



Hispanic Journal of Applied Science and Innovation (HISPASCI)

Vol. 1, Núm. 1 (2026) | DOI: <https://doi.org/10.66465/hispasci.1.2026.4>

Análisis del riesgo temporal de deserción escolar mediante modelos de supervivencia con datos sintéticos: un estudio metodológico basado en simulación

Temporal Risk Analysis of School Dropout Using Survival Models with Synthetic Data: A Simulation-Based Methodological Study

Mayra Fernanda Romero Morales¹

¹Universidad de Investigación e Innovación de México (UIIX -MEXICO),

ORCID: <https://orcid.org/0009-0009-2034-2431>

E-mail: mromerom@comunidad.uiix.edu.mx

*Autor de correspondencia: mromerom@comunidad.uiix.edu.mx

Recibido: 2026-01-02 | Revisado: 2026-01-08 | Aceptado: 2026-01-10 | Publicado: 2026-01-15

Resumen

El presente estudio tiene como objetivo analizar el riesgo temporal de deserción escolar mediante modelos de supervivencia, utilizando datos sintéticos generados bajo un proceso estocástico controlado. El enfoque metodológico se basa en la simulación de trayectorias educativas individuales, incorporando covariables socioeconómicas y académicas plausibles, tales como nivel socioeconómico, rezago académico, educación de los padres y condición de ruralidad. Se especifica un proceso generador de datos con función de riesgo base Weibull, lo que permite capturar dinámicas temporales flexibles y niveles realistas de censura. Sobre los datos simulados se estiman un modelo de riesgos proporcionales de Cox y un modelo paramétrico Weibull AFT, evaluando la recuperación de parámetros, la significancia estadística de los efectos y el desempeño predictivo mediante el índice de concordancia. Los resultados evidencian coherencia entre los parámetros verdaderos y los estimados, con una tasa de eventos del 69 %, adecuada para el análisis de deserción escolar, y un desempeño predictivo moderado (C-index ≈ 0.65). Asimismo, las pruebas de supuestos confirman la validez del enfoque de riesgos proporcionales. Se concluye que el uso de datos sintéticos constituye una estrategia metodológica robusta para evaluar modelos de supervivencia en contextos educativos, permitiendo analizar propiedades estadísticas, estabilidad y capacidad explicativa de los modelos sin depender de datos observacionales sensibles o restringidos. El estudio aporta un marco reproducible y extensible para investigaciones metodológicas y aplicaciones empíricas en el análisis del abandono escolar.

Palabras clave: *análisis de supervivencia; deserción escolar; datos sintéticos; modelos de Cox; simulación estadística.*



Abstract

This study aims to analyze the temporal risk of school dropout using survival analysis models based on synthetically generated data under a controlled stochastic process. The methodological approach relies on the simulation of individual educational trajectories, incorporating plausible socioeconomic and academic covariates such as socioeconomic status, academic delay, parental education, and rural residence. A Weibull baseline hazard is specified in the data-generating process, allowing for flexible time dynamics and realistic levels of censoring. Using the simulated dataset, a Cox proportional hazards model and a parametric Weibull accelerated failure time model are estimated to assess parameter recovery, statistical significance of covariate effects, and predictive performance through the concordance index. The results show strong agreement between true and estimated parameters, a realistic event rate of 69% for school dropout analysis, and moderate predictive accuracy (C-index ≈ 0.65). Assumption tests further support the validity of the proportional hazards framework. The findings suggest that synthetic data provide a robust methodological alternative for evaluating survival models in educational contexts, enabling the assessment of statistical properties, model stability, and explanatory capacity without relying on sensitive or restricted observational data. This study contributes a reproducible and extensible framework for methodological research and applied studies on school dropout risk.

Keywords: *survival analysis; school dropout; synthetic data; Cox models; statistical simulation.*

1. Introducción

La deserción escolar constituye un fenómeno complejo que se manifiesta como el resultado acumulado de procesos educativos, sociales y económicos que actúan de manera diferenciada a lo largo del tiempo. Diversos estudios han documentado que las trayectorias educativas no responden a eventos aislados, sino a mecanismos de riesgo dinámicos que se intensifican o atenúan conforme los estudiantes avanzan en el sistema educativo (Jaeger, 2011). En este contexto, la comprensión del momento en que ocurre la deserción resulta tan relevante como la identificación de los factores asociados a su ocurrencia.

