

Perfil de alumnos con abandono escolar en una universidad privada mexicana

Profile of Dropout Students at a Private Mexican University

Gisel Hernández Chávez^a
Cuauhtémoc López Martín^b
Jesús Eladio Sánchez García^c

Recibido: 4 de diciembre de 2019
Aceptado: 21 de septiembre de 2020

Resumen: El objetivo de este trabajo fue determinar el perfil de alumnos con mayor abandono de la universidad ITESO, en programas de pregrado de nueve semestres, con el fin de mejorar la retención. Estimamos el abandono aplicando el método de Análisis de Supervivencia de Kaplan-Meier con ocho variables explicativas y un seguimiento de 12 semestres. La aplicación de pruebas de rango logarítmico a las curvas estimadas para cada categoría de las variables determinó que existen perfiles diferenciados. El perfil de mayor abandono es el de alumnos que ingresan con más de 20 años, masculinos, con promedio de preparatoria menor a 80, no procedentes de la zona metropolitana de Guadalajara, de preparatorias públicas, sin beca ni crédito, con el menor rango de puntuación en examen de ingreso y que iniciaron en programas de administración o ingenierías. Con estos hallazgos es posible detectar a los alumnos con mayor probabilidad de abandono desde su ingreso.

Palabras clave: Deserción; abandono escolar; universidades; análisis estadístico.

^a Maestra en Sistemas de Información por el Instituto Tecnológico y de Estudios Superiores de Occidente (ITESO). ✉ ghernand@iteso.mx | ORCID: <https://orcid.org/0000-0002-0904-4785>.

^b Doctor en Ciencias Computacionales por el Instituto Politécnico Nacional. Profesor-investigador en la Universidad de Guadalajara. ✉ cuauhtemoc@cucea.udg.mx | ORCID: 0000-0001-6172-9899.

^c Doctor en Ciencias Matemáticas. Investigador titular en el Instituto de Matemática y Estadística, La Habana, Cuba. ✉ jesanch64@yahoo.com | ORCID: 0000-0001-8137-2882.

Abstract: The objective of this work was to determine the student profile with the highest dropout at ITESO university in undergraduate programs of nine semesters, to improve retention. We estimated student dropout by applying the Kaplan-Meier Survival Analysis method with 8 explanatory variables and a follow-up time of 12 semesters. The application of log-rank tests to the estimated curves, for each category of variables, found that there are differentiated profiles. The profile with the highest dropouts is that of students who enter with more than 20 years, males, with a high school average lower than 80, from high schools which are not located in Guadalajara metropolitan area, from public high schools, without scholarship or credit, with the lowest scores in the entrance exam and who were admitted in administration or engineering bachelor programs. With these findings, it is possible to identify students with the highest probability of dropping out since their admission.

Key words: Dropout; student attrition; universities; statistical analysis.

Introducción

El abandono escolar universitario, entendido como el evento de interrupción de estudios sin regreso antes de un período dado¹, es un fenómeno que se presenta a nivel mundial, tanto en instituciones públicas como privadas. México no está exento de esta problemática, misma que ocasiona pérdidas económicas institucionales, familiares y personales, así como problemas psicológicos y otros impactos sociales negativos (Bound & Turner, 2011) (Sosu & Pheunpha, 2019).

Los análisis sobre abandono, o de su opuesto que es la retención, aparecen en numerosos documentos de organismos internacionales, dado el impacto que tienen a nivel global y territorial, tal como se puede observar en los trabajos que realizan los países miembros de la Organización para la Cooperación y el Desarrollo Económico (OCDE), de la cual México es miembro (OECD, 2019).

Como parte de la atención a esta problemática, el Consejo para la Acreditación de la Educación Superior (COPAES) considera estos indicadores para evaluar la calidad los programas de estudio o “carreras”, como se les llama normalmente en México (COPAES, 2016). En este contexto es que el Instituto Tecnológico y de Estudios Superiores de Occidente (ITESO), interesado en mantener sus acreditaciones y mejorar sus programas de admisión y retención, decide abrir un programa de investigación de analítica académica.

Una de las líneas del programa de investigación es el estudio del abandono universitario, tanto con fines explicativos como predictivos. Los resultados presentados en este trabajo son consecuencia de este esfuerzo.

¹ Normalmente de dos o tres semestres, en dependencia de la institución.

En el ITESO, entre 2011 y 2018, alrededor del 65% de los alumnos que matricularon su primer programa de pregrado egresaron en 12 semestres o menos². Aunque este resultado es muy superior al promedio nacional³, la aspiración es de elevar la cifra. Para ello debe disminuirse el porcentaje promedio de abandono a la universidad que es alrededor del 18%, el porcentaje promedio de cambio de carrera que es de un 6.5% y el porcentaje de alumnos activos después de 12 semestres que es de 10.5%.

Nuestro objetivo es determinar el perfil de los alumnos del ITESO con más probabilidad de abandonar. Para ello utilizamos información académica, demográfica y financiera que se obtiene durante el proceso de admisión. Esto permitirá mejorar los programas de acompañamiento, así como afinar las políticas de admisión, becas y créditos, esperando una disminución del abandono.

Para lograr el objetivo nos preguntamos:

- 1) ¿Cuáles son las variables que permiten modelar el abandono universitario en el ITESO?
- 2) ¿Pueden categorizarse las variables para determinar si existen diferencias entre ellas?
- 3) ¿Estas categorías permiten diferenciar las poblaciones y encontrar aquellos perfiles con mayor probabilidad de abandono?
- 4) ¿Cuál es el comportamiento del abandono durante el período de seguimiento, de manera que se pueda identificar el semestre más crítico para cada perfil?

Para responder a estas preguntas, estudiamos modelos teóricos y empíricos desarrollados en torno al proceso de retención y abandono universitario (Bean & Metzner, 1985; Braxton, Hirschy, & McClendon, 2004; A. F. Cabrera, Nora, & Castaneda, 1993; Tinto, 1975). La revisión de los modelos teóricos contribuyó con la identificación de las posibles variables explicativas del fenómeno y sus categorías para diferenciar los perfiles.

Los modelos empíricos por su parte, han contribuido con la confirmación y mejora de los modelos teóricos, pero hay mucho por hacer. Estos se concentran en Estados Unidos y Europa, como lo muestra la revisión sistemática de literatura llevada a cabo por van der Zanden, Denessen, Cillessen, y Meijer (2018). En México estos tipos de investigaciones son escasas (Muñoz, 2017). Tenemos entonces una

² Empleando datos de la población de ocho cohortes (2004-2011) . Por otra parte, la tasa de egreso de la universidad es superior al 80 %, pues hay alumnos que egresan en más de 12 semestres y la mayoría de los que cambian de programa dentro de la universidad logran terminar.

³ En su reporte del 2018 la OCDE publica que la eficiencia terminal de México en el 2016 fue alrededor del 28%, con base en el patrón de egreso de México (OECD, 2018).

oportunidad adicional de contribuir con nuevos datos y modelos que permitan hacer estudios comparados. La revisión de los hallazgos empíricos también contribuyó con la selección de variables y el método a aplicar en nuestro modelo. El enfoque de minería de datos en el abordaje metodológico de la solución fue también producto de esta revisión.

