



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**ANÁLISIS DE SUPERVIVENCIA EN LOS ALUMNOS
DE LA FACULTAD DE MEDICINA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

PABLO DANIEL ISLAS URRUTIA



**TUTORA
MAT. MARGARITA ELVIRA CHÁVEZ CANO**

CIUDAD UNIVERSITARIA, CD. MX 2017



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice General

Introducción	4
Planteamiento	7
Objetivo	8
Método	8
Capítulo 1. Introducción Análisis de Supervivencia	10
1.1 Variable Aleatoria: Tiempo de supervivencia	11
1.2 Funciones de Supervivencia	11
1.3 Función de densidad de probabilidad	12
1.4 Función de Supervivencia	12
1.5 Función de Riesgo	13
1.6 Función de Riesgo acumulado.....	13
Capítulo 2. Censura y Truncamiento	15
2.1 Censura.....	15
2.2 Censura por la derecha.....	15
2.3 Censura Tipo I.....	15
2.4 Censura por la izquierda	20
2.5 Truncamiento	22
2.6 Truncamiento por la izquierda.....	22
2.7 Truncamiento por la derecha	22
Capítulo 3. Estimación de la función de supervivencia	23
3.1 Modelos No Paramétricos	23
3.1.1 Estimador Kaplan-Meier	23
3.1.2 Varianza del Estimador Kaplan-Meier	25
3.1.3 Intervalos de confianza basados en el Estimador Kaplan-Meier	26

3.1.4 Estimador Nelson-Aalen	26
3.1.5 Varianza del Estimador Nelson-Aalen	27
3.1.6 Intervalos de confianza basada en el Estimador Nelson-Aalen.....	27
3.2 Modelos Semiparamétricos	28
3.2.1 Modelos de Riesgos Proporcionales	28
3.2.2 Modelos de Riesgos Proporcionales de dos grupos.....	28
3.2.3 Modelo de Riesgos Proporcionales general	29
3.2.4 Ajuste de la función de supervivencia en el modelo de riesgos proporcionales	31
3.2.5 Inclusión de variables numéricas al modelo de riesgos proporcionales	32
3.2.6 Inclusión de variables categóricas al modelo de riesgos proporcionales....	32
3.2.7 Estimación del modelo de riesgos proporcionales.....	33
Capítulo 4. Residuos	38
4.1 Residuales de Cox-Snell.....	39
4.2 Residuales de martingala.....	39
4.3 Residuales de devianza	40
Capítulo 5. Análisis Descriptivo	40
Capítulo 6. Análisis curvas de Supervivencia	61
Capítulo 7. Comparación de las funciones de supervivencia por variable	63
Capítulo 8. Regresión de Cox.....	85
Capítulo 9. Verificación del modelo de Cox	88
 Conclusiones	 90
 Bibliografía.....	 91

Introducción

La deserción escolar y la no acreditación en la educación media superior y superior es uno de los retos más grandes de la educación en México, por lo que es de gran valía estudiar estos fenómenos.

En nuestro país el rezago educativo y la deserción escolar son conflictos que tienen severas consecuencias, tanto sociales como económicas, y esto perjudica en gran medida el desarrollo de México. Por ello el avance industrial y económico, que desencadenarían un avance social, de nuestro país no ha podido generarse y no es posible competir en niveles educativos, sociales, ni tecnológicos, con países desarrollados.

La deserción escolar la entenderemos como el abandono de la escuela por los alumnos, y sus factores van desde lo económico hasta lo social, desde el salir a trabajar para ayudar a mantener a su familia como el ahora llamado “bullying”, mientras que el rezago educativo lo denotaremos como el grupo de personas mayores a 14 años que no han podido concluir sus estudios que le corresponderían a su edad.

La UNAM, al observar la deserción escolar que tiene en sus aulas, se preocupó por la situación y generó un proyecto encabezado por el Departamento de Salud Mental de la Facultad de Medicina, donde se encuestó a más de 6 mil alumnos de diferentes áreas y se encontraron varios factores que influyen de una manera determinante en la deserción escolar, siendo la depresión y factores económicos las más graves. En este estudio se observó que las carreras con más niveles de depresión, y por lo tanto deserción escolar, son Medicina, las de Ciencias, Derecho, Música y Arquitectura.¹

La Licenciatura en Medicina es una de las carreras con mayor demanda estudiantil, sin embargo, también es uno de los campos de estudio a nivel superior con mayor carga de trabajo debido a su alta exigencia académica. Por ello, una de las

¹ Dirección General de Comunicación Social (2015, noviembre). Estimulación magnética transcraneal, opción para tratar depresión moderada.

preocupaciones esenciales en la Facultad de Medicina de la UNAM es el alto índice de estudiantes con bajo rendimiento académico, así como la deserción escolar que sucede particularmente entre los alumnos de los primeros años de la licenciatura (Informe de la Facultad de Medicina 2008-2016). Lo anterior, demanda una serie de retos institucionales que implican la búsqueda de soluciones a dicho problema por medio de estudios tanto en el ámbito pedagógico, psicológico y actuarial.

Haciendo hincapié en la gran exigencia académica que demandan los estudios de medicina, es imprescindible analizar lo referente a las necesidades en la enseñanza para la formación en el sector de la salud, sin dejar de lado el excesivo número de estudiantes que ingresan a la carrera de medicina a quienes se les pretende dar una adecuada formación.

Posterior a la implementación del Plan de Estudios 2010 de médico cirujano en la generación 2011, se ha observado un alto índice de alumnos que no acreditan las asignaturas del primer año, dando como consecuencia un considerable porcentaje de alumnos repetidores; lo que ha motivado a varios académicos a desarrollar investigación para estudiar esta problemática. Así, se ha demostrado que los antecedentes académicos desempeñan un papel muy importante para acreditar las diferentes asignaturas.²

Aunado a lo anterior, la carrera de medicina, debido a su carga curricular exige a sus estudiantes una alta entrega y dedicación, mismas que para algunos de ellos puede convertirse en un problema con el que no pueden lidiar y que en algunos casos puede producirse una alteración del estado de salud física y mental de los alumnos. Otros factores que afectan el rendimiento académico son los señalados por algunos autores, como el estrés, la ansiedad y la depresión experimentados por los estudiantes previo al inicio de la carrera y a lo largo del ciclo escolar. Estos factores originan un círculo vicioso en el estudiante ya que son causa y consecuencia de problemas como el bajo desempeño académico.³

² Urrutia et al, 2014

³ Foulloux et al, 2013

Inventario de Depresión de Beck

Según la Organización de la Salud la depresión es una enfermedad en la cual existe una tristeza persistente en el individuo que la padece, generando una pérdida de interés en cosas que disfruta realizar.⁴ Afectando actividades cotidianas, de este punto surge la importancia en nuestro estudio.

El Inventario de Depresión de Beck creado en 1961 por Aaron T. Beck, Robert A. Steer y Garbin⁵, surge con el objetivo de detectar los síntomas depresivos y evaluar la gravedad de la depresión. Es un cuestionario auto aplicado que contiene 21 ítems.

Para Méxicoc en 1996 se obtuvo la última versión del Inventario actualizada y revisada, es el BDI-II. Se estandarizó el Inventario de Depresión de Beck, en esta se manejan cuatro escalas de depresión que son: mínima, leve, moderada y severa.⁶

Estilos de Afrontamiento al Estrés

La Organización Mundial de la Salud define como estrés al fenómeno de reacciones fisiológicas que en conjunto preparan al organismo para una acción. Cuando el fenómeno se genera en exceso, produce una sobrecarga de tensión que impide el normal desarrollo y funcionamiento del cuerpo humano.⁷

También se podría definir como el resultado de la interacción entre los estímulos estresantes y los recursos de la persona.

La licenciatura en Medicina al ser de las más demandantes es causante de estrés entre sus estudiantes.

⁴ OMS, 2003

⁵ Beck T A, Steer A R, Garbin G M; Psychometric properties of the Beck Depression Inventory. Clinical Psychology Review, 1988

⁶ Jurado S, Villegas M E, Méndez L, Rodríguez F, Loperena V, Varela R; La estandarización del inventario de depresión de Beck para los residentes de la ciudad de México, 1998

⁷ OMS, 1994

Es importante entender el modo en que los alumnos de nuevo ingreso enfrentarán las situaciones de estrés a partir del inicio escolar. De este modo es trascendental conocer los modos de afrontamiento al estrés de los alumnos.

Lazarus y Folkman definen a los modos de afrontamiento como “aquellos esfuerzos cognitivos y conductuales constantemente cambiantes que se desarrollan para manejar las demandas específicas externas o internas”. Y construyeron un instrumento que mide el afrontamiento a partir de una serie de entrevistas.⁸

En México se tiene la traducción del instrumento al español. Sotelo y Maupome tradujeron y estandarizaron la escala de modos de afrontamiento de Lazarus y Folkman, y contiene seis factores: 1. Afrontamiento dirigido al problema, 2. Pensamiento mágico, 3. Distanciamiento, 4. Afrontamiento dirigido a lo positivo, 5. Revaloración positiva y 6. Flexibilidad.⁹

Planteamiento

Con el propósito de diseñar un modelo que ayude a identificar los posibles factores que generan un bajo desempeño escolar, los datos que se analizan en el presente trabajo, fueron proporcionados por la Facultad de Medicina de la UNAM. La población fue la generación 2015 y sus primeros dos años de carrera.

El primer año de la carrera de Medicina es uno de los más importantes, ya que los resultados obtenidos en éste serán de suma importancia para tomar la decisión de seguir adelante o abandonar la carrera. Es claro, que cada estudiante tiene una formación académica y psicológica distinta que influirá en el proceso. Es por lo que es de suma importancia realizar estudios estadísticos que identifiquen los factores que ponen en riesgo académico a la población de estudiantes. El estudio se justifica por la preocupación de las autoridades académicas de evitar la deserción e implementar medidas que ayuden a los alumnos a continuar con sus estudios.

⁸ Lazarus R, Folkman S; El concepto de afrontamiento en estrés y procesos cognitivos, 1984

⁹ Sotelo A C, Maupome S V; Traducción y estandarización del cuestionario modos de afrontamiento al estrés de Lazarus y Folkman para una población de adolescentes mexicanos. Tesis de licenciatura. Universidad Nacional Autónoma de México, 2000

Además, será un estudio con información fidedigna y de apoyo a la institución para que con esto se puedan ejercer medidas preventivas y evitar el abandono escolar.

Objetivo

El presente estudio tiene como propósito identificar las variables asociadas al problema de reprobación en los primeros dos años para los alumnos de la licenciatura de medicina de la facultad de Medicina por medio de un estudio de supervivencia, a través de una base de datos con la información correspondiente a todos los estudiantes inscritos.

El objetivo de este trabajo es estimar la función de supervivencia a través del estimador Kaplan Meier y la regresión de Cox, de un grupo de la muestra en un intervalo definido de tiempo.

Método

El periodo de estudio fue de 24 meses (agosto 2014- abril 2016).

La base de datos se conformó por 1,256 estudiantes inscritos en la licenciatura de Médico Cirujano de la generación 2015, de los cuales el 64% eran del sexo femenino y el 36% masculino.

Para poder realizar este estudio, se consideraron la aplicación de diferentes instrumentos:

- Examen Diagnóstico al inicio del año escolar.

Se aplicó al inicio del año escolar. Permite identificar el grado de nivel académico que poseen los alumnos al ingresar a la Facultad de Medicina, este consta de 120 reactivos en los que se evalúan las áreas de matemáticas, física, química, biología, historia universal, historia de México, literatura y geografía; clasificándolos como 'Muy Alto', 'Alto', 'Medio', 'Bajo' y 'Muy Bajo'.

- Inventario de Depresión de Beck.

Es uno de los instrumentos más empleados para la evaluación de los síntomas depresivos en adolescentes y adultos. Consta de 21 reactivos autoaplicables, cada reactivo consiste en una serie de cuatro afirmaciones de las cuales una tiene que ser seleccionada en relación a la forma en la que se ha sentido el paciente en la última semana. La puntuación total varía de 0 a 63. Se consideró una puntuación mayor o igual a 13 puntos como punto de corte para identificar a los casos con probable depresión y se clasifica en: 'Leve', 'Mínima', 'Moderada' y 'Severa'.

- Cuestionario de Afrontamiento al Estrés, medida con 66 reactivos clasificados en los seis factores antes descritos, y en la cual se describe cuál es el tipo de afrontamiento predominante de cada alumno.
- Evaluaciones del rendimiento académico, con registro de calificación numérico.

Para la evaluación del rendimiento escolar se tomaron en cuenta los resultados obtenidos por los estudiantes de agosto 2015 a mayo 2016, en los exámenes departamentales de primer año en las asignaturas de Anatomía, Biología Celular e Histología Médica, Bioquímica y Biología Molecular y Embriología. Así como las obtenidas en el segundo año en las asignaturas de Inmunología, Microbiología y Parasitología, Farmacología y Fisiología.

Capítulo 1. Introducción Análisis de Supervivencia

Existen estudios en los cuales, para uno o varios individuos, la variable de interés es el tiempo transcurrido entre el inicio del estudio hasta la ocurrencia de un evento puntual definido, el evento marcará el término del estudio para cada individuo y se denominará falla.

El análisis de supervivencia es el conjunto de técnicas que tiene como objetivo modelar el tiempo de interés, también conocido como: tiempo de falla, tiempo de supervivencia o tiempo de muerte, debe ser una medida cuantitativa como periodos, minutos, días, meses, años, entre otros.¹⁰

El término supervivencia se debe a que en las primeras aplicaciones del método se utilizaba como evento la muerte de un paciente.

En campos como la medicina el análisis de supervivencia se aplica para, por ejemplo, conocer el tiempo que le toma a un organismo morir o el tiempo en el que un paciente tendrá respuesta a un tratamiento.

La utilización de modelos específicos paramétricos, no paramétricos y semi-paramétricos es de suma importancia para una buena descripción del comportamiento de los datos del estudio.

Una característica importante en el análisis de supervivencia es la existencia de información parcial sobre el tiempo de falla de uno o más individuos. Es frecuente que se presente la desaparición de individuos durante el estudio que no completarán el seguimiento, como también existirán los casos en los que no se produzca el suceso durante la ventana de tiempo del estudio. Lo anterior, indica que para algunos sujetos no se sabrá el momento exacto en el que se presenta el suceso a estudiar a los que se llama datos censurado de la variable tiempo, por lo que los métodos de estimación tienen que ser adaptados, permitiendo que estos datos no se consideren como perdidos y puedan ser incluidos en el estudio.

¹⁰ Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

1.1 Variable Aleatoria: Tiempo de supervivencia

Como se mencionó anteriormente, se denomina tiempo de supervivencia al periodo transcurrido desde el inicio de estudio hasta la ocurrencia del suceso (falla) o evento previamente definido (figura 1).

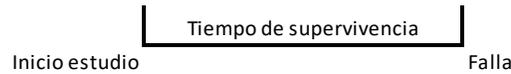


Figura 1 Representación Tiempo de supervivencia

Es importante señalar que el tiempo origen no necesariamente tiene que ser el mismo para todos los individuos del estudio, este tiempo puede ser escalonado para cada uno, es decir, cada individuo podrá entrar al estudio en cualquier momento. La falla deberá ocurrir a lo más una vez en cada individuo.

El tiempo de supervivencia se llamará la variable T , siendo una variable aleatoria no negativa ($T \geq 0$) que podrá caracterizar un modelo de supervivencia.¹¹

1.2 Funciones de Supervivencia

John P. Klein y Melvin L. Moeschberger explican que la distribución de la variable T puede ser caracterizada o descrita por las siguientes funciones: *función de supervivencia* que es la probabilidad de supervivencia o de que el sujeto no presente el evento de interés antes de un momento o tiempo t , la *función de riesgo* indica la posibilidad de que a algún individuo le ocurra el evento en un intervalo muy corto, *función de densidad de probabilidad* representa la probabilidad de ocurrencia del evento al tiempo t .¹²

Las tres funciones anteriores están relacionadas entre sí, es decir, si se conoce alguna de ellas, se puede obtener cualquiera de las restantes.

¹¹, ¹² Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

1.3 Función de densidad de probabilidad

El tiempo de supervivencia T al ser una variable aleatoria continua, tiene asociada una función de densidad de probabilidad, la cual permite calcular la probabilidad de que la variable tome valores en un intervalo, a través de:

$$P(a < T < b) = \int_a^b f(t)dt$$

1.4 Función de Supervivencia

De acuerdo con lo publicado por David Collet, es importante mencionar que la variable T también tiene asociada una función de distribución que representa la probabilidad de que el tiempo de supervivencia sea menor que algún valor t ¹³ y está dado por:

$$F(t) = P(T \leq t) = \int_0^t f(u)du$$

Una vez definida la función de distribución podemos mencionar a la función de supervivencia denotada por $S(t)$ y que refiere a la probabilidad de que un individuo a partir del tiempo de entrada al estudio sobreviva más allá del tiempo t , definida de la siguiente manera:

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

En otras palabras, representa la probabilidad de que algún individuo experimente el evento de interés o falla después del tiempo t .

$$S(t) = P(T > t) = 1 - F(t) = 1 - P(T \leq t)$$

La función de supervivencia es la integral de la función de densidad de probabilidad y es:

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

¹³ Collet D; Modelling survival data in medical research (2003)

Cabe resaltar, que la probabilidad de supervivencia pasado el tiempo cero o tiempo de entrada es igual a 1, es decir, $S(0)=1$. Y si el tiempo se aproxima a infinito la probabilidad es igual a cero, es decir, $S(t)=0$ si $t \rightarrow \infty$.

1.5 Función de Riesgo

La función de riesgo se define como la tasa instantánea de que el evento ocurra, condicionado a que el individuo no ha presentado dicho evento al inicio del instante. La función de riesgo es obtenida como la probabilidad de que un individuo falle dentro del intervalo t y $t+\Delta t$, dado que el individuo ha sobrevivido hasta el tiempo t , es decir,

$$P (t \leq T \leq t+\Delta t | T > t)$$

La función de riesgo denotada por $h(t)$, es la tasa de la probabilidad condicional por unidad de tiempo, y es el límite de ese valor cuando Δt tiende a cero. La fórmula matemática de la función de riesgo se define a continuación:

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P (t \leq T \leq t + \Delta t | T > t)}{\Delta t} \right\}$$

1.6 Función de Riesgo acumulado

La Función de Riesgo acumulado, $H(t)$, como su nombre lo indica es el riesgo acumulado de ocurrencia del evento al tiempo t . Ésta es una función no decreciente y corresponde a:

$$H(t) = \int_0^t h(u)du$$

En otras palabras, la función de riesgo acumulado al tiempo t es el valor esperado de eventos que ocurren en el intervalo a partir del tiempo de inicio al tiempo t .

Como se mencionó anteriormente, las funciones antes descritas están relacionadas y son equivalentes entre sí, conociendo una de ellas, las restantes pueden ser obtenidas.

1. Se sabe que la función de supervivencia está dada por

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

entonces,

$$f(t) = -\frac{dS(t)}{dt}$$

2. La función de riesgo se define como

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t \leq T \leq t + \Delta t | T > t)}{\Delta t} \right\}$$

entonces,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log S(t)$$

3. La función de riesgo acumulado está dada por

$$H(t) = \int_0^t h(u)du = -\ln(S(t))$$

4. Por último, se tiene que

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(u)du\right]$$

Capítulo 2. Censura y Truncamiento

2.1 Censura

Como anteriormente se mencionó existen datos para los cuales los métodos de estimación tienen que ser modificados, esto con la finalidad de no tener que rescindir de información.

Estos datos son llamados datos censurados y existen distintos tipos de censura, diferenciadas por la forma en que han sido obtenidos los datos de supervivencia. Los tipos de censura que se tomaron en cuenta para realizar este trabajo son según los descritos por Klein y Moeschberger en 2003.

2.2 Censura por la derecha

Previo conocer la censura por la derecha, es necesario definir la censura de tipo I.

2.3 Censura Tipo I

Existen estudios, en los cuales previo a su comienzo se define el tiempo de duración total. Cuando el tiempo de supervivencia de algún individuo es observado sólo si éste es menor al tiempo de duración definido, se conoce como *Censura tipo 1*.¹⁴ En caso de no ser observado, el individuo u observación se considera censura por la derecha.

A continuación, se define un ejemplo. Un investigador realiza un estudio, por situaciones de tiempo o ahorro, éste decide terminar anticipadamente con el estudio antes de que el evento se presente en todos los individuos. En este caso, si no se tuvieron pérdidas o abandonos, todas las observaciones de los individuos que fueron forzados a terminar, son datos censurados por la derecha con una longitud de tiempo igual a la del estudio.

¹⁴ Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

Dicho lo anterior, se supone que para un individuo específico que se encuentra bajo estudio tiene un tiempo de supervivencia T y un tiempo fijo de censura Cr ¹⁵.

El tiempo de supervivencia T exacto de un individuo será conocido, si y sólo si, $T \leq Cr$. Si T es mayor que Cr , el individuo es un sobreviviente con un tiempo de vida censurado igual a Cr .

En presencia de censura por la derecha se tendrá la pareja de variables aleatorias (X, δ) para cada individuo. Donde δ indica si el tiempo de supervivencia T es observado ($\delta=1$) o si es censurado ($\delta=0$). X es igual a T , si el tiempo de supervivencia es observado, e igual a Cr si es censurado, i.e., $X=\min(T, Cr)$.

Como siguiente ejemplo, considere un experimento en el cual un grupo de ratones fueron alimentados con una dosis específica de cancerígeno. El objetivo del estudio es evaluar el efecto del cancerígeno en la supervivencia de cada ratón. Los ratones fueron seguidos desde el inicio del estudio hasta su muerte o hasta que se alcanzó un tiempo preestablecido y todos los sobrevivientes fueron sacrificados (censura¹⁶).

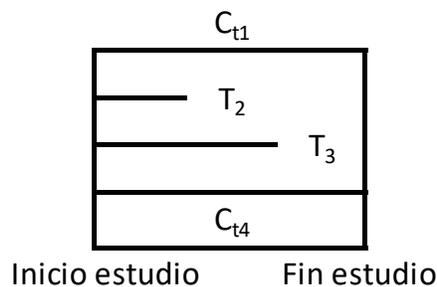


Figura 2 Ejemplo de censura Tipo I.

En la figura 2 se observa que en el estudio hay cuatro individuos, T_2 y T_3 presentaron el evento durante el estudio, mientras que el resto C_{t1} y C_{t4} no presentaron el evento y se consideran datos censurados.

¹⁵ Right Censoring time (Censura por la derecha)

¹⁶ Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

Por otro lado, existen estudios en los cuales previo al comienzo del estudio se tienen fijados distintos tiempos de censura, este tipo de censura es conocido como Censura Tipo I progresiva¹⁷.

