PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO

MODELAMIENTO DEL TIEMPO A LA OCURRENCIA DE UN EVENTO CON TIEMPOS DISCRETOS

TESIS PARA OPTAR POR EL GRADO DE MAGÍSTER EN ESTADÍSTICA

Presentado por:

Anthony Enrique Huertas Quispe

Asesor: Dr. Cristian Luis Bayes Rodriguez

Miembros del jurado:
Dra. Rocio Paola Maehara Aliaga.
Dr. Luis Valdivieso Serrano.
Dr. Cristian Luis Bayes.

Lima, Diciembre 2018

Dedicatoria

Esta tesis va dedicada a mi madre, cuyo apoyo continuo, a lo largo de mi vida personal y profesional, ha logrado mantener en mi la motivación suficiente de superarme día a día.

Agradecimentos

Agradezco a mis maestros de toda la vida por haberme inculcado tan buenos conocimientos que al día de hoy puedo decir que es la base de muchos éxitos profesionales y agradezco enormemente a mi familia por la admiración y respeto mutuo.

Resumen

En este trabajo de tesis, se plantea estudiar el tiempo a la ocurrencia de un evento en un proceso discreto. Para ello, se considera un modelo mixtura de fracción de cura sobre una población segmentada en dos tipos de individuos: sujetos curados, o también denominados sobrevivientes a largo plazo, haciendo referencia a aquellos sujetos que no alcanzarán el evento de interés en estudio; y sujetos no curados, o también denominados sujetos susceptibles, quienes en un tiempo específico, experimentarán dicho evento de interés.

Los objetivos principales de esta tesis, son el de estimar la fracción de cura, la cual está definida como la proporción de individuos curados al final del estudio, y estimar el tiempo de falla para los individuos susceptibles, entendiéndose como el tiempo a la ocurrencia del evento. Este análisis se llevará a cabo con la presencia de covariables y datos censurados, siendo la simulación e inferencia de los datos efectuados vía el software estadístico R, en donde los procesos de simulación abordarán distintos escenarios para evaluar la performance del modelo propuesto.

Palabras-clave: modelo mixtura, censura, fracción de cura.

Índice general

Lista de Símbolos

T_0	Variable aleatoria de tiempo de susceptibilidad.
T	Variable aleatoria de tiempo de susceptibilidad bajo $\mathbf{x}$.
T_c	Variable aleatoria de tiempo de censura.
T_*	Variable aleatoria mixta bajo una población de susceptibles y no susceptibles.
$f_0(t)$	Función de densidad de probabilidad de T_0.
$f_c(t)$	Función de densidad de probabilidad de T_c.
$F(t)$	Función de distribución acumulada de T_0.
$S_0(t)$	Función de supervivencia.
$S_c(t)$	Función de supervivencia de T_c.
$S_0(t-)$	Función de supervivencia con límite izquierdo hacia t en T_0.
$S_c(t-)$	Función de supervivencia con límite izquierdo hacia t en T_c.
$S_*(t-)$	Función de supervivencia con límite izquierdo hacia t en T_*.
$\lambda_0(t)$	Tasa de riesgo basal en t.
λ_{t0}	Tasa de riesgo basal en t (Otra notación).
Δ	Indicador de censura.
t_f	Tiempo límite de estudio para la población susceptible.
$\mathcal{L}(\lambda)$	Función de verosimilitud sobre los parámetros $\lambda_0(t)$.
$\mathbf{x}$	Vector de covariables.
β	Vector de efectos de las covariables.
$\mathcal{L}(\theta)$	Función de verosimilitud sobre los parámetros $\lambda_0(t), p$ y β.
$\lambda(t \mid \mathbf{x})$	Tasa de riesgo en t bajo las covariables $\mathbf{x}$ en T.
$\lambda_*(t \mid \mathbf{x})$	Tasa de riesgo en t bajo las covariables $\mathbf{x}$ en T_*.
$H(t)$	Función de riesgo acumulada.
HR	radio de riesgos.
OR	radio de odds de riesgos.
η	Variable dicotómica que representa a los individuos curados.
p	Proporción de individuos curados.
$p(\mathbf{x})$	Proporción de individuos curados bajo efecto de las covariables.

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

1.1. Consideraciones preliminares

Existe una gran variedad de modelos de supervivencia, los cuales vienen enfocados a la susceptibilidad de los individuos, cualidad asignada a la ocurrencia de un evento y representada por un tiempo que es punto de interés, y denominada *tiempo de falla*. Comúnmente se engloba una población homogénea bajo la característica de susceptibilidad pues se asume que eventualmente todo individuo alcanzará en un tiempo particular el evento de interés, permitiéndose técnicas de análisis estándar [Kumar et al., 2016]. Sin embargo, el estudio que se llevará a cabo en la presente tesis, es el análisis sobre un modelo de fracción de cura, particularmente un *modelo de mixtura con supervivencia*, la cual segmenta la población en dos tipos de individuos, los susceptibles y los no susceptibles, siendo estos últimos denominados sobrevivientes a largo plazo o curados, siendo aquellos individuos que nunca alcanzarán el evento de interés después de haberse concluido un límite de seguimiento, definiéndose la proporción de estos como *fracción de cura* [Zhao and Zhou, 2008].

Los 50's fue cuna del primer modelo de fracción de cura, propuesto por Boag [1949] quien diseñó un modelo que introdujo un componente representando la fracción de cura, definido desde un punto de vista médico como aquellos individuos con una específica enfermedad, y una distribución latente representando la distribución de supervivencia de los individuos susceptibles.

Si bien, modelos de tiempo continuo, como el diseñado por Boag [1949], son tratados con mayor frecuencia en el área de supervivencia, modelos de tiempo discreto se ajustan mejor cuando los datos son recolectados período tras período; como lo es por ejemplo en estudios demográficos en donde se recolecta información retrospectivamente, o en ciertos estudios médicos como el realizado por Scheike and Kold [1997], quienes modelan el tiempo como ciclos menstruales hasta lograr el embarazo, o cuando tiempos continuos han sido agrupados en intervalos.

Una característica principal de este tipo de modelo de fracción de cura es la presencia de censura, siendo esta asignada a datos con información incompleta respecto a su tiempo de falla, que a su vez tienen que ser incorporadas en el modelo y no ignoradas evitándose una incorrecta inferencia. Es importante la inserción del mecanismo de censura en el modelo, indicando que existen varios tipos de este y dándose a conocer posteriormente.

Cabe mencionar que en el área de la biomedicina los modelos de fracción de cura son altamente estudiados a causa del surgimiento de datos que permiten extender diseños de

varios tipos [Schmid et al., 2016]. Sin embargo, en la presente tesis, se propone un modelo de fracción de cura, estudiado bajo distintos escenarios y aplicado sobre datos que corresponden a estudiantes de una institución destinada a la enseñanza superior, en donde la variable de interés es el tiempo de permanencia en dicha institución hasta su abandono. Esta información asignará los tiempos de forma discreta cuyos valores corresponderán a los ciclos cursados por un estudiante. Además, para efectos prácticos, se asume que el estudio sobre todos los estudiantes parte en un tiempo establecido y que aquellos que logren estar presentes hasta el máximo tiempo de estudio serán considerados individuos no susceptibles, y aquellos que abandonen la universidad tendrán asignado un tiempo menor en la cual el efecto de susceptibilidad fue llevado a cabo, determinando a este evento como *abandono*.

El esquema en el que se trabajará trabajado es construido bajo la presencia de los datos previamente descritos, covariables y tiempos censurados, siendo este último el que genera un mecanismo de censura al modelo.

El objetivo general de la tesis es estimar el tiempo a la ocurrencia de un evento de interés permitiendo construir un modelo de mixtura con datos censurados observándose en análisis previos el efecto de esta característica sobre la estimación. Finalmente, se discutirán los resultados y puntos de mayor importancia, así como la aplicación del modelo a un conjunto de datos reales.

Detallando lo incluido en la presente tesis, tenemos que en el capítulo 2, se presentan conceptos generales que serán la base principal para el desarrollo del modelo, llevando a estructurar un modelo de supervivencia usual, extendido bajo ciertas mecanismos de censura y efectos de las covariables; en el capítulo 3, se estructura un modelo de mixtura de fracción de cura con covariables y datos censurados; en el capítulo 4, se presenta el estudio de simulación bajo diferentes escenarios que permitirán mostrar el desempeño del modelo propuesto; en el capítulo 5, se llevará a cabo la aplicación sobre un conjunto de datos reales; finalmente en el capítulo 6, se presentan las conclusiones respectivas.

Capítulo 2

Modelos de supervivencia

Inicialmente, se asume una población compuesta solo por individuos susceptibles, aquellos quienes lograrán alcanzar el evento de interés en un tiempo determinado, particularmente discreto. En un primer paso, no se tomará en cuenta el efecto de las covariables, luego se ampliará el concepto bajo dichos efectos y la inserción de un mecanismo de censura.

2.1. Función de densidad de probabilidad y Supervivencia

Siendo T_0 una variable aleatoria discreta, positiva, que mide el tiempo a la ocurrencia de un evento de interés, denominada como *tiempo de falla*, se define su función de probabilidad dada por

$$f_0(t) \quad = \quad P(T_0 = t), \;\; t = 1, 2, \ldots \tag{2.1}$$

El uso de la función de distribución acumulada $F_0(t) = P(T_0 \leq t)$ no se vuelve frecuente en el área de supervivencia, sino la *función de supervivencia*, definida como la probabilidad de que un individuo presente un tiempo de falla después de un tiempo t y está denotada por

$$S_0(t) \quad = \quad 1 - F(t) \quad = \quad P(T_0 > t) \quad , \quad t = 1, 2, \ldots$$

Es claro que dado el caso discreto, tenemos que $S_0(t-1) = P(T_0 > t-1) = P(T_0 \geq t)$. Sin embargo, denotamos $S_0(t-) = S_0(t-1)$, manteniendo una notación generalizada para el caso continuo y discreto.