A continuación, en la sección “Orientación teórica del estudio”, presentamos los fundamentos conceptuales alrededor de los eventos de abandono, la descripción del tipo de investigación llevada a cabo y los principales modelos teóricos revisados. En la sección “Métodos”, explicamos los criterios de selección de las variables y del Análisis de Supervivencia (AS), mismo que fue escogido para determinar los perfiles de alumnos que abandonan. También profundizamos en el método específico de Kaplan Meier (KM), que permite el estudio univariado para la comparación de perfiles. La sección “Descripción de los datos” explica las características de la muestra, así como las acciones de selección y categorización de variables explicativas. Seguidamente, en la sección “Resultados”, exponemos el proceso de aplicación del método KM, incluyendo la validación de sus supuestos y comentamos los hallazgos asociados al perfil de los alumnos que más y menos abandonan. Finalmente, en “Discusión” hacemos conclusiones y recomendaciones asociadas a cada perfil de abandono y de carácter general. También señalamos las limitaciones del estudio y esbozamos los trabajos futuros.

1 Orientación teórica del estudio

1.1 Definiciones

El abandono de estudios o deserción estudiantil, como establecimos anteriormente, es el evento de interrupción estudios sin regreso antes de un periodo establecido (L. Cabrera, Bethencourt, Alvarez Pérez, & González Afonso, 2006). Este evento incluye al suceso de irse involuntariamente por incumplimientos administrativos o académicos, al suceso de dejar la carrera para iniciar otra en la misma institución, a la salida de la institución para continuar o no estudios en otra universidad y a la interrupción de formación por tres periodos consecutivos, tengan o no la intención de retomarla en el futuro.

El egreso, por su parte, lo definimos como el evento de culminación de todos los cursos y actividades que conforman un plan de estudios, quedando pendiente la obtención del título o grado y los trámites administrativos correspondientes. El periodo de seguimiento al alumno debe acotarse en el tiempo. Aquí lo definimos de 12 semestres, por normativa de la institución bajo estudio. Durante el mismo, un alumno puede egresar, abandonar o terminar como alumno activo.

El perfil lo hemos conceptualizado como una característica del alumno que lo hace pertenecer a una subpoblación o grupo que se diferencia de otro, por su comportamiento en cuanto al abandono durante el período de seguimiento. Se obtienen dos o más perfiles por cada variable explicativa.

1.2 Tipo de investigación

Partimos del supuesto de que, conociendo el comportamiento del abandono durante el tiempo de seguimiento y las características de los alumnos que abandonaron (variables explicativas categóricas), se pueden determinar los perfiles. Por ello, usamos bases de datos de alumnos de los que se conocía toda su trayectoria de 12 semestres, imprimiendo un enfoque no experimental y de tipo observacional cuantitativo.

No seguimos un modelo teórico de abandono-retención específico, pero sí partimos de considerar la inclusión de variables sociodemográficas, financieras y educativas (académicas). Esto nos permitió imprimir un carácter explicativo a la investigación, no sólo para verificar que los tres tipos de variable tienen una influencia en el abandono, sino también para determinar el perfil o categoría de cada variable que caracteriza a los alumnos desertores.

La fortaleza del método de Análisis de Supervivencia seleccionado (KM) permitió, como beneficio colateral, estimar la probabilidad de abandono de cada subpoblación en cada semestre. Es de subrayar que el método KM aplicado no es el mejor para obtener estimaciones de gran exactitud, pero el objetivo en nuestro estudio no fue encontrar el mejor modelo predictivo de abandono estudiantil.

1.3 Modelos teóricos revisados

Entre los modelos teóricos que revisamos destacan:

- 1) La teoría de la integración del estudiante (Tinto, 1975, 2012), que explica el proceso que motiva al abandono y lo atribuye fundamentalmente a la falta de congruencia entre los objetivos del estudiante y de la institución.
- 2) El modelo de deserción de Bean (1980), que enfatiza la importancia de la calidad de la institución (cursos, maestros), así como de los amigos, en las intenciones de quedarse o no.
- 3) El modelo de deserción del estudiante no tradicional (Bean y Metzner, 1985), que explica el proceso de abandono de estudiantes no

residenciales⁴, de tiempo parcial o de más edad que la común, en los cuales pesan más los factores fuera del campus.

- 4) El modelo de habilidad de pago (Cabrera, A., Stampen, & Hansen, 1990), que profundiza en factores de índole financiero.
- 5) El modelo de nexo entre la elección de universidad y persistencia (St. John, Paulsen, & Starkey, 1996), que argumenta que las razones por las cuales un estudiante escoge una universidad pueden ser consideradas como dimensiones de los compromisos iniciales.
- 6) La teoría del abandono del estudiante no residente (Braxton, Hirschy, & McClendon, 2004), que conceptualiza la multitud de fuerzas económicas, organizacionales, psicológicas y sociológicas que influyen a este tipo de alumno en su persistencia.

Un común denominador de estos modelos, independientemente del énfasis en una u otra causa de abandono, es la complejidad de la interacción de factores de diferente índole (académica, demográfica, económica, social, actitudinal, etc.).

2 Método

2.1 Criterios de selección de variables

Empleamos dos criterios para seleccionar las variables:

- 1) Que pertenecieran a una de las siguientes categorías identificadas en los modelos teóricos: demográficas, académicas y socio-económicas. Esto fue motivado porque sólo de ellas se registran datos en el proceso de admisión del ITESO.
- 2) Que los expertos de la Oficina de Admisión y de la Dirección de Servicios Escolares las consideraran relevantes y su información también estuviera registrada en las bases de datos del ITESO.

Consideramos como variables explicativas del abandono las presentadas en la Tabla 1. Las variables demográficas fueron: sexo, edad y localidad de la preparatoria. Las académicas fueron: promedio de preparatoria, puntuación de examen de ingreso y tipo de carrera matriculada. Las socio-económicas fueron: financiamiento (beca-crédito) y tipo de preparatoria.

Transformamos todas las variables originalmente continuas a categóricas, para poder hacer el estudio de grupos poblacionales, siguiendo las recomendaciones de Agresti (2003) para definir rangos.

⁴ No viven en el campus universitario.

