



**ANÁLISIS DE SUPERVIVENCIA APLICADO
A LA DESERCIÓN EN ESTUDIANTES DE
INGENIERÍA**

Sergio Yesid Aroca Pedroza

Lida Rubiela Fonseca Gómez

Rafael Eduardo Borges Peña

ANÁLISIS DE SUPERVIVENCIA APLICADO A LA DESERCIÓN EN ESTUDIANTES DE INGENIERÍA

SURVIVAL ANALYSIS APPLIED TO DROPOUT IN ENGINEERING STUDENTS

Sergio Yesid Aroca Pedroza
sergioyesid.aroca@unincca.edu.co

Universidad Incca de Colombia

Lida Rubiela Fonseca Gómez

lrfonsecag@libertadores.edu.co

Fundación Universitaria Los Libertadores

Rafael Eduardo Borges Peña

borgesr@gmail.com

Universidad de los Andes - Venezuela

RESUMEN

El presente trabajo tiene como finalidad el realizar e interpretar un análisis de supervivencia aplicado a la problemática de la deserción en estudiantes de la Facultad de Ingeniería de una Universidad de la ciudad de Bogotá, Colombia. Para esto se ha considerado los datos obtenidos por parte de la oficina de Registraduría de la Universidad, tomando como énfasis desde el primer semestre del 2010 hasta el segundo semestre del 2018. Para la ejecución de este análisis se ha hecho uso de la técnica del estimador de Kaplan y Meier, seguido de un modelo de Regresión de Cox. El análisis permite enfocar cuales variables son más influyentes al estudiante

al momento de tomar la decisión de desertar de su formación académica, y por lo tanto permitirá tomar a la Universidad acciones correctivas que puedan disminuir este factor y así asegurar una estabilidad de permanencia estudiantil con respecto al tiempo.

Palabras clave: Deserción, análisis de supervivencia, estimación Kaplan-Meier, modelo Regresión de Cox.

ABSTRACT

The purpose of this work is to carry out and interpret a survival analysis applied to the problem of desertion in students of the

Faculty of Engineering of a University in the city of Bogotá, Colombia. For this, the data obtained by the University Registrar's office has been considered, taking as emphasis from the first semester of 2010 to the second semester of 2018. For the execution of this analysis, the estimator technique has been used of Kaplan and Meier, followed by a Cox Regression model. The analysis allows focusing on which variables are most influential to the student when making the decision to drop out of their academic training, and therefore will allow the University to take corrective actions that may decrease this factor and thus ensure stability of student permanence with respect to at the time.

Keywords: Dropout, survival analysis, Kaplan-Meier estimation, Cox Regression model.

INTRODUCCIÓN

La Educación a nivel nacional ha ido presentando un inconveniente latente con el pasar de los años, y es el tema de la deserción estudiantil, es decir cuando un estudiante por diversos motivos no culmina su formación académica. El problema de la deserción se puede evaluar desde diferentes aspectos, y como tal ocasiona diversos efectos, no solo en el entorno educativo, sino también en el entorno social nacional. Es por esto que cualquier esfuerzo por disminuir la deserción se

refleja en el aumento de la cobertura y el mejoramiento de la eficiencia de la educación superior (González, 2015).

La deserción estudiantil es una problemática significativa la cual debe afrontar cada institución educativa no solo a nivel local sino también nacional, debido al hecho de que no solo pueden presentarse resultados adversos en la institución educativa, sino también en la mano de obra capacitada para el país. Al ser esta una problemática significativa es importante evaluar cada aspecto relevante que lleve a un estudiante a tomar esta decisión, la medición y estudio deben ser parte de la evaluación de la eficiencia del sistema educativo, de la calidad de los procesos y de los programas que ofrecen las instituciones (Sánchez & Márquez, 2012). Es por esto que el diseño de un modelo que permita identificar dichos aspectos se hace necesario para una institución educativa, de manera tal que se puedan tomar medidas correctivas y preventivas para poder disminuir esta problemática en un futuro (Guzmán, Durán, & Franco, 2009). En este contexto, el presente trabajo se ha orientado hacia la formulación de un modelo estadístico de supervivencia por medio de la estimación Kaplan y Meier, seguido de un modelo de regresión de Cox mediante una serie de variables significativas, que pueda determinar los principales factores a considerar al momento de presentarse la deserción por parte de los estudiantes de Ingeniería.

REFERENTES TEORICOS

Desde hace algún tiempo esta problemática ha sido analizada a profundidad en varias Universidades, en casos generales como el de la Universidad Nacional de Colombia (sede Medellín) con el desarrollo de un modelo por medio de “tablas de vida dinámicas” (Ochoa, 2015), o más específicos como lo han sido:

- La Universidad Santo Tomas de Bogotá realizó un modelo en la facultad de Ingeniería Civil (Hernández, 2017) y en las facultades de Estadística y Diseño Gráfico (Ocampo, 2018).
- La Universidad Tecnológica de Pereira en donde se realizó un modelo generalizado para las carreras de dicha Universidad (Barrera, 2008).

Algo a considerar, además de las metodologías realizadas en estos proyectos, es el hecho del análisis a través de la información usada, dado que, si bien hay factores similares en todos los análisis, otros son propios del entorno en donde se estén ejecutando.

Deserción Estudiantil

Primero es necesario comprender el concepto de deserción, esta se puede entender como: “una situación a la que se enfrenta un estudiante cuando aspira y no logra concluir su proyecto educativo,

considerándose como desertor a aquel individuo que siendo estudiante de una institución de educación superior no presenta actividad académica durante dos semestres académicos consecutivos, lo cual equivale a un año de inactividad académica” (Tinto, 1982).

Dado que la anterior definición es la adoptada por el Ministerio de Educación Nacional, es de especial interés para este trabajo, al igual que las siguientes definiciones.

Tipo de deserciones con respecto al tiempo:

- **Deserción precoz:** individuo que habiendo sido admitido por la institución de educación superior no se matricula.
- **Deserción temprana:** individuo que abandona sus estudios en los primeros semestres del programa.
- **Deserción tardía:** individuo que abandona los estudios en los últimos semestres.

Tipo de deserción con respecto al espacio:

- **Deserción institucional:** caso en el cual el estudiante abandona la institución.
- **Deserción interna o del programa académico:** se refiere al alumno que decide cambiarse a otro programa que

ofrece la misma institución de educación superior.

Si bien de momento se ha dado relevancia al tipo de deserciones más comunes que se pueden encontrar a nivel educativo, otros autores denotan los factores asociados al entorno educativo. Dichos factores se pueden estudiar de dos formas: Internos y Externos a la Universidad (Salcedo, 2010). Desde estos puntos de vista, se pueden distinguir las características denotadas en la Tabla 1.

