

## **Aplicación del análisis de supervivencia para determinar el tiempo hasta la deserción escolar en la universidad**

Application of survival analysis to determine the time until school dropout at the university

Héctor Jesús Unzueta Lazo 0000-0003-1227-3631<sup>1\*</sup>

Angel Alberto Vazquez Sánchez 0000-0002-3130-7983<sup>1</sup>

<sup>1</sup>Universidad de las Ciencias Informáticas, La Habana, Cuba.

\*Autor para correspondencia: ([hjzunzeta@uci.cu](mailto:hjunzueta@uci.cu))

## RESUMEN

En el contexto cubano la calidad de sus programas de estudios universitarios se miden a través de los indicadores de eficiencia académica y retención, que se ven directamente afectados por la deserción escolar. La deserción escolar es un problema que se ha incrementado aun cuando existen acciones para su disminución debido a las diferentes aristas que la componen. Como solución a este problema se utilizó la técnica de análisis de supervivencia como método para determinar el tiempo en que un estudiante puede llegar a causar baja de la universidad. Además se utilizan factores explicativos basados en las diferentes variables que afectan a un estudiante para causar baja. Las técnicas estadísticas de Kaplan-Meier y Modelo de riesgos proporcionales de Cox son utilizadas para obtener las medidas de supervivencia y riesgo necesarias para estimar las probabilidades de deserción con respecto al tiempo de un estudiante. Se obtiene una toma de decisiones más ajustada y una alerta temprana en posibles casos de deserción una vez determinado el tiempo crítico y los factores más significativos para la deserción.

**Palabras clave:** análisis de supervivencia; deserción escolar; factores explicativos; kaplan-meier; modelo de riesgos proporcionales de cox.

## ABSTRACT

In the Cuban context, the quality of its university study programs is measured through the indicators of academic efficiency and retention, which are directly affected by school dropouts. School dropout is a problem that has increased even though there are actions to reduce it due to the different aspects that make it up. As a solution to this problem, the survival analysis technique was used as a method to determine the time in which a student can withdraw from the university. In addition, explanatory factors based on the different variables that affect a student are used to cause withdrawal. The Kaplan-Meier and Cox Proportional Hazards Model statistical techniques are used to obtain the survival and risk measures necessary to estimate the probabilities of dropping out with respect to a student's time. A tighter decision-making and early warning in possible cases of dropout is obtained once the critical time and the most significant factors for dropout have been determined.

**Keywords:** survival analysis; school dropout; explanatory factors; kaplan-meier; cox proportional hazards model.

Recibido: 21/06/2024

Aceptado: 18/12/2024

## Introducción

La deserción escolar es un tema complejo y sus efectos no abarcan solamente la arista académica sino que afecta también aristas socio-económicas tanto personales como de la universidad. En la sociedad también son visibles las afectaciones que causa la deserción, donde los autores Ruiz-Ramírez et al. (2014), Rojas (2020) y Morales (2019) han investigado en esta temática. Estos coinciden en que la incidencia directamente en el desarrollo del capital humano, la pérdida de los valores del individuo y su posterior vulnerabilidad a caer en conductas delictivas, así como limitar el desarrollo económico y social del país son las principales consecuencias sociales que acarrea la deserción de una institución educativa.

En el contexto cubano, el Ministerio de Educación Superior (MES), organismo rector de la educación superior cubana, analiza varios indicadores relacionados con la deserción escolar o “baja” que es como comúnmente se denomina en Cuba. Estos indicadores son la eficiencia académica y la retención, que demuestran la calidad del proceso docente en las instituciones educativas y nos permite estimar cuantos estudiantes se gradúan con respecto a la matrícula inicial (Lázaro, 2020). La baja de un estudiante puede declararse por varias razones: por insuficiencia docente, por sanción disciplinaria, voluntaria, por deserción, por pérdida de requisitos, por inasistencia y definitiva (Ministerio de Educación Superior, 2022). Cada una de ellas viene dada por diversos factores que pueden ser académicos, personales, familiares, psicológicos o socio-económicos.

El contexto socio-económico actual en el país y la necesidad de elevar el indicador de eficiencia académica actual de las universidades es el impulsor principal de la búsqueda de soluciones que permitan mejorar estos aspectos. Uno de los pasos más importantes a la hora de tomar alguna decisión está relacionada directamente con la forma de predecir de forma temprana, no solo la probabilidad de que un estudiante en concreto deserte de la carrera sino además el tiempo estimado hasta que ocurra la deserción. Esto conllevaría a tomas de decisiones más precisas en aras de determinar el momento más crítico para cada

estudiante y evitar que ocurra la deserción de la universidad.

La obtención de la probabilidad de que un estudiante cause baja de una carrera, se puede obtener a través de algunos métodos estadísticos que utilizan variables predictivas para el cálculo como son el análisis discriminante (Polonio, 2022), las redes neuronales (Espino, 2017) o la regresión logística (Ventura et al., 2022). Aún así estos métodos no permiten obtener el tiempo que demora en causar baja un estudiante en un período determinado. Dentro de estos métodos destaca uno de ellos: el análisis de supervivencia (Ruiz, 2021). La utilización de esta serie de procedimientos nos permite una comprensión de los eventos que transcurren durante un determinado período de tiempo y se utiliza principalmente en el campo de la medicina (Ojeda et al., 2020). Además del campo médico tiene diversos usos en el campo del marketing, la ingeniería, la sociología y la economía. Para que el análisis de supervivencia se realice de forma eficaz es necesario contar con un conjunto de datos del proceso que se quiere estudiar. Estos datos recolectados llamados “datos de supervivencia” se refieren a las observaciones que se recogen en un estudio donde se mide el tiempo hasta que ocurre un evento de interés como el tiempo de seguimiento, estado del evento o una serie de variables asociadas al estudio (Rebasa, 2005). Estos pueden ser analizados a partir del modelo de riesgos proporcionales de Cox (Arias et al., 2023) y la función de supervivencia del método de Kaplan-Meier (Arias et al., 2022) para estimar la probabilidad de deserción en función del tiempo (Kleinbaum and Klein, 2012). Estas técnicas usan los eventos ocurridos y los tiempos de ocurrencia de los mismos, así como las características de los estudiantes para generar los resultados.

