MODELACIÓN POISSON CON ENFOQUE BAYESIANO PARA EXPLICAR EL EFECTO DE LA EDUCACIÓN Y DEL ÁREA DE RESIDENCIA SOBRE LA MORTALIDAD DURANTE LOS PRIMEROS AÑOS DE VIDA

Lic. Loza Cruz, Patricia
lcruzpatricia@gmail.com

RESUMEN

El presente trabajo de investigación es parte de un estudio mucho más amplio en el que se analiza, con el uso de modelos tipo Poisson, los factores determinantes de la mortalidad durante los primeros años de vida a partir de dos enfoques estadísticos: el bayesiano y el frecuentista. En este trabajo sólo se presentan algunos resultados basados en el enfoque bayesiano para determinar el efecto de la educación de las madres y el lugar de su residencia sobre los niveles de mortalidad neonatal (MNN), post-neonatal (MPNN) y post-infantil (MPI). La información usada proviene de la encuesta nacional de demografía y salud realizada en 2008. Se concluye que la educación de las madres tiene un alto impacto positivo sobre los tres tipos de mortalidad, luego de controlar las demás covariables. La tasa de MNN en hijos de madres con educación alta es 38.1% menos que en hijos de madres con educación baja o sin educación. Esta diferencia se incrementa a 47.8% para la MPNN y a 60.4% en el caso de la MPI. El problema de sobredispersión fue superado a través de dos mecanismos: una especificación adecuada del modelo y el uso de un modelo binomial negativo. Se recomienda impulsar políticas orientadas a incrementar la cobertura de la educación principalmente secundaria, de mujeres y focalizada en áreas rurales del país.

PALABRAS CLAVE

Enfoque bayesiano, mortalidad neo-natal, mortalidad post-neonatal, mortalidad post-infantil.

1. INTRODUCCIÓN

Es de gran importancia para los planificadores del sector de la salud del país conocer tanto el nivel como la evolución de la mortalidad en los primeros años de vida, pues ello les permite evaluar la eficacia de las acciones pasadas en el área, así como delinear nuevas políticas y programas. Empero, también es de vital importancia conocer los factores socioeconómicos que determinaron tanto esos niveles de mortalidad como su tendencia.

En el orden metodológico, un modelo de regresión Poisson es adecuado para abordar el análisis de los factores determinantes de la mortalidad en los primeros años de vida. El problema de sobredispersión, típico en datos tipo Poisson, puede superarse recurriendo o bien a un "mejor" tratamiento de la información en la etapa de modelación o bien a modelos alternativos.

2. OBJETIVO

Determinar el efecto de la educación y lugar de residencia sobre el nivel de la mortalidad neonatal, post-neonatal y post-infantil, durante el periodo de análisis (1993-2007).

3. INFORMACIÓN

La Encuesta Nacional de Demografía y Salud, principal fuente de información del País con relación al sector salud, es la más adecuada para realizar la presente investigación. En el módulo de "historia de nacimientos" de la ENDSA 2008 se obtiene información acerca de la fecha de nacimiento, edad actual y edad al morir si es el caso, entre otros datos, para cada uno de los hijos nacidos vivos de las mujeres en edad fértil entrevistadas. A partir de esta información se elaboró el Cuadro 1, información que se usó en el análisis posterior.

En el Cuadro 1 se muestran las tasas de mortalidad clasificadas por grupo de edad del niño, cohorte de nacimiento del niño, área de residencia y nivel de educación de las madres. Se han definido tres grupos de edad: 0 meses, 1-11 meses y 12-59 meses. Estos grupos de edad corresponden a las edades consideradas para calcular las tasas de mortalidad neonatal (MNN), mortalidad post-neonatal (MPNN) y mortalidad pos-infantil (MPI), respectivamente. Por esta razón, cuando por ejemplo se hace referencia a la tasa de MNN se estará haciendo mención a la mortalidad en el primer mes de vida. La variable cohorte también tiene tres categorías: la cohorte de nacidos vivos en el periodo 1993-1997, la cohorte de 1998-2002 y la cohorte de nacimientos en el periodo 2003-2007.