Factores asociados a la Universidad	Características
Internos	<ul style="list-style-type: none"> • Escasos recursos y defectos de organización. • Deficiencias docentes. • Falta de ayuda organizada para estudiantes (bienestar estudiantil, becas, subsidios, entre otros).
Externos	<ul style="list-style-type: none"> • Sistema educativo nacional. • Tipo de actividades y medio económico ambiental reinante es desfavorable.

Tabla 1. Factores y características asociados a la Universidad sobre la deserción estudiantil

Lo anterior expuesto puede adecuarse dependiendo de la Universidad objeto de estudio, además de poder incluir y/o excluir características que pueden llegar a ser relevantes.

Se puede llegar a la conclusión de que la deserción estudiantil es una problemática que afecta netamente a la Universidad. Sin

embargo, esta problemática se puede amplificar no solo a nivel local sino también nacional; la deserción estudiantil puede obstaculizar el incremento para la cobertura en la educación superior, además de dilatar la formación de recurso humano con capacidad profesional, que pueda ser clave para llevar un desarrollo ideal y un crecimiento económico en el país (Sánchez & Márquez, 2012).

Hasta este punto se ha dado a conocer las diversas perspectivas que pueden conllevar a que un estudiante deserte de su labor académica. Analizar cada una de estas perspectivas puede lograrse desde un punto de vista estadístico, el cual logre hacer más sencilla la tarea de identificar que variable tiene mayor prioridad en cada Universidad.

Además de esto, en diversos estudios se ha logrado usar esta información a favor para poder llevar a cabo modelos estadísticos que puedan determinar no solo el causante principal de la deserción en dicha Universidad, sino también establecer un aproximado del número de estudiantes que pueden desertar en el siguiente período académico o también llevar a cabo un análisis de supervivencia para determinar el número mínimo de estudiantes que permanecerán en la Universidad el siguiente periodo académico.

METODOLOGIA

Análisis de Supervivencia



Los análisis de supervivencia se pueden entender como un conjunto de técnicas estadísticas que permiten, a través del tiempo, el análisis y seguimiento hasta la ocurrencia de un evento de interés (Borges, 2004), en donde el tiempo de seguimiento de manera inicial hasta la ocurrencia del evento de interés se conoce como tiempo de vida o tiempo de supervivencia, y puede generarse de manera parcial o completa.

Para el presente trabajo, el tiempo de supervivencia se define como: el tiempo que un estudiante universitario permanece en la carrera, medido a través de la condición de activo en el sistema de la Universidad, y el evento de interés es la deserción medida a partir del momento en que el estudiante posee varios semestres con una condición de inactividad en el sistema.

Es común que al momento de realizar el análisis no se cuente con la información sobre el momento en que ocurre el evento de interés desde el tiempo inicial hasta el tiempo final, lo cual denotaría que el análisis se realizaría de forma parcial. Este hecho es conocido como censura o truncamiento y es de vital importancia considerar su presencia dado que esta censura disminuye el riesgo de que el modelo estadístico proporcione estimaciones sesgadas, lo cual reduce el sentido práctico del análisis.

Los tipos de censura más comunes que ocurren al momento de realizar un análisis de supervivencia son los siguientes:

Censura por la derecha: Se presenta cuando hasta la última observación que se le hace al individuo, aún no se ha ocurrido el evento que se desea observar. En este tipo de censura algunos motivos que permiten su ejecución son los siguientes:

- Que hasta el momento de la finalización del estudio no haya ocurrido el evento.
- Que el individuo haya abandonado el estudio.
- Que haya ocurrido en el individuo otro evento que imposibilite la ocurrencia del evento que se desea observar.

Censura por la izquierda: Se presenta cuando para la primera observación que se realiza sobre el individuo ya ha ocurrido el evento que se desea observar. Dado el análisis que se esté realizando, es poco común encontrar este tipo de censura.

Censura por intervalos: Se presenta cuando se sabe que al individuo le ocurre el evento de interés en un intervalo de tiempo dado, es decir que el evento ocurre desde un tiempo inicial hasta un tiempo final.

Para el análisis de este trabajo el evento de interés es la deserción medida a partir del momento en que el estudiante posee varios semestres seguidos sin matrícula, el tipo de

cesura que se ha tenido en cuenta es el de censura por derecha.

$$H(t) = \int_0^t h(s) ds = -\ln [S(t)] \quad (3)$$

Definiciones generales

La función de supervivencia se define como la probabilidad de que una persona sobreviva (no le ocurra el evento de interés) al menos hasta el tiempo t . Sea T una variable aleatoria continua positiva (o no negativa) con función de densidad de probabilidad $f(t)$ y función de distribución acumulada $F(t) = P(T \leq t)$, la función de supervivencia $S(t)$ se define como $S(t) = 1 - F(t) = P(T > t)$, que representa la probabilidad de que se produzca el evento en el tiempo mayor que t .

De modo que para funciones continuas:

$$H(t) = \int_0^t \frac{dF(s)}{S(s)} ds \quad (4)$$

Y, por lo tanto, la función de supervivencia se puede entender como:

$$S(t) = \exp \left[- \int_0^t h(s) ds \right] = \exp [-H(t)] \quad (5)$$

Esta función es apropiada para determinar distribuciones de riesgo utilizando variables cualitativas, para describir las formas como el evento cambia con el tiempo.

Función de Riesgo

La función de riesgo $h(t)$ representa el riesgo instantáneo de que un evento ocurra en un intervalo de tiempo, dado que el individuo ha sobrevivido hasta el tiempo t . De esta forma, la función de riesgo se define como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (1)$$

Así, se puede deducir que la función de riesgo:

$$h(t) = \frac{f(t)}{S(t)} \quad (2)$$

Con $h(t) \geq 0$. En donde la función de riesgo acumulada $H(t)$:

Estimación de la función de supervivencia

Debido a la existencia de datos censurados, la función de supervivencia no puede ser generada directamente, por lo tanto, se hace necesario el uso de estimadores.

Existen diversos métodos para obtener la función de supervivencia a partir de dichos estimadores, entre los más conocidos son los basados en tablas de vida, en donde se incluye el estimador de Kaplan y Meier (1958), el cual es el estimador de la función de supervivencia más utilizado dado que hace uso de la información con la que se dispone para realizar la estimación de la función de supervivencia, es decir que

puede trabajar a partir de casos donde se presente censura e incluso en donde no se presente. También se puede resaltar que el modelo en cierto sentido es más práctico dado que los tiempos de observación contribuyen a la estimación de la función de supervivencia.