Observemos que $S_0(1-) = S_0(0) = P(T_0 > 0) = 1$ debido a que implícitamente se asume que ningún individuo está experimentando el evento del interés al inicio del estudio.

2.1.1. Tasas de riesgo basal

Una función de riesgo basal, o *tasa de riesgo basal* en t, denotada por $\lambda_0(t)$, es definida como la probabilidad condicional del tiempo de falla de t, en f_0, dado que el individuo ha sobrevivido por lo menos hasta t, esto es

$$\lambda_0(t) \quad = \quad P(T_0 = t \mid T \geq t) \quad = \quad \frac{f_0(t)}{S_0(t-)}, \; t = 0, 1, 2, \ldots \tag{2.2}$$

La expresión mostrada en (2.2), nos permite deducir que si un individuo ha sobrevivido por lo menos hasta t, la probabilidad de que su tiempo de falla sea mayor a t es de $(1 - \lambda_0(t))$, pudiéndose generar un desarrollo secuencial en cada tiempo de falla previo para redefinir

$S(t)$. Matemáticamente, esto sería

$$\begin{aligned} S_0(t) &= P(T_0 > t) \\ &= P(T_0 > 1 \cap T_0 > 2 \cap \ldots \cap T_0 > t-1 \cap T_0 > t) \\ &= P(T_0 > 1)P(T_0 > 2 \mid T_0 > 1) \ldots P(T_0 > t \mid T_0 > t-1) \\ &= P(T_0 > 1 \mid T_0 \geq 1)P(T_0 > 2 \mid T_0 \geq 2) \ldots P(T_0 > t \mid T_0 \geq t) \\ &= (1 - P(T_0 = 1 \mid T_0 \geq 1))\,(1 - P(T_0 = 2 \mid T_0 \geq 2)) \ldots (1 - P(T_0 = t \mid T_0 \geq t)) \\ &= \prod_{j=1}^{t} (1 - \lambda_0(j))\,,\ t = 1, 2, \ldots \end{aligned} \tag{2.3}$$

Además, $f_0(t)$, partiendo de (2.2), puede también redefinirse de la forma

$$\begin{aligned} f_0(t) &= \lambda_0(t) S_0(t-) \\ &= \lambda_0(t) \prod_{j=1}^{t-1} (1 - \lambda_0(j)),\ t = 1, 2, \ldots \end{aligned} \tag{2.4}$$

Pudiéndose interpretar de que las tasas de riesgo regidas bajo T_0, logran determinar su distribución correspondiente, en una relación biunívoca.

2.2. Categorías de Censura

A causa de la alta complejidad tanto en la recolección de datos como en el estudio a realizarse, existen distintos mecanismos o categorías de censura, que de acuerdo al artículo de Zhao and Zhou [2008] uno de ellos es la censura aleatoria por la derecha, la cual será discutida en la presente tesis por ser de primordial relevancia en el modelamiento a tratar en el capítulos posteriores.

Otros tipos de censura son *censura por la izquierda* y *censura por intervalos*, las cuales son ampliamente discutidos en Klein and Moeschberger [1997].

Censura aleatoria por la derecha

Este tipo de censura se presenta por la inclusión de no solo la falta de ocurrencia del evento de interés al finalizar el estudio sino también de la pérdida del seguimiento del individuo, debido a que pudo decidir excluirse a sí mismo del estudio manteniéndose solo la información del tiempo al cual se le observó por última vez, o su exclusión del estudio por parte del investigador, debido a que el individuo experimentó algún proceso ajeno al evento de interés.

En la Figura 2.1, el tiempo límite de estudio es de 5 periodos medidos en forma discreta en donde se visualizan los seguimientos para cuatro individuos: El individuo 1 no es censurado y presenta un tiempo de falla igual a 2; el individuo 2 tiene un tiempo de falla igual a 4 sin embargo este hecho no fue observado pues fue censurado en un tiempo igual a 3; el individuo 3 presenta un tiempo de falla igual al tiempo límite de estudio; y el individuo 4 no alcanzó el evento de interés al final del estudio.

Sea f_c la función de densidad de la variable aleatoria discreta T_c, representando el tiempo de censura asociado a T_0, variable cuya función de densidad viene definida en (2.1) y suponiendo una censura no informativa, es decir, que T_c y T_0 son independientes, lo que se

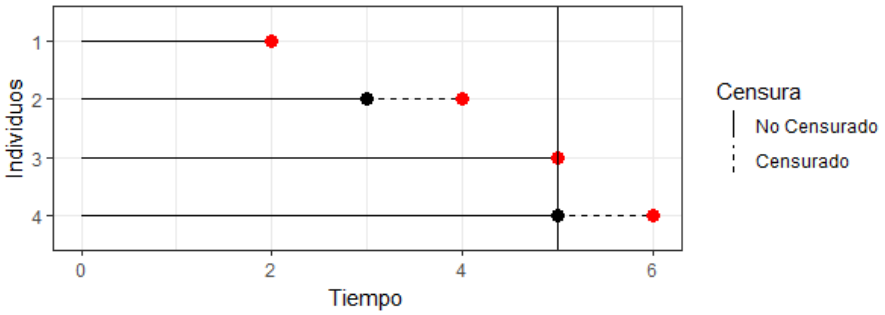

Figura 2.1: Censura aleatoria por la derecha.

observa es el par aleatorio (T, Δ) definido por

$$T \;=\; \text{mín}(T_0, T_c) \quad , \quad \Delta \;=\; \mathbf{1}(T_0 \leq T_c).$$

donde $\mathbf{1}(\cdot)$ devuelve 1 si se evalua sobre una sentencia verdadera y 0 en caso contrario.

El enfoque continuo de la distribución de este par (T, Δ) puede observarse en Matadamas [2010]; sin embargo, bajo este escenario, la distribución puede determinarse de forma análoga y menos compleja, para una observación censurada (con $\Delta = 0$) como:

$$\begin{aligned} P(T = t, \Delta = 0) &= P(T_c = t, T_c < T_0) \\ &= P(t < T_0) P(T_c = t) \\ &= S_0(t) f_c(t), \end{aligned}$$

y para una observación no censurada (con $\Delta = 1$):

$$\begin{aligned} P(T = t, \Delta = 1) &= P(T_0 = t, T_0 \leq T_c) \\ &= P(t \leq T_c) P(T_0 = t) \\ &= S_c(t-) f_0(t). \end{aligned}$$

La función de densidad de probabilidad para un par observado (t_i, Δ_i) tendría la forma

$$P(T = t_i, \Delta = \Delta_i) \;=\; [S_c(t_i-) f_0(t_i)]^{\Delta_i} \, [S_0(t_i) f_c(t_i)]^{1-\Delta_i}$$

Tomándose una muestra de n individuos, se define la verosimilitud para las tasas de riesgo basales $\lambda_0(t)$, solo tomando las contribuciones de $f_0(t)$, $S_0(t)$ y los indicadores Δ, a causa de la independencia por parte de la variable T_c

$$\begin{aligned} \mathcal{L}(\lambda) &\propto \prod_{i=1}^{n} [f_0(t_i)]^{\Delta_i} \, [S_0(t_i)]^{1-\Delta_i} \\ &= \prod_{i=1}^{n} \left[\lambda_0(t_i) \prod_{j=1}^{t_i - 1} (1 - \lambda_0(j)) \right]^{\Delta_i} \left[\prod_{j=1}^{t_i} (1 - \lambda_0(j)) \right]^{1-\Delta_i}. \end{aligned} \tag{2.5}$$

Suponiendo la contribución de un vector de covariables $\mathbf{x_i} = \{x_{1i}, \ldots, x_{mi}\}$ por cada individuo i sobre las tasas de riesgo, representadas por $\lambda(t \mid \mathbf{x})$, se pueden formular modelos de regresión que permitan ajustes adecuados además de una no tan compleja interpretación de estas.

El modelo de supervivencia (2.5) bajo el efecto de estas covariables, se redefiniría como

$$\mathcal{L}(\lambda) \quad \propto \quad \prod_{i=1}^{n} \left[\lambda(t_i \mid \mathbf{x}_i) \prod_{j=1}^{t_i - 1} (1 - \lambda(j \mid \mathbf{x}_i)) \right]^{\Delta_i} \left[\prod_{j=1}^{t_i} (1 - \lambda(j \mid \mathbf{x}_i)) \right]^{1-\Delta_i}.$$

2.3. Modelo de Odds proporcionales

Cox [1972] siguiendo ciertos ajustes de flexibilidad propuso un modelo de estructura de riesgos proporcionales definida por

$$\lambda(t \mid \mathbf{x}) \quad = \quad \lambda_0(t) \exp\left(\mathbf{x}^T \beta\right), \tag{2.6}$$

donde $\beta^T = (\beta_1, \ldots, \beta_m)$ es un vector de parámetros que representan los efectos de las covariables, $\lambda_0(t)$ las tasas de riesgo basales y $\lambda(t \mid \mathbf{x})$ las tasas de riesgos bajo covariables. Con el objetivo se eviten restricciones en las tasas de riesgo por parte del modelo de regresión (2.6), pues necesariamente se tendría que tener que $\lambda_0(t) \exp\left(\mathbf{x}^T \beta\right) \in (0, 1]$, Cox [1972] sugiere el uso de los odds de riesgo, definidos por

$$\text{odds}(t \mid \mathbf{x}) \quad = \quad \frac{\lambda(t \mid \mathbf{x})}{1 - \lambda(t \mid \mathbf{x})}.$$

los cuales miden la proporcionalidad del riesgo a ser susceptible en un tiempo específico con respecto a no serlo, y propone el siguiente modelo lineal logístico

$$\frac{\lambda(t \mid \mathbf{x})}{1 - \lambda(t \mid \mathbf{x})} \quad = \quad \frac{\lambda_0(t)}{1 - \lambda_0(t)} \exp\left(\mathbf{x}^T \beta\right). \tag{2.7}$$

equivalente a

$$\text{logit}(\lambda(t \mid \mathbf{x})) \quad = \quad \text{logit}(\lambda_0(t)) + \mathbf{x}^T \beta. \tag{2.8}$$

permitiendo que $\lambda(t \mid \mathbf{x}) \in (0, 1)$ como corresponde. Condicionamos el caso en el que se cuente con un límite de estudio, es decir un tiempo final t_f, entonces la tasas de riesgo en t_f deberán cumplir con que $\lambda(t_f \mid \mathbf{x}) = \lambda_0(t_f) = 1$.

