

MODELO DE COX MULTINIVEL ESTRATIFICADO CON INTERACCIÓN POR ÍNDICE DE RIQUEZA PARA ANÁLISIS DE RIESGO DE MUERTE DE NIÑOS MENORES DE CINCO AÑOS - ENDSA 2008**STRATIFIED MULTI-LEVEL COX MODEL WITH INTERACTION BY RICHNESS INDEX FOR RISK ANALYSIS OF DEATH OF CHILDREN UNDER FIVE YEARS – ENDSA 2008**

Pamela Córdova Olivera y Marina Yurevna Nikolaeva
Centro de Investigaciones Económicas y Empresariales (CIEE)
Universidad Privada Boliviana

pcordova@upb.edu

(Recibido el 23 noviembre 2019, aceptado para publicación el 21 diciembre 2019)

RESUMEN

Este estudio busca determinar en qué medida el estatus socioeconómico familiar, vinculado a un indicador de riqueza, afecta la probabilidad que los niños menores de cinco años tienen de sobrevivir, utilizando la información de la Encuesta Nacional de Demografía y Salud de Bolivia 2008 (ENDSAB-2008). La población objeto de estudio fueron los niños nacidos vivos, de mujeres entre 15 a 49 años, en los cinco años anteriores a la ENDSAB 2008. La variable dependiente en este estudio es la mortalidad de menores de cinco años (5q0), definida como el riesgo de morir entre el nacimiento y el quinto año de vida. Se estimó un modelo multinivel de Cox estratificado con interacción por las categorías de una variable de estrato denominada como "Índice de Riqueza". Se estimaron 3 modelos multinivel, uno por cada estrato (pobre, medio y rico). El modelo estimado para el estrato pobre del Índice de Riqueza logra explicar una proporción mayor de la varianza entre clústeres, más del 60%, determinado por la zona de residencia, características a nivel niño-madre, hogar y contextual-comunitario. Lo contrario ocurre en el modelo estimado para el estrato rico del Índice de riqueza ya que solo logra explicar el 31% de la varianza entre clústeres.

Palabras Clave: Mortalidad, Modelo de Cox Estratificado con Interacciones, Modelo Multinivel, Índice de Riqueza.

ABSTRACT

This study seeks to determine to what extent family socioeconomic status, linked to an indicator of wealth, affects the probability that children under five years old have to survive, using the information from the National Survey of Demography and Health of Bolivia 2008 (ENDSAB 2008). The population under study was children born alive, from women between 15 and 49 years old, in the five years prior to ENDSAB 2008. The dependent variable in this study is the mortality of children under five years old (5q0), defined as the risk of dying between birth and the fifth year of life. A stratified multilevel Cox model was estimated with interaction by the categories of a stratum variable called "Wealth Index". Three multilevel models were estimated, one for each stratum (poor, medium and rich). The estimated model for the poor stratum of the Wealth Index manages to explain a greater proportion of the variance between clusters, more than 60%, determined by the area of residence, characteristics at the child-mother, household and community-contextual level. The opposite occurs in the model estimated for the rich stratum of the wealth index since it only manages to explain 31% of the variance between clusters.

Keywords: Mortality, Stratified Cox Model with Interactions, Multilevel Model, Wealth Index.

1. INTRODUCCIÓN

Este artículo es parte de una línea de investigación que intenta establecer el efecto de los factores de riesgo individual, del hogar y contextuales de la comunidad sobre la salud relacionada con la mortalidad de niños menores de cinco años en Bolivia. En este intento en un estudio anterior [1], se analizaron los datos de Encuesta Nacional de Demografía y Salud de Bolivia (ENDSAB) 2008 utilizando el método no paramétrico de estimación de la función de supervivencia de Kaplan-Meier y se comparó el tiempo de supervivencia de niños menores de un año, menores entre uno y cuatro años y menores de cinco años. Se presentaron modelos multinivel de Cox y estratificados sin interacciones de Cox para el análisis de supervivencia con el fin de comparar el riesgo de muerte de niños menores de cinco años para diferentes covariables en comparación con las categorías de referencia y se determinó la varianza del riesgo de muerte entre clústeres que es explicada por factores a nivel niño-madre, hogar y contextual.

Este estudio también se identificaron variables que teóricamente debían tener efectos en la mortalidad de menores de cinco años, como el "Índice de Riqueza" según datos de la ENDSA 2008, que no fueron incluidas en las estimaciones de supervivencia desarrolladas por ser covariables que cambian en el tiempo, es decir, por no cumplir con el supuesto de riesgos proporcionales de Cox [1].

La presente investigación busca determinar en que medida el estatus socioeconómico familiar, vinculado a un indicador de riqueza, afecta la probabilidad que los niños menores de cinco años tienen de sobrevivir, utilizando la información de la Encuesta Nacional de Demografía y Salud de Bolivia 2008 (ENDSAB 2008). Para esto se emplearon técnicas estadísticas y econométricas adecuadas específicamente al tratamiento de este tipo de análisis.

El modelo de regresión simple de Cox se basa en que los riesgos son proporcionales, es decir, el efecto de una covariable determinada no cambia con el tiempo (las tasas de riesgo no cambian con el tiempo, los valores de una o de varias de las covariables no son diferentes en los distintos puntos temporales), es muy importante verificar que las covariables cumplan con el supuesto de proporcionalidad. Si se viola esta suposición, el modelo de regresión de Cox simple no es válido y se requieren análisis más complicados, como el modelo de regresión de Cox estratificado [2].

Therneau et al. (2000), destacan que el precio de omitir la heterogeneidad y dependencia del evento se paga, muchas veces, en estimaciones sesgadas e ineficientes de los efectos de las covariables, cuya naturaleza y alcance varían según los contextos. En el estudio de supervivencia, ese precio puede ser importante, tanto empírico como sustantivo. La supervivencia de la niñez menor de cinco años es un foco de preocupación y concentración de políticas económicas y sociales y como sus consecuencias afectan al desarrollo y la economía doméstica, la urgencia con la que presionamos nuestra comprensión de la problemática aumenta [3].

Si existe la sospecha de que las variables dependientes del tiempo aumentan o disminuyen el riesgo de muerte, entonces la estimación de un modelo simple de Cox producirá estimaciones sesgadas de los efectos de las covariables. Es necesario permitir este cambio en el riesgo de referencia y para ello necesitamos realizar una estratificación en función de una o más variables dependientes del tiempo consideradas como relevantes para este estudio.

Además, la teoría estadística establece que puede haber heterogeneidad entre los casos o, alternativamente, (1) homogeneidad compartida por (dentro) observaciones a lo largo del tiempo o (2) heterogeneidad compartida por todas las observaciones dentro de un clúster a lo largo del tiempo. Las muertes ocurren dentro de los hogares, que "pertenecen" a un clúster en particular. Los factores únicos de un individuo, pero omitidos en el modelo, que influyen en la ocurrencia de la muerte probablemente afecten la probabilidad de supervivencia posterior, creando correlaciones entre los eventos dentro de los clústeres. Cada observación está anidada dentro de un clúster y un modelo multinivel estratificado con dos niveles, donde las observaciones (niños menores de cinco años) del primer nivel están anidadas dentro de los clústeres considerados como el segundo nivel, es el proceso de estimación que seguimos en esta investigación con el fin de mejorar las estimaciones del riesgo de muerte al considerar la presencia de heterogeneidad y la dependencia mediante la estratificación para una variable clave en la economía.

Se utilizaron modelos multinivel estratificados de Cox con interacciones para el análisis del riesgo de muerte de niños menores de cinco años con datos de la ENDSA 2008, donde la variable de estrato es el "Índice de Riqueza" que no cumple con el supuesto de proporcionalidad de riesgos según datos de la ENDSAB 2008, con el fin de mostrar como se debe proceder en presencia de variables dependientes del tiempo, evitando así problemas de sesgo por omisión de variable relevante, y heterogeneidad.

2. METODOLOGÍA

2.1 Modelo Multinivel

Griffiths et al. (2014) [4], establece que el análisis multinivel es una respuesta a la necesidad de analizar la relación entre los individuos y los diversos contextos en los que se desenvuelven. Las hipótesis de partida de estos modelos establecen que los individuos pertenecientes a un mismo contexto tenderán a ser más similares en su comportamiento entre sí, que respecto a su pertenencia a distintos contextos.

El objetivo fundamental del análisis multinivel, es modelar estadísticamente la influencia de variables contextuales sobre las actitudes o los comportamientos medidos a nivel individual. Separar los efectos composicionales (características a nivel del niño-madre) de los contextuales para explicar las causas de la mortalidad de la niñez menor de cinco años no es una tarea fácil. Si bien puede ser posible abordarlos con metodologías tradicionales como la regresión lineal múltiple o la regresión logística, se presentan algunos problemas en los supuestos básicos para el ajuste de los modelos, lo que hace que se pierda explicación útil proveniente del contexto del individuo, ya que con estos modelos se hacen análisis sin tener en cuenta la jerarquía en los datos y se analizan las variables de interés como si todas perteneciesen solo al nivel composicional [5].

