

¿CUÁL ES EL MODELO MÁS APROPIADO PARA ANALIZAR DATOS DE SOBREVIVENCIA INFANTIL EN BOLIVIA?

WHAT IS THE MOST APPROPRIATE MODEL TO ANALYZE CHILD SURVIVAL DATA IN BOLIVIA?

Ramiro Coa Clemente¹

Instituto de Estadística Teórica y Aplicada, Universidad Mayor de San Andrés, La Paz-Bolivia

✉ clementecoa@gmail.com

Artículo recibido: 28-01-2023

Artículo aceptado: 03-04-2023

RESUMEN

El objetivo principal de este estudio es determinar cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia. Además del modelo semiparamétrico de riesgos proporcionales de Cox, se evalúan tres modelos paramétricos: exponencial, Weibull y gamma generalizado. Para abordar el objetivo se comparan tres resultados: los coeficientes estimados a partir de los distintos modelos, la significancia de las covariables en términos del valor-p y el criterio de información de Akaike (AIC). Se concluye que el modelo Weibull es más apropiado que el modelo de Cox para analizar datos de sobrevivencia infantil en Bolivia, si bien ambos modelos conducen a resultados parecidos. También se recomienda contrastar siempre los resultados de los modelos paramétricos con los del modelo de Cox.

Palabras clave: Modelo de Cox; Modelos paramétricos; Modelo más apropiado para datos de sobrevivencia.

ABSTRACT

The main objective of this study is to determine which is the most appropriate model to analyze child survival data in Bolivia. In addition to Cox's proportional hazards semiparametric model, three parametric models are evaluated: exponential, Weibull, and generalized gamma. To address the objective, three results are compared: the coefficients estimated from the different models, the significance of the covariates in terms of the p-value, and the Akaike Information Criterion (AIC). It is concluded that the Weibull model is more appropriate than the Cox model to analyze child survival data in Bolivia, although both models lead to similar results. It is also recommended to always contrast the results of the parametric models with those of the Cox model.

Keywords: Cox model; Parametric models; Most appropriate model for survival data.

1. INTRODUCCIÓN

La literatura existente sobre el tema de análisis de sobrevivencia generalmente clasifica en tres grandes grupos a los distintos métodos de análisis de datos en esta área: no-paramétricos, semiparamétricos y paramétricos (Cleves, 2016). Los modelos no-paramétricos son libres de supuestos y de otros parámetros, lo cual es equivalente

a permitir que los resultados del análisis sean determinados completamente por los datos. Los modelos semiparamétricos se caracterizan porque no requieren de supuestos acerca de la distribución del tiempo transcurrido hasta la ocurrencia del evento; empero, debido a que se parametriza el efecto de las covariables, es decir, debido a que se incluye un componente paramétrico en el análisis, este tipo de modelos reciben tal

¹ Director del Instituto de Estadística Teórica y Aplicada (IETA), Carrera de Estadística, Universidad Mayor de San Andrés. ORCID: [0000-0002-2955-0204](https://orcid.org/0000-0002-2955-0204).

¿Cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia?

denominativo. Los modelos paramétricos, sin embargo, requieren de un conjunto de supuestos acerca de la distribución de los tiempos de falla, por lo que los resultados del análisis son determinados por los supuestos y por sus cumplimientos, más que por los datos mismos.

Un método semiparamétrico muy importante tiene que ver con el modelo de regresión de riesgos proporcionales de Cox. La razón de su enorme popularidad es que este modelo no requiere que se elija una distribución de probabilidad particular para representar los tiempos de sobrevivencia, si bien se parametriza el efecto de las covariables. Otra innovación importante es que Cox propuso un nuevo método de estimación que posteriormente fue denominado verosimilitud parcial (Kalbfleisch and Prentice, 2002). Lo que es destacable o extraordinario de la verosimilitud parcial es que se puede estimar el vector de coeficientes del modelo de riesgos proporcionales sin tener que especificar la función de riesgo de línea de base. Otra interesante propiedad de los estimadores de verosimilitud parcial es que ellos dependen solo de los rangos de tiempos de falla, no de sus valores numéricos (Cleves, 2016). Sin embargo, algunos autores advierten que, a pesar de todas estas cualidades deseables, la regresión de Cox no debiera ser vista como el método universal para el análisis de regresión de datos de sobrevivencia. Hay situaciones en las que es preferible un método paramétrico.

