad con término cuadrático

:

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = -8,481 - 0,035 * X_{1ij} + 1,736 * X_{2ij} - 0,753 * X_{3ij} + u_j,$$

con los efectos aleatorios $u_j \sim N(0; 1,103)$ AIC: 287.1

Variable	Coefficiente	Std.	Error z	$Pr(> z)$	OR
(Intercepto)	-8.480	0.001854	-4575	$< 2e - 16$ ***	0.0002
(X_2)Autoregulación	1.736	0.001856	935	$< 2e - 16$ ***	5.6760
(X_3)Sexo	-0.752	0.001914	-393	$< 2e - 16$ ***	0.4711
(X_1)Edad en meses	-0.0349	0.001911	-18	$< 2e - 16$ ***	0.9656

Cuadro 3: Modelo de Regresión con 3 variables

Se observa que, para la variable *Autorregulación*, $\exp^{1,736} = 5,675$, al aumentar una categoría en la variable aumenta en 5.7 el odds de tener un óptimo logro académico, en cambio para la variable *Sexo*, vemos que el ser varón disminuye el ratio en un 52,9% ($\exp^{-0,753} = 0,471$)³

³1-hombre, 0-mujer

4. Conclusiones y futuros pasos

Se estimaron modelos de interceptos aleatorios para estudiar los efectos de los grupos, donde los resultados preliminares muestran evidencia de que la varianza entre grupos es diferente de cero; luego se modeló la variable ELA a través de la autoregulación, el sexo y la edad en meses, usando para eso un modelo de intercepto aleatorio con efectos fijos en las variables regresoras.

Se propone seguir ensayando modelos multinivel que sean mas complejos, incorporando otras posibles fuentes de variación

4.1. Modelo con pendiente aleatoria

Se puede a su vez trabajar con un modelo logit con intersección aleatoria que está dado por la ecuación (16), el que se puede extender a un modelo con pendiente aleatoria:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \underbrace{\beta_0 + \beta_1 x_{ij}}_{\text{efectos fijos}} + \underbrace{u_{0j} + u_{1j} x_{ij}}_{\text{efectos aleatorios}} \quad (22)$$

en donde los efectos aleatorios $u = \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim NMV(\vec{0}, \Omega_u)$, con $\Omega_u = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}$

4.2. Modelo de tres niveles

Al observar la figura 3 con los grupos agrupados por escuela, (las líneas verdes punteadas dividen los grupos por escuela), se puede ver que el comportamiento de los grupos en las escuelas no es uniforme, característica que se puede estudiar, en forma más avanzada, con un análisis que contemplando un modelo de tres niveles (Snijders and Bosker, 1999).