El siguiente caso es considerado un ejemplo de este tipo de censura. Considere un estudio con 200 ratones que fueron divididos aleatoriamente en dos grupos (ver figura 3). Para la reducción de gastos de mantenimiento se decidió previamente que los ratones del primer grupo que sobrevivan a la semana 42 serán sacrificados y los del segundo grupo que lleguen vivos a la semana 104 también serán sacrificados. Es decir, el tiempo de censura para el primer grupo es la semana 42 y para el segundo la semana 104.

1er tiempo de sacrificio= semana 42

2do tiempo de sacrificio=semana 104

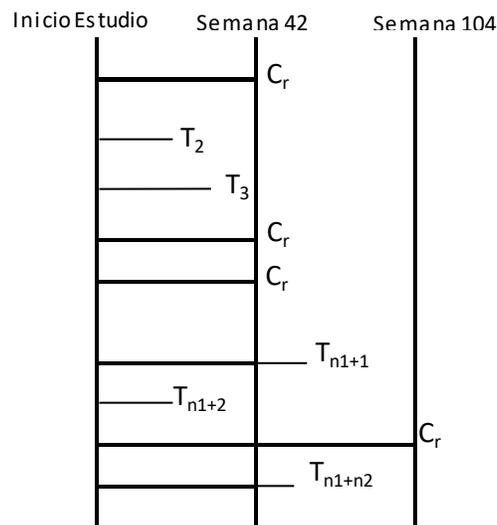


Figura 3 Ejemplo de censura Tipo I progresiva (2 distintos tiempos de sacrificio).

En los casos anteriores el comienzo del estudio es el mismo para todos los ratones, pero cuando los ratones entran al estudio en distintos tiempos y tienen un fin de estudio (predeterminado) igual para todos, es decir, el tiempo de censura es

¹⁷ Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

conocido por el investigador al momento de entrar al estudio (ver figura 4). Este tipo de censura es llamado de Tipo I generalizada¹⁸.

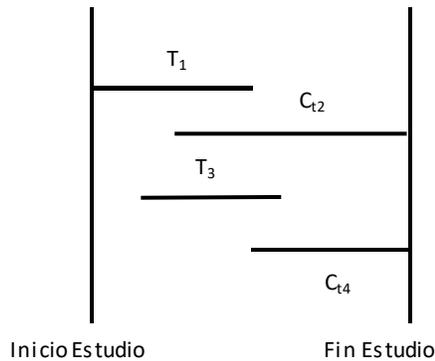


Figura 4 Ejemplo de censura Tipo I generalizada (tiempo de inicio distinto en cada individuo).

Otra forma conveniente de graficar la censura Tipo I generalizada es moviendo a cada individuo al tiempo de inicio cero, como se muestra en la figura 5.

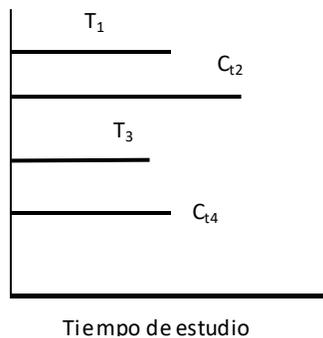


Figura 5 Ejemplo de censura Tipo I generalizada para los cuatro individuos de la Figura con un Inicio de estudio en cero. $X_1=T_1$ (tiempo de supervivencia para el 1er individuo) ($\delta=1$); $X_2=C_{t2}$ (tiempo de censura por la derecha para el 2do individuo) ($\delta=0$); $X_3=T_3$ (tiempo de supervivencia para el 3er individuo) ($\delta=1$); $X_4=C_{t4}$ (tiempo de censura por la derecha para el 4to individuo) ($\delta=0$)

Existe otro método con el cual podemos representar los resultados de distinta manera, es el diagrama de Lexis¹⁹. Este cuenta con dos principales componentes:

^{18,19} Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

el tiempo calendario, que se encuentra en el eje horizontal, y la longitud del tiempo de supervivencia se representa por una línea de 45°.

La altura de cada rayo representa el tiempo que cada individuo se encuentra dentro del estudio. La figura 6 se muestra la representación del diagrama de Lexis del ejemplo anterior (la censura de tipo I generalizada).

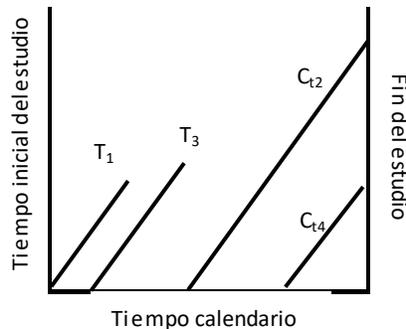


Figura 6 *Ejemplo Diagrama de Lexis para la censura tipo I generalizada*

Ahora suponga que previo al inicio del estudio, el investigador decide que el tiempo de duración del estudio sea hasta que se presente el evento en los primeros r individuos ($0 < r < n$).

Este tipo de censura por la derecha es llamada censura tipo II, y ejemplo claro de ellos son las pruebas de vida de equipos o máquinas, todos ellos ingresan al estudio al mismo tiempo, pero esperar a que suceda la falla en todos podría tomar demasiado tiempo, por lo que se decide terminar cuando r de n equipos hayan fallado.

Es importante hacer notar que tanto r (número de fallas) y $n-r$ (número de observaciones censuradas) son enteros fijos y el tiempo de censura $C(r)$ es aleatorio.

Una generalización de la Censura Tipo II²⁰ similar a la generalización en la Censura Tipo I con distintos tiempos de censura, es la Censura Tipo II progresiva. En ella,

²⁰ Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

las primeras r_1 fallas (entero escogido previamente al inicio del estudio) de una muestra de n elementos son anotadas. Después, de los $n-r_1$ supervivientes n_1-r_1 individuos serán retirados del experimento, quedando $n-n_1$ individuos en el estudio. Cuando las siguientes r_2 fallas (entero escogido previamente al inicio del estudio) se presenten, n_2-r_2 individuos serán excluidos de los supervivientes. Este proceso continuará hasta que se hayan completado una serie de repeticiones predeterminadas. Los tiempos de censura $C(r_1)$ y $C(n_1+r_2)$ son aleatorios.

Se puede notar que en la Censura Tipo I progresiva²¹ los tiempos de censura son fijados previos al inicio del estudio, mientras que en la Censura Tipo II progresiva los tiempos de censura son aleatorios a la ocurrencia de un predeterminado número de fallas.

Por último, existe otro tipo de censura llamada censura aleatoria, la cual surge cuando algún individuo del estudio experimenta un evento ajeno al estudio, causando que sea excluido del estudio y el evento de interés no es observado en estos individuos.

2.4 Censura por la izquierda

Contrario a las definiciones anteriores, el tiempo de supervivencia de un individuo dentro de un estudio es considerado censurado por la izquierda, si su tiempo de supervivencia T es menor que el tiempo de censura C ²² (Klein y Moeschberger en 2003), es decir, que el evento de interés le ha ocurrido al individuo antes de que el sujeto haya sido observado por el investigador al tiempo C y su tiempo de evento exacto es desconocido. El tiempo de supervivencia T exacto será conocido, si y sólo si, $T \geq C$.

En presencia de censura por la izquierda se tendrá la pareja de variables aleatorias (X, \mathcal{E}) , para cada individuo, como vimos anteriormente, donde X es igual T si el tiempo de supervivencia es observado y \mathcal{E} indica si el tiempo de supervivencia

²¹ Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

²² Left Censoring time (censura por la izquierda)

exacto T es observado ($\mathcal{E}=1$) o no lo es ($\mathcal{E}=0$). En la censura por la izquierda en contraste con la censura por la derecha, $X=\max(T, C)$.

A continuación, se muestra un ejemplo de censura por la izquierda (Klein y Moeschberger en 2003). Suponga que se realiza un estudio en estudiantes de preparatoria para determinar la distribución del tiempo hasta el primer contacto con la marihuana. La pregunta que se hace es: ¿Cuándo fue la primera vez que probaste la marihuana? Uno de los individuos responde “La he consumido, pero no recuerdo el primer contacto con ella.” Por la respuesta se sabe que el evento de interés le ha ocurrido al individuo antes de que se realizara la encuesta, pero se desconoce el momento exacto en cual el evento ocurrió.

Dadas las definiciones anteriores, podremos determinar el siguiente tipo de censura en la cual en algunos individuos del estudio se presenta censura por la derecha, mientras que otros presentan censura por la izquierda y los tiempos de supervivencia son considerados doblemente censurados. De la misma forma que los casos anteriores, los datos pueden ser representados por la pareja de variables aleatorias (X, δ) , donde X representa el tiempo de estudio, $X=\max[\min(T, C_r), C_l]$; y δ indica si X es el tiempo en el que ocurre el evento ($\delta=1$), si X es el tiempo censurado por la derecha ($\delta=0$) o si X es el tiempo censurado por la izquierda ($\delta=-1$). Aquí, C_l es el tiempo anterior al que los individuos presenten el evento y C_r es el tiempo posterior a que los individuos presenten el evento. El tiempo de supervivencia T será conocido si éste es menor o igual que C_r y mayor o igual que C_l .

Por último, un tipo de censura más general es cuando se conoce que el tiempo de supervivencia ocurre dentro de un intervalo. Esta censura por intervalo ocurre en estudios longitudinales con seguimientos periódicos y el tiempo en el que ocurre el evento en los pacientes sólo es conocido que ocurrió en el intervalo $(L_i, R_i]$. Ejemplo de este tipo de censura, son los experimentos industriales en donde se tienen revisiones periódicas para conocer el estado funcional de las herramientas de los equipos, aquí existirán individuos o equipos los cuales presentan la falla entre una revisión y otra, es decir, se conoce que la falla ocurrió dentro de un intervalo.

2.5 Truncamiento

Otra característica de los datos de tiempos de supervivencia, a veces confundido con la censura, es el truncamiento.

En el truncamiento son considerados únicamente los individuos cuyo evento de interés ocurre dentro de la ventana de observación (Y_L, Y_R). Para los individuos que presenten el evento de interés fuera del intervalo no se tiene información disponible y no son considerados en el estudio por el investigador²³.

2.6 Truncamiento por la izquierda

En este tipo de truncamiento sólo son observados aquellos individuos cuyo tiempo de supervivencia T exceda el tiempo de truncamiento Y_L , y aquellos individuos cuyo tiempo de supervivencia sea inferior al de truncamiento no son considerados como parte del estudio.

En el truncamiento por la izquierda para que los individuos entren al estudio es necesario que ocurra el evento con tiempo de truncamiento Y_L , es por ello que se le conoce como 'tiempo de retraso de entrada'. Una vez ocurrido este tiempo, los individuos serán observados hasta que el evento de interés del estudio ocurra o hasta que el evento sea censurado.

Un ejemplo claro son los estudios del diámetro de partículas microscópicas. En este estudio sólo las partículas suficientemente grandes que puedan ser vistas por la resolución del microscopio serán observadas y las pequeñas no podrán ser observadas por el investigador.

2.7 Truncamiento por la derecha

En este truncamiento son observados los tiempos de supervivencia T sólo cuando es menor al tiempo de truncamiento Y_R ($T \leq Y_R$).

²³ Klein J, Moeschberger M; Survival Analysis Techniques for Censored and Truncated Data (2003)

Ejemplo de ello, son los estudios de distancias de la Tierra a las estrellas. Aquellas estrellas tan lejanas que no pueden ser observadas por el investigador no serán consideradas dentro del estudio.

En los datos censurados, a diferencia del truncamiento, existe información parcial de cada individuo.

Capítulo 3. Estimación de la función de supervivencia

3.1 Modelos No Paramétricos

En ocasiones los datos del estudio nos permiten utilizar métodos analíticos y gráficos que no requieren de supuestos distribucionales, estos métodos son los No Paramétricos.

3.1.1 Estimador Kaplan-Meier

El estimador de Kaplan-Meier²⁴ introducido por Edward L. Kaplan y Paul Meier se convirtió en uno de los métodos más utilizados para la estimación de la función de supervivencia.

Esta técnica de estimación no paramétrica toma en cuenta toda la información disponible en el estudio, casos censurados y no censurados.

Suponga que en el estudio hay n individuos y k fallas en el estudio con $k \leq n$, lo cual indica que $c = n - k$ es el número de casos censurados en el estudio.

Los tiempos de supervivencia estarán ordenados de forma ascendente y formarán intervalos entre ellos, estos intervalos no comienzan en tiempos de censura. Los tiempos de supervivencia se ordenan como se muestra a continuación:

$$t_1 < t_2 < t_3 < \dots < t_k$$

²⁴ Kaplan E. L., Meier P.; Nonparametric estimation from incomplete observations (1958)

Observe que el intervalo entre el tiempo origen del estudio (t_0) y el tiempo justo antes de la presencia de la primera falla (t_1) no incluirá ninguna falla.

El método de Kaplan-Meier estima la probabilidad de supervivencia en el tiempo t_j mediante el producto de la probabilidad de supervivencia en el tiempo t_{j-1} por la probabilidad condicional de sobrevivir al tiempo t_j si se ha sobrevivido hasta el tiempo t_{j-1} .

A continuación, se muestra con detalle lo mencionado anteriormente. Sea el j -ésimo tiempo de falla t_j con $j= 1,2,\dots,r$ y el intervalo formado desde $t_{j-\delta}$ con $\delta > 0$. Se sabe que hay n_j (número de individuos en riesgo inmediatamente antes del tiempo t_j) y d_j (número de fallas o muertes al instante t_j), entonces la probabilidad de supervivencia de un individuo en este intervalo es $\frac{d_j}{n_j}$. Por lo tanto, la probabilidad de sobrevivir en el intervalo es $\frac{(n_j - d_j)}{n_j}$.

Es importante señalar que, al momento de calcular el número de individuos en riesgo en cada punto, los datos censurados son calculados como si ocurrieran inmediatamente después del tiempo de falla cuando se calcula el número de individuos n_j .

Así, el estimador de Kaplan-Meier de la función de supervivencia se define de la siguiente forma:

$$S(t) = \prod_{j=1}^k \frac{n_j - d_j}{n_j} = \prod_{j=1}^k \left(1 - \frac{d_j}{n_j}\right)$$

En caso de que en el estudio no se hayan observado todas las fallas y el último dato registrado es un tiempo de supervivencia censurado, t^* , entonces la estimación de la función de supervivencia esta indefinida, de tal forma que no se podrá estimar la curva de supervivencia hasta el final del estudio.

3.1.2 Varianza del Estimador Kaplan-Meier

Para la estimación de la función de supervivencia es importante la construcción de un buen estimador de la varianza. De esta forma, se tendrá una medida de precisión del estimador Kaplan-Meier de $S(t)$ en distintos instantes del tiempo.

A continuación, se tratará el enfoque de la fórmula de Greenwood (1926).

El estimador Kaplan-Meier de la función de supervivencia dado por:

$$S(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right)$$

Ahora,

$$\log(S(t)) = \log \left(\prod_{j=1}^k \frac{n_j - d_j}{n_j} \right) = \sum_{j=1}^k \log \left(\frac{n_j - d_j}{n_j} \right) = \sum_{j=1}^k \log(p_j)$$

Entonces,

$$\text{Var}(\log(S(t))) = \sum_{j=1}^k \text{Var}(\log(p_j))$$

Aplicando el método delta se tiene:

$$\text{Var}(\log(p_j)) \approx \left(\frac{1}{\pi_j} \right)^2 \frac{\pi_j(1 - \pi_j)}{n_j} = \left(\frac{1}{\pi_j} \right) \frac{1 - \pi_j}{n_j}$$

Ahora,

$$\text{Var}(S(t)) \approx [S(t)]^2 \text{Var}(\log(S(t))) = [S(t)]^2 \sum_{j=1}^k \left(\frac{1}{\pi_j} \right) \frac{1 - \pi_j}{n_j}$$

Reemplazando $p_j = \frac{n_j - d_j}{n_j}$ por π_j , se tiene

$$Var(S(t)) = [S(t)]^2 \sum_{j=1}^k \frac{dj}{nj [nj - dj]}$$

3.1.3 Intervalos de confianza basados en el Estimador Kaplan-Meier

Bajo el supuesto de normalidad asintótica de los estimadores de máxima verosimilitud, el intervalo de $(1 - \alpha) \times 100\%$ de confianza para la función de supervivencia $S(t)$ para un valor de t , puede ser calculado de la siguiente manera:

$$\tilde{S}(t) \pm Z_{1-\alpha/2} \sqrt{Var(\tilde{S}(t))}$$

Donde $Z_{1-\alpha/2}$ denota el cuantil de $1 - \frac{\alpha}{2}$ de una distribución normal estándar, es decir,

$$\mathbb{P}\left(Z < Z_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} \quad \text{con } Z \sim N(0,1)$$

3.1.4 Estimador Nelson-Aalen

Estimador de gran utilidad para los casos o estudios en los cuales se tienen tamaño de muestras pequeñas.

Basado en la función de Riesgo Acumulada, ya que la función de supervivencia se puede expresar en términos de ésta ($H(t) = -\ln(S(t))$), y en este caso un estimador natural de la función de Riesgo Acumulada es donde $\tilde{S}(t)$ es el estimador Kaplan-Meier.

Ahora, el estimador definido por Nelson y Aalen²⁵ puede obtenerse de la suma acumulada de la estimación empírica de la función de riesgo

²⁵ Nelson W; Hazard plotting for incomplete failure data (1969)

Nelson W; Theory and applications of hazard plotting for censored failure data (1972)

Aalen O.; Nonparametric inference for a family of counting processes (1978)

$$H(t) = \sum_{j=1}^k \frac{d_j}{n_j}$$

Donde d_j representa el número de fallas ocurridas en el momento t_j y n_j es el número de individuos en riesgo al instante t_j .

Entonces, el estimador Nelson-Aalen puede ser usado como estimador de la función de supervivencia mediante la transformación:

$$S(t) = e^{-H(t)}$$

3.1.5 Varianza del Estimador Nelson-Aalen

Para el cálculo de la varianza de la función de supervivencia basada en la función de riesgo acumulado se tiene:

$$\text{Var}(\widehat{H}(t)) = \sum_{j=1}^k \frac{d_j}{n_j^2}$$

Aalen y Johansen (1978) proponen para la estimación de la varianza de $S(t)$ como:

$$\text{Var}(\widehat{S}(t)) = S(t)^2 \sum_{j=1}^k \frac{d_j}{n_j^2}$$

Intervalos de confianza basada en el Estimador Nelson-Aalen

El intervalo de confianza de $100(1-\alpha) \%$ para la función de supervivencia $S(t)$, mediante el estimador Nelson-Aalen está dado por:

$$\widehat{S}(t) \pm Z_{1-\alpha/2} \sqrt{\text{Var}(\widehat{S}(t))}$$

3.2 Modelos Semiparamétricos

Los modelos semiparamétricos permiten evaluar el impacto o efecto provocado por las covariables sobre el tiempo de falla del individuo en estudio.

3.2.1 Modelos de Riesgos Proporcionales

Modelo semiparamétrico propuesto por Cox, también conocido como “Regresión de Cox”.²⁶

Para explicarlo a detalle, primero se planteará el caso cuando se quiere comparar la función de riesgo para individuos de 2 distintos grupos. Después se podrá generalizar el modelo.

3.2.2 Modelos de Riesgos Proporcionales de dos grupos

Suponga que existe un estudio con pacientes que aleatoriamente recibirán un tratamiento estándar y un nuevo tratamiento. Sean $h_S(t)$ y $h_N(t)$ las funciones de riesgo respectivamente al tiempo t .

Entonces, el modelo proporcional de ambas funciones se expresa como:

$$h_N(t) = \varphi h_S(t); \text{ donde } \varphi \text{ es una constante}$$

Si $\varphi < 1$, entonces la función de riesgo al tiempo t es menor para el individuo del nuevo tratamiento.

Por otro lado, si $\varphi > 1$ la función de riesgo al tiempo t es mayor para el individuo del nuevo tratamiento.

Se debe mencionar que φ no podrá ser negativa, por lo cual es conveniente denotarla como $\varphi = \exp^{\beta} \Rightarrow \beta = \log(\varphi)$

Ahora, suponga que tenemos n individuos y que la función de riesgo para el i -ésimo individuo se denota por $h_i(t)$; $i = 1, 2, 3, \dots, n$, $h_0(t)$ es la función de riesgo para

²⁶ Cox D. R.; Regression models and lifetables (1972)

algún individuo del tratamiento estándar y $\varphi h_0(t)$ la función de riesgo para un individuo con riesgo proporcional al estándar.

Ahora, denotemos a la variable X como:

$$X = \begin{cases} 0, & \text{si el individuo pertenece al estudio estándar} \\ 1, & \text{si el individuo pertenece al nuevo estudio} \end{cases}$$

Si x_i es el valor que toma la variable X para el i -ésimo individuo ($i = 1, 2, 3, \dots, n$), la función de riesgo puede ser escrita como:

$$h_i(t) = \exp^{\beta x_i} h_0(t)$$

Se dice que es un modelo semiparamétrico, ya que se cuenta con la parte paramétrica ($\exp^{\beta x_i}$) y la parte no paramétrica con $h_0(t)$ que es una función arbitraria y no especificada.

3.2.3 Modelo de Riesgos Proporcionales general

Para generalizar el modelo anterior donde la función de riesgo en algún momento dependa de los valores x_1, x_2, \dots, x_p de p variables explicativas (X_1, X_2, \dots, X_p).

Los valores de las variables explicativas serán representados por el vector

$$x = (x_1, x_2, \dots, x_p)'$$

Ahora, sea $h_0(t)$ la función de riesgo para algún individuo en el caso estándar, para el cual los valores de las variables explicativas del vector x valen cero.

Entonces, la función de riesgo para el i -ésimo individuo ($i = 1, 2, 3, \dots, n$), se representa:

$$h_i(t) = \varphi(x_i) h_0(t)$$

Donde $\varphi(x_i)$ representa a la función de los valores del vector de variables explicativas para el i -ésimo individuo.

Como se mencionó en el modelo de dos grupos $\varphi(x_i)$ no puede ser negativa, por lo cual será conveniente escribirla como \exp^{η_i} , donde η_i es la combinación lineal de las p variables explicativas x_i .

$$\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$$

$$\eta_i = \sum_{j=1}^p \beta_j x_{ji}$$

En notación matricial:

$$\eta_i = \beta' x_i$$

Donde β es el vector de parámetros que representa el coeficiente que mide el impacto de las variables explicativas.