El estimador de Kaplan y Meier que es utilizado para el caso en que se presente censura por la derecha es el siguiente:

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \left[\frac{r(t_i) - d(t_i)}{d(t_i)} \right] \quad (6)$$

Donde $d(t_i)$ representa el número total de ocurrencia en el momento t_i y $r(t_i)$ representa el número de unidades en riesgo, justo antes del tiempo t_i .

Varianza del estimador de Kaplan y Meier

La varianza del estimador de la fórmula (6) se puede obtener mediante la fórmula de Greenwood (1926):

$$\hat{V}[\hat{S}_{KM}(t)] = \hat{S}_{KM}^2 \sum_{t_i \leq t} \frac{d(t_i)}{r(t_i)[r(t_i) - d(t_i)]} \quad (7)$$

Error estándar del estimador de Kaplan y Meier

Por medio de este se puede obtener el intervalo de confianza para el estimador de Kaplan y Meier con un nivel de

significancia del 5%, mediante la siguiente fórmula:

$$\hat{S}_{KM}(t) \pm 1.96 [EE(\hat{S}_{KM}(t))] \quad (8)$$

Donde $EE(\hat{S}_{KM}(t))$ representa el error estándar del estimador de Kaplan y Meier.

Comparación de las funciones de supervivencia

Al momento de considerar las curvas de supervivencia entre dos o más funciones se tienen diversos métodos para poder llevarlo a cabo. Entre los más conocidos se encuentran: el test de Mantel y Haenzel (mejor conocido como test de los rangos de logaritmos o Test Log-Rank) (Mantel, 1966), o el Test de Peto y Peto (1972). Sin embargo, una forma de considerar estos dos métodos fue propuesto por Harrington y Fleming (1982). Para dichos métodos se trabaja bajo una prueba de hipótesis, donde la hipótesis nula describe la igualdad entre las funciones de supervivencia, el objetivo es el de rechazar dicha hipótesis con un determinado nivel de significancia, para poder concluir que existe una diferencia entre las funciones y por lo tanto existe por lo menos una de estas que presentan un mayor riesgo de muerte a comparación de las otras. Dicha prueba se puede generar por medio de un estadístico Chi cuadrado con $n - 1$ grados de libertad, donde n representa el número de factores de la variable.

Modelo para el análisis de supervivencia

El modelo de regresión de Cox fue propuesto por David Cox (1972), es el modelo de regresión más utilizado para datos de supervivencia. A diferencia de otros modelos, el modelo de regresión de Cox supone que existe un conjunto de covariables $X = (X_1, X_2, \dots, X_K)^T$, que influyen en el comportamiento del tiempo de ocurrencia de los eventos.

Este modelo trabaja bajo la función de riesgo $h[t|x_i(t)]$ que define el momento t dado un individuo con una característica en el tiempo t , $x_i(t)$. Dicha función de riesgo es de la forma:

$$h[t|x_i(t)] = h_0(t)[\exp(\beta'x_i(t))] \quad (9)$$

El modelo de la fórmula (9) se establece como un modelo semiparamétrico dado los componentes de este:

- 1) La expresión “ $\exp(\beta'x_i(t))$ ” se entiende como la parte paramétrica del modelo, y define el puntaje de riesgo (conocido como Risk Score), donde β es el vector de parámetros que miden el efecto de las covariables.
- 2) La expresión “ $h_0(t)$ ” se entiende como la parte no paramétrica del modelo, y define la función de riesgo base.

Contraste de hipótesis para el modelo de regresión de Cox

Para verificar la significancia de las covariables en el modelo de Cox, existen tres contrastes de hipótesis que permiten realizar dicha verificación:

- Test de razón de verosimilitud.
- Test de Wald.
- Test de puntajes.

El objetivo mediante los anteriores contrastes es el de rechazar la hipótesis nula que representa la no significancia del modelo.

Supuesto de riesgos proporcionales

El modelo de regresión de Cox también es conocido como el modelo de supuestos de riesgos proporcionales, dado que, las observaciones al ser independientes, el cociente entre el riesgo para dos individuos con el mismo vector de covariables es constante en el tiempo, y por lo tanto la proporcionalidad de los riesgos para los dos individuos no debe variar con respecto al tiempo. Por lo tanto, dicho supuesto se establece como:

$$\frac{h(t|x_i(t))}{h(t|x_j(t))} = \frac{h_0(t)[\exp(\beta'x_i(t))]}{h_0(t)[\exp(\beta'x_j(t))]} \quad (10)$$

$$= \frac{\exp(\beta'x_i(t))}{\exp(\beta'x_j(t))}$$

El objetivo de dicho supuesto es el de no rechazar la hipótesis nula que representa el cumplimiento del supuesto de riesgo proporcional, lo que significa el vector de

covariables es constante con respecto al tiempo.

Modelo de Cox estratificado

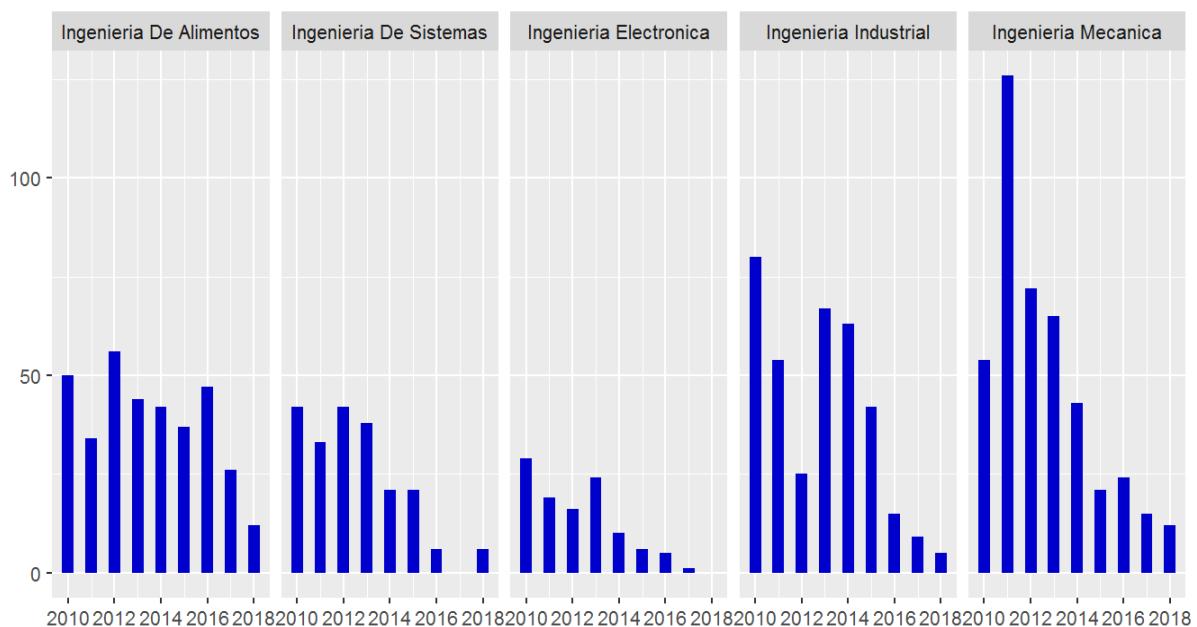
Representa una extensión del modelo de regresión de Cox, y es usado principalmente cuando en el modelo original de Cox no se cumple con el supuesto de riesgo proporcional en alguna de las covariables. La idea es estratificar aquellas covariables que no cumplen con el supuesto de riesgo proporcional, de esta forma al generar la comprobación del supuesto se permitirá comprobar que las covariables son constantes a través del tiempo. La única desventaja que presenta este proceso es el de no generarse ningún parámetro β con el que se pueda demostrar la estimación sobre el efecto de la covariable ya estratificada (Borges, 2002).

