Efecto de valores faltantes en estudios longitudinales en adultos mayores

Franklin Fernando Massa Mandagarán

Maestría en Ingeniería Matemática
Instituto de Matemática y Estadística Rafael Laguardia
Facultad de Ingeniería
Universidad de la República

Noviembre 2015.
Montevideo.

Efecto de valores faltantes en estudios longitudinales en adultos mayores

Franklin Fernando Massa Mandagarán

Tesis presentada con el objetivo de obtener el título de Magíster en Ingeniería Matemática

Directora de Tesis: Graciela Muniz
Director Académico: Marco Scavino

Noviembre 2015.
Montevideo.

Defensa realizada el 16 de diciembre frente al comité de examinadores:

Dr. Graciela Muniz	Directora de Tesis
Dr. Marco Scavino	Director Académico
Dr. Mathías Bourel	
Dr. Juan Gil	
Dr. Juan Kalemkerián	

Agradecimientos

Me gustaría agradecer a todos los que colaboraron en este proceso e hicieron que fuera una gran instancia de aprendizaje. En primer lugar agradezco a mi directora de tesis, Graciela Muniz, quien me brindó la posibilidad de trabajar con ella, aportando todo su conocimiento para poder llevar adelante este trabajo. Asimismo, quisiera agradecer a mi director académico, Marco Scavino, por estar presente en las distintas etapas de este proceso. También quisiera agradecer a mis compañeros del Insituto de Estadística de la Facultad de Ciencias Económicas, UdelaR, por su gran apoyo y colaboración. Finalmente agradezco a mis amigos y mi familia por su comprensión y por brindarme consejos dándome la mejor guía. Por último pero no por ello menos importante agradezco a Natalia Berberian por estar a mi lado.

Resumen

Comprender el proceso de deterioro cognitivo global de los adultos mayores es de gran relevancia para una mejor planificación no sólo a nivel económica/familiar, repercutiendo de forma directa en los individuos afectados, sino también a nivel del sistema de salud y seguridad social de un país. Puesto que el deterioro se manifiesta a lo largo del tiempo, para su estudio son implementados modelos longitudinales. Sin embargo estos modelos tienen como desventaja que a lo largo del período de seguimiento, por diversas causas, algunos individuos desertan del estudio. Este abandono no sólo reduce el tamaño muestral sino que genera importantes sesgos en los resultados de los análisis estadísticos.

El objetivo de este trabajo fue modelar el deterioro cognitivo global de un conjunto de adultos mayores a lo largo del tiempo. Para llevar a cabo este objetivo se trabajó con datos provenientes del estudio "Origins of Variance in the Old-old: Octagenarian Twins" que contaba con una cohorte de 702 individuos entre 79 y 98 años de edad. Para medir el deterioro global se utilizó el *Mini Mental State Examination* ($MMSE$), evaluado a lo largo del tiempo en intervalos de dos años. Se implementaron modelos conjuntos puesto que estos parten de la base de que el tiempo de sobrevida de los sujetos y sus resultados de MMSE están relacionados.

Se observó que los valores de $MMSE$ son afectados tanto por la edad de los individuos al inicio del estudio como por el nivel educativo de los mismos. Incluir el análisis de sobrevida como parte del proceso de deterioro cognitivo corrige las inferencias.

No existen trabajos previos en el Uruguay relacionados con la metodología desarrollada en este trabajo por lo que se considera que este trabajo representa una primera aproximación a la metodología a ser aplicada tanto para la planificación en el sistema de salud y seguridad social como para el sistema educativo e incluso para el contexto de trasplante de órganos.

Palabras claves: Modelo conjunto; Datos faltantes; Deterioro cognitivo.

Abstract

Understanding the process of global cognitive decline in older adults is of great importance for better planning not only in a personal level for the affected individuals but also for health service and social security service providers. Since the cognitive decline becomes evident over time, longitudinal models are implemented for its study. However these models have the disadvantage that during the follow up period, due to several reasons, some individuals drop out of the study. This drop out not only reduces the sample size but are also likely to be an importatn source of bias in results from the statistical analysis.

The objective of this study was to model the global cognitive decline of a set of older adults over time. To accomplish this objective data from the study "Origins of Variance in the Old-old: Octagenarian Twins" with a cohort of 702 individuals between 79 and 98 years old was used. The global cognitive function was assessed using the Mini Mental State Examination ($MMSE$), evaluated over time in two-year intervals. Joint models were implemented since these are based on the assumption that the survival time of individuals and their results of MMSE are related.

Based on results we conclude that the baseline $MMSE$ values depended on the age of individuals and the rate of decline was affected by their educational level. The inclusion of survival analysis as part of the cognitive decline process updates inferences.

To the best of our knowledge, this is the first implementation of this modelling approach in Uruguay. We hope that it can be extended and implemented in this area to improve existing knowledge about cognitive ageing. Furthermore, this statistical methodology may be used in other research areas including organ donation, HIV research, and others.

Keywords: Joint model, Missing data; Cognitive decline.

Índice general

Capítulo 1

Introducción

Naciones Unidas ha estimado que en el año 2050 el número de individuos mayores de 60 años superará los 2 billones y que por primera vez, la proporción de la población de adultos mayores superará a la de niños menores de 14 años. Las consecuencias de este cambio en la pirámide poblacional son vastas desde el punto de vista social e individual. Algunos países están tomado previsiones al respecto con el fin de reducir el impacto en las estructuras sociales.