Modelos como los mencionados estiman la variación dentro de los contextos sociales en una población y entre ellos mismos; tienen la capacidad de evaluar las pequeñas fluctuaciones que los métodos tradicionales no alcanzan a detectar

y, más específicamente, de introducir las características de los individuos en un modelo que relacione las variables del ambiente cercano con las características del contexto social que rodea al individuo, evitando fenómenos como la falacia ecológica (consistiría, fundamentalmente, en atribuir a los miembros de un agregado estadístico las propiedades de tal agregado) y la falacia individualista (inferir incorrectamente a partir de unidades de nivel inferior la condición de sistemas de más alto orden, se trata muy raramente en la literatura metodológica), en los que se presenten relaciones espurias por no tener en cuenta la variabilidad de los contextos. Con los modelos tradicionales pueden ser adaptados a distintas propiedades estadísticas, como son el de varianza constante e independencia entre las unidades de estudio. En individuos agrupados en un contexto, la variabilidad es distinta entre diferentes grupos [6].

Desde un punto de vista puramente estadístico, el análisis multinivel, también denominado análisis jerárquico, permite resolver la limitación del uso de modelos de regresión múltiple que invalidan la hipótesis de independencia cuando se presenta mayor homogeneidad entre individuos de un mismo grupo respecto a individuos de distintos grupos. Esta similitud entre los individuos dentro de los grupos establece una estructura de correlación intracontextual (intracluster) que impide el cumplimiento de la hipótesis de independencia sobre la que están basados los modelos de regresión tradicionales e invalida, por tanto, sus métodos de estimación, lo que se traduce en estimaciones incorrectas de los errores estándar. Según Sandoval (2004) [6] los modelos multinivel resuelven dos problemas que se presentan cuando se usan análisis de un único nivel a datos que son jerárquico: 1º.- Problemas estadísticos de correlación entre los individuos en la estimación de los mínimos cuadrados ordinarios ineficientes y con significaciones espurias. 2º.- Problemas conceptuales ya que se emplea el nivel equivocado (analizar los datos a un nivel y extraer conclusiones a otro). Desde esta segunda problemática nos podemos encontrar con dos tipos de falacia, a) falacia ecológica (interpretar datos agregados a nivel individual) y la falacia atomística (interpretación agregada a partir de datos individuales). Además resuelven otros problemas tales como fijar el efecto directo de las variables explicativas individuales de grupo, determinar si las variables de grupo “moderan” las relaciones a nivel individual (interacciones entre niveles) y establecer qué porcentaje de variabilidad de la variable explicada o dependiente, una vez controlada por las variables explicativas, es imputable al individuo y qué porcentaje es imputable al grupo. El análisis multinivel tiene por lo tanto como objetivo, modelizar estadísticamente la influencia de variables contextuales sobre las actitudes o los comportamientos medidos a nivel individual. En este sentido, nos permite, en ciencias sociales, tener en cuenta el efecto de las variables de la estructura social y económica sobre el individuo [7].

Un número muy escaso de investigaciones han empleado el análisis multinivel para identificar correlaciones de mortalidad de la niñez menor de cinco años. El supuesto es que "los individuos (nivel 1) están anidados dentro de los hogares (nivel 2) y los hogares están anidados dentro de las comunidades (nivel 3)" [8]. Hasta la fecha, con la excepción de uno o dos estudios, los estudios bolivianos sobre la mortalidad de menores de cinco años han utilizado rara vez el enfoque de modelos multinivel. El enfoque multinivel es bueno para identificar los amplios contextos sociales, económicos y ambientales en los que el niño vive y experimenta un determinado resultado de salud [3]. Esto sugiere que los individuos con similares características del hogar pueden tener diferentes resultados de salud cuando residen en diferentes comunidades con características diferentes. Griffith et al (2014), opinaron que sería metodológicamente incorrecto ajustar un modelo de regresión estándar de un solo nivel en el análisis de la supervivencia de menores de cinco años. Esto se debe a que los modelos de regresión estándar no pueden manejar la estructura jerárquica en los conjuntos de datos debido a su suposición de independencia. En este estudio, por lo tanto, se llevó a cabo un análisis de regresión de riesgos proporcionales de múltiples niveles de Cox para determinar hasta qué punto los determinantes contextuales explican la mortalidad de menores de cinco años en Bolivia. El modelo de riesgos proporcionales de Cox (es decir, el análisis de supervivencia) es apropiado para analizar observaciones censuradas. Esto significa que, utilizando el análisis de regresión de riesgos proporcionales de Cox, tanto la ocurrencia de mortalidad de menores de cinco años, menores de un año y menores entre uno a cuatro años, fueron consideradas como las variables de resultado. Además, el análisis del modelo de riesgos puede hacerse para cuidar la estructura multinivel o jerárquica de los datos de la ENDSAB (es decir, la interdependencia de los determinantes, porque los individuos (nivel 1) están anidados dentro de las comunidades (nivel 2). La modelación multinivel proporciona un mejor análisis que producirá resultados más sólidos sobre los factores asociados con la mortalidad de la niñez menor de cinco años.

Se hace referencia a modelos jerárquicos donde las unidades se clasifican por algún factor en clusters (grupos) de nivel superior. Las unidades en este nivel superior son subclasificadas por un factor adicional en un cluster de nivel más bajo y así sucesivamente.

Para relajar la suposición de independencia condicional entre las respuestas para el mismo niño analizado dadas las covariables, podemos incluir un intercepto aleatorio específico del niño ζ_j en el predictor lineal para obtener un modelo de regresión logística de intercepto aleatorio. El análisis de regresión multinivel de riesgos proporcionales de Cox, es la extensión de una regresión logística de intercepto aleatorio, donde varía los niveles de clusters. Por efectos prácticos, se determinará como primer nivel i (niños) y j (comunidad).

Para empezar, una regresión logística de dos niveles puede escribirse como [9]:

$$\text{logit}\{\Pr(y_{ij} = 1|x_{ij}, \zeta_j)\} = \beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij} + \zeta_j \quad (1)$$

En donde ζ_j es un intercepto aleatorio normal e independiente distribuido $(0, \psi)$, entre los individuos i e independientes de las covariables $x_{ij} = (x_{2j}, x_{3ij})'$ es un vector que contiene a dos variables. Dado ζ_j y x_{ij} , las respuestas y_{ij} se distribuyen de forma independiente Bernoulli, $\pi_{ij} \equiv \Pr(y_{ij}|\zeta_j, x_{ij})$, dando:

$$\text{logit}(\pi_{ij}) = \beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij} + \zeta_j$$

$$y_{ij}|\pi_{ij} \sim \text{Binomial}(1, \pi_{ij})$$

Este es un ejemplo simple de un modelo mixto lineal generalizado (GLMM) porque es un modelo lineal generalizado con ambos efectos fijos β_1 a β_4 y un efecto aleatorio ζ_j . El modelo también se conoce a veces como un modelo lineal generalizado jerárquico (HGLM) en contraste con un modelo lineal jerárquico (HLM). La interceptación aleatoria puede considerarse como el efecto combinado de las covariables omitidas específicas del niño (tiempo constante) que causan que algunos niños sean más propensos a la morir que otros (más precisamente, el componente de este efecto combinado que es independiente de las covariables no es un problema si las covariables son exógenas). Es atractivo modelar esta heterogeneidad no observada de la misma forma que la heterogeneidad observada simplemente agregando la interceptación aleatoria al predictor lineal. Hay que tener en cuenta que los odds ratios obtenidos al exponenciar los coeficientes de regresión en este modelo deben interpretarse condicionalmente en la interceptación aleatoria y, por lo tanto, a menudo se los denomina odds ratios condicional o específicos del sujeto.

Usando la formulación de respuesta latente, el modelo puede escribirse de manera equivalente

$$y^*_{ij} = \beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij} + \zeta_j + \epsilon_{ij} \quad (2)$$

ϵ_{ij} sigue una distribución logística estándar. Las respuestas binarias de y_{ij} están determinadas por las respuestas continuas latentes a través del modelo de umbral:

$$y_{ij} = \begin{cases} 1 & \text{si } y^*_{ij} > 0 \\ 0 & \text{otros} \end{cases}$$

En ambas formulaciones del modelo (a través de un enlace logit o en términos de una respuesta latente), se supone que los ζ_j son independientes entre los individuos e independientes de las covariables x_{ij} en la i . También se asume que las covariables no afectan las probabilidades de respuesta dada la interceptación aleatoria (llamada exogeneidad estricta condicional de la interceptación aleatoria). Para la formulación de respuesta latente, se supone que el ϵ_{ij} es independiente en ambas ocasiones y en los niños, e independiente de ambos, j y x_{ij} . En la formulación del modelo lineal generalizado, las suposiciones análogas están implícitas al suponer que las respuestas están distribuidas de forma independiente por Bernoulli (con probabilidades determinadas por ζ_j y x_{ij}).