Por otra parte, con relación a los modelos paramétricos, hay muchas distribuciones teóricas ampliamente usadas para describir los tiempos de sobrevivencia, los cuales dan origen a distintos modelos de regresión paramétricos. Entre ellos se pueden citar los modelos de regresión exponencial, Weibull, gamma generalizado, Gompertz, lognormal y loglogística (Kalbfleisch and Prentice,

2002). Algunos de estos modelos pueden ser planteados tanto en la métrica de tiempo de falla acelerado como en la métrica de riesgos proporcionales. Pero, un punto importante a resaltar sobre los modelos paramétricos es que requieren suponer una distribución de probabilidades para los tiempos de sobrevivencia y también requieren asumir alguna forma funcional para el riesgo de línea de base. Los resultados y conclusiones dependerán del grado de cumplimiento de esos supuestos.

Por lo anterior, el objetivo de este trabajo es responder las siguientes dos interrogantes: (i) ¿cuán próximos son los resultados derivados con el modelo de Cox y con los modelos paramétricos? y, entre los modelos evaluados en este artículo, (ii) ¿cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia?

2. MÉTODO

2.1 Información

La única fuente de información en Bolivia que permite abordar el objetivo de este trabajo es la encuesta nacional de demografía y salud (ENDSA). En el país, tal encuesta se realizó en seis oportunidades, la última se llevó a cabo en el año 2016. Empero, en la encuesta del año 2016 se detectó algunos problemas en los datos referidos a historia de nacimientos, razón por la que se usa la información de la penúltima encuesta nacional de demografía y salud realizada en el año 2008.

La sección de historia de nacimientos de la ENDSA de 2008 permite analizar el riesgo de muerte de los niños menores de un año (Coa y Ochoa, 2008). Se ha procesado el tiempo de vida durante el primer año de los niños nacidos vivos en los cinco años previos a la encuesta, es decir, el tiempo de vida durante

los primeros 11 meses después de haber nacido, para los 9.112 nacimientos en los 5 años previos a la encuesta. Adicionalmente, para cada nacimiento se tiene identificado su condición de sobrevivencia, es decir, si murió o estuvo vivo en el primer año de vida. En caso de haber muerto se tiene información de la edad a la que murió el niño, es decir el tiempo de vida; mientras en el caso de no haber muerto se tiene información de la edad en meses. Notar que una muerte representa la ocurrencia del evento, en cambio si el niño no ha muerto representa una censura por la derecha.

Adicionalmente, se han considerado 6 variables explicativas relacionadas con el tiempo de vida durante el primer año de vida: el nivel de educación de la madre, el quintil de riqueza del hogar, el sexo del niño, el área de residencia (urbano y rural), la edad de la madre y el departamento de residencia.

2.2 Modelos

Además del modelo semiparamétrico de riesgos proporcionales de Cox, de la amplia variedad de modelos paramétricos que se encuentra en la literatura sobre análisis de sobrevivencia se han seleccionado tres: exponencial, Weibull y gamma generalizado. En un extremo, el modelo exponencial con un parámetro tiene asociado una función de riesgo de muerte constante en el tiempo y, en el otro extremo, el modelo gamma generalizado con tres parámetros es bastante flexible por adecuarse a diversas trayectorias de la función de riesgo. En una situación intermedia podría considerarse al modelo Weibull con dos parámetros.

El modelo semiparamétrico de riesgos proporcionales de Cox

Considerando que T representa una variable

de tiempo de falla absolutamente continua y t un valor de tal variable, el modelo de riesgos proporcionales de Cox (1972), para la persona j , se lo expresa como

$$h(t/X_j) = h_0(t) e^{X_j(t)\beta},$$

donde $h_0(t)$ denota el riesgo de muerte de línea de base o riesgo de muerte común para todos los niños y que no es necesario especificarlo, $X_j(t)$ es el vector de covariables para la persona j en el momento t , β representa el vector de coeficientes de la regresión asociado a las covariables y $h(t/X_j)$ representa el riesgo instantáneo de muerte en el momento t dada las características particulares de la persona j . En este modelo, las covariables actúan multiplicativamente sobre la función de riesgo. Si las covariables son independientes del tiempo, entonces el modelo se reduce a $h(t/X_j) = h_0(t) e^{X_j\beta}$.

Para estimar el vector de coeficientes β sólo se consideran los rangos de los tiempos de falla, no sus valores numéricos. Asumiendo que los datos de tiempos de falla son censurados a la derecha, la verosimilitud parcial (Kalbfleisch y Prentice, 2002) es expresada como

$$L(\beta) = \prod_{j=1}^k \frac{e^{X_j(t_j)\beta}}{\sum_{l \in R(t_j)} e^{X_l(t_j)\beta}}$$

donde $R(t_j)$ representa el conjunto de personas en riesgo de muerte en el momento en que muere la persona j y k es el número de tiempos de falla o muerte no-censuradas, esto es, $t_1 < \dots < t_k$, ignorando por el momento el caso de empates. Los restantes $n-k$ individuos son censurados a la derecha, donde n es el tamaño de muestra. El vector de estimadores de máxima verosimilitud de

¿Cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia?