Se impone la búsqueda de soluciones que ayude a las universidades cubanas a determinar qué factores inciden más sobre los estudiantes para trazar un plan efectivo de enfrentamiento a estas cuestiones. El análisis de supervivencia proporciona herramientas únicas para estudiar eventos a lo largo del tiempo y manejar datos censurados, lo que lo diferencia significativamente de otros métodos estadísticos más generales (José, 2009). La aplicación de estas técnicas en un conjunto de datos estadísticos relacionados con las bajas permite analizar qué factores influyen más sobre la deserción estudiantil en la universidad en función del tiempo y obtener una detección temprana de los posibles casos de deserción escolar.

## Métodos

Se entiende por deserción la definición que se establece como “baja” en el Reglamento de Organización

Docente en el contexto cubano (Lázaro, 2020): “...se entiende por baja la suspensión temporal o definitiva de la condición de estudiante universitario”. La deserción se refiere al abandono prematuro de un programa de estudios antes de alcanzar el título o grado, y considera un tiempo suficientemente largo como para descartar la posibilidad de que el estudiante se reincorpore.

Respecto de la deserción, es necesario distinguir, además, entre la deserción voluntaria y la involuntaria. La deserción voluntaria puede adoptar la forma de renuncia a la carrera por parte del estudiante o del abandono no informado a la institución de educación superior. La deserción involuntaria, en cambio, se produce como consecuencia de una decisión institucional, fundamentada en sus reglamentos vigentes, que obliga al alumno a retirarse de los estudios. En este último caso, la deserción puede estar fundamentada en un desempeño académico insuficiente o responder a razones disciplinarias de diversa índole (Himmel, 2002). El abandono escolar se presenta frecuentemente por la desmotivación, que imposibilita el desarrollo del proceso académico, cabe destacar que los diversos factores de esta situación son muy frecuentes; y las diversas estrategias aplicadas en el ámbito educativo no solventan la eficacia para poder frenarla (Pachay-López and Rodríguez -Gámez, 2021).

Una perspectiva para comprender la deserción corresponde a la individual, entendiéndola en relación con las metas y propósitos que tienen los estudiantes al ingresar a una institución de educación superior, significando, para quien ejerce el acto de desertar, un fracaso individual en el proceso de completar un determinado curso de acción que propicie el logro de un objetivo. Otro de los focos desde los que es posible mirar este fenómeno es desde la institución, pues la deserción para una organización educativa implica una serie de intereses, donde el estudiante que deserta genera una vacante que podría haber sido ocupada por otro alumno que si persistiera y no provocara una pérdida de recursos para la institución. Adicionalmente, se propone una tercera perspectiva, la que dice relacionarse con el estado; desde este foco la deserción se comprende como el abandono del estudiante del sistema educativo en general, haciendo la salvedad de que en algunos casos se trata de transferencias entre instituciones o cambios al interior del sistema (Catalán, 2018).

La deserción estudiantil tiene múltiples aristas y diferentes enfoques para el análisis de este problema. Estos se pueden ubicar en 4 categorías diferentes, siendo estas: psicológicas, económicas, sociológicas, organizacionales o de interacciones. De la mayoría de estos enfoques se pueden encontrar modelos que han

recibido diferentes grados de sustentación empírica por parte de un gran número de investigaciones (Himmel, 2002).

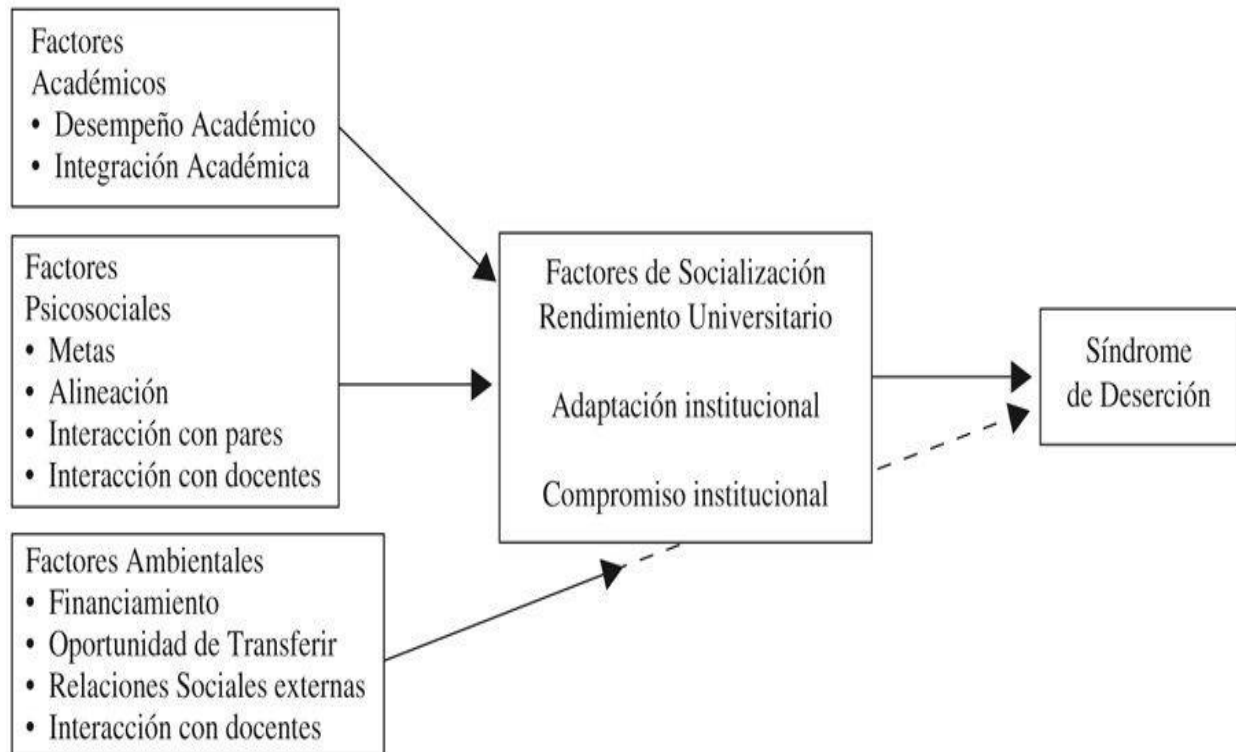
Existen diversos modelos como por ejemplo: Modelo de Tinto (García, 2024), Modelo de Rumberger (Cervero, 2020), Modelo de Cabrera (Borges et al., 2023), Modelo de Bean (Fonseca and García, 2016), entre otros. En la tabla 1 se muestra una comparación entre los diferentes modelos mencionados basado en sus características:

**Tabla 1** - Comparación de los modelos teóricos de deserción escolar.

Modelo	Enfoque Principal	Factores Clave	Limitaciones
Tinto	Integración social y académica	Compromiso, integración social-académica	No aplicable a todos los contextos
Bean	Satisfacción estudiantil	Factores académicos, psicosociales, ambientales	Enfoque más amplio pero complejo
Rumberger	Factores socioeconómicos y demográficos	Individuales, escolares, contextuales	Dificultad en la Implementación
Cabrera	Compromiso institucional	Compromiso inicial, experiencias universitarias	Cambios en expectativas a lo largo del tiempo no siempre considerados

Para la investigación tomaremos como guía el Modelo de Bean, que se centra en identificar los factores que influyen en la decisión de un estudiante de abandonar sus estudios basados en la teoría del capital humano, que sostiene que los individuos toman decisiones racionales sobre su educación en función de los costos y beneficios asociados. En la figura 1 se observa el Modelo de Bean, donde se propone que los estudiantes consideren tres tipos de costos al decidir abandonar la escuela: costos académicos, costos sociales y costos económicos. Los costos académicos se refieren a la dificultad percibida de los estudios y la falta de éxito académico. Los costos sociales se refieren a la falta de integración social en la escuela y la falta de apoyo de compañeros y profesores. Los costos económicos se refieren a las oportunidades laborales disponibles y la percepción de que la educación no ofrece beneficios económicos a corto plazo.

## Modelo de Bean (1985)



**Fig. 1** - Modelo de Bean.

Fuente: (Himmel, 2002)

Además de los costos, el modelo de Bean también considera los beneficios percibidos de abandonar la escuela, como evitar los costos mencionados anteriormente y obtener ingresos inmediatos a través del trabajo. También se incluyen los factores personales, como el nivel socioeconómico, el género, la edad y el nivel educativo de los padres, que pueden influir en la decisión de abandonar la escuela. Esta investigación se centrará sobre todo en los costos académicos y costos económicos como parte de los posibles factores explicativos que causan la probabilidad de que un estudiante esté en riesgo de desertar.

## **Características principales de un análisis de supervivencia**

El análisis de supervivencia es una colección de procedimientos estadísticos para el análisis de datos donde la variable de salida de interés es el tiempo hasta que un evento ocurre. Por tiempo podemos definir años, meses, semanas o días desde el comienzo del seguimiento de un individuo hasta que un evento ocurre. Alternativamente también se puede referir a la edad de un individuo cuando ocurre un evento. Por evento podemos definir muerte, incidencia de la enfermedad, de vuelta a remisión, recuperación o alguna experiencia designada de interés que le puede pasar a un individuo. Además, más de un evento puede ser considerado en el mismo análisis, nosotros asumiremos que solo un evento es de interés designado. Cuando más de un evento es considerado, por ejemplo, muerte por cualquiera de diversas causas, el problema estadístico puede ser caracterizado como un evento recurrente o un problema de competencia de riesgos (Kleinbaum and Klein, 2012).

En un análisis de supervivencia, usualmente nos referimos a la variable de tiempo como tiempo de supervivencia, que nos permite conocer el tiempo que un individuo ha sobrevivido en cierto periodo de seguimiento. El punto importante para definir un problema como parte de un análisis de supervivencia es precisamente que la variable de salida en estudio debe ser el tiempo hasta que ocurre un evento (Kleinbaum and Klein, 2012). Su objetivo es estimar, teniendo en cuenta la variable tiempo, la probabilidad de que ocurra un suceso determinado. Tiene la particularidad de aceptar tiempos incompletos de participación y asumir que todos los factores implicados en el estudio son homogéneos (Pérez and Martínez, 2023).

Se utiliza el método de Kaplan-Meier para estimar la probabilidad de que un evento ocurra en un tiempo determinado, considerando tanto los eventos observados como aquellos sujetos que no han experimentado el evento al final del estudio. Este método es útil en estudios donde se desea observar la duración hasta un evento sin considerar múltiples factores simultáneamente. Pero cuando se busca entender cómo diferentes variables influyen en el riesgo de ocurrencia del evento se utiliza el Modelo de riesgos proporcionales de Cox, que a diferencia del método Kaplan-Meier, este modelo no solo estima la función de supervivencia, sino que también evalúa cómo las covariables influyen en el riesgo de ocurrencia del evento. Ambos métodos son complementarios y se utilizan en conjunto para proporcionar una comprensión más completa del análisis de supervivencia. Mientras Kaplan-Meier se centra en



estimar las probabilidades acumuladas, el Modelo de Cox permite investigar los efectos relativos de diferentes factores sobre estas probabilidades.

El análisis de supervivencia toma en consideración la censura de los datos, situación que la mayoría de las técnicas estadísticas no incluyen en su análisis. Por ejemplo, si se tuviera el tiempo hasta que sucede el evento de todos y cada uno de los participantes, bien podría recurrirse a una regresión lineal o una regresión logística; sin embargo, esta información no está disponible en la mayoría de las ocasiones en las que se realiza un estudio prospectivo (Vega-Cauchich, 2019).

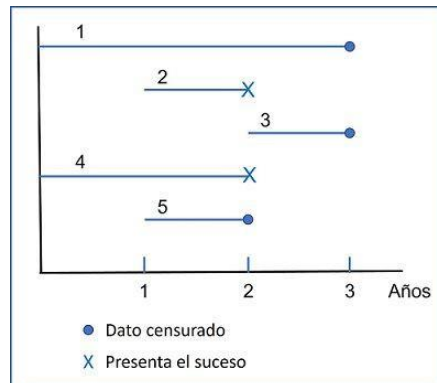
### **Datos censurados**

La mayoría de los análisis de supervivencia deben considerar una variable clave del problema analítico que se llama censurado. En esencia, el censurado ocurre cuando tenemos alguna información acerca del tiempo de supervivencia del individuo, pero no conocemos el tiempo de supervivencia exactamente. Hay tres razones por las que puede ocurrir un censurado generalmente (Kleinbaum and Klein, 2012):

- 1. Una persona no experimenta el evento designado antes de que termine el seguimiento o estudio.
- 2. Una persona se perdió (o sus datos se perdieron) durante el periodo de estudio.
- 3. Una persona se retira del estudio por diversas razones como puede ser muerte (si la muerte no es el evento de interés) o cualquier otra razón como la reacción adversa a una droga u otro riesgo de competencia.