Entonces, tenemos al modelo general como:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) h_0(t)$$

$$\frac{h_i(t)}{h_0(t)} = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi})$$

$$\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$$

Se puede notar que no existe un término constante β_0 en la componente lineal, ya que, si lo incluyera, la función podría $h_0(t)$ ser dividida entre $\exp(\beta_0)$, generando que la constante quede eliminada.

3.2.4 Ajuste de la función de supervivencia en el modelo de riesgos proporcionales

Tomando el ejemplo antes mencionado del estudio con pacientes que aleatoriamente recibirán un tratamiento estándar y un nuevo tratamiento, se tiene el modelo de riesgos proporcionales²⁷ de ambas funciones como:

$$h_N(t) = \varphi h_S(t);$$

donde φ es una constante que no depende del tiempo t

Integrando y multiplicando por -1 ambos lados, se tiene:

$$-\int_0^t h_N(u)du = -\int_0^t \varphi h_S(u)du$$

Exponenciando,

$$\exp\left[-\int_0^t h_N(u)du\right] = \exp\left[-\int_0^t \varphi h_S(u)du\right]$$

Dada la relación anteriormente explicada de:

$$S(t) = \exp\left[-\int_0^t h(u)du\right]$$

Se tiene:

$$S_N(t) = [S_S(t)]^\varphi$$

Por lo tanto, el modelo de riesgos proporcionales puede expresarse en términos de la función de supervivencia como:

$$S_{(t)} = [S_0(t)]^{\exp(X_i\beta)}$$

²⁷ Kaplan E. L., Meier P.; Nonparametric estimation from incomplete observations (1958)

3.2.5 Inclusión de variables numéricas en el modelo de riesgos proporcionales

Existen estudios en los cuales las variables definidas al inicio sean del tipo numéricas, por ejemplo, edad, presión sanguínea y estatura.

La definición antes vista, muestra que estas variables son introducidas al modelo de riesgos proporcionales y que cada una de ellas tiene su correspondiente coeficiente β .

Se tendrán casos en los cuales el investigador decidirá tratar las variables numéricas como binarias, por ejemplo, la variable que indica la presión sanguínea de algún paciente y se decide clasificarlas como categóricas en: hipertensión e hipotenso.

Las variables categóricas podrán ser introducidas al modelo al igual que las numéricas con un tratamiento distinto.

3.2.6 Inclusión de variables categóricas en el modelo de riesgos proporcionales

En caso de tener variables categóricas²⁸, por ejemplo, sexo, escuelas y niveles de glucosa, es conveniente incluir un término que cuantifique el efecto que toma cada valor.

Para las variables que pueden tomar únicamente dos valores posibles, como sexo (hombre o mujer) o si el individuo fuma o no fuma, la solución obvia es definir las arbitrariamente con variables ficticias con valor uno o cero.

Tomando como ejemplo el sexo de cada individuo en estudio, la variable ficticia $D_1 = 1$ si es hombre, $D_1 = 0$ si es mujer. La función de riesgo para los hombres quedaría como:

²⁸ Kaplan E. L., Meier P.; Nonparametric estimation from incomplete observations (1958)

$$h(t) = \exp^{\beta_1(1)} h_0(t)$$

Para las mujeres:

$$h(t) = \exp^{\beta_0(0)} h_0(t) = \exp^{(0)} h_0(t) = h_0(t)$$

Las variables ficticias creadas, indicarán en cada caso a que grupo o factor pertenece.

Ahora, suponga que se tienen k categorías, entonces se crearán las variables ficticias D_1, D_2, \dots, D_k que se definen como:

$$D_j = \begin{cases} 1, & \text{si el individuo pertenece a la } j\text{-ésima categoría} \\ 0, & \text{si el individuo no pertenece a la } j\text{-ésima categoría} \end{cases}$$

Una vez conocidos los valores de las $(k-1)$ variables, entonces se sabrá el valor de la categoría k .

Los coeficientes $\beta_1, \beta_2, \dots, \beta_{k-1}$ medirán el impacto de las variables categóricas explicativas en el tiempo de supervivencia, teniendo el siguiente modelo para el i -ésimo individuo:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_{k-1} x_{k-1(i)}) h_0(t)$$

3.2.7 Estimación del modelo de riesgos proporcionales

Una vez definido la expresión del modelo de riesgos proporcionales²⁹:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) h_0(t)$$

Los coeficientes $\beta_1, \beta_2, \dots, \beta_p$ en el modelo de riesgos proporcionales son parámetros desconocidos, se busca estimarlos sin la necesidad de suponer una forma particular

²⁹ Kaplan E. L., Meier P.; Nonparametric estimation from incomplete observations (1958)

para la función de riesgo $h_0(t)$, ya que ésta última no está especificada y es desconocida.

Supondremos que existen n individuos disponibles con los tiempos de supervivencia $t_1, t_2, t_3, \dots, t_n$ y que no existen empates en estos tiempos, es decir, solamente un individuo muere en cada tiempo de supervivencia. Los tiempos de supervivencia ordenados de los n individuos quedan como: $t_{(1)} < t_{(2)} < t_{(3)} < \dots < t_{(n)}$, siendo $t_{(j)}$ el j -ésimo tiempo de supervivencia o falla ordenado.

Se denotará $\mathbf{R}(t_{(j)})$ al conjunto de individuos no censurados que se encuentran en riesgo al tiempo $t_{(j)}$, es decir, estos individuos no han presentado el evento de interés al tiempo exactamente anterior a $t_{(j)}$. A este conjunto se le denomina conjunto en riesgo y se presentan de la siguiente manera:

$\mathbf{R}(t_{(0)})$: Todos los n individuos

$\mathbf{R}(t_{(1)})$: Individuos en riesgo antes de la primera falla ($t_{(1)}$)

⋮

$\mathbf{R}(t_{(n)})$: Individuos en riesgo antes de la n – ésima falla ($t_{(n)}$)

Ahora, la probabilidad de que el i -ésimo individuo presente el evento de interés en $t_{(j)}$ está dada por la probabilidad condicional de que el individuo falle en el instante $t_{(j)}$ dado que pertenece al conjunto en riesgo $\mathbf{R}(t_{(j)})$ y se puede expresar como:

$$\begin{aligned} & \mathbb{P}(\text{individuo } i \text{ presente el evento al tiempo } t_j / \text{algún individuo presente el evento al tiempo } t_{(j)}) \\ &= \frac{\mathbb{P}(\text{el individuo } i \text{ presente el evento al tiempo } t_{(j)})}{\mathbb{P}(\text{algún individuo de } \mathbf{R}(t_{(j)}) \text{ presente el evento al tiempo } t_{(j)})} \\ &= \frac{\mathbb{P}(\text{el individuo } i \text{ presente el evento al tiempo } t_{(j)})}{\sum_{k \in \mathbf{R}(t_{(j)})} \mathbb{P}(\text{algún individuo de } \mathbf{R}(t_{(j)}) \text{ presente el evento al tiempo } t_{(j)})} \end{aligned}$$

Reemplazando el intervalo de tiempo por $(t_j, t_{j+\delta_t})$, entonces:

$$= \lim_{\delta_t \rightarrow 0} \left\{ \frac{\mathbb{P}(i \in \mathbf{R}(t_{(j)}) \text{ presente el evento en } (t_j, t_{j+\delta_t}))}{\sum_{k \in \mathbf{R}(t_{(j)})} \mathbb{P}(k \in \mathbf{R}(t_{(j)}) \text{ presente el evento en } (t_j, t_{j+\delta_t}))} \right\}$$

Dividiendo entre δ_t , se tiene que:

$$= \lim_{\delta_t \rightarrow 0} \frac{\frac{\mathbb{P}(i \in \mathbf{R}(t_{(j)}) \text{ presente el evento en } (t_j, t_{j+\delta_t}))}{\delta_t}}{\frac{\sum_{k \in \mathbf{R}(t_{(j)})} \mathbb{P}(k \in \mathbf{R}(t_{(j)}) \text{ presente el evento en } (t_j, t_{j+\delta_t}))}{\delta_t}}$$

Obteniendo así:

$$= \frac{h_i(t_{(j)})}{\sum_{k \in \mathbf{R}(t_{(j)})} h_k(t_{(j)})}$$

Finalmente, si se toma el producto sobre los tiempos de supervivencia y no los tiempos censurados, se obtiene una función de verosimilitud parcial:

$$\mathcal{L}(\beta) = \prod_{j=1}^n \frac{\exp(\beta' x_j)}{\sum_{k \in \mathbf{R}(t_{(j)})} \exp(\beta' x_k)}$$

Una vez definida la función de verosimilitud parcial que no depende de la función de riesgo $h_0(t)$, se busca estimar los coeficientes β y como primer paso se obtiene la función log de la verosimilitud parcial:

$$\ln(\mathcal{L}(\beta)) = \sum_{j=1}^n \left[\beta' x_j - \ln \left(\sum_{k \in \mathbf{R}(t_{(j)})} \exp(\beta' x_k) \right) \right]$$

Ahora, para hallar la función score o puntaje, derivamos respecto a β , entonces:

$$\begin{aligned}\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta_r} &= \sum_{j=1}^d \left[\frac{\partial \beta' x_j}{\partial \beta_r} - \frac{\partial \ln \left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) \right)}{\partial \beta_r} \right] \\ &= \sum_{j=1}^d \left[x_{rj} - \sum_{k \in R(t_{(j)})} x_{rj} \exp(\beta' x_k) \right]\end{aligned}$$

Para la obtención del estimador máximo verosimil parcial, se deberá resolver:

$$\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta_r} = 0$$

Ahora, se buscan las derivadas parciales de segundo orden y éstas serán utilizadas más adelante, entonces:

$$\begin{aligned}& - \frac{\partial^2 \ln(\mathcal{L}(\beta))}{\partial \beta_r \partial \beta_s} \\ &= \sum_{j=1}^d \left[\frac{\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) x_{jr} x_{js} \left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) \right) - \left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) x_{jr} \right) \left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) x_{js} \right)}{\left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) \right)^2} \right]\end{aligned}$$

Finalmente, tenemos:

$$\begin{aligned}- \frac{\partial^2 \ln(\mathcal{L}(\beta))}{\partial \beta_r \partial \beta_s} &= \sum_{j=1}^d \left[\frac{\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) x_{jr} x_{js}}{\sum_{k \in R(t_{(j)})} \exp(\beta' x_k)} \right. \\ & \left. - \frac{\left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) x_{jr} \right) \left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) x_{js} \right)}{\left(\sum_{k \in R(t_{(j)})} \exp(\beta' x_k) \right)^2} \right]\end{aligned}$$

Ahora, como se mencionó anteriormente para la solución de los estimadores de β , se debe hallar la solución de $\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta_r} = 0$. Mediante el método de Newton Raphson,

el cual es utilizado para resolver la ecuación $f(x) = 0$ mediante la expansión en series de Taylor de la función $f(x)$ alrededor del punto x_1 , se tiene la ecuación:

$$f(x) = f(x_1) + (x - x_1) f'(x_1) + \frac{1}{2} (x - x_1)^2 f''(x_1) + \dots$$

Tomando los dos primeros términos de la expansión, se tiene la aproximación:

$$f(x) \approx f(x_1) + (x - x_1) f'(x_1)$$

Igualando a cero:

$$f(x_1) + (x - x_1) f'(x_1) = 0$$

Para x , se tendría:

$$x = x_1 - \frac{f(x_1)}{f'(x_1)}$$

Generalizando, se obtiene el método iterativo de Newton:

$$x_i = x_{i-1} - \frac{f(x_{i-1})}{f'(x_{i-1})} \text{ para } i \in \mathbb{N}$$

Dado lo anterior, para la función de los estimadores de β , se tiene:

$$\hat{\beta}_{(r)} = \hat{\beta}_{(r-1)} - \left[\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta_r} \right] \left[-\frac{\partial^2 \ln(\mathcal{L}(\beta))}{\partial \beta_r \partial \beta_s} \right]^{-1}$$

Es de interés saber si las covariables contribuyen información a la función de riesgo, es decir, si el vector de coeficientes β es distinto de cero.

Para el contraste de $H_0: \beta = \beta_0$ vs $H_a: \beta \neq \beta_0$, se podrá utilizar la prueba de Wald, basado en que $\hat{\beta}$ presenta asintóticamente una distribución normal y la matriz de varianzas y covarianzas estimadas es $I^{-1}(\hat{\beta})$.

El estadístico de Wald está dado por:

$$X_N = (\hat{\beta} - \beta_0)' I^{-1}(\hat{\beta}) (\hat{\beta} - \beta_0)$$

Bajo H_0 , X_N presenta una distribución *Ji-cuadrada* con p grados de libertad.

La siguiente prueba se hace bajo el contraste de la razón de verosimilitudes, tomando en cuenta el valor de la función de verosimilitud parcial evaluada en $\hat{\beta}$, es decir, $\mathcal{L}(\hat{\beta})$ y en β_0

$$X_{LR} = 2 (\ln(\hat{\beta}) - \ln(\beta_0))$$

Bajo H_0 , X_{LR} sigue una distribución *Ji-cuadrada* con p grados de libertad.

Por último, bajo el contraste score o de puntaje basado en las derivadas del logaritmo de verosimilitud parcial evaluada en la hipótesis $H_0: \beta = \beta_0$, el estadístico está dado por:

$$X_s = \left(\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta_r} \right)' \left(-\frac{\partial^2 \ln(\mathcal{L}(\beta))}{\partial \beta_r \partial \beta_s} \right)^{-1} \left(\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta_r} \right)$$

Donde el estadístico X_s también presenta una distribución *Ji-cuadrada* con p grados de libertad, bajo H_0 .

Capítulo 4. Residuos

Los residuos son la diferencia entre el valor observado y el estimado por el modelo, calculado para cada individuo del estudio.

Se analizarán los cinco tipos de residuos de interés: los residuos de Cox-Snell, los de martingala y los de *deviance*.

4.1 Residuales de Cox-Snell

Desarrollados por Cox & Snell³⁰. Los residuos para n individuos se calculan como:

$$r_i = \exp(\beta' X_i) \hat{H}_0(t_i) \text{ con } i = 1, \dots, n$$

Donde $\hat{H}_0(t)$ es el estimador de la función de riesgo acumulado, conocida como estimador de Breslow:

$$\hat{H}_0(t) = \int_0^t \frac{\sum_{i=1}^n d N_i(s)}{\sum_{i=1}^n Y_i(s) \exp(\hat{\beta}' X_i(s))}$$

Basados en el proceso de conteo $N_i \equiv \{N_i(t), t \geq 0\}$ para el i -ésimo individuo es el número de eventos observados al tiempo t .

La variable $Y_i(t)$ indica si el i -ésimo individuo está en riesgo en el tiempo t :

$$Y_i(t) = \begin{cases} 1, & \text{si el individuo ha experimentado el evento al tiempo } t \\ 0, & \text{si el individuo no ha presentado el evento al tiempo } t \end{cases}$$

Como forma sencilla, se debe graficar los residuales Cox-Snell, ya que su principio radica en que tengan una distribución exponencial con media igual a 1.

4.2 Residuales de martingala

Los residuales de martingala tienen una ligera modificación a los de Cox-Snell previamente mostrados.

Estos residuales están basados en la diferencia entre el proceso de conteo y la integral de la función de intensidad, como se muestra a continuación:

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \exp(\beta' X_i(s)) h_0(s) ds \text{ con } i = 1, \dots, n$$

³⁰ Cox D. R., Snell E. J.; A general definition of residuals (1968)

Teniendo el estimador de máximo verosímil $\hat{\beta}$ y el estimador de Breslow (\hat{H}_0), entonces los residuales de martingala se podrán estimar como:

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) \exp(\hat{\beta}' X_i(s)) d\hat{H}_0(s)$$

4.3 Residuales de devianza

Utilizados para obtener residuales que tengan una distribución normal de los residuales de martingala.

Son definidos como:

$$d_i = \text{signo}(\hat{M}_i) \{-2 [\hat{M}_i + N_i \log(N_i - \hat{M}_i)]\}^{1/2}$$

Los residuales tienen un valor de 0 cuando \hat{M}_i es cero.

Estos residuales son útiles para la detección de valores atípicos.

Capítulo 5. Análisis Descriptivo

Como se mencionó anteriormente, el estudio contiene información de 1,256 alumnos y el evento de interés es el tiempo en el cual el alumno reprueba la primera materia. El tiempo de falla se observará de acuerdo con el orden en los cuales fueron realizados los exámenes departamentales de los primeros dos años de la licenciatura. Se denominará *periodo* al intervalo entre cada uno de los exámenes. Por ejemplo, al intervalo entre el inicio del estudio (inicio de clases) y el primer examen departamental, será el periodo 1.

A continuación, se muestra una tabla de cada uno de los periodos:

Examen	Periodo
Embriología Humana (1er Examen)	1
Bioquímica y Biología Molecular (1er Examen)	2
Anatomía (1er Examen)	3
Biología Celular e Histología Médica (1er Examen)	4
Embriología Humana (2do Examen)	5
Bioquímica y Biología Molecular (2do Examen)	6
Anatomía (2do Examen)	7
Biología Celular e Histología Médica (2do Examen)	8
Embriología Humana (3er Examen)	9
Bioquímica y Biología Molecular (3er Examen)	10
Anatomía (3er Examen)	11
Biología Celular e Histología Médica (3er Examen)	12
Inmunología (1er Examen)	13
Microbiología y Parasitología (1er Examen)	14
Farmacología (1er Examen)	15
Fisiología (1er Examen)	16
Microbiología y Parasitología (2do Examen)	17
Inmunología (2do Examen)	18
Microbiología y Parasitología (3er Examen)	19
Farmacología (2do Examen)	20
Fisiología (2do Examen)	21
Inmunología (3er Examen)	22
Fisiología (3er Examen)	23
Farmacología (3er Examen)	24

El tiempo de duración del estudio fue de 24 periodos, por lo que los alumnos que no reprobaron alguna materia al momento 24, presentaron un tiempo de censura por la derecha al periodo 24.

Otro tipo de datos censurados observados en el estudio, fueron los alumnos que por motivos ajenos al estudio no se presentaron alguno de los exámenes, presentando censura aleatoria y por lo cual el evento de interés no fue observado en estos alumnos.

En total se presentaron 186 (15%) casos de censura por lo que fue favorable. A continuación, se muestra el comportamiento de nuestros datos totales.

N° Total de alumnos	N° de alumnos que reprobaron alguna materia	N° censurados	Porcentaje censurados
1,256	1,070	186	15%

Tabla 1 Total de alumnos, alumnos que reprobaron alguna materia y datos censurados

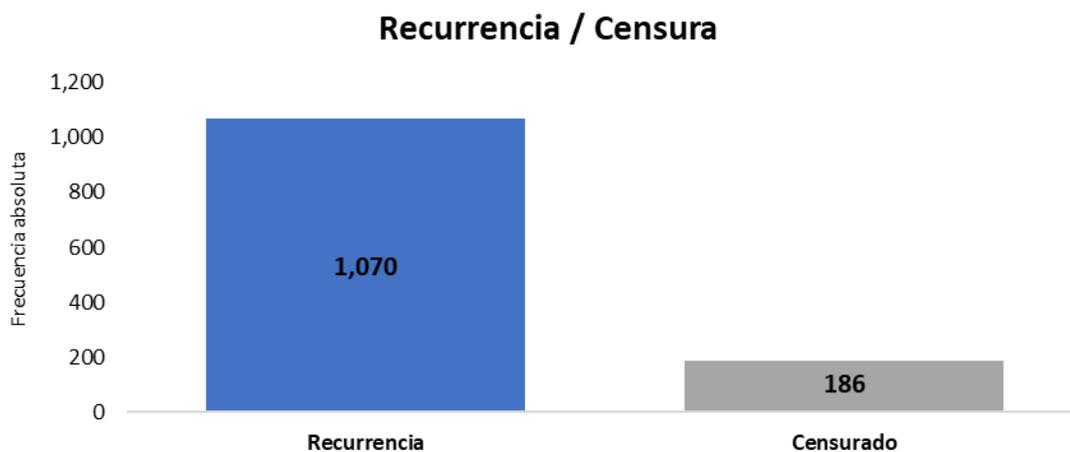


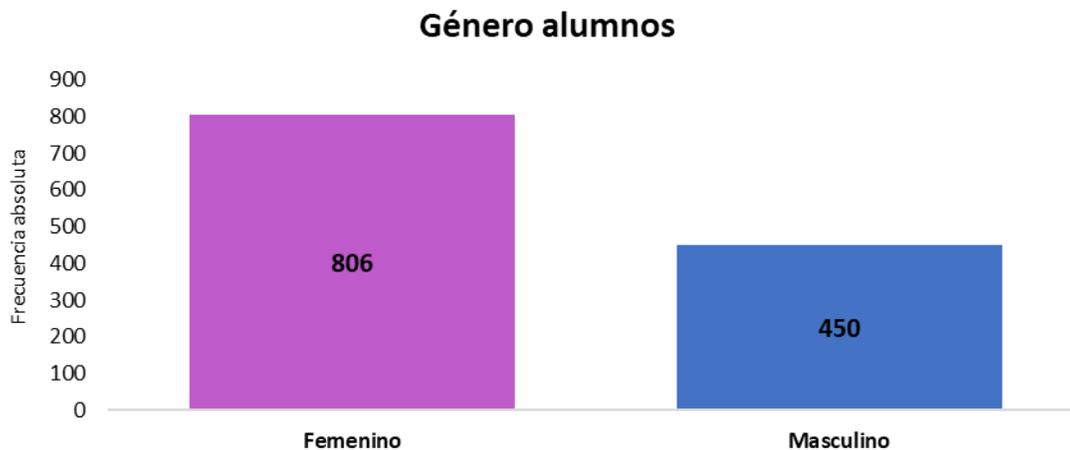
Figura 7 Total de alumnos que reprobaron alguna materia y datos censurados

Los alumnos tienen variables que ayudan al estudio:

- Sexo del estudiante.