Cuadro 1
Bolivia: Tasas de mortalidad (por mil) por grupo de edad, según cohorte de nacimiento, área de residencia y nivel de educación, ENDSA 2008

Área de residencia	Nivel de educación	Grupo de edad (en meses)	Cohorte de Nacimiento 1993-1997	1998-2002	2003-2007
Urbana	Baja	0	418.2	362.2	295.0
		1-11	50.3	30.0	27.0
		12-59	7.1	5.6	1.7
Urbana	Alta	0	197.0	265.7	210.0
		1-11	23.0	19.6	11.2
		12-59	3.7	1.5	1.7
Rural	Baja	0	570.1	573.1	465.5
		1-11	57.7	47.1	36.8
		12-59	10.7	7.4	6.4
Rural	Alta	0	215.5	305.2	393.4
		1-11	24.0	21.9	22.7
		12-59	3.7	2.5	1.7

Fuente: Elaboración propia

Con relación a las variables área de residencia y nivel de educación, ambas tienen dos categorías: residencia urbana y rural para la variable área de residencia y educación baja y alta para la variable nivel de educación. A madres sin educación formal o con educación primaria se las ha denominado con educación baja, mientras a las madres con educación secundaria o superior se las ha denominado con educación alta.

4. EL MÉTODO BAYESIANO

A diferencia del enfoque clásico, el enfoque bayesiano considera al parámetro como una variable aleatoria e interpreta la probabilidad desde el punto de vista subjetivo. Para realizar el proceso de inferencia bayesiana es necesario la especificación de una distribución previa o *apriori* de probabilidad, la cual representa el conocimiento acerca del parámetro antes de obtener cualquier información respecto a los datos. Este enfoque tiene como punto central el Teorema de Bayes el cual se detalla a continuación:

Sea $y = y_1, y_2, \dots\dots, y_n$ un vector n de observaciones cuya distribución de probabilidad $f(y, \theta)$ depende de k parámetros involucrados en el vector $\theta = \theta_1, \theta_2, \dots\dots, \theta_k$. Se supone también que θ tiene una distribución de probabilidad previa $f(\theta)$. Entonces la distribución conjunta de θ e y es:

$$f(y,\theta) = f(y|\theta) * f(\theta)$$

$$= f(\theta|y) * f(y)$$

donde la distribución de probabilidad condicional de θ dado el vector de observaciones y tiene la forma

$$f(\theta|y) = \frac{f(y|\theta) * f(\theta)}{f(y)}$$

$$= c * f(y|\theta) * f(\theta)$$

$$= l(\theta|y) * f(\theta)$$

Siendo $c = \frac{1}{f(y)}$ la constante normalizadora, $l(\theta|y)$ representa la función de verosimilitud, $f(\theta)$ la distribución previa y $f(\theta|y)$ la distribución posterior del parámetro.

La distribución de probabilidad previa $f(\theta)$ puede tener información específica acerca del parámetro, así como también puede tener información general y ser considerado como una distribución previa no informativa.

Una selección común para una previa no informativa es la previa propuesta por Jeffreys (1961) quien definió una distribución previa no informativa considerando tranformaciones uno a uno del parametro $\phi = h(\theta)$. Mediante transformacion de las variables se puede ver que la densidad previa $f(\theta)$ es proporcional a la densidad previa para ϕ, esto es

$$f(\theta) \propto \overline{J(\phi)}$$

donde $\overline{(J(\phi)}$ es la matriz de información de Fisher.

Para una variable aleatoria y con una distribución Poisson (θ), la previa no informativa de Jeffreys está dada por:

$$f(\theta) = \overline{J(\phi)} = \frac{1}{\theta}$$

La función de verosimilitud es la función a través de la cual los datos modifican el conocimiento previo de θ. El modelo lineal generalizado Poisson asume que se distribuye Poisson con media θ y con una función de enlace logarítmico, tal que $\ln \theta = X\beta$. En esta situación, la distribución muestral para los datos $y = y_1, y_2, \dots\dots, y_n$ queda expresada como:

$$f(y,\beta) = \prod_{i=1}^{n} \frac{1}{y_1!} e^{\exp(\eta_i)} \exp(\eta_i)^{y_i}$$

donde $\eta_i = X\beta$ es el predictor lineal para el i-ésimo caso.

