



LOS LIBERTADORES  
FUNDACIÓN UNIVERSITARIA

# MODELO DE SUPERVIVENCIA APLICADO A LA DESERCIÓN ESTUDIANTIL EN LA FACULTAD DE INGENIERÍA DE UNA UNIVERSIDAD EN LA CIUDAD DE BOGOTÁ EN LA COHORTE 2017-1 A 2021-1

## SURVIVAL MODEL APPLIED TO STUDENT DESERTION IN THE FACULTY OF ENGINEERING OF A UNIVERSITY IN THE CITY OF BOGOTÁ IN THE COHORT 2017-1 TO 2021-1

Carlos Eduardo Pineda Ruiz  
cepinedar@libertadores.edu.co

Junio 2021

### RESUMEN

El siguiente trabajo busca analizar la deserción estudiantil de la facultad de ingeniería de una universidad privada en la ciudad de Bogotá a través de un modelo de supervivencia de Kaplan–Meier y Box Cox, para ello se analizó la matrícula del año 2017-1 de la universidad y se formó una nueva base de datos únicamente con los estudiantes matriculados en la Facultad de Ingeniería, posteriormente se realizó seguimiento a la matrícula de los estudiantes hasta el primer semestre del año 2021-1. En el análisis de supervivencia la variable respuesta es el tiempo que transcurre entre un evento inicial (que determina la inclusión del individuo en el estudio) y un evento final (genéricamente llamado falla) que ocurre cuando el individuo presenta la característica para terminar el estudio (muerte, alta de la enfermedad, deserción, etc.). En este tipo de estudios de cohorte puede ocurrir que algún individuo lo abandone antes de que le ocurra el evento de interés, registrándose sólo información parcial (censura) sobre la variable de interés (tiempo de falla). Con esta información recopilada se quiere establecer las variables que inciden al momento en el cual un estudiante decide desertar de la Facultad de Ingeniería y si existe diferencias entre la deserción por sexos, edad y por carreras en dicha facultad.

**Palabras clave:** Deserción, Análisis de Supervivencia, Censura, Kaplan-Meier, Box Cox.

### ABSTRAC

The following work seeks to analyze the student dropout from the engineering faculty of a private university in the city of Bogotá through a survival model of Kaplan – Meier and Box Cox, for this purpose the enrollment of the year 2017-1 of the university and a new database was formed only with students enrolled in the Faculty of Engineering, subsequently the enrollment of students was followed up until the first semester of the year 2021-1.

In the survival analysis, the response variable is the time that elapses between an initial event (which determines the inclusion of the individual in the study) and a final event (generically called failure) that occurs when the individual presents the characteristic to finish the study ( death, discharge from illness, desertion, etc.). In this type of cohort studies, it may happen that an individual abandons it before the event of interest occurs, registering only partial information (censorship) on the variable of interest (failure time). With this information collected, we want to establish the variables that affect the monument in which a student decides to drop out of the Faculty of Engineering and if there are differences between the desertion by sex, age and by careers in said faculty.

**Keywords: Desertion, Survival Analysis, Censorship, Kaplan-Meier, Box Cox.**

## INTRODUCCIÓN

En los últimos años la deserción estudiantil en el sistema de educación superior en Colombia se ha convertido en un problema que ha llamado la atención del Gobierno Nacional y las Universidades tanto privadas como públicas. El Ministerio de Educación Nacional MEN entiende por desertor, a aquel estudiante que no termina su formación académica. “De acuerdo con lo anterior, y al conjugar las definiciones de deserción dadas por Tinto (1982) y Giovagnoli (2002), se puede entender la deserción 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”. Según el tiempo en que la deserción ocurra esta se puede clasificar de la siguiente manera de acuerdo al MEN así:

- i) Deserción precoz: es aquel individuo que habiendo sido admitido por la institución de educación superior no se matricula.
- ii) Deserción temprana: es el individuo que abandona sus estudios en los primeros semestres del programa.
- iii) Deserción tardía: es el individuo que abandona los estudios en los últimos semestres.

La deserción es un fenómeno educativo bastante complejo, ya que involucra una gran cantidad de variables que van desde las de tipo socio-económico, pasando por situaciones de tipo individual, familiar, institucional, e inclusive la falta de vocación hacia la carrera elegida. Un estudiante que deserta, no solo afecta su núcleo social y familiar debido a la frustración que trae consigo la deserción, debido a que se pierde la posibilidad de acceder a mejores empleos y una mejor calidad de vida del estudiante y su entorno, sino también se ve afectada la sociedad y el país debido a que se pierde mano de obra calificada, también se afectan las universidades debido a la baja de sus matrículas haciendo que sus recursos económicos se vean disminuidos. El gobierno nacional también se afecta porque los recursos que se invierten en la educación de este estudiante no se pueden recuperar en el caso que la deserción se de en la universidad pública. Por esta razón el Gobierno Nacional y las universidades públicas y privadas están encaminado acciones para disminuir la deserción estudiantil, ya que de nada sirve aumentar la cobertura en educación si los índices de deserción estudiantil no se reducen.