Femenino	Masculino
806	450

Tabla 2 Total de alumnos por sexo

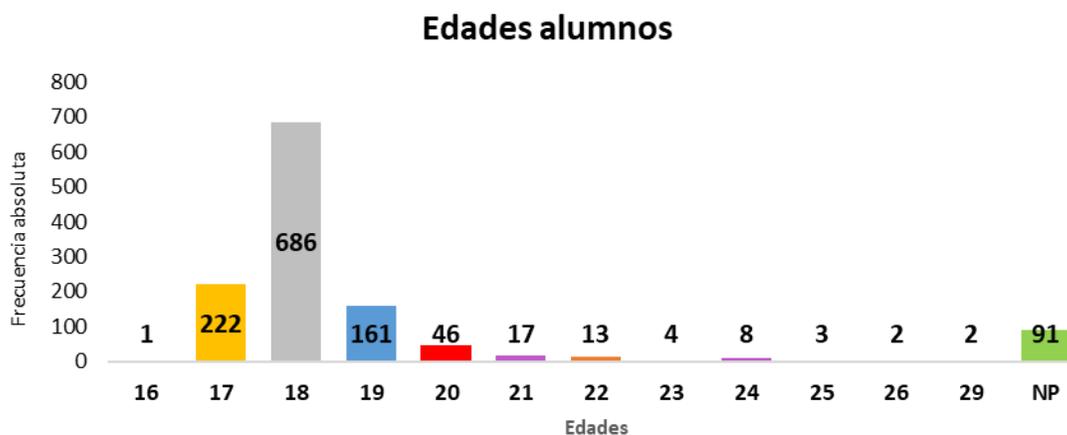


Cuadro 1. Total de censuras y eventos observados por sexo

Sexo	#	# Eventos	Censurados	% Censura
Femenino	806	706	100	12%
Masculino	450	364	86	19%
Total	1,256	1,070	186	15%

Del total de alumnos analizados, el 64% son mujeres y 36% hombres. En las mujeres el 12% son datos censurados, es decir, el 88% de las mujeres presentaron el evento. En hombres el 81% reprobaron alguna materia.

- Edad estudiante.



Edades	16	17	18	19	20	21	22	23	24	25	26	29	NP:Sin datos
#	1	222	686	161	46	17	13	4	8	3	2	2	91

Tabla 3 Total de alumnos por edades

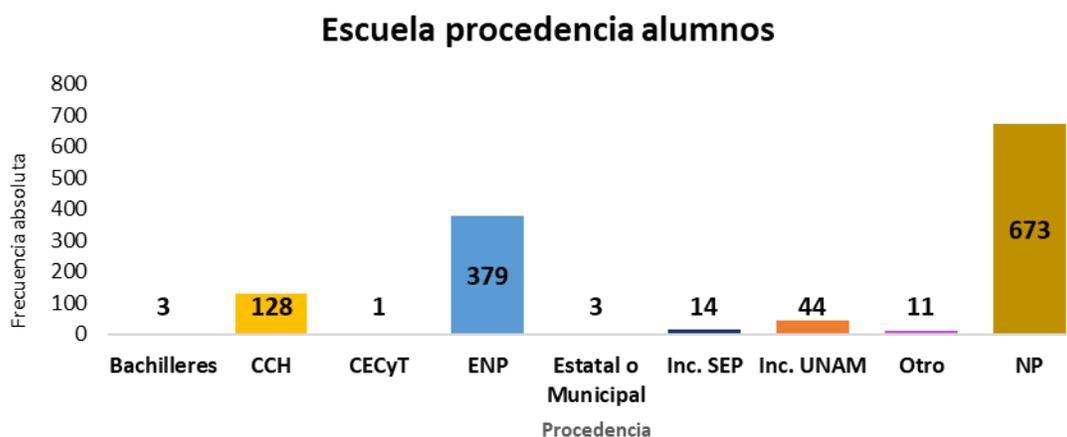
Cuadro 2. Total de censuras y eventos observados por edades

Edad	#	# Eventos	Censurados	% Censura
16	1	1	0	0%
17	222	191	31	14%
18	686	570	116	17%
19	161	135	26	16%

20	46	43	3	7%
21	17	17	0	0%
22	13	11	2	15%
23	4	4	0	0%
24	8	6	2	25%
25	3	3	0	0%
26	2	2	0	0%
29	2	1	1	50%
NP:Sin datos	91	86	5	5%
Total	1,256	1,070	186	15%

Para el 93% de los alumnos analizados se tuvo información de la edad con la que ingresaron a la carrera, el 85% de esos alumnos presentaron el evento. La población se encuentra concentrada con 72% en edades de 17 y 18 años.

- Escuela de Procedencia



Procedencia	Bachilleres	CCH	CECyT	ENP	Estatal o Municipal	Inc. SEP	Inc. UNAM	Otro	NP:Sin datos
#	3	128	1	379	3	14	44	11	673

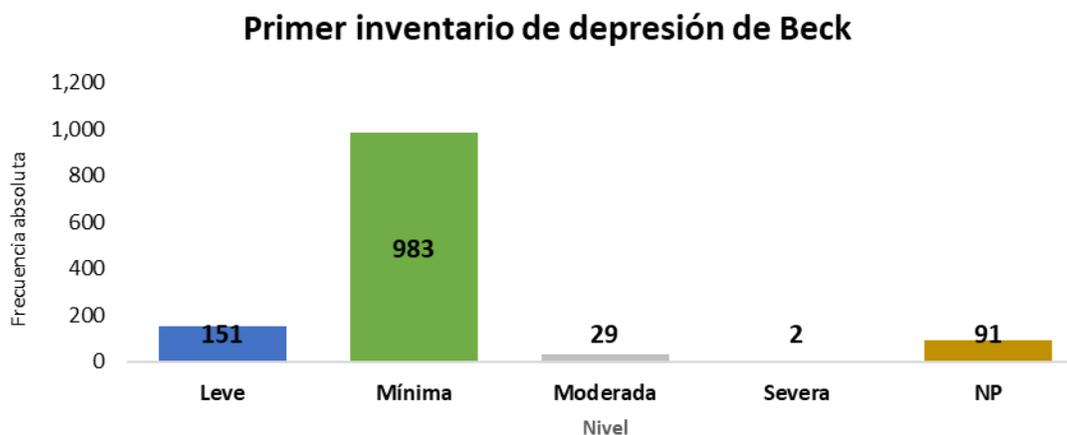
Tabla 4 Total de alumnos por escuela de procedencia

Cuadro 3. Total de censuras y eventos observados por escuela de procedencia.

Procedencia	#	# Eventos	Censurados	% Censura
Bachilleres	3	1	2	67%
CCH	128	105	23	18%
CECyT	1	0	1	100%
ENP	379	270	109	29%
Estatad o Municipal	3	2	1	33%
Inc. SEP	14	9	5	36%
Inc. UNAM	44	28	16	36%
Otro	11	8	3	27%
NP:Sin datos	673	647	26	4%
Total	1,256	1,070	186	15%

Por escuela de procedencia no se cuenta con el 54% de datos, la mayor concentración de procedencia son alumnos de ENP (Escuela Nacional Preparatoria) con el 30%.

- Resultados Inventario de Depresión de Beck



Casos	Leve	Mínima	Moderada	Severa	NP:Sin datos
#	151	983	29	2	91

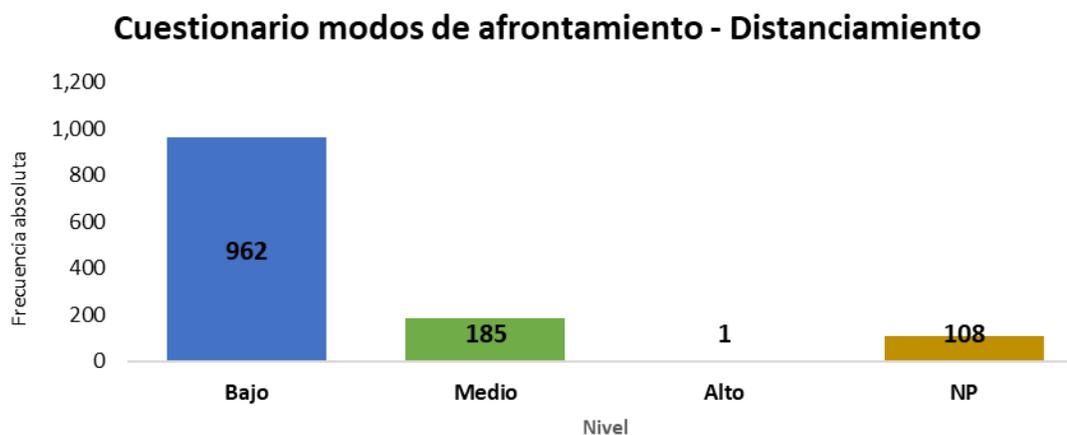
Tabla 5 Resultados del inventario de depresión de Beck

Cuadro 4. Total de censuras y eventos observados por resultados del inventario de depresión de Beck

	#	# Eventos	Censurados	% Censura
Leve	151	129	22	15%
Mínima	983	843	140	14%
Moderada	29	25	4	14%
Severa	2	2	0	0%
NP:Sin datos	91	71	20	22%
Total	1,256	1,070	186	15%

En los resultados de inventario de depresión de Beck, el 7% no se tuvo información, el 78% de la población tuvieron de resultado “mínima”, el 12% “leve”, 2% “moderada” y 0.2% “severa”.

- Cuestionario Modos de Afrontamiento-Distanciamiento



Casos	Bajo	Medio	Alto	NP:Sin datos
#	962	185	1	108

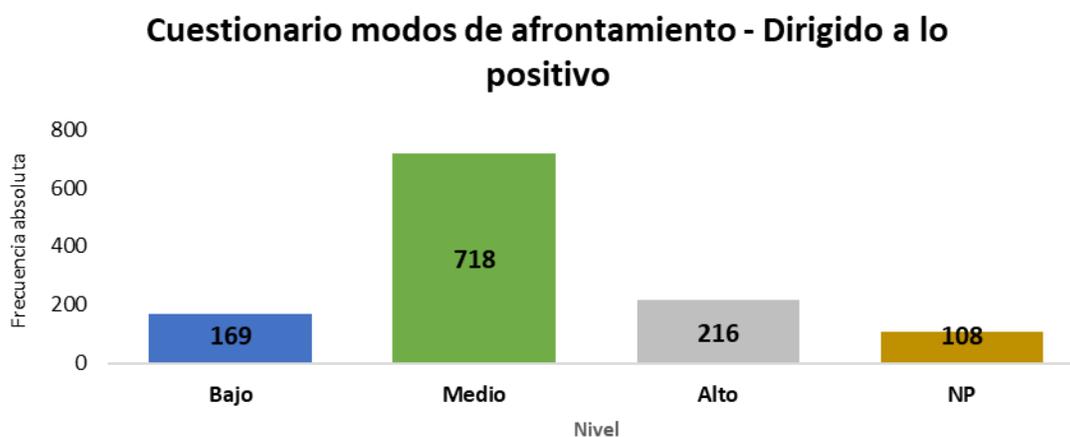
Tabla 6 Resultados del cuestionario modos de afrontamiento-Distanciamiento

Cuadro 5. Total de censuras y eventos observados por resultados del cuestionario modos de afrontamiento-Distanciamiento

	#	# Eventos	Censurados	% Censura
Bajo	962	822	140	15%
Medio	185	161	24	13%
Alto	1	1	0	0%
NP:Sin datos	108	86	22	20%
Total	1,256	1,070	186	15%

Del cuestionario de modos de afrontamiento- distanciamiento, el 9% no se tuvo información, el 77% de la población tuvieron de resultado “bajo”, el 15% “medio” y 0.1% “alto”.

- Cuestionario Modos de Afrontamiento-Dirigido a lo Positivo



Casos	Bajo	Medio	Alto	NP:Sin datos
#	169	718	216	108

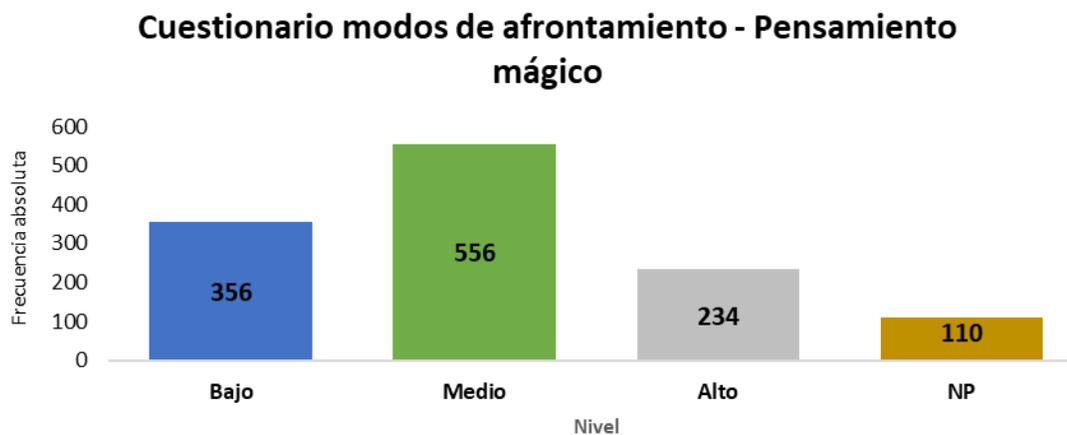
Tabla 7 Resultados del cuestionario modos de afrontamiento- Dirigido a lo positivo

Cuadro 6. Total de censuras y eventos observados por resultados del cuestionario modos de afrontamiento-Dirigido a lo positivo

	#	# Eventos	Censurados	% Censura
Bajo	169	136	33	20%
Medio	718	615	103	14%
Alto	261	233	28	11%
NP:Sin datos	108	86	22	20%
Total	1,256	1,070	186	15%

Del cuestionario de modos de afrontamiento- dirigido a lo positivo, el 9% no se tuvo información, el 13% de la población tuvieron de resultado “bajo”, el 57% “medio” y 21% “alto”.

- Cuestionario Modos de Afrontamiento-Pensamiento Mágico



Casos	Bajo	Medio	Alto	NP:Sin datos
#	356	556	234	110

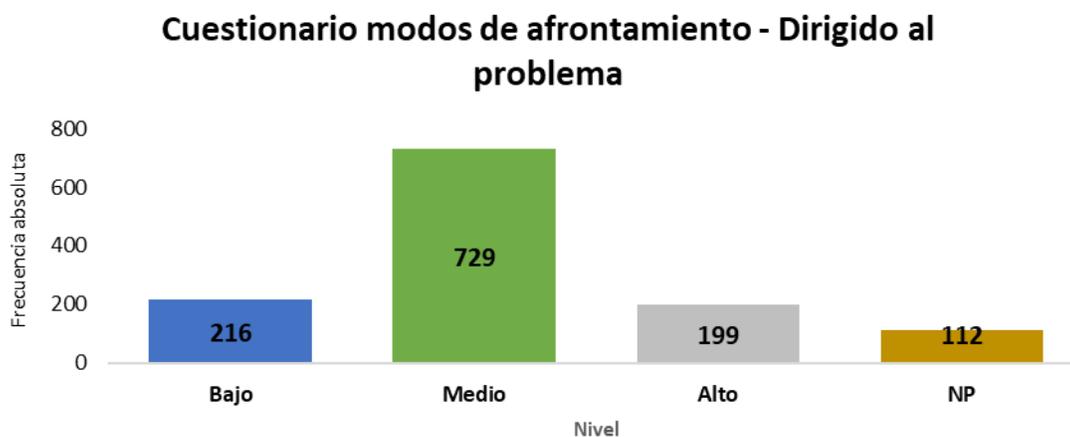
Tabla 8 Resultados del cuestionario modos de afrontamiento- Pensamiento mágico

Cuadro 7. Total de censuras y eventos observados por resultados del cuestionario modos de afrontamiento-Pensamiento mágico

	#	# Eventos	Censurados	% Censura
Bajo	356	309	47	13%
Medio	556	473	83	15%
Alto	234	202	32	14%
NP:Sin datos	110	86	24	22%
Total	1,256	1,070	186	15%

Del cuestionario de modos de afrontamiento- pensamiento mágico, el 9% no se tuvo información, el 28% de la población tuvieron de resultado “bajo”, el 44% “medio” y 19% “alto”.

- Cuestionario Modos de Afrontamiento-Dirigido al Problema



Casos	Bajo	Medio	Alto	NP:Sin datos
#	216	729	199	112

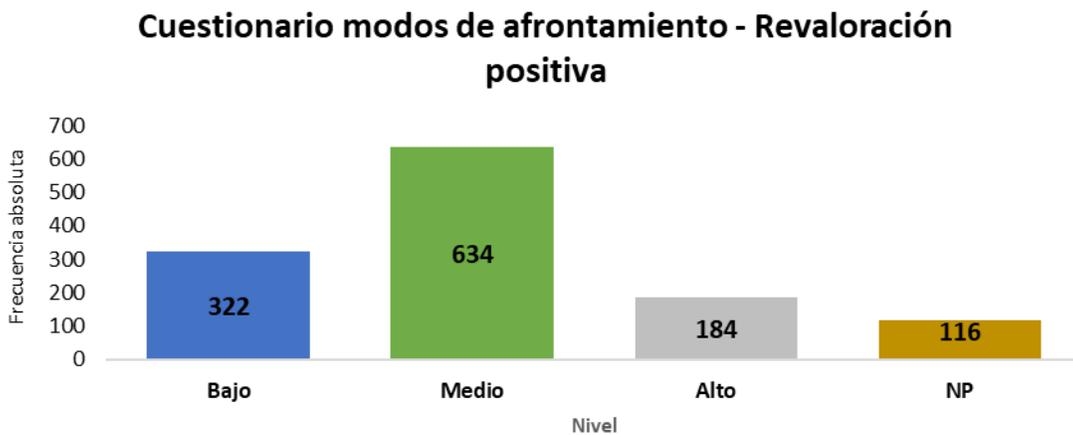
Tabla 9 Resultados del cuestionario modos de afrontamiento- Dirigido al problema

Cuadro 8. Total de censuras y eventos observados por resultados del cuestionario modos de afrontamiento-Dirigido al problema

	#	# Eventos	Censurados	% Censura
Bajo	216	188	28	14%
Medio	729	622	107	15%
Alto	199	171	28	14%
NP:Sin datos	112	89	23	21%
Total	1,256	1,070	186	15%

Del cuestionario de modos de afrontamiento- dirigido al problema, el 9% no se tuvo información, el 17% de la población tuvieron de resultado “bajo”, el 58% “medio” y 16% “alto”.

- Cuestionario Modos de Afrontamiento-Revaloración Positiva



Casos	Bajo	Medio	Alto	NP:Sin datos
#	322	634	184	116

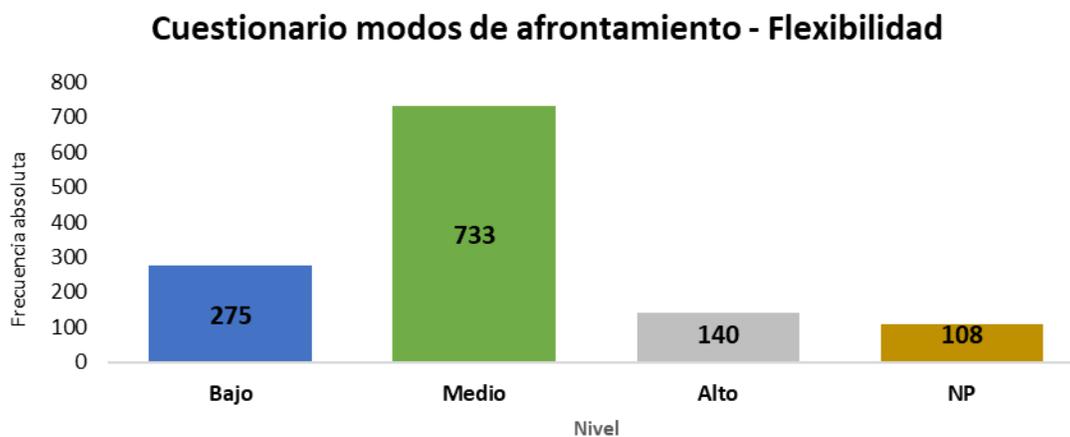
Tabla 10 Resultados del cuestionario modos de afrontamiento-Revaloración positiva

Cuadro 9. Total de censuras y eventos observados por resultados del cuestionario modos de afrontamiento-Revaloración positiva

	#	# Eventos	Censurados	% Censura
Bajo	322	278	44	14%
Medio	634	531	103	16%
Alto	184	168	16	9%
NP:Sin datos	116	93	23	20%
Total	1,256	1,070	186	15%

Del cuestionario de modos de afrontamiento- revaloración positiva, el 9% no se tuvo información, el 26% de la población tuvieron de resultado “bajo”, el 50% “medio” y 15% “alto”.

- Cuestionario Modos de Afrontamiento-Flexibilidad



Casos	Bajo	Medio	Alto	NP:Sin datos
#	275	733	140	108

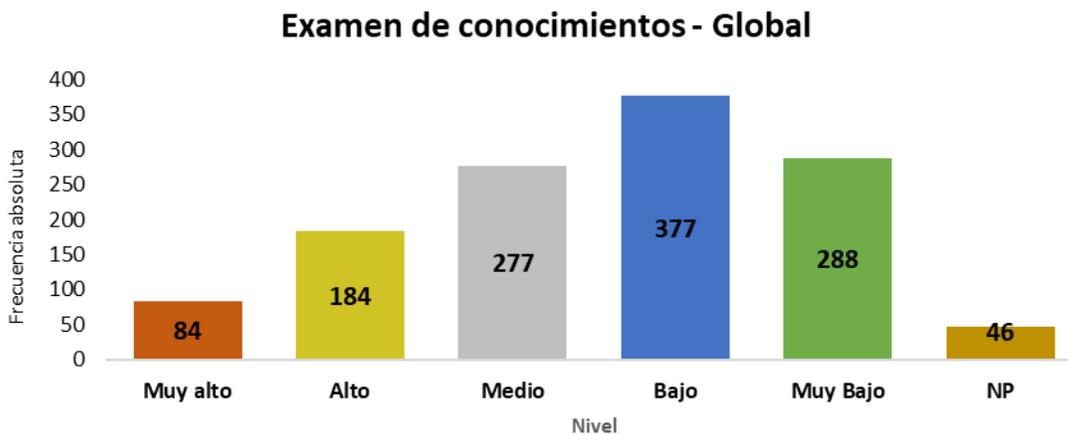
Tabla 11 Resultados del cuestionario modos de afrontamiento- Flexibilidad

Cuadro 10. Total de censuras y eventos observados por resultados del cuestionario modos de afrontamiento-Flexibilidad

	#	# Eventos	Censurados	% Censura
Bajo	275	232	43	16%
Medio	733	633	100	14%
Alto	140	119	21	15%
NP:Sin datos	108	86	22	20%
Total	1,256	1,070	186	15%

Del cuestionario de modos de afrontamiento- flexibilidad, el 9% no se tuvo información, el 22% de la población tuvieron de resultado “bajo”, el 58% “medio” y 11% “alto”.