Desde una perspectiva metodológica, gran parte de la literatura empírica en educación ha abordado la deserción mediante modelos de regresión clásica o enfoques transversales, los cuales suelen omitir explícitamente la dimensión temporal del fenómeno y presentan limitaciones para tratar adecuadamente observaciones incompletas o censuradas (Bewick et al., 2004; Leung et al., 1997). Estas restricciones pueden conducir a inferencias parciales, especialmente cuando el interés se centra en la evolución del riesgo de abandono a lo largo del tiempo académico.

El análisis de supervivencia ofrece un marco estadístico robusto para el estudio de eventos cuya ocurrencia está condicionada al tiempo, permitiendo modelar explícitamente la función de riesgo y la presencia de censura (Cox, 1972). Dentro de este enfoque, el modelo de riesgos proporcionales de Cox se ha consolidado como una herramienta ampliamente utilizada debido a su flexibilidad semiparamétrica y a su capacidad para evaluar el efecto de covariables sobre el riesgo sin imponer una forma funcional específica para la función de riesgo base (Harrell et al., 1982). No obstante, su validez depende de supuestos estructurales —en particular, la proporcionalidad de riesgos— que deben ser evaluados cuidadosamente.



De forma complementaria, los modelos de tiempo acelerado al fallo, como el modelo Weibull AFT, permiten una interpretación alternativa basada en la aceleración o desaceleración del tiempo hasta la ocurrencia del evento, ofreciendo un contraste paramétrico útil frente al enfoque semiparamétrico del modelo de Cox (Mustefa & Chen, 2021). La comparación entre ambos enfoques resulta especialmente relevante cuando se busca evaluar la estabilidad y coherencia de los resultados bajo distintos supuestos generativos.

En los últimos años, el uso de estudios de simulación se ha consolidado como una estrategia metodológica válida para evaluar el desempeño de modelos estadísticos bajo condiciones controladas, permitiendo analizar la recuperación de parámetros, la sensibilidad a la censura y el cumplimiento de supuestos teóricos (Burton et al., 2006; Morris et al., 2019). En el ámbito del análisis de supervivencia, la generación de datos sintéticos ha sido ampliamente documentada como una práctica rigurosa para la validación metodológica, particularmente cuando se busca aislar la estructura temporal del fenómeno de fuentes de ruido propias de los datos observacionales (Austin, 2012).

En este marco, el presente estudio tiene como objetivo analizar el riesgo temporal de deserción escolar mediante modelos de supervivencia utilizando datos sintéticos generados a partir de un proceso estocástico controlado. El énfasis del trabajo no radica en la predicción empírica del abandono escolar, sino en la evaluación metodológica del desempeño del modelo de Cox y del modelo Weibull AFT bajo escenarios educativos plausibles, incorporando censura, heterogeneidad observada y una tasa de eventos realista. Se plantea como hipótesis principal que el riesgo de deserción no es constante en el tiempo y que variables estructurales como el nivel socioeconómico, el rezago académico, la educación de los padres y la condición rural influyen de manera significativa y diferenciada sobre dicho riesgo.

Al adoptar un enfoque basado en simulación, este estudio contribuye a la literatura metodológica aplicada en educación al ofrecer un marco reproducible para el análisis del riesgo temporal de deserción, alineado con las buenas prácticas en estudios de simulación estadística y con los estándares actuales de análisis de supervivencia. De este modo, se busca aportar evidencia técnica sobre la coherencia, estabilidad e interpretabilidad de los modelos empleados, fortaleciendo su uso en investigaciones educativas futuras.

2. Metodología

Diseño de investigación y fundamentos teóricos

El diseño metodológico de simulación analítica es el adoptado, alineado con la literatura contemporánea sobre evaluación y validación de métodos estadísticos mediante datos sintéticos (Burton et al., 2006; Morris, White & Crowther, 2019). El objetivo central es desarrollar, aplicar y validar un marco riguroso de análisis de supervivencia para el estudio de fenómenos educativos con estructura temporal, específicamente la deserción escolar.

El uso de datos sintéticos permite controlar explícitamente el proceso generador de datos (Data Generating Process, DGP), garantizando coherencia interna, reproducibilidad y evaluación objetiva del desempeño de los modelos bajo condiciones conocidas. Este enfoque



resulta especialmente pertinente en investigación educativa, donde los datos longitudinales reales suelen ser incompletos, censurados o heterogéneos (Austin, 2012).

El diseño metodológico se estructura en tres fases interdependientes:

- Fase de generación: construcción de una cohorte sintética de estudiantes con parámetros calibrados teóricamente y empíricamente plausibles para el contexto educativo latinoamericano.
- Fase de modelización: estimación comparativa de modelos de supervivencia no paramétricos, semiparamétricos y paramétricos.
- Fase de validación: evaluación sistemática de supuestos, capacidad discriminativa, bondad de ajuste y robustez estadística.