La interpretación del modelo de regresión logístico, se da bajo la definición de un radio de odds OR que mide tanto el incremento, estabilidad o descenso de los odds de riesgo. Sean dos individuos i y j, con covariables $\mathbf{x}_i$ y $\mathbf{x}_j$ respectivamente, entonces el OR en un tiempo t, sería

$$\text{OR} \quad = \quad \frac{\text{odds}(t \mid \mathbf{x}_i)}{\text{odds}(t \mid \mathbf{x}_j)} \quad = \quad \exp\left((\mathbf{x}_i - \mathbf{x}_j)^T \beta\right). \tag{2.9}$$

Como logra observarse este radio de riesgo solo depende de las covariables medidas,

independiente del tiempo, interpretando este radio como sigue a continuación:

- Si OR > 1: El odds de riesgo incrementa en una observación con covariables $\mathbf{x}_i$ respecto a otra con covariables $\mathbf{x}_j$.
- Si OR $= 1$: El odds de riesgo se mantiene estable en una observación con covariables $\mathbf{x}_i$ respecto a otra con covariables $\mathbf{x}_j$.
- Si OR < 1: El odds de riesgo disminuye en una observación con covariables $\mathbf{x}_i$ respecto a otra con covariables $\mathbf{x}_j$.

La función de supervivencia y función de probabilidad, independientemente del modelo de regresión en uso, bajo el efecto de las covariables y una estructura análoga de cálculo como en (2.3) vendrían dadas por

$$\begin{aligned} S(t \mid \mathbf{x}) &= \prod_{j=1}^{t}(1-\lambda(j \mid \mathbf{x})), \\ f(t \mid \mathbf{x}) &= \lambda(t \mid \mathbf{x})\prod_{j=1}^{t-1}(1-\lambda(j \mid \mathbf{x})). \end{aligned} \tag{2.10}$$

Sin embargo, bajo el modelo propuesto (2.7), se tiene que

$$\lambda(t \mid \mathbf{x}) = \frac{1}{1+\dfrac{1-\lambda_0(t)}{\lambda_0(t)}\exp(-\mathbf{x}^T\beta)}.$$

Capítulo 3

Modelo de fracción de cura

En este capítulo, se amplia el concepto del modelo de supervivencia visto anteriormente, planteándose el estudio sobre una población con dos tipos de individuos, *susceptibles* y *curados* (no susceptibles), donde los segundos se caracterizan por nunca alcanzar el evento de interés pudiéndose indicar, para efectos prácticos, que lo alcanzan solo en un tiempo infinito, además de tener en cuenta la característica particular de censura sobre algunos individuos.

Fijando K como el tiempo límite de estudio sobre los individuos, se define la variable aleatoria discreta T_0 denominada como *tiempo de falla para los individuos susceptibles* con función de densidad f_0, y de igual forma $\boldsymbol{\infty}$ como *tiempo de falla para los individuos curados.*

Asumimos inicialmente para efectos prácticos, que no hay influencia de las covariables en el modelo y no existe censura. Zhao and Zhou [2008] definen una variable aleatoria que representa la proporción de individuos susceptibles; sin embargo, en la presente tesis se define una variable aleatoria η que representa la proporción de individuos curados de la forma $P(\eta = 1) = p$, donde $p \in (0, 1]$, y por consiguiente la variable T_*, definida como *tiempo de falla*, representada por una estructura mixta vendría dada de la siguiente forma

$$T_* \quad = \quad (1 - \eta)T_0 + \eta\boldsymbol{\infty}. \tag{3.1}$$

donde $T_0 < \infty$, luego la distribución de T_* vendría dada por

$$f_*(t) \quad = \quad \begin{cases} (1-p)f_0(t) & ;\ t = 1, 2, \ldots, K. \\ p & ; t = \infty \end{cases} \tag{3.2}$$

La función de supervivencia de T_* en relación con $S_0(t)$, función de supervivencia para T_0, tendría la forma

$$S_*(t) \quad = \quad p + (1 - p)S_0(t). \tag{3.3}$$

El modelo bajo esta variable T_*, que se verá posteriormente, se conoce como modelo de fracción de cura.

3.1. Modelo de Mixtura odds de riesgos proporcionales

Asumiendo que las tasas de riesgo se ven influenciadas por un vector de covariables $\mathbf{x}$, se pueden formular distintos modelos de regresión; sin embargo el objetivo primordial será determinar un modelo de fracción de cura que permita interpretaciones sobre estructuras proporcionales sobre las tasas de riesgo u odds de riesgo.

Zhao and Zhou [2008] siguiendo el modelo de riesgos proporcionales propuesto en (2.6), construyen un modelo análogo para el caso en el que se tiene una mixtura, de la siguiente forma

$$\begin{aligned}\lambda_*(t \mid \mathbf{x}) &= \lambda_*(t)\exp(\beta^T\mathbf{x}) \\ &= \frac{f_*(t)}{S_*(t-)}\exp(\beta^T\mathbf{x}) \\ &= \frac{(1-p)f_0(t)}{p+(1-p)S_0(t-)}\exp(\beta^T\mathbf{x}) \end{aligned} \tag{3.4}$$

donde el flujo en el que se descompone la ecuación es debido a (3.2) y (3.3); siendo λ_* la tasa de riesgo basal para T_*.

En esta sección se propone un modelo basado en (3.4), pudiéndose hacer uso del modelo logístico (2.8) sobre la población heterogénea con el fin mantener una estructura con odds de riesgos proporcionales en la variable mixta T_*

$$\text{logit}(\lambda_*(t \mid \mathbf{x})) = \text{logit}(\lambda_*(t)) + \mathbf{x}^T\beta, \tag{3.5}$$

con

$$\lambda_*(t) = \frac{f_*(t)}{S_*(t-)} = \frac{(1-p)f_0(t)}{p+(1-p)S_0(t-)} = \frac{(1-p)\lambda_0(t)\prod_{j=1}^{t-1}(1-\lambda_0(t))}{p+(1-p)\prod_{j=1}^{t-1}(1-\lambda_0(t))} \tag{3.6}$$

Luego la función de supervivencia y probabilidad, de la variable mixtura T_* bajo el efecto de las covariables, vendrían dadas, respectivamente, por

$$S_*(t \mid \mathbf{x}) = \prod_{j=1}^{t}(1-\lambda_*(j \mid \mathbf{x})), \tag{3.7}$$

$$f_*(t \mid \mathbf{x}) = \lambda_*(t \mid \mathbf{x})\prod_{j=1}^{t-1}(1-\lambda_*(j \mid \mathbf{x})), \tag{3.8}$$

donde

$$\lambda_*(t \mid \mathbf{x}) = \frac{1}{1+\dfrac{1-\lambda_*(t)}{\lambda_*(t)}\exp(-\mathbf{x}^T\beta)}. \tag{3.9}$$

3.2. Proporción de individuos curados

Si bien trabajos realizados como el de Kalbfleisch and Prentice [2002], realizan modelos sin usar la proporción de individuos curados bajo el efecto de las covariables, en esta tesis analizaremos las representaciones de ellas, con el objetivo sean incluidas posteriormente en la estimación en caso sea necesario.

Se determina que la proporción de individuos curados tendría la forma

$$\begin{aligned} p(\mathbf{x}) &= P(T_* = \infty) \\ &= P(T_* > K) \\ &= S_*(K \mid \mathbf{x}) \\ &= \prod_{j=1}^{K} (1 - \lambda_*(j \mid \mathbf{x})) \end{aligned} \tag{3.10}$$

donde K representa el tiempo de falla límite para un individuo susceptible y $\lambda_*(t \mid \mathbf{x})$ viene representada por (3.9). Se tendrá que tener claro que $\lambda_*(t \mid \mathbf{x})$ es la tasa de riesgo para la variable T_* provista de covariables, por lo que dichas tasas se extienden no solo hasta K sino hasta ∞. Esto, da a entender que $S_*(K \mid \mathbf{x}) > S_*(\infty \mid \mathbf{x}) = 0$, y he ahí el hecho de que (3.10) está bien definido. Además, se tendrá que tener en cuenta que $p(\mathbf{x})$ difiere de p, por el hecho de que este último representa la proporción de individuos curados sin efecto de las covariables .

3.3. Modelo fracción de cura con tiempos censurados

Siguiendo el trabajo realizado por Zhao and Zhou [2008], el esquema a tratar será el de una censura aleatoria por la derecha, visualizada en Figura 2.1, por lo que se asume que se desconoce cuando se presentará la censura pues ésta puede variar de un individuo a otro. Por tanto, definimos una variable aleatoria discreta T_c, con función de probabilidad f_c y función de supervivencia $S_c(t)$, que denote el tiempo de censura, que posteriormente se asociará a T_*, suponiéndose independencia de T_c con T_* y con las covariables $\mathbf{x}$.