Como se observa en la figura 2, se puede conocer que el verdadero tiempo de supervivencia de una persona se vuelve incompleta en el lado derecho del periodo de seguimiento en algunos momentos. Esto ocurre generalmente cuando termina el seguimiento o estudio (sin generar el evento), o cuando la persona se perdió durante el periodo de estudio, o la persona se retira del estudio por diversas razones. Esto se conoce como censurado- derecho (Kleinbaum and Klein, 2012).

Para estos datos el tiempo de supervivencia completo ha sido cortado (lo que llamamos censurado) en el lado derecho del intervalo del tiempo de supervivencia observado. Aunque los datos pueden tener censurado- izquierdo también, la mayoría de los datos de supervivencia están “censurados por la derecha”.



**Fig. 2** - Tipos de censurado por la derecha.

Fuente: (Arias et al., 2022)

## **Factores explicativos necesarios para el análisis de supervivencia de la deserción estudiantil**

En el contexto del análisis de supervivencia, las covariables son variables que se consideran como posibles factores que pueden influir en la supervivencia o el tiempo hasta un evento de interés. Estas variables se utilizan para ajustar los modelos de supervivencia y controlar los efectos de otros factores en el evento de interés. Las covariables pueden ser características individuales de los sujetos, como edad, género o nivel socioeconómico, o características contextuales, como el tamaño de la institución académica o los recursos disponibles. Estas variables se incluyen en el análisis como posibles factores de confusión que podrían sesgar las estimaciones de supervivencia. Al ajustar los modelos de supervivencia con las covariables asociadas al estudio, se puede determinar cómo influyen estas variables en la probabilidad de experimentar el evento de interés (Borges, 2005).

Para realizar la investigación se toman en cuenta distintos factores relacionados con el contexto específico en que se desarrolla el estudio. Se pueden agrupar estas variables a partir de dos categorías: rendimiento académico y socio-demográficas.

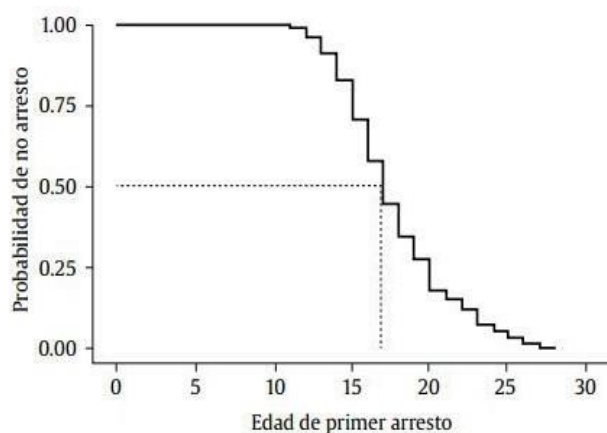
## Técnicas estadísticas

### Método de Kaplan-Meier

Si se asume que el evento terminal es independiente para cada individuo, las probabilidades de sobrevivir en un tiempo  $t$  determinado se calculan gracias a una ley multiplicativa de probabilidades. La característica distintiva es que la proporción acumulada que sobrevive se calcula para el tiempo de supervivencia individual de cada individuo y no se agrupan los tiempos de supervivencia en intervalos. Debido a ello, es especialmente útil en estudios con número pequeño de individuos (Pérez and Martínez, 2023).

Aunque la supervivencia en un tiempo dado depende de la supervivencia en todos los periodos previos, la posibilidad de la misma en un período de tiempo es independiente de la probabilidad de supervivencia en los demás períodos. Para ello se toma en consideración el número de eventos que suceden en un determinado tiempo y el número de sujetos en riesgo antes del tiempo, tomando en consideración la existencia de censura en los datos. La función estimada a partir del método de Kaplan-Meier puede ser representada gráficamente, en donde el eje X representa el tiempo desde que inicia el estudio hasta que finaliza el seguimiento y en el eje y se representa la probabilidad de que un individuo no presente el evento, como se representa en la figura 3.

Este tipo de gráfico se denomina gráfico de supervivencia o curva de supervivencia.



**Fig. 3** - Ejemplo de curva de supervivencia Kaplan-Meier.

Fuente: (Vega-Cauchich, 2019)

Cuando son estimadas, estas curvas de supervivencia se representan a través de funciones escalonadas.

Los escalones se forman a medida que transcurre el tiempo, de modo que cada vez que ocurre el evento de estudio se genera un escalón hacia abajo, lo cual conlleva una reducción de la supervivencia acumulada hasta ese instante. Se asume que los datos censurados se comportarán del mismo modo que los que han sido seguidos hasta el final y, aunque en la gráfica estos datos no disminuyan la supervivencia acumulada, provocan un mayor tamaño en el próximo escalón (Armesto, 2011).

Una de las ventajas que tiene el método de Kaplan-Meier es que permite la estimación de las funciones de supervivencia de diferentes grupos. De esta forma, se pueden utilizar medidas estadísticas para determinar si las funciones de supervivencia de un grupo son iguales o diferentes a las de otro. Además se puede calcular la mediana de supervivencia, que no requiere conocer el tiempo de supervivencia de todos los pacientes y que se estima con el percentil 50 de la distribución, que corresponde al primer tiempo con una proporción de supervivencia (probabilidad acumulada de supervivencia) igual o inferior a 0.5: sólo se trata de proyectar la probabilidad de supervivencia del 0.5 sobre la curva y comprobar a qué tiempo le corresponde.

### **Modelo de riesgos proporcionales de Cox**

Existen otras formas de estimar la función de supervivencia de una muestra, si bien el método de Kaplan-Meier es por mucho el más utilizado, también existe el estadístico de Nelson-Aalen (Ra, 2015), que permite estimar esta función a partir de la función de riesgo de la muestra. La técnica anterior si bien permite identificar la función de supervivencia de uno o varios grupos, e incluso compararlos, no nos permiten conocer la magnitud de las diferencias de éstos, o considerar otras covariables para conocer su efecto en el tiempo hasta que sucede un evento. Una técnica que permite considerar esto es utilizar modelos de regresión que tomen en cuenta la existencia de datos censurados (Vega-Cauich, 2019). Uno de los modelos más extendidos en su uso es el modelo de regresión de Cox de riesgos proporcionales (Díaz, 2018).