RESULTADOS

Análisis Descriptivo

Para el análisis de este estudio se ha considerado la información suministrada por la oficina de Registraduría de la Universidad, tomando en cuenta a los estudiantes de la facultad de Ingeniería que ingresaron desde el primer semestre del año 2010 hasta el segundo semestre del año 2018. En esta ocasión, se ha contado con las siguientes variables de estudio:

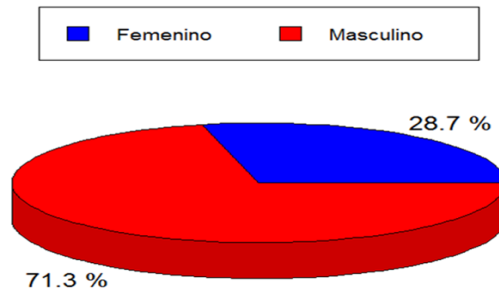
- Carrera: Representa la carrera profesional de Ingeniería a la cual está inscrita el estudiante. Se cuenta con las siguientes carreras: Ingeniería de Alimentos, Ingeniería de Sistemas, Ingeniería Electrónica, Ingeniería Industrial e Ingeniería Mecánica.



Gráfica 1. Deserción por año para cada carrera.

- Sexo: Representa el género del estudiante (F para estudiantes de género Femenino y M para estudiantes de género Masculino).
- Estado: Representa el estado financiero del estudiante (Paz y Salvo para estudiantes que no poseen deudas con la Universidad y Moroso para estudiantes que si poseen deudas con la Universidad).

Electrónica presento una deserción menor en comparación a las demás.

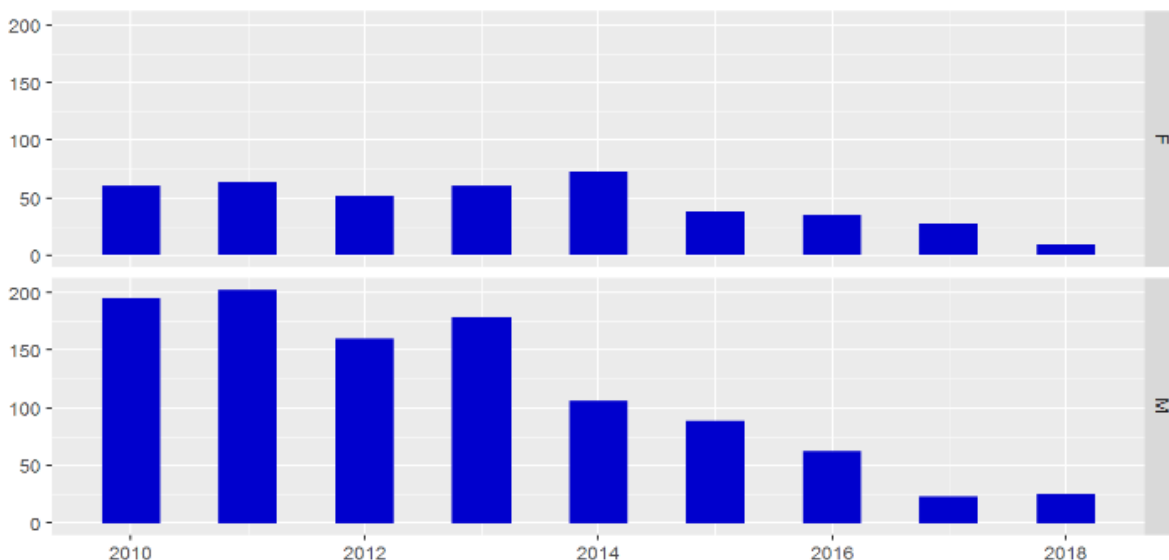


Gráfica 2. Porcentaje de estudiantes desertados por sexo.

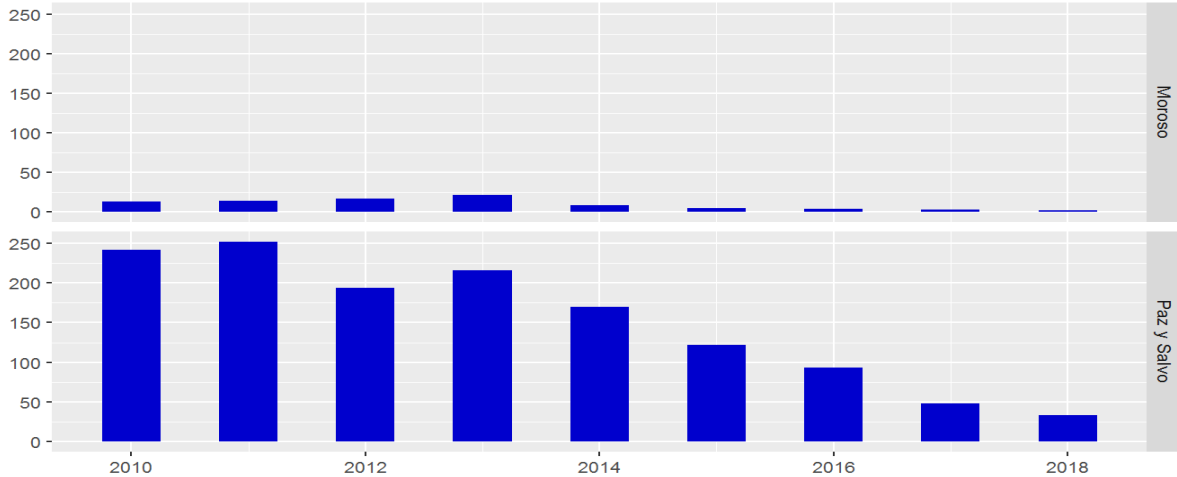
El análisis descriptivo para cada una de las variables anteriormente mencionadas se ha realizado mediante el lenguaje de programación R, al igual que, como se observará más adelante, el análisis de supervivencia.