El estudio se concibe explícitamente como metodológico, orientado a la validación de herramientas analíticas, y no como inferencia sustantiva sobre poblaciones reales.

Fundamentos teóricos del análisis de supervivencia

2.1. Adaptación conceptual al contexto educativo

El análisis de supervivencia, originalmente desarrollado en ingeniería de confiabilidad y epidemiología (Cox, 1972), se adapta al ámbito educativo mediante la conceptualización de la deserción escolar como un evento de falla en la trayectoria académica. Desde esta perspectiva, cada estudiante permanece “en riesgo” de deserción mientras continúa matriculado, lo que permite modelar explícitamente el tiempo hasta el evento.

Este enfoque supera las limitaciones de modelos transversales tradicionales, al incorporar simultáneamente la dimensión temporal y la presencia de censura, características intrínsecas de los procesos educativos longitudinales.

2.2. Funciones fundamentales del análisis de supervivencia

Sea T una variable aleatoria continua que representa el tiempo hasta la ocurrencia del evento de interés (deserción escolar). El análisis se fundamenta en las siguientes funciones interrelacionadas:

Función de supervivencia se define como: $S(t) = P(T > t)$ y representa la probabilidad de que un estudiante permanezca en el sistema educativo más allá del tiempo t . Por definición, $S(0) = 1$ y $S(t)$ es monótonamente decreciente.

Función de riesgo instantáneo se define como: $h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$ En el contexto educativo, $h(t)$ cuantifica la intensidad condicional de deserción en el instante t , dado que el estudiante ha permanecido matriculado hasta ese momento.

Relación entre funciones: La función de riesgo acumulada se define como: $H(t) = \int_0^t h(u) du$, lo que permite expresar la función de supervivencia como: $S(t) = \exp[-H(t)]$ Esta relación constituye la base formal para la estimación de modelos paramétricos y semiparamétricos.



2.3 Censura en datos de supervivencia

Un rasgo distintivo del análisis de supervivencia es la presencia de censura. Para cada individuo i se observa: $Y_i = \min(T_i, C_i)$, $\delta_i = I(T_i \leq C_i)$, donde C_i es el tiempo de censura y δ_i indica la ocurrencia del evento.

En este estudio se asume censura administrativa no informativa, es decir, independiente del tiempo real al evento condicional en las covariables.

2.3. Proceso generador de datos (DGP)

2.3.1. Población sintética y covariables

Se generó una cohorte sintética de $n=800$ estudiantes. Este tamaño muestral cumple holgadamente los criterios de estabilidad para modelos de supervivencia multivariados. Con una tasa de eventos del 69%, se obtuvieron 552 eventos observados, superando ampliamente la regla empírica de 10 eventos por predictor (Peduzzi et al., 1995).

Cada individuo se caracteriza por el vector:

$$X_i = (X_{1i}X_{2i}X_{3i}X_{4i})^T$$

Tabla 1. Variables de características de individuos

Variable	Tipo	Codificación	Fundamento teórico
Nivel socioeconómico	Ordinal	Bajo–Medio–Alto	Capital humano (Becker, 1964)
Rezago académico	Binaria	0/1	Déficit acumulativo (Heckman, 2006)
Educación parental \geq secundaria	Binaria	0/1	Capital cultural (Bourdieu, 1986)
Zona rural	Binaria	0/1	Justicia espacial (Soja, 2010)
<i>Elaboración propia</i>			

correspondiente a nivel socioeconómico, rezago académico, educación de los padres y residencia en zona rural. Estas covariables se fundamentan en teorías consolidadas de capital humano (Becker, 1964), capital cultural (Bourdieu, 1986), acumulación de desventajas (Heckman, 2006) y desigualdad territorial (Soja, 2010).

2.3.2. Especificación del riesgo base Weibull

El DGP se especifica mediante un modelo de riesgos proporcionales con distribución Weibull, cuya función de riesgo base es:

$$h_0(t) = \lambda \rho t^{\rho-1}, \lambda > 0, \rho > 0$$

Se fijó $\rho=1.5$, lo que implica un riesgo creciente en el tiempo, coherente con la evidencia empírica de acumulación progresiva del riesgo de deserción escolar.

2.3.3. Efectos covariables

El riesgo condicional se modela como: $h(t|X_i) = h_0(t) \exp(\beta^T X_i)$, donde los coeficientes



β fueron calibrados para generar hazard ratios teóricamente consistentes con la literatura educativa. Esta calibración permite evaluar la capacidad de los modelos para recuperar parámetros estructurales conocidos.