A nivel individual, la pérdida de la salud física y de la capacidad cognitiva son los mayores desafíos que los adultos mayores enfrentan. Entender estos procesos es entonces fundamental, y para eso, los estudios longitudinales representan una excelente oportunidad. Científicamente hay varias preguntas de interés, siendo algunas de ellas ¿cuál es el cambio marginal de la poblacion total a lo largo del tiempo?, ¿cuál es el cambio experimentado a nivel individual?, ¿cuáles son los principales factores de riesgo que afectan la pérdida de la capacidad cognitiva?, ¿son los mismos para individuos en el segmento más jóven (75-79 años), 80-84 y los más ancianos (89+)?, ¿cuáles de estos factores son modificables?, ¿cuál es el efecto de haber sufrido un accidente cardiovascular, o de estar clínicamente deprimido en la salud mental?, ¿existe un efecto de los años de educación en la pérdida de la capacidad cognitiva?, ¿personas con mayor escolaridad declinan a la misma velocidad que los menos educados? Estas últimas interrogantes son de suma importancia debido a que la educación es un factor modificable por el propio individuo (Lenehan et al., 2015).

Cuando buscamos respuestas a estas preguntas, nos enfrentamos con uno

de los mayores problemas de los estudios longitudinales de individuos mayores; el gran número de observaciones perdidas. Entender las razones por las cuales los individuos cesan su participación en los estudios es fundamental para modelar adecuadamente el cambio experimentado por estos individuos, ya sea a nivel individual como marginal. Por ejemplo, algunos individuos cesan su participación debido a condiciones de salud física, otros dejarán de participar por razones totalmente ajenas a su situación de salud (por ejemplo, porque se mudan a otra zona del país), otros estarán muy impedidos mentalmente como para comprender las preguntas que se les hacen en los tests y finalmente otros morirán durante el desarrollo del estudio.

Otra manera de entender el fenómeno de pérdida de datos es que conforme avanza el tiempo, los individuos se deterioran y simultáneamente aumenta la probabilidad de fallecimiento, esto hace pensar en un mecanismo donde el tiempo hasta experimentar el evento bajo estudio (en este caso el fallecimiento) depende del estado cognitivo del individuo. Ignorar la información aportada por algunos de estos casos quizá no afecte los estimadores (y por ende las conclusiones del estudio), pero ciertamente ignorar la información aportada por otros sesgará los resultados masivamente.
Por estos motivos, este trabajo se basa en describir el deterioro cognitivo teniendo en cuenta características como la edad, el sexo y la educación sin dejar de prestarle atención al proceso de sobrevida de los individuos. Para lograr una adecuada exposición, la tesis se organiza de la siguiente manera. El capítulo 2 presenta una breve descripción del fenómeno bajo estudio, el deterioro cognitivo. Los capítulos del 3 al 7 presentan las diferentes técnicas estadísticas utilizadas. En los capítulos 8.1 al 8.3 se presenta la aplicación de los modelos descriptos en los capítulos anteriores a una caso particular. Finalmente, en el capítulo 9.1 se presentan las conclusiones obtenidas y posibles líneas de investigación a futuro.

1.1. Antecedentes

En los últimos 20 años se han formado diversas líneas de investigación dentro del campo del estudio de la evolución del envejecimiento. Más concretamente, en lo que refiere al deterioro cognitivo, algunas aplicaciones de estas técnicas han tocado temas como mortalidad, memoria, deterioro y habilidad motora.

En Terrera et al. (2011) se presta particular atención a los cambios en la memoria y como su deterioro podría presentar una aceleración en su trayectoria. Uno de los principales hallazgos fue que la tasa de decaimiento de la memoria consiste en un predictor de la proximidad del fallecimiento. En el estudio de Ghisletta (2008) se estudian la velocidad de percepción y la fluidez verbal de ancianos como posibles predictores del riesgo de fallecimiento. El trabajo de Hughes et al. (1997) se encarga de estudiar los factores que predicen el deterioro en la habilidad manual de los ancianos.

Fuera del ámbito del envejecimiento, existe una vasta literatura dedicada a la descripción de los avances en el área de modelos conjuntos. En la gran mayoría de los casos, está dedicada al análisis de biomarcadores relacionados con la evolución del virus del síndrome de inmunodeficiencia adquirida (SIDA). En este entorno se destacan los trabajos de Self and Pawitan (1992) y de DeGruttola and Tu (1994). El primero considera el desarrollo de un modelo conjunto para la descripción de la evolución del número de células de los biomarcadores T4 y T8 así como el tiempo desde la seroconversión hasta el diagnóstico de SIDA. El segundo se encarga de analizar la incidencia de diversos factores en la evolución del conteo de linfocitos CD4 y su asociación con el tiempo de sobrevida. En este trabajo, los autores adoptan la metodología propuesta por Wu and Carroll (1988) incluyendo efectos aleatorios para modelar la asociación entre el proceso longitudinal y el proceso de sobrevida. En contrapartida a los casos anteriores, Faucett and Thomas (1996) se valen de métodos Bayesianos, más concretamente *Gibbs-Sampling* para estimar la distribución a posteriori de los parámetros del modelo conjunto. En última instancia es muy valioso destacar el trabajo de Wulfsohn and Tsiatis (1997) quienes sientan las bases del que hoy se considera el modelo conjunto standard. En su artículo, los autores proponen el uso del algoritmo *Expectation-Maximization* (EM) (Dempster et al., 1977) asegurando que su método es superior a los anteriores por dos motivos; no se basa en una estimación en dos etapas y no maximiza la porción de la verosimilitud del modelo de sobrevida utilizando los valores observados del biomarcador (contaminados por el error de medición).