β , denotado por $\hat{\beta}$, es la solución de un sistema de ecuaciones no-lineales.

El modelo de riesgos proporcionales de Cox asume que la razón de riesgos de dos personas j y m son proporcionales, esto es, la razón

$$\frac{h(t/X_j)}{h(t/X_m)} = \frac{e^{X_j\beta}}{e^{X_m\beta}}$$

es constante en el tiempo. En la aplicación se recurre a una prueba formal para evaluar este supuesto.

Los modelos paramétricos exponencial, Weibull y gamma generalizado

Además del modelo de Cox, en este trabajo se consideran tres modelos paramétricos: exponencial, Weibull y gamma generalizado. Los primeros dos pueden modelarse tanto en términos de riesgo de muerte instantáneo (denominado modelo de riesgos proporcionales) como en términos del logaritmo natural del tiempo de sobrevivencia (denominado tiempo de falla acelerado); en cambio, el modelo gamma generalizado sólo puede formularse en términos del logaritmo natural del tiempo de sobrevivencia. Estos tres modelos son expresados en el Cuadro 1.

Si bien los dos modelos paramétricos de riesgos proporcionales del Cuadro 1 (exponencial y Weibull) tienen la misma forma funcional que el modelo semiparamétrico de Cox, es importante notar que en el caso paramétrico se define una función para el riesgo de línea de base h_0 , mientras que en el modelo de Cox no es necesario especificar su forma funcional. Por otra parte, para los tres modelos paramétricos de tiempo de

falla acelerado también expuestos en el Cuadro 1, el término $\ln \varepsilon_j$ representa el error aleatorio del modelo. Estos modelos requieren definir alguna distribución de probabilidades para este término (o para ε_j) y, consecuentemente, se tiene una distribución correspondiente para el tiempo de sobrevivencia.

Las funciones de verosimilitud de los modelos paramétricos, independientemente del caso particular, tienen la misma forma general (Cleves et al, 2016):

$$L(\beta, \theta) = \prod_{j=1}^n \frac{S(t_j/X'_j \beta, \theta)^{1-d_j} f(t_j/X'_j \beta, \theta)^{d_j}}{S(t_{0j}/X'_j \beta, \theta)}$$

donde $f()$ es la función de densidad de la distribución asumida, $S()$ es la correspondiente función de sobrevivencia, β es el vector de coeficientes asociado al vector X , θ es el vector de parámetros auxiliares, d_j es un indicador de que el niño murió ($d_j = 1$) o quedó censurado ($d_j = 0$) y n es el tamaño de muestra. Diferente de lo que acontece en el modelo de Cox, en el que la magnitud del tiempo de sobrevivencia no importa, en los modelos paramétricos sí importa la magnitud o los valores de los tiempos de sobrevivencia y, en particular, interesa el momento en que se ingresa al estudio. Por esta razón, la función de verosimilitud para el niño j incluye el término $S(t_{0j}/X'_j \beta, \theta)$ en su denominador. Cuando $t_{0j} = 0$, $S(t_{0j}/X'_j \beta, \theta) = 1$. El vector de estimadores de máxima verosimilitud de β también se obtiene resolviendo el sistema de ecuaciones no lineales con métodos iterativos.

Cuadro 1.
Tres Modelos de Regresión Paramétricos para ser Comparados con el Modelo de Cox

Modelo de Regresión	Métrica	Modelo	Parametrización	Parámetro Auxiliar
Exponencial	RP	$h(t_j) = h_0(t)e^{X_j'\beta}$	$h_0(t) = 1$ $\lambda_j = e^{X_j'\beta}$ escala	—
	TFA	$\ln t_j = X_j'\beta + \ln \varepsilon_j$	$t_j \sim E(\lambda_j = e^{-X_j'\beta})$ λ_j escala	—
Weibull	RP	$h(t_j) = h_0(t)e^{X_j'\beta}$	$h_0(t) = pt^{p-1}$ $\lambda_j = e^{X_j'\beta}$ λ_j escala p forma	p
	TFA	$\ln t_j = X_j'\beta + \sigma \ln \varepsilon_j$	$t_j \sim W(\lambda_j, p)$ $\lambda_j = e^{-pX_j'\beta}$ escala $p = \frac{1}{\sigma}$ forma σ escala	p
Gamma Generalizado	TFA	$\ln t_j = X_j'\beta + \sigma \ln \varepsilon_j$	$t_j \sim GG(\lambda_j, \alpha, \gamma)$ $\lambda_j = e^{-X_j'\beta}$ escala $\alpha = \frac{\delta}{\sigma}$ forma $\gamma = \frac{1}{\delta^2}$ forma σ escala	σ δ