Esta es una técnica semiparamétrica que permite estimar la función de supervivencia mediante un modelo de regresión cuya interpretación es similar a la que se hace en una regresión logística. El modelo de riesgos proporcionales de Cox hace uso de la estimación no paramétrica propuesta por la técnica de Nelson-Aalen y permite la inclusión de covariables. Además de que requiere el cumplimiento de un

supuesto importante: riesgos proporcionales. Este supuesto significa que la función de riesgo de diferentes valores de una covariable es proporcional a lo largo del tiempo, es decir, el riesgo es independiente del tiempo (Vega-Cauchich, 2019).

Existen varios parámetros para la interpretación de los datos utilizando el Modelo de Cox, como se muestran a continuación:

- **Coefficientes de regresión:** Estos coeficientes representan el efecto de cada variable explicativa en el riesgo de ocurrencia del evento de interés. Un coeficiente positivo indica que la variable aumenta el riesgo, mientras que un coeficiente negativo indica que la variable disminuye el riesgo.
- **Razón de riesgo (HR):** El *hazard ratio* o razón de riesgo es una medida de la razón de riesgos entre dos grupos con diferentes valores de una variable predictora. Un HR mayor a 1 indica un mayor riesgo en el grupo con valores más altos de la variable, mientras que un HR menor a 1 indica un menor riesgo.
- **Intervalo de confianza:** El intervalo de confianza proporciona una estimación de la incertidumbre asociada a los parámetros estimados. Es importante tener en cuenta que si el intervalo de confianza incluye el valor 1 para el HR, no se puede concluir que hay una diferencia significativa en el riesgo entre los grupos comparados.
- **Valor p:** El valor p es una medida de la evidencia en contra de la hipótesis nula de que no hay efecto de una variable explicativa en el riesgo. Un valor p menor a un umbral predefinido (por ejemplo, 0.05) indica que hay evidencia suficiente para rechazar la hipótesis nula y concluir que existe un efecto significativo.

Además de los parámetros mencionados anteriormente, el modelo de Cox también permite generar curvas de supervivencia ajustadas. Estas curvas representan la probabilidad de supervivencia a lo largo del tiempo para diferentes grupos definidos por los valores de las variables explicativas.

Las curvas de supervivencia ajustadas muestran cómo cambia la probabilidad de supervivencia a medida que pasa el tiempo y permiten comparar la supervivencia entre diferentes grupos (Armesto, 2011). Por ejemplo, se puede comparar la supervivencia entre hombres y mujeres, o entre pacientes con diferentes valores de una variable explicativa.

En la figura 4 se representan las curvas de supervivencia ajustadas. Estas curvas son útiles para visualizar y comunicar los resultados del modelo de Cox, ya que proporcionan una representación gráfica de la relación entre las variables explicativas y la supervivencia. Además, pueden ayudar a identificar patrones o tendencias en los datos y a identificar posibles diferencias en la supervivencia entre grupos.

Cox2.png Cox2.pdf Cox2.jpg Cox2.mps Cox2.jpeg Cox2.jbig2 Cox2.jb2 Cox2.PNG Cox2.PDF Cox2.JPG  
Cox2.JP

**Fig. 4** - Ejemplo de curva de supervivencia ajustadas para el modelo de riesgos proporcionales de Cox.

Fuente: (Vega-Cauchich, 2019)

## Resultados y discusión

A partir de no contar con la posibilidad de acceso a los datos de las universidades cubanas, para la experimentación se decidió utilizar un conjunto de datos disponible en una base de datos pública que reúne la información básica para el análisis de supervivencia a realizar. La base de datos es accesible en <https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success>

in- cluye 4424 instancias y 36 características de diferentes tipologías. Se incluye información conocida en el momento de la inscripción de los estudiantes (ruta académica, demografía y factores socioeconómicos) y el rendimiento académico de los estudiantes al final del primer y segundo semestre Realinho et al. (2021). El

análisis de supervivencia se aplicó a una muestra de 1000 estudiantes matriculados en diferentes carreras de pregrado, como agronomía, diseño, educación, enfermería, periodismo, administración, servicio social y tecnologías (Realinho et al., 2021).

Se realiza el análisis en dos etapas fundamentales, primero se obtiene la función de supervivencia de Kaplan- Meier referente al evento de desertar de la universidad. Posteriormente en una segunda etapa se

construyen los modelos con factores explicativos para el uso de los modelos semiparamétricos de regresión de Cox.

### **Estructura de datos utilizada**

- Id: Identificador asociado a cada estudiante del conjunto de datos
- Tiempo: Tiempo en meses de la duración en la universidad de un estudiante.
- Evento observado: Permite conocer si el evento "desertar de la universidad fue observado o no.
- Conjunto seleccionado de 24 covariables presentes en el conjunto de datos:

1. Necesidad de educación especial
2. Internacionalidad
3. Asignaturas del 1er semestre recibidas
4. Asignaturas del 2do semestre matriculadas
5. Género
6. Asignaturas del 2do semestre evaluadas
7. Asignaturas del 2do semestre no evaluadas
8. Asignaturas del 1er semestre no evaluadas
9. Asignaturas del 1er semestre evaluadas
10. Asignaturas del 2do semestre recibidas
11. Ocupación del padre
12. Calificación del padre
13. Carrera
14. Calificación previa
15. Ocupación de la madre

16. Calificación de la madre
17. Edad al matricularse
18. Tipo de curso diurno/nocturno
19. Asignaturas del 1er semestre matriculadas
20. Nacionalidad
21. Estado civil
22. Asignaturas del 1er semestre aprobadas
23. Asignaturas del 2do semestre aprobadas
24. Modalidad en que aplicó a la carrera

### **Función de supervivencia de Kaplan-Meier**

Una vez aplicado la técnica estadística de Kaplan-Meier, se obtienen los resultados necesarios para estimar el tiempo crítico hasta la deserción escolar, como se muestra en la figura 5.

Estadísticas básicas asociadas a la función de supervivencia del evento “desertar de la universidad”

Total	Cantidad Eventos	Censurados
1000	378	622

Supervivencia.png Supervivencia.pdf Supervivencia.jpg Supervivencia.mps Supervivencia.jpeg  
Supervivencia.jbig2

**Fig. 5** - Curva de supervivencia Kaplan-Meier.

A partir de la estimación de la función de supervivencia de Kaplan-Meier se observan los siguientes datos:

- El 91,31 % de los estudiantes logran sobrevivir en la carrera 12 meses que es el instante de tiempo donde cierra el primer curso académico.