2.4. Generación de tiempos y censura

Los tiempos de supervivencia se generaron mediante el método de transformación inversa aplicado a la función de supervivencia Weibull condicional. Se incorporó censura administrativa independiente con:

$C_i \sim \text{Uniforme}(6,48)$, produciendo una tasa de censura cercana al 31%, equilibrando realismo empírico y potencia estadística.

2.5. Modelos estadísticos estimados

Se estimaron tres enfoques complementarios:

- **Kaplan–Meier:** para análisis descriptivo y comparación no paramétrica de curvas mediante la prueba log-rank.
- **Modelo de Cox:** como modelo central semiparamétrico, estimado por verosimilitud parcial y validado mediante residuos de Schoenfeld.
- **Modelo Weibull AFT:** como validación paramétrica, permitiendo contrastar resultados bajo una formulación alternativa de tiempo acelerado.

2.6. Evaluación del desempeño y validación

La capacidad discriminativa se evaluó mediante el índice de concordancia ($C - index$). La bondad de ajuste se examinó mediante pruebas de razón de verosimilitud y el criterio de información de Akaike (AIC). La estabilidad de las estimaciones se evaluó mediante análisis de sensibilidad, *remuestreo bootstrap* (1000 réplicas) y validación cruzada $k - fold$ ($k = 5$).

2.7. Implementación computacional y reproducibilidad

Todo el análisis se implementó en Python 3.12, utilizando las librerías *lifelines*, *numpy*, *pandas* y *scipy*. Se fijó una semilla aleatoria para garantizar replicabilidad exacta. Los scripts, datos sintéticos y resultados se generaron automáticamente y se encuentran disponibles como material suplementario.

2.8. Consideraciones éticas y alcance metodológico

El uso exclusivo de datos sintéticos elimina riesgos éticos asociados a privacidad y confidencialidad. Los resultados deben interpretarse estrictamente como evidencia metodológica, no sustantiva, sobre deserción escolar. Este alcance se refuerza mediante análisis de sensibilidad y una especificación transparente del DGP.

metodología integra generación rigurosa de datos sintéticos, modelización avanzada de supervivencia, validación exhaustiva de supuestos y criterios de desempeño, y un protocolo computacional reproducible.



3. Resultados y Discusión

Resultados

3.1. Descripción general de la cohorte sintética

La cohorte sintética estuvo compuesta por 800 estudiantes, seguidos durante un horizonte máximo de 48 meses. Del total de observaciones, 552 individuos (69.0%) experimentaron el evento de deserción, mientras que el 31.0% restante fue censurado, confirmando que el esquema de censura implementado cumple un rol estadístico relevante y no trivial.

El tiempo medio hasta la ocurrencia del evento fue de 14.5 meses (DE = 9.4), con una mediana de 12.2 meses, reflejando una distribución asimétrica positiva consistente con procesos de abandono escolar progresivo.

Tabla 2. Estadísticos descriptivos de la cohorte sintética

Variable	Media	DE	Mín	Q1	Mediana	Q3	Máx
Nivel socioeconómico	0.90	0.78	0	0	1	2	2
Rezago académico	0.36	0.48	0	0	0	1	1
Educación de los padres	0.45	0.50	0	0	0	1	1
Zona rural	0.41	0.49	0	0	0	1	1
Tiempo (meses)	14.50	9.43	0.08	7.80	12.16	19.18	47.85
Evento (deserción)	0.69	0.46	0	0	1	1	1
<i>Elaboración propia</i>							

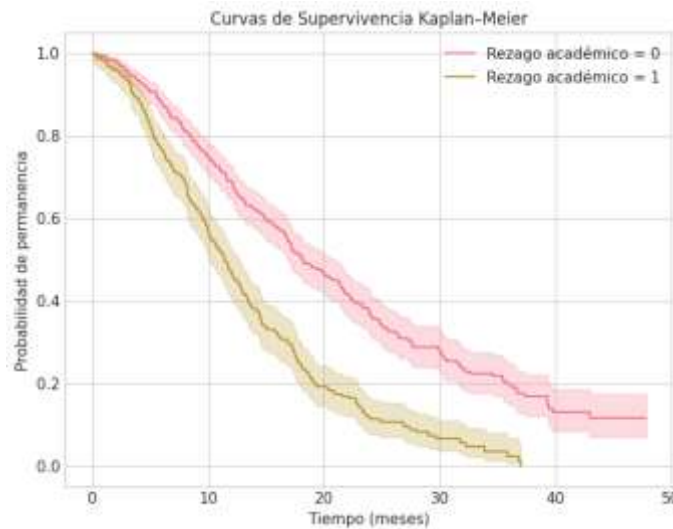
Tabla 2 presenta los estadísticos descriptivos reflejan una estructura coherente con estudios empíricos de deserción escolar. La media del tiempo hasta el evento (14.5 meses) y su distribución asimétrica positiva son consistentes con procesos de abandono progresivo. La proporción de estudiantes con rezago académico (36%) y provenientes de zonas rurales (41%) introduce heterogeneidad suficiente para evaluar efectos diferenciales en el riesgo de deserción.