El modelo de censura, lo que observa es el par aleatorio (T, Δ) definido por

$$T = min(T_*, T_c) \quad , \quad \Delta = \mathbf{1}(T_* \leq T_c)$$

donde

$$T_* = (1 - \eta)T_s + \eta\infty$$

siendo T_s y η variables aleatorias que representan el tiempo de susceptibilidad y la fracción de cura, ambos bajo el efecto de las covariables. Cabe indicar, que dado este caso, T_* difiere del determinado en (3.1) por el hecho de presentar efecto de covariables.

Dado que el enfoque del modelo final incorpora a los individuos curados, entonces se analizarán los tiempos observados en T para 3 tipos de individuos, siendo estos los (i) individuos curados (censurados pero bajo el tiempo límite de estudio como lo es el individuo 4 en Figura 2.1), (ii) individuos susceptibles (como lo son el individuo 1 e individuo 3 en Figura 2.1) e

(iii) individuos censurados (con tiempo de censura por debajo del límite de estudio como lo es el individuo 2 en Figura 2.1).

Por tanto, asumiendo un tiempo límite de estudio K definiendo la cota superior en el dominio de la variable T el cual representa el tiempo de falla para los individuos susceptibles, entonces la distribución para una observación $(T_i = t_i, \Delta_i)$ sería:

Si $T_i = t_i, \Delta_i = 0$ con $t_i > K$ entonces

$$\begin{aligned}
P(T_i = t_i, \Delta_i = 0) &= P(T_i = t_i, \Delta_i = 0 \mid \eta = 0)P(\eta = 0) + P(T_i = t_i, \Delta_i = 0 \mid \eta = 1)P(\eta = 1) \\
&= P(T_{c,i} = t_i, T_{c,i} < T_{s,i})P(\eta = 0) + P(T_{c,i} = t_i, T_{c,i} < \boldsymbol{\infty})P(\eta = 1) \\
&= \underbrace{P(t_i < T_{s,i})}_{=0} P(T_{c,i} = t_i)(1 - p(\mathbf{x})) + \underbrace{P(t_i < \boldsymbol{\infty})}_{=1} P(T_{c,i} = t_i)p(\mathbf{x}) \\
&= f_c(t_i)p(\mathbf{x}).
\end{aligned}$$

Si $T_i = t_i, \Delta_i = 0$ con $t_i \leq K$ entonces

$$\begin{aligned}
P(T_i = t_i, \Delta_i = 0) &= P(T_i = t_i, \Delta_i = 0 \mid \eta = 0)P(\eta = 0) + P(T_i = t_i, \Delta_i = 0 \mid \eta = 1)P(\eta = 1) \\
&= P(T_{c,i} = t_i, T_{c,i} < T_{s,i})P(\eta = 0) + P(T_{c,i} = t_i, T_{c,i} < \boldsymbol{\infty})P(\eta = 1) \\
&= P(t_i < T_{s,i})P(T_{c,i} = t_i)(1 - p(\mathbf{x})) + P(t_i < \boldsymbol{\infty})P(T_{c,i} = t_i)p(\mathbf{x}) \\
&= S_{s,i}(t_i \mid \mathbf{x})f_c(t_i)(1 - p(\mathbf{x})) + f_c(t_i)p(\mathbf{x}) \\
&= f_{c,i}(t_i)\left((1 - p(\mathbf{x}))S_{s,i}(t_i \mid \mathbf{x}) + p(\mathbf{x})\right) \\
&= f_{c,i}(t_i)S_*(t_i \mid \mathbf{x}).
\end{aligned}$$

Si $T_i = t_i, \Delta_i = 1$ con $t_i \leq K$ entonces

$$\begin{aligned}
P(T_i = t_i, \Delta_i = 1) &= P(T_i = t_i, \Delta_i = 1 \mid \eta = 0)P(\eta = 0) + P(T_i = t_i, \Delta_i = 1 \mid \eta = 1)P(\eta = 1) \\
&= P(T_{s,i} = t, T_{s,i} \leq T_{c,i})P(\eta = 0) + \underbrace{P(\boldsymbol{\infty} = t_i, \boldsymbol{\infty} \leq T_{c,i})P(\eta = 1)}_{=0} \\
&= P(T_{s,i} = t, T_{s,i} \leq T_{c,i})(1 - p(\mathbf{x})) \\
&= P(t_i \leq T_{c,i})P(T_{s,i} = t_i)(1 - p(\mathbf{x})) \\
&= S_{c,i}(t_i-)f_{s,i}(t_i \mid \mathbf{x})(1 - p(\mathbf{x})) \\
&= S_{c,i}(t_i-)f_*(t_i \mid \mathbf{x}).
\end{aligned}$$

Por lo que, la función de probabilidad para los pares observados (T_i, Δ_i) vendría definida por

$$f(t_i) = P(t_i, \Delta_i) = \begin{cases} \left[S_c(t_i-)f_*(t_i \mid \mathbf{x})\right]^{\Delta_i}\left[f_c(t_i)S_*(t_i \mid \mathbf{x})\right]^{1-\Delta_i} & , \ t_i = 1, \ldots, K \ ; \ \Delta_i = 0, 1. \\ f_c(t_i)p(\mathbf{x}) & , \ t_i = K+1, \ldots \ ; \ \Delta_i = 0. \end{cases} \tag{3.11}$$

3.4. Función de verosimilitud

Considerando nuestro modelo de cura con datos de censura, para n_1 individuos curados y $n - n_1$ individuos susceptibles y censurados, observados de la forma $(t_i, \Delta_i, \mathbf{x}_i)$, denotando las tasas de riesgo basales por $\lambda_{t0} = \lambda_0(t)$, y asumiendo que $\mathbf{x}$ representa un vector de M covariables, la función de verosimilitud de $\theta = (p, \lambda_{10}, \lambda_{20}, \ldots, \lambda_{(K-1)0}, \beta^T)$ es dada por:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n_1} f_c(t_i) p(\mathbf{x}) \prod_{i=n_1+1}^{n} \left[S_c(t_i-) f_*(t_i \mid \mathbf{x})\right]^{\Delta_i} \left[f_c(t_i) S_*(t_i \mid \mathbf{x})\right]^{1,-\Delta_i}$$

donde $S_*(t \mid \mathbf{x})$ y $f_*(t \mid \mathbf{x})$ son las funciones de supervivencia y probabilidad definidas en (3.7) y (3.8), respectivamente.

Observándose que no se estima la tasa de riesgo basal en el punto K dado que por definición es $\lambda_{K0} = 1$.

Por otro lado, el efecto provocado por la censura es importante en la inferencia del modelo; sin embargo, es un efecto intrínseco dado que estas no dependen de los parámetros que necesitamos estimar por lo que nuestro modelo opta por la representación siguiente

$$\begin{aligned}
\mathcal{L}(\theta) &\propto \prod_{i=1}^{n_1} p(\mathbf{x}_i) \prod_{i=n_1+1}^{n} \left[f_*(t_i \mid \mathbf{x})\right]^{\Delta_i} \left[S_*(t_i \mid \mathbf{x})\right]^{1-\Delta_i} \\
&\propto \prod_{i=1}^{n_1} p(\mathbf{x}_i) \prod_{i=n_1+1}^{n} \left[\lambda_*(t_i \mid \mathbf{x}) \prod_{j=1}^{t_i-1} (1 - \lambda_*(j \mid \mathbf{x}))\right]^{\Delta_i} \left[\prod_{j=1}^{t_i} (1 - \lambda_*(j \mid \mathbf{x}))\right]^{1-\Delta_i} \\
&\propto \prod_{i=1}^{n_1} \left[\prod_{j=1}^{K} (1 - \lambda_*(j \mid \mathbf{x}))\right] \prod_{i=n_1+1}^{n} \left(\left[\frac{\lambda_*(t_i \mid \mathbf{x})}{1 - \lambda_*(t_i \mid \mathbf{x})}\right]^{\Delta_i} \prod_{j=1}^{t_i} (1 - \lambda_*(j \mid \mathbf{x}))\right). \quad (3.12)
\end{aligned}$$

3.5. Estimación del Modelo

Siendo lo propuesto un modelo paramétrico, la estimación de sus parámetros se llevarán a cabo mediante el método de máxima verosimilitud (EMV), la cual consiste en la determinación de los parámetros óptimos del modelo maximizando la función de verosimilitud definida en (3.12), o el logaritmo de dicha función, por efectos prácticos.