- Examen de Conocimientos-Global



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	84	184	277	377	288	46

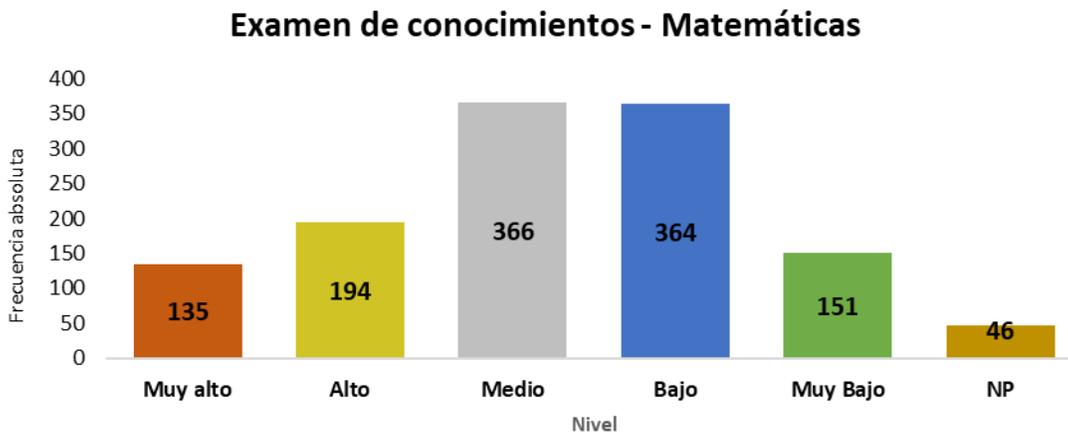
Tabla 12 Resultados del examen de conocimientos - Global

Cuadro 11. Total de censuras y eventos observados por resultados del examen de conocimientos- Global

	#	# Eventos	Censurados	% Censura
Muy alto	84	44	40	48%
Alto	184	132	52	28%
Medio	277	229	48	17%
Bajo	377	355	22	6%
Muy bajo	288	280	8	3%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- global, el 4% no se tuvo información, el 7% de la población tuvieron de resultado “muy alto”, el 15% “alto”, 22% tuvo resultado “medio”, 30% resultado “bajo” y 23% “muy bajo”.

- Examen de Conocimientos-Matemáticas



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	135	194	366	364	151	46

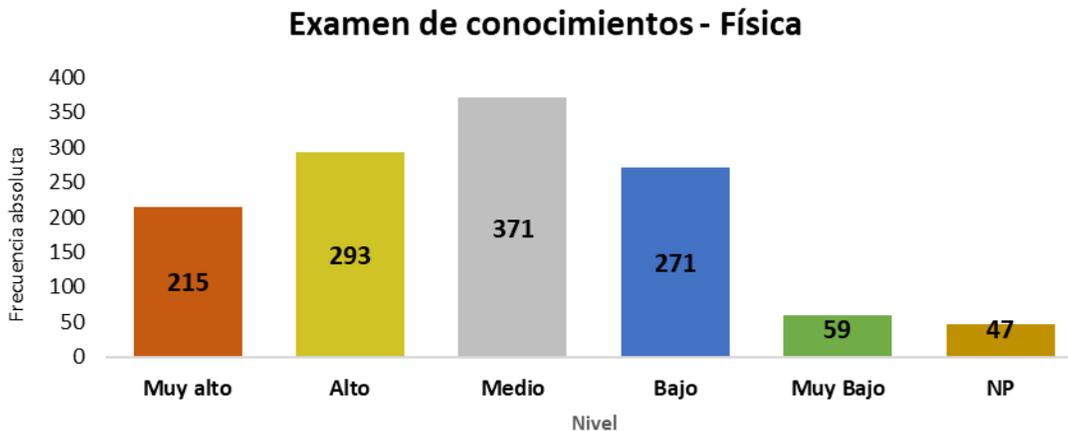
Tabla 13 Resultados del examen de conocimientos - Matemáticas

Cuadro 12. Total de censuras y eventos observados por resultados del examen de conocimientos- Matemáticas

	#	# Eventos	Censurados	% Censura
Muy alto	135	80	55	41%
Alto	194	151	43	22%
Medio	366	317	49	13%
Bajo	364	348	16	4%
Muy bajo	151	144	7	5%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- matemáticas, el 4% no se tuvo información, el 11% de la población tuvieron de resultado “muy alto”, el 15% “alto”, 29% tuvo resultado “medio”, 29% resultado “bajo” y 12% “muy bajo”.

- Examen de Conocimientos-Física



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	215	293	371	271	59	47

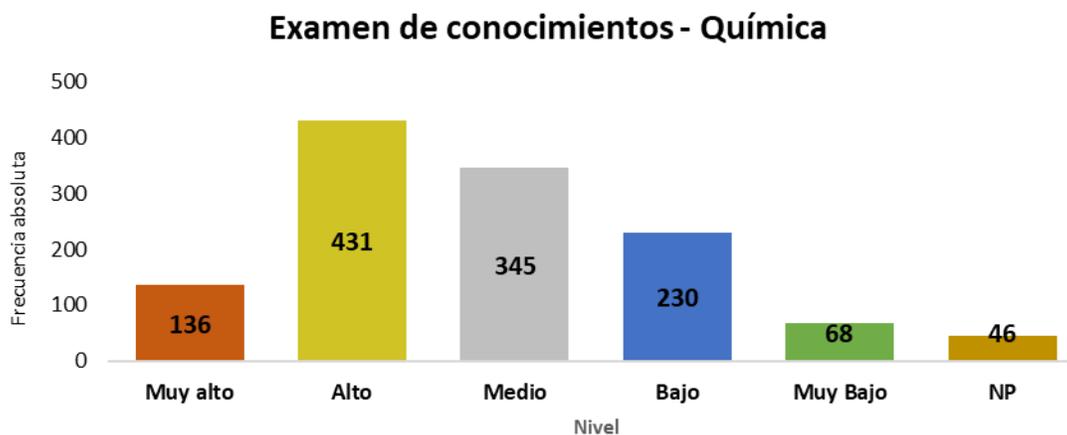
Tabla 14 Resultados del examen de conocimientos - Física

Cuadro 13. Total de censuras y eventos observados por resultados del examen de conocimientos- Física

	#	# Eventos	Censurados	% Censura
Muy alto	215	143	72	33%
Alto	293	249	44	15%
Medio	371	328	43	12%
Bajo	271	261	10	4%
Muy bajo	59	58	1	2%
NP:Sin datos	47	31	16	34%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- física, el 4% no se tuvo información, el 17% de la población tuvieron de resultado “muy alto”, el 23% “alto”, 30% tuvo resultado “medio”, 22% resultado “bajo” y 5% “muy bajo”.

- Examen de Conocimientos-Química



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	136	431	345	230	68	46

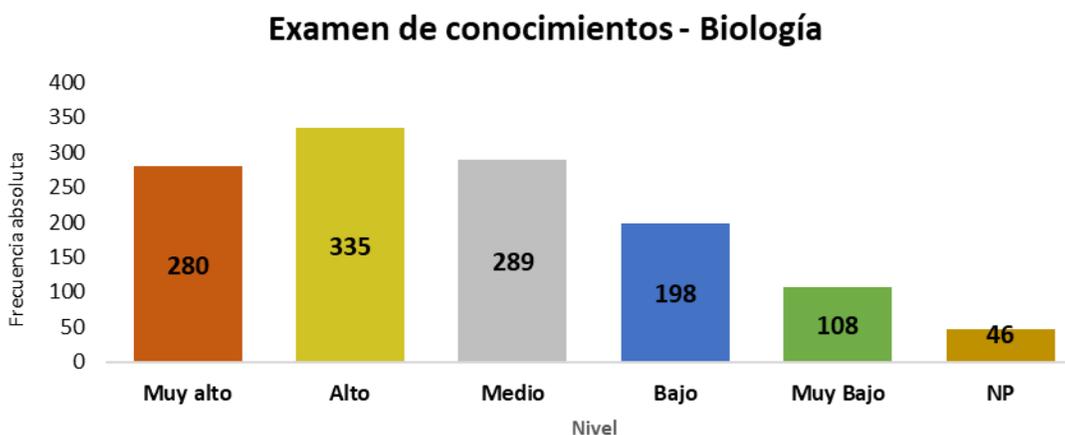
Tabla 15 Resultados del examen de conocimientos - Química

Cuadro 14. Total de censuras y eventos observados por resultados del examen de conocimientos- Química

	#	# Eventos	Censurados	% Censura
Muy alto	136	88	48	35%
Alto	431	352	79	18%
Medio	345	314	31	9%
Bajo	230	220	10	4%
Muy bajo	68	66	2	3%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- química, el 4% no se tuvo información, el 11% de la población tuvieron de resultado “muy alto”, el 34% “alto”, 27% tuvo resultado “medio”, 18% resultado “bajo” y 5% “muy bajo”.

- Examen de Conocimientos-Biología



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	280	335	289	198	108	46

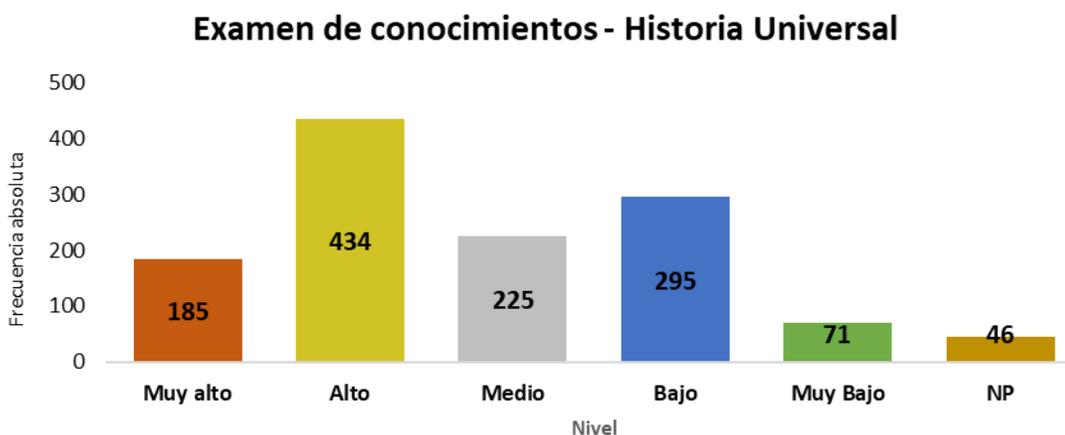
Tabla 16 Resultados del examen de conocimientos - Biología

Cuadro 15. Total de censuras y eventos observados por resultados del examen de conocimientos- Biología

	#	# Eventos	Censurados	% Censura
Muy alto	280	193	87	31%
Alto	335	291	44	13%
Medio	289	263	26	9%
Bajo	198	186	12	6%
Muy bajo	108	107	1	1%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- biología, el 4% no se tuvo información, el 22% de la población tuvieron de resultado “muy alto”, el 27% “alto”, 23% tuvo resultado “medio”, 16% resultado “bajo” y 9% “muy bajo”.

- Examen de Conocimientos-Historia Universal



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	185	434	225	295	71	46

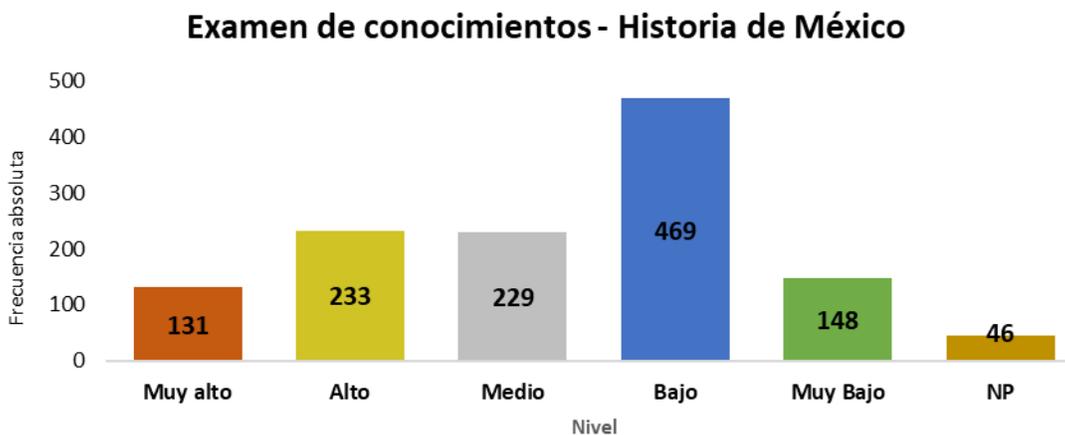
Tabla 17 Resultados del examen de conocimientos – Historia Universal

Cuadro 16. Total de censuras y eventos observados por resultados del examen de conocimientos- Historia Universal

	#	# Eventos	Censurados	% Censura
Muy alto	185	137	48	26%
Alto	434	356	78	18%
Medio	225	202	23	10%
Bajo	295	278	17	6%
Muy bajo	71	67	4	6%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- historia universal, el 4% no se tuvo información, el 15% de la población tuvieron de resultado “muy alto”, el 35% “alto”, 18% tuvo resultado “medio”, 23% resultado “bajo” y 6% “muy bajo”.

- Examen de Conocimientos-Historia de México



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	131	233	229	469	148	46

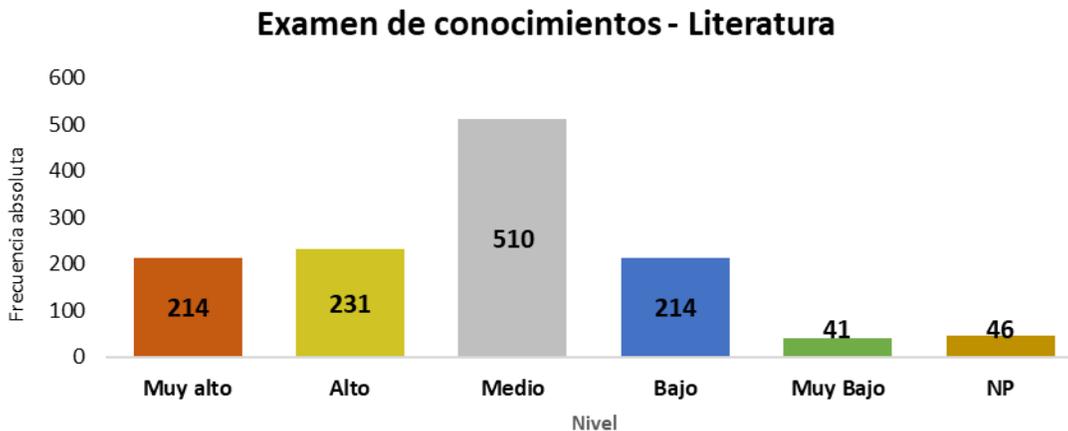
Tabla 18 Resultados del examen de conocimientos – Historia de México

Cuadro 17. Total de censuras y eventos observados por resultados del examen de conocimientos- Historia de México

	#	# Eventos	Censurados	% Censura
Muy alto	131	95	36	27%
Alto	233	185	48	21%
Medio	229	200	29	13%
Bajo	469	420	49	10%
Muy bajo	148	140	8	5%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- historia de México, el 4% no se tuvo información, el 10% de la población tuvieron de resultado “muy alto”, el 19% “alto”, 18% tuvo resultado “medio”, 37% resultado “bajo” y 12% “muy bajo”.

- Examen de Conocimientos-Literatura



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	214	231	510	214	41	46

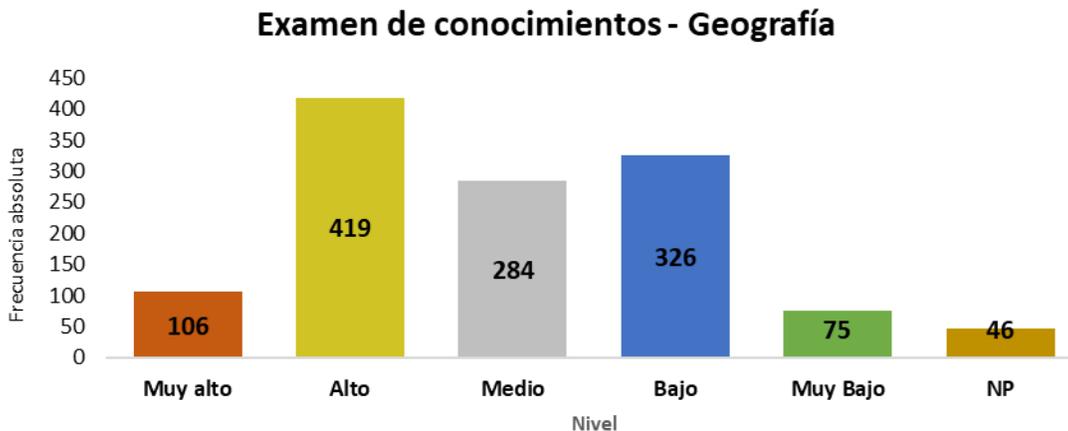
Tabla 19 Resultados del examen de conocimientos – Literatura

Cuadro 18. Total de censuras y eventos observados por resultados del examen de conocimientos- Literatura

	#	# Eventos	Censurados	% Censura
Muy alto	214	158	56	26%
Alto	231	189	42	18%
Medio	510	456	54	11%
Bajo	214	196	18	8%
Muy bajo	41	41	0	0%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- literatura, el 4% no se tuvo información, el 17% de la población tuvieron de resultado “muy alto”, el 18% “alto”, 41% tuvo resultado “medio”, 17% resultado “bajo” y 3% “muy bajo”.

- Examen de Conocimientos-Geografía



Casos	Muy alto	Alto	Medio	Bajo	Muy bajo	NP:Sin datos
#	106	419	284	326	75	46

Tabla 20 Resultados del examen de conocimientos – Geografía

Cuadro 19. Total de censuras y eventos observados por resultados del examen de conocimientos- Geografía

	#	# Eventos	Censurados	% Censura
Muy alto	106	74	32	30%
Alto	419	349	70	17%
Medio	284	243	41	14%
Bajo	326	303	23	7%
Muy bajo	75	71	4	5%
NP:Sin datos	46	30	16	35%
Total	1,256	1,070	186	15%

Para el examen de conocimientos- geografía, el 4% no se tuvo información, el 8% de la población tuvieron de resultado “muy alto”, el 33% “alto”, 23% tuvo resultado “medio”, 26% resultado “bajo” y 6% “muy bajo”.

Capítulo 6. Análisis curvas de Supervivencia

Primero, se estima la función de supervivencia mediante el estimador Kaplan Meier, anteriormente explicado.

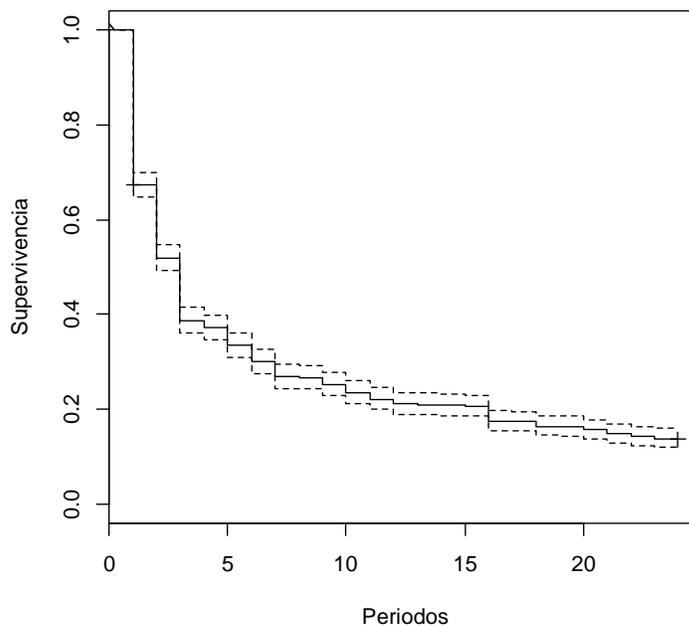
En la tabla 21 se observan la probabilidad de falla o supervivencia en cada uno de los periodos, en el cual el evento de interés es la primera materia reprobada de los alumnos.

time	n.risk	n.event	survival	std.err	lower	95% CI upper	95% CI
1	1256	411	0.673	0.01324		0.647	0.699
2	825	188	0.519	0.01418		0.492	0.548
3	637	163	0.387	0.01386		0.360	0.415
4	474	19	0.371	0.01375		0.345	0.399
5	455	45	0.334	0.01343		0.309	0.362
6	410	43	0.299	0.01305		0.275	0.326
7	367	39	0.267	0.01261		0.244	0.293
8	328	1	0.267	0.01260		0.243	0.293
9	327	17	0.253	0.01239		0.230	0.278
10	310	22	0.235	0.01208		0.212	0.260
11	288	17	0.221	0.01183		0.199	0.245
12	271	13	0.210	0.01162		0.189	0.234
13	258	1	0.210	0.01161		0.188	0.234
14	257	3	0.207	0.01156		0.186	0.231
15	254	2	0.206	0.01152		0.184	0.229
16	252	39	0.174	0.01081		0.154	0.196
17	213	1	0.173	0.01079		0.153	0.195
18	212	11	0.164	0.01056		0.144	0.186
19	201	1	0.163	0.01054		0.144	0.185
20	200	8	0.157	0.01037		0.138	0.178
21	192	11	0.148	0.01012		0.129	0.169
22	181	8	0.141	0.00993		0.123	0.162
23	173	4	0.138	0.00983		0.120	0.159
24	169	3	0.135	0.00976		0.118	0.156

Tabla 21 Resumen estimación de la función de supervivencia mediante el estimador Kaplan Meier

Se observa para el primer periodo de los 1,256 alumnos, 411 alumnos (datos no censurados) reprobaron por primera vez y existe una probabilidad de supervivencia de 0.67 en ese día. Para el segundo periodo de los 825 alumnos que no habían reprobado materias, 188 reprobaron, generando una probabilidad de supervivencia de 0.519. La probabilidad de supervivencia para el tercer periodo es de 0.387, para el cuarto periodo es de 0.371. A partir del periodo cinco con probabilidad de supervivencia de 0.334, la caída en la probabilidad disminuye paulatinamente. En el último periodo, el periodo 24, de los 169 alumnos que no habían reprobado, la probabilidad de supervivencia es de 0.135.