Tabla 1. Variables explicativas y frecuencia de cada categoría

Nombre Variable	Descripción	Tipo de variable	Valores y significado e las categorías	Frecuencias (%)
IdSexo	sexo	Categórica nominal/demográfica	1: Masculino	4,206 (54.42 %)
			2: Femenino	3,523 (45.58 %)
CatEdad	edad	Categórica ordinal/demográfica	1: Menor a 19 años	4,387 (56.76 %)
			2: De 19 y 20 años	2,949 (38.16 %)
			3: Mayor o igual a 21 años	393 (5.08 %)
CatCarrera	tipo de carrera matriculada	Categórica nominal/académica	1: Ingenierías	2,373 (30.70 %)
			2: Administrativas	2,866 (37.08 %)
			3: Ciencias Sociales y Humanidades	2,490 (32.22 %)
CatPAA	puntuación examen de ingreso	Categórica semi- ordinal (el valor 4 rompe el orden) /académica	1: De 400 a 1092 puntos	1,629 (21.08 %)
			2: De 1093 a 1224 puntos	1,679 (21.72 %)
			3: De 1225 a 1600	2,302 (29.98 %)
			4: Pase automático; no hacen examen de ingreso	2,104 (27.22 %)
CatPromPre	promedio de Preparatoria	Categórica ordinal/académica	1: Menor a 80 puntos	2,685 (34.74 %)
			2: De 80 a 89 puntos	3,469 (44.88 %)
			3: De 90 a 100 puntos	1,575 (20.38 %)
LocalForaneo	localidad de la preparatoria	Categórica nominal/demográfica	1: Local (Guadalajara, Tlaquepaque, Tonalá, Zapopan, Tlajomulco, Zapotlanejo, Chapala)	5,235 (67.73%)
			2: Foráneo (Cualquier otro)	2,49 (32.27 %)4
CatPrepa	tipo de preparatoria	Categórica nominal/socio-económica	1: Pública	1,513 (19.58 %)
			2: Privada	6,216 (80.42 %)
CatBecaCred	financiamiento (Beca-Crédito)	Categórica nominal/socio-económica	1: ni beca ni crédito	3,471 (44.91%)
			2: beca o crédito	1,902 (24.61%)
			3: beca y crédito ⁵	2,356 (30.48%)

Fuente: Elaboración propia.

⁵ Se modificó primero con 4 valores, pero la prueba de *logrank* mostró que no había diferencias entre los grupos 2 y 3:

1: No tiene ni beca ni crédito; 2: Solo beca; 3: Solo crédito; 4: Beca y crédito.

Contamos con información sobre el bienestar financiero de los estudiantes a través de los montos de beca y crédito que tenían a su ingreso, misma que fue recogida en la variable con mnemónico CatBecaCred. Dicha variable tiene al menos dos connotaciones: financiera y de compromiso compartido con la universidad. Muchos de los becarios tienen esta prestación por ser hijos de empleados, por tener hermanos que estudian o han estudiado en la universidad o por excelencia académica. Por otra parte, el crédito es otorgado directamente por la casa de estudios y no por instituciones financieras externas, con los compromisos de corto y mediano plazo que implican.

Consideramos la inclusión del tipo de preparatoria (reflejada en la variable CatPrepa), por estar asociado con el nivel socioeconómico, registrando dos categorías: pública y privada. Incluimos el origen geográfico de la preparatoria (reflejado en la variable LocalForaneo), porque influye en la calidad de vida de un estudiante en la universidad. Los foráneos (proviene de preparatorias fuera de la Zona Metropolitana de Guadalajara) tienen más gastos de vivienda y alimentos, así como menos acompañamiento familiar.

También escogimos la variable tipo de carrera (CatCarrera) debido a que se ha probado en estudios con poblaciones de semejante composición y tamaño, que el abandono está influenciado por las características de los programas hacia los cuales los estudiantes son inicialmente atraídos (Imbrie, Lin & Reid, 2010; Murtaugh, Burns & Schuster, 1999). Como se observa en la Tabla 1, esta variable se consideró con tres categorías: Ingenierías, Administración y Ciencias Sociales y Humanidades (CSH), mismas que usa la universidad bajo estudio como agrupación en sus programas de admisión.

2.2 Selección del método

Existe una amplia posibilidad de aplicación de métodos estadísticos y de aprendizaje de máquina que revelan diferentes aspectos del abandono universitario (Ameri, Fard, Chinnam & Reddy, 2016). Es el caso del Análisis de Supervivencia (AS) que, a diferencia de métodos como la regresión logística, permite modelar cuando la variable dependiente es el tiempo hasta la ocurrencia del evento de interés (Kleinbaum & Klein, 2012). Es por ello que se adecua a nuestras preguntas de investigación.

La revisión de la literatura corroboró la pertinencia de aplicar AS, tanto con fines explicativos como predictivos, como se constata en estudios empíricos latinoamericanos y de otras regiones (Murtaugh et al., 1999; Bahi, Higgins & Staley, 2015; Ameri et al., 2016; Vallejos & Steel, 2017; Lassibille & Gómez, 2008).

Para seleccionar un método de AS específico consideramos la ventaja de KM de ser no paramétrico, por lo que no es necesario comprobar que los datos sigan una distribución específica. Por otra parte, es adecuado y suficiente para el análisis univariado. El último criterio fue el haber encontrado en la literatura evidencia de la calidad de sus hallazgos, lo que constatamos en estudios como el de Murtaugh y colegas (Murtaugh et al., 1999) y el de Muñoz y colegas (Muñoz, 2017).

2.3 Característica y ventajas del AS

El AS es el análisis de los datos del tiempo transcurrido hasta que ocurre un evento de interés. Dichos datos describen el período desde un origen temporal hasta un punto final que llamamos tiempo de observación o de seguimiento.

El AS incluye métodos estadísticos paramétricos, semi paramétricos y no paramétricos, así como adaptaciones a técnicas de aprendizaje de máquina (Wang, Li & Reddy, 2019). Puede emplearse con fines explicativos, como es el caso de nuestro estudio, o con fines predictivos. Todo esto le da una amplia gama de usos.

Otra ventaja del AS es que permite la correcta manipulación de los individuos estudiados en los que no es posible saber si ocurre el evento de interés, pues “salen” del estudio y no pueden ser considerados en riesgo. A esos eventos de salida del estudio se les llama censura y a los individuos se les clasifica como censurados.

2.4 Análisis de Supervivencia con Kaplan-Meier (KM)

KM es un método de estimación no paramétrico (Kaplan & Meier, 1958) utilizado para estimar la función de supervivencia a partir de datos de “vida”.

KM permite realizar análisis de una variable (univariado) y es útil para la comprensión preliminar del fenómeno, independientemente de si después se desarrolla un modelo predictivo con otras técnicas que permitan análisis multivariados y predicciones más precisas, como es el caso del modelo de riesgos proporcionales de Cox (1972).

La formulación general de KM es:

$$\text{Función de supervivencia: } S(t) = P(t < T) \tag{1}$$

Para $t = 0$, $S(t) = 1$. Esto quiere decir que al inicio del seguimiento la probabilidad de “sobrevivir” (no abandonar la universidad) es 1.

$$\text{Función de riesgos: } h(t) = \lim_{\Delta t \rightarrow 0} \left(\frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \right) \quad (2)$$

En nuestro caso

T : variable aleatoria que denota el tiempo de supervivencia del alumno en semestres desde que matricula hasta que abandona la carrera (se va de la universidad o se cambia de carrera) o se termina el tiempo de observación.

t : cualquier valor específico de interés para la variable T

Δt : intervalo de tiempo entre eventos (un semestre)

$\delta = (0,1)$: variable aleatoria que indica si el alumno abandonó o no la carrera

$$\delta = \begin{cases} 1 & \text{estudiante abandona} \\ 0 & \text{es censurado} \end{cases}$$

$S(t)$: función de supervivencia. Probabilidad de que un alumno permanezca (no abandone) más allá del tiempo t . O sea, la variable T excede a t .

$h(t)$: función de riesgos. Es el potencial instantáneo por unidad de tiempo para que el abandono ocurra si el alumno ha permanecido hasta el semestre t . O sea, la variable $T \leq t$.