- La matrícula decrece hasta el 82,9% con respecto a la inicial a los 24 meses de haber comenzado el curso académico.
- La mayor ocurrencia del evento es observada a los 22 meses de iniciado el estudio. La segunda mayor ocurrencia ocurre en el instante de tiempo donde han transcurrido 3 meses.
- El 62,2% de los estudiantes logra sobrevivir los 60 meses. Este resultado puede oscilar entre un 57% y un 63% de supervivencia real.

**Tabla 2** - Función de supervivencia del evento “desertar de la universidad”.

Tiempo	Supervivencia	Límite inf.	Límite sup.	Eventos obser.
0	1	1	1	0
1	0.992	0.984	0.996	8
2	0.986	0.976	0.992	6
3	0.975	0.963	0.983	11
4	0.969	0.956	0.978	6
5	0.960	0.946	0.970	9
6	0.952	0.936	0.963	8
7	0.948	0.932	0.960	4
8	0.938	0.921	0.951	10
9	0.929	0.912	0.944	8
10	0.925	0.907	0.940	4

*Continúa en la próxima página*

**Tabla 2** - Función de supervivencia del evento “desertar de la universidad”

Tiempo	Supervivencia	Límite inf.	Límite sup.	Eventos obser.
11	0.919	0.901	0.935	6
12	0.913	0.894	0.929	6
13	0.903	0.883	0.920	10
14	0.896	0.875	0.913	7
15	0.891	0.870	0.909	5
16	0.888	0.866	0.906	3
17	0.880	0.858	0.899	7
18	0.872	0.850	0.892	8
19	0.869	0.846	0.889	3
20	0.863	0.840	0.883	6
21	0.854	0.830	0.874	9
22	0.841	0.817	0.863	12
23	0.836	0.811	0.858	5
24	0.829	0.803	0.851	7
25	0.825	0.800	0.848	3
26	0.817	0.791	0.840	8
27	0.812	0.786	0.835	5
28	0.810	0.784	0.833	2
29	0.802	0.776	0.826	7
30	0.795	0.768	0.819	7
31	0.787	0.760	0.812	7
32	0.781	0.754	0.806	6
33	0.776	0.748	0.801	5
34	0.768	0.740	0.793	7
35	0.760	0.732	0.786	7
36	0.755	0.727	0.781	5
37	0.750	0.721	0.776	5

*Continúa en la próxima página*

**Tabla 2** - Función de supervivencia del evento “desertar de la universidad”.

Tiempo	Supervivencia	Límite inf.	Límite sup.	Eventos obser.
38	0.746	0.718	0.773	3
39	0.734	0.705	0.761	4
...	...	...	...	...
60	0.622	0.574	0.636	0

La tabla 2 ilustra la función de supervivencia del evento “desertar de la universidad” donde puede observarse la supervivencia y los eventos observados a través de los distintos instantes de tiempo.

### Modelos de regresión de riesgos proporcionales de Cox

Para el Modelo de riesgos proporcionales de Cox se elaboró un primer modelo con todas las covariables, de las cuales solo siete resultaron significativas al 5 %. Con estas variables se construyó un segundo modelo, el cual resultó significativo.

**Tabla 3** - Covariables y sus exponentes del coeficiente para el Modelo de Cox 0.

Covariables	coef	exp(coef)	es(coef)	z	p
Estado civil	-0.11	0.89	0.10	-1.09	0.27
Modalidad en que aplicó a la carrera	0.00	1.00	0.00	0.44	0.66
Carrera	0.00	1.00	0.00	3.11	¡0.005
Tipo de curso diurno/nocturno	-0.15	0.86	0.18	-0.87	0.39
Calificación previa	0.00	1.00	0.00	-0.46	0.65
Nacionalidad	-0.01	0.99	0.01	-0.74	0.46
Calificación de la madre	0.00	1.00	0.00	-0.63	0.53
Calificación del padre	-0.01	0.99	0.00	-1.61	0.11
Ocupación de la madre	0.00	1.00	0.00	0.14	0.89
Ocupación del padre	0.00	1.00	0.00	-0.42	0.68
Necesidad de educación especial	0.34	1.40	0.43	0.80	0.43
Género	0.23	1.26	0.12	1.89	0.06

Edad al matricularse	-0.01	0.99	0.01	-1.62	0.05
Internacionalidad	0.27	1.31	0.60	0.46	0.65
Asig. 1er semestre recibidas	0.06	1.06	0.10	0.64	0.52
Asig. 1er semestre matriculadas	-0.05	0.95	0.09	-0.54	0.59
Asig. 1er semestre evaluadas	0.02	1.02	0.02	0.90	0.04
Asig. 1er semestre aprobadas	-0.19	0.83	0.05	-3.90	¡0.005
Asig. 1er semestre no evaluadas	-0.15	0.86	0.09	-1.73	0.08
Asig. 2do semestre recibidas	0.32	1.37	0.11	2.83	0.10
Asig. 2do semestre matriculadas	0.15	1.17	0.09	1.65	¡0.005
Asig. 2do semestre evaluadas	0.04	1.04	0.02	1.96	0.05
Asig. 2do semestre aprobadas	-0.41	0.67	0.04	-9.60	¡0.005
Asig. 2do semestre no evaluadas	0.01	1.01	0.06	0.16	0.88

Donde:

**coef:** es el coeficiente estimado mediante el Modelo de Cox.

**exp(coef):** es el exponente del coeficiente y se interpreta como el riesgo de fracaso(HR).

**es(coef):** es el error estándar del coeficiente

**z:** es el estadístico contraste para la significación del coeficiente.

**p:** es el valor p o el valor de significación del coeficiente.

Covariables incluidas en el Modelo de Cox 1:

- Asignaturas del 2do semestre matriculadas
- Asignaturas del 2do semestre evaluadas
- Asignaturas del 1er semestre aprobadas
- Asignaturas del 2do semestre aprobadas
- Carrera
- Edad al matricularse
- Asignaturas del 1er semestre evaluadas