Gráfico No. 1. Estimador de Kaplan y Meier



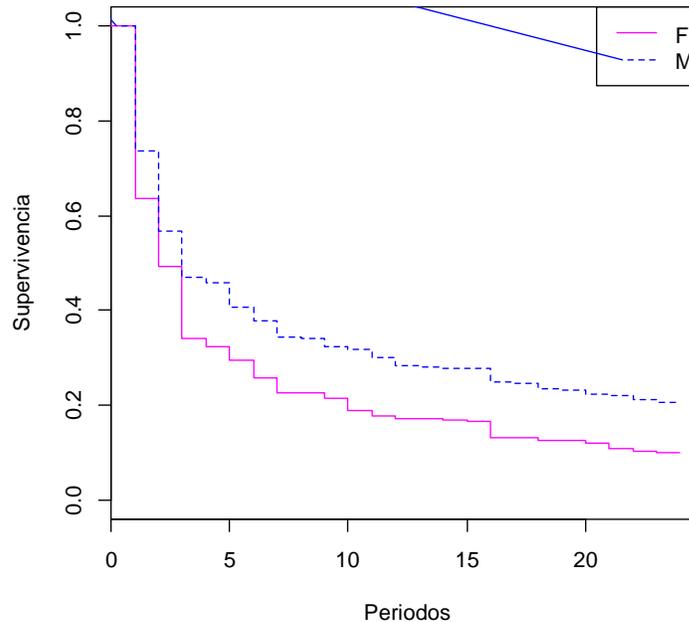
De forma visual se puede mencionar que a partir del periodo cinco la caída en supervivencia es más lenta para la población analizada, es decir, a partir de este punto la probabilidad de no reprobado alguna materia es baja y la diferencia entre cada periodo es mínima.

Capítulo 7. Comparación de las funciones de supervivencia por variable

En esta sección se presenta cada una de las variables mostradas anteriormente y se comparan sus respectivas funciones de supervivencia, donde se muestran las estimaciones de Kaplan Meier.

En la figura No.2 se muestra el estimador Kaplan Meier de las funciones de supervivencia por sexo de estudiantes.

Gráfico No. 2. Estimador de Kaplan y Meier por Género



Se observa que las funciones de supervivencia son distintas, la probabilidad estimada de no reprobar alguna materia es mayor en el grupo del sexo masculino.

```
Call:
survdif(formula = Surv(Survt, status) ~ genero, data = base,
        rho = 0)

      N Observed Expected (O-E)^2/E (O-E)^2/V
genero=F 806      717      643      8.61      28.1
genero=M 450      353      427     12.95      28.1

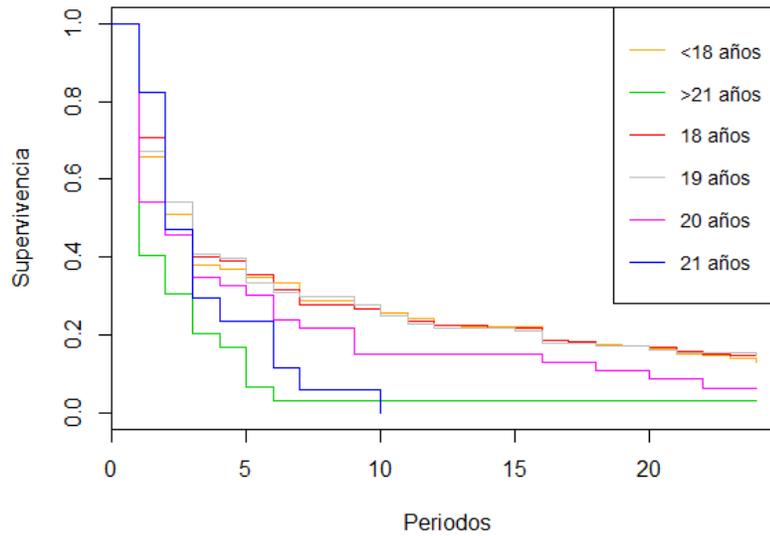
Chisq= 28.1 on 1 degrees of freedom, p= 1.17e-07
```

Tabla 22 Resumen estimación de las funciones de supervivencia por sexo mediante el estimador Kaplan Meier

De manera formal se observa en la tabla 22 que existen diferencias significativas en la variable sexo, ya que se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia a un nivel de significancia del 5%.

Por otro lado, considerando la edad de los estudiantes con la cual ingresaron a la licenciatura, se observa que las funciones de supervivencia son distintas.

Gráfico No.3.Estimador de Kaplan y Meier por Edad



Al comparar las funciones de supervivencia de manera formal se tiene la tabla 23.

```
Call:
survdif(formula = msurv ~ edad2, data = base, na.action = na.omit)

n=1165, 91 observations deleted due to missingness.

      N Observed Expected (O-E)^2/E (O-E)^2/V
edad2=<18 223      193   192.5  0.00125  0.002
edad2=>21  32       30    17.0  9.98212 14.029
edad2=18  686      577   600.0  0.88076  2.856
edad2=19  161      137   140.9  0.10553  0.158
edad2=20   46       43    34.6  2.01221  2.726
edad2=21   17       17    12.0  2.06648  2.831

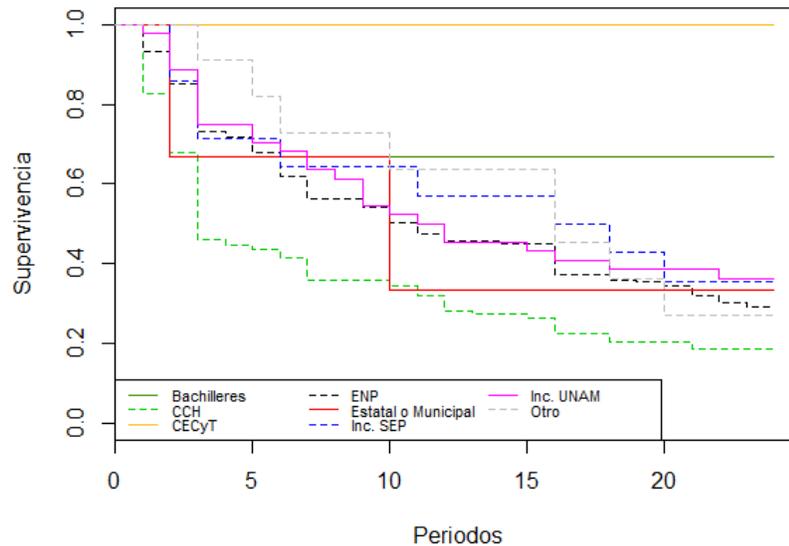
ChiSq= 20.6 on 5 degrees of freedom, p= 0.001
```

Tabla 23 Resumen estimación de las funciones de supervivencia por edades mediante el estimador Kaplan Meier

Con lo cual se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia, es decir, existen diferencias significativas por edades.

Dada la escuela de procedencia de los estudiantes, se muestran las distintas curvas de supervivencia.

Gráfico No. 4. Estimador de Kaplan y Meier por Escuela de Procedencia



Se observa que los alumnos que no indicaron la escuela de procedencia (NP) presentan la probabilidad estimada más baja de supervivencia, seguida de los alumnos procedentes del CCH.

```
Call:
survdifff(formula = msurv ~ procedencia, data = base)

n=583, 673 observations deleted due to missingness.

procedencia=Bachilleres      N Observed Expected (O-E)^2/E
procedencia=CCH              128    105    71.98  15.14836
procedencia=CECyT            1       0      1.23  1.23218
procedencia=ENP              379    270   288.54  1.19095
procedencia=Estatal o Municipal 3       2      2.13  0.00755
procedencia=INC. SEP         14       9     11.67  0.61063
procedencia=Inc. UNAM        44      28    34.98  1.39254
procedencia=otro             11       8     9.82  0.33689

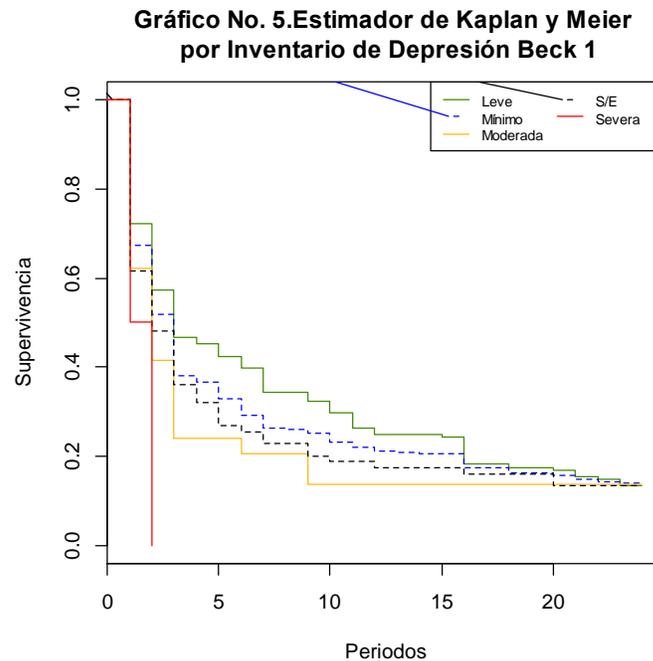
(0-E)^2/V
procedencia=Bachilleres      1.14133
procedencia=CCH              20.34203
procedencia=CECyT            1.35510
procedencia=ENP              4.15215
procedencia=Estatal o Municipal 0.00836
procedencia=INC. SEP         0.69384
procedencia=Inc. UNAM        1.67770
procedencia=otro             0.38199

Chisq= 23.3 on 7 degrees of freedom, p= 0.001
```

Tabla 24 Resumen estimación de las funciones de supervivencia por escuela de procedencia mediante el estimador Kaplan Meier

La significancia estadística es menor a 0.05, por lo cual se rechaza la hipótesis nula (H_0) de igualdad de funciones, se concluye que existen diferencia por escuela de procedencia.

Ahora, se comparan las funciones de supervivencia por resultados del Inventario de Depresión de Beck, aplicado al inicio del curso escolar.



Se observa en el gráfico no. 5 que los alumnos con resultado “Leve” presentan una mayor probabilidad estimada de no presentar el evento de interés. Mientras que los alumnos con “Severa” y “Moderada” tienen una probabilidad estimada baja de no reprobar alguna materia.

```
Call:
survdif(formula = Surv(Survt, status) ~ Psico_l, data = base,
         rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Psico_l=Leve	151	129	140.943	1.01207	1.4982
Psico_l=Mínima	983	843	840.935	0.00507	0.0307
Psico_l=Moderada	29	25	21.737	0.48973	0.6580
Psico_l=S/E	91	71	65.502	0.46155	0.6475
Psico_l=Severa	2	2	0.882	1.41575	2.0282

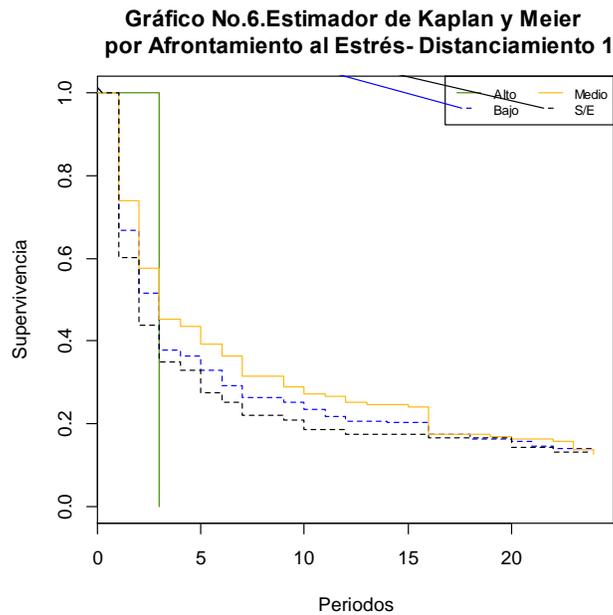
Chisq= 4.6 on 4 degrees of freedom, p= 0.331

Tabla 25 Resumen estimación de las funciones de supervivencia por resultados del inventario de depresión de Beck mediante el estimador Kaplan Meier

Formalmente, la hipótesis nula (H_0) de igualdad de funciones de supervivencia no se rechaza dado el $pvalue = 0.331$, lo que indica que no existe diferencia significativa entre las funciones de supervivencia.

A continuación, se muestra el análisis realizado a cada uno de los estilos de afrontamiento al estrés también aplicados al inicio del curso escolar.

El primer estilo es el Distanciamiento.



En la gráfica No. 6 se observa que el grupo de alumnos con resultado “Medio” tienen una mayor probabilidad estimada de no reprobado alguna materia.

```
Call:
survdif(formula = Surv(Survt, status) ~ Distanciamientol, data = base,
         rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Distanciamientol=Alto	1	1	0.811	0.04405	0.0609
Distanciamientol=Bajo	962	822	819.257	0.00919	0.0507
Distanciamientol=Medio	185	161	172.055	0.71029	1.0901
Distanciamientol=S/E	108	86	77.878	0.84713	1.2008

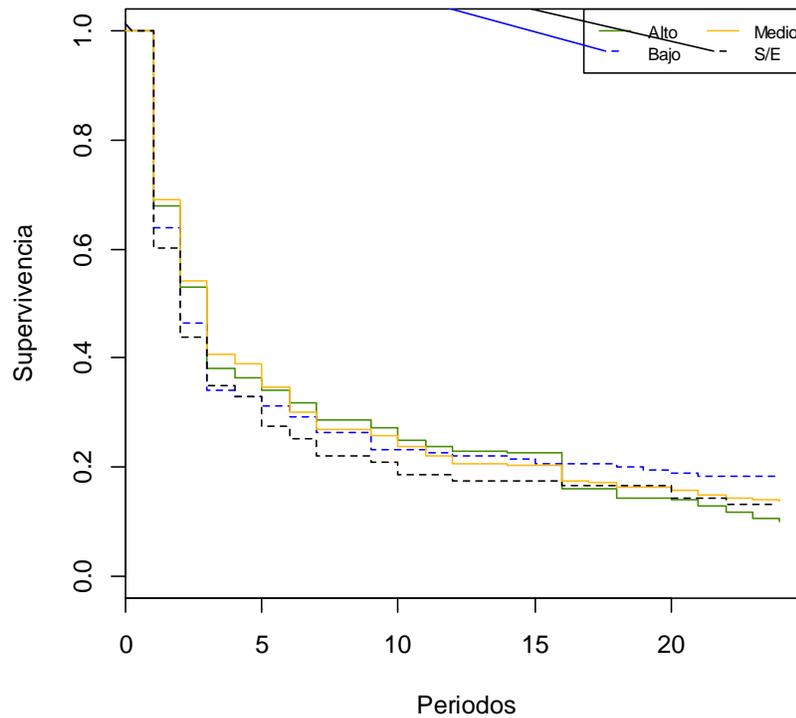
```
Chisq= 2.1 on 3 degrees of freedom, p= 0.551
```

Tabla 26 Resumen estimación de las funciones de supervivencia por resultados del cuestionario modos de afrontamiento-Distanciamiento mediante el estimador Kaplan Meier

La significancia estadística es mayor a 0.05 por lo que no se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia, es decir, la aplicación del estilo distanciamiento no influye en el tiempo de supervivencia.

El siguiente estilo de afrontamiento al estrés es dirigido a lo positivo, los resultados se presentan en la siguiente gráfica.

Gráfico No.7. Estimador de Kaplan y Meier por Afrontamiento al Estrés- Dirigido a lo Positivo 1



No se observa alguna diferencia significativa en cada una de las funciones de supervivencia.

```
Call:
survdif(formula = Surv(Survt, status) ~ DirigPositivo1, data = base,
        rho = 0)
```

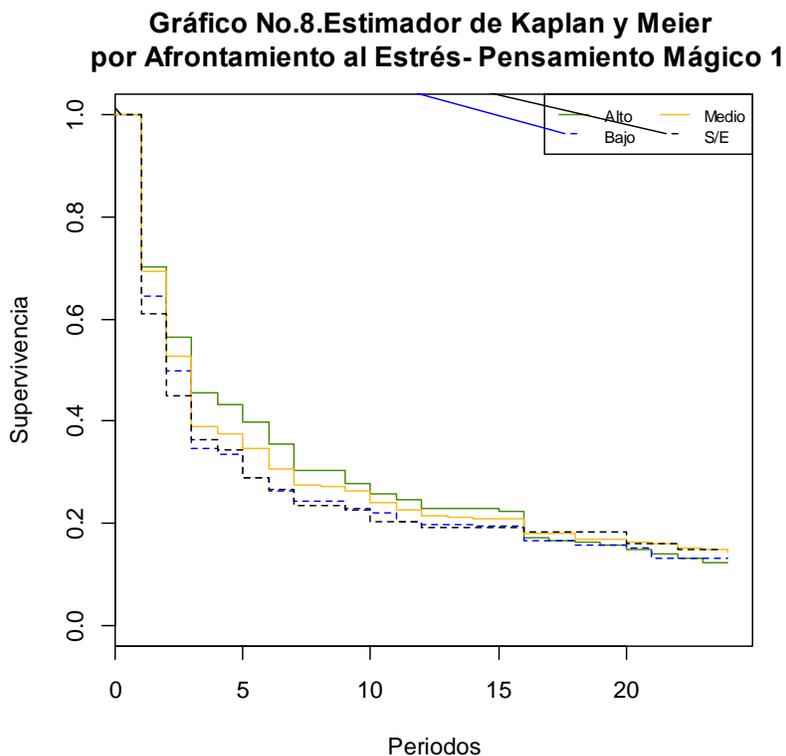
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
DirigPositivo1=Alto	261	233	226.3	0.1994	0.327
DirigPositivo1=Bajo	169	136	139.6	0.0921	0.137
DirigPositivo1=Medio	718	615	626.3	0.2022	0.632
DirigPositivo1=S/E	108	86	77.9	0.8471	1.201

Chisq= 1.8 on 3 degrees of freedom, p= 0.625

Tabla 27 Resumen estimación de las funciones de supervivencia por resultados del cuestionario modos de afrontamiento-Dirigido a lo positivo mediante el estimador Kaplan Meier

Formalmente, no se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia, se concluye que no hay diferencia significativa en la aplicación del estilo dirigido a lo positivo.

En relación con la aplicación del estilo de afrontamiento pensamiento mágico, se presentan los estimadores de Kaplan Meier de las funciones de supervivencia para cada uno de los resultados.



De la gráfica se concluye que significativamente no hay diferencia en cada una de las funciones.

```
Call:
survdif(formula = Surv(Survt, status) ~ Pens_Magicol, data = base,
        rho = 0)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Pens_Magicol=Alto  234      202    210.0     0.307     0.494
Pens_Magicol=Bajo  356      309    293.8     0.786     1.408
Pens_Magicol=Medio 556      473    484.7     0.280     0.664
Pens_Magicol=S/E   110       86     81.5     0.247     0.350

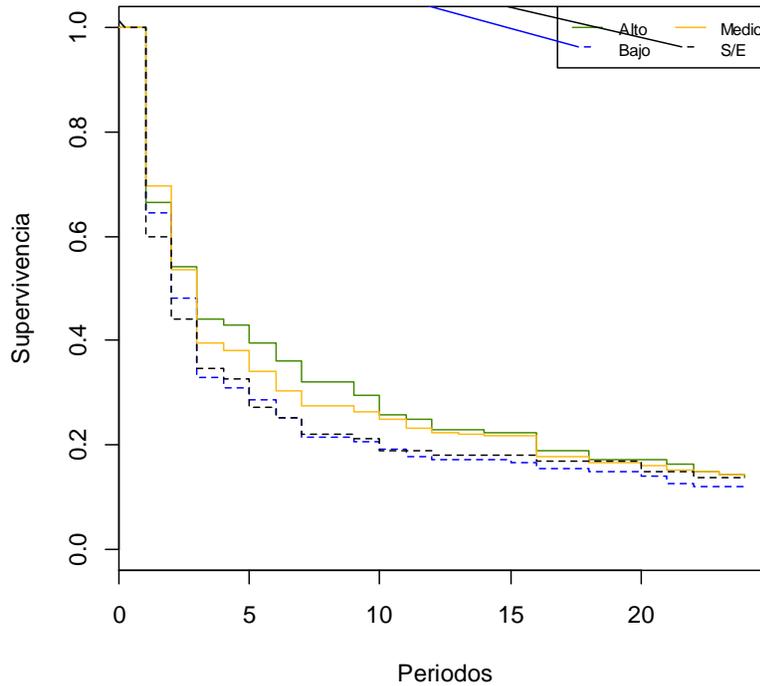
Chisq= 2.1  on 3 degrees of freedom, p= 0.55
```

Tabla 28 *Resumen estimación de las funciones de supervivencia por resultados del cuestionario modos de afrontamiento-Pensamiento mágico mediante el estimador Kaplan Meier*

Dado el pvalue= 0.55 no se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia, es decir, los resultados de la aplicación no influyen en el tiempo de supervivencia.

En la aplicación del estilo de afrontamiento al estrés Dirigido al Problema, se construye el gráfico, donde se muestran las estimaciones de Kaplan Meier.

Gráfico No. 9 .Estimador de Kaplan y Meier por Afrontamiento al Estrés- Dirigido al Problema 1



Del gráfico se puede concluir que no existe diferencia significativa en la primera aplicación del estilo de afrontamiento al estrés Dirigido al Problema.

```
Call:
survdif(formula = Surv(Survt, status) ~ DirigProbl, data = base,
rho = 0)
```

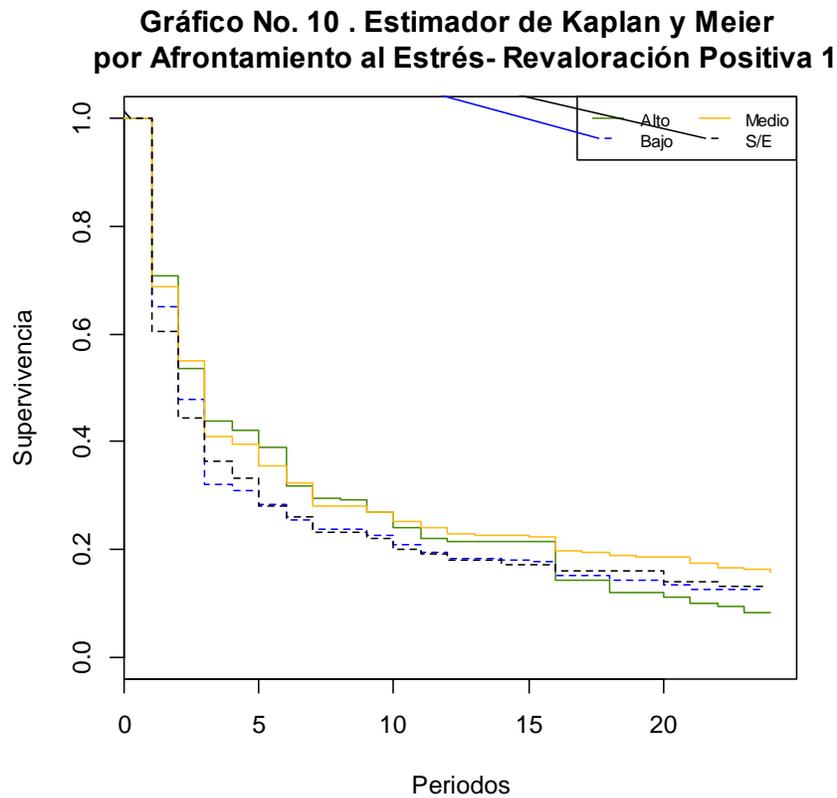
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
DirigProbl=Alto	199	171	178.9	0.351	0.543
DirigProbl=Bajo	216	188	171.4	1.614	2.510
DirigProbl=Medio	729	622	638.5	0.429	1.379
DirigProbl=S/E	112	89	81.2	0.757	1.076

Chisq= 4.1 on 3 degrees of freedom, p= 0.249

Tabla 29 Resumen estimación de las funciones de supervivencia por resultados del cuestionario modos de afrontamiento-Dirigido al problema mediante el estimador Kaplan Meier

La significancia estadística es mayor a 0.05, por lo cual no se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia.