Dado que en el numerador de la ecuación (2) tenemos la probabilidad condicional $P(t \leq T < t + \Delta t \mid T \geq t)$, en ocasiones se le llama tasa de falla condicional. Resaltamos que es una tasa y no una probabilidad, por lo que su valor puede ser mayor que 1, pero siempre mayor o igual a cero.

En este estudio estamos interesados en $S(t)$, pero es importante considerar que existe una estrecha relación entre ambas funciones y si se conoce una, se puede derivar la otra. A mayor tasa de riesgo promedio, menor probabilidad de supervivencia.

Para estimar la función de supervivencia empleamos la ecuación:

$$\hat{S}(t) = \prod_{i: t_i < t} \left(1 - \frac{d_i}{n_i} \right) \quad (3)$$

d_i : cantidad de eventos de abandono en el tiempo t

n_i : cantidad de sujetos en riesgo de abandonar en el tiempo anterior a t

El modelo de KM permite comparar las curvas estimadas de supervivencia para cada grupo o categoría de variable explicativa. Una forma de hacerlo es graficar las curvas de supervivencia segmentadas por valores de las variables. Si tienen la misma forma y difieren sólo por un factor constante, no se desecha la asunción de que son grupos diferentes.

Un segundo método es graficar las curvas log-log⁶ o curvas de logaritmos. Si las mismas son paralelas y no se cruzan, entonces muy probablemente la variable satisface la asunción. Si las curvas se cruzan, no se sostiene la asunción.

Un tercer método más robusto es aplicar una prueba de *logrank* (rango logarítmico), misma que es una prueba de hipótesis para comparar las distribuciones de supervivencia de dos grupos. El estadístico de *logrank* tiene una distribución aproximada de chi-cuadrada con un grado de libertad, y el *p*-valor se calcula utilizando la distribución de chi-cuadrada. La hipótesis nula H_0 es que los grupos tienen la misma supervivencia y normalmente con $p > 0.05$ se rechaza. Tiene en cuenta todo el período de seguimiento y la ventaja de ser no paramétrica, por lo que no requiere que se sepa algo sobre la distribución de los tiempos de supervivencia. El número esperado de sujetos que sobreviven en cada punto del tiempo se ajusta según el número de sujetos en riesgo en cada punto del tiempo en cada evento. La prueba determina si el número observado de eventos en cada grupo es significativamente diferente del número esperado. Cuando el estadístico de *logrank* es grande hay evidencia de una diferencia en los tiempos de supervivencia entre los grupos.

En la sección 4 detallamos la aplicación del primer método con fines visuales y del tercer método como recurso decisional definitivo, para determinar si las categorías de cada variable explicativa representan poblaciones diferentes en cuanto a abandono. También mostramos la función de supervivencia estimada considerando todas las variables.

2.5 Herramientas de software empleadas

Desarrollamos una aplicación que usa una biblioteca de funciones en lenguaje Python especializada en AS llamada *lifelines* (Davidson-Pilon et al., 2019) y otra biblioteca de funciones especializada en análisis de datos llamada *pandas* (*pandas*)

⁶ Los gráficos log-log muestran datos en dos dimensiones donde ambos ejes usan escalas logarítmicas. Cuando una variable cambia como una potencia constante de otra, un gráfico log-log muestra la relación como una línea recta. El término de potencia corresponde a la pendiente y la constante corresponde a la intersección de la línea.

development team, 2020). Esta última permitió realizar los trabajos de preparación y analítica preliminar.

3 Descripción de los datos

La información anónima de los estudiantes de la universidad proviene de bases de datos de la Dirección de Servicios Escolares, la Oficina de Admisión y la Oficina de Finanzas del ITESO. También empleamos catálogos de escuelas preparatorias que proporciona la Secretaría de Educación Pública. De los 9,850 inscritos por primera vez en la universidad en los semestres de otoño de 2004 a 2011 en 32 programas de estudios, eliminamos un programa con una estructura curricular diferente y dos programas con duración de 10 semestres, quedando 7,729 alumnos de 29 programas.

El evento bajo estudio, como ya hemos dicho, es el acto de abandonar definitivamente el primer programa de pregrado matriculado, ya sea por cambio de programa o por abandono de la universidad, sin hacer distinción entre estas dos causas. En el caso del cambio de programa existe un proceso formal que se registra en la base de datos. En el caso del abandono, el alumno puede dejar de asistir sin avisar y no se registra hasta que ocurran tres semestres consecutivos sin efectuar matrícula⁷. La razón es que los alumnos, de manera común, toman uno o dos semestres para atender asuntos financieros o desarrollar otras experiencias de vida.

Nuestra variable dependiente, medida en semestres, es el tiempo transcurrido hasta que ocurre el evento de abandono. El tiempo final de observación, como ya se ha señalado, se fijó en 12 semestres, por ser el tiempo máximo de permanencia que marca el reglamento de la universidad bajo estudio. Normalmente los programas se actualizan cada 5 años y los alumnos que no egresan en menos de 12 semestres deben hacer cambio al programa actualizado.

Consideramos que no abandona quien egresa por concluir la totalidad de los créditos escolares, aunque no se haya titulado, o quien llega como alumno activo al semestre 12. La cantidad de alumnos retenidos es igual a la cantidad de egresados más los alumnos activos.

La Tabla 2 proporciona información sobre el estado de supervivencia de los estudiantes bajo estudio. El 24.96% de ellos abandonó la carrera inicialmente matriculada, ya fuera por cambio de carrera o abandono de la universidad, durante

⁷ Esto quiere decir que como máximo se admiten dos semestres consecutivos sin matricular. Esta es una política interna del ITESO basada en estudios estadísticos internos y coincidente con estudios internacionales reportados desde hace muchos años. Tómese como ejemplo el trabajo de Shuman (Shuman, 1956) de mediados del siglo pasado.

los primeros 12 semestres. El mayor porcentaje se dio después del primer semestre y alcanzó el 8%. Es por ello que nuestro estudio desarrolla el modelo con los datos de admisión.

Tabla 2. Tabla de vida de los 7,729 alumnos estudiados

Semestre	Entra- da	En riesgo al inicio del semestre	Observa- do	Censur a-dos ⁸	Elimina- do	% abandon o por semestre ⁹	% abandono acumulado	% egreso por semestre ¹⁰	% de egreso acumu- lado
0	7729	7729	0	0	0	0.00	0.00		
1	0	7729	618	0	618	8.00	8.00	0.00	0.00
2	0	7111	301	0	301	3.89	11.89	0.00	0.00
3	0	6810	236	0	236	3.05	14.94	0.00	0.00
4	0	6574	191	0	191	2.47	17.41	0.00	0.00
5	0	6383	140	1	141	1.81	19.23	0.01	0.01
6	0	6242	72	2	74	0.93	20.16	0.03	0.04
7	0	6168	77	29	106	1.00	21.15	0.38	0.41
8	0	6062	69	110	179	0.89	22.05	1.42	1.84
9	0	5883	58	2262	2320	0.75	22.80	29.27	31.10
10	0	3563	68	1864	1932	0.88	23.68	24.12	55.22
11	0	1631	38	596	634	0.49	24.17	7.71	62.93
12	0	997	61	936	997	0.79	24.96	9.52	72.45

Fuente: Elaboración propia.