RP: Riesgos proporcionales

TFA: Tiempo de falla acelerado

2.3 Criterios para comparación de resultados

Para lograr el objetivo de este artículo se usan tres criterios: (i) la comparación de los coeficientes estimados con los distintos modelos, (ii) la comparación de la significancia de las covariables en términos del valor-p y (iii) el criterio de información de Akaike (AIC). Para comparar tanto los coeficientes estimados como los valores-p, se obtiene la diferencia relativa absoluta (DRA) para la j -ésima covariable, definida como

$$DRA_j = \left| \frac{Y_j^{\text{modelo 1}} - Y_j^{\text{modelo 2}}}{Y_j^{\text{modelo 2}}} * 100 \right|,$$

donde Y_j representa el coeficiente estimado o el valor-p asociado a la covariable j en el modelo. La diferencia relativa absoluta promedio, DRAP, es simplemente el promedio aritmético de las DRA_j , esto es, $DRAP = \frac{1}{p} \sum_{j=1}^p DRA_j$, denotando p el número de coeficientes estimados en el modelo. En cambio, para obtener el AIC se usa la siguiente expresión

$$AIC = -2 \ln L + 2(p + c),$$

denotando c el número de parámetros de la distribución específica al modelo. Un valor más pequeño del AIC sugiere mejor ajuste del modelo.

¿Cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia?

Cuadro 2
Coefficientes Estimados con Distintos Modelos de Regresión

Variable/Categoría	Modelo					
	Cox RP (rr)	Exponencial RP TFA (rr) (rt)		Weibull RP TFA (rr) (rt)		Gamma Generalizado TFA (rt)
Nivel de Educación						
sin educación	3,11	3,28	0,30	3,18	0,01	0,01
primaria	1,73	1,75	0,57	1,74	0,09	0,09
secundaria	1,46	1,48	0,68	1,46	0,20	0,20
Quintil de Riqueza						
Q. más pobre	1,39	1,41	0,71	1,40	0,24	0,25
Q. pobre	1,21	1,22	0,82	1,21	0,43	0,45
Q. medio	0,85	0,84	1,19	0,85	2,05	2,08
Q. rico	0,72	0,71	1,40	0,72	4,07	4,06
Sexo del Niño						
hombre	1,27	1,27	0,79	1,27	0,36	0,37
Área de Residencia						
urbano	1,09	1,10	0,91	1,09	0,68	0,70
Edad de la Madre	0,99	0,99	1,01	0,99	1,03	1,03
Departamento						
Chuquisaca	0,66	0,66	1,52	0,66	5,87	5,85
La Paz	1,74	1,77	0,57	1,75	0,09	0,09
Cochabamba	1,58	1,60	0,62	1,58	0,14	0,14
Oruro	1,63	1,66	0,60	1,63	0,12	0,13
Potosí	1,97	2,00	0,50	1,99	0,05	0,06
Tarija	0,78	0,78	1,28	0,78	2,86	2,92
Santa Cruz	0,75	0,74	1,35	0,75	3,52	3,52
Beni	0,71	0,70	1,42	0,71	4,32	4,25
Constante	-	0,00225	444,66	0,01230	157000000,0	70400000,0

Fuente: Elaboración propia

RP = riesgos profesionales, TFA = tiempo de falla acelerado

rr = razón de riesgos, rt = razón de tiempos

3. RESULTADOS

Comparación basada en el primer criterio:

En el Cuadro 2 se presentan los coeficientes estimados con distintos modelos de regresión. Los valores de las tres columnas resaltadas con rojo son directamente comparables. Esto es, permiten comparar los coeficientes de los modelos expresados en términos de razón de riesgos.

En general, las diferencias relativas absolutas

entre los coeficientes estimados a partir de los modelos exponencial, Weibull y Cox son pequeñas. Sin embargo, las diferencias entre los coeficientes de los modelos Cox y Weibull son mucho más pequeñas que las diferencias entre los coeficientes de los modelos Cox y exponencial. En efecto, para el primer caso, la diferencia relativa absoluta promedio es sólo 0.4 por ciento, mientras para el segundo caso asciende a 1.2 por ciento (Cuadro 2a). La razón por la que los coeficientes del modelo de Cox

son más próximos a los del modelo Weibull que a los del modelo exponencial se debe a que la distribución exponencial tiene una tasa instantánea de riesgo de muerte constante en el tiempo, mientras la evolución de esta tasa para la distribución Weibull es más flexible,

puede ser creciente o decreciente, motivo por el que esta distribución se acomoda mucho más que la exponencial a los datos reales, al menos en el caso boliviano. En suma, los coeficientes de los modelos Cox y Weibull son los más parecidos.