Sin embargo, este tipo de modelos no ha sido implementado únicamente en esta área, Henderson et al. (2002) proponen un estadístico para contrastar la asociación entre los procesos longitudinal y de sobrevida, aplicándolo al caso del índice de protrombina como un posible indicador de la sobrevi-

da en pacientes afectados por cirrosis. Un segundo ejemplo se presenta en el artículo de Taylor et al. (2013) donde los modelos conjuntos se utilizan para estimar la probabilidad de recurrencia de cáncer de próstata valiéndose del valor del antígeno prostático específico (PSA). Las aplicaciones que han surgido en los últimos años son de diversa índole, algunas son: transplantes de riñón (Musoro et al., 2014), cuidados paliativos en pacientes con cancer (Li et al., 2013), transplantes de pulmón en pacientes con fibrosis quística (Thabut et al., 2013), enfermedad de Parkinson (He and Luo, 2013), entre otras.

Pese a no existir antecedentes en el Uruguay, estos modelos podrían ser de especial utilidad debido a que Uruguay es uno de los países más envejecidos de la región (Cabella and Pellegrino, 2009). En este sentido, estudiar los posibles determinantes que intervienen en el proceso de envejecimiento de los ancianos uruguayos sería un punto a tener en cuenta para futuras líneas de investigación ya que tener un mejor entendimiento de proceso de deterioro podría permitir una mejor planificación económica/familiar. En este sentido, dado que estos modelos permiten realizar predicciones (tanto de la trayectoria de las variables asociadas al deterioro como de la probabilidad de fallecimiento) esto permitiría al sistema de salud/seguridad social planificar con más información.

Otras posibles aplicaciones de estos modelos, fuera del ámbito del deterioro cognitivo, serían en el contexto de los trasplantes de órganos (algunos biomarcadores podrían predecir la probabilidad de rechazo del órgano y el tiempo de sobrevida del mismo y del paciente) o para estudiar la evolución (y las causas) del abandono de los niños al sistema educativo.

1.2. Objetivos

Este trabajo se llevó a cabo con la finalidad de aplicar las técnicas referentes al modelado conjunto de datos longitudinales y de sobrevida. Para ello se utilizaron los datos del estudio "*Origins of Variance in the Old-old: Octogenarian Twins*" (*OCTO-Twin Study*). El mismo consistía de mellizos suecos cuya edad era al menos 80 años en el año 1991, cuando se inició el estudio. El período de seguimiento de los individuos consistió de 4 visitas (posteriores a la evaluación inicial) en períodos de 2 años, donde se recabaron datos sobre

memoria, capacidad funcional, y salud entre otras mediciones. El presente estudio se focalizó sobre la evolución del resultado del "Mini-Mental State Examination" ($MMSE$). Se trata de un cuestionario (véase anexo B.1) con un puntaje máximo de 30 puntos desarrollado por Folstein et al. (1975) que mide la función cognitiva general y es utlizado muchas veces como herramienta en contextos clínicos que pretende determinar el deterioro cognitivo y diagnosticar la demencia

El objetivo general que se persigue en este trabajo es el de estimar la velocidad de cambio del $MMSE$ y determinar los factores que puedan alterarla, prestando especial atención al proceso de fallecimiento de los individuos ya que el mismo puede sesgar los resultados.

Capítulo 2

Deterioro cognitivo

Todas las personas desarrollan cierto grado de deterioro en sus funciones cognitivas a medida que transcurre el tiempo. El impedimento cognitivo es un término clínico utilizado para describir una condición asociada a problemas de la función cognitiva como lo son pensar, recordar, razonar, problemas en el lenguaje, en la atención o en ciertas actividades visuales o espaciales. Las personas que lo sufren suelen tener mayores problemas en el "día a día" en actividades relacionadas principalmente con la memoria, pero dichos problemas no son lo suficientemente graves como para diagnosticar un caso de demencia.

En estas personas, el proceso de deterioro es más pronunciado que en personas que envejecen de manera natural. En el caso particular de la pérdida de memoria, esto puede indicar un primer signo del desarrollo de demencia o incluso Alzheimer. En pacientes que presentan signos de deterioro en otras actividades cognitivas, esto puede resultar en el desarrollo de otras enfermedades como demencia vascular, demencia fronto-temporal o demencia de cuerpos de Lewy.

Diversos estudios sugieren que entre 5 y 20 por ciento de la población anciana sufre de algún tipo de impedimento cognitivo en algún momento. La importancia en este tipo de datos radica en que este grupo de personas posee un riesgo de entre tres y cinco veces mayor de desarrollar algún tipo de demencia. Es por esto que resulta de vital importancia la detección temprana de esta enfermedad, de esta manera el paciente puede recibir el apoyo necesario para acompañar el proceso de deterioro de la mejor manera posible. Se prevee que para el futuro se desarrollen drogas que prevengan la progresión

de esta condición hacia la demencia.

El deterioro cognitivo no afecta a todos los individuos por igual, se han identificado asociaciones entre la tasa y la severidad del deterioro con diversos factores como el stress (Ansari and Derakshan, 2011; Newberg et al., 2010), descenso de ciertas hormonas (Hogervorst et al., 2010; Ryan et al., 2012), exceso de peso (Kerwin et al., 2011; Abbatecola et al., 2010), hipertensión (Bellew et al., 2004; Swan and Larue, 1998), diabetes (Biessels et al., 2006; van Elderen et al., 2010) y el estilo de vida (Wang et al., 2002; Atti et al., 2010) entre otras. Sin embargo, muchos de estos factores son modificables por el sujeto y, junto con un entrenamiento cognitivo, una nutrición adecuada puede resultar en un descenso de la tasa del deterioro.