En contraste con los modelos de efectos aleatorios lineales, la estimación consistente en la regresión logística de efectos aleatorios requiere que la parte aleatoria del modelo se especifique correctamente además de la parte fija. Específicamente, la consistencia requiere formalmente (1) un predictor lineal correcto (como incluir interacciones relevantes), (2) una función de enlace correcta, (3) una correcta especificación de covariables que tienen coeficientes aleatorios, (4) independencia condicional de las respuestas dados los efectos aleatorios y covariables, (5) independencia de los efectos aleatorios y covariables (para la inferencia causal), y (6) efectos aleatorios normalmente distribuidos. Por lo tanto, los supuestos son más fuertes que los discutidos para los modelos lineales. Al igual que en la regresión logística estándar, el estimador ML no es necesariamente imparcial en muestras finitas, incluso si todas las suposiciones son ciertas como [9]

Rabe-Hesketh *et al.* [9] y otros escriben modelos de dos niveles en términos de un modelo de nivel 1 y uno o más modelos de nivel 2. En los modelos mixtos lineales generalizados, la necesidad de especificar una función de enlace y distribución conduce a dos etapas adicionales de la especificación del modelo. Utilizando la notación y la terminología de citado en Rabe-Hesketh *et al.*, 2012 [9], el modelo de muestreo de nivel 1, la función de enlace y el modelo estructural se escriben como:

$$y_{ij} \sim \text{Bernoulli}(\varphi_{ij})$$

$$\text{logit}(\varphi_{ij}) = \eta_{ij}$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j}x_{2j} + \beta_{2j}x_{3ij} + \beta_{3j}x_{2j}x_{3ij} \quad (3)$$

respectivamente. El modelo de nivel 2 para la intersección β_{0j} se escribe como:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

donde γ_{00} es un intercepto fijo y u_{0j} es un intercepto residual o aleatorio. Los modelos de nivel 2 para los coeficientes β_{1j} , β_{2j} y β_{3j} no tienen residuos para un modelo de intersección aleatoria,

$$\beta_{pj} = \gamma_{p0} \quad p = 1,2,3 \quad (4)$$

Al conectar los modelos de nivel 2 al modelo estructural de nivel 1, obtenemos:

$$\begin{aligned} \eta_{ij} &= \gamma_{00} + u_{0j} + \gamma_{01}x_{2j} + \gamma_{02}x_{3ij} + \gamma_{03}x_{2j}x_{3ij} \\ &\equiv \beta_1 + \zeta_{0j} + \beta_2x_{2j} + \beta_3x_{3ij} + \beta_4x_{2j}x_{3ij} \end{aligned} \quad (5)$$

La extensión del modelo a tres niveles en donde los niños i están anidado en madres j y al mismo tiempo ellas están anidadas en comunidades k , se obtiene lo siguiente:

$$\begin{aligned} \text{logit}\{\Pr(y_{ijk} = 1|x_{ijk}, \zeta_{jk}^{(2)}, \zeta_k^{(3)})\} &= \beta_1 + \beta_2x_{2ijk} + \dots + \beta_nx_{n,k} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} \\ &= (\beta_1 + \zeta_{jk}^{(2)} + \zeta_k^{(3)}) + \beta_2x_{2ijk} + \dots + \beta_nx_{n,k} \end{aligned} \quad (6)$$

En donde $x_{ijk} = (x_{2ijk}, \dots, x_{n,k})$ es un vector de covariables, $\zeta_{jk}^{(2)} \sim N(0, \psi^{(2)})$ es un intercepto aleatorio que varía de acuerdo al nivel 2 (madres), y $\zeta_k^{(3)} \sim N(0, \psi^{(3)})$ varía sobre el nivel 3 (comunidades). Los interceptos $\zeta_{jk}^{(2)}$ y $\zeta_k^{(3)}$ son independientes uno de otro, de su respectivo nivel y de las covariables x_{ijk} . Las respuestas para y^*_{ijk} son independientemente distribuidas dado los efectos aleatorios y las covariables.

Este modelo puede escribirse como:

$$y^*_{ijk} = \beta_1 + \beta_2x_{2ijk} + \dots + \beta_nx_{n,k} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} + \epsilon_{ij} \quad (7)$$

donde ϵ_{ij} sigue una distribución logística estándar.

Dos indicadores son determinantes para el análisis de los modelos multinivel: i) Coeficiente de Correlación Intraclase CCI, (The intraclass coefficient ICC), entendido como la proporción de la varianza total explicada por diferencias entre las macro unidades y; ii) Coeficiente de Partición de la Varianza CPV, (Variance Partition Coefficient VPC), que mide el porcentaje cambio de varianza explicada en cada modelo especificado.

El CCI es una proporción la cual es resultado de un cociente de varianzas en el cual su denominador es la varianza total, se puede entender como el porcentaje de varianza que se desea explicar en determinado nivel respecto al porcentaje de varianza total. Este coeficiente se denota como " ρ " y su valor se encuentra entre cero y uno. El grado de semejanza entre las micro-unidades que pertenecen a la misma macro-unidad esta puede ser expresada por el coeficiente de correlación intraclase [10]. Si el modelo fuera de dos niveles entonces el CCI sería calculado de la siguiente manera:

$$\rho = \frac{\psi^{(2)}}{\psi^{(2)} + \pi^2/3} = \frac{\text{Varianza poblacional entre las macro unidades}}{\text{Varianza total}}$$

$\psi^{(2)}$: es la varianza entre los grupos (es decir la varianza entre las macro-unidades)

$\pi^2/3$: es la varianza dentro de los grupos (varianza residual-es decir la varianza de los sujetos dentro de la macro-unidad)

Por lo tanto la varianza total sería la suma de estos dos componentes:

$$\text{Var}(y_{ij}) = \psi^{(2)} + \pi^2/3$$

En el caso de un análisis multinivel de dos niveles el coeficiente de correlación intraclase representa el porcentaje de varianza explicada por el nivel 2.

2.2 Medidas de calidad para el ajuste global de los modelos multinivel

Se presentan los estadísticos de ajuste global cuando se desea cuantificar en qué medida un modelo representa la variabilidad de los datos (variaciones observadas en los datos), sin embargo, estos indicadores no son directamente interpretables, su importancia radica en la comparación de los modelos.

Se plantean las siguientes hipótesis:

$$\begin{aligned} H_0: & \text{El modelo ajustado es significativo} \\ H_1: & \text{El modelo ajustado no es significativo} \end{aligned}$$

Los estadísticos de ajuste global son [10]:

2.2.1 Criterio de Información de Akaike (AIC)

$$AIC = -2\ln L + 2d \quad (8)$$

El modelo con AIC mínimo, en un conjunto de modelos anidados, será el más parsimonioso de acuerdo con este criterio. Donde d representa el número de parámetros estimados.

2.2.2 Criterio de Información Bayesiano BIC (Schwarz)

$$BIC = -2\ln L + d\log(n) \quad (9)$$

A menor valor de este criterio indicaría el modelo más parsimonioso entre los modelos comparados. Por lo general n indica el número de unidades en el nivel más alto de la jerarquía.

Al comparar modelos alternativos esperamos que estos estadísticos sean lo menor posible. Detallamos sus componentes de la siguiente manera:

$$\ln L: \text{Logaritmo de la verosimilitud} = \ln \left(\frac{\text{verosimilitud del modelo ajustado}}{\text{verosimilitud del modelo saturado}} \right)$$

$$-2\ln L: \text{Desviación (D)} = -2\ln \left(\frac{\text{verosimilitud del modelo ajustado}}{\text{verosimilitud del modelo saturado}} \right)$$

(D) es la medida del grado de diferencia entre las frecuencias predichas y las observadas del modelo, el mejor modelo será aquel que tenga menor varianza, este estadístico sigue una distribución Chi – Cuadrada con $n - (k+1)$ grados de libertad, donde k es el número de variables [9].

2.3 Modelos de riesgos proporcionales de Cox

El modelo de riesgo proporcional es el modelo más utilizado para representar los efectos de un conjunto de variables explicativas sobre la variable tiempo de cambio (tiempo de supervivencia), o más bien sobre la probabilidad condicional de cambio, es decir sobre la función de riesgo $h(t)$ [5]. Suponemos que para cada sujeto tenemos un vector X de variables explicativas, concomitantes o pronóstico.

El modelo de Cox expresa la función de riesgo en función del tiempo y de un conjunto de covariables que definen al sujeto en estudio del siguiente modo:

$$h(t|x_j) = h_0(t) \exp \left(\sum_{j=1}^p \beta_j x_j \right) \quad (10)$$

donde los coeficientes de la regresión β_j son estimados de los datos.

El modelo de Cox asume que las covariables cambian multiplicativamente al riesgo de base (también se interpreta que la función de riesgo basal sería aquella función “básica” del modelo si éste no incorporara predictores) de la función de riesgo y no se asume que el riesgo tiene una forma en particular a través del tiempo. Este afirma que el ratio de riesgo para el j -ésimo sujeto en los datos es (10).