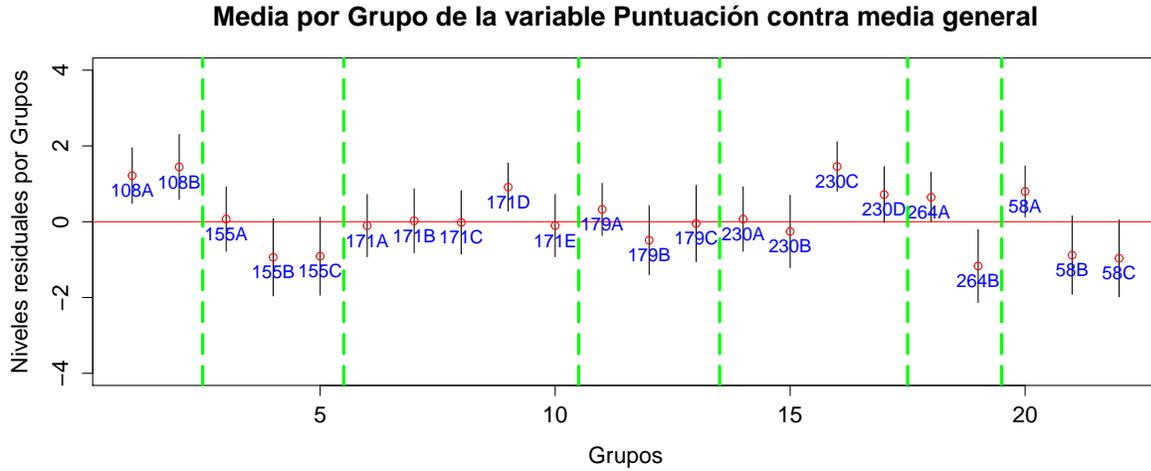


Figura 3: Grupos ordenados por escuela a partir de los valores estimados de los u_j .

4.3. Modelo de Umbral con interceptos aleatorios en dos niveles

Se puede describir a su vez un modelo de umbral para una estructura de dos niveles, como una representación alternativa del modelo lineal generalizado de interceptos aleatorios. Se supone que a la variable y_{ij} de respuesta binaria subyace una variable continua latente y_{ij}^* que está relacionada con la observada de la siguiente manera:

$$y_{ij} = \begin{cases} 1 & \text{si } y_{ij}^* \leq 0 \\ 0 & \text{si } y_{ij}^* > 0 \end{cases} \quad (23)$$

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

$$y_{ij}^* = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}^* \quad (24)$$

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, hay que establecer su escala, lo que se hace fijando $\sigma_{e^*}^2$. También es necesario especificar la distribución de e_{ij}^* . Suponer una distribución normal con $\sigma_{e^*}^2 = 1$ conduce a un modelo probit con intercepto aleatorio, mientras que, una distribución logística con $\sigma_{e^*}^2 = 3.29$, lleva a un modelo logit con intercepto aleatorio. Este 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.

A. Apéndice

Una alternativa para el método de máxima verosimilitud es el uso de una aproximación que reemplace la ecuación (17) por un modelo lineal.

Procedimientos de cuasi-verosimilitud implican la aproximación π_{ij} en 17 por una expresión lineal, utilizando un desarrollo en serie de Taylor de $\pi_{ij} = F(\beta_0 + \beta_1 x_{ij} + u_j)$. (Goldstein, 1991a) 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: (Rodríguez and Goldman, 1995)).

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:

$$\pi^{(m+1)} \approx \pi^{(m)} + w^{(m)}[(\beta_0 - \beta_0^{(m)}) + (\beta_1 - \beta_1^{(m)})x] + w^{(m)}u \quad (25)$$

Después de sustituir ((25)) en $y_{ij} = \pi_{ij} + \epsilon_{ij}$ y reordenando, se obtiene 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 $\tilde{u}_j = w^{(m)}u_j$.

Referencias

Alvarez, R. and Gadea, S. (2014). Aplicación de modelos multinivel para variables binarias en estudios sobre logros académicos en escolares. Poster presentado en el XI Congreso de Sociedades Latinoamericanas de Estadística.

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv e-prints*.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *me4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.
- de Leeuw, J. and Meijer, E. (2008). *Handbook of Multilevel Analysis*. Springer.
- Finch, W. (2014). *Multilevel modeling using R*. CRC Press, Boca Raton, Florida.
- Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition.
- Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Goldstein, H. (1991a). Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78(1):45–51.
- Goldstein, H. (2003). *Multilevel Statistical Models*. Arnold, 3rd edition.
- Goldstein, L. P. H. (1991b). New statistical methods for analysing social structures: An introduction to multilevel models. *Educational Research Journal*, 17(4):387–393.
- Hox, J. (1995). *Multilevel Analysis Techniques and Applications*. Lawrence Erlbaum Associates, Mahwah, N.J.
- McCulloch, C. (2001). *Generalized, linear, and mixed models*. John Wiley & Sons, New York.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- 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.
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5.
- Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications Ltd.
- Szmaragd, C. and Leckie, G. (2014). Module 5: Introduction to multilevel modelling.
- Twisk, J. W. (2006). *Applied Multilevel Analysis*. Cambridge University Press.