No hubo datos faltantes para los 7,729 registros de la población estudiada.

3.1 Tratamiento de la censura

En la Tabla 2 se observa que hasta el séptimo semestre sólo hay 32 censurados (por egresar y ya no estar en riesgo de abandono). Al final del estudio hubo 936 censurados, lo que representa el 12.11 % de los 7,729 bajo observación. De ellos, 291 fueron censurados por egreso y 584 (7.56%) quedaron activos.

⁸ En los semestres del 1 al 11 la cifra corresponde con los egresados. En el semestre 12 corresponde a la suma de 736 egresados más otros 200 que estaban activos hasta ese momento.

⁹ En relación a los 7,729 matriculados.

¹⁰ En relación a los 7,729 matriculados.

A pesar de las críticas a la consideración del egreso como censura (Vallejos y Steel, 2017), consideramos la misma pertinente en este caso, debido a que podemos centrar el análisis en los primeros semestres donde se da la cantidad de abandono más alta y no existe censura.

Uno de supuestos de aplicación de KM es que los que se censuran tienen una probabilidad de sobrevivir igual a los que siguen en riesgo. Su incumplimiento es especialmente peligroso si los que se censuran tienen una probabilidad menor de sobrevivir. Este no es el caso, pues los egresados tienen 100% de probabilidad de no abandonar, por lo que nuestros estimados no fueron optimistas.

4 Resultados

4.1 Cumplimiento de supuestos

El primer paso fue analizar el cumplimiento de supuestos, comenzando por el patrón de censura. En este caso hay censura por egreso de la carrera o por fin de la observación pasados 12 semestres desde el ingreso. El egreso de un alumno no aumenta la probabilidad de egreso de otros, lo que conduciría a una falta de independencia entre la censura y los sujetos. Sin embargo, como se fundamentó en el apartado anterior, se decidió obviar la condición de que los que quedan en riesgo deben tener la misma probabilidad de sobrevivir que los que se censuran.

Seguidamente, verificamos que las probabilidades de supervivencia fueran las mismas para los sujetos independientemente del momento en el que entren al estudio. En todos los casos los alumnos entran en $t=0$ (inicio del primer semestre), independientemente de la fecha real de matrícula (cualquier otoño entre 2004 y 2011), considerando que esto no tiene efecto en la respuesta.

4.2 Estimación de la supervivencia general

La función de supervivencia estimada con KM, sin diferenciar subpoblaciones, se muestra en la Figura 1. Los valores de probabilidad de supervivencia estimada para los 12 semestres aparecen a la derecha de la gráfica.

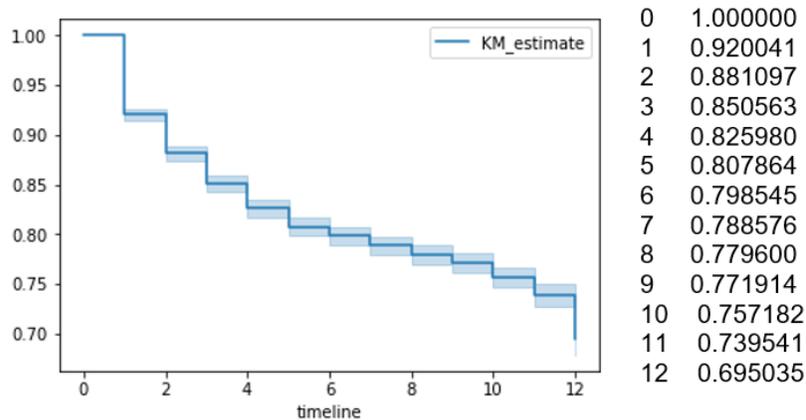


Figura 1. Función de supervivencia estimada con KM
Fuente: Elaboración propia

4.3 Análisis univariado para determinar perfiles de abandono

Para evaluar si dos o más grupos difieren en cuanto a abandono, estimamos la función de supervivencia para cada categoría de variable explicativa, visualizamos las curvas y aplicamos la prueba de *logrank*.

Las hipótesis para la prueba de *logrank*, en cualquier t y con p -valor de 0.05 son:

H_0 : Los grupos no difieren (las curvas son casi idénticas);

H_1 : Los grupos difieren (las curvas no son idénticas)

Un p -valor de 0.05 significa que aproximadamente una prueba de cada veinte rechazará falsamente la hipótesis nula. Si es menor que 0.05 se rechaza la hipótesis nula; o sea, que los grupos son diferentes.

4.3.1 Resultados con la primera versión de categorías

Las variables CatPAA (categorías de examen de ingreso), CatCarrera y CatBecaCred resultaron con algunas poblaciones casi idénticas en cuanto a abandono utilizando la primera versión de categorías mostrada en la Tabla 1. En el caso de CatPAA no hubo diferencias significativas entre los alumnos que obtuvieron puntuación entre 400 y 1092 y los que obtuvieron entre 1093 y 1224. En el caso de CatCarrera sólo fue significativa la diferencia entre carreras administrativas y las de Ciencia Sociales y Humanidades. En el caso de CatBecaCred no hubo diferencias entre los que tenían sólo un apoyo o los que tenían los dos. Con esos resultados, se agruparon las categorías que no se diferenciaban.

4.3.2 Resultados con la segunda versión de categorías

Después de recategorizar las variables, repetimos la prueba de *logrank* y los resultados se muestran en la Tabla 3. En todos los casos p-valor < 0.05, indicando que las tasas de riesgo de los grupos no son idénticas para todos los valores de tiempo y rechazándose la hipótesis nula H_0 . Los grupos en esta tabla corresponden a poblaciones diferentes en cuanto a abandono.

Tabla 3. Pruebas de *logrank* para categorías de variables explicativas

Variable	Grupos contrastados	p-valor
IdSexo	Masculino con Femenino	<0.005
LocalForaneo	Local con Foráneo	<0.005
CatPrepa	Pública con Privada	< 0.005
CatEdad	Grupo edad <19 con grupo entre [19 y 21)	< 0.000
	Grupo edad <19 con grupo edad>21	< 0.000
	Grupo entre [19 y 21 con grupo >21	< 0.000
CatPAA	Grupo [400 ,1224] con grupo [1225,1600]	< 0.000
	Grupo [400 ,1224] con grupo pase directo	< 0.000
	Grupo [1225,1600] con grupo pase directo	< 0.000
CatCarrera	Grupo Ingenierías + Administrativas con grupo Ciencias Sociales y Humanidades	0.0079
CatPromPre	Grupo prom<80 con grupo de 80<=prom< 89	< 0.000
	Grupo prom<80 con grupo prom>89	< 0.000
	Grupo de 80 <=prom < a 89 con grupo prom>89	< 0.000
CatBecaCred	Grupo sin apoyo con grupo con apoyo	< 0.000

Fuente: Elaboración propia.

La Figura 2 muestra curvas separadas y con paralelismo de todos los grupos recategorizados de la Tabla 3.