Cuadro 2a.
Diferencias Relativas Absolutas entre los Coeficientes de Distintos Modelos

Variable/ Categoría	Cox v.s. Exponencial (rr)	Cox v.s. Weibull (rr)	Exponencial v.s. Gamma Generalizado (rt)	Weibull v.s. Gamma Generalizado (rt)
Nivel de Educación				
sin educación	5,5	2,2	97,4	12,8
primaria	1,0	0,3	84,0	1,8
secundaria	1,6	0,2	71,0	0,7
Quintil de Riqueza				
Q. más pobre	1,1	0,2	64,6	5,2
Q. pobre	0,6	0,0	44,7	4,3
Q. medio	1,0	0,3	74,5	1,5
Q. rico	1,5	0,4	189,6	0,2
Sexo del Niño				
hombre	0,3	0,2	53,1	2,9
Área de Residencia				
urbano	0,5	0,0	23,6	1,6
Edad de la Madre	0,3	0,0	2,4	0,0
Departamento				
Chuquisaca	0,6	0,0	285,5	0,3
La Paz	1,5	0,8	83,5	4,2
Cochabamba	1,4	0,4	77,2	2,6
Oruro	1,6	0,2	79,1	3,6
Potosí	1,7	0,9	88,9	5,0
Tarija	0,2	0,1	128,8	2,1
Santa Cruz	0,7	0,2	160,6	0,2
Beni	1,4	0,3	198,8	1,7
Promedio	1,2	0,4	100,4	2,8

Fuente: Elaboración propia

rr = razón de riesgos, rt = razón de tiempos

Cabe notar que, si bien los coeficientes del modelo gamma generalizado no son comparables a los del modelo de Cox por estar representados en distintas métricas, se puede

percibir la similitud entre los coeficientes de los modelos gamma generalizado y Weibull cuando se hace la comparación en la métrica de tiempo de falla acelerado (Cuadro 2).

¿Cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia?

En cambio, los coeficientes del modelo exponencial distan considerablemente de los correspondientes a estos modelos; por ejemplo, la diferencia relativa absoluta promedio entre los coeficientes de los modelos exponencial y gamma generalizado es muy alta, 100.4 por ciento (Cuadro 2a). De todas maneras, la utilidad de haber incorporado el modelo gamma generalizado en el análisis se verá más adelante.

Cuadro 3
Valores-p para los Coeficientes Estimados con Distintos Modelos de Regresión

Variable/Categoría	Modelo					
	Cox RP (rr)	Exponencial RP (rr) TFA (rt)		Weibull RP (rr) TFA (rt)		Gamma Generalizado TFA (rt)
Nivel de Educación						
sin educación	0,000	0,000	0,000	0,000	0,000	0,000
primaria	0,023	0,021	0,021	0,023	0,024	0,023
secundaria	0,115	0,101	0,101	0,113	0,114	0,113
Quintil de Riqueza						
Q. más pobre	0,253	0,238	0,238	0,250	0,250	0,264
Q. pobre620	0,463	0,449	0,449	0,463	0,463	0,482
Q. medio	0,481	0,455	0,455	0,474	0,475	0,464
Q. rico	0,180	0,161	0,161	0,174	0,175	0,175
Sexo del Niño						
hombre	0,019	0,018	0,018	0,018	0,019	0,021
Área de Residencia						
urbano	0,621	0,600	0,600	0,620	0,620	0,630
Edad de la Madre	0,304	0,142	0,142	0,278	0,278	0,271
Departamento						
Chuquisaca	0,244	0,238	0,238	0,244	0,245	0,245
La Paz	0,062	0,056	0,056	0,059	0,060	0,063
Cochabamba	0,134	0,122	0,122	0,131	0,132	0,136
Oruro	0,124	0,112	0,112	0,123	0,124	0,129
Potosí	0,025	0,021	0,021	0,023	0,024	0,026
Tarija	0,501	0,505	0,505	0,503	0,504	0,495
Santa Cruz	0,361	0,349	0,349	0,358	0,358	0,0359
Beni	0,357	0,337	0,337	0,353	0,353	0,358
Constante	-	0,000	0,000	0,000	0,000	0,000

Fuente: Elaboración propia

RP = riesgos profesionales, TFA = tiempo de falla acelerado

rr = razón de riesgos, rt = razón de tiempos

Comparación basada en el segundo criterio: el modelo gamma generalizado. La comparación se hace en términos del valor-p. Por otra parte, el Cuadro 3 permite comparar la significancia de cada una de las variables para los distintos tipos de modelos, incluido el modelo gamma generalizado. La comparación se hace en términos del valor-p. Notar que todos los valores-p en el Cuadro 3 son comparables. Se puede observar que los valores-p para el modelo de Cox son mucho

más parecidos a los del modelo Weibull que a los valores-p de los modelos exponencial y gamma generalizado. En efecto, la diferencia relativa absoluta promedio entre los valores-p para los modelos Cox y Weibull es sólo 2.2 por ciento, inferior a la diferencia promedio entre los valores-p para las otras comparaciones de modelos (Cuadro 3a).