Las Gráficas 2 y 3 permiten determinar que la deserción fue más predominante en estudiantes del género masculino entre todas las ingenierías, denotando que en el 2011 se presentó una mayor tasa de deserción para este género.

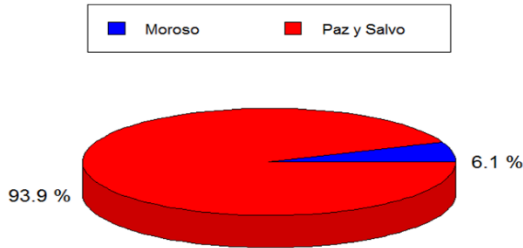
La Gráfica 1 permite visualizar que durante el año 2011 se presentó una mayor deserción en una carrera específica la cual fue Ingeniería Mecánica, al igual que se puede observar que la carrera de Ingeniería



Gráfica 3. Estudiantes desertados por año y por sexo.



Gráfica 5. Estudiantes desertados por año y por estado.



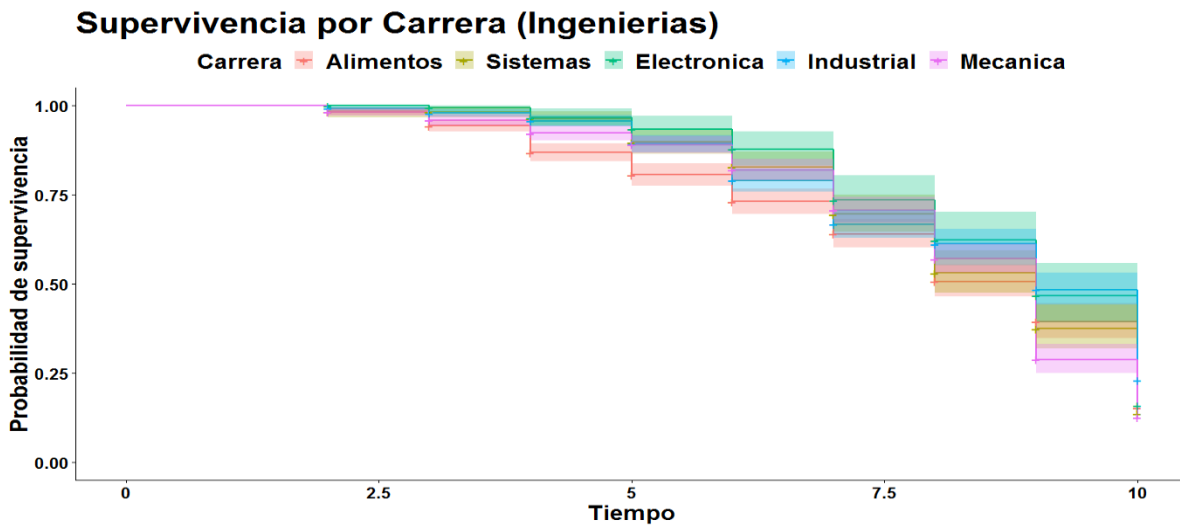
Gráfica 4. Porcentaje de estudiantes desertados por estado.

Las Gráficas 4 y 5 permiten determinar que la deserción fue mucho más predominante

en estudiantes que se encontraban a Paz y Salvo, presentando de igual forma que en el año 2011 se presentó una mayor deserción en estudiantes con esta condición financiera hacia la universidad.

Análisis de Supervivencia

El análisis de supervivencia se ha realizado considerando las variables anteriormente expuestas, de forma tal que se pueda detectar cuales factores son más



Gráfica 6. Curvas de supervivencia por Carreras.

influyentes en la deserción estudiantil de la Universidad. Para iniciar se ha realizado un análisis mediante el estimador Kaplan y Meier para cada una de las variables, generando sus debidas curvas de supervivencia y el test de Log-Rank que permita identificar si la función de supervivencia para cada factor de las variables es igual o no.

Análisis por Carreras

La Gráfica 6 indica las curvas de supervivencia para cada una de las carreras de ingeniería, donde se puede evidenciar que no hay demasiada diferencia entre algunas de las curvas. Sin embargo, teniendo en cuenta el *p-valor* generado a través del test de Log-Rank en la Tabla 2 hay evidencia suficiente para confirmar que las curvas de supervivencia son diferentes con un nivel de significancia del 5%, lo que asegura que la variable Carreras

presenta un factor influyente al evaluar la deserción entre las carreras de ingeniería.

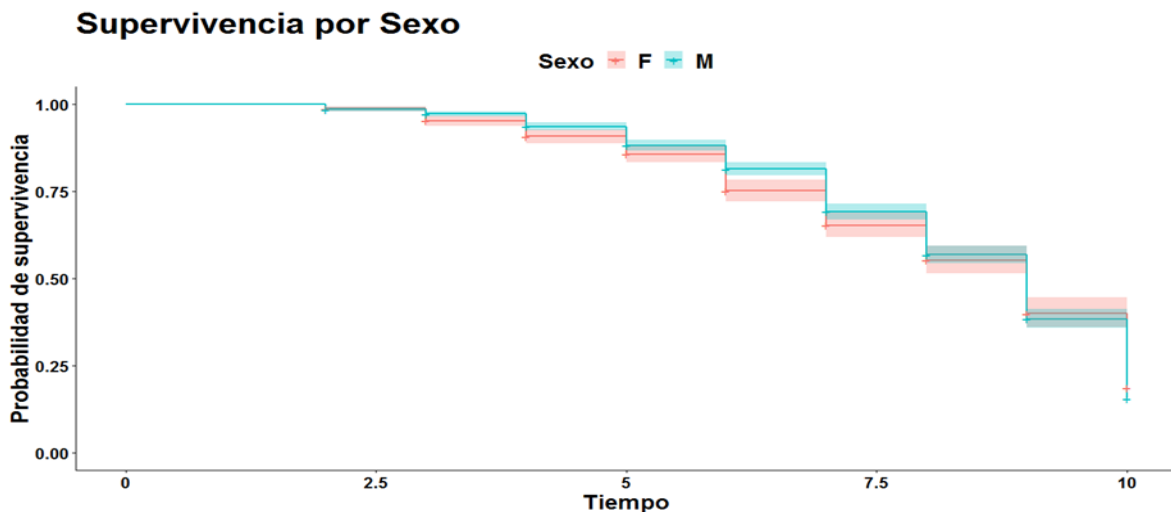
Carrera de Ingeniería	N	Observado	Esperado	(O-E) ² /E	(O-E) ² /N
Alimentos	705	348	310	4.748	7.678
Sistemas	338	209	204	0.105	0.159
Electrónica	195	110	127	2.359	3.394
Industrial	701	360	419	8.301	15.303
Mecánica	675	432	399	2.786	4.918