Tanto por inspección visual como por resultados estadísticos observamos que la estimación de abandono es menor en: mujeres; alumnos locales; alumnos de preparatorias privadas; con edad menor a 19 años; con pase automático (sin examen de ingreso); que matriculan carreras de Ciencias Sociales y Humanidades; con promedio de preparatoria mayor o igual a 89 y con apoyo financiero de beca o crédito de la universidad.

La variable en la que encontramos diferencias más marcadas entre perfiles es la de financiamiento y en la que menos diferencia encontramos es en la del tipo de carrera.

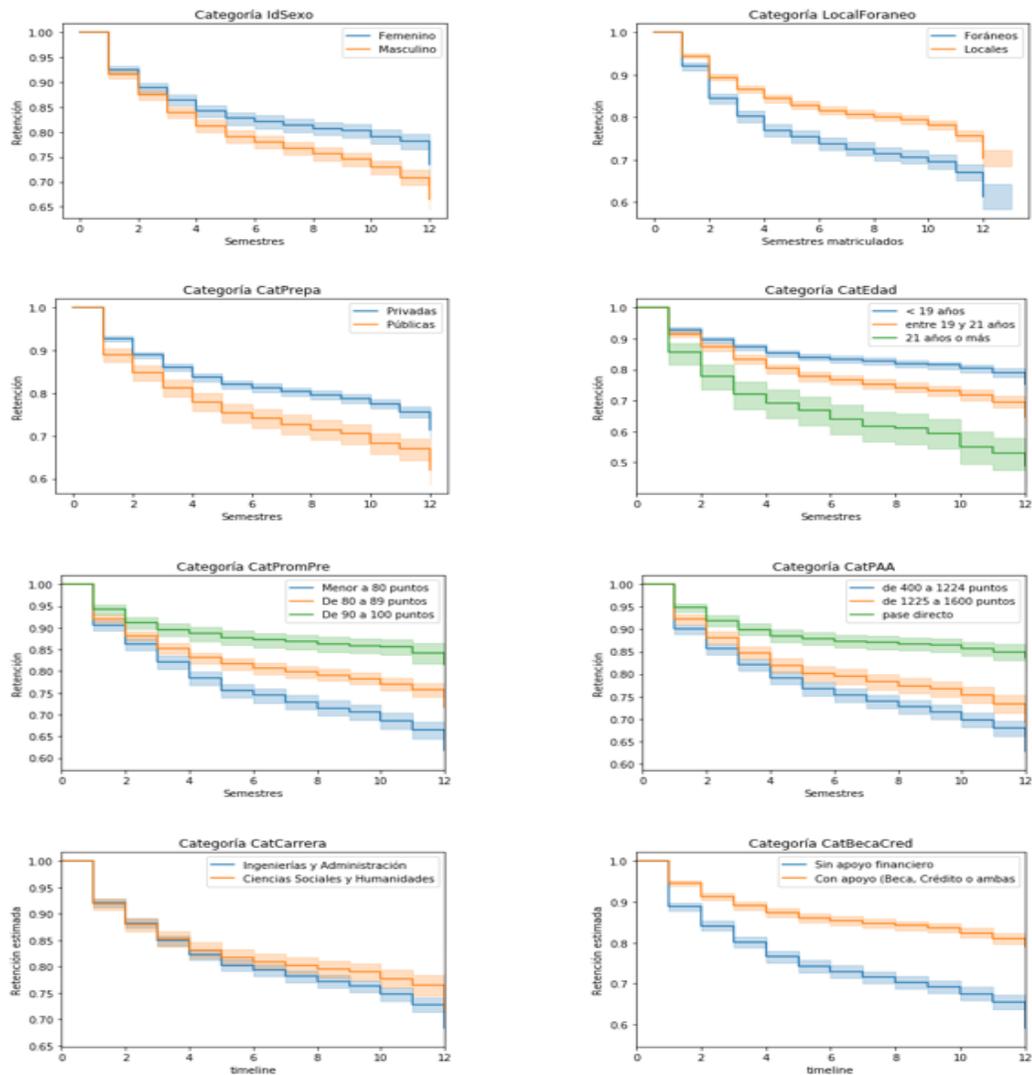


Figura 2. Curvas de supervivencia para variables explicativas y sus categorías
Fuente: Elaboración propia.

5 Discusión

Es importante señalar que la institución bajo estudio no busca excluir a los alumnos con más riesgo de abandonar, sino mejorar el acompañamiento, los apoyos financieros, la retroalimentación a las preparatorias y demás acciones encaminadas a lograr que disminuya el abandono. Con ello en mente, presentamos a continuación conclusiones y recomendaciones por cada una de las variables explicativas e incluimos comparaciones con estudios previos y modelos teóricos.

Cerramos este apartado con conclusiones y recomendaciones generales, limitaciones y trabajos futuros.

5.1 Perfil según el género (variable IdSexo)

Concluimos que el perfil de género es influyente. Las diferencias entre las poblaciones de mujeres y hombres en cuanto a abandono se acentúan cada semestre. Después de un seguimiento de 12 semestres, se estima una diferencia de un 7.3% mayor en los varones. Recomendamos analizar si existe un comportamiento similar en todas las carreras en cuanto al género e indagar de manera más detallada en el origen de las diferencias.

5.2 Perfil según la procedencia geográfica (variable LocalForaneo)

Se estima que los locales, al igual que los foráneos, tienen el mayor abandono en el primer semestre, siendo el de los locales de 7.0% y el de los foráneos de 10.2%.

Los foráneos tienen más riesgo de no permanecer. Estos alumnos viven normalmente rentando casas particulares y no en una residencia universitaria. Según la Teoría del abandono del alumno no residente (Braxton et al., 2004) y los estudios de Bean (1983, 2012), en los alumnos que no viven en una residencia universitaria inciden múltiples factores entre los que destacan problemas económicos y psicológicos. También estudios empíricos demuestran que los alumnos de universidades residenciales tienen una retención mayor que la de las universidades no-residenciales (Pascarella y Terenzini, 1980; Braxton et al., 2004; Pascarella y Chapman, 1983; Torres, 2006).

Recomendamos estudiar este fenómeno más a fondo para determinar las causas y desarrollar programas de atención especializados para foráneos. También planteamos la pregunta a las autoridades universitarias en relación a la conveniencia de la creación de una residencia.

Concluimos que el perfil de localidad de la preparatoria es influyente y prioritario de consideración porque más del 32% de los alumnos son foráneos y la diferencia estimada entre ambos perfiles después de 12 semestres es de más del 10%.

5.3 Perfil según la categoría de la preparatoria (variable CatPrepa)

Se estima que los alumnos de preparatorias públicas, al igual que los de las privadas, tienen el mayor abandono en el primer semestre, con diferencia aproximada de un 5%.

Recomendamos un análisis detallado separando a los que provienen del sistema abierto y del extranjero, incluidos aquí en el grupo de las públicas. También vemos conveniente analizar con profundidad si factores como el insuficiente

financiamiento o la falta de orientación profesional pudieran estar asociadas a este perfil de alumno.

Concluimos que esta es una variable de influencia importante, pues después de un seguimiento de 12 semestres, la probabilidad estimada de abandono de las privadas es alrededor de 10% superior que la de las públicas.