En cuanto a la detección del impedimento en sí, existen diversas alternativas, una de las cuales es el $MMSE$. Pese a que este test no constituye una herramienta de diagnóstico, sí es un instrumento que permite detectar y estimar cuantitativamente la severidad del deterioro cognitivo y sus cambios a lo largo del tiempo. La prueba en sí es sencilla y de rápida aplicación pero su aplicación es limitada ya que presenta ciertos sesgos en tanto que no logra captar pérdidas leves en la memoria de sujetos con alto nivel educativo. La prueba se compone de varias secciones, las cuales pretenden evaluar la orientación, memoria de corto plazo y lenguaje. El resultado de la prueba es un puntaje (con máximo de 30 puntos), valores superiores a 25 sugieren un estado normal en la persona, valores entre 21 y 24 sugieren un deterioro leve de las funciones cognitivas, valores entre 10 y 20 son signos de un deterioro moderado mientras que valores menores a 10 puntos son característicos de personas con un deterioro severo.

Capítulo 3

Análisis de datos longitudinales

Un estudio longitudinal es aquel en el cual las variables de interés se registran en múltiples instancias sobre un conjunto de individuos. La principal característica de esta manera de investigar es que permite caracterizar el cambio y los factores que lo determinan a través del tiempo. Sin embargo, la metodología de análisis clásico presenta limitaciones en este tipo de estudios debido a que, dada la naturaleza secuencial del relevamiento de datos, existe correlación entre las observaciones de un mismo individuo. No obstante, el hecho de que las medidas se registren de manera secuencial presenta el inconveniente de que las mediciones de un mismo individuo suelen presentar correlación (generalmente positiva) por lo tanto requieren de un tratamiento específico que permita realizar inferencias válidas. De todas maneras, al hacer un balance entre "ventajas" y "desventajas", el resultado neto es que los datos longitudinales proveen de una mayor cantidad de información para cada individuo que un análisis de tipo "*cross-section*".
En la mayoría de los estudios que utilizan este tipo de análisis, algunas de las preguntas que interesa responder son las siguientes:

- ¿Qué tanto difieren 2 (o más tratamientos) a lo largo del tiempo?

- ¿Cómo evoluciona la diferencia entre tratamientos a lo largo de cierto período y que factores afectan dicha diferencia?

3.1. Definiciones

Una idea bastante natural en este ámbito es pensar que las mediciones de cada individuo del estudio siguen una forma funcional específica, la cual es un caso especial de un cierto perfil poblacional. Antes de dar una representación estadística, es necesario introducir algunos elementos. Sea $y(t_{ij})$ la medición del individuo i en el momento t_{ij}. El modelo más sencillo para describir trayectorias longitudinales es el siguiente modelo de regresión lineal:

$$y(t_{ij}) = \beta_{0i} + \beta_{1i} * t_{ij} + \epsilon(t_{ij}). \tag{3.1}$$

Donde se asume que los errores $\epsilon(t_{ij})$ siguen una distribución normal con media nula y varianza constante. En cuanto a las covarianzas de estos errores se suele asumir que son nulas, sin embargo es posible especificar una estructura que permita que distintos errores del mismo individuo estén correlacionados, no así para los errores de distintos individuos. Hasta aquí, este modelo lineal no tiene ninguna particularidad. Pero si tenemos en cuenta que los individuos representan una muestra aleatoria de una cierta población, es razonable suponer que los coeficientes de regresión β_{0i} y β_{1i} también sean realizaciones de una variable aleatoria bidimensional (β_0, β_1). Al suponer que estos coeficientes también tienen distribución normal (independiente de los errores), el modelo anterior se puede escribir de la siguiente manera:

$$y(t_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) * t_{ij} + \epsilon(t_{ij}). \tag{3.2}$$

Esta reformulación permite distinguir entre efectos fijos (β_0 y β_1) y efectos aleatorios(b_0 y b_1), donde los primeros son comunes a toda la población y por lo tanto constituyen cantidades a *estimar* mientras que los últmos son realizaciones de variables aleatorias relativas a cada individuo y el objetivo de la etapa inferencial será de *predecirlas*. Dada la presencia de efectos fijos y aleatorios, es que estos modelos suelen llamarse *mixtos*.
Una de la características más notorias que posee la formulación (3.2) es que distintas observaciones de un mismo individuo son factibles de presentar correlación. A dicha correlación, comúnmente se la llama "correlación intraclase".

3.2. Modelos

El modelo presentado en la sección anterior puede generalizarse fácilmente de modo de incluir covariables que describen adecuadamente variaciones en

la media de los perfiles longitudinales:

$$\begin{cases} Y_i = X_i\beta + Z_i b_i + \epsilon_i \\ b_i \sim N\left(0, D(\gamma)\right) \\ \epsilon \sim N\left(0, \Sigma(\gamma)\right) \end{cases} \quad (3.3)$$

En este caso, Y_i es un vector que contiene las n_i observaciones del i-ésimo individuo, X_i es la matriz que contiene las covariables que afectan los coeficientes poblacionales, Z_i contiene las covariables aociados a los parámetros aleatorios, β y b representan los efectos fijos y aleatorios respectivamente y por último ϵ_i es el vector de errores, de la misma dimensión de Y_i que pueden o no presentar varianzas heterogéneas y covarianzas nulas para un mismo individuo. La matriz de covarianzas $\Sigma(\gamma)$ suele ser diagonal, por otro lado, $D(\gamma)$ puede no tiene por qué serlo y los elementos de ambas suelen ser función de "pocos" parámetros, contenidos en el vector γ. Vale aclarar que uno de los supuestos iniciales de estos modelos es que se asume independencia entre las variables no observables ϵ y b así como también se supone independencia entre los vectores de respuesta de distintos individuos. Desde el punto de vista del modelo esto implica que:

$$Cov(\epsilon_i, \epsilon_{i'}) = \begin{cases} \Sigma(\gamma) & , \quad i = i' \\ \mathbf{0} & , \quad i \neq i' \end{cases} \quad (3.4)$$

Es sencillo notar que al condicionar en los efectos aleatorios, la distribución del perfil del sujeto i es normal con vector de medias $X_i\beta + Z_i b_i$ y varianza $\Sigma(\gamma)$, por otro lado, la distribución marginal del vector Y_i también es normal, pero con vector de medias $X_i\beta$ y matriz de covarianzas $\Sigma(\gamma) + Z_i' D(\gamma) Z_i$. De esta manera se puede ver que al manipular adecuadamente la matriz Z_i, es posible captar diversos patrones de correlación entre las medidas de un mismo individuo.