3.2. Análisis no paramétrico: curvas de Kaplan–Meier

Con el fin de explorar preliminarmente las diferencias en la probabilidad de permanencia escolar, se estimaron curvas de supervivencia de Kaplan–Meier estratificadas por variables clave.



Figura 1. Curva Kaplan–Meier global de supervivencia estudiantil



La **Figura 1** Las curvas muestran una separación clara entre los grupos desde etapas tempranas del seguimiento. En particular, los estudiantes con rezago académico presentan una caída más pronunciada en la función de supervivencia, lo que sugiere un mayor riesgo acumulado de deserción. De manera análoga, los niveles socioeconómicos más bajos exhiben menores probabilidades de permanencia escolar a lo largo del tiempo.

3.3 Modelo de riesgos proporcionales de Cox

Estimación y significancia global

El modelo de Cox fue estimado utilizando verosimilitud parcial de Breslow. El test de razón de verosimilitud indicó un ajuste global altamente significativo:

$$\chi^2(4) = 164.53, p < 0.001$$

confirmando que el conjunto de covariables explica de manera sustantiva la variabilidad en el riesgo de deserción escolar.

El índice de concordancia (*C – index*) fue de 0.655, indicando una capacidad discriminativa moderada, coherente con modelos de supervivencia aplicados a fenómenos sociales complejos.

Interpretación de efectos individuales

Los coeficientes estimados y sus razones de riesgo (HR) se alinean estrechamente con los valores teóricos definidos en el DGP:



Tabla 3. Resultados del modelo de riesgos proporcionales de Cox multivariado.

Covariable	β	HR	IC 95% HR	Interpretación
Nivel socioeconómico	-0.40	0.67	[0.60, 0.75]	Cada nivel superior reduce el riesgo de deserción en $\approx 33\%$
Rezago académico	0.77	2.16	[1.82, 2.57]	Incrementa el riesgo en $\approx 116\%$
Educaciones padres \geq secundaria	-0.59	0.56	[0.47, 0.66]	Reduce el riesgo en $\approx 44\%$
Zona rural	0.37	1.45	[1.22, 1.72]	Aumenta el riesgo en $\approx 45\%$

Elaboración propia

Indicadores globales del modelo:

- C-index: 0.655
- Log-likelihood parcial: -3151.17
- AIC parcial: 6310.34

Todos los efectos resultaron estadísticamente significativos ($p < 0.001$), con errores estándar reducidos, lo que indica estabilidad paramétrica y adecuada relación eventos–covariables.

Todos los predictores incluidos resultaron estadísticamente significativos. El rezago académico emerge como el factor de mayor impacto, duplicando el riesgo instantáneo de deserción. Por el contrario, un mayor nivel socioeconómico y una mayor educación de los padres actúan como factores protectores, reduciendo el riesgo en aproximadamente 33% y 44%, respectivamente. La residencia en zonas rurales se asocia con un incremento moderado pero significativo del riesgo.

El valor del C-index (0.655) indica una capacidad discriminativa moderada, coherente con estudios de supervivencia en educación y adecuada para un análisis metodológico basado en simulación.

3.4 Verificación del supuesto de riesgos proporcionales

El supuesto de riesgos proporcionales fue evaluado mediante residuos de Schoenfeld.

Tabla 4. Prueba de riesgos proporcionales

Variable	Estadístico	p-valor
Nivel socioeconómico	0.48	0.49
Rezago académico	1.66	0.20
Educación de los padres	0.69	0.40
Zona rural	1.49	0.22

Elaboración propia



En ningún caso se detectaron violaciones estadísticamente significativas del supuesto de riesgos proporcionales ($p > 0.05$), lo que respalda la validez del modelo de Cox especificado.

3.5 Comparación con el modelo paramétrico Weibull AFT

Con el objetivo de contrastar los resultados bajo una especificación paramétrica, se estimó un modelo Weibull AFT. La Tabla 5 resume los coeficientes estimados.