El riesgo de un sujeto es una réplica multiplicativa de otro sujeto, comparando a los sujetos j y m se obtiene:

$$HR = \frac{h(t, x^*)}{h(t, x)} \quad (11)$$

Siendo constante, asumiendo que las covariables x y x^* no cambian en el tiempo. La función de riesgo basal o de base es la única parte de la expresión del modelo de Cox que depende del tiempo t . La otra parte solo depende del vector de covariables de los sujetos que se suponen independientes del tiempo.

Una variable independiente del tiempo se define como una variable cuyos valores no varían a lo largo del tiempo. Por ejemplo el sexo, la raza o el grupo de tratamiento son variables fijas, sólo toman un valor, el inicial. Por ejemplo la edad y el peso de los sujetos sí varían con el tiempo, pero puede ser apropiado tratarlas como independientes del tiempo en análisis determinados. Esto es posible siempre que los valores de estos predictores no varíen en exceso a lo largo del tiempo, o bien si el efecto de dichas variables en el riesgo de supervivencia depende esencialmente de un único valor de medición.

Existe la posibilidad de considerar predictores dependientes del tiempo t . En tal caso es posible utilizar la modelización de Cox pero no se suele satisfacer la condición de “riesgos proporcionales” que más abajo definimos. En esta situación en que se tienen en cuenta predictores que dependen del tiempo la regresión se denomina “modelo de Cox Estratificados” [9].

2.4 Hipótesis de riesgos proporcionales

En el modelo de Cox se busca como primer paso la relación entre los riesgos de muerte de dos individuos expuestos a factores de riesgo diferentes. Para ello, el modelo parte de una hipótesis fundamental, la de que los riesgos son proporcionales. Para comprender esta noción definiremos previamente la denominada “razón de riesgos” (Hazard Ratio, HR) entre dos sujetos con diferente vector de covariables $x = (x_1, \dots, x_p)$ y $x^* = (x_1^*, \dots, x_p^*)$ como [9]:

$$HR = \frac{h(t, x^*)}{h(t, x)} = \exp \left(\sum_{j=1}^p \beta_j (x_j^* - x_j) \right) \quad (12)$$

Se observa que el resultado de la razón de riesgo (12) no depende de la función de riesgo basal, tan sólo del valor de los predictores y de las betas estimadas, es decir, no depende del tiempo. Por lo tanto, en el modelo de Cox se supone la hipótesis de que los riesgos son proporcionales, ya que se suponen covariables no dependientes del tiempo. La hipótesis de riesgos proporcionales significa explícitamente que la razón de riesgo (12) es constante en el tiempo: $h(t, x^*) = constante * h(t, x)$. Si lo aplicamos a la expresión resultante en el modelo de Cox (12), denominando θ a la constante y una vez estimados los coeficientes de la regresión por máxima verosimilitud parcial, tenemos que la razón de proporcionalidad es constante en el tiempo e igual a:

$$\theta = \exp \left(\sum_{j=1}^p \beta_j (x_j^* - x_j) \right)$$

A continuación presentamos los principales mecanismos para testeo del supuesto de riesgos proporcionales [9]:

- **Métodos gráficos:** El más popular es el uso de los gráficos denominados “loglog”. Si se cumple la hipótesis de riesgos proporcionales estos gráficos deben ser visualmente paralelos para las diferentes clases de las covariables del modelo por separado. Otro método es el de comparar las curvas de supervivencia “observada” y “esperada” para las diferentes clases de las covariables por separado. Éstas curvas deberían ser parecidas visualmente.
- **Estudio de la bondad de ajuste (goodness-of-fit, GOF) de la hipótesis de riesgos proporcionales mediante un test estadístico para cada variable del modelo:** Mediante el test se obtienen p-valores con los que decidir si se cumple o no la hipótesis de riesgos proporcionales. De este modo tenemos una medida menos subjetiva que con el método gráfico y menos complicada computacionalmente que con la inclusión de variables dependientes del tiempo. Sin embargo, el método GOF es demasiado global como para detectar algunas desviaciones de la hipótesis de riesgos proporcionales que el método gráfico y el de inclusión de variables dependientes del tiempo sí pueden detectar y que pueden ser de ayuda en las conclusiones e interpretaciones de la hipótesis.
- **Estudio mediante el uso de variables dependientes del tiempo:** Este camino implica utilizar el modelo de Cox ampliado con los datos introducidos en formato “counting process”. Concentra su análisis en el supuesto

de riesgos proporcionales utilizando la significancia de la interacción entre la variable de tiempo y la covariables. La hipótesis nula es que la interacción entre covariable y el tiempo de análisis no es significativa. Se agrega la interacción de una covariable con función de variable de tiempo: la nueva variable agregada no debe ser estadísticamente significativa.

2.5 Modelo de Cox estratificado

El modelo Multinivel de Cox Estratificado es una modificación del modelo de Cox de riesgos proporcionales que permite incluir un control mediante una estratificación de un predictor que no cumple la hipótesis de riesgos proporcionales. Los predictores que se supone que satisfacen la suposición de riesgos proporcionales se incluyen en el modelo, mientras que el predictor que se está estratificando no está incluido (variable de estrato). Existe la posibilidad de estratificar por un único predictor o por varios a la vez [11].

Los estratos son las diferentes categorizaciones de la variable de estrato. Si la función de riesgo de línea de base es diferente para cada uno de los estratos y los coeficientes son los mismos estamos ante un modelo estratificado sin interacciones, caso contrario cuando los coeficientes son diferentes entre los estratos estamos ante un modelo estratificado con interacciones.

2.6 Modelo de Cox Estratificado sin interacciones

Suponemos que tenemos k variables que no satisfacen el supuesto de proporcionalidad de riesgos y P variables que satisfacen el supuesto. Las variables que no satisfacen el supuesto de se denotan con de Z_1, Z_2, \dots, Z_k ; Las variables que satisfacen el supuesto que dan determinadas como x_1, x_2, \dots, x_P .

Para realizar el procedimiento de Cox estratificado, definimos una variable que no cumple con el supuesto de proporcionalidad de riesgos que llamamos Z^* .

Los estratos son las categorías de la variable de estrato, k^* . La función de riesgo para el modelo de Cox estratificado sin interacciones esta denotado de la siguiente manera:

$$h_g(t|x_j) = h_{0g}(t) \exp(\sum_{j=1}^P \beta_j x_j) \tag{13}$$

donde g denota el estrato #, $g=1,2,\dots,k^*$, entonces las funciones de riesgo para cada estrato quedan especificadas como:

$$\begin{aligned} h_1(t|x_j) &= h_{01}(t) \exp(\sum_{j=1}^P \beta_j x_j) \\ h_2(t|x_j) &= h_{02}(t) \exp(\sum_{j=1}^P \beta_j x_j) \\ h_3(t|x_j) &= h_{03}(t) \exp(\sum_{j=1}^P \beta_j x_j) \\ &\vdots \\ h_{k^*}(t|x_j) &= h_{0k^*}(t) \exp(\sum_{j=1}^P \beta_j x_j) \end{aligned}$$

Sin embargo, debido a que la variable de estrato no está incluida en el modelo, no es posible obtener un valor de relación de riesgo para el efecto de la misma ajustado para las otras variables que si cumplen con el supuesto de riesgos proporcionales. Este es el precio a pagar por la estratificación de la variable de estrato. Un solo valor para la relación de riesgos de la variable de estrato no es apropiado si la variable de estrato no cumple con la suposición de riesgos proporcionales, porque la relación de riesgo debe variar con el tiempo [11].

Las funciones de riesgo por estrato solo difieren en la medida en que tienen un riesgo basal diferente, a saber, $h_{01}(t), h_{02}(t), \dots, h_{0k^*}(t)$. Sin embargo, los coeficientes β_j son los mismos para los modelos por estrato.

Debido a que existen diferentes funciones de riesgo basal (tantas como estratos de la variable de estrato), el modelo de Cox estratificado ajustado dará diferentes curvas de supervivencia estimadas.

Ya que los coeficientes de x_j son los mismos para cada estrato, las estimaciones de las razones de riesgo, para el $\hat{\beta}_j$, son las mismas para cada estrato de la variable estrato. Esta característica del modelo estratificado de Cox se denomina suposición de "no interacción". Es posible evaluar si esta suposición es sostenible y modificar el análisis si no lo es.

Para obtener estimaciones de β_j , se forma una función de verosimilitud parcial (L) a partir del modelo y los datos. La función de verosimilitud (L) para el modelo estratificado de Cox es diferente del modelo de Cox no estratificado. Para

el modelo estratificado, L se obtiene al multiplicar las funciones de probabilidad para cada estrato. Por lo tanto, L es igual al producto de las L_g , que denotan las funciones de verosimilitud de cada estrato, respectivamente, que se derivan de sus respectivas funciones de riesgo $h_g(t|x_j)$.