Algunas de las ventajas que postula el modelo descrito en (3.3) para analizar patrones longitudinales es que no presenta inconvenientes cuando diferentes individuos tienen distinto número de observaciones y que es capaz de manejar adecuadamente el hecho de que, por lo general, en estudios longitudinales las mediciones no suelen llevarse a cabo para toda la muestra de individuos en los mismos momentos. En cuanto a las ventajas del modelo propiamente dicho, permitiría la predicción de la trayectoria futura de cada individuo en particular así como de la población en su conjunto. Sin embargo, una gran de-

ventaja de este modelo es que, postulado de esta manera, el modelo permite realizar la predicción de los valores del individuo aún luego de fallecimiento.

3.3. Inferencia

Según lo expuesto en el apartado anterior, el vector con las mediciones del i-ésimo individuo se distribuye

$$Y_i \sim N\left(X_i\beta, Z_i^{'}D(\gamma)Z_i + \Sigma(\gamma)\right). \tag{3.5}$$

Las inferencias realizadas bajo este modelo marginal no necesariamente asumen la presencia de efectos aleatorios, pese a que estos se hayan usado para modelar la heterogeneidad entre los individuos. Antes de plantear la función de verosimilitud de la muestra de n individuos, es conveniente particionar el vector de parámetros (como en Verbeke and Molenberghs (2000)) en dos conjuntos. De esta forma, el vector θ contendrá todos los parámetros del modelo, siendo β el vector de parámetros asociados a la media del vector Y_i y γ el vector cuyas componentes intervienen en los elementos de D y Σ. De esta manera la función de log-verosimilitud será la indicada en la ecuación (3.6).

$$\mathcal{L}_{ML}(\theta) \propto \sum_{i=1}^{n}\left\{\frac{1}{2}\left|V_i(\gamma)\right| - \frac{1}{2}(Y_i - X_i\beta)^{'}V_i^{-1}(\gamma)(Y_i - X_i\beta)\right\} \tag{3.6}$$

donde $V(\gamma) = Z^{'}D(\gamma)Z + \Sigma(\gamma)$.
Al maximizar esta función con respecto a γ y β se obtienen los estimadores de máxima verosimilitud "marginal". Sin embargo estos estimadores no poseen una forma cerrada, por lo que, un posible procedimiento a seguir es obtener el estimador de mínimos cuadrados generalizados de β como:

$$\hat{\beta}(\gamma) = \left(\sum_{i=1}^{n} X_i^{'}V_i^{-1}(\gamma)X_i\right)^{-1}\sum_{i=1}^{n} X_i^{'}V_i(\gamma)Y_i. \tag{3.7}$$

De esta manera, suponiendo conocido el vector de parámetros de covarianza, es posible estimar el vector β, no obstante para esto se hace necesario estimar al vector γ, proceso que suele llevarse a cabo mediante métodos numéricos. Es necesario agregar que la estimación de los parámetros de covarianza puede llevarse a cabo maximizando la verosimilitud o la versión restringida de

esta última, a estos estimadores se los llama $REML$ por su sigla en inglés *Restricted Maximum Likelihood* (véase anexo A.1).

La ventaja de estos estimadores es que, a diferencia de la versión de máxima verosimilitud, son insesgados. Es bien sabido que al estimar la varianza de los residuos de una regresión, se pierde un grado de libertad por cada parámetro estimado en la media de los datos (véase Walker (1943)). Por lo cual el estimador insesgado surge como el estimador correspondiente a un conjunto de contrastes que "eliminen" los parámetros correspondientes a la media. En el ámbito de los modelos longitudinales esto implica transformar los datos con una matriz de contrastes A, de modo que $AY_i \sim N(0, AV_i(\gamma)A^{'})$. Es importante notar que al cambiar el método de estimación (y por ende los valores estimados $\hat{\gamma}$), es de esperar que las estimaciones de los parámetros asociados a la media también cambien, contrario a lo que sucede en los modelos de regresión de efectos fijos donde las estimaciones de β y σ^2 son independientes.

Al ser derivados bajo el mismo planteamiento, tanto los estimadores obtenidos por ML como por $REML$ gozan de propiedades deseables como consistencia, normalidad asintótica y eficiencia (Richardson and Welsh, 2008). En general, es de esperar que la diferencia entre ambos tipos de estimadores "crezca" al aumentar el número de covariables en la media de la distribución, debido a que esto aumentaría el rango de la matriz de contrastes A.

En cuanto al procedimiento para realizar inferencias, éste se puede dividir en dos apartados: el referente a los parámetros de la media y el asociado a a los parámetros de covarianza.

3.3.1. Inferencia sobre β

A la hora de realizar inferencias sobre los elementos del vector β existen diferentes procedimientos. Los más comunes consisten en realizar intervalos de confianza basados en la aproximación normal, para lo cual se hacen necesarias las estimaciones del desvío estándar de cada parámetro, o llevar a cabo pruebas de hipótesis. En este último caso existen diversos enfoques. Aquí se presentarán las dos metodologías más comunes que son las pruebas basadas en los estadísticos de Wald y de cociente de verosimilitud.

- Pruebas basadas en el estadístico de Wald.