En cuanto a la problemática de la deserción en educación superior se han elaborado un sinnúmero de artículos, y trabajos en los cuales se describen, caracterizan y analizan las causas de esta, desde diferentes ópticas. Sin embargo, son pocos los estudios que se han realizado en los cuales se aborde la deserción a través de la construcción de modelos matemáticos y estadísticos que permitan describir este fenómeno y de paso permitir que las universidades puedan diseñar estrategias de retención de estudiantes y de esta forma mitigar el abandono de los estudiantes de la educación superior. Por esa razón en este trabajo se busca analizar la deserción a través de un modelo de Kaplan-Meier y Box Cox que permita establecer cuál es el periodo en el cual los estudiantes de la facultad de ingeniería de esta universidad tienen mayor probabilidad de desertar y si el sexo y la edad tienen alguna incidencia en dicha deserción

## REFERENTES TEÓRICOS

El gobierno de Nacional y las universidades tienen el reto de aumentar la cobertura de educación superior en el siglo XXI, ya que actualmente tan solo el 53% de la población que está en edad de estudiar, es decir, jóvenes entre 17 y 21 años, se encuentra matriculados en programas de pregrado, pero para ello el gobierno y las universidades deben primero reducir la deserción en la educación superior. En un informe del MEN la población que culmina sus estudios en educación básica y media e ingresa inmediatamente a la educación superior es cerca del 35%. (LEE, 2020). Allí se establece que este porcentaje está por encima de la media de ingreso de estudiantes de educación básica en Latinoamérica. En este informe se muestra cómo han evolucionado las matriculas de estudiantes de pregrado desde el 2000-1 hasta el año 2018-1, en el cual se establece que durante estos años la matrícula ha aumentado durante este periodo, sin embargo, en términos porcentuales se refleja una disminución año tras año. Aunque en términos absolutos la matrícula de estudiantes de pregrado ha aumentado durante estos años, el gobierno nacional y las universidades enfrentan el desafío de disminuir la deserción escolar en la educación superior, la cual ronda el 50% a nivel nacional.

Con estos datos el gobierno se puso la meta en el 2009 de disminuir la deserción escolar en las instituciones de educación superior al 25% para el 2019, pero lastimosamente dicha meta no se ha cumplido y todo parece indicar que para el año 2020 y 2021 la deserción aumentará. Un estudio realizado por el Ministerio de Educación Nacional MEN (MEN, 2009) establece que el factor que más incide en la deserción en la educación superior es el académico, seguido por factores financieros y socioeconómicos y en tercer lugar por factores relacionados con las instituciones universitarias y la falta de vocación por parte del estudiante. Por esta razón el gobierno nacional en cabeza del MEN y las universidades colombianas se han dado a la tarea de analizar este fenómeno que se está presentando en la educación superior, dichos análisis se hacen desde diferentes perspectivas tales como institución, programa, y estudiantes, entre otros.

Según el MEN(2009) existen cuatro factores que son determinantes de la deserción escolar en las universidades, que son el factor individual, académico, institucional y socioeconómico y cada uno de ellos posee a su vez variables que inciden de una u otra forma en la deserción en las universidades. Por esta razón para poder disminuir la deserción el Gobierno Nacional y las universidades tanto públicas como privadas deben caracterizarla, ya que para poder aplicar políticas gubernamentales e institucionales que sean efectivas se necesita establecer el tipo de deserción, el cual es diferente en muchos casos por región, por universidad y por estudiante.

En estudios realizados por el MEN se estableció que el factor económico no es el más relevante en la deserción estudiantil y que por el contrario factores como la vulnerabilidad académica y la falta de vocación eran más relevantes al momento de abandonar las aulas de educación superior. Un aspecto crucial para disminuir la deserción en las universidades sobre todo en los primeros semestres, es la articulación entre la educación media y la educación superior, con el objetivo de cerrar la brecha académica y evitar que los nuevos estudiantes que pasan del colegio a las universidades deserten de manera temprana. En estos momentos, diferentes universidades del país han diseñado, implementado y desarrollado planes de alertas tempranas para detectar aquellos estudiantes que se encuentran en riesgo de desertar de los diferentes programas académicos, por su parte el gobierno nacional ha implementado el Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior -Spadies- el cual consolida y ordena información que permite hacer seguimiento a las condiciones académicas y socioeconómicas de los estudiantes que han ingresado a la educación superior en el país.

Sobre deserción estudiantil en la educación superior se han hecho diferentes estudios en los cuales se abordan las causas de esta, como es el caso de Zarate (2009) quien elabora un artículo cuyo propósito es comprender los motivos

por los cuales los estudiantes de la Universidad Industrial de Santander –UIS– desertan. Allí se encontró que las motivaciones significativas y decisivas a la hora de desertar obedecen a múltiples motivaciones de tipo personal, socio-económico, cultural, institucional y académico, que influyen sobre la toma de decisión de abandono por parte del estudiante.

Por su parte Hernández (2020) aborda la deserción estudiantil a través de un modelo para predecir el riesgo de abandono escolar en estudiantes de nivel medio superior en México. En él desarrolló una investigación cuantitativa, no experimental y transversal, donde la variable independiente fue el riesgo de abandono escolar, mientras que las variables predictoras fueron los hábitos de estudio, la autorregulación del aprendizaje y los estilos de aprendizaje entre otros.

Bejarano (2017) realizó un estudio de caso de deserción en la Fundación Universitaria Los Libertadores, en este artículo se hace un acercamiento a las causas de deserción en esta universidad por facultades durante los periodos 2014-1 a 2016-1, para dicha caracterización se realizó una encuesta a los estudiantes desertores durante este periodo. Entre las muchas conclusiones que allí se encuentran se puede observar que la deserción en la facultad de Ingeniería se da en mayor medida en los hombres y que el semestre donde hay mayor probabilidad de abandono es el primero, además, en dicho artículo se concluye que la deserción también se da en relación al bajo rendimiento académico, expectativas no satisfechas con la universidad o con el programa, sin embargo la variable socioeconómica se consolida como una de las variables más significativas que determina la deserción en esta facultad.