A continuación, se comparan las funciones de supervivencia de la primera aplicación del estilo de afrontamiento al estrés Revaloración Positiva.



En la cual no se observa diferencia significativa entre cada uno de los resultados de la primera aplicación.

```
Call:
survdif(formula = Surv(Survt, status) ~ Revaloracion_positival,
        data = base, rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Revaloracion_positival=Alto	184	168	162.0	0.219	0.335
Revaloracion_positival=Bajo	322	278	258.9	1.416	2.434
Revaloracion_positival=Medio	634	531	564.3	1.962	5.389
Revaloracion_positival=S/E	116	93	84.8	0.785	1.119

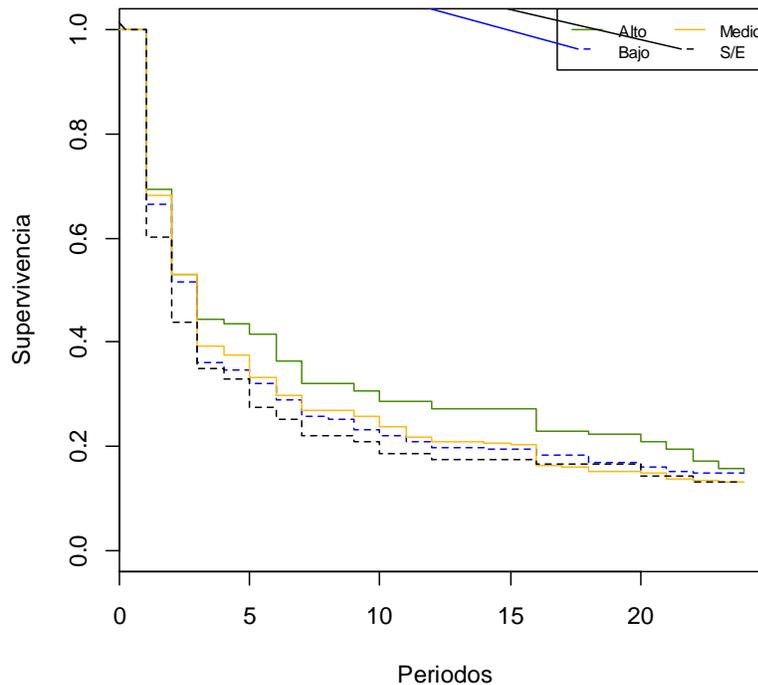
Chisq= 5.7 on 3 degrees of freedom, p= 0.127

Tabla 30 Resumen estimación de las funciones de supervivencia por resultados del cuestionario modos de afrontamiento-Revaloración positiva mediante el estimador Kaplan Meier

De manera formal no hay diferencias significativas en la variable, no se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia.

Para la última aplicación del estilo de afrontamiento al estrés de la primera aplicación, Flexibilidad, se muestran en la siguiente gráfica las funciones de supervivencia.

Gráfico No. 11 . Estimador de Kaplan y Meier por Afrontamiento al Estrés- Flexibilidad 1



Se observa que el segmento marcado como “Alto” presenta una probabilidad de supervivencia mayor que el resto de la población.

```
Call:
survdif(formula = Surv(Survt, status) ~ Flexibilidad1, data = base,
rho = 0)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
Flexibilidad1=Alto	140	119	130.5	1.00619	1.46969
Flexibilidad1=Bajo	275	232	231.0	0.00396	0.00655
Flexibilidad1=Medio	733	633	630.6	0.00897	0.02830
Flexibilidad1=S/E	108	86	77.9	0.84713	1.20080

Chisq= 2.4 on 3 degrees of freedom, p= 0.49

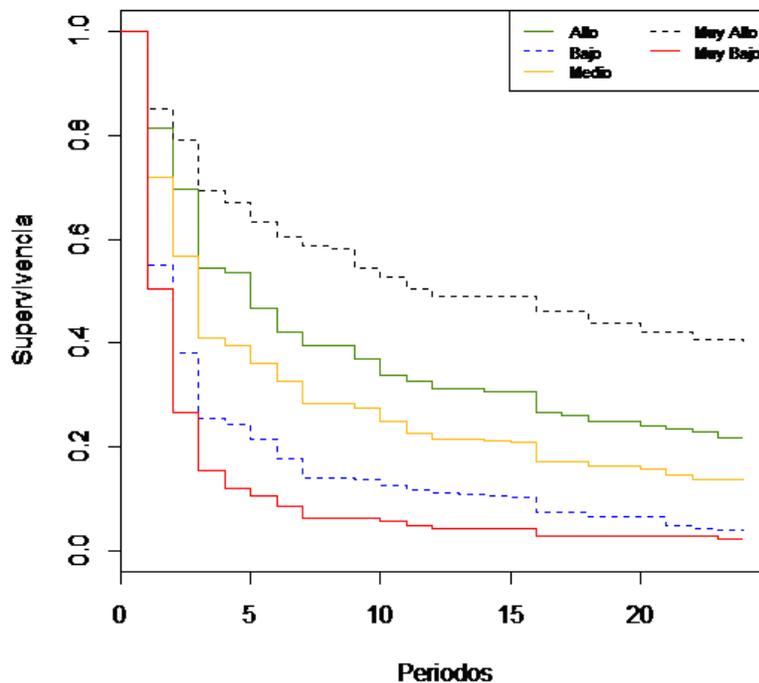
Tabla 31 Resumen estimación de las funciones de supervivencia por resultados del cuestionario modos de afrontamiento-Flexibilidad mediante el estimador Kaplan Meier

Sin embargo, dado el $pvalue=0.49$, formalmente no se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia, es decir, no existe diferencia significativa entre los segmentos.

Ahora, se muestra el análisis de los exámenes de conocimientos realizados al inicio del ciclo escolar, segmentado por materias.

Primero se muestra la función de supervivencia de la materia de matemáticas.

Gráfico No. 13. Examen de Conocimientos - Matemáticas



A diferencia de las pruebas de afrontamiento al estrés, en esta prueba de conocimientos si se observa una diferencia significativa entre los grupos. Los alumnos con resultado “Muy Alto” presentan una mayor probabilidad estimada de no reprobado alguna materia. Mientras que los alumnos con “Bajo” y “Muy Bajo” tienen una probabilidad estimada baja de no reprobado alguna materia.

```
Call:
survdif(formula = Surv(Survt, status) ~ Clasif_Mat, data = base,
        rho = 0)
```

n=1210, 46 observations deleted due to missingness.

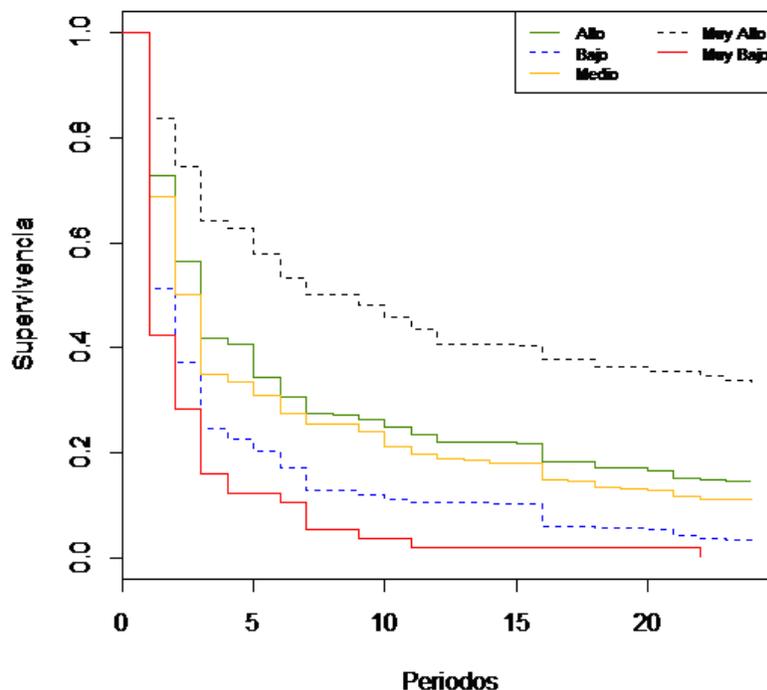
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Clasif_Mat=Alto	194	151	204.9	14.158	22.67
Clasif_Mat=Bajo	364	348	249.5	38.848	68.93
Clasif_Mat=Medio	366	317	330.8	0.574	1.09
Clasif_Mat=Muy alto	135	80	169.6	47.337	72.72
Clasif_Mat=Muy bajo	151	144	85.2	40.536	61.80

Chisq= 197 on 4 degrees of freedom, p= 0

Tabla 32 Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- matemáticas mediante el estimador Kaplan Meier

Formalmente, la hipótesis nula (H_0) de igualdad de funciones de supervivencia se rechaza, lo que indica que existe diferencia significativa entre las funciones de supervivencia.

Gráfico No. 14. Examen de Conocimientos - Física



En el examen de conocimiento para la materia de Física, los alumnos con resultado “Muy Alto” presentan una mayor probabilidad estimada de no reprobado alguna

materia y se encuentra muy alejado de los demás alumnos. Mientras que los alumnos con “Bajo” y “Muy Bajo” tienen una probabilidad estimada baja de no reprobado alguna materia.

```
Call:
survdif(formula = Surv(Survt, status) ~ Clasif_Fis, data = base,
         rho = 0)
```

n=1209, 47 observations deleted due to missingness.

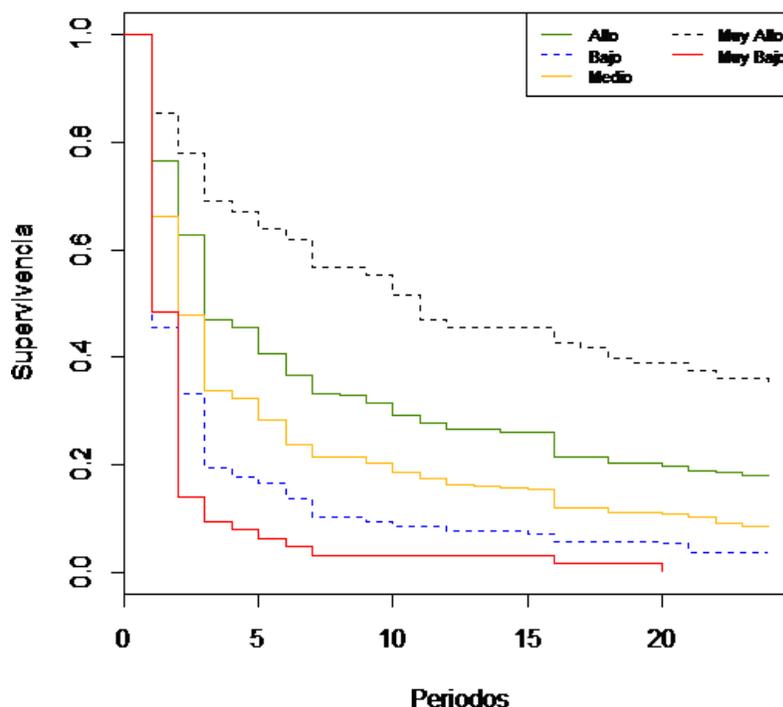
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Clasif_Fis=Alto	293	249	263.5	0.798	1.38
Clasif_Fis=Bajo	271	261	180.3	36.117	59.07
Clasif_Fis=Medio	371	328	308.6	1.213	2.25
Clasif_Fis=Muy alto	215	143	254.3	48.717	83.96
Clasif_Fis=Muy bajo	59	58	32.2	20.579	29.58

Chisq= 147 on 4 degrees of freedom, p= 0

Tabla 33 Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- física mediante el estimador Kaplan Meier

La hipótesis nula (H_0) de igualdad de funciones de supervivencia se rechaza, es decir, existe diferencia significativa entre las funciones de supervivencia.

Gráfico No. 15. Examen de Conocimientos - Química



De igual forma para la materia de Química los grupos se observan bien segmentados. Los alumnos con resultado “Muy Alto” presentan una mayor probabilidad estimada de no reprobado alguna materia y se encuentra muy alejado de los demás alumnos. Mientras que los alumnos con “Bajo” y “Muy Bajo” tienen una probabilidad estimada baja de no reprobado alguna materia.

```
Call:
survdif(formula = Surv(Survt, status) ~ Clasif_Quim, data = base,
        rho = 0)

n=1210, 46 observations deleted due to missingness.

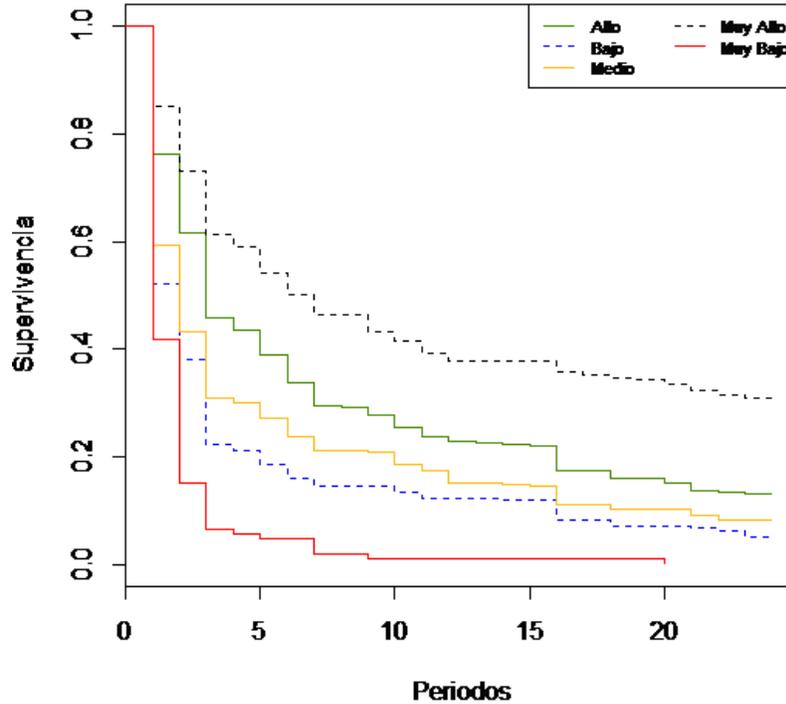
      N Observed Expected (O-E)^2/E (O-E)^2/V
Clasif_Quim=Alto    431     352    418.9    10.69    23.26
Clasif_Quim=Bajo    230     220    141.7    43.21    68.61
Clasif_Quim=Medio   345     314    274.8     5.58     9.94
Clasif_Quim=Muy alto 136      88    170.4    39.84    61.16
Clasif_Quim=Muy bajo  68      66     34.1    29.81    43.69

Chisq= 181 on 4 degrees of freedom, p= 0
```

Tabla 34 Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- química mediante el estimador Kaplan Meier

La significancia estadística es menor a 0.05, por lo cual se rechaza la hipótesis nula (H_0) de igualdad de funciones de supervivencia.

Gráfico No. 16. Examen de Conocimientos - Biología



Para la materia de Biología, los alumnos con resultado “Muy Alto” presentan una mayor probabilidad estimada de no reprobado alguna materia. Los alumnos con “Muy Bajo” tienen una probabilidad estimada baja de no reprobado alguna materia. Se observa una diferencia mínima entre los alumnos “Bajo” y “Medio”.

```
Call:
survdif(formula = Surv(Surv, status) ~ Clasif_Biol, data = base,
rho = 0)
```

n=1210, 46 observations deleted due to missingness.

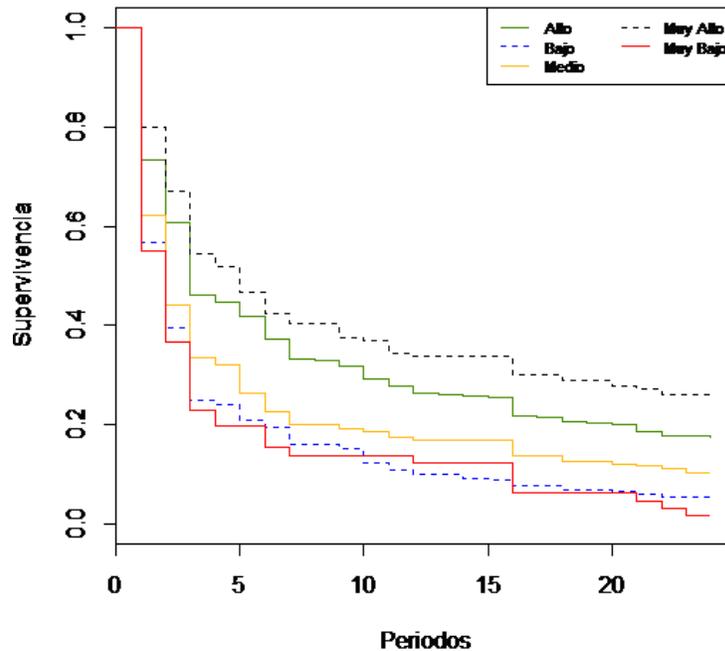
	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
Clasif_Biol=Alto	335	291	313.9	1.66	3.09
Clasif_Biol=Bajo	198	186	133.8	20.32	31.31
Clasif_Biol=Medio	289	263	218.9	8.88	14.79
Clasif_Biol=Muy alto	280	193	321.3	51.22	97.32
Clasif_Biol=Muy bajo	108	107	52.1	57.81	87.92

Chisq= 198 on 4 degrees of freedom, p= 0

Tabla 35 Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- biología mediante el estimador Kaplan Meier

La hipótesis nula (H_0) de igualdad de funciones de supervivencia se rechaza, es decir, existe diferencia significativa entre las funciones de supervivencia.

Gráfico No. 17. Examen de Conocimientos - Historia Universal



Para el examen de conocimientos de Historia Universal no se observa una buena segmentación, en algún punto los alumnos con resultado “Bajo” y “Muy Bajo” se cruzan, llegando a puntos cercanos a los alumnos con resultado “Medio”.

```
Call:
survdif(formula = Surv(Survt, status) ~ Clasif_Hist_Univ, data = base,
rho = 0)
```

n=1210, 46 observations deleted due to missingness.

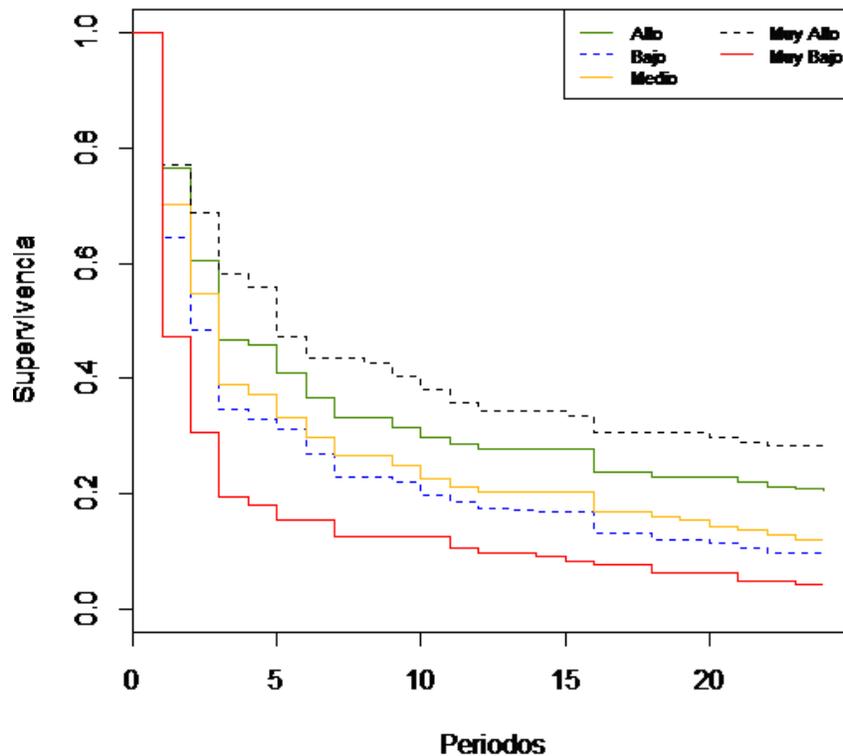
	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
Clasif_Hist_Univ=Alto	434	356	414.4	8.22	17.75
Clasif_Hist_Univ=Bajo	295	278	205.8	25.30	42.36
Clasif_Hist_Univ=Medio	225	202	176.7	3.64	5.74
Clasif_Hist_Univ=Muy alto	185	137	196.7	18.12	28.72
Clasif_Hist_Univ=Muy bajo	71	67	46.4	9.11	12.83

Chisq= 86.5 on 4 degrees of freedom, p= 0

Tabla 36 Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- historia universal mediante el estimador Kaplan Meier

Formalmente, la hipótesis nula (H_0) de igualdad de funciones de supervivencia se rechaza, lo que indica que existe diferencia significativa entre las funciones de supervivencia.

Gráfico No. 18. Examen de Conocimientos - Historia México



A diferencia del examen de conocimientos de Historia Universal, los grupos por calificaciones para el examen de Historia de México no se cruzan, pero visualmente se detecta que no presentan una buena segmentación como los anteriores exámenes de conocimientos.

```
Call:
survdif(formula = Surv(Survt, status) ~ Clasif_Hist_Mx, data = base,
        rho = 0)
```

n=1210, 46 observations deleted due to missingness.