Antes de pasar al cuerpo de la prueba, vale la pena notar que el vector de estimaciones $\hat{\beta}(\gamma)$ estimado por mínimos cuadrados generalizados (asumiendo que se conocen los parámetros contenidos en γ) se distribuye normal con la siguiente media y matriz de covarianzas:

$$\hat{\beta}(\gamma) \sim N\left(\beta, X'V^{-1}(\gamma)X\right). \tag{3.8}$$

Trabajando sobre esta base, se pueden construir pruebas de hipótesis basadas en restricciones lineales del vector β de la siguiente manera:

$$\begin{array}{ll} H_0) & L\beta = \underline{c} \\ H_1) & L\beta \neq \underline{c} \end{array}$$

donde L es una matriz de contrastes (lineales) que indica las restricciones que se desea poner a prueba. El estadístico de prueba es consecuencia de la normalidad del vector $L\beta$ y su forma es:

$$F = \frac{(\hat{\beta} - \beta)' L' \left[L(X'V^{-1}(\gamma)X)^{-1}L'\right] L(\hat{\beta} - \beta)}{rango(L)}. \tag{3.9}$$

Bajo el cumplimiento de la hipótesis nula, el estadístico (3.9) sigue una distribución F con grados de libertad del numerador equivalentes al rango de la matriz L, mientras que los grados de libertad del denominador deben ser estimados a partir de los datos (véase Waseem (2007)). En el caso especial de que se quisiera contrastar

$$\begin{array}{ll} H_0) & \beta_k = 0 \\ H_1) & \beta_k \neq 0 \end{array}$$

basta con utilizar $L = (0, 0, \ldots, 1, \ldots, 0, 0)$ y $c = 0$. Donde el 1 ocupa el lugar k-ésimo del vector fila L. En dicho caso, la raíz cuadrada del estadístico sigue una distribución de Student cuyos grados de libertad son los correspondientes a los del denominador del estadístico F (los cuales suelen ser estimados por el procedimiento de Satterthwaite (1946), Welch (1947), Kenward and Roger (1997)).

- Pruebas basadas en el estadístico de cociente de verosimilitud (LRT). En el caso del estadístico de cociente de verosimilitud, a diferencia del caso anterior, es necesario re-estimar los parámetros del modelo bajo

el cumplimiento de la hipótesis nula (modelo que suele denominarse "reducido" o "restringido"), lo cual puede resultar costoso si la estimación es computacionalmente demandante. Otros dos aspectos a tener en cuenta son que el método de estimación debe ser máxima verosimilitud (no $REML$) y que se utilice el mismo número de observaciones en la estimación del modelo "completo" y en la estimación del modelo "reducido". Este último punto es de especial importancia en los casos en que algunas covariables presenten datos faltantes en distintas observaciones. El procedimiento dictamina la comparación de las log-verosmiilitudes de ambos modelos ($\mathcal{L}_{completo}(\theta)$ y $\mathcal{L}_{reducido}(\theta)$), y se espera que cuanto mayor sea la diferencia entre ambos, mayor es la evidencia en contra de la hipótesis nula.

Finalmente el estadístico de prueba se construye de la siguiente manera:

$$LRT = -2(\mathcal{L}_{reducido}(\theta) - \mathcal{L}_{completo}(\theta)). \tag{3.10}$$

La distribución del estadístico (3.10) se aproxima (asintóticamente) a una χ^2 con tantos grados de libertad como restricciones se hayan impuesto sobre el modelo reducido (Wilks, 1938). Un uso adicional de este procedimiento (aunque computacionalmete es aún más demandante) es la construcción de intervalos de confianza mediante el "perfilado" de la verosimilitud.

Pese a que esta alternativa requiere una mayor carga desde el punto de vista computacional (tanto para las pruebas de hipótesis como para los intervalos de confianza) las propiedades de estos procedimientos suelen ser mejores a las correspondientes a los basados en el estadístico de Wald, sobre todo (como se presenta en el siguiente apartado) en el caso de las inferencias realizadas sobre los elementos de γ.

3.3.2. Inferencia sobre γ

Pese a que en la gran mayoría de las situaciones, el interés de estos modelos recae sobre las inferencias realizadas sobre la media, modelar los parámetros correspondientes a la covarianza de manera adecuada es igual de importante, tanto para realizar una descripción más rica del fenómeno bajo estudio (modelando la variación intra-individual) como para obtener inferencias válidas

de los elementos del vector β. La estructura de los elementos de γ es clave en el sentido de que una especificación incorrecta o muy restrictiva puede invalidar las inferencias realizadas, mientras que una excesiva flexibilidad (o demasiados parámetros) puede sesgar la estimación de la matriz de covarianza de los estimadores de la media (Altham, 1984).

La interrogante más común a la hora de realizar inferencias sobre estos parámetros es la siguiente:

$$\begin{array}{ll} H_0) & \sigma_e = 0 \\ H_1) & \sigma_e > 0 \end{array} . \tag{3.11}$$

La literatura sobre este tipo de pruebas es amplia y existen diversos procedimientos que permiten extraer conclusiones sobre este planteo. A continuación se exponen las alternativas más comunes.

- Pruebas basadas en el estadístico de Wald.
  Para los parámetros de covarianza existen problemas al utilizar el estadístico de Wald (problema que se acentúa en muestras pequeñas) debido a que, cuando la varianza es cercana a cero, las inferencias realizadas con este procedimiento se encuentran muy cerca del "borde" del espacio paramétrico.

- Pruebas basadas en el estadístico de cociente de verosimilitud (LRT).
  Mientras que el uso de este tipo de estadísticos no plantea mayores problemas cuando se utiliza sobre elementos de β, su uso para contrastar la hipótesis de ausencia de heterogeneidad intra-sujetos ($\sigma_e = 0$), requiere ciertos ajustes para realizar inferencias adecuadas.