2.7 Modelo de Cox Estratificado con interacción mediante ajuste de modelos separados

Si la interacción es posible, entonces se obtienen diferentes coeficientes β_j para cada uno de los estratos (g). Es decir, se ajustan modelos de riesgo separados para cada estrato.

Una forma de especificar la función de riesgo cuando hay interacción es:

$$h_g(t|x_j) = h_{0g}(t) \exp(\sum_{j=1}^p \beta_{jg} x_j) \quad (14)$$

Observe que cada variable x_j en este modelo tiene un coeficiente diferente para cada estrato, como lo indican los subíndices g de los coeficientes, β_{jg} .

La prueba es una prueba de razón de verosimilitud (LR) basada en la relación de las probabilidades de los modelos sin interacción y con interacción. Es decir, el estadístico de prueba LR (logaritmo de la verosimilitud-probabilidad logarítmica) es de la forma $-2 \ln L_R$ menos $-2 \ln L_F$, donde R denota el modelo reducido, sin interacción, que en este caso es el modelo sin interacción, y F denota el modelo completo, que es el modelo con interacción. El estadístico de prueba LR tiene aproximadamente una distribución de chi-cuadrado con $(k^* - 1)$ grados de libertad bajo la hipótesis nula de que el modelo sin interacción es correcto ($LR \sim \chi_{p(k^*-1)}^2$).

Este estudio empleó una metodología combinada propia de un Modelo Multinivel de Cox y Estratificación con Interacciones por un Índice de Riqueza (pobre, medio y rico).

3. MODELO DE COX ESTRATIFICADO POR “ÍNDICE DE RIQUEZA”

La variable utilizada como variable de estrato fue el “Índice de Riqueza”, variable a nivel del hogar que resultó no cumplir con el supuesto de riesgos proporcionales en el modelo de riesgos proporcionales de Cox para datos de la ENDSA 2008 y para el caso de menores de cinco años.

El índice de riqueza se basa en datos de la encuesta de hogares de la ENDSA 2008. La variable de estrato es el “Índice de Riqueza” que es una medida compuesta del nivel de vida acumulativo de un hogar, se calcula utilizando datos fáciles de recopilar sobre fuente de agua potable, tipo de inodoro, uso compartido de baños, material del piso principal, paredes, techo, combustible para cocinar, servicios domésticos y posesiones, tales como electricidad, televisión, radio, reloj, tipos de vehículos, agricultura tamaño de la tierra, propiedad, tipo y número de animales, cuenta bancaria, tipos de ventanas. Los ítems varían algo entre las encuestas y las categorías de los ítems del cuestionario principal. Todos estos elementos se utilizarán para distinguir un hogar de otro [12]. Este índice es generado con un procedimiento estadístico conocido como análisis de componentes principales, el índice de riqueza coloca a los hogares individuales en una escala continua de riqueza relativa. Separa a todos los hogares entrevistados en grupos de riqueza (pobre, medio, rico) para comparar la influencia de la riqueza en varios indicadores de población, salud y nutrición.

Los estratos son las diferentes categorizaciones de la variable de estrato “Índice de Riqueza”. La variable está categorizada en: i) Pobre; ii) Medio y iii) Rico. La variable de estrato no se incluye implícitamente en el modelo, mientras que las que sí cumplen con el supuesto de riesgos proporcionales se incluyen en el modelo como la zona de residencia, intervalo de nacimiento, lugar de parto, edad de la madre al primer nacimiento, afiliación étnica, educación de la madre, nivel de educación de la mujer en la comunidad, parto en el hospital de la comunidad, distancia al centro de salud, diversidad étnica, atención prenatal en la comunidad y edad actual (al momento de la encuesta) de la madre.

La Figura 1 muestra las curvas de supervivencia estimadas para cada una de las categorías de la covariables de Índice de Riqueza para datos de la ENDSA 2008. Observamos que estas no cumplen con el supuesto de riesgos proporcionales en 2008 para niños menores de cinco años.

La Tabla 1, muestra los resultados del test de log Rank para la variable de estrato, se observa que la supervivencia de la niñez menor de cinco años es diferente para cada uno de los estratos del Índice de Riqueza (rechazamos la hipótesis nula de que la función supervivencia para cada estrato sea la misma).

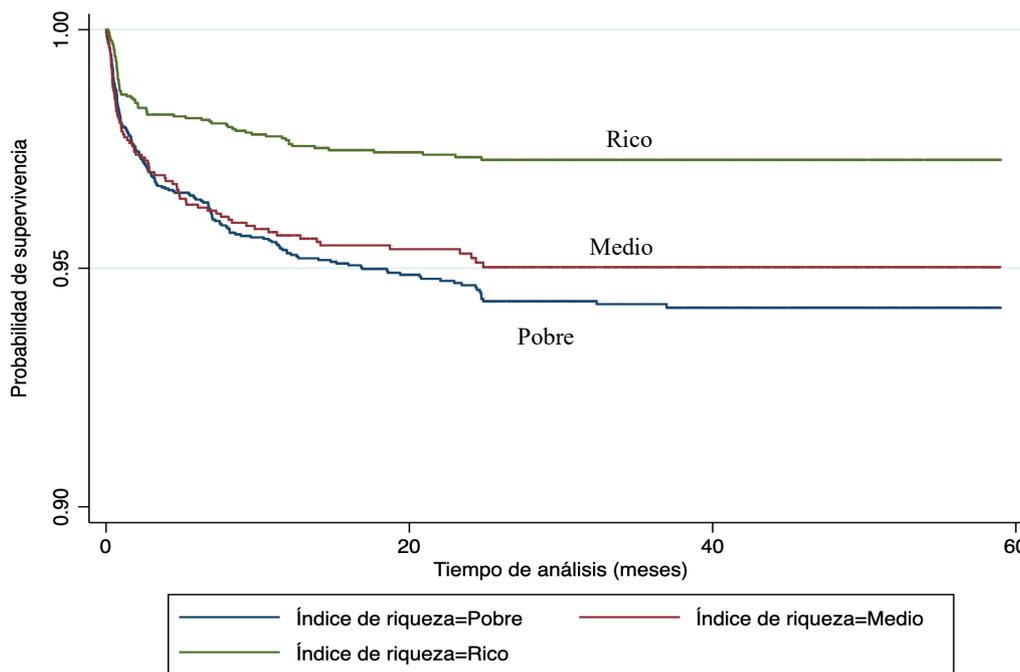


Figura 1: Curvas de supervivencia de menores de cinco años según datos de las ENDSA 2008, por estrato de Índice de Riqueza.

TABLA 1 - TEST DE LOG RANK VARIABLE DE ESTRATO ÍNDICE DE RIQUEZA

ENDSA	EVENTOS OBSERVADOS	EVENTOS ESPERADOS
POBRE	239	174,54
MEDIO	55	71,88
RICO	58	105,58
Total	352	352,00

chi2(1)= 49,21
Pr>chi2= 0,000

Fuente: Elaboración propia ENDSAB 2008.

En la Tabla 2, columna 3, se observa que el Índice de Riqueza es una variable dependiente del tiempo¹ y, por lo tanto, una variable de estrato que justifica la estimación de un modelo de Cox estratificado (en la Figura 1 se ve que hay momentos en el tiempo que el ratio de riesgos de muerte de los menores de cinco años no es constante). La segunda columna informa los resultados (cocientes de riesgo) para las categorías de la covariable que no varían con el tiempo, la tercera columna informa los resultados de la interacción entre la covariable Índice de Riqueza y tiempo de análisis.

TABLA 2 - ANÁLISIS DEL SUPUESTO DE RIESGOS PROPORCIONALES

VARIABLE QUE VARIA CON EL TIEMPO	RESULTADOS SIN DEPENDENCIA DEL TIEMPO (COEFICIENTES)	RESULTADOS CON DEPENDENCIA DEL TIEMPO
Índice de Riqueza		
Medio	-0,58***	-0,016**
Rico	-0,91***	
Rechazamos la hipótesis nula de que la interacción entre covariable Índice de Riqueza y el tiempo de análisis no es significativa		

Fuente: Elaboración propia ENDSAB 2008.

Una vez que se demostró que la covariable Índice de Riqueza es una variable candidata a variable de estrato se procedió a la estimación de un modelo estratificado multinivel de Cox. Inicialmente se estimó un modelo estratificado sin

¹ Se especificó la variable que varía continuamente con respecto al tiempo, es decir, la covariable Índice de Riqueza. La hipótesis nula es que la interacción entre la variable de tiempo de análisis y el Índice de Riqueza es no es significativa.

interacciones con un $-2 \ln L_R = 574,456$ y posterior a este ejercicio se procedió a la estimación de un modelo estratificado con interacciones con un $-2 \ln L_F = 508,656$, lo que generó una diferencia de 65,80 y un valor de chi-cuadrado crítico en tablas de 36,42 con 24 grados de libertad que permitió al 5% de significancia rechazar la hipótesis nula de que el modelo sin interacción es preferible. Se optó entonces por el modelo multinivel estratificado de Cox con interacción donde tanto la función de riesgo de base y los coeficientes son diferentes para cada estrato. Dado que los coeficientes y, por lo tanto, las relaciones de riesgos son diferentes para cada estrato, se procedió a la estimación de un modelo de dos niveles para cada uno de los estratos, esto permitió realizar un análisis más preciso de la supervivencia de la niñez menor de cinco años para datos de la ENDSA 2008.