Siguiendo dicha representación, se formula el logaritmo de la función de verosimilitud $\log(\mathcal{L}(\theta))$, bajo la siguiente estructura

$$\log(\mathcal{L}(\theta)) = \sum_{i=1}^{n_1} \log(\mathcal{L}_{1,i}) + \sum_{i=n_1+1}^{n} \log(\mathcal{L}_{2,i}), \quad (3.13)$$

donde

$$\begin{aligned}
\log(\mathcal{L}_{1,i}) &= \sum_{j=1}^{K} \log\left(1 - \lambda_*(j \mid \mathbf{x}_i)\right), \\
\log(\mathcal{L}_{2,i}) &= \Delta_i \left[\log\left(\lambda_*(t_i \mid \mathbf{x}_i)\right) - \log\left(1 - \lambda_*(t_i \mid \mathbf{x}_i)\right)\right] + \sum_{j=1}^{t_i} \log\left(1 - \lambda_*(j \mid \mathbf{x}_i)\right).
\end{aligned}$$

La estimación de los parámetros se obtienen igualando a cero las primeras derivadas de la función (3.13), también denominadas funciones de score, siendo las siguientes:

$$\begin{aligned}
\frac{\partial \log(\mathcal{L}(\theta))}{\partial p} &= \sum_{i=1}^{n_1}\sum_{j=1}^{K} \frac{1}{1-\lambda_*(j \mid \mathbf{x}_i)} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \lambda_*(j)} \cdot \frac{\partial \lambda_*(j)}{\partial p} \\
&+ \sum_{i=n_1+1}^{n} \frac{\Delta_i}{\lambda_*(t_i \mid \mathbf{x}_i)(1-\lambda_*(t_i \mid \mathbf{x}_i))} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \lambda_*(j)} \cdot \frac{\partial \lambda_*(j)}{\partial p} \\
&\sum_{i=n_1+1}^{n}\sum_{j=1}^{t_i} \frac{1}{1-\lambda_*(j \mid \mathbf{x}_i)} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \lambda_*(j)} \cdot \frac{\partial \lambda_*(j)}{\partial p}. \\
\frac{\partial \log(\mathcal{L}(\theta))}{\partial \lambda_{t0}} &= \sum_{i=1}^{n_1}\sum_{j=1}^{K} \frac{1}{1-\lambda_*(j \mid \mathbf{x}_i)} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \lambda_*(j)} \cdot \frac{\partial \lambda_*(j)}{\partial \lambda_{t0}} \\
&+ \sum_{i=n_1+1}^{n} \frac{\Delta_i}{\lambda_*(t_i \mid \mathbf{x}_i)(1-\lambda_*(t_i \mid \mathbf{x}_i))} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \lambda_*(j)} \cdot \frac{\partial \lambda_*(j)}{\partial \lambda_{t0}} \\
&\sum_{i=n_1+1}^{n}\sum_{j=1}^{t_i} \frac{1}{1-\lambda_*(j \mid \mathbf{x}_i)} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \lambda_*(j)} \cdot \frac{\partial \lambda_*(j)}{\partial \lambda_{t0}}. \\
\frac{\partial \log(\mathcal{L}(\theta))}{\partial \beta_m} &= \sum_{i=1}^{n_1}\sum_{j=1}^{K} \frac{1}{1-\lambda_*(j \mid \mathbf{x}_i)} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \beta_m} \\
&+ \sum_{i=n_1+1}^{n} \frac{\Delta_i}{\lambda_*(t_i \mid \mathbf{x}_i)(1-\lambda_*(t_i \mid \mathbf{x}_i))} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \beta_m} \\
&\sum_{i=n_1+1}^{n}\sum_{j=1}^{t_i} \frac{1}{1-\lambda_*(j \mid \mathbf{x}_i)} \cdot \frac{\partial \lambda_*(j \mid \mathbf{x}_i)}{\partial \beta_m}.
\end{aligned}$$

para $t = 1, \ldots, K$ y $m = 1, \ldots, M$.

Así mismo las segundas derivadas $\dfrac{\partial^2 \log(\mathcal{L}(\theta))}{\partial \theta_i \partial \theta_j}$ permiten definir la siguiente matriz denominada matriz de Información

$$I(\theta) = -E\left[\frac{\partial^2 \log(\mathcal{L}(\theta))}{\partial \theta_i \partial \theta_j}\right] \tag{3.14}$$

cuya inversa, evaluada sobre la estimación de los parámetros, sea $\hat{\theta}$, corresponde a la matriz de covarianza de la distribución asintótica para θ

$$\sqrt{n}(\theta - \hat{\theta}) \sim N(0, I^{-1}(\hat{\theta})). \tag{3.15}$$

Capítulo 4

Estudio de Simulación

En esta sección se evaluará el desempeño del modelo de fracción de cura estructurado en (3.12) a través de un estudio de simulación considerando distintos escenarios de censura,que se verá posteriormente.

Los criterios que se tomarán en cuenta para el desempeño de un intervalo de confianza (IC) son:

- **Cobertura**: Se define el estimador de Monte Carlo de la cobertura como

$$\widehat{\text{Cobertura}} \quad = \quad \frac{\#(\theta \in [LI_i, LS_i])}{M}$$

  la cual mide el buen desempeño de que un intervalo de confianza contenga el verdadero valor del parámetro θ.

- **Amplitud**: Se define el estimador de Monte Carlo de la amplitud como

$$\widehat{\text{Amplitud}} \quad = \quad \overline{LS} - \overline{LI}$$

  la cual compara varios intervalos de confianza para un mismo parámetro, donde un IC con menor amplitud sería más apropiado.

Para el proceso de simulación se considera una covariable unidimensional x_i dividida en dos grupos ($x_i = 0$ para el grupo 1 y $x_i = 1$ para el grupo 2) y 5 tiempos de falla para individuos susceptibles (1,2,3,4,5) donde $K = 5$ representa el tiempo límite de estudio. Los valores teóricos de los parámetros son tomados como

$$(\lambda_{10}, \lambda_{20}, \lambda_{30}, \lambda_{40}, p, \beta) \quad = \quad (0.2, 0.375, 0.3, 0.7143, 0.3, 0.375)$$

Los escenarios que generan las función de probabilidad (3.11) para un modelo de fracción de cura con datos censurados, son los siguientes:

- Escenario I: Los puntos de censura son $(3, 4, 5, 6, 7)$ con probabilidad $(0.2, 0.2, 0.2, 0.2, 0.2)$, la cual genera una tasa de censura de aproximadamente 30 %.
- Escenario II: Los puntos de censura son $(1, 2, 3, 4, 5)$ con probabilidad $(0.2, 0.2, 0.2, 0.2, 0.2)$, la cual genera una tasa de censura de aproximadamente 47 %.
- Escenario III: Los puntos de censura son $(1, 2, 3, 4)$ con probabilidad $(0.6, 0.1, 0.1, 0.2)$, la cual genera una tasa de censura de aproximadamente 66 %.

Escenario I		λ_{10}	λ_{20}	λ_{30}	λ_{40}	p	β
n=100	Media	0.201	0.372	0.299	0.723	0.299	0.388
	Amplitud	0.187	0.260	0.307	0.441	0.262	1.117
	Cobertura	0.957	0.970	0.959	0.977	0.951	0.953
n=300	Media	0.199	0.376	0.300	0.712	0.299	0.377
	Amplitud	0.108	0.152	0.181	0.256	0.154	0.636
	Cobertura	0.954	0.961	0.953	0.956	0.947	0.947
n=500	Media	0.200	0.376	0.301	0.715	0.300	0.378
	Amplitud	0.084	0.118	0.141	0.200	0.120	0.492
	Cobertura	0.941	0.952	0.951	0.946	0.964	0.947
Escenario II		λ_{10}	λ_{20}	λ_{30}	λ_{40}	p	β
n=100	Media	0.201	0.376	0.309	0.738	0.300	0.392
	Amplitud	0.202	0.317	0.321	0.724	0.251	1.316
	Cobertura	0.952	0.950	0.934	0.969	0.934	0.943
n=300	Media	0.199	0.376	0.301	0.717	0.298	0.376
	Amplitud	0.116	0.187	0.245	0.413	0.208	0.743
	Cobertura	0.956	0.940	0.958	0.963	0.952	0.946
n=500	Media	0.201	0.378	0.300	0.718	0.299	0.376
	Amplitud	0.090	0.146	0.191	0.320	0.163	0.573
	Cobertura	0.958	0.951	0.936	0.967	0.955	0.947
Escenario III		λ_{10}	λ_{20}	λ_{30}	λ_{40}	p	β
n=100	Media	0.202	0.388	0.315	0.693	0.299	0.395
	Amplitud	0.222	0.410	0.330	0.822	0.290	1.332
	Cobertura	0.950	0.978	0.976	0.970	0.976	0.940
n=300	Media	0.197	0.371	0.297	0.697	0.292	0.383
	Amplitud	0.121	0.191	0.310	0.510	0.220	0.636
	Cobertura	0.978	0.970	0.960	0.986	0.966	0.950
n=500	Media	0.198	0.372	0.299	0.694	0.294	0.376
	Amplitud	0.091	0.151	0.189	0.327	0.180	0.492
	Cobertura	0.962	0.952	0.972	0.985	0.974	0.934

Cuadro 4.1: Parámetros estimados bajo el modelo (3.12).

Se realizarán 1000 simulaciones con tamaños de muestra $n = 100, 300$ y 500, sobre cada escenario que a su vez mantendrán cantidades aproximadamente iguales para cada grupo ($n_1 = n_2$), pudiendo esto generarse mediante una distribución binomial con probabilidad de 0.5. El proceso de simulación para un tamaño de muestra n se logra resumir de la siguiente forma:

1. Se genera un vector de covariables $\mathbf{x}$ de longitud n mediante $X \sim Binomial(n, 0.5)$.

2. Se determinan, $\lambda_*(t)$ para $t = 1, \ldots, 5$, las cuales son las tasas de riesgo basales de la variable mixtura T_* en ausencia de covariables, estructurados en (3.6).

3. Para cada $t = 1, \ldots, 5$, se determinan, $\lambda_*(t \mid \mathbf{x})$, mediante (3.9), las cuales son las tasas de riesgo de la variable mixtura T_* en presencia de covariables y posteriormente la función de probabilidad $f_*(t \mid \mathbf{x})$, mediante (3.8). Se sabe que $f_*(\infty \mid \mathbf{x})$ se encuentra definido determinando la probabilidad de individuos curados influenciados por $\mathbf{x}$, por lo que $f_*(\infty \mid \mathbf{x}) = 1 - \sum_{j=1}^{5} f_*(j \mid \mathbf{x})$.

4. Sobre un individuo con covariable $\mathbf{x}$, se selecciona aleatoriamente un tiempo de fallo entre $1, 2, 3, 4, 5, \infty$ bajo las probabilidades determinadas por $f_*(t)$, siendo estos tiempos los representados por T_*.