  El estadístico de prueba es el mismo que el que se presentó en el apartado de inferencias sobre β sin embargo su distribución es asintóticamente χ^2 sólo si se cumplen ciertas condiciones, una de las cuales es que H_0 no se encuentre sobre el "borde" del espacio paramétrico, ya que en dicho caso, este estadístico padece los mismos problemas que el de Wald. Sin embargo, el estadístico LRT es capaz de realizar el contraste indicado en (3.11) si se realiza un ajuste sobre la distribución del estadístico. El procedimiento sugerido por Stram and Lee (1994) intenta determinar secuencialmente, cuál es la especificación correcta de los efectos aleatorios contenidos en la matriz D. El procedimiento sugiere llevar a cabo la siguiente prueba sobre una matriz D de dimensiones $(q+k)\times(q+k)$:

$$H_0)D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix}$$
$$H_1)D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}.$$

De esta manera se plantea que el q-ésimo efecto aleatorio tiene varianza igual a cero (y por ende covarianzas iguales a cero). Como se describió anteriormente se estima el modelo completo y el modelo asumiendo H_0 como cierta, para luego construirse el estadístico de prueba. Pese a que, en primera instancia, se compararía este valor con el valor correspondiente a una distribución χ^2 con tantos grados de libertad como parámetros iguales a cero, estudios de simulación han demostrado que este procedimiento es demasiado conservador. Por lo tanto Verbeke and Molenberghs (2000) proponen que la distribución del estadístico no es χ^2 sino que es una mezcla de estas distribuciones, según se indica en (3.12).

$$P(LRT \geq x) = 0{,}5P(\chi^2_q \geq x) + 0{,}5P(\chi^2_{q+k} \geq x). \tag{3.12}$$

También es importante aclarar que, a diferencia del apartado que concierne a la inferencia sobre β, la discusión anterior también es válida utilizando $REML$ en vez de ML, de hecho el estadístico basado en $REML$ presenta niveles de rechazo de H_0 levemente más cercanos al nivel nominal[1].

[1]Se habla de nivel nominal debido a que la distribución del estadístico es asintótica y en la práctica el número de observaciones es finita, por lo que la probabilidad de error de tipo I puede ser mayor a la deseada.

Capítulo 4

Análisis de datos de sobrevida

El análisis de sobrevida comprende un conjunto de métodos para estudiar la duración hasta que suceda un eventos de interés (o varios). El caso más común se da en la biología, el fallecimiento. Sin embargo sus aplicaciones pueden surgir en diversas áreas, algunos ejemplos son el análisis del tiempo de estadía en un hospital, la duración de una huelga, etc. Es más, pese a que el nombre sugiere la intervención del tiempo, en el campo de ensayo de materiales, estas técnicas son útiles para responder preguntas como ¿cuánta carga resistirá tal o cual material hasta fallar? La teoría en la que se basa este tipo de análisis asume eventos que se puedan definir adecuadamente en momentos específicos, en este contexto el fallecimiento constituye un "evento" definido sin ambigüedades en el sentido de que un solo evento sucede en cada individuo. Pese a que hay casos que relajan este supuesto y permiten múltiples "eventos" por individuo, esta tesis dedica especial atención al primer caso.

Los métodos de análisis tradicionales no son aplicables en este tipo de datos debido a que (comúnmente) este tipo de estudios involucra variables positivas que por lo general presentan asimetría. Otra característica muy común es la "censura". Se dice que una observación está censurada cuando se sabe que excede (o está por debajo de) cierto umbral pero no se sabe cuanto. Para lidiar con estas peculiaridades, modelos específicos han sido desarrollados. En las secciones siguientes se detallan algunas definiciones, modelos y aspectos inferenciales aplicados de este tipo de datos.

4.1. Definiciones

Antes de introducir los elementos básicos con las que se trabajará en este apartado, es necesario definir la variable aleatoria de interés, a la que se llamará T. En el caso del análisis de sobrevida dicha variable aleatoria debe ser estrictamente positiva, medida a partir de un punto de origen y hasta un final (que se denominará *evento*) bien determinados y con una cierta escala (generalmente temporal). Sin embargo, puede ocurrir que la variable T se observe parcialmente debido a alguna de las siguientes situaciones:

- individuos que comienzan su participación en el estudio de manera tardía,
- individuos que dejan de ser observados durante el período de seguimiento,
- individuos que al final del período de seguimiento, aún no hayan experimentado el evento.

Todos estos casos son distintos tipos de censura (el primero a la izquierda y los últimos dos a la derecha) y deben ser tenidos en cuenta de manera adecuada en el análisis.

En esta tesis sólo se trabajó con situaciones de censura a la derecha, para hacer frente a esta situación se introdujo la siguiente notación. Sea $T = \text{mín}(T^*, C)$, donde T es el tiempo efectivamente observado, T^* es el tiempo de sobrevida (sea este observado por el investigador o no) y C es el tiempo hasta el momento de la censura. Adicionalmente se introduce la variable δ que indica si el dato de un individuo ha sido observado ($\delta = 1$) o si corresponde a una censura ($\delta = 0$). Estas cantidades son de particular utilidad al construir la función de verosimilitud de una muestra.
El objetivo primordial de estudio en este tipo de análisis es la llamada función de sobrevida:

$$S(t) = P(T > t). \tag{4.1}$$

En la gran mayoría de las aplicaciones se supone que $S(0) = 1$, lo cual indica que la probabilidad de sobrevivir al inicio es cierta. Es trivial notar que si $u > t$ entonces $S(u) \leq S(t)$. Adicionalmente a la función de sobrevida, también se suele trabajar con la función de riesgo $h(t)$, la cual denota la probabilidad