Chi cuadrado = 23.9 con 4 grados de libertad, p-valor = 9e-05

Tabla 2. Test Log-Rank para la variable Carrera.

Análisis por Sexo

La Gráfica 7 indica que no hay demasiada diferencia entre las curvas de supervivencia para los géneros Femenino y Masculino. De igual forma, al verificar el *p-valor* de la Tabla 3 no hay evidencia suficiente para confirmar que las curvas de supervivencia son diferentes con un nivel de significancia del 5%, lo que asegura que la variable Sexo no presenta un factor



Gráfica 7. Curvas de supervivencia por Sexo.

influyente al evaluar la deserción entre las carreras de ingeniería.

Sexo	N	Observado	Esperado	$\frac{(O-E)^2}{E}$	$\frac{(O-E)^2}{N}$
F	839	419	411	0.1708	0.305
M	1775	1040	1048	0.0669	0.305

Chi cuadrado = 0.3 con 1 grado de libertad, p-valor = 0.6

Tabla 3. Test Log-Rank para la variable Sexo.

Análisis por Estado

La Gráfica 8 indica que de manera inicial no hay demasiada diferencia entre las curvas de supervivencia para los estados Moroso y Paz y Salvo, sin embargo, al momento de finalizar la gráfica se evidencia que los estudiantes que se encuentran a paz y salvo tienen una mayor probabilidad de sobrevivir (continuar con sus estudios). De igual forma, al verificar el *p-valor* de la Tabla 4 hay evidencia

suficiente para confirmar que las curvas de supervivencia son diferentes con un nivel de significancia del 5%, lo que asegura que la variable Estado presenta un factor influyente al evaluar la deserción entre las carreras de ingeniería, esto podría indicar que aquellos estudiantes que poseen alguna deuda con la Universidad presentan mayor riesgo de deserción.

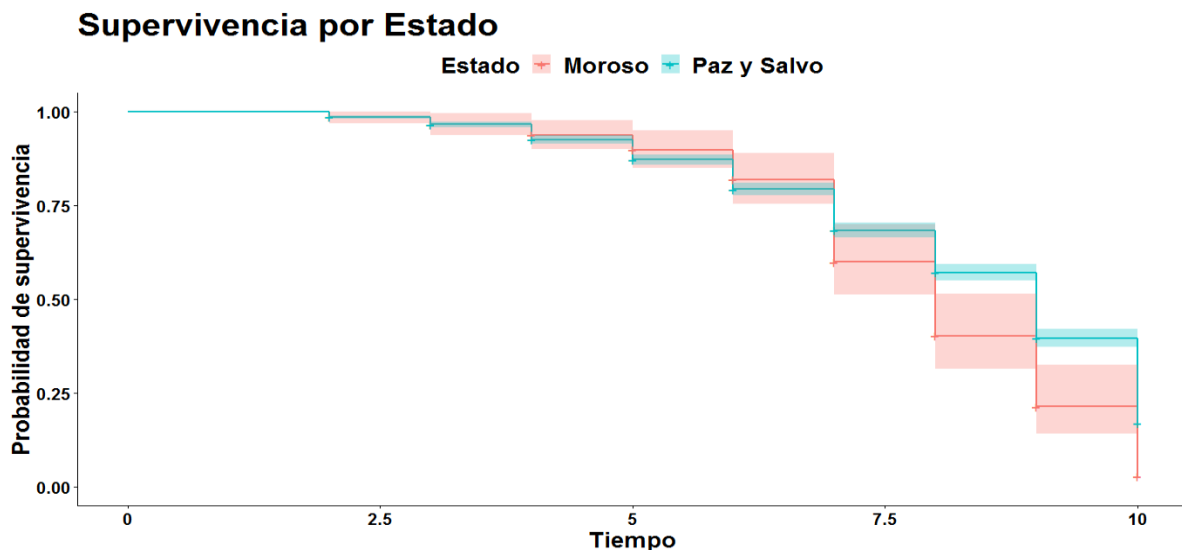
Estado	N	Observado	Esperado	$\frac{(O-E)^2}{E}$	$\frac{(O-E)^2}{N}$
Moroso	153	89	65.4	8.539	11.1
Paz y Salvo	2461	1370	1393.6	0.401	11.1

Chi cuadrado = 11.1 con 1 grado de libertad, p-valor = 9e-04

Tabla 4. Test Log-Rank para la variable Estado.

Modelo de Regresión de Cox

Para ejecutar el modelo de regresión de Cox se ha considerado las covariables de



Gráfica 8. Curvas de supervivencia por Estado.

Carrera, Sexo y Estado, teniendo en cuenta principalmente aquellas que, mediante el análisis de supervivencia, presentaron una influencia en cuanto al nivel de deserción. El análisis se ejecuta inicialmente para cada una de las covariables, y posteriormente se ejecuta un análisis general, considerando únicamente los factores significativos para el modelo. Finalmente se comprueba si el modelo cumple con el supuesto de riesgos proporcionales.

Modelo para la covariable Carrera

Carrera de Ingeniería	coef	exp (coef)	se (coef)	z	Pr(> z)
Sistemas	-0.095	0.909	0.087	-1.088	0.276
Electrónica	-0.264	0.767	0.109	-2.416	0.015 *
Industrial	-0.270	0.763	0.075	-3.586	0.000335 ***
Mecánica	-0.036	0.964	0.072	-0.507	0.611

Test de razón de verosimilitud = 18.7 con 4 gl, p-valor = 9e-04
Test de Wald = 18.34 con 4 gl, p-valor = 0.001
Test de puntajes = 18.43 con 4 gl, p-valor = 0.001

Tabla 5. Modelo de Regresión de Cox para la covariable Carreras.