5. Se escoge un escenario de impacto de la censura, y se generan n tiempos $t_{c,i}$ representando puntos de censura T_c.

6. Por cada individuo i, se calcula su tiempo observado t_i por $min(t_{*,i}, t_{c,i})$ y su indicador de censura $\Delta_i = \mathbf{1}(t_{*,i} \leq t_{c,i})$, obteniendo un conjunto de datos de la forma (T_i, Δ_i, x_i).

Los resultados de la simulación y las estimaciones bajo los distintos escenarios, se presentan en el Cuadro 4.1, donde se obtienen estimaciones cercanas a su valor teórico con coberturas estables de cerca del 95 % como correspondería por la consideración de intervalos de confianza del 95 % en cada simulación, y amplitudes que van decreciendo conforme aumenta el tamaño de muestra.

Capítulo 5

Aplicación

Los datos que han sido tomados para la aplicación del modelo, son de aquellos alumnos de la Pontificia Universidad Católica del Peru (PUCP), que han ingresado desde el año 2002 hasta el año 2012. Cabe mencionar que existen dos periodos de ingreso por año, denominados ciclos, siendo las cantidades de alumnos evaluados por periodo reflejados en el Cuadro 5.1, representando un total de 31665.

Ciclo	**Cantidad**	**Ciclo**	**Cantidad**
2002-1	1673	2007-2	623
2002-2	923	2008-1	2220
2003-1	1639	2008-2	701
2003-2	855	2009-1	2483
2004-1	1511	2009-2	764
2004-2	841	2010-1	2357
2005-1	1659	2010-2	968
2005-2	764	2011-1	2221
2006-1	1986	2011-2	845
2006-2	663	2012-1	2624
2007-1	2539	2012-2	806

Cuadro 5.1: Frecuencia de alumnos ingresantes por ciclo de ingreso.

El periodo de estudio para un alumno corresponderá a 4 ciclos desde su ingreso y el evento de interés al que puede llegar corresponde al hecho de abandonar la universidad. Este evento se llevará a cabo en un tiempo, denominado como tiempo de falla, el cual hará referencia a los ciclos que el alumno logró permanecer en la universidad hasta su abandono. Se definirán como individuos curados a aquellos alumnos que no abandonen la universidad el periodo de estudio. Si un alumno, no logra culminar el cuarto ciclo entonces se dirá que alcanzó el evento de interés de abandono, por tanto se le asigna un tiempo de falla de 1, 2 ,3 o 4 indicando la cantidad de ciclos que logró mantenerse hasta su abandono. En caso no se tenga la información completa para un alumno en el periodo de estudio y siendo el hecho de que bajo lo observado no alcanza el evento de interés, entonces éste tendrá un tiempo de censura dado por el último ciclo en el que pudo ser observado. Es decir, si un alumno tiene un tiempo de 2 y asu vez presenta la característica de censurado, es que solo se sabe que logró permanecer en al menos su primer ciclo en la universidad desconociéndose su información posterior. Cabe mencionar que se han excluido a aquellos alumnos que han abandonado la universidad sin haber culminado su primer ciclo pues estos corresponden a un 0.02 % de

los registros, además se han excluido a aquellos alumnos cuya variable "CRAEst" presente valores nulos en su primer ciclo de estudio, con objeto sea una variable incluida en el modelo siendo estos un 0.7 % de los registros totales.

Los datos evaluados son de 31665 registros, donde se presenta una tasa de censura de 20.40 %, siendo estructurados como en el Cuadro 5.2 donde la variable "**CENSURA**" representa un indicador tal que si es 0 entonces la variable "**TIEMPO DE FALLA**" representa un tiempo censurado, y si es 1 entonces representa el tiempo en el cual se alcanzó el evento de interés, la variable "**AREA**" considera las áreas de estudios a las cuales se puede ingresar, dividiéndose en 5 categorías: Ciencias; Letras; Arte; Arquitectura y Urbanismo; y, Educación. La variable "**CRAEst**" se define como el coeficiente de rendimiento académico estándar que un alumno ha obtenido, tomándose en particular el coeficiente calculado en el primer ciclo de estudio.

numsec	**CICLO**	**AREA**	**CRAEst**	**TIEMPO_FALLA**	**CENSURA**
30752	2012-1	CIENCIAS	48.12	1	0
30753	2012-1	CIENCIAS	34.21	2	0
30754	2012-1	LETRAS	55.63	2	0
10784	2006-2	LETRAS	15.81	1	1
10785	2006-2	LETRAS	57.61	4	1
10786	2007-2	LETRAS	46.80	3	1
10788	2006-2	LETRAS	48.30	4	1
10789	2006-2	LETRAS	46.57	4	1

Cuadro 5.2: Estructura de los datos

Las variables tomadas por el modelo serán las variables "**CRAEst**", estandarizada, y "**AREA**" estructurada como dummy, tomando como referencia a la categoría Educación, teniéndose un modelo de la forma

$$\begin{aligned}\text{logit}(\lambda_*(t \mid \mathbf{x})) &= \text{logit}(\lambda_*(t)) + \text{AREA\_LETRAS} * \beta_1 + \text{AREA\_CIENCIAS} * \beta_2 \\ &\quad + \text{AREA\_ARQUITECTURA\_Y\_URBANISMO} * \beta_3 \\ &\quad + \text{AREA\_ARTE} * \beta_4 + \frac{\text{CRAEst} - \mu_{CRAEst}}{\sigma_{CRAEst}} * \beta_5\end{aligned}$$

donde $\mu_{CRAEst} = 50.247$ y $\sigma_{CRAEst} = 8.142$ representan la media y desviación estándar de la variable "**CRAEst**"; los coeficientes de las variables del modelo vienen determinados en el cuadro 5.3, y las tasas de riesgos basales como la proporción de individuos curados en el cuadro 5.4

Parámetros	Estimación	Odds Ratio (OR)	OR (95 % IC)
β_1	0.584	1.793	[1.102,2.917]
β_2	0.744	2.105	[1.294 ,3.426]
β_3	-0.201	0.818	[0.470 ,1.423]
β_4	-0.946	0.388	[0.204 ,0.738]
β_5	-1.57	0.207	[0.199 ,0.217]

Cuadro 5.3: Coeficientes del Modelo de regresión logística para odds de riesgo proporcionales.

Parámetros	Estimación	(95 % IC)
λ_{01}	0.000	[0.000,0.001]
λ_{02}	0.170	[0.159,0.183]
λ_{03}	0.453	[0.430,0.476]
p	0.963	[0.941,0.977]

Cuadro 5.4: Tasas de riesgos basales(λ_{01},λ_{02},λ_{03}) y proporción de individuos curados(p).

De acuerdo al cuadro 5.3, los intervalos de confianza de los odds de riesgos presentan un nivel de confianza del 95 %, teniéndose que el odds de riesgo aumenta en un 79.3 % (OR= 1.793, 95 % IC = [1.102,2.917]), en un alumno que ha ingresado al área de Letras con respecto a un alumno que ha ingresado al área de Educación; el odds de riesgo es de 2.105 veces más (95 % IC = [1.294,3.426]), en un alumno que ha ingresado al área de Ciencias con respecto a un alumno que ha ingresado al área de Educación; a diferencia de un alumno que ha ingresado al área Arte, pues su odds de riesgo disminuye en un 61.2 % con respecto a un alumno que ha ingreso al área de Educación. Con respecto a la variable "**CRAEst**", se tiene un coeficiente de $\beta_5/\sigma_{CRAEst} = -0.193$, generando un OR = 0.824 (OR = exp(-0.193)), indicando que el odds de riesgo disminuye en un 17.6 % si aumenta en un punto la variable "**CRAEst**". Además, siguiendo los resultados del cuadro 5.4, y analizando solo a los alumnos que alcanzaron el evento de interés, la probabilidad de que un alumno haya abandonado la universidad en el primer periodo es 0 (95 % IC: [0.000,0.001]), mientras la probabilidad de que abandonde la universidad en el segundo periodo habiendo concluido el primero es de 0.170 (95 % IC: [0.159,0.183]). Si el alumno logró concluir el segundo periodo, la probabilidad de que abandone en el tercer periodo es de 0.453 (95 % IC: [0.430,0.476]); y si logra concluir el tercer periodo, la probabilidad de que abandone en el cuarto periodo es de 1. Por otro lado, la proporción de alumnos que no alcanzaron el evento de interés es de un 0.963 (95 % IC: [0.941,0.977]).

La función de supervivencia $S(t-)$, bajo el modelo mixtura, se observa en la Figura 5.1, que es determinada por (3.3), resultando (1.000,0.994,0.980,0.963) para los tiempos de falla 1,2,3 y 4.

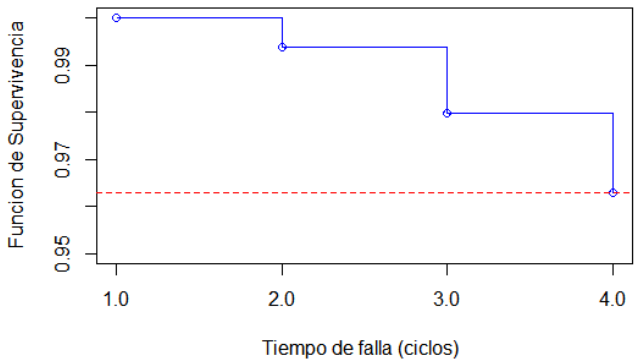

Figura 5.1: Función de Supervivencia de los alumnos en los periodos de 2002 a 2012.