La distribución posterior contiene toda la información concerniente al parámetro de interés θ. En consecuencia, cualquier inferencia con respecto a θ se hará a partir de dicha distribución. En el caso del modelo Poisson, la distribución posterior llega a ser

$$f(\theta|y) \propto \theta^{\sum_{i=1}^{n} y_i - \frac{1}{2}} e^{-n\theta}$$

la cual es el kernel de una distribución Gamma con parámetros

$$\left(\sum_{i=1}^{n} y_i - \frac{1}{2}, n\right)$$

En el contexto bayesiano, las estimaciones de los parámetros del modelo se obtienen generalmente con el procedimiento denominado muestreador de Gibbs, un procedimiento específico del denominado método MCMC. En términos generales, el procedimiento consiste en extraer muestras en forma sucesiva de las probabilidades condicionales de los parámetros del modelo y consta de los siguientes pasos:

1. Definir $t = 0$, y elegir un valor inicial arbitrariamente de
$$\theta^0 = \theta_1^0, \theta_2^0, \ldots\ldots, \theta_k^0$$
2. Generar cada componente de θ como sigue:

- Mostrar θ_1^{t+1} desde
$$f(\theta_1|\theta_2^{(t)}, \ldots\ldots, \theta_k^{(t)}, y)$$
- Mostrar θ_2^{t+1} desde
$$f(\theta_2|\theta_1^{t+1}, \theta_3^{(t)} \ldots\ldots, \theta_k^{(t)}, y)$$
- ………..
- Mostrar θ_k^{t+1} desde
$$f(\theta_k|\theta_1^{t+1}, \ldots\ldots, \theta_{k-1}^{t+1}, y)$$

3. Generar $t = t+1$. Si $t<T$ el número de muestras deseadas, retornar al paso 2. En otro caso, parar.

5. ALGUNOS RESULTADOS

Con base en criterios estadísticos para la comparación de modelos, el modelo que mejor se ajusta a la información y que, a la vez, supera el problema de sobredispersión es el que en la parte sistemática toma en cuenta, además de los efectos individuales, los efectos de interacción entre el grupo de edad y la cohorte de nacimientos, y la interacción entre el grupo de edad y el nivel de educación. El área de residencia no interactúa con ninguna de las variables analizadas. El modelo final para analizar las tasas de mortalidad, entonces, es representado como:

$$Edad * Cohorte + Edad * Educación + Residencia$$

Cuadro 2
Estimación Bayesiana de los Parámetros para el Modelo Elegido “EDA*COH+EDA*EDU+RESI”

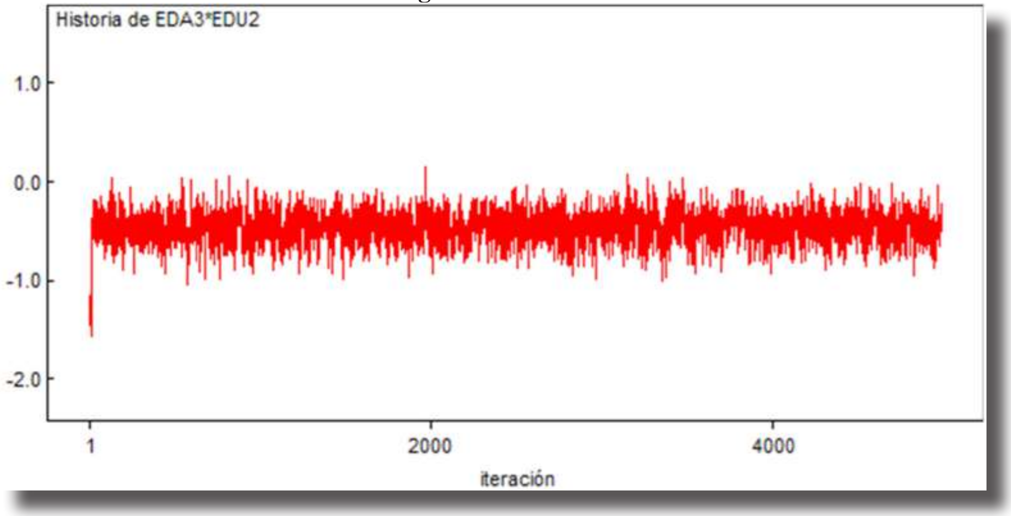
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
eda2	-2.195	0.077	0.005	-2.348	-2.196	-2.040	1001	4000
eda3	-3.946	0.084	0.004	-4.108	-3.946	-3.785	1001	4000
coh2	0.011	0.076	0.004	-0.141	0.011	0.162	1001	4000
coh3	-0.168	0.083	0.004	-0.326	-0.168	0.000	1001	4000
edu2	-0.481	0.088	0.003	-0.654	-0.482	-0.311	1001	4000
res2	-0.356	0.048	0.001	-0.449	-0.357	-0.261	1001	4000
eda2coh2	-0.298	0.108	0.006	-0.504	-0.297	-0.084	1001	4000
eda2coh3	-0.361	0.121	0.006	-0.608	-0.359	-0.129	1001	4000
eda3coh2	-0.399	0.124	0.005	-0.637	-0.401	-0.151	1001	4000
eda3coh3	-0.546	0.169	0.005	-0.879	-0.545	-0.217	1001	4000
eda2edu2	-0.171	0.125	0.004	-0.420	-0.172	0.071	1001	4000
eda3edu2	-0.456	0.162	0.005	-0.786	-0.452	-0.151	1001	4000
cons	-0.587	0.057	0.004	-0.698	-0.587	-0.473	1001	4000

Fuente: Elaboración propia

Las estimaciones de los parámetros en el modelo elegido fueron realizadas tanto con el software Winbugs como con programas propios elaborados por la autora. Los resultados son expuesto en el Cuadro 2.