Con base en este segundo criterio, nuevamente los modelos de Cox y Weibull presentan los mayores parecidos. Consecuentemente, las

conclusiones a cerca de la significancia de efectos de las covariables son las mismas. Por ejemplo, para un nivel de significancia de 5 por ciento, el riesgo de muerte infantil para niños de madres sin educación es un poco más de tres veces que para niños de madres con educación superior (primera fila de Cuadro 3), una diferencia altamente significativa que es revelada con ambos modelos (Cox y Weibull), incluso con el modelo exponencial, puesto que se registra un valor-p = 0 (primera fila de Cuadro 3a).

Cuadro 3a.
Diferencias Relativas Absolutas entre los Valores-p de Distintos Modelos

Variable/ Categoría	Cox v.s. Exponencial	Cox v.s. Weibull	Cox v.s. Gamma Generalizado	Weibull v.s. Gamma Generalizado
Nivel de Educación				
sin educación	0,0	0,0	0,0	0,0
primaria	8,7	0,0	0,0	4,2
secundaria	12,2	1,7	1,7	0,9
Quintil de Riqueza				
Q. más pobre	5,9	1,2	4,3	5,6
Q. pobre	3,0	0,0	4,1	4,1
Q. medio	5,4	1,5	3,5	2,3
Q. rico	10,6	3,3	2,8	0,0
Sexo del Niño				
hombre	5,3	5,3	10,5	10,5
Área de Residencia				
urbano	3,4	0,2	1,4	1,6
Edad de la Madre	53,3	8,6	10,9	2,5
Departamento				
Chuquisaca	2,5	0,0	0,4	0,0
La Paz	9,7	4,8	1,6	5,0
Cochabamba	9,0	2,2	1,5	3,0
Oruro	9,7	0,8	4,0	4,0
Potosí	16,0	8,0	4,0	8,3
Tarija	0,8	0,4	1,2	1,8
Santa Cruz	3,3	0,8	0,6	0,3
Beni	5,6	1,1	0,3	1,4
Promedio	9,1	2,2	2,9	3,1

Fuente: Elaboración propia

¿Cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia?

Comparación basada en el tercer criterio:

En el Cuadro 4 se exhiben los resultados relativos al tercer criterio de comparación, el índice AIC. El valor más pequeño de este índice corresponde al modelo Weibull (5140,6), valor muy por debajo del correspondiente al modelo de Cox (7130,4), incluso levemente inferior al valor del

índice para el modelo gamma generalizado (5144,9). En consecuencia, el modelo paramétrico Weibull es más apropiado que el modelo semiparamétrico de Cox, incluso más apropiado que los otros dos modelos paramétricos para analizar los datos de sobrevivencia infantil en Bolivia.

Cuadro 4
Valores de AIC para distintos modelos

Modelo	Métrica	Log-verosimilitud	p	c	AIC
Cox	RP	-3547,19	18	0	7130,4
Exponencial	RP	-3308,90	18	1	6655,8
	TFA	-3308,90	18	1	6655,8
Weibull	RP	-2550,29	18	2	5140,6
	TFA	-2550,29	18	2	5140,6
Gamma Generalizado	TFA	-2551,45	18	3	5144,9

Fuente: Elaboración propia

RP = riesgos proporcionales, TFA= tiempo de falla acelerado

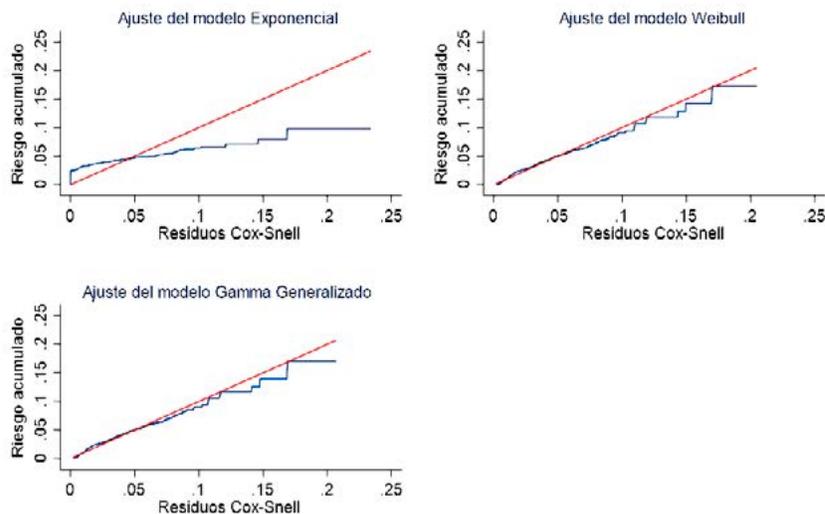
Evaluación de supuestos:

Para evidenciar el cumplimiento del supuesto de riesgos proporcionales en el modelo de Cox, cada una de las covariables se interactuó con el tiempo y se comprobó que los efectos de estas variables interactuadas no son estadísticamente distintos de cero (valor-p = 0.0954), un resultado que corrobora el cumplimiento del supuesto. Adicionalmente, el test basado en los residuos de Schoenfeld también confirma el supuesto de riesgos proporcionales, puesto que para 17 de las 18 covariables en el modelo su valor-p es superior a 0.05.

Para evaluar la bondad de ajuste a los datos de cada uno de los tres modelos paramétricos (exponencial, Weibull y gamma generalizado) se usaron los residuos de Cox-Snell. Si el modelo paramétrico Weibull, por ejemplo, es apropiado para los datos analizados, se esperaría que los residuos de Cox-Snell se comporten como una muestra

censurada proveniente de la distribución Weibull, en cuyo caso, un gráfico del riesgo acumulado basado en estos residuos debería aproximarse a una línea recta a través del origen (Kalbfleisch and Prentice, 2002). Basado en este criterio, el Gráfico 1 muestra la bondad de ajuste de los tres modelos de sobrevivencia paramétricos. Claramente se puede advertir que el modelo paramétrico exponencial – tanto de riesgos proporcionales como de tiempo de falla acelerado – no se ajusta a los datos, mientras el gráfico sugiere un buen ajuste de los modelos Weibull y gamma generalizado. Este último resultado es coherente con los valores del índice AIC (Cuadro 4). En efecto, los valores del índice para los modelos Weibull y gamma generalizado son muy próximos, si bien el valor del AIC correspondiente al modelo Weibull es levemente inferior. En suma, a partir del Gráfico 1, ambos modelos, Weibull y gamma generalizado, se ajustan bastante bien a los datos.

Gráfico 1
Evaluación de Bondad de Ajuste de Tres Modelos Paramétricos



Fuente: Elaboración propia

4. CONCLUSIÓN

A partir de los dos primeros criterios de comparación (similitud de los coeficientes en términos de razón de riesgos y similitud de los valores-p), se pudo evidenciar que los modelos Cox y Weibull proporcionan los resultados más parecidos cuando se analiza datos de supervivencia infantil en el caso boliviano. Empero, con base en el tercer criterio de comparación (el índice AIC), el modelo Weibull es más apropiado que el modelo de Cox para analizar esos datos, incluso más apropiado que los otros dos modelos paramétricos evaluados en este artículo (exponencial y gamma generalizado).

5. DISCUSIÓN

Varios estudios se han realizado, principalmente en el área médica, con el propósito de encontrar el modelo de regresión más apropiado para el análisis de datos de supervivencia. Por ejemplo, en el estudio de Hosseini y Taghi (2017) se compararon los resultados del modelo de regresión de Cox con los de varios modelos paramétricos en la

evaluación de la supervivencia de niños con leucemia aguda en el sur de Irán. Se concluyó que los modelos de regresión paramétrica demostraron un mejor rendimiento que el modelo de Cox para identificar factores de riesgo y recomiendan que no se deben ignorar los modelos paramétricos. En el estudio realizado por Amin et al (2007) también se comparó el modelo de Cox con varios modelos paramétricos para analizar datos de supervivencia de pacientes con cáncer gástrico en el hospital Taleghani de Tehran. En este estudio se concluyó que, entre los modelos paramétricos, los datos respaldaron el modelo log-normal y que éste puede conducir a resultados más precisos como una alternativa a Cox. Otros estudios también demostraron que los métodos paramétricos tienen mejor desempeño que el modelo de Cox. Wang et al (2022) compararon los resultados del modelo de Cox con los de varios modelos paramétricos en el análisis de supervivencia de variantes genéticas en el gen Hnf1b relacionado con la edad de aparición del cáncer, se concluyó que el modelo Weibull es el mejor para el análisis de sus datos. También hay estudios

¿Cuál es el modelo más apropiado para analizar datos de sobrevivencia infantil en Bolivia?

que recomiendan contrastar siempre los resultados de un análisis paramétrico con los del modelo de Cox. Por ejemplo, Nardi y Schemper (2003), en su estudio de análisis de sobrevivencia con conjuntos de datos de diferentes pruebas clínicas realizadas en la escuela médica de la universidad de Vienna, concluyeron que los resultados de un análisis paramétrico deben ser contrastados siempre con los del modelo de Cox.