Estas investigaciones sobre deserción se realizan por lo general utilizando estadística descriptiva allí se determinan las causas de abandono de la universidad, sin embargo son pocas las investigaciones que traten de explicar la deserción en la educación superior usando modelos de supervivencia como el de Kaplan–Meier o el de regresión de Box Cox, en donde la variable determinante en el estudio es el tiempo.

Barrera (2009) realiza un estudio sobre la deserción en la Universidad de Pereira utilizando un modelo de Kaplan-Meier y un modelo de regresión de Cox, en dicho trabajo se construyen modelos por cada tipo de programa académico. Allí se establece por ejemplo que en el caso de las ingenierías Industrial y de Sistemas tienen una mayor supervivencia con respecto a las ingenierías Mecánica y Eléctrica y que en general las mujeres tienen un mayor nivel de supervivencia que los hombres; mientras que los estudiantes que proviene de colegios privados tienen una leve ventaja de supervivencia dentro de la universidad frente a los estudiantes que son egresados de colegios públicos.

Álvarez (2016) plantea un modelo de sobrevida para el análisis de la deserción en los programas de ingeniería Electromecánica y Diseño Industrial de la Universidad Pedagógica y Tecnológica de Colombia UPTC, en este se busca establecer a través de un modelo de sobrevida las variables determinantes en la disminución del riesgo de deserción y del aumento en los niveles de graduación usando una función de riesgo del modelo de sobrevida para cada uno de los estudiantes activos en el primer semestre de 2015 y calcular así la probabilidad de abandono voluntario de sus estudios en el segundo semestre de 2015.

Un estudio de supervivencia aplicado a la deserción lo hace Aroca (2020) en el cual plantea un modelo Kaplan-Meier y uno de Cox para analizar la deserción de estudiantes de una facultad de ingeniería de una universidad en Colombia, en este estudio se analiza cuales son las variables que más influyen en el estudiante al momento de tomar la decisión de desertar de la facultad. Allí se encuentra que los estudiantes de Ingeniería de Alimentos presentan un mayor riesgo de deserción frente a los estudiantes de ingeniería de Sistemas y Mecánica, también pudo observar que los estudiantes morosos presentan mayor deserción comparado con los estudiantes que no lo son, y por ultimo concluye que el sexo no es un factor determinante frente a la deserción en la universidad.

## METODOLOGÍA

Esta universidad bogotana que pertenece al sector privado no es ajena al fenómeno del ausentismo intersemestral y la deserción estudiantil, por esta razón se contemplaron las bases de datos de matrícula desde el año 2017-1 al año 2021-1 con el fin de analizar el comportamiento de la deserción estudiantil durante este periodo, para ello se contó con la ayuda de la oficina de matrículas de esta universidad.

La metodología de análisis que se va a emplear en esta investigación es de tipo descriptiva de corte interpretativo con enfoque mixto, usando un modelo estadístico de supervivencia de Kaplan-Meier y de Box Cox usando el programa estadístico R que permita determinar el periodo intersemestral en el cual un estudiante de la facultad de ingeniería es más probable que deserte y además establecer si la deserción está asociada al programa académico, a la edad o al género del estudiante. La base de datos con la cual se trabajó inicialmente estaba compuesta por 1614 estudiantes matriculados de diferentes facultades en el año 2017-1, esta base se desagregó y se construyó una nueva formada por 234 estudiantes matriculados en la facultad de ingeniería en el año 2017-1, en la cual se encuentran los programas de Ingeniería de Sistemas, Mecánica, Electrónica, Industrial y Aeronáutica; con esta nueva información se buscaron los estudiantes de la Facultad de Ingeniería matriculados en cada semestre desde el 2017-1 hasta el 2021-1.

Con esta base de datos reducida se tomaron las variables código de carrera, sexo, edad y tiempo de permanencia en la universidad y se propuso realizar un análisis de la deserción en la facultad de ingeniería de esta universidad. En la información desglosada se tienen 9 semestres de matrícula debido a que no se trabajó con la base 2021-2 porque aún no se poseía esta información; se tomó la siguientes condición, si el estudiante permanece activo durante 8 semestres se considera que no desertó y se le asigna el valor de 2 (2= Censura) y por el contrario si el estudiante dejó de matricularse en alguno de los semestres se considera como desertor y se le asigna el valor de 1 (1=Desertó) en la base de datos. Esta nueva base de datos contiene las variables que se muestran en cuadro 1.

Variable	Descripción de las variables
Código de Programa	Es el código interno de la universidad para identificar la carrera.
Código SNIES	Es el código de la universidad para identificar la carrera ante el ICFES.
Documento	Es la identificación del estudiante (CC o TI).
Nombre	Nombres y apellidos del estudiante.
Código Estudiante	Es la identificación interna del estudiante en la universidad.
Sexo	Hace referencia al género (1= Hombre, 0= Mujer)
Edad	Años cumplidos del estudiante.
Tiempo	Semestres en los cuales el estudiante permanece matriculado la universidad.
Status	Hace referencia a si el estudiante deserto o no (1= Deserto, 2= Censura)

Cuadro 1:

## Análisis de Supervivencia

El análisis de supervivencia se puede entender como un conjunto de técnicas estadísticas que permiten hacerle seguimiento y analizar una variable a través del tiempo hasta la ocurrencia de esta (Borges, 2005), 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. En el análisis de supervivencia la variable respuesta es el tiempo que transcurre entre un evento inicial (que determina la inclusión del individuo en el estudio) y un evento final (genéricamente llamado falla) que ocurre cuando el individuo presenta la característica para terminar el estudio (muerte, alta de la enfermedad, deserción, etc.). En este tipo de estudios de cohorte puede ocurrir que algún individuo lo abandone antes de que le ocurra el evento de interés, registrándose sólo información parcial (censura) sobre la variable de interés (tiempo de falla). El objetivo principal del análisis de supervivencia es incorporar esta información parcial que proporcionan los individuos censurados mediante métodos desarrollados para ese fin. En otras áreas estadísticas esta información parcial es ignorada (datos faltantes); esta manera de proceder es incorrecta, ya que se desconoce el aporte parcial de estos individuos al estudio, lo que es contrario a la filosofía estadística de incorporar toda la información disponible dentro del análisis. Los métodos estadísticos en el análisis de supervivencia son similares a los que se utilizan en otras áreas donde no se presentan datos censurados, por ejemplo: análisis descriptivo, comparación de poblaciones, modelos de regresión, entre otros.

Los tipos de censura más comunes en esta área son:

1. **Censura tipo I.** En muchos estudios el investigador debe determinar un tiempo máximo de observación para que ocurra la falla en los individuos, por lo que aquellos a los cuales aún no les ocurría la falla al concluir el periodo de observación representan una censura.
2. **Censura tipo II.** En este caso el investigador decide prolongar el periodo de observación hasta que ocurran  $k$  fallas de  $n$  posibles ( $k \leq n$ ), registrando este último valor de falla para el resto de los individuos (censuras) que no observó. Una razón común para determinar el número de fallas a observar es la potencia que se requiere para el estudio. En estos dos casos, la censura está controlada por el investigador.
3. **Censura aleatoria.** En este tipo de censura el investigador no tiene ningún control sobre la misma. Las censuras pueden ocurrir porque el individuo abandona el estudio, muere por una causa que no es de interés o permanece vivo al término del mismo.

Estos tres tipos de censura son clasificados como censura por la derecha, cuya característica es que el tiempo de censura observado es menor que el tiempo de falla, sólo que este último no es observado.

### La función de supervivencia

Esta función representa la probabilidad de que un individuo sobreviva (no experimente la falla) a un tiempo determinado  $t$ . En términos matemáticos se escribe así:

$$S(t) = P(T > t)$$

De su definición se tiene que

$$S(t) = 1 - F(t)$$

con  $F(t)$  la función de distribución acumulada. Además se tiene que:

$S(0) = 1, \quad S(t) = 0 \quad \text{si} \quad t \rightarrow \infty \quad S(t_1) \geq S(t_2) \quad \text{si} \quad t_1 < t_2$

Si  $\mathbf{T}$  es continua, con función de densidad  $\mathbf{f}(\mathbf{t})$ , entonces

$$S(t) = P(T > t) = \int_t^\infty f(u)du$$

Si es  $T$  discreta con función de masa de probabilidad

$$f(t_j) = P(T = t_j), \quad j = 1, 2, \dots \quad \text{y} \quad t_1 < t_2 < \dots$$

Entonces la función de supervivencia está dada por:

$$S(t) = P(T \geq t) = \sum_{t_j > t} f(t_j)$$

El estimador de Kaplan–Meier de la función de supervivencia es el estimador más común en los paquetes de estadística y para realizar los cálculos se utilizó el programa estadístico R, De acuerdo con Barrera (2008) el estimador de Kaplan-Meier utiliza los casos censurados y no censurados para realizar la estimación de la función de supervivencia; este estimador se obtiene de una secuencia de probabilidades condicionales de supervivencia estimada.

Cada probabilidad condicional estimada se obtiene del número de casos observados y el número de “muertes” en el instante de tiempo y se calcula como

$$\frac{n - b}{n}$$

donde  $n$  es el número de casos en riesgo y  $d$  es el número de “muertos” observados.

Para una muestra de  $n$  observaciones independientes si  $t_1 < t_2 < t_3 < \dots < t_k$  son los tiempos de vida observados y  $T_l$  es el tiempo de vida de la observación de mayor duración en la muestra. Se define entonces:

$n_i$  = Número de sujetos en riesgo en el instante  $t_i$

$d_i$  = Número de muertes en el instante  $t_i$

$c_i$  = Número de censuras en el intervalo  $[t_i, t_i + 1)$

Se puede ver que:  $n_0 = n, \quad t_0 = 0, \quad d_0 = 0, \quad c_0 = 0, \quad t_{k+1} = \infty$

$n_{i+1} = n_i - d_i - c_i \quad \text{para} \quad i = 1, 2, 3, \dots, k - 1$

El estimador de la función de supervivencia Kaplan-Meier  $\widehat{S}(t)$  se calcula de la siguiente manera:

$$\widehat{S}(t) = \prod_{t_i < t_1} \frac{n_i - b_i}{n_i} = \prod_{t_i < t_1} \left( 1 - \frac{b_i}{n_i} \right)$$

Algunas de las propiedades del estimador de Kaplan-Meier son las siguientes:

- Es insesgado para muestras grandes
- Es consistente
- Converge asintóticamente para un proceso gaussiano
- Es un estimador de máxima verosimilitud de  $\widehat{S}(t)$

## Modelo de regresión de Cox

De acuerdo con Borges (2005) el modelo de regresión de Cox es el modelo de regresión más utilizado para datos de supervivencia en el área médica. El modelo de Cox posee la ventaja de que permite modelar covariables que dependen del tiempo.