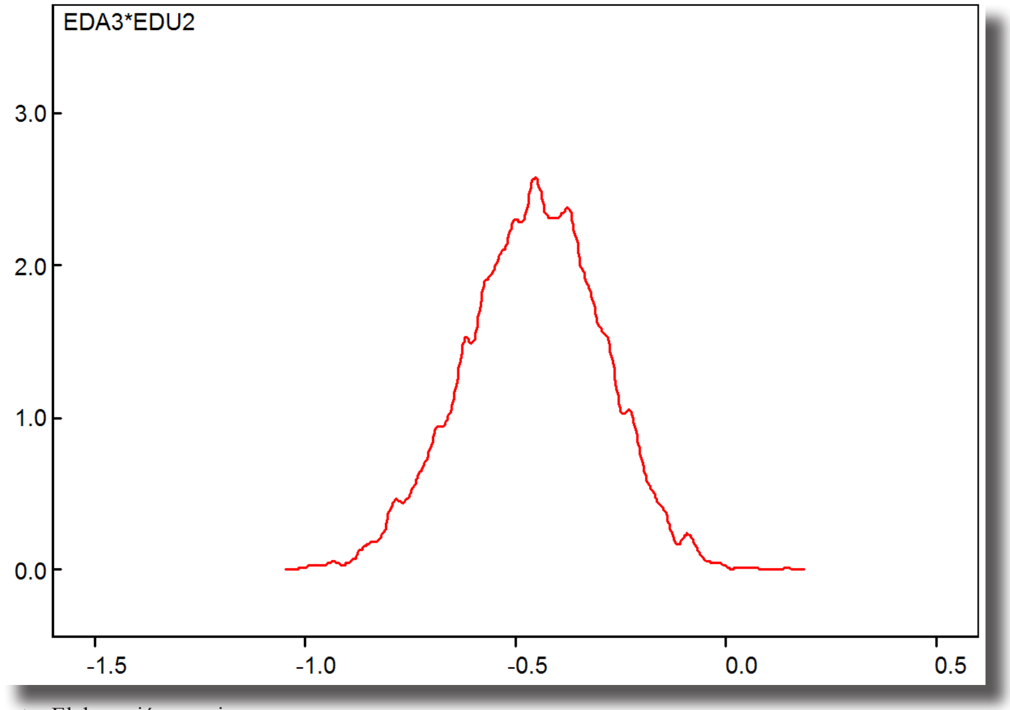
El Gráfico 1 muestra la convergencia del estimador para el coeficiente de la interacción entre el tercer grupo de edad (12-59 meses) y el segundo nivel de educación (educación alta), simbolizado por *EDA3*EDU2*, luego de eliminar ("burn") las primeras 1000 simulaciones. El valor estimado de tal coeficiente es -0.456. La distribución suavizada para el mismo estimador se muestra