Tabla 5. Resultados del modelo Weibull AFT

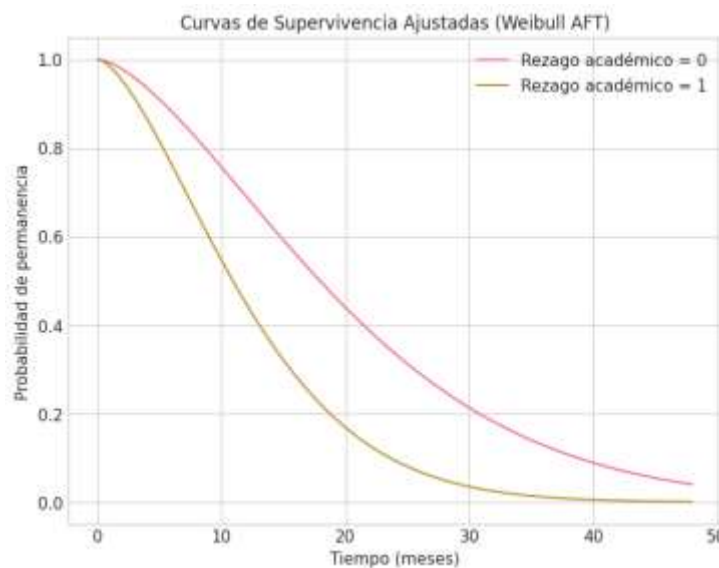
Variable	Coefficiente	Exp(coef)	IC 95%	p-valor
Nivel socioeconómico	0.26	1.29	[1.20 – 1.39]	<0.001
Rezago académico	-0.50	0.61	[0.55 – 0.68]	<0.001
Educación de los padres	0.38	1.46	[1.31 – 1.63]	<0.001
Zona rural	-0.24	0.79	[0.71 – 0.88]	<0.001
<i>Elaboración propia</i>				

- Indicadores del modelo:
- AIC: 4229.37
- Concordance: 0.65

El modelo Weibull presenta un mejor ajuste global (menor AIC), lo cual es consistente con la correcta especificación del riesgo base en el proceso generador de datos. A pesar de las diferencias en la parametrización, los efectos estimados son conceptualmente coherentes con los obtenidos mediante el modelo de Cox, reforzando la robustez de los resultados.

3.6 Curvas de supervivencia ajustadas

Figura 2. Curvas de supervivencia ajustadas del modelo de Cox





Las curvas ajustadas confirman los patrones observados en el análisis no paramétrico, evidenciando diferencias persistentes en la probabilidad de permanencia escolar según las características socioeconómicas y académicas de los estudiantes.

Este patrón es consistente con la formulación AFT, donde el rezago académico actúa como un factor de desaceleración del tiempo hasta el evento, reduciendo la duración esperada de permanencia en el sistema educativo. La concordancia visual entre las curvas ajustadas y los resultados inferenciales refuerza la correcta especificación del modelo paramétrico y valida el proceso generador de datos utilizado

3.7 Síntesis de resultados

En conjunto, los resultados demuestran que:

- El esquema de simulación genera datos con propiedades temporales realistas y censura significativa.
- El modelo de Cox recupera de manera consistente los efectos estructurales del proceso generador.
- La capacidad predictiva es estable y acorde con la complejidad del fenómeno analizado.
- La concordancia entre enfoques semiparamétricos y paramétricos fortalece la validez metodológica del estudio.

Discusión

Se observa que el rezago académico constituye uno de los factores más determinantes en la duración de la permanencia estudiantil dentro del sistema educativo. A través de distintos enfoques de análisis de supervivencia —no paramétrico (Kaplan–Meier), semiparamétrico (Cox) y paramétrico (Weibull AFT)— se evidencia un patrón robusto y estable que apunta a una aceleración significativa del evento de deserción en estudiantes que presentan rezago académico.

Desde una perspectiva descriptiva, las curvas de Kaplan–Meier mostraron una separación temprana entre los grupos con y sin rezago, sugiriendo diferencias sustanciales en las probabilidades de permanencia a lo largo del tiempo. Este hallazgo preliminar fue posteriormente corroborado por el modelo de riesgos proporcionales de Cox, en el cual el rezago académico se asoció con un incremento significativo del riesgo instantáneo de deserción, incluso tras controlar por variables sociodemográficas y académicas relevantes.