En síntesis, con base en las conclusiones de estos y otros estudios, en algunas situaciones particulares los modelos paramétricos

podrían tener mejor desempeño que el modelo de Cox y, en otras, a la inversa. Por esta experiencia, la recomendación más apropiada es contrastar siempre los resultados de los modelos paramétricos con los del modelo de Cox. En el presente artículo, luego de tal contraste, se pudo evidenciar que el modelo de regresión paramétrico Weibull tuvo el mejor desempeño.

CONFLICTO DE INTERESES

El autor declara que no hay conflicto de intereses con respecto a la publicación de este documento

REFERENCIA BIBLIOGRÁFICA

- Amin P. M., Hajizadeh E., Moghimi D. B., Safae A., Abadi A. and Reza Z. M. (2007). Comparison of Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma. *Asian Pacific Journal of Cancer Prevention*, Vol. 8:412-416.
http://journal.waocp.org/article_24627_ebe3312e5286863e0b3f2b84277d66fb.pdf
- Bradburn M.J., Clark T.G., Love S.B and Altman, D.G. (2003). Survival Analysis Part II: Multivariate Data Analysis – An Introduction to Concepts and Methods. *British Journal of Cancer* 89: 431 – 436.
- Collett, D. (1994). *Modelling Survival Data in Medical Research*. Chapman Hall, London, U.K.
- Cleves M., Gould W, and Marchenko Y. (2016). *An Introduction to Survival Analysis*. Stata Press, Revised Third Edition.
- Coa R. y Ochoa L. (2008). *Encuesta Nacional de Demografía y Salud 2008*. Informe Final.
- Cox, D.R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B* 34: 187-220.
- Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman Hall, London, U.K.
- De Pascoa M., Ortega E. and Cordeiro G. (2011). The Kumaraswamy generalized gamma distribution with application in survival análisis. *Statistical Methodology* 8:411-433.
<https://web.stanford.edu/~lutian/coursepdf/reading-generalized-gamma.pdf>
- Efron B. (1977). The Eciency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association*, Vol. 72:557 – 565.
<http://www.medicine.mcgill.ca/epidemiology/hanley/bios601/SurvivalAnalysis/Efron1977jasa.pdf>
- Elandt-johnson, R.C. and Johnson, N.L. (1980). *Survival Models and Data Analysis*. Wiley, New York

- Hosseini T. S., Taghi A. S. (2017). Comparison of Cox Regression and Parametric Models: Application for Assessment of Survival of Pediatric Cases of Acute Leukemia in Southern Iran. *Asian Pacific Journal of Cancer Prevention*, Vol. 18(4):981-985.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5494248/>
- Kalbfleisch J. D., Prentice R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
- Kleinbaum, D.G., Klein, M. (2012). *Survival Analysis*. Springer, 3ra Ed.
- Kumar P., Vijay M. P, Vanita N., Amit J., Atanu B., (2016). Bayesian Accelerated Failure Time and its Application in Chemotherapy Drug Treatment Trial. *Statistics in Transition New Series*, Vol. 17, No. 4, pp. 671–690.
- Nardi A. and Schemper M. (2003). Comparing Cox and Parametric Models in Clinical Studies. *Statistics in Medicine*, Vol. 22:3597-3610.
https://cemsii.meduniwien.ac.at/fileadmin/user_upload/imported/fileadmin/msi_akim/CeMSIIS/KB/volltexte/Nardi_Schemper_2003_Statistics_in_Medicine.pdf
- Oxford Spring School. (2007). *An Introduction to Event History Analysis*.
https://spia.uga.edu/faculty_pages/rbakker/pols8501/OxfordTwoNotes.pdf
- SAS/STAT 13.2. (2014). *User's Guide The Phreg Procedure*. SAS Institute Inc.
- Srividhya K. and Radhika A. (2019). Comparison of Different Parametric Modeling for Time-to-Event Data among Cancer Patients. *International Journal of Scientific Research in Mathematical and Statistical Sciences* Vol. 6:187-192.
https://www.isroset.org/pub_paper/IJSRMSS/23-IJSRMSS-01716.pdf
- Stevenson M. (2007). *An Introduction to Survival Analysis*. EpiCentre, IVABS, Massey University.
http://www.biecek.pl/statystykaMedyczna/Stevenson_survival_analysis_195.721.pdf
- Stata. (2019). *Statistical Software*. College Station, TX: StataCorp LLC.
- Wang K., Liu X., Pan Y., Owusu D. and Xu C. (2022). Comparison of Cox Regression and Parametric Models for Survival Analysis of Genetic Variants in Hnf1b Gene Related to Age at Onset of Cancer. *Journal of Data Science*, Vol. 15: 423-442.
<https://jds-online.org/journal/JDS/article/508>