5.4 Perfil según la edad de ingreso (variable CatEdad)

El abandono estimado del primer semestre para los alumnos menores de 19 años es del 7.1%, para los que están entre 19 y 21 se estima de 8.49% y para los de 21 o más se estima de 14.5%. Es en este semestre donde se dan las mayores caídas.

Confirmamos los planteamientos del modelo de deserción del estudiante no tradicional de Bean y Metzner (1985), pues la población de alumnos de más edad, que no entra terminando la preparatoria, tiene más riesgo de abandono. Esto también se observa en los estudios de Murtaugh y colegas (1999).

Recomendamos una revisión individual de casos para los alumnos con 21 o más años al ingreso. Al ser pocos (en 8 años fue de sólo 49 alumnos como promedio al año), y ser la población con mayor probabilidad de abandono después de 12 semestres (48.8%), amerita el tratamiento caso por caso.

Concluimos que esta es una variable de influencia importante. Después de un seguimiento de 12 semestres, estimamos que la diferencia entre el grupo de menos de 19 y el siguiente es inferior 10.9%, y entre el primero y los de 21 o más años es de 26.6%.

5.5 Perfil según la puntuación en examen de ingreso (variable CatPAA)

En todos los casos el semestre de mayor abandono fue el primero. Estimamos que, en dicho semestre, los de pase directo abandonan en 5.2%, los que obtienen entre 1,225 y 1,600 puntos en 7.7 % y los que obtienen entre 400 y 1,224 puntos abandonan en un 10.0 %.

Hasta donde revisamos en la literatura, no encontramos estudios con un diseño de variable exactamente igual. Normalmente se categorizan sólo los rangos de puntuaciones del examen de ingreso, pues todos los alumnos lo realizan, mientras que en el ITESO sólo lo hacen aquellos que provienen de prepas no reconocidas por la universidad como de alta calidad, y todos los que tienen promedio menor a 80. Los resultados asociados a esta variable confirman lo acertado de la política de convenio con preparatorias de alta calidad. También es acertada porque permite

establecer nexos previos (pláticas vocacionales, talleres sobre la universidad, etc.) que facilitan la elección de la universidad y la carrera. También promueve que grupos de amigos continúen juntos, elementos todos ellos recomendados en los modelos de Tinto, Bean y St. John mencionados anteriormente en este trabajo.

El hecho de que el mayor abandono se presenta en los que obtienen menor puntuación en el examen de ingreso también confirma lo adecuado de este filtro.

Concluimos que esta es una variable de influencia importante. Después de un seguimiento de 12 semestres, estimamos que la diferencia en la probabilidad de abandono entre el grupo de pase directo y el de mayor puntaje en examen de ingreso es inferior en un 14.1%. La diferencia de este último perfil con el de menor puntaje en examen de ingreso es inferior en un 6.7%.

5.6 Perfil según la categoría de programa de estudio (variable CatCarrera)

Esta es la variable en la que las diferencias son menos significativas. Las ingenierías y las carreras administrativas tienen probabilidades de abandono similares, siendo estas últimas ligeramente menores. Al no ser grupos significativamente diferentes, las categorizamos como una sola población y resultó que la diferencia con las Ciencias Sociales y Humanidades es significativa, aunque pequeña. Recomendamos estudiar por separado las carreras, pues en estudios como los de Muñoz y Vallejo (Muñoz, 2017; Vallejos & Steel, 2017) se encontraron diferencias entre cada una.

5.7 Perfil según el promedio de preparatoria (variable CatPromPre)

La variable CatPromPre muestra curvas diferenciadas. A menor promedio de preparatoria observamos menor supervivencia. Nótese que el límite de rango de la población con menor retención la establecimos por debajo de 80 puntos, pues no encontramos diferencias significativas en poblaciones de alumnos con promedios entre 60 y 80.

En todos los casos el semestre de mayor abandono fue el primero. Estimamos que en dicho semestre, en los alumnos con menos de 80 puntos, el abandono es de un 9.5%; en los que obtienen entre 80 y 89 puntos es de un 7.9 % y en los que obtienen 90 o más es de un 5.7 %.

Esto confirma que esta variable académica es una buena diferenciadora y que el ITESO debe mantener la obligatoriedad del examen de ingreso para todos aquellos con menos de 80 de promedio.

Concluimos que esta es una variable de influencia importante. Después de un seguimiento de 12 semestres, estimamos que la diferencia de probabilidad de abandono entre el grupo de menos de 80 de promedio y el que tiene entre 80 y 89 de promedio es inferior en 9.79%. La diferencia entre este último y el grupo de 90 o más de promedio es de 9.9%.

5.8 Perfil según el tipo de financiamiento (variable CatBecaCred)

Estimamos que los que tienen apoyo financiero, al igual que los que no lo tienen, presentan el mayor abandono en el primer semestre, siendo el de los primeros de 5.43% y el de los segundos de 11.16 %. Como se observa en la Figura 2, esta es una de las variables más importante en la discriminación de los perfiles de abandono. Estas diferencias se acentúan en el tiempo, pues después de 12 semestres es alrededor del 20%.

Concluimos que continuar con la mejora de las políticas de becas y créditos para los alumnos con perfil de más riesgo de abandono es prioritario, aunque los apoyos del ITESO benefician ya a más del 40% de la población estudiantil.

5.9 Conclusiones generales

Nuestro estudio tiene implicaciones importantes en los esfuerzos del ITESO por mejorar la retención. Para todas las categorías de variables explicativas, el semestre de mayor abandono fue el primero, demostrando la pertinencia de usar un modelo con sólo variables de admisión. Esto permite una temprana detección de los que tienen mayor riesgo de abandonar.

Todas las variables permitieron diferenciar las poblaciones de alumnos, encontrando siempre un grupo o perfil con diferencias significativas. Corroboramos que, tanto los factores académicos previos a la universidad, como los demográficos y los relativos al soporte financiero, deben ser parte de nuestra aproximación teórica y empírica para el abordaje del problema del abandono universitario.

5.10 Limitaciones y recomendaciones relacionadas

El estudio no diferenció los tipos de abandono. Recomendamos continuar con la aplicación de modelos de riesgos en competencia, diferenciando al abandono por cambio de carrera de aquel que lleva a la salida de la universidad, e incluyendo al egreso como evento y no como censura. Con ello tendríamos un modelo integral de historia de vida.

Los resultados presentados son producto de un análisis univariado. Es necesario realizar un análisis multivariado, pues en el mismo pueden desaparecer diferencias entre perfiles. Por ejemplo, podría no haber diferencias entre hombres y mujeres en la población de menos de 19 años.

5.11 Trabajo futuro

Si bien concluimos que en el primer semestre es donde se da la mayor cantidad de abandonos, el fenómeno se sigue manifestando a lo largo de la trayectoria estudiantil. Esto nos lleva a continuar la investigación desarrollando modelos que incluyan variables medidas durante el curso de los estudios.

Es conveniente hacer el modelo para las carreras de Diseño y Arquitectura, que son de 10 semestres y que representan casi el 25% de la matrícula total de la universidad.

Si bien en el ITESO las cohortes de primavera son pequeñas comparadas con las de otoño (la relación es de 1 a 3, aproximadamente), debemos estudiarlas y compararlas, para así obtener un panorama más integral de la universidad.