En el modelo de regresión de Cox, el riesgo para el  $i$ -ésimo individuo se define mediante la siguiente expresión:

$$\lambda(t; Z_i(t)) = \lambda_o(t)e^{\beta' Z_i(t)}$$

Donde  $Z_i(t)$  es el vector de covariables para el  $i$ -ésimo individuo en el tiempo  $t$ . El modelo de Cox establecido anteriormente se dice que es un modelo semiparamétrico debido a que incluye una parte paramétrica y otra no paramétrica.

1. La parte paramétrica es  $r_i(t) = e^{\beta' Z_i(t)}$ , llamada puntaje de riesgo (risk score), y  $\beta$  es el vector de parámetros de la regresión.
2. La parte no paramétrica es  $\lambda_o(t)$  es llamada función de riesgo base, es una función arbitraria no especificada.

## RESULTADOS

En la Tabla 1 se muestran algunos resultados, donde se observan los estudiantes matriculados en la cohorte 2017-1 por cada una de las carreras de la facultad de ingeniería por sexo y la deserción hasta el año 2021-1.

Carrera	Sexo	Matriculados 2017-1	No Eventos	No Censura	% Supervivencia
Ingeniería Electronica	Hombres	13	10	3	21,1 %
	Mujeres	2	1	1	50 %
	Subtotal	15	11	4	26,67 %
Ingeniería Sistemas	Hombres	38	34	4	10,53 %
	Mujeres	4	2	2	50 %
	Subtotal	42	36	6	14,29 %
Ingeniería Mecánica	Hombres	46	33	13	28,26 %
	Mujeres	0	0	0	NA
	Subtotal	46	33	13	28,26 %
Ingeniería Industrial	Hombres	31	26	5	16,13 %
	Mujeres	27	25	2	7,41 %
	Subtotal	58	51	7	12,07 %
Ingeniería Aeronáutica	Hombres	53	34	19	35,85 %
	Mujeres	20	12	8	40 %
	Subtotal	73	46	27	36,99 %
Total	Hombres	181	137	44	24,31 %
	Mujeres	53	40	13	24,53 %
	Total	234	177	57	24,36 %

Tabla 1



El No de eventos representa los estudiantes desertores por carrera y por sexo y No de censura representa los estudiantes que han sobrevivido, por ejemplo para la carrera de Ingeniería de Sistemas se matricularon 42 estudiantes en el periodo 2017-1, de los cuales 38 eran hombres y 4 mujeres, al final del periodo 2021-1 habían desertado 36 estudiantes de los cuales 34 fueron hombres y 2 mujeres. Los casos censurados corresponden a los estudiantes que a fecha del cohorte 2021-1 no habían desertado de la carrera, es decir no habían desertado 4 hombres y 2 mujeres; lo que quiere decir que solo el 14,29% de los estudiantes de Ingeniería de Sistemas matriculados inicialmente aún sobreviven en la universidad para este último periodo.

Para el caso de la Ingeniería Aeronáutica se matricularon 73 estudiantes en el periodo 2017-1, de los cuales 53 eran hombres y 20 mujeres, al final del periodo 2021-1 ya habían desertado 46 estudiantes de los cuales 34 fueron hombres y 12 mujeres. Los casos censurados corresponden a los estudiantes que a la fecha de la cohorte 2021-1 no han desertado de la carrera, es decir no habían desertado 19 hombres y 8 mujeres; lo que quiere decir que solo el 36,99% de los estudiantes de Ingeniería Aeronáutica aún sobreviven en la universidad para este último periodo.

En general en la cohorte 2017-1 se matricularon 234 estudiantes, 181 hombres y 53 mujeres, para la cohorte 2018-1 han sobrevivido 57 estudiantes de los cuales 44 son hombres 13 son mujeres, lo que quiere decir que han sobrevivido 24.64% de los estudiantes de la Facultad de Ingeniería y han desertado un 74.64%.

### Análisis de Supervivencia para todos los estudiantes

Para abordar el análisis de supervivencia con la función de Kaplan-Meier se va a iniciar con todos los estudiantes, en la figura 1 se puede apreciar que la probabilidad de que un estudiante sobreviva en el tiempo es decir que no deserte va disminuyendo y la probabilidad de que un estudiante no deserte después del 5 semestre es menor a 0.5. También se observa que la mediana de permanecía en la universidad es de 5 semestres.

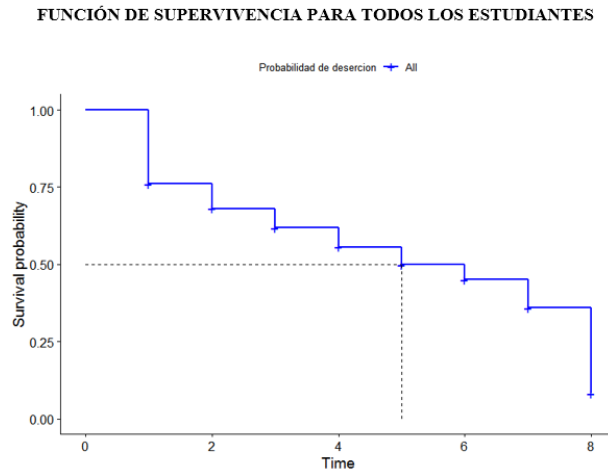


Figura 1:

n	events	median	0.95LCL	0.95UCL
234	181	5	4	7

### Análisis de Supervivencia por sexo

En la figura 2 se muestra la función de supervivencia por sexo, allí se observa que el p-valor es de 0.16, lo cual muestra que hay diferencia significativa entre los dos sexos; existe una alta probabilidad de supervivencia por parte de las mujeres con respecto a los hombres, ya que la curva de deserción de las mujeres siempre está por encima de la curva de los hombres durante todo el tiempo de observación, lo que quiere decir que la variable sexo representa un factor determinante al momento de la deserción, aquí se aprecia por ejemplo que la probabilidad de supervivencia de las mujeres en los primeros 6 semestres es superior al 0.75.