Los resultados de la estimación de los modelos multinivel estratificados de Cox con interacciones que mejor se ajustan a la mortalidad de menores de cinco años para datos de la ENDSA 2008 son presentados en las Tablas 3 a la 5.

La Tabla 3 muestra los resultados de la composición de factores a nivel niño-madre, hogar y comunidad-contexto asociados a las variaciones regionales en la mortalidad de menores de cinco años en el estrato pobre del Índice de Riqueza en Bolivia, 2008. El modelo 0 (vacío) no sirve para explicar la varianza del riesgo de muerte, este solo descompone la varianza en dos componentes independientes: i) varianza del error del 1º nivel, y ii) varianza del error del nivel contextual o 2º nivel [13]. El modelo nulo es un modelo muy importante porque provee una partición básica de la variabilidad de los datos entre los 2 niveles [14]. Para el análisis de la mortalidad de la niñez menor de cinco años, que no contiene variables explicativas, la varianza entre grupos resulta ser significativa ($p < 0,05$), es decir, podemos proceder con la inclusión de variables explicativas en el modelo que reduzcan la varianza no explicada. Es decir, el anidamiento de los individuos dentro de las comunidades es significativo para explicar las variaciones de la mortalidad de menores cinco años en las zonas de residencia. Las diferencias regionales en la mortalidad de menores de cinco años pueden ser explicadas por diferencias entre los contextos y diferencias entre los individuos al interior de cada contexto.

La varianza de la mortalidad de menores de cinco años, en el modelo vacío (modelo 0) es de 0,81 ($p < 0,05$) con un error estándar de 0,48, lo que justifica el uso de un modelo multinivel para el análisis de la mortalidad de la niñez menor de cinco años en Bolivia para datos de la ENDSAB 2008. El Coeficiente Correlación Intracluster (CCI), alcanza a 32,99% de la varianza total, es decir, el 33% de la varianza total esta explicada por variaciones entre comunidades (contexto-clústeres).

De acuerdo a los resultados de la mortalidad de menores de cinco años, según datos ENDSAB 2008 (Tabla 3), la inclusión de la zona de residencia como única covariable en el análisis multinivel resulta en un riesgo 67% menor de morir para los niños de los valles ($p < 0,01$), en comparación con los niños del Altiplano y menor en 44% para los niños del los Llanos con respecto a los niños del altiplano. El cambio proporcional en la varianza (CPV) en el Modelo 1 indica que el 17,28% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia.

La inclusión de las características del niño-madre en el modelo 2 como covariables en el análisis multinivel, manteniendo la zona de residencia en el modelo, explican el 32,10% de la de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia y características a nivel del niño-madre ($p < 0,1$). Se incluyeron variables como intervalo de nacimiento, lugar de parto, edad de la madre al primer nacimiento, afiliación étnica de la madre y educación de la madre que resultaron ser significativas para explicar el riesgo de muerte de menores de cinco años según datos de la ENDSAB 2008.

Se evidencia que la incorporación de variables a nivel del hogar (Modelo 3) aportan a la explicación de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades. El 27,16% de la de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia y características a nivel del hogar ($p < 0,1$).

La inclusión de las características de la comunidad como covariables en el análisis multinivel, manteniendo la zona de residencia en el modelo, indica que el 28,4% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia y características a nivel comunidad como la educación de mujer en la comunidad, diversidad étnica, atención prenatal en la comunidad y parto en el hospital de la comunidad.

TABLA 3 - COMPOSICIÓN DE FACTORES A NIVEL NIÑO-MADRE, HOGAR Y COMUNIDAD-CONTEXTO ASOCIADOS A LAS VARIACIONES REGIONALES EN LA MORTALIDAD DE MENORES DE CINCO AÑOS EN EL ESTRATO POBRE DEL ÍNDICE DE RIQUEZA EN BOLIVIA, 2008

CARACTERÍSTICAS	Modelo 0	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
EFFECTOS FIJOS	Modelo vacío	Región	Niño/Madre	Hogar	Comunidad	Final
	CHR	CHR	CHR	CHR	CHR	CHR
Zona de Residencia						
Altiplano		1	1	1	1	1
Valles		-0,67***	-0,67***	-0,68***	-0,65***	-0,66***
Llanos		-0,44***	-0,44***	-0,45***	-0,44***	-0,44***
Intervalo de nacimiento						
<2 años			1			1
>=2 años			-0,85**			-0,84**
Lugar de parto						
Casa			1			1
Hospital			-0,88***			-0,79***
Edad de la madre al primer nacimiento						
<19 años			1			1
19+			-0,37**			-0,36**
Afiliación étnica de la madre						
Alguna			1			1
Ninguna			-0,80***			-0,78*
Educación de la madre						
Ninguna			1			1
Primaria			-0,20***			-0,21**
Secundaria			-0,35**			-0,34*
Superior			-1,42***			-1,25***
Nivel de educación de la mujer en la comunidad						
Bajo					1	1
Medio					-0,60*	-0,59*
Alto					-0,71*	-0,70*
Parto en el hospital de la comunidad						
Bajo					1	1
Medio					-0,30**	-0,28**
Alto					-0,42**	-0,36**
Distancia al Centro de Salud						
Problema					1	1
No es un problema					-0,50***	-0,43***
Diversidad Étnica						
Homogénea					1	1
Mixta					0,23**	0,20**
Heterogénea					-0,35*	-0,33*
Atención prenatal en la comunidad-proveedor especializado						
Baja					1	1
Media					-0,32*	-0,30**
Alta					-0,36*	-0,34*
Nivel Comunidad						
Varianza (SE)	0,81(0,48)**	0,67(0,39)**	0,48(0,28)*	0,59(0,32)*	0,58(0,32)**	0,31(0,17)**
CCI%	32,99	28,94	22,59	26,40	26,07	15,86
CPV(%)	Referencia	17,28	32,10	27,16	28,40	61,73
Nivel niño/madre CPV%	Referencia	15,4	45,5	15,4	25,5	57,2
Ajuste del moldeo AIC	1232,15	1201,34	1140,32	1201,34	1180,25	1103,54

***p<0.01, **p<0.05, *p<0.10

Fuente: Elaboración propia ENDSAB 2008.

El modelo final (Modelo 5), incluye solo aquellas variables que fueron estadísticamente significativas para explicar el riesgo de muerte de menores de cinco años a nivel niño-madre, hogar y contextual. El cambio proporcional en la varianza (CPV) en el Modelo 5 indica que el 61,73% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia, características a nivel del niño-madre como lugar de parto y afiliación étnica de madre, características a nivel del hogar (índice de riqueza como variable de estrato), características a nivel contextual de la comunidad como la educación de mujer en la comunidad.

Para analizar la capacidad predictiva del modelo calculamos el coeficiente de determinación total R^2 para el 2do nivel (comunidad-clústeres), con la siguiente fórmula:

$$R^2 = 1 - \frac{\text{Varianza del modelo final}}{\text{Varianza del modelo nulo}}$$

donde la varianza del modelo final (modelo 5) representa la varianza residual en el modelo cuyo poder explicativo se pretende evaluar a través de R^2 , y la varianza del modelo nulo es la varianza del modelo vacío (modelo 0-sin covariables).

$$R^2 = 1 - \frac{0,31}{0,81} = 0,62$$

El modelo explica el 62% de la varianza entre clusteres-comunidades.

La Tabla 4 muestra los resultados de la composición de factores a nivel niño-madre, hogar y comunidad-contexto asociados a las variaciones regionales en la mortalidad de menores de cinco años en el estrato medio del Índice de Riqueza en Bolivia, 2008. Para el análisis de la mortalidad de la niñez menor de cinco años, que no contiene variables explicativas, la varianza entre grupos resulta ser significativa ($p < 0,05$), es decir, podemos proceder con la inclusión de variables explicativas en el modelo que reduzcan la varianza no explicada.

La varianza de la mortalidad de menores de cinco años, en el modelo vacío (modelo 0) es de 0,50 ($p < 0,05$) con un error estándar de 0,27, lo que justifica el uso de un modelo multinivel para el análisis de la mortalidad de la niñez menor de cinco años en Bolivia para datos de la ENDSAB 2008. El Coeficiente Correlación Intracluster (CCI), llega a 23,31% de la varianza total *i.e.* el 23% de la varianza total esta explicada por variaciones entre comunidades (contexto-clústeres).