También trabajaremos en el desarrollo de modelos predictivos multivariados, buscando aquel que permita determinar lo más exactamente posible a los alumnos en riesgo de abandonar.

Finalmente creemos que la línea de trabajo más importante es la incorporación de estos modelos a los procesos cotidianos de la universidad. Son fundamentalmente los actores de la Oficina de Admisión, los coordinadores de las carreras y los compañeros de la Coordinación de Acompañamiento para la Excelencia Académicas, los que deben tener esta información para su trabajo del día a día. 

Referencias

- Agresti, A. (2003). *Categorical Data Analysis* (3rd ed.). Hoboken: John Wiley & Sons Inc.
- Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016). Survival Analysis based Framework for Early Prediction of Student Dropouts. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management - CIKM '16*, 903–912. doi: [10.1145/2983323.2983351](https://doi.org/10.1145/2983323.2983351)
- Bahi, S., Higgins, D., & Staley, P. (2015). A Time Hazard Analysis of Student Persistence: a US University Undergraduate Mathematics Major Experience. *International Journal of Science and Mathematics Education*, 13(5), 1139–1160. doi: [10.1007/s10763-014-9538-9](https://doi.org/10.1007/s10763-014-9538-9)
- Bean, J. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12(2), 155–187. doi: [10.1007/BF00976194](https://doi.org/10.1007/BF00976194)
- Bean, J. (1983). The Application of a Model of Turnover in Work Organizations to the Student Attrition Process. *The Review of Higher Education*, 6(2), 129–148. doi: [10.1353/rhe.1983.0026](https://doi.org/10.1353/rhe.1983.0026)
- Bean, J., & Metzner, B. (1985). A Conceptual Model of Nontraditional Undergraduate Student Attrition. *Review of Educational Research*, 55(4), 485–540. doi: [10.2307/1170245](https://doi.org/10.2307/1170245)
- Bound, J., & Turner, S. (2011). Dropouts and Diplomas. The Divergence in Collegiate Outcomes. In *Handbook of the Economics of Education* (Vol. 4, pp. 573–613). doi: [10.1016/B978-0-444-53444-6.00008-0](https://doi.org/10.1016/B978-0-444-53444-6.00008-0)
- Braxton, J. M., Hirschy, A. S., & McClendon, S. A. (2004). Understanding and reducing college student departure. *ASHE-ERIC Higher Education Report*, 30(3), 108.
- Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College Persistence: Structural Equations Modeling Test of an Integrated Model of Student Retention. *The Journal of Higher Education*, 64(2), 123. doi: [10.2307/2960026](https://doi.org/10.2307/2960026)
- Cabrera, A. F., Stampen, J. O., & Hansen, W. L. (1990). Exploring the Effects of Ability to Pay on Persistence in College. *Review of Higher Education*, 13(3), 303–336.
- Cabrera, L., Bethencourt, J. T., Alvarez Pérez, P., & González Afonso, M. (2006). The problem of University dropout. *RELIEVE - Revista Electronica de Investigacion y Evaluacion Educativa*, 12(2), 171–203.
- COPAES. (2016). *Marco General de Referencia para los Procesos de Acreditación de Programas Académicos de Tipo Superior*. Mexico: Consejo Nacional para la Acreditación de la Educación Superior.

- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187–220.
- Davidson-Pilon, C., Kalderstam, J., Zivich, P., Kuhn, B., Fiore-Gartland, A., AbdealiJK, ... Rendeiro, A. F. (2019). *CamDavidsonPilon/lifelines: v0.23.1*. doi: 10.5281/ZENODO.3555617
- Imbrie, P. K., Lin, J. J. J., & Reid, K. (2010). Comparison of four methodologies for modeling student retention in engineering. *ASEE Annual Conference and Exposition, Conference Proceedings*.
- Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. In *Source: Journal of the American Statistical Association* (Vol. 53).
- Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis : a self-learning text*. New York, NY: Springer.
- Lassibille, G., & Gómez, L. N. (2008). Why do higher education students drop out? Evidence from Spain. *Education Economics*, 16(1), 89–105. doi: [10.1080/09645290701523267](https://doi.org/10.1080/09645290701523267)
- Muñoz, B. (2017). *Análisis de la deserción en las licenciaturas de la FCFM-BUAP mediante modelos de supervivencia*. Benemérita Universidad Autónoma de Puebla.
- Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the Retention of University Students. *Research in Higher Education*, 40(3), 355–371. doi: [10.1023/A:1018755201899](https://doi.org/10.1023/A:1018755201899)
- OECD. (2019). Education at a Glance 2019. In *Education at a Glance 2019: OECD Indicators*. doi: [10.1787/f8d7880d-en](https://doi.org/10.1787/f8d7880d-en)
- pandas development team, T. (2020). *pandas-dev/pandas: Pandas*. doi: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134)
- Pascarella, E. T., & Chapman, D. W. (1983). A multiinstitutional, path analytic validation of Tinto ' s model of college withdrawal. *American Educational Research Journal*, 20(1), 87–102. doi: [10.3102/00028312020001087](https://doi.org/10.3102/00028312020001087)
- Pascarella, E. T., & Terenzini, P. T. (1980). Predicting freshman persistence voluntary dropout decisions from a theoretical model. *The Journal of Higher Education*, 51(1), 60–75. doi: [10.2307/1981125](https://doi.org/10.2307/1981125)
- Shuman, R. B. (1956). *College Dropouts : An Overview*. 29(8), 347–350.
- Sosu, E. M., & Pheunpha, P. (2019). Trajectory of University Dropout: Investigating the Cumulative Effect of Academic Vulnerability and Proximity to Family Support. *Frontiers in Education*, 4, 6. doi: [10.3389/feduc.2019.00006](https://doi.org/10.3389/feduc.2019.00006)
- St. John, E. P., Paulsen, M. B., & Starkey, J. B. (1996). The nexus between college choice and persistence. *Research in Higher Education*, 37(2), 175–220. doi:

[10.1007/BF01730115](https://doi.org/10.1007/BF01730115)

- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45(1), 89–125. doi: [10.3102/00346543045001089](https://doi.org/10.3102/00346543045001089)
- Tinto, V. (2012). *Leaving college. [electronic resource] : rethinking the causes and cures of student attrition.*
- Torres, V. (2006). A Mixed Method Study Testing Data-Model Fit of a Retention Model for Latino/a Students at Urban Universities. *Journal of College Student Development*. doi: [10.1353/csd.2006.0037](https://doi.org/10.1353/csd.2006.0037)
- Vallejos, C. A., & Steel, M. F. J. (2017). Bayesian survival modelling of university outcomes. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 180(2), 613–631. doi: [10.1111/rssa.12211](https://doi.org/10.1111/rssa.12211)
- van der Zanden, P. J. A. C., Denessen, E., Cillessen, A. H. N., & Meijer, P. C. (2018, February 1). Domains and predictors of first-year student success: A systematic review. *Educational Research Review*, Vol. 23, pp. 57–77. doi: [10.1016/j.edurev.2018.01.001](https://doi.org/10.1016/j.edurev.2018.01.001)
- Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys*, 51(6), 1–39. doi: [10.1145/3214306](https://doi.org/10.1145/3214306)