Gráfico 1
Convergencia del estimador



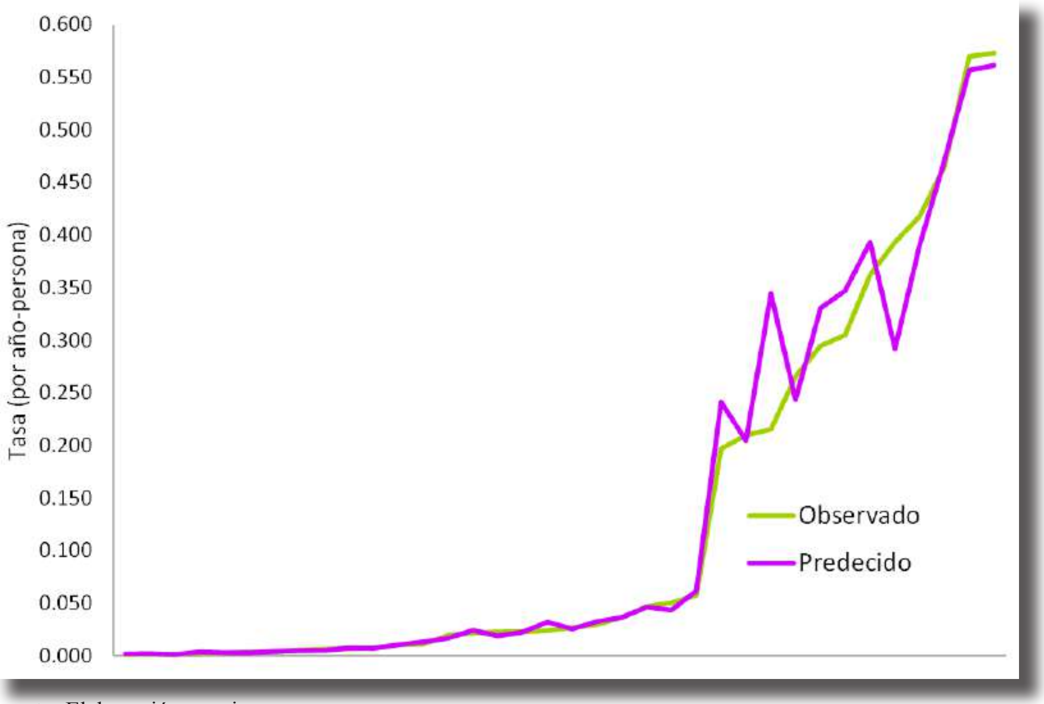
Fuente: Elaboración propia

Gráfico 2
Distribución suavizada para el mismo estimador



Fuente: Elaboración propia

Gráfico 3
Tasas de mortalidad observadas y tasas predecidas con el modelo elegido "EDA*COH+EDA+RESI



Fuente: Elaboración propia

en el Gráfico 2. Similares convergencias y distribuciones suavizadas se tienen para cada uno de los estimadores de parámetros.

Tasas de mortalidad observadas y predecidas con el modelo elegido son comparadas en el Gráfico 3. Si bien se aprecia algunas diferencias en niveles intermedios de mortalidad, el ajuste del modelo es casi perfecto en niveles extremos.

A diferencia del informe completo, en este documento se resaltan sólo dos de los resultados de la investigación. Por una parte, la educación tiene un efecto altamente significativo sobre la tasa de mortalidad en los primeros años de vida. En efecto, la tasa de mortalidad neonatal (MNN) en hijos de madres con educación alta es 38.1% menos que en hijos de madres con educación baja, controlado el efecto de cohorte de nacimiento y de área de residencia. Esta diferencia porcentual se incrementa a 47.8% en el caso de la mortalidad post-neonatal (MPNN) y a 60.4% en el caso de la mortalidad post-infantil (MPI).

En consecuencia, la educación tiene un alto impacto positivo en los tres tipos de mortalidad, principalmente en la tasa de mortalidad post-infantil (ver Cuadro 3).

Cuadro 3
Efecto de la educación sobre el nivel de mortalidad, por tipo de mortalidad

Tipo de mortalidad	Educación Baja	Educación Alta (%)*
MNN	-	-38.1 (0.0000)
MPNN	-	-47.8 (0.0000)
MPI	-	-60.4 (0.0000)

Fuente: Elaboración propia
*En paréntesis el valor-p del efecto

El segundo resultado tiene que ver con el efecto del área de residencia. Si bien su efecto es inferior en comparación con el efecto de la educación, también es altamente significativo. Esto es, la mortalidad en el área urbana es 29.9% menos que en el área rural, controlado el efecto de cohorte de nacimiento y de educación. Tal diferencia porcentual es la misma en cada uno de los tres tipos de mortalidad (MNN, MPNN y MPI), en cada nivel de educación y en cada cohorte de nacimientos (ver Cuadro 4).

Cuadro 4
Efecto de área de residencia sobre el nivel de mortalidad

Tipo de mortalidad	Educación Baja	Alta (%)*
MNN	-	-38.1 (0.0000)
MPNN	-	-47.8 (0.0000)
MPI	-	-60.4 (0.0000)

Fuente: Elaboración propia
*En paréntesis el valor-p del efecto

BIBLIOGRAFÍA

Box, G.E.P and Tiao, G. C.(1973), "*Bayesian Inference in Statistical Analisys*".Bretthorst,

G. Larry (1988), "*Bayesian Spectrum Analisys and Parameter Estimation*"

Lee, Peter (1997), "*Bayesian Statistics: An Introduction*".

Winkler, R. (2003) "*Introduction to Bayesian Inference and Decision*".