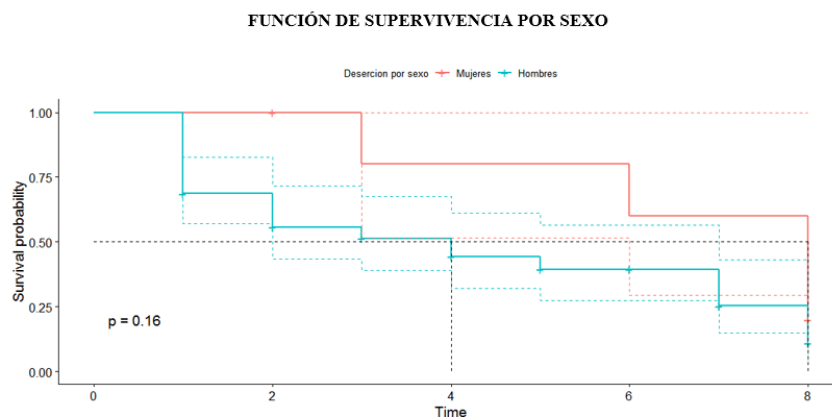


Figura 2:

Por último, el test de LogRank permite establecer si las curvas por sexo son iguales o diferentes, los resultados son los siguientes:

sexo	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
mujeres	53	0	44.4	44.4	84.2
hombres	181	181	136.6	14.4	84.2

$$\text{Chisq} = 84.2 \text{ on } 1 \text{ degrees of freedom, } p = < 2e^{-16}$$

En este caso el p-valor es menor de 0.05, es decir es muy pequeño, por lo que se puede afirmar con un nivel de significancia del 5% que las dos curvas son diferentes, es decir que el sexo es un factor que influye en la deserción estudiantil en la universidad.

### Análisis de Supervivencia por edad

En la figura 3 Histograma por Edad se observa la matrícula por edades en la cohorte 2016-1, allí se puede apreciar que la mayor cantidad de estudiantes matriculados se encuentra entre los 20 y 25 años de edad. Mientras que en las edades superiores a los 30 años son pocos, con tan solo 22 estudiantes matriculados.

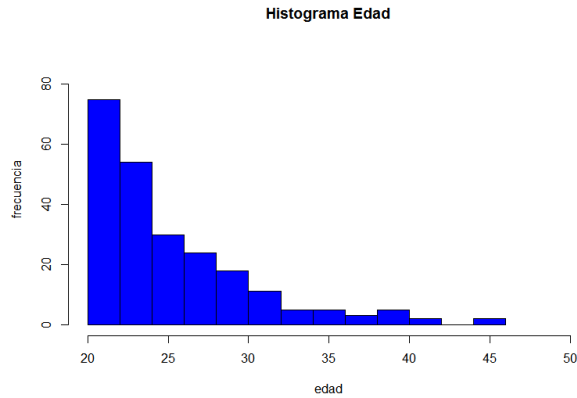


Figura 3:

La figura 4 revela que la deserción es mayor en los estudiantes cuyas edades son mayores a 30 años, mientras que los estudiantes con edades entre los 20 y 30 años su deserción es mucho menor. A continuación se muestran las deserciones por edad donde la edad mínima es 20 y la máxima es 45 años.

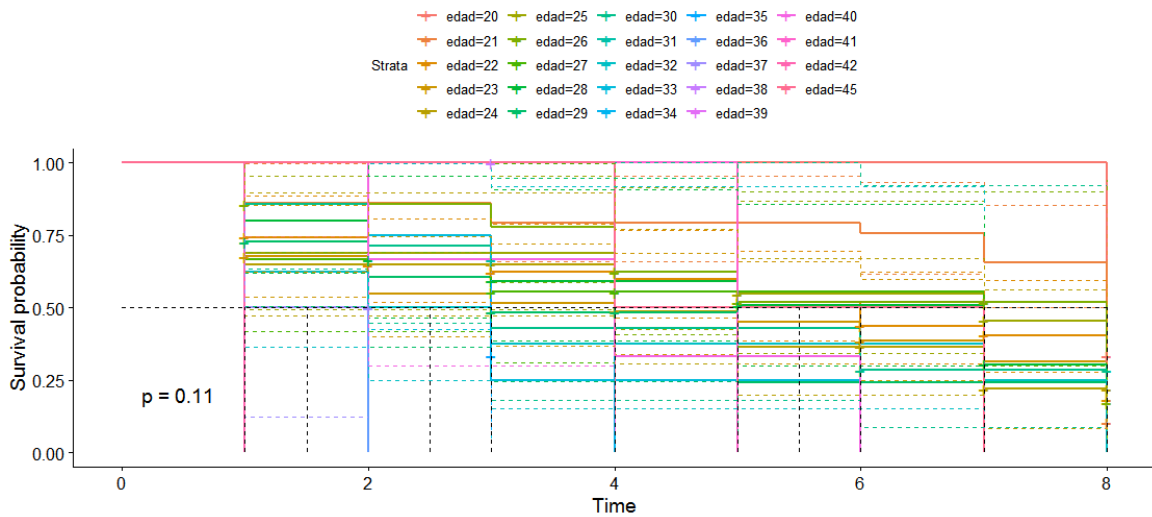


Figura 4:

Seguidamente se muestra un resumen en la figura 5 que hace el programa R de la deserción por edades donde n representa las edades de los estudiantes matriculados en 2017-1 y events representa los estudiantes que han ido desertando por edad hasta la cohorte 2021-1.

	n	events	median	0.95LCL	0.95UCL
edad=20	3	2	8.0	8	NA
edad=21	29	26	8.0	8	8
edad=22	43	30	6.0	3	8
edad=23	34	31	4.0	2	NA
edad=24	20	14	4.0	1	NA
edad=25	16	12	7.0	1	NA
edad=26	14	8	8.0	4	NA
edad=27	9	6	7.0	1	NA
edad=28	15	11	7.0	2	NA
edad=29	11	7	3.0	2	NA
edad=30	7	5	3.0	2	NA
edad=31	3	2	4.0	2	NA
edad=32	8	8	2.5	1	NA
edad=33	4	4	3.0	2	NA
edad=34	1	1	4.0	NA	NA
edad=35	3	2	3.0	2	NA
edad=36	2	2	1.5	1	NA
edad=37	2	1	1.0	1	NA
edad=38	1	0	NA	NA	NA
edad=39	2	2	5.5	4	NA
edad=40	3	3	4.0	2	NA
edad=41	1	1	5.0	NA	NA
edad=42	1	1	1.0	NA	NA
edad=45	2	2	5.5	4	NA

Figura 5:

### Estimador de Kaplan-Meier

Veamos la tabla de la figura 6 que el programa R arroja del estimador de Kaplan –Meier, donde time es el semestre, n.risk son los estudiantes que han sobrevivido, es decir no han desertado, n.events son los estudiantes que han desertado y survival es la probabilidad de supervivencia en la universidad, todo esto con margen de confianza del 95%. Aquí observamos por ejemplo que en el semestre 5 han sobrevivido 108 estudiantes y desertaron 11 estudiantes y la probabilidad de que un estudiante sobreviva es decir que no deserte después del semestre 5 es de 0.499.

Time	n. risk	n. event	survival	std. err	lower 95% CI	upper 95% CI
1	234	56	0.761	0.0279	0.7079	0.817
2	172	18	0.681	0.0306	0.6236	0.744
3	151	14	0.618	0.0321	0.5581	0.684
4	129	13	0.556	0.0332	0.4943	0.625
5	108	11	0.499	0.0339	0.4368	0.670
6	93	9	0.451	0.0342	0.3884	0.523
7	79	16	0.359	0.0341	0.2985	0.433
8	57	44	0.082	0.0214	0.0491	0.137

Figura 6:

### Análisis de Supervivencia por carrera

La grafica dela figura 8 Histograma Semestre revela que la mayor cantidad de estudiantes que desertan se presenta en el primer y en el último semestre, mientras que en el semestre cuarto y quinta esta la menor cantidad de deserción estudiantil.

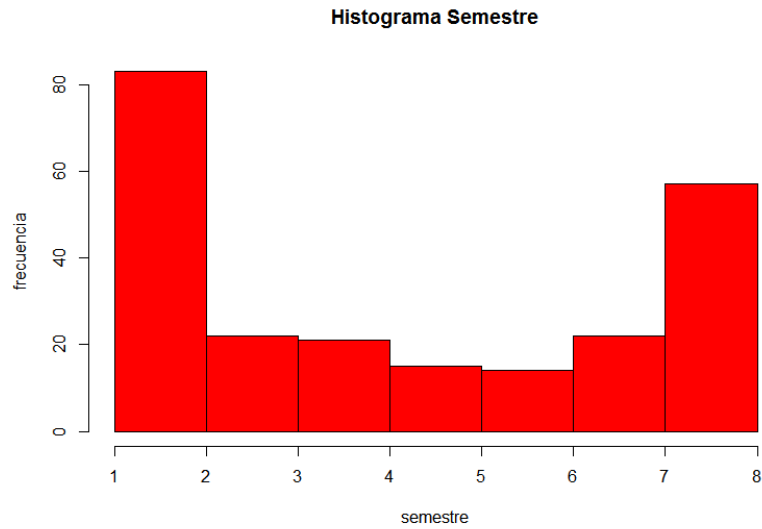


Figura 7:

En la figura 8 se observan las curvas de supervivencia por carrera, aquí se pueden apreciar que existen diferencias significativas entre los programas de Ingeniería Industrial y Aeronáutica, los cuales tienen una probabilidad mayor de deserción que los programas de Ingeniería Electrónica, Mecánica y Sistemas. En la gráfica se ve como la probabilidad de que un estudiante de Ingeniería Industrial deserte durante los primeros cuatro semestres es del 0.75, mientras que la probabilidad de que un estudiante de Electrónica deserten en los primeros cuatro semestres está por debajo del 0.50.