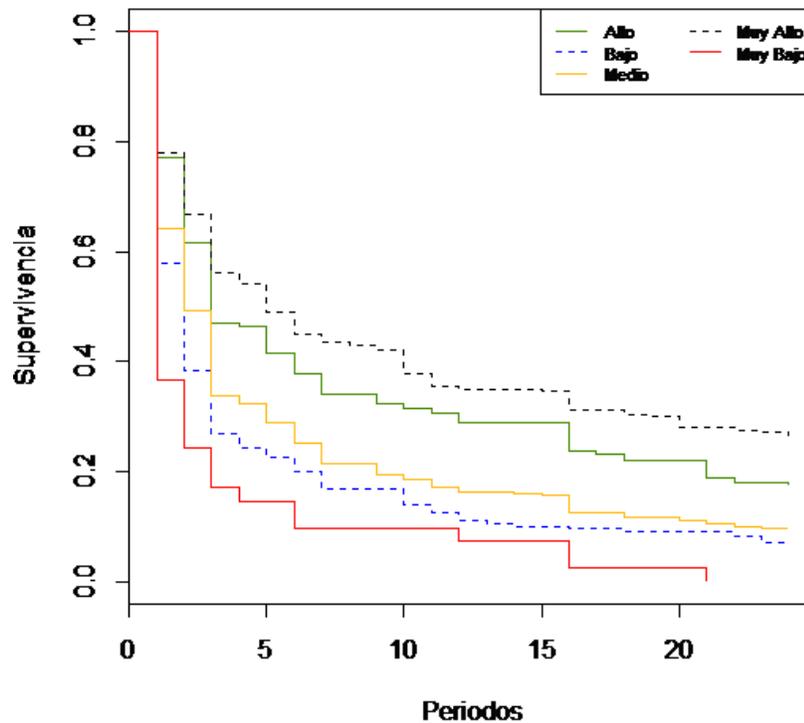
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Clasif_Hist_Mx=Alto	233	185	228.0	8.11085	13.38402
Clasif_Hist_Mx=Bajo	469	420	378.6	4.52765	9.29207
Clasif_Hist_Mx=Medio	229	200	199.4	0.00179	0.00288
Clasif_Hist_Mx=Muy alto	131	95	141.8	15.42249	22.82096
Clasif_Hist_Mx=Muy bajo	148	140	92.2	24.72959	36.79315

Chisq= 70.5 on 4 degrees of freedom, p= 1.8e-14

Tabla 37 Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- historia de México mediante el estimador Kaplan Meier

La hipótesis nula (H_0) de igualdad de funciones de supervivencia se rechaza, es decir, existe diferencia significativa entre las funciones de supervivencia.

Gráfico No. 19. Examen de Conocimientos - Literatura



En el examen de conocimiento para la materia de Literatura no se observa una buena segmentación entre los grupos de resultados. Sin embargo, los alumnos con

resultado “Muy Alto” presentan una mayor probabilidad estimada de no reprobado alguna materia y se encuentra muy alejado de los demás alumnos. Mientras que los alumnos con “Bajo” y “Muy Bajo” tienen una probabilidad estimada baja de no reprobado alguna materia.

```
Call:
survdif(formula = Surv(Survt, status) ~ Clasif_Lit, data = base,
         rho = 0)

n=1210, 46 observations deleted due to missingness.

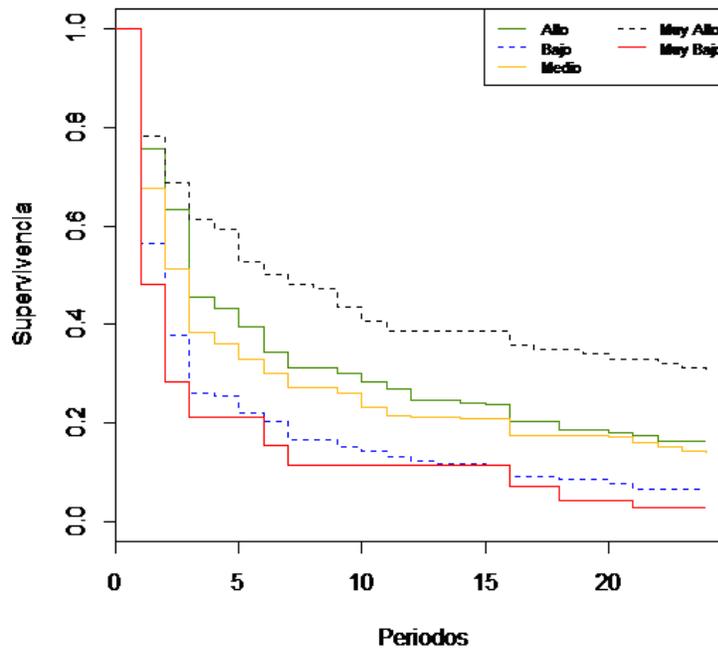
      N Observed Expected (O-E)^2/E (O-E)^2/V
Clasif_Lit=Alto    231     189     227      6.43     10.6
Clasif_Lit=Bajo    214     196     151     13.38     20.8
Clasif_Lit=Medio   510     456     407      5.83     12.6
Clasif_Lit=Muy alto 214     158     231     23.30     38.7
Clasif_Lit=Muy bajo  41      41      23     13.98     19.6

Chisq= 84.3 on 4 degrees of freedom, p= 0
```

Tabla 38 *Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- literatura mediante el estimador Kaplan Meier*

La hipótesis nula (H_0) de igualdad de funciones de supervivencia se rechaza, es decir, existe diferencia significativa entre las funciones de supervivencia.

Gráfico No. 20. Examen de Conocimientos - Geografía



Para el último examen de conocimientos, Geografía, se observa que las funciones de supervivencia de los grupos se cruzan en varios puntos.

Call:

```
survdif(formula = Surv(Survt, status) ~ Clasif_Geo, data = base,
        rho = 0)
```

n=1210, 46 observations deleted due to missingness.

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
Clasif_Geo=Alto	419	349	399.7	6.43680	13.566
Clasif_Geo=Bajo	326	303	229.9	23.23699	39.816
Clasif_Geo=Medio	284	243	244.3	0.00651	0.011
Clasif_Geo=Muy alto	106	74	119.6	17.39928	25.026
Clasif_Geo=Muy bajo	75	71	46.5	12.92947	18.280

Chisq= 80.3 on 4 degrees of freedom, p= 1.11e-16

Tabla 39 Resumen estimación de las funciones de supervivencia por resultados del examen de conocimientos- geografía mediante el estimador Kaplan Meier

Formalmente, la hipótesis nula (H_0) de igualdad de funciones de supervivencia se rechaza, lo que indica que existe diferencia significativa entre las funciones de supervivencia.

De acuerdo con el análisis de estimaciones de Kaplan Meier para cada una de las variables del estudio, se observa que existen variables significativas como el sexo, edades con la que ingresan a la carrera, la escuela de procedencia, exámenes de conocimientos de matemáticas, física, química, biología, historia universal, historia de México, literatura y geografía. Por lo cual, las variables que no muestran diferencia significativa son el inventario de depresión de Beck y los resultados de afrontamiento al estrés.

Capítulo 8. Regresión de Cox

Como anteriormente se mencionó, el modelo introduce un conjunto de variables, que permita estimar la probabilidad de que ocurra el evento de la primera reprobada durante el curso escolar.

Se realizaron varias regresiones con todas las variables del estudio, tanto significativas o no significativas en las estimaciones de Kaplan Meier.

A continuación, se muestra el mejor resultado de la aplicación de la regresión de Cox, se incluyeron las siguientes variables: sexo (F,M) , nivel obtenido en el examen diagnóstico en el área de matemáticas (Muy bajo, Bajo, Medio, Alto, Muy alto), edad, nivel obtenido en el examen diagnóstico en el área de física (Muy bajo, Bajo, Medio, Alto, Muy alto), nivel obtenido en el examen diagnóstico en el área de biología (Muy bajo, Bajo, Medio, Alto, Muy alto) y nivel obtenido durante la evaluación del Inventario de Depresión de Beck (Leve, Mínima, Moderada, Severa).

```
Call:
coxph(formula = surv(Surv(t, status) ~ genero + Clasif_Mat + edad3 +
  Clasif_Fis + Clasif_Biol + Beck_1, data = base)
```

```
n= 1111, number of events= 952
(145 observations deleted due to missingness)
```

	coef	exp(coef)	se(coef)	z	Pr(> z)	
generoM	-0.10792	0.89770	0.07255	-1.487	0.13689	
Clasif_MatBajo	0.57691	1.78053	0.10601	5.442	5.27e-08	***
Clasif_MatMedio	0.22544	1.25288	0.10437	2.160	0.03077	*
Clasif_MatMuy alto	-0.31711	0.72825	0.15479	-2.049	0.04051	*
Clasif_MatMuy bajo	0.70141	2.01659	0.13035	5.381	7.41e-08	***
edad3	0.23091	1.25975	0.03574	6.461	1.04e-10	***
Clasif_FisBajo	0.21373	1.23829	0.09643	2.217	0.02665	*
Clasif_FisMedio	-0.04697	0.95412	0.08928	-0.526	0.59884	
Clasif_FisMuy alto	-0.29349	0.74566	0.11727	-2.503	0.01233	*
Clasif_FisMuy bajo	0.27784	1.32027	0.16677	1.666	0.09571	.
Clasif_BiolBajo	0.31549	1.37094	0.10109	3.121	0.00180	**
Clasif_BiolMedio	0.22282	1.24959	0.08960	2.487	0.01289	*
Clasif_BiolMuy alto	-0.28799	0.74977	0.10381	-2.774	0.00554	**
Clasif_BiolMuy bajo	0.80862	2.24481	0.12857	6.290	3.18e-10	***
Beck_1Mínima	0.21089	1.23477	0.09845	2.142	0.03219	*
Beck_1Moderada	0.35521	1.42648	0.22981	1.546	0.12219	
Beck_1Severa	1.57529	4.83213	0.72281	2.179	0.02930	*

```
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
generoM	0.8977	1.1140	0.7787	1.0349
Clasif_MatBajo	1.7805	0.5616	1.4465	2.1917
Clasif_MatMedio	1.2529	0.7982	1.0211	1.5373
Clasif_MatMuy alto	0.7283	1.3731	0.5377	0.9864
Clasif_MatMuy bajo	2.0166	0.4959	1.5619	2.6036
edad3	1.2597	0.7938	1.1745	1.3511
Clasif_FisBajo	1.2383	0.8076	1.0250	1.4959
Clasif_FisMedio	0.9541	1.0481	0.8010	1.1366
Clasif_FisMuy alto	0.7457	1.3411	0.5925	0.9383
Clasif_FisMuy bajo	1.3203	0.7574	0.9522	1.8307
Clasif_BiolBajo	1.3709	0.7294	1.1245	1.6714
Clasif_BiolMedio	1.2496	0.8003	1.0483	1.4895
Clasif_BiolMuy alto	0.7498	1.3337	0.6117	0.9190
Clasif_BiolMuy bajo	2.2448	0.4455	1.7448	2.8881
Beck_1Mínima	1.2348	0.8099	1.0181	1.4976
Beck_1Moderada	1.4265	0.7010	0.9092	2.2381
Beck_1Severa	4.8321	0.2069	1.1719	19.9249

```
Concordance= 0.698 (se = 0.01 )
Likelihood ratio test= 327.2 on 17 df, p=<2e-16
wald test = 321.7 on 17 df, p=<2e-16
score (logrank) test = 346.4 on 17 df, p=<2e-16
```

De lo anterior se puede reportar que los pvalores de las tres pruebas (razón de verosimilitudes, prueba de Wald y Score) son significativas ya que son menores a

0.05, rechazando así la hipótesis nula de que todos los coeficientes son cero. Por lo cual este modelo es significativo.

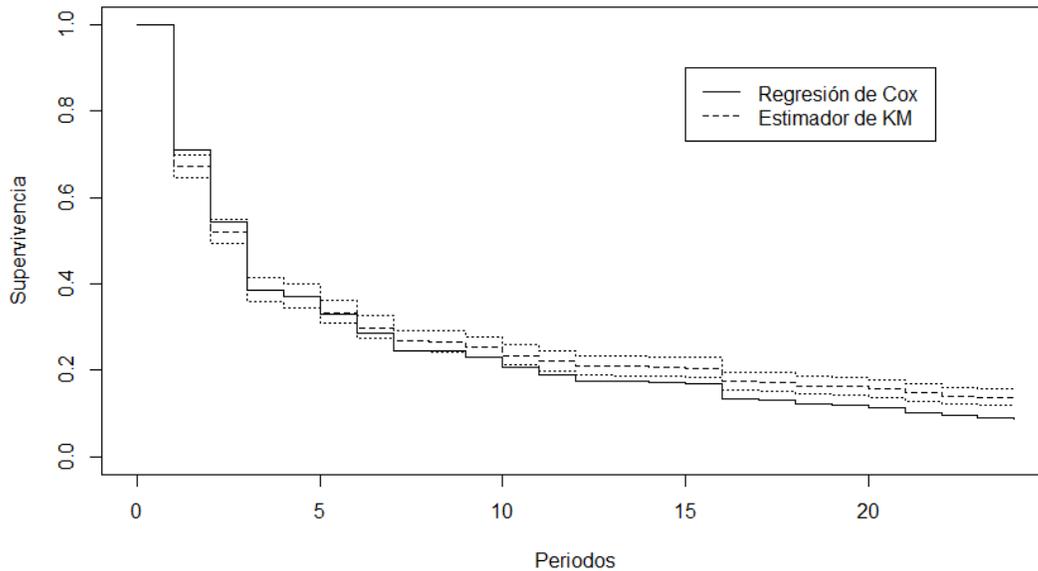
De igual forma, sabiendo que el reporte obtenido mediante R ordena los factores alfabéticamente, el exponencial de los coeficientes muestra que los alumnos de sexo masculino son 10% menos riesgosos a reprobar una materia que las alumnas. Por nivel obtenido en el examen diagnóstico en el área de matemáticas se observa una segmentación adecuada, los alumnos de nivel Muy Bajo tienen 100% más de probabilidad de reprobar que los de nivel Alto, los de nivel Bajo tienen 78% de probabilidad mayor de presentar el evento que los Alto, los Medio 25% mayor que los Alto y los de nivel Muy Alto presentan 19% de probabilidad menor de reprobar que los Alto. Por cada año que los alumnos son mayores la probabilidad de reprobar una materia es 26% mayor. Para el examen diagnóstico en el área de Física los alumnos Muy Bajo tienen 32% de probabilidad mayor de presentar el evento de interés, los Bajo 23% mayor que los Alto, los Medio presentan 4% de probabilidad menor de presentar el evento que los Alto y los Muy Alto 25% menor que los Alto.

Por nivel obtenido en el examen diagnóstico en el área de biología se observa que los de nivel Muy Bajo tienen 114% de mayor probabilidad de reprobar que los Alto, los Bajo 37% mayor, Medio 25%, mientras que los Muy Alto presentan 25% de probabilidad menor de presentar el evento.

Finalmente, los resultados del inventario de Depresión de Beck, los alumnos con clasificación Severa tienen 4 veces mayor probabilidad de reprobar una materia, los Mínima y Moderada presentan 23% y 42% respectivamente de probabilidad mayor de reprobar una materia que los Leve.

Tomando el modelo anterior, se obtiene la función de supervivencia ajustada y se compara con el estimador de Kaplan Meier, el resultado es el siguiente.

Gráfico No. 21. Comparación del ajuste del modelo de Cox y el estimador de Kaplan y Meier



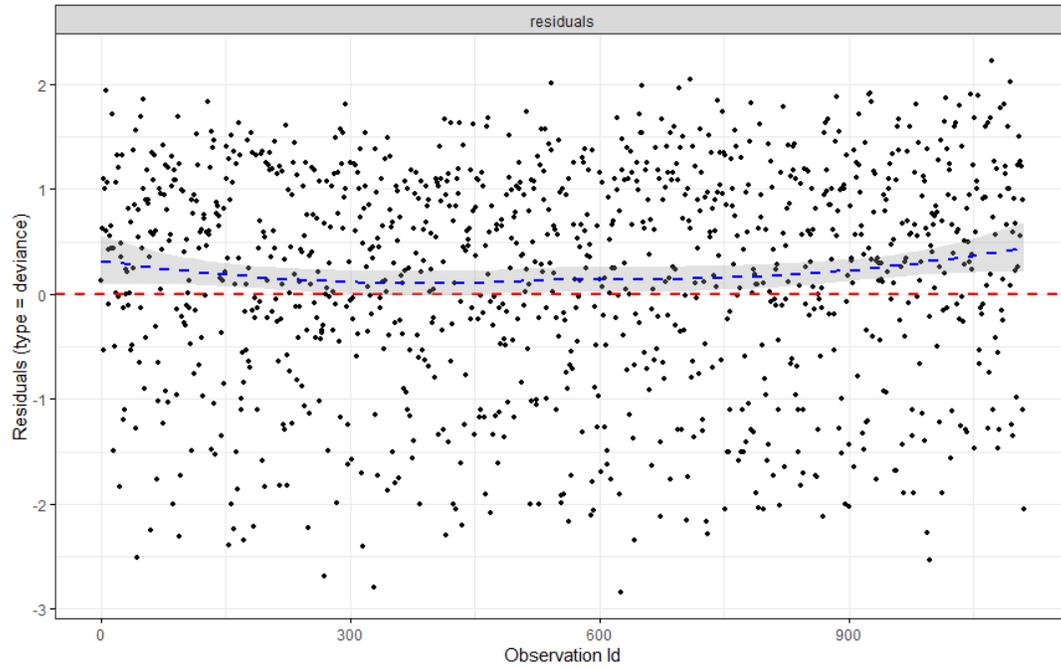
Se observa que la regresión de Cox presenta una función de supervivencia inferior a la del estimador de Kaplan y Meier.

Capítulo 9. Verificación del modelo de Cox

A continuación, se muestran los resultados para el modelo que contempla todas las variables.

	chisq	df	p
genero	2.346	1	0.13
Clasif_Mat	4.500	4	0.34
edad3	0.482	1	0.49
Clasif_Fis	1.058	4	0.90
Clasif_Biol	13.238	4	0.01
Beck_1	1.771	3	0.62
GLOBAL	23.097	17	0.15

De la tabla de resultados se observa el indicador GLOBAL indica que en conjunto no existe evidencia significativa de que no se cumpla el supuesto de riesgos proporcionales. Para este modelo la forma de los residuos deviance se observan en el siguiente gráfico.



En el gráfico de residuales del tipo deviance no es posible observar patrones definidos, por lo cual ningún individuo del estudio influye en la estimación del modelo.

Conclusiones

El presente estudio contribuye como una alternativa estadística para la identificación temprana al bajo rendimiento académico.

El universo analizado consistía en 1,256 alumnos de la licenciatura de medicina, a los cuales al momento de comenzar su ciclo escolar se les realizaron diversas pruebas de conocimientos, inventario de depresión y afrontamiento al estrés. Se realizó un seguimiento por dos años para saber el momento en los cuales reprobaban la primera materia.

En este análisis no es de interés conocer el momento en el que ocurrirá el evento. El interés u objetivo radica en obtener las variables que están influyendo en el evento de interés, reprobación de una materia, por lo cual el modelo de Cox nos permitió observar cómo influyen las variables significativas en el evento de estudio, ya que se comprobaron los supuestos de riesgos proporcionales.

Se concluye que, para alumnos de nuevo ingreso, las variables que influyen en el modelo de Cox fueron el sexo, el nivel obtenido en el examen diagnóstico en el área de matemáticas, la edad con la que ingresan los alumnos, nivel obtenido en el examen diagnóstico en el área de física, nivel obtenido en el examen diagnóstico en el área de biología y el nivel obtenido en la evaluación del Inventario de Depresión de Beck. Para la variable sexo las mujeres presentan mayor probabilidad a reprobación de una materia que los hombres; para los exámenes diagnóstico de matemáticas, física y biología los alumnos con el nivel “muy bajo” tienen mayor probabilidad a reprobación de una materia; por edades por cada año que los alumnos son mayores la probabilidad de reprobación crece. Finalmente, los alumnos con clasificación “severa” en el inventario de depresión de Beck presentan mayor probabilidad de reprobación de una materia.

Éstas son las variables que influyen el riesgo de reprobación de una materia durante los primeros dos años de licenciatura.

De esta forma las variables mostradas, ayudarán a identificar los factores de bajo desempeño académico y así se podrán implementar esfuerzos desde el principio del ciclo en los alumnos con mayor probabilidad de reprobación alguna materia en el transcurso del tiempo.

Bibliografía

Allison P. D. Survival Analysis Using the SAS System. SAS Institute, 2 edition, 2010.

Collet David. Modelling Survival Data in Medical Research. Chapman-Hall, 2 edition, 1994.

Cox D. R. and Snell E. J. A general definition of residuals. J.R. Statist. Soc., 1968.

Folkman, S and Lazarus R.S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of a college examination. Journal of Personality and Social Psychology; 48: 150-170.

Fouilloux, M.C., Barragan, P.V., Ortiz, L.S., Jaimes, M.A., Urrutia, A.E., Guevara, GR. (2013). Síntomas depresivos y rendimiento académico en estudiantes de medicina. Revista Salud Mental. 36: 59-65.

Hosmer W.D. and Lemeshow S. Applied Survival Analysis: Regression Modeling of Time to Event Data. Wiley Series in Probability and Statistics, 1998.

Informe de la Facultad de Medicina 2008-2012. Disponible en:

http://www.facmed.unam.mx/_documentos/informe/2011/index.html

Informe de la Facultad de Medicina 2012-2015. Disponible en:

http://www.facmed.unam.mx/_documentos/informe/2015/inf_2k12_2k15.pdf

Jurado, S., Villegas, M.E., Mendez, L., Rodriguez, F. et al. La estandarización del Inventario de Depresión de Beck para los residentes de la ciudad de México. Salud Mental 1998;21(3):26-31.

Klein P. J. and. Moeschberger M. L. Survival Analysis: Techniques for Censored and Truncated Data. Springer, 1997.

Klein, P. J., and Moeschberger, L. M. Survival Analysis Techniques for Censored and Truncated Data. Springer, 2003

Organización Mundial de la Salud. Disponible en:

https://www.who.int/mental_health/management/depression/es/

Secretaría de Salud. Día Mundial de la Salud Mental: Estudian prevalencia de trastornos depresivos en universitarios; 2015. Disponible en: http://www.conacytprensa.mx/index.php/ciencia/salud/3269-estudian-prevalencia-de-trastornos-depresivos-en-universitarios?utm_source=newsletter_1466&utm_medium=email&utm_campaign=prevalencia-de-trastornos-depresivos-en-universitarios

Sotelo, C.M y Maupone V. (2000). Traducción y estandarización del cuestionario modos de afrontamiento al estrés de Lazarus y Folkman, para una población de adolescentes mexicanos. México, D.F. Facultad de Psicología, Universidad Nacional Autónoma de México

Urrutia-Aguilar, M.E, Ortiz-León, S., Fouilloux M. C., Ponce R, E.R y Guevara-Guzmán R. (2014). El rendimiento académico en el primer año de la carrera de médico cirujano: modelo multivariado explicativo. Gaceta Médica de México.150 (S 3): 324-30.