**Tabla 4** - Covariables y sus exponentes del coeficiente para el Modelo de Cox 1.

Covariables	coef	exp(coef)	es(coef)	z	p
Asig. 1er semestre aprobadas	-0.15	0.86	0.03	-4.45	<0.005
Asig. 2do semestre matriculadas	0.26	1.29	0.03	9.32	<0.005
Asig. 2do semestre evaluadas	0.08	1.09	0.01	7.32	<0.005
Asig. 2do semestre aprobadas	-0.30	0.74	0.03	-9.80	<0.005
Carrera	0.00	1.00	0.00	-0.03	0.97
Edad al matricularse	-0.01	0.99	0.01	-2.11	0.03
Asig. 1er semestre evaluadas	0.03	1.03	0.02	1.63	0.10

La tabla 4 nos muestra la siguiente información:

- La cantidad de asignaturas aprobadas en el 1er semestre por los estudiantes indican un menor riesgo; siendo el riesgo de fracaso de 0.86 por cada asignatura aprobada en el semestre.
- La cantidad de asignaturas matriculadas en el 2do semestre por los estudiantes aumenta el factor de riesgo en 1,29 veces por cada asignatura matriculada.
- La cantidad de asignaturas evaluadas en el 2do semestre por los estudiantes aumenta el riesgo en 1,09 veces por cada asignatura evaluada.
- La cantidad de asignaturas aprobadas en el 2do semestre por los estudiantes tienen un riesgo menor de 0.74 veces por cada asignatura evaluada.
- La carrera específica donde se matriculó el estudiante así como las asignaturas del primer semestre evaluadas no resultaron covariables significativas para este modelo.
- La edad al matricularse comprende un riesgo menor al sólo reportar un rango de riesgo de 0.99 por cada año que tenga el estudiante.

## Conclusiones

Se aprecia que hay una reducción considerable de la matrícula en los primeros 24 meses, llegando casi a un 83 % de la matrícula inicial. Por lo tanto las acciones para mitigar la deserción deben estar dirigidas a los

primeros meses de la carrera. Del evento “desertar de la universidad” en función del tiempo de permanencia medido en meses, se obtuvieron dos modelos significativos, en el segundo, construido con aquellas covariables del primer modelo que fueron significativas al 5 % se determinaron los factores más influyentes en la deserción. Las covariables incluidas en el segundo modelo de Cox fueron:

- Asignaturas del 2do semestre matriculadas
  - Asignaturas del 2do semestre evaluadas
  - Asignaturas del 1er semestre aprobadas
  - Asignaturas del 2do semestre aprobadas
- Carrera
- Edad al matricularse
- Asignaturas del 1er semestre evaluadas

En las covariables se evidencia que el número de asignaturas del 2do semestre matriculadas y evaluadas afecta principalmente el riesgo de presentar el evento “desertar de la universidad”. Evidenciando un riesgo mucho menor en las asignaturas aprobadas tanto en el 1er semestre como en el 2do semestre. De igual forma, la carrera escogida y las asignaturas evaluadas en el 1er semestre no resultaron significativas para el modelo.

La edad al matricularse representa solo un riesgo menor para el evento “desertar de la universidad”. Esto evidencia que la toma de decisiones para evitar la deserción para estos estudiantes deben aplicarse en el 2do semestre de la carrera donde tienen mayor riesgo de presentar el síndrome de deserción. Sería de utilidad aplicar este método estadístico en un conjunto de bases de datos de estudiantes universitarios cubanos para ayudar en la toma de decisiones y visualizar cuales son los factores que actualmente son determinantes al momento de desertar. Este método es útil no solo para datos estudiantiles, es una herramienta que puede aplicarse a diversos sectores y áreas como la economía, la ingeniería, las ciencias sociales, el marketing, entre otros. Por lo tanto es recomendable extender su uso a otras disciplinas y obtener así conclusiones significativas sobre el tiempo hasta eventos críticos.

## Referencias

- M. Molina Arias, E. Ortega Páez, and C. Ochoa Sangrador. *Método de Kaplan-Meier*. [ref. de 6 de abril 2024], 2022. Disponible en Web: <https://evidenciasenpediatria.es/articulo/7991/estudios-de-supervivencia-metodo-de-kaplan-meier>.
- M. Molina Arias, E. Ortega Páez, and C. Ochoa Sangrador. *Estudios de supervivencia. Modelo de riesgos proporcionales. Regresión de Cox*. [ref. de 4 de abril 2024], 2023. Disponible en Web: <https://evidenciasenpediatria.es/articulo.php?lang=es&id=8249&tab=>.
- D. Armesto. Análisis de supervivencia. *Revista electrónica de Biomedicina*, vol 2:pp. 53–58, 2011. Disponible en Web: <https://biomed.uninet.edu/2011/n2/armesto.html>.
- C. Borges, J. L. Lordelo de Sales, and S. A. Franco. “Modelos teóricos da evasão no ensino superior e notas sobre o contexto nacional”. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, vol 28, 2023. ISSN 1414-4077. URL <https://doi.org/10.1590/S1414-40772023000100022>.
- R. E. Borges. “Análisis de supervivencia de pacientes con diálisis peritoneal”. *Rev. Colomb. Estad*, vol 28 (núm 2):pp. 243–259, Diciembre 2005.
- F. González Catalán. Deserción estudiantil en la educación superior técnico-profesional: Explorando los factores que inciden en alumnos de primer año. *Scielo México*, 2018.
- A. Cervero. *Factores relacionados con el abandono en la educación superior*. PhD thesis, Universidad de Oviedo, 2020. Disponible en Web: [https://digibuo.uniovi.es/dspace/bitstream/handle/10651/57252/TD\\_AntonioCerveroFernandezCasta%  
c3%b1on.pdf?sequence=1&isAllowed=y](https://digibuo.uniovi.es/dspace/bitstream/handle/10651/57252/TD_AntonioCerveroFernandezCasta%c3%b1on.pdf?sequence=1&isAllowed=y).
- J. L. Díaz. *Trabajo final de grado: Modelo de Cox de riesgos proporcionales*. PhD thesis, Universitat de Barcelona, 2018. Disponible en Web: <https://diposit.ub.edu/dspace/bitstream/2445/125687/2/memoria.pdf>.
- C. Espino. “Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo-herramientas Open Source que permiten su uso”. PhD thesis, Universitat Oberta de Catalunya, 2017.
- G. Fonseca and F. García. Permanencia y abandono de estudios en estudiantes universitarios: un análisis desde la teoría organizacional. *Revista de la Educación Superior*, vol 45(núm 179):pp. 25–39, Jul. 2016. [https://www.researchgate.net/publication/308273410\\_Permanencia\\_y\\_abandono\\_de\\_estudios\\_en\\_estudiantes\\_universitarios\\_un\\_analisis\\_desde\\_la\\_teor%C3%ADa\\_organizacional](https://www.researchgate.net/publication/308273410_Permanencia_y_abandono_de_estudios_en_estudiantes_universitarios_un_analisis_desde_la_teor%C3%ADa_organizacional).
- J. A González García, S. L. Canchola Magdaleno, and R. L Ulloa Cazarez. Enfoques, estudios y perspectivas

teóricas sobre la deserción escolar en la educación superior: propuesta de un modelo teórico. [ref. de 8 de abril 2024], sep. 2024. Disponible en Web: [https://mail.rediech.org/ojs/2017/index.php/ie\\_rie\\_rediech/article/view/1959](https://mail.rediech.org/ojs/2017/index.php/ie_rie_rediech/article/view/1959).