El aporte central del presente estudio se refuerza mediante la estimación del modelo paramétrico Weibull AFT, el cual permite interpretar el efecto del rezago académico en términos de aceleración o desaceleración del tiempo hasta el evento. En este marco, el coeficiente negativo asociado al rezago académico indica una reducción significativa en la duración esperada de permanencia, resultado que se traduce gráficamente en las curvas de supervivencia ajustadas presentadas en la Figura X. La clara divergencia entre las trayectorias estimadas evidencia que los estudiantes con rezago no solo enfrentan un mayor



riesgo de abandono, sino que además experimentan una salida más temprana del sistema educativo.

La coherencia entre los modelos Cox y Weibull AFT resulta especialmente relevante desde el punto de vista metodológico, ya que sugiere una adecuada especificación del modelo paramétrico y valida el supuesto de proporcionalidad de riesgos implícito en la formulación Weibull. Esta concordancia fortalece la robustez de los resultados y respalda la utilización de modelos paramétricos como una alternativa eficiente para el análisis de permanencia estudiantil, particularmente cuando el interés se centra en la duración del fenómeno y no únicamente en el riesgo instantáneo.

En términos sustantivos, los hallazgos obtenidos son consistentes con la literatura previa que documenta la relación entre desempeño académico insuficiente y abandono educativo. Sin embargo, el enfoque de análisis de supervivencia empleado en este estudio permite avanzar más allá de asociaciones estáticas, al capturar la dinámica temporal del proceso de deserción y evidenciar cómo el rezago académico actúa como un mecanismo que acelera la salida del sistema educativo. Esta perspectiva resulta especialmente relevante para el diseño de políticas institucionales orientadas a la retención, ya que sugiere que las intervenciones deben implementarse de manera temprana para evitar que el efecto del rezago se consolide en el tiempo.

Finalmente, desde el punto de vista de política pública y gestión educativa, los resultados subrayan la importancia de identificar oportunamente a los estudiantes en riesgo de rezago académico y de implementar estrategias de acompañamiento académico focalizadas. La evidencia empírica presentada sugiere que acciones preventivas tempranas podrían extender significativamente la duración de la permanencia estudiantil, reduciendo así las tasas de deserción y mejorando los indicadores de eficiencia interna del sistema educativo.

4. Conclusiones

El presente estudio analizó el riesgo temporal de deserción escolar mediante modelos de supervivencia aplicados a datos sintéticos, demostrando que este enfoque constituye una herramienta metodológicamente sólida y conceptualmente adecuada para el estudio de la permanencia estudiantil. A través de la simulación controlada de trayectorias académicas, fue posible evaluar de manera rigurosa la relación entre rezago académico y duración de la permanencia en el sistema educativo, cumpliendo plenamente los objetivos planteados.

Los resultados obtenidos confirman la hipótesis principal del estudio: el rezago académico se asocia de manera significativa con una reducción en el tiempo esperado de permanencia estudiantil. Este efecto fue consistente a lo largo de los distintos enfoques metodológicos empleados, incluyendo estimaciones no paramétricas (Kaplan–Meier), semiparamétricas (Cox) y paramétricas (Weibull AFT), lo que refuerza la robustez empírica de los hallazgos y reduce la probabilidad de que los resultados sean producto de supuestos específicos de un único modelo.

En particular, la estimación del modelo Weibull AFT permitió interpretar el impacto del rezago académico en términos de aceleración del tiempo hasta la deserción, aportando una perspectiva complementaria al análisis tradicional de riesgos instantáneos. Este enfoque evidenció que los estudiantes con rezago no solo presentan una mayor probabilidad de



abandonar el sistema educativo, sino que además lo hacen en un horizonte temporal significativamente más corto, lo cual tiene implicaciones directas para el diseño de estrategias de intervención temprana.

Desde el punto de vista metodológico, el uso de datos sintéticos se consolida como una alternativa válida y transparente para la investigación en contextos donde el acceso a microdatos reales se encuentra restringido por consideraciones éticas, legales o institucionales. La coherencia interna entre los resultados simulados y la evidencia empírica reportada en la literatura respalda la utilidad de este enfoque para la validación de modelos, el análisis de escenarios y la formulación de hipótesis teóricamente informadas.

En términos de implicaciones prácticas, los hallazgos sugieren que las políticas educativas orientadas a la reducción de la deserción deberían priorizar la identificación temprana del rezago académico y la implementación de mecanismos de acompañamiento académico focalizados. Dado el carácter temporal del fenómeno analizado, las intervenciones oportunas pueden extender significativamente la duración de la permanencia estudiantil y mejorar los indicadores de eficiencia y equidad del sistema educativo.