La Tabla 5 evidencia el modelo de regresión para la covariable Carreras, y que, con un nivel de significancia del 5%, dicha covariable es significativa según los tres criterios (razón de verosimilitud, Wald y puntajes), por lo tanto, se puede decir que:

- Las carreras de Ingeniería Electrónica e Ingeniería Industrial son las más significativas para el modelo, a un nivel del 5%.
- El riesgo de que un estudiante deserte de sus estudios sabiendo que este pertenece a la carrera de Ingeniería Electrónica es 0.767 veces menor que un estudiante de otra ingeniería.
- El riesgo de que un estudiante deserte de sus estudios sabiendo que este pertenece a la carrera de Ingeniería Industrial es 0.763 veces menor que un estudiante de otra ingeniería.
- Al no evidenciarse la carrera de Ingeniería de Alimentos en el modelo, esta carrera puede que presente un mayor riesgo de deserción, a comparación de las demás ingenierías.

Modelo para la covariable Sexo

Sexo	coef	exp (coef)	se (coef)	z	Pr(> z)
M	-0.028	0.972	0.057	-0.488	0.626

Test de razón de verosimilitud = 0.24 con 1 gl, p-valor=0.6
Test de Wald = 0.24 con 1 gl, p-valor=0.6
Test de puntajes = 0.24 con 1 gl, p-valor=0.6

Tabla 6. Modelo de Regresión de Cox para la covariable Sexo.

La Tabla 6 evidencia que el modelo de regresión para la covariable Sexo, con un nivel de significancia del 5%, no es significativo para ninguno de los tres criterios, y por lo tanto se confirma lo

evaluado mediante el análisis de supervivencia para dicha covariable, sobre que la variable Sexo no es influyente al momento de analizar la deserción en estudiantes de ingeniería.

Modelo para la covariable Estado

Estado	coef	exp (coef)	se (coef)	z	Pr(> z)
Paz y Salvo	-0.326	0.721	0.109	-2.97	0.00289***

Test de razón de verosimilitud = 8.08 con 1 gl, p-valor = 0.004
Test de Wald = 8.88 con 1 gl, p-valor = 0.003
Test de puntajes = 8.96 con 1 gl, p-valor = 0.003

Tabla 7. Modelo de Regresión de Cox para la covariable Estado.

La Tabla 7 evidencia que el modelo de regresión para la covariable Estado, con un nivel de significancia del 5%, es significativo para los tres criterios. Además de que a partir de dicho modelo se puede decir que el riesgo de que un estudiante deserte de sus estudios sabiendo que este se encuentra a paz y salvo con la Universidad es 0.721 veces menor que un estudiante que se encuentra en mora.

Los anteriores modelos permiten concluir que un modelo general puede desarrollarse a partir de las covariables Carreras y Estado, las cuales fueron significativas en sus respectivos modelos individuales.

Modelo general

Covariables	coef	exp (coef)	se (coef)	z	Pr(> z)
Ingeniería Electrónica	-0.272	0.761	0.109	-2.484	0.012984*
Ingeniería Industrial	-0.271	0.762	0.075	-3.596	0.000324***
Paz y Salvo	-0.335	0.715	0.109	-3.062	0.002201**

Test de razón de verosimilitud = 27.22 con 5 gl, p-valor = 5e-05
Test de Wald = 27.66 con 5 gl, p-valor = 4e-05
Test de puntajes = 27.82 con 5 gl, p-valor = 4e-05

Tabla 8. Modelo general de Regresión de Cox.

La Tabla 8 permite comprobar que el modelo de regresión general con las variables Carreras y Estado es significativo para los tres criterios, a un nivel del 5%.

Supuesto de riesgos proporcionales

Para determinar si las covariables anteriormente analizadas pueden permanecer constantes con respecto al tiempo se ha puesto a prueba esto mediante el supuesto de riesgos proporcionales.

Covariables	Chi cuadrado	Grados de libertad	p-valor
Carrera	23.73	4	9.1e-05
Sexo	7.86	1	0.00504
Estado	2.56	1	0.10991
GLOBAL	26.87	6	0.00015

Tabla 9. Validación del supuesto de riesgos proporcionales.

Considerando los *p-valores* generados en la Tabla 9 se puede concluir, con un nivel de significancia del 5%, que el modelo no cumple con el supuesto de riesgos proporcionales, tanto a nivel global como por variable (exceptuando la covariable Estado). Esto significa que existe alguna tendencia en las covariables con respecto al tiempo, y por lo tanto se puede afirmar que estas no son constantes. Esto último se puede corregir realizando una estratificación del modelo de Regresión de Cox.

Modelo de Cox estratificado

La estratificación de un modelo de regresión de Cox se efectúa siempre y cuando no se cumpla con el supuesto de riesgos proporcionales. Para esto se inicia estratificando aquella covariable que representa una significancia menor a las demás, para este caso la covariable Carrera, en donde al estratificarla se contara a cada una de los factores (las carreras de Ingeniería) como estratos. De manera inicial, se empieza realizando los modelos de regresión de Cox estratificados por Carreras para las covariables, para posteriormente dar validez al supuesto de riesgos proporcionales.

Segundo modelo para la covariable Sexo

Sexo	coef	exp (coef)	se (coef)	z	Pr(> z)
M	0.00250	1.002	0.064	0.039	0.969

Test de razón de verosimilitud = 0 con 1 gl, p-valor = 1
Test de Wald = 0 con 1 gl, p-valor = 1
Test de puntajes = 0 con 1 gl, p-valor = 1

Tabla 10. Modelo de Regresión de Cox para la covariable Sexo estratificando por Carreras.

La Tabla 10 evidencia que el modelo de regresión para la covariable Sexo, estratificando las carreras y con un nivel de significancia del 5%, no es significativa para ninguno de los tres criterios, y por lo tanto se reafirma lo evaluado mediante el modelo generado a partir de la Tabla 6.

Segundo modelo para la covariable Estado

Estado	coef	exp (coef)	se (coef)	z	Pr(> z)
Paz y Salvo	-0.340	0.711	0.109	-3.103	0.00191 **

Test de razón de verosimilitud = 8.74 con 1 gl, p-valor = 0.003
Test de Wald = 9.63 con 1 gl, p-valor = 0.002
Test de puntajes = 9.72 con 1 gl, p-valor = 0.002

Tabla 11. Modelo de Regresión de Cox para la covariable Estado estratificando por Carreras.



La Tabla 11 evidencia que el modelo de regresión para la covariable Estado, estratificando las carreras y con un nivel de significancia del 5%, es significativo para los tres criterios, reafirmando lo evaluado mediante el modelo generado a partir de la Tabla 7.