Capítulo 6

Conclusiones

Del presente trabajo podemos concluir que se ha logrado desarrollar un modelo de mixtura de fraccion de cura con odds de riesgos proporcionales que siendo simulados bajo distintos escenarios de censura, con el objetivo de medir el efecto adquirido, se pudieron obtener estimaciones con una buena cobertura.
Si bien estudios anteriores no parten del hecho de incluir la proporción de individuos curados en la función de verosimilitud como en el construido en Zhao and Zhou [2008], en la presente tesis se hace uso de ello.
El análisis con tiempo discreto ha permitido cálculos matemáticos que evitan la complejidad en el desarrollo. Es de notarse que ciertos autores proponen modelos de fracción de cura de tiempo discreto particularizando un estudio de tiempo continuo, esto se debe a que en base a un proceso complejo de continuidad, posteriormente se puede flexibilizar el diseño.
Se aplicó el modelo a un conjunto de datos reales vistos en una ventana de comportamiento de 4 periodos, sobre los alumnos de la Pontificie Universidad Católica del Peru (PUCP) que han ingresado en los años de 2002 a 2012, donde el evento de interés corresponde al abandono de la universidad, obteniéndose odds de riesgo interpretables sobre cada variable en el modelo y tasas de riesgo basales. Además de obtener una proporción de individuos que no alcanzaron el evento de interés en el estudio.

6.1. Sugerencias para investigaciones futuras

Algunas sugerencias de gran importancia son las de

- Comparar performance con otros modelos, bajo distintos escenarios de censura.
- Hacer uso de un enfoque bayesiano para el modelo propuesto.
- Inclusión de un conjunto más amplio de variables predictoras, y un método de selección de variables.

Bibliografía

J. Boag. Mixture and non-mixture cure fraction models based on generalized gompertz distribution under bayesian approach. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11:15–44, 1949.

D. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society*, 34(2): 187–220, 1972. URL `http://www.jstor.org/stable/2985181`.

J. Kalbfleisch and R. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Statistics. Wiley, New York, 2002.

J. Klein and M. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New York, 1997. doi: 10.1007/978-1-4757-2728-9.

P. Kumar, G. Grover, and K. Goel. Mixture and non-mixture cure fraction models based on generalized gompertz distribution under bayesian approach. *Tatra Mountains*, 66:121–135, 2016.

M. Matadamas. Inferencia para modelos de supervivencia de un solo evento y extensiones para modelos de riesgos competitivos. Tesis de maestria en ciencias, Departamento de Matematica, Universidad Autonoma Metropolitana, Mexico, 2010.

T. Scheike and T. Kold. A discrete survival model with random efects: An application to time to pregnancy. *Biometrics*, 53:318–329, 1997.

M. Schmid, H. Kuchenhoff, A. Hoerauf, and G. Tutz. A survival tree method for the analysis of discrete event times in clinical and epidemiological studies. *Statistics in Medicine*, 35: 734–751, 2016.

X. Zhao and X. Zhou. Discrete-time survival models with long-term survivors. *Statistics in Medicine*, 27:1261–1281, 2008. doi: 10.1002/sim.3018.

Códigos R

Listing 1: Simulación

```
1
2 ##########################################################################
3 ###-----------------------------LIBRERIAS-------------------------------------###
4
5 library(gtools)
6 library(matrixcalc)
7 library(numDeriv)
8
9 ##########################################################################
10 ###-----------------------------DISENO DE FUNCIONES---------------------------###
11
12 # Funcion: tasas de riesgo -> Probabilidades
13 f <- function(lambda){
14 f <- c(lambda[1], numeric(dim(array(lambda))-1))
15 for (i in 2:dim(array(lambda))) { f[i] <- lambda[i]*prod(1-lambda[1:(i-1)])} 16 f
17 }
18
19 # Funcion: Probabilidades -> Tasas de riesgo
20 lambda.from.prob <- function(prob){
21 m = length(prob)
22 lambda = numeric(m)
23 for (i in 1:m) lambda[i] = prob[i]/sum(prob[i:m])
24 return(lambda)
25 }
26
27
28 # Funcion de Supervivencia
29 S <- function(lambda){f(lambda)*(1-lambda)/lambda}
30 S2 <- function(p,t,lambda){
31 a=p+(1-p)*f(lambda)*(1-lambda)/lambda
32 a[t]
33 }
34 S.lim <- function(lambda){c(1,S(lambda)[1:(dim(array(lambda))-1)])}
35 S.lim2 <- function(p,t,lambda){
36 a=p+(1-p)*c(1,S(lambda)[1:(dim(array(lambda))-1)])
37 a[t]
38 }
39
40 #------------------------------------- SIMULACION N.1 ----------------------------
41
42
43
44 ##########################################################################
45 ###-------GENERACION DE DATOS (CENSURADOS) CON COVARIABLES PARA ----------###
46 ###-------TASAS DE RIESGO DE LA SUSCEPTIBILIDAD---------------------------###
47 ###
48 ### Tasas de riesgos (lambda)                    : Individuos Totales
49 ### logit(lambda t)=logit(lambda basal)+beta*X
50 ### p                                             : Individuos curados
51
```

```
gen.data.6 <- function(n=500, lambda=c(0.2,0.375,0.3,0.7143,1),beta=0.375,p=0.3,
          censura=c(3,4,5,6,7),prob c=rep(0.2,5)){
          X = rbinom(n,1,0.5)
          t = numeric(n)
          la.s = numeric(5)
          la.s[1]=(1-p)*lambda[1]
          for(i in 2:5) la.s[i]=(1-p)*lambda[i]*prod(1-lambda[1:(i-1)])/(p+(1-p)*
              prod(1-lambda[1:(i-1)]))
          la=matrix(rep(0,5*n),ncol=5,nrow=n)
          for (j in 1:5) {
          la[,j]=exp(logit(la.s[j])+beta*X)/(1+exp(logit(la.s[j])+beta*X))
          }
          for(i in 1:n) {
                   a = f(la[i,])
                   t[i] <- sample(c(1:5,Inf),1,replace=TRUE,prob=c(a,1-sum(a)))
          }
          y <- sample(censura,n,replace=TRUE,prob=prob c) t_i <
          - pmin(t,y)
          ind <- as.numeric(t<=y)
          dat <- data.frame(t=t,t i=t i,ind=ind,X=X)
          return(dat)
}

# Verosimilitud
lik6 <- function(x,t,ind,X){
          n=length(t)
          term = numeric(n)
          p = exp(x[5])/(1+exp(x[5]))
          lambda = numeric(5)
          lambda[1:4] = exp(x[1:4])/(1+exp(x[1:4]))
          lambda[5] = 1
          la.s = numeric(5)
          la.s[1]=(1-p)*lambda[1]
          for(i in 2:5) la.s[i]=(1-p)*lambda[i]*prod(1-lambda[1:(i-1)])/(p+(1-p)*
              prod(1-lambda[1:(i-1)]))
          lambda2=matrix(rep(0,5*n),ncol=5,nrow=n)
          for (j in 1:5) {
                   lambda2[,j]=exp(logit(la.s[j])+x[6]*X)/(1+exp(logit(la.s[j])+x
                       [6]*X))
          }
          for (i in 1:n){
                   if (t[i]>5)      term[i]=sum(log(1-lambda2[i,1:5]))
                   {}else{
                            term[i]=ind[i]*(log(lambda2[i,t[i]])-log(1-lambda2[i,t[i
                                ]]))+sum(log(1-lambda2[i,1:t[i]]))
                   }
          }
          -sum(term)
}

#lik6(c(logit(c(lambda.b[1:4],p)),beta),dat$t i,dat$ind,dat$X)

lambda.b = c(0.2,0.375,0.3,0.7143,1)
beta=0.375
p=0.3

# SIMULACIONES
M <- 1000
n = c(100,300,500)

estimador=matrix(0,M,6*3)
colnames(estimador) <- c("lambda1.n1","lambda2.n1","lambda3.n1","lambda4.n1","p.
      n1","beta.n1",
```

```
110         "lambda1.n2","lambda2.n2","lambda3.n2","lambda4.n2","p.n2","beta.n2",
111         "lambda1.n3","lambda2.n3","lambda3.n3","lambda4.n2","p.n3","beta.n3")
112 desviacion = matrix(0,M,6*3)
113 IC = matrix(0,M,2*6*3)
114 colnames(IC) <- c("lambda1.LI.n1","lambda1.LS.n1","lambda2.LI.n1","lambda2.LS.n1"
        ,
115         "lambda3.LI.n1","lambda3.LS.n1","lambda4.LI.n1","lambda4.LS.n1",
116         "p.LI.n1","p.LS.n1","beta.LI.n1","beta.LS.n1",
117         "lambda1.LI.n2","lambda1.LS.n2","lambda2.LI.n2","lambda2.LS.n2",
118         "lambda3.LI.n2","lambda3.LS.n2","lambda4.LI.n2","lambda4.LS.n2",
119         "p.LI.n2","p.LS.n2","beta.LI.n2","beta.LS.n2",
120         "lambda1.LI.n3","lambda1.LS.n3","lambda2.LI.n3","lambda2.LS.n3",
121         "lambda3.LI.n3","lambda3.LS.n3","lambda4.LI.n3","lambda4.LS.n3",
122         "p.LI.n3","p.LS.n3","beta.LI.n3","beta.LS.n3")
123
124
125
126 for (i in 1:length(n)){
127         for (j in 1:M ){
128                 # cens = matrix(rbind(c(3,4,5,6,7),rep(0.2,5)),2,5)
129                 # cens = matrix(rbind(c(1,2,3,4,5),rep(0.2,5)),2,5)
130                 cens = matrix(rbind(c(1,2,3,4),c(0.6,0.1,0.1,0.2)),2,4)
131                 dat = gen.data.6(n=n[i],lambda=c(0.2,0.375,0.3,0.7143,1),beta
                        =0.375,p=0.3,
132                 censura=cens[1,],prob =cens[2,])
133
134                 res <- optim(par=rep(0,6),fn=lik6,t=dat$t i,
135                 ind = dat$ind,
136                 X=dat$X,method="L-BFGS-B",hessian=T)
137                 hess = optim(par = res$par,fn = lik6,t = dat$t i,
138                 ind = dat$ind,
139                 X=dat$X,method="L-BFGS-B",hessian=T)$hess
140
141                 param = c(exp(res$par[1:5])/(1+exp(res$par[1:5])),res$par[6])
142
143                 hess = res$hess
144
145                 Matriz Fisher <- solve(hess) ## Informacion de Fisher observada std <-
146                 sqrt(diag(Matriz Fisher))
147
148                 alpha=0.05
149                 z<-qnorm(1-alpha/2)
150
151                 estimador[j,(6*i-5):(6*i)]<-param
152                 # estimador[j,(6*i-5):(6*i)]<-res$par
153                 desviacion[j,(6*i-5):(6*i)]<- std
154
155                 LI.lam1 <- exp(res$par[1]-z*std[1])/(1+exp(res$par[1]-z*std[1]))
156                 LS.lam1 <- exp(res$par[1]+z*std[1])/(1+exp(res$par[1]+z*std[1]))
157                 LI.lam2 <- exp(res$par[2]-z*std[2])/(1+exp(res$par[2]-z*std[2]))
158                 LS.lam2 <- exp(res$par[2]+z*std[2])/(1+exp(res$par[2]+z*std[2]))
159                 LI.lam3 <- exp(res$par[3]-z*std[3])/(1+exp(res$par[3]-z*std[3]))
160                 LS.lam3 <- exp(res$par[3]+z*std[3])/(1+exp(res$par[3]+z*std[3]))
161                 LI.lam4 <- exp(res$par[4]-z*std[4])/(1+exp(res$par[4]-z*std[4]))
162                 LS.lam4 <- exp(res$par[4]+z*std[4])/(1+exp(res$par[4]+z*std[4]))
163                 LI.p <- exp(res$par[5]-z*std[5])/(1+exp(res$par[5]-z*std[5])) LS.p <
164                 - exp(res$par[5]+z*std[5])/(1+exp(res$par[5]+z*std[5]))
165                 LI.beta <- res$par[6]-z*std[6]
166                 LS.beta <- res$par[6]+z*std[6]
167
168                 # if(is.na(LS.lam4)==T){
169                 #   LS.lam4 = 1
170                 # }
```

```
171
172                     IC[j,(12*i-11):(12*i)]<-c(LI.lam1,LS.lam1,
173                             LI.lam2,LS.lam2,
174                             LI.lam3,LS.lam3,
175                             LI.lam4,LS.lam4,
176                             LI.p,LS.p,
177                             LI.beta,LS.beta)
178           }
179 }
180
181 round df <- function(x, digits) {
182 # round all numeric variables
183 # x: data frame
184 # digits: number of digits to round
185 numeric columns <- sapply(x, mode) == 'numeric'
186 x[numeric columns] <- round(x[numeric columns], digits) 187 x
188 }
189
190 estimador3 = estimador
191 IC3 = IC
192
193 alpha=0.05
194 z<-qnorm(1-alpha/2)
195
196 estimador3 = data.frame(estimador3)
197 mean estimador3=data.frame(lapply(estimador3,mean))
198 mean_estimador3 = round df(mean estimador3,3)
199
200 std estimador3=data.frame(lapply(estimador3,sd))
201
202
203 cobertura = matrix(0,3,6)
204 para = c(lambda.b[1],

205   lambda.b[2],
206   lambda.b[3],
207 lambda.b[4],
208 p,
209 beta)
210 for (i in 1:3){
211 for( j in 1:6){
212 cobertura[i,j]<-mean((para[j]>IC3[,2*(j-1)+12*(i-1)+1])&(para[j]<IC3[,2*(j-1)
      +12*(i-1)+2]))
213 }
214 }
215
216
217 amplitud = matrix(0,3,6)
218
219 for (i in 1:3){
220           for( j in 1:6){
221           amplitud[i,j]<-mean(IC3[,2*(j-1)+12*(i-1)+2] - IC3[,2*(j-1)+12*(i-1)+1]
                )
222           }
223 }
224
225 cobertura = round df(cobertura,3) 226
amplitud = round df(amplitud,3)
227
228 print(mean_estimador3)
229 print(cobertura)
230 print(amplitud)
```