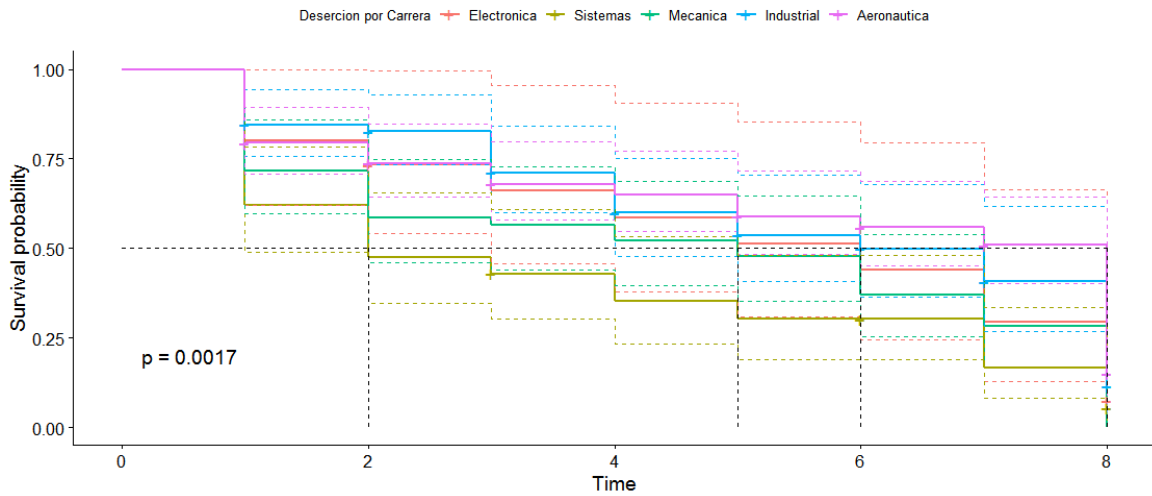


Figura 8:

Ahora al realizar el test de LogRank para determinar si las curvas por carrera son iguales o diferentes, los resultados se muestran en la figura 9.

carrera	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Electrónica	13	13	12.5	0.0221	0.0345
Sistemas	42	38	25.8	5.7850	9.4670
Mecánica	46	46	37.6	1.8582	3.4568
Industrial	31	31	36.7	0.8918	1.5554
Aeronáutica	53	53	68.4	3.4589	8.3374
Chisq= 17.3 on 4 degrees of freedom, p= 0.002					

Figura 9:

En este caso el p-valor es menor de 0.05, es decir es muy pequeño, por lo que se puede afirmar con un nivel de significancia del 5% que las curvas son diferentes, es decir que la carrera es un factor que influye en la deserción estudiantil.

### Modelo de Box Cox

Ajustamos un modelo de Box Cox la edad, el sexo y el semestre como variables. La siguiente tabla muestra los resultados para la deserción de estudiantes.

El modelo de Box Cox nos confirma que la edad es significativa al momento en el cual un estudiante decide

desertar de alguna de las carreras de la facultad de Ingeniería de la universidad. Se puede establecer que el riesgo de deserción por la variable edad es 1,129 mayor que la deserción por la variable sexo.

n = 234 , number of events=57					
	coef	exp(coef)	se(coef)	Z	Pr(> z )
edad	1.210e-01	1.129e+00	5.206e-02	2.325	0.201 *
sexo	2.307e-01	1.260e+00	3-164e-01	0.729	0.4658
semestre	1.844e+01	1.023e+08	1.362e+03	0.014	0.9892

Figura 10:

## CONCLUSIONES

Partiendo del análisis de la función de supervivencia de Kaplan-Meier y del test de LogRank y Box Cox se pueden llegar a las siguientes conclusiones:

- El sexo es considerado como un factor de riesgo para la deserción en la facultad de ingeniería de esta universidad, es decir la probabilidad de deserción es mayor en los hombres que las mujeres.
- La deserción está asociada en mayor medida a la variable edad, ya que los estudiantes mayores de 30 tienen mayor probabilidad de deserción.
- Las variables carrera y sexo no están correlacionada en la deserción de los estudiantes en la facultad de ingeniería.
- La probabilidad de deserción estudiantil en la facultad de ingeniería es menor en las carreras de Ingeniería Industrial y Aeronáutica, y donde existe una mayor probabilidad de abandonar los estudios es en la carrera de Sistemas.
- La universidad debe establecer políticas de retención de los estudiantes de más de 30 años ya que aquí se está presentando la mayor probabilidad de deserción en este grupo de estudiantes. Además debe concentrar los esfuerzos en la retención de las mujeres en la facultad de ingeniería.

## REFERENCIAS

- [1] Aroca, et al (2020) Análisis de supervivencia aplicado a la deserción en estudiantes de ingeniería. Bogotá Colombia.
- [2] Bejarano, et al. Caso de estudio: Caracterización de la deserción estudiantil en la Fundación Universitaria Los Libertadores

- [3] Barrera, M (2008) Análisis de Supervivencia aplicado al problema de deserción estudiantil en la Universidad Tecnológica de Pereira, Pereira. Colombia.
- [4] Borges, E. (2005). Análisis de supervivencia usando lenguaje R Simposio de Estadística 2005 Paipa, Boyacá, Colombia.
- [5] 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.
- [6] Hernández (2020) Modelo predictivo del riesgo de abandono escolar en educación media superior en México. México.
- [7] Matallana, et al (2020) Modelo sobrevida para la deserción estudiantil de los programas del nivel técnico en una IES de formación para el trabajo en la ciudad de Bogotá. Bogotá. Colombia.
- [8] MEN (2009) Deserción en la educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos para su prevención. Colombia.
- [9] Ocampo et al (2019) Modelo de sobrevida aplicado a deserción estudiantil de las facultades de Estadística y Diseño gráfico de la Universidad Santo Tomás. Bogotá. Colombia.
- [10] Palacios, Y (2016). Aplicación de un modelo de sobrevida para el análisis de la deserción en los programas de ingeniería electromecánica y diseño industrial de la Universidad Pedagógica y Tecnológica de Colombia. Colombia.
- [11] Zárate, R. Socha, M (2009) Estudio sobre las motivaciones de deserción estudiantil en la Universidad Industrial de Santander. Colombia.