Supuesto de riesgos proporcionales

Covariables	Chi cuadrado	Grados de libertad	p-valor
Sexo	0.943	1	0.33
Estado	3.061	1	0.08
GLOBAL	3.792	2	0.15

Tabla 12. Validación del supuesto de riesgos proporcionales estratificando por Carreras.

Considerando los p-valores generados en la Tabla 12 se puede concluir, con un nivel de significancia del 5%, que el modelo cumple con el supuesto de riesgos proporcionales, tanto a nivel global como por covariable, lo que demuestra que estas son constantes con respecto al tiempo.

CONCLUSIONES

Por medio de un análisis de supervivencia haciendo uso del estimador de Kaplan y Meier, y un modelo de Regresión de Cox aplicado a la facultad de Ingeniería de una Universidad de Bogotá, Colombia se puede determinar que:

- Existe un mayor riesgo de deserción estudiantil en la carrera de Ingeniería de Alimentos, esto claro

comparándola con las demás ingenierías, dado que de igual forma las carreras de Ingeniería de Sistemas y Mecánica también poseen un riesgo significativo de deserción estudiantil.

- La deserción es mayor en estudiantes de ingeniería que se encuentran en estado “Moroso”, por lo que se hace necesario que la Universidad tome planes de acción frente a esta problemática, dado que además de que el estudiante ya no se encuentre en la institución es fundamental evaluar la deuda que este ha dejado, y que puede ser significativa para el estado financiero de la Universidad.
- En estudiantes de ingeniería, el sexo no es algo influyente al momento de que estos deserten de sus estudios. Esto se ha podido demostrar a lo largo del trabajo, en donde se ha visto que la variable no es significativa bajo ningún análisis.
- El hecho de haber estratificado la covariable Carreras se justifica al realizar el supuesto de riesgos proporcionales, donde se ha demostrado que tanto esta como las demás covariables eran dependientes del tiempo; una vez realizada la estratificación se demostró que las covariables cumplían con el supuesto, lo que permite determinar que ahora son constantes con respecto al tiempo. Esto demuestra que el análisis es más funcional si se toma a cada carrera de ingeniería como un estrato diferente

entre estos, y que la interpretación real para las otras covariables (en este caso Sexo y Estado) se da para cada uno de los estratos.

RECOMENDACIONES

- Gracias a la información suministrada por parte de la Universidad esta investigación pudo ser llevada a cabo. Sin embargo, esta información puede llegar a ser mejor ordenada y recopilada para estudios posteriores, dado que se observaron falencias que posee actualmente la base de datos con la información de los estudiantes, por lo que se hace necesario un mejor control y orden de dicha información.
- Si bien por medio del análisis anteriormente expuesto se pueden llegar a conclusiones interesantes sobre la deserción en estudiantes de ingeniería para la Universidad, estudios posteriores pueden llegar a generar resultados más precisos por medio de otras variables de interés, como lo pueden ser la Edad de los estudiantes, el Estrato socioeconómico de estos, la Condición laboral, el Índice de asignaturas aprobadas y/o reprobadas, entre otras.

REFERENCIAS BIBLIOGRÁFICAS

- Barrera, M. (2008). *Análisis de supervivencia aplicado al problema de la deserción estudiantil en la Universidad Tecnológica de Pereira*. Pereira, Colombia: Universidad Tecnológica de Pereira.
- Borges, R. (2002). *Análisis de Supervivencia aplicado a un caso: Diálisis peritoneal en el Hospital Clínico Universitario de Caracas y Hemodiálisis en el Hospital de Clínicas Caracas, 1980-2000*. Mérida, Venezuela: Universidad de Los Andes.
- Borges, R. (2004). *Análisis de Supervivencia básico utilizando el lenguaje R*. Caracas, Venezuela: Universidad Central de Venezuela. Escuela de Estadística y Ciencias Actuariales.
- Cox, D. (1972). *Regression models and life tables (with discussion)*. Journal of the Royal Statistical Society: Series B. Vol. 34, N° 2: 187 – 220.
- Gonzalez, L. (2015). *Estrategias para la permanencia en Educación Superior: Experiencias Significativas*. Ministerio de Educación Nacional.
- Greenwood, M. (1926). *The natural duration of cancer*. Londres: Her Majesty's Stationery Office.: Reports on Public Health and Medical Subjects 33, 1–26. .
- Guzmán, C., Durán, D., & Franco, J. (2009). *Deserción estudiantil en la*

- educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos para la prevención.* Bogotá, Colombia: Ministerio de Educación Nacional.
- Harrington, D., & Fleming, T. (1982). *A class of rank procedures for censored survival data.* *Biometrika*, 69: 553-566.
- Hernández, N. (2017). *Un modelo de deserción en los estudiantes de la facultad de Ingeniería Civil de la Universidad Santo Tomás.* Bogotá, Colombia: Universidad Santo Tomás.
- Kaplan, E., & Meier, P. (1958). *Nonparametric estimation from incomplete observations.* *Journal of the American Statistical Association*. Vol. 53, N° 282: 457 – 481.
- Mantel, N. (1966). *Evaluation of survival data and two new rank order statistics arising in its consideration.* *Cancer Chemotherapy Reports*, 50: 163-170.
- Ocampo, A. (2018). *Modelo de sobrevida aplicado a deserción estudiantil de las Facultades de Estadística y Diseño gráfico de la Universidad Santo Tomás.* Bogotá, Colombia: Universidad Santo Tomás.
- Ochoa, C. (2015). *El modelo Lee-Carter para estimar y pronosticar mortalidad: Una aplicación para Colombia.* Medellín, Colombia: Universidad Nacional de Colombia.
- Peto, R., & Peto, J. (1972). *Asymptotically efficient rank invariant test procedures (with discussion).* *Journal of the Royal Statistical Society: Series A*, 135: 195-206.
- Salcedo, A. (2010). *Deserción universitaria en Colombia.* Bogotá, Colombia: Universidad Militar Nueva Granada. *Revista Academia y Virtualidad*.
- Sánchez, F., & Márquez, J. (2012). *La Deserción en la Educación Superior en Colombia durante la Primera Década del Siglo XXI: ¿Por qué ha aumentado tanto?* Bogotá, Colombia: Universidad de los Andes: Facultad de Economía.
- Tinto, V. (1982). *Dropout from Higher Education: A Theoretical Synthesis of Recent Research.* *Journal of Higher Education*. Vol. 3, N° 6: 687 – 700.