Listing 2: Aplicación

```
lik6 <- function(x,t,ind,X,n_la,n_cov){
          n=length(t)
          term = numeric(n)

          lambda = numeric(n_la)

          # tasas de riesgo basales: son n_la cantidades
          lambda[1:(n_la-1)] = exp(x[1:(n_la-1)])/(1+exp(x[1:(n_la-1)]))
          lambda[n_la] = 1


          #probabilidad de cura
          p = exp(x[n_la])/(1+exp(x[n_la]))


          la.s = numeric(n_la)
          la.s[1]=(1-p)*lambda[1]

          #betas
          x2 = x[(n_la+1):(n_la+n_cov)]

                  for(i in 2:n_la) la.s[i]=(1-p)*lambda[i]*prod(1-lambda[1:(i-1)])/(p+(1-p
              )*prod(1-lambda[1:(i-1)]))
          lambda2 = matrix(rep(0,n_la*n),ncol=n_la,nrow=n)
          for (j in 1:n_la) {
                  lambda2[,j] = exp(logit(la.s[j])+as.vector(as.matrix(X) %*% %as.
                      matrix(x2,ncol=1,nrow=n_cov)))/(1+ exp(logit(la.s[j])+as.
                      vector(as.matrix(X) %*% %as.matrix(x2,ncol=1,nrow=n_cov))))
          }
          for (i in 1:n){
                  if (t[i]>n_la){     term[i] = sum(log(1-lambda2[i,1:n_la]))
          }else{
                  term[i]=ind[i]*(log(lambda2[i,t[i]])-log(1-lambda2[i,t[i]])) +
                      sum(log(1-lambda2[i,1:t[i]]))
                  }
          }
          - sum(term)
}

lik6(rep(0.02,9),dat$TIEMPO_FALLA,dat$CENSURA,dat[c(2,5,6,7,8)],n_la=4,n_cov = 5)



library(fastDummies)
library(data.table)

alumnos <- read.table("C:/Users/thonyehuertas/Documents/PUCP/TESIS/Tesis de
    Maestr"a - Anthony/VERSION 5/Alumnos TESIS_F 2.csv",sep = '|',header=TRUE)

alumnos = alumnos[,c(4,5,12,13)]
knitr::kable(head(alumnos,8))

dat <- fastDummies::dummy_cols(alumnos)
knitr::kable(head(dat,8))

dat = as.data.frame(dat)


dat$CRAEST2 = (dat$CRAEST-mean(dat$CRAEST))/sqrt(var(dat$CRAEST))

mean(dat$CRAEST)/sqrt(var(dat$CRAEST))

```

```
58
59 res <- nlminb(start=rep(0,9),
60           objective =lik6,
61           t = dat$TIEMPO FALLA,
62           ind = dat$CENSURA,
63           X=dat[c(5,6,7,8,10)], n
64           la=4,
65           n_cov = 5)
66
67 param = c(exp(res$par[1:4])/(1+exp(res$par[1:4])),res$par[5:10])
68 param
69
71 res2 <- aptim(par   es$par fn   ik6
72           t = dat$TIEMPO FALLA,
73           ind = dat$CENSURA,
74           X=dat[c(5,6,7,8,10)],
75           n_la=4,
76           n_cov = 5,
77           method = "L-BFGS-B",hessian=T)
78
79 atriz isher <- aolve(res2$hess) # nformacion e isher bservada 80
std <- sqrt(diag(Matriz isher))
81
82
83 alpha=0.05
84 <-qnorm(1-alpha 2)
85
86 I.lam1 <- axp(res$par[1]-z*std 1]) (1+axp(res$par[1]-z*std 1]))
87 S.lam1 <- axp(res$par 1]+z*std 1]) (1+axp(res$par 1]+z*std 1]))
88 I.lam2 <- axp(res$par[2]-z*std 2]) (1+axp(res$par[2]-z*std 2]))
89 S.lam2 <- axp(res$par 2]+z*std 2]) (1+axp(res$par 2]+z*std 2]))
90 I.lam3 <- axp(res$par[3]-z*std 3]) (1+axp(res$par[3]-z*std 3]))
91 S.lam3 <- axp(res$par 3]+z*std 3]) (1+axp(res$par 3]+z*std 3]))
92 I.p <- axp(res$par[4]-z*std 4]) (1+axp(res$par[4]-z*std 4]))
93 S.p <- axp(res$par 4]+z*std 4]) (1+axp(res$par 4]+ z*std 4]))
94 I.beta1 <- es$par[5]-z*std[5]
95 S.beta1 <- es$par 5]+z*std[5]
96 I.beta2 <- es$par[6]-z*std[6]
97 S.beta2 <- es$par 6]+z*std[6]
98 I.beta3 <- es$par[7]-z*std[7]
99 S.beta3 <- es$par 7]+z*std[7]
100            LI.beta4          <-          res$par[8]-z*std[8]
101          LS.beta4          <-       res$par          [8]+z*std[8]
102            LI.beta5          <-          res$par[9]-z*std[9]
103 LS.beta5 <- res$par[9]+z*std[9]
104
105
106 odds = round df(data.frame(c(exp(res$par[5:9]))),3)
107
108 I dds  round af(data.frame(c(exp(res$par[5:9]-z*std 5:9]))) 3)
109 S dds  round af(data.frame(c(exp(res$par 5:9]+z*std 5:9]))) 3)
```