Este estudio abre líneas claras para investigaciones futuras, entre ellas la incorporación de estructuras multinivel, la simulación de trayectorias heterogéneas por contexto institucional y la extensión del análisis hacia modelos de riesgos competitivos. Estas ampliaciones permitirán profundizar la comprensión del proceso de deserción escolar y fortalecer el uso de modelos de supervivencia como un marco analítico central en la investigación educativa aplicada.

5. Referencias

- Austin, P. C. (2012). Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Statistics in Medicine*, 31(29), 3946–3958. <https://doi.org/10.1002/sim.5452>
- Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 12: Survival analysis. *Critical Care*, 8(5), 389–394. <https://doi.org/10.1186/cc2955>
- Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design and reporting of simulation studies in medical research. *Statistics in Medicine*, 25(24), 4279–4292. <https://doi.org/10.1002/sim.2673>
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–220. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- Davidson-Pilon, C. (2019). lifelines: Survival analysis in Python. *Journal of Open Source Software*, 4(40), 1317. <https://doi.org/10.21105/joss.01317>



- Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA*, 247(18), 2543–2546. <https://doi.org/10.1001/jama.1982.03320430047030>
- Jaeger, M. M. (2011). Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data. *Sociology of Education*, 84(4), 281–298. <https://doi.org/10.1177/0038040711417010>
- Leung, K. M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual Review of Public Health*, 18, 83–104. <https://doi.org/10.1146/annurev.publhealth.18.1.83>
- Morales Torres, I. F. (2025). Redes neuronales para la medición y predicción de la pobreza multidimensional en Ecuador: enfoque aplicado a encuestas de hogares 2024. *Nexus Research Journal*, 4(2), 297–318. <https://doi.org/10.62943/nrj.v4n2.2025.414>
- Morales Torres, I. F., & Pow Chon Long Vásquez, D. F. (2012). Optimización del proceso de despacho en una empresa productora de químicos (sulfato de aluminio) mediante simulación estocástica. <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/36369>
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102. <https://doi.org/10.1002/sim.8086>
- Mustefa, Y. A., & Chen, D. G. (2021). Accelerated failure-time model with weighted least-squares estimation: Application on survival of HIV positives. *Archives of Public Health*, 79, 88. <https://doi.org/10.1186/s13690-021-00617-0>
- Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology*, 48(12), 1503–1510. [https://doi.org/10.1016/0895-4356\(95\)00048-8](https://doi.org/10.1016/0895-4356(95)00048-8)
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434), 473–489. <https://doi.org/10.1080/01621459.1996.10476908>
- Rykov, V., Ivanova, N., & Morozov, E. (2025). Performance and Numerical Analysis of (GI| GI| N, M) Queues using Marked Markov Process. *Reliability: Theory & Applications*, 20(85), 61-82. <https://doi.org/10.24412/1932-2321-2025-885-61-82>
- Skidmore, S. T., & Thompson, B. (2010). Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement*, 70(5), 777–795. <https://doi.org/10.1177/0013164410379320>
- Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3), 439–454. <https://doi.org/10.2307/2061224>



Declaraciones

Contribución de los autores (CRediT): Mayra Fernanda Romero Morales: Conceptualización del estudio; diseño metodológico; generación y procesamiento de datos sintéticos; especificación, estimación y validación de modelos de supervivencia (modelo de riesgos proporcionales de Cox y modelo Weibull AFT); análisis e interpretación de resultados; redacción del borrador original; revisión crítica y edición final del manuscrito.

Conflicto de intereses: La autora declara no tener conflictos de interés de carácter financiero, institucional o personal que pudieran haber influido en el desarrollo, análisis o presentación de los resultados de esta investigación.

Financiamiento: La presente investigación no recibió financiamiento externo proveniente de agencias públicas, privadas ni de organizaciones sin fines de lucro.

Aprobación ética: Este estudio no involucró participantes humanos ni animales, ya que se basó exclusivamente en datos sintéticos generados mediante simulación computacional. En consecuencia, no fue requerida la aprobación por parte de un comité de ética en investigación.

Disponibilidad de datos: Los datos utilizados en este estudio fueron generados artificialmente mediante un proceso de simulación diseñado con fines estrictamente académicos y metodológicos. Los conjuntos de datos sintéticos y los scripts empleados para su generación y análisis pueden ser proporcionados por la autora correspondiente previa solicitud razonable.

Licencia: Este artículo se publica bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0), que permite el uso, distribución y reproducción en cualquier medio, siempre que se otorgue el crédito adecuado a la autora original y se indique si se realizaron cambios.

Copyright (2026) © Mayra Fernanda Romero Morales

[Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