De acuerdo a los resultados de la mortalidad de menores de cinco años, según datos ENDSAB 2008 (Tabla 4), la inclusión de la zona de residencia (modelo 1) como única covariable en el análisis multinivel resulta en un riesgo 55% menor de morir para los niños de los valles ($p < 0,01$), en comparación con los niños del Altiplano y menor en 32% para los niños del los Llanos con respecto a los niños del altiplano. El cambio proporcional en la varianza (CPV) en el Modelo 1 indica que el 18,01% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia.

La inclusión de las características del niño-madre en el modelo 2 como covariables en el análisis multinivel, manteniendo la zona de residencia en el modelo, explican el 36,02% de la de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia y características a nivel del niño-madre ($p < 0,1$). Se incluyeron variables como intervalo de nacimiento, lugar de parto, afiliación étnica de la madre y educación de la madre que resultaron ser significativas para explicar el riesgo de muerte de menores de cinco años según datos de la ENDSAB 2008 para el estrato medio del Índice de Riqueza.

Se evidencia que la incorporación de variables a nivel del hogar (Modelo 3) no aportan a la explicación de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades.

La inclusión de las características de la comunidad como covariables en el análisis multinivel, manteniendo la zona de residencia en el modelo, indica que el 34,04% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia y características a nivel comunidad como la educación de mujer en la comunidad y distancia al centro de salud.

El modelo final (Modelo 5), incluye solo aquellas variables que fueron estadísticamente significativas para explicar el riesgo de muerte de menores de cinco años a nivel niño-madre, hogar y contextual. El cambio proporcional en la varianza (CPV) en el Modelo 5 indica que el 52,05% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia, características a nivel del niño-madre como lugar de parto y afiliación étnica de madre y características a nivel contextual- comunidad como la educación de mujer en la comunidad y distancia al centro de salud. El modelo explica el 52% de la varianza entre clusteres-comunidades.

TABLA 4 - COMPOSICIÓN DE FACTORES A NIVEL NIÑO-MADRE, HOGAR Y COMUNIDAD- CONTEXTO ASOCIADOS A LAS VARIACIONES REGIONALES EN LA MORTALIDAD DE MENORES DE CINCO AÑOS EN EL ESTRATO MEDIO DEL ÍNDICE DE RIQUEZA EN BOLIVIA, 2008

CARACTERÍSTICAS	Modelo 0	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
EFECTOS FIJOS	Modelo vacío	Región	Niño/Madre	Hogar	Comunidad	Final
	CHR	CHR	CHR	CHR	CHR	CHR
Zona de Residencia						
Altiplano		1	1	1	1	1
Valles		-0,55***	-0,55***	-0,55***	-0,55***	-0,54***
Llanos		-0,32***	-0,32***	-0,30***	-0,32***	-0,31***
Intervalo de nacimiento						
<2 años			1			1
>=2 años			-0,60**			-0,58**
Lugar de parto						
Casa			1			1
Hospital			-0,35***			-0,30***
Edad de la madre al primer nacimiento						
<19 años			1			1
19+			-0,23**			-0,22**
Afiliación étnica de la madre						
Alguna			1			1
Ninguna			-0,43***			-0,40*
Educación de la madre						
Ninguna			1			1
Primaria			-0,12			-0,11
Secundaria			-0,25**			-0,20*
Superior			-0,75**			-0,73**
Nivel de educación de la mujer en la comunidad						
Bajo					1	1
Medio					-0,41*	-0,40*
Alto					-0,51*	-0,38*
Parto en el hospital de la comunidad						
Bajo					1	1
Medio					-0,26**	-0,25**
Alto					-0,39**	-0,38**
Distancia al Centro de Salud						
Problema					1	1
No es un problema					-0,30***	-0,29***
Atención prenatal en la comunidad-proveedor especializado						
Baja					1	1
Media					-0,29*	-0,25**
Alta					-0,32*	-0,30*
Nivel Comunidad						
Varianza (SE)	0,50(0,27)**	0,41(0,22)**	0,32(0,17)*	0,41(0,22)**	0,33(0,18)**	0,24(0,13)**
CCI%	23,31	19,95	16,28	19,95	16,71	12,73
CPV(%)	Referencia	18,01	36,02	18,01	34,04	52,05
Nivel niño/madre CPV%	Referencia	18,4	35,2	18,4	23,4	40,5
Ajuste del moldeo AIC	1437,11	1410,24	1363,45	1410,24	1421,42	1320,5

***p<0.01, **p<0.05, *p<0.10

Fuente: Elaboración propia ENDSAB 2008.

$$R^2 = 1 - \frac{0,24}{0,50} = 0,52$$

La Tabla 5 muestra los resultados de la composición de factores a nivel niño-madre, hogar y comunidad-contexto asociados a las variaciones regionales en la mortalidad de menores de cinco años en el estrato rico del Índice de Riqueza en Bolivia, 2008.

TABLA 5 - COMPOSICIÓN DE FACTORES A NIVEL NIÑO-MADRE, HOGAR Y COMUNIDAD-CONTEXTO ASOCIADOS A LAS VARIACIONES REGIONALES EN LA MORTALIDAD DE MENORES DE CINCO EN EL ESTRATO RICO DEL ÍNDICE DE RIQUEZA EN BOLIVIA, 2008

CARACTERÍSTICAS	Modelo 0	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
EFECTOS FIJOS	Modelo vacío	Región	Niño/Madre	Hogar	Comunidad	Final
	CHR	CHR	CHR	CHR	CHR	CHR
Zona de Residencia						
Altiplano		1	1	1	1	1
Valles		-0,43***	-0,42***	-0,43***	-0,42***	-0,42***
Llanos		-0,32***	-0,30***	-0,32***	-0,31***	-0,31***
Edad actual de la madre						
<=19 años			1			1
>=20 y <=34 años			-0,20**			-0,15**
>=35			-0,35***			-0,32***
Educación de la madre						
Ninguna			1			1
Primaria			-0,06			-0,06
Secundaria			-0,15**			-0,15**
Superior			-0,22**			-0,22**
Nivel de educación de la mujer en la comunidad						
Bajo					1	1
Medio					-0,30*	-0,29*
Alto					-0,43*	-0,41*
Atención prenatal en la comunidad-proveedor especializado						
Baja					1	1
Media					-0,23*	-0,20**
Alta					-0,30*	-0,30*
Nivel Comunidad						
Varianza (SE)	0,45(0,25)**	0,40(0,22)**	0,35(0,19)*	0,40(0,22)*	0,36(0,20)**	0,31(0,17)**
CCI%	21,50	19,56	17,54	19,56	17,96	15,86
CPV(%)	Referencia	11,11	22,22	11,11	20,00	31,11
Nivel niño/madre CPV%	Referencia	8,6	28,2	8,6	20,5	39,8
Ajuste del moldeo AIC	1834,67	1750,22	1701,56	1750,22	1730,56	1667,63

***p<0.01, **p<0.05, *p<0.10

Fuente: Elaboración propia ENDSAB 2008.

Para el análisis de la mortalidad de la niñez menor de cinco años, que no contiene variables explicativas, la varianza entre grupos resulta ser significativa (p<0,05), es decir, podemos proceder con la inclusión de variables explicativas en el modelo que reduzcan la varianza no explicada.

La varianza de la mortalidad de menores de cinco años, en el modelo vacío (modelo 0) es de 0,45 (p<0,05) con un error estándar de 0,25, lo que justifica el uso de un modelo multinivel para el análisis de la mortalidad de la niñez menor de cinco años en Bolivia para datos de la ENDSAB 2008. El Coeficiente Correlación Intracluster (CCI), alcanza a 21,50%

de la varianza total, es decir, el 22% de la varianza total esta explicada por variaciones entre comunidades (contexto-clústeres).

De acuerdo a los resultados de la mortalidad de menores de cinco años, según datos ENDSAB 2008 (Tabla 5), la inclusión de la zona de residencia como única covariable en el análisis multinivel resulta en un riesgo 43% menor de morir para los niños de los valles ($p < 0,01$), en comparación con los niños del Altiplano y menor en 32% para los niños de los Llanos con respecto a los niños del altiplano. El cambio proporcional en la varianza (CPV) en el Modelo 1 indica que el 11,11% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia.

La inclusión de las características del niño-madre en el modelo 2 como covariables en el análisis multinivel, manteniendo la zona de residencia en el modelo, explican el 22,22% de la de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia y características a nivel del niño-madre ($p < 0,1$). Se incluyeron variables como edad actual de la madre y educación de la madre que resultaron ser significativas para explicar el riesgo de muerte de menores de cinco años según datos de la ENDSAB 2008 para el estrato rico del Índice de Riqueza.

Se evidencia que la incorporación de variables a nivel del hogar (Modelo 3) no aportan a la explicación de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades.

La inclusión de las características de la comunidad (modelo 4) como covariables en el análisis multinivel, manteniendo la zona de residencia en el modelo, indica que el 20% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia y características a nivel comunidad como la educación de mujer en la comunidad y atención prenatal en la comunidad.