E. Himmel. “Modelo de análisis de la deserción estudiantil en la educación superior”. *Calidad en la Educación*, vol 1(núm. 17):pp. 91–108, 2002.

Belén San José, Elia Pérez, and Rosario Madero. Métodos estadísticos en estudios de supervivencia. *Anales de Pediatría Continuada*, 7(1):55–59, Feb 2009. Disponible en Web: <https://www.elsevier.es/es-revista-anales-pediatria-continuada-51-articulo-metodos-estadisticos-est>

D. G. Kleinbaum and M. Klein. *Introduction to survival analysis*. Springer, 2012. ISBN 978-1-4419-6645-2.

N. Lázaro. *Acciones Tutoriales con TIC atendiendo a factores predictivos de la deserción estudiantil en carreras de Ingeniería Informática*. PhD thesis, Universidad de Granada, 2020.

Ministerio de Educación Superior. RES. No 47/2022 BAJAS. [ref. de 2 de diciembre 2024], 2022. Disponible en Web: [https://siteal.iiep.unesco.org/sites/default/files/sit\\_accion\\_files/cuba\\_resolucion47-2022.pdf](https://siteal.iiep.unesco.org/sites/default/files/sit_accion_files/cuba_resolucion47-2022.pdf).

A. Morales. *Deserción escolar: que es, causas y consecuencias* [en línea]. Toda Materia, [ref. de 16 de marzo 2024], 2019. URL <https://www.todamateria.com/desercion-escolar/>. Disponible en Web: <https://www.todamateria.com/desercion-escolar/>.

J. L. Matos Ojeda, L. R. Águila Crespo, and I. Rodríguez Soto. “El estudio de la estimación del Análisis de Supervivencia: su importancia en las especialidades médicas”. *EDUMECENTRO*, vol 12(núm 4):pp. 257 – 260, Dic. 2020. ISSN 2077-2874. URL [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2077-28742020000400257&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742020000400257&nrm=iso).

M. J. Pachay-López and M. Rodríguez-Gámez. “La deserción escolar: Una perspectiva compleja en tiempos de pandemia”. *Polo del Conocimiento*, vol 6(núm 1):pp. 1–4, 2021.

J.A. Pérez and P.S. Martínez. Análisis de supervivencia. *Medicina de Familia. SEMERGEN*, vol 49(núm 5), 2023.

M. Polonio. “*Estudio del Análisis Discriminante. Aplicación a datos reales*”. PhD thesis, Universidad de Granada, 2022. Disponible en Web: [https://digibug.ugr.es/bitstream/handle/10481/76620/TFG\\_Maria.pdf?sequence=1](https://digibug.ugr.es/bitstream/handle/10481/76620/TFG_Maria.pdf?sequence=1).

K. Ra. “A Comparison study of Kaplan Meier and Nelson-Aalen Methods in Survival Analysis”. *ResearchGate*, vol 2(núm 11):pp. 34–38, November 2015. ISSN 2349-7610.

V. Realinho, M. Vieira, J. Machado, and L. Baptista. Predict students’ dropout and academic success. [ref. de



- 16 de marzo 2024], 2021. Disponible en Web: <https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success>, UCI Machine Learning Repository.
- P. Rebas. Conceptos básicos del análisis de supervivencia. *Cirugía española*, 78(4):222–230, 2005.
- M. E. Rojas. “La deserción escolar en Costa Rica: Un estudio de causas y consecuencias en una institución educativa”. *Diálogos Revista Electrónica de Historia*, vol. 1, núm. 4, 2020.
- Carlos Eduardo Pineda Ruiz. Modelo de supervivencia aplicado a la deserción estudiantil en la facultad de ingeniería de una universidad en la ciudad de bogota en la cohorte 2017-1 a 2021-1, 2021. Disponible en Web: <https://repository.libertadores.edu.co/bitstreams/5b6e94b9-7272-411a-8aff-49e17d3358c4>.
- R. Ruiz-Ramírez, J. L. García-Cué, and M. A. Pérez-Olvera. “Causas y consecuencias de la deserción escolar en el bachillerato: Caso Universidad Autónoma de Sinaloa”. *Ra Ximhai*, vol 10(núm. 5), Dic. 2014.
- J. I. Vega-Cauich. El análisis de supervivencia como técnica para la evaluación de la validez predictiva en la psicología jurídica. *Anuario de Psicología Jurídica*, vol 29:pp. 1–10, 2019.
- J. C. Solís Ventura, S. Quiroz, and O. Fosado. “Modelo de regresión logística para la estimación de la deserción escolar del posgrado en la Universidad Técnica de Manabí, Ecuador”. *Bases de la Ciencia*, vol 7(núm 3), 2022. ISSN 2588-0764.

### **Conflicto de interés**

El autor autoriza la distribución y uso de su artículo.

### **Contribuciones de los autores**

Conceptualización: Angel Alberto Vazquez Sánchez

Curación de datos: Héctor Jesús Unzueta Lazo

Análisis formal: Héctor Jesús Unzueta Lazo

Investigación: Héctor Jesús Unzueta Lazo

Metodología: Héctor Jesús Unzueta Lazo

Recursos: Angel Alberto Vazquez Sánchez

Software: Héctor Jesús Unzueta Lazo

Supervisión: Angel Alberto Vazquez Sánchez

Validación: Angel Alberto Vazquez Sánchez

**Visualización:** Héctor Jesús Unzueta Lazo

**Redacción - borrador original:** Héctor Jesús Unzueta Lazo

**Redacción - revisión y edición:** Angel Alberto Vazquez Sánchez