El modelo final (Modelo 5), incluye solo aquellas variables que fueron estadísticamente significativas para explicar el riesgo de muerte de menores de cinco años a nivel niño-madre, hogar y contextual. El cambio proporcional en la varianza (CPV) en el Modelo 5 indica que el 31,11% (2008) de la varianza asociada con los riesgos de morir de menores de cinco años en las comunidades se explicaron por la zona de residencia, características a nivel del niño-madre como edad de la madre y educación de la madre y características a nivel contextual- comunidad como la educación de mujer en la comunidad y atención prenatal en la comunidad. El modelo explica el 31% de la varianza entre clusteres-comunidades.

4. CONCLUSIONES

En este estudio se optó por la estimación de un modelo multinivel de Cox estratificado con interacción por las categorías de la variable de estrato “Índice de Riqueza”. Se estimaron 3 modelos multinivel, uno por cada estrato (pobre, medio y rico), con el propósito de evitar la omisión de la heterogeneidad y dependencia del evento (muerte de los niños menores de cinco años), que muchas veces resultan en estimaciones sesgadas e ineficientes de los efectos de las covariables.

El estatus socioeconómico de los hogares donde nacen o crían los niños también es un importante predictor del riesgo de muerte de menores de cinco años en el país. Se estableció que la concentración de la pobreza dentro de los hogares parece aumentar significativamente los riesgos de mortalidad de menores de cinco años. Los factores que contribuyen a los contextos socioeconómicos pobres en Bolivia incluyen una infraestructura deficiente. Bolivia es uno de los países América Latina que enfrentan el desafío de infraestructura más crítico. Se espera que el efecto negativo del contexto socioeconómico deficiente sobre la supervivencia de menores de cinco años sea grave en un país como Bolivia, que es altamente deficiente en infraestructura. Dado que los sucesivos gobiernos de Bolivia no han logrado totalmente proporcionar a toda la población acceso a servicios sociales tales como carreteras, electricidad, agua potable instalaciones de salud, así como valores adecuados, entre otros, los resultados de supervivencia en la niñez menor de cinco años en contextos socioeconómicos pobres es poco alentadores, tal vez porque los niños pequeños probablemente sean muy vulnerables a contextos comunitarios desfavorables derivados de la falta de servicios sociales esenciales. Esto establece una de las razones por las que los grupos poblacionales pobres tienden a tener resultados de salud muy malos con respecto a los grupos ricos (Ver Figura 1). Como era de esperar, es probable que los niños de hogares sin refrigerador, medios de transporte y mala calidad den los materiales de construcción de la vivienda, debido al limitado acceso a los recursos del hogar sufran de mayores riesgos de muerte. Esto establece la importancia de la riqueza familiar como un fuerte predictor de la mortalidad de la niñez menor de cinco años.

Se determinó que vivir en comunidades que tenían un alto porcentaje de madres que atendían su parto en un hospital estaba asociado con menores riesgos de mortalidad. El contexto socioeconómico de la comunidad donde los niños

nacieron o se criaron fue otro importante predictor de la mortalidad en Bolivia según datos de la ENDSAB 2008. Finalmente la diversidad étnica es un determinante importante del riesgo de muerte de los menores de cinco años el 2008 (estrato pobre), este resultado se encuentra asociado a costumbres, tradiciones, etc. de la comunidad.

En cuanto al análisis de varianza se estableció que el modelo estimado para el estrato pobre del Índice de Riqueza es en el que más se logra explicar la varianza entre clústeres, más del 60%, determinado por la zona de residencia, características a nivel niño-madre, hogar y contextual-comunitario. Parece ser un resultado lógico, ya que la mortalidad de menores de cinco años está concentrada en estratos pobres y donde sus determinantes son plausibles de control y mejora a partir de la política pública y otro tipo de intervenciones.

Lo contrario ocurre en el modelo estimado para el estrato rico del Índice de riqueza ya que solo se logra explicar el 31% de la varianza entre clústeres. Recordemos que la mortalidad de la niñez menor de cinco años está fuertemente concentrada en muertes de neonatos y los factores que explican estas muertes están asociados a factores congénitos, de atención oportuna e infecciones recurrentes que afectan a los niños menores de 28 días. Es decir, que el estrato rico puede evitar la mortalidad por otros factores controlables a nivel niño-madre, hogar y comunidad, pero no puede evitar la muerte de los niños por otras causas no predecibles o controlables. Consideramos que este es el caso y por ese motivo nuestro modelo no logra explicar más del 32% de la varianza entre clústeres. La varianza en el riesgo de muerte entre los clústeres de los niños pertenecientes a este estrato muere por causas no posibles de identificar en importante proporción con los datos de la ENDSAB 2008.

Una serie de hallazgos de este estudio tienen importantes implicaciones políticas. Se descubrió que los riesgos de muerte de la niñez menor de cinco años, así como los riesgos de morir a temprana edad, eran más elevados en hogares que pertenecían al estrato de riqueza pobre. Como resultado, el riesgo de muerte de niños menores de 5 años seguirán siendo altos en Bolivia si las intervenciones y políticas no se formulan ni implementan para abordar las variaciones o desigualdades en salud generadas por el estatus socioeconómico. Es necesario que los gobiernos amplíen las estrategias para agilizar el desarrollo socioeconómico en las comunidades y regiones desfavorecidas social y económicamente; de lo contrario, la mortalidad de la niñez menor de cinco años continuará aumentando en algunas regiones de Bolivia, mientras que disminuye en algunas otras regiones.

La implicación de las variaciones étnicas observadas en la mortalidad en Bolivia es que las tasas de mortalidad pueden seguir disminuyendo en grupos sin afiliación étnica, mientras que las tasas pueden seguir aumentando en grupos con afiliación étnica.

Es necesario comprender mejor la importancia de las características de los contextos comunitarios para lograr una reducción razonable de la mortalidad de la niñez menor de cinco años. Las estrategias deben ampliarse para abordar los factores contextuales que amplifican los riesgos de mortalidad.

5. REFERENCIAS BIBLIOGRÁFICAS

- [1] Córdova Olivera, P., Román Eyzaguirre, S., & Soria Galvarro Ferrufino, Z. (2018). Mortalidad de la niñez menor de cinco años en Bolivia: análisis de supervivencia y sus factores de riesgo asociados ENDSA 2003, 2008 y EDSA 2016. *Investigación & Desarrollo*, 18(2), 73-92.
- [2] Boj del Val, E. (2017). El modelo de regresión de Cox. Departamento de Matemática Económica, Financiera y Actuarial. Facultad de Economía y Empresa. Universidad de Barcelona.
- [3] Therneau, T. M., & Patricia M. G. 2000. Modeling Survival Data: Extending the Cox Model. *Statistics for Biology and Health New York: Springer-Verlag*.
- [4] Griffiths, P., Madise N, Whitworth A, & Mathew Z. (2014). A tale of two continents: a multilevel comparison of the determinants of child nutritional status from selected African and Indian regions. *Health & Place*, 10, 183–199.
- [5] Cleves, C., Gould, W., & Gutiérrez, R. (2004). *An Introduction to Survival Analysis using Stata*. 2004.
- [6] Sandoval, J. (2004). Construcción de un modelo multinivel para el análisis de la agresividad indirecta en escolares; comuna nororiental, Medellín, Colombia, 2001. *Facultad Nacional de Salud Pública*, 22(2).
- [7] Ucedo Silva, V. H. 2003. Comparación de los modelos logit y probit del análisis multinivel, en el estudio de renimiento escolar.
- [8] Harttgen, K., & Misselhorn, M. (2006). *A multilevel approach to explain child mortality and undernutrition in South Asia and Sub-Saharan Africa* (No. 152). Discussion papers//Ibero America Institute for Economic Research.
- [9] Rabe-Hesketh, S., Skrondal, A. (2012). Multilevel and Longitudinal Modeling Using Stata, Volume II: Categorical Responses, Counts and Survival, Edición 3rd. [Bookshelf Online]. Recuperada de <https://bookshelf.vitalsource.com/#/books/1-59718-204-4>.
- [10] Cebolla, B., (2013). Introducción al análisis multinivel. Madrid, Centro de Investigaciones Sociológicas (CIS).

- [11] Zhang, X., Loberiza, F. R., Klein, J. P., & Zhang, M. J. (2007). A SAS macro for estimation of direct adjusted survival curves based on a stratified Cox regression model. *Computer methods and programs in biomedicine*, 88(2), 95-101.
- [12] Boj del Val, E. (2017). El modelo de regresión de Cox. Departamento de Matemática Económica, Financiera y Actuarial. Facultad de Economía y Empresa. Universidad de Barcelona.
- [13] Therneau, T. M., & Patricia M. G. 2000. Modeling Survival Data: Extending the Cox Model. *Statistics for Biology and Health New York: Springer-Verlag*.
- [14] Sear, R., Steele, F., McGregor, I. A., & Mace, R. (2002). The effects of kin on child mortality in rural Gambia. *Demography*, 39(1), 43-63.