instantánea de que un individuo experimente un evento, condicional a haber sobrevivido t unidades de tiempo, la cual tiene la forma:

$$\begin{aligned} h(t) &= \lim_{dt\to 0} \frac{P(t \leq T < t+dt | T > t)}{dt} \\ &= \lim_{dt\to 0} \frac{P(t \leq T < t+dt)}{dtS(t)} = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} \end{aligned} \tag{4.2}$$

donde $f(t)$ es la función de densidad de la variable T. La función de riesgo (también llamada "fuerza de la mortalidad") es no negativa en todo el recorrido de la variable aleatoria ($h(t) \geq 0 \quad \forall \quad t > 0$), pero a diferencia de la función de sobrevida puede ser creciente, decreciente o ni siquiera ser monótona, de esta manera representa una cantidad mucho más flexible a la hora de modelar la variable T.
Otra manera de vincular estas dos funciones es la siguiente:

$$S(t) = e^{-\int_0^t h(s)ds}. \tag{4.3}$$

A partir de esta ecuación se define la función de riesgo acumulado:

$$\Lambda(t) = \int_0^t h(s)ds. \tag{4.4}$$

Una última relación que se puede obtener entre $f(t)$, $S(t)$ y $h(t)$ es la siguiente:

$$f(t) = h(t)S(t). \tag{4.5}$$

Utilizando las relaciones indicadas en (4.3)) y (4.5), es posible construir de la función de verosimilitud de una muestra aleatoria simple. Para ello se utilizará (para cada individuo) la pareja de variables aleatorias (T_i, δ_i). Utilizando esta notación, se puede afirmar que la contribución a la verosimilitud del i-ésimo individuo, cuyo evento ha sido observado, es $f(t)$ mientras que en el caso de un individuo cuyo tiempo corresponde a una censura, su contribución es $S(t)$. De esta manera, la función de verosimilitud de una muestra de n individuos es:

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{4.6}$$

siendo t_i el tiempo en el que el individuo i experimenta el evento o, en el caso de que el individuo haya sido censurado, el último instante donde se tuvo contacto con él. Haciendo uso de las relaciones (4.3) y (4.5), se puede expresar del siguiente modo:

$$L = \prod_{i=1}^{n} h(t_i)^{\delta_i} e^{-\int_0^{t_i} h(s)ds}. \tag{4.7}$$

Esta última expresión resulta de mayor practicidad debido a que sólo requiere expresar la función de riesgo, la cual (como se mencionó anteriormente) permite mayor flexibilidad que $f(t)$ o $S(t)$.

4.2. Modelos

En los modelos de regresión para datos de sobrevida, el punto de atención se centra en el proceso de envejecimiento al que se someten los objetos que se siguen a través del tiempo. A diferencia de los modelos convencionales de regresión, donde la componente sistemática modela cambios en la media de la distribución, en los modelos de sobrevida, dicho componente suele ser utilizado para describir la función de riesgo. Adicionalmente se debe tener en cuenta la posible presencia de datos censurados, ya que, pese a que en primera instancia esto sugiere un problema al cual se le debe prestar especial atención, puede generar conclusiones muy interesantes desde el punto de vista del investigador. En los siguientes apartados se describen dos metodologías que plantean, desde diferentes perspectivas, modelos capaces de describir variaciones en la función de riesgo.

A continuación se presentan el modelo semiparamétrico de Cox (1972) y el modelo paramétrico de Weibull. Estos parten de la base de especificar la función de riesgo del i-ésimo individuo de la siguiente manera:

$$h(t; x_i) = h_0(t)c(x_i^{'}\beta) \tag{4.8}$$

siendo x_i el vector de covariables del i-ésimo individuo y β un vector de parámetros. Ambas funciones deben ser elegidas de manera que $h(t; x_i) > 0$. Nótese que cuando $c(x\beta) = 1$ entonces $h(t; x_i) = h_0(t)$, por lo cual esta última suele ser llamada función de riesgo de referencia. La característica más sobresaliente de estos modelos (por la cual adoptan su nombre) se basa en el

cociente de la función de riesgo de dos individuos. Este cociente, denominado cociente de riesgos (CR) se calcula de la siguiente manera:

$$CR(t; x_1; x_2) = \frac{h_0(t)c(x_1\beta)}{h_0(t)c(x_2\beta)} = \frac{c(x_1\beta)}{c(x_2\beta)} \tag{4.9}$$

aquí se puede ver que el riesgo de un individuo con respecto a otro es una función independiente del tiempo (al menos en el caso de que las covariables sean fijas). Por este motivo suelen llamarse *modelos de riesgo proporcional*. Ambas metodologías se basan en el supuesto de "riesgos proporcionales" y se diferencian en el tratamiento de la función de riesgo de referencia. Mientras que el modelo Cox deja sin especificar al componente $h_0(t)$, el modelo Weibull lo especifica incluyendo un parámetro más a la distribución.

4.2.1. Modelo semiparamétrico

Tal como se estableció anteriormente, en los modelo de sobrevida, la clave se encuentra en especificar la función de riesgo. De manera más concisa, se puntualizó que el modelo de Cox asume que dicha función adopta la siguiente especificación:

$$h(t; x_i) = h_0(t)e^{x_i\beta} \tag{4.10}$$

dejando sin especificar la función $h_0(t)$, debido a esta elección en particular, la literatura clasifica a este modelo como "semiparamétrico". Al calcular el logaritmo de la función de riesgo presentada en (4.10) se puede ver que:

$$\log h(t; x_i) = \log h_0(t)e^{x_i\beta} = \log h_0(t) + x_i\beta = \eta(t) + x_i\beta. \tag{4.11}$$

Que es una función que sigue las mismas reglas básicas de los modelos lineales. Por este motivo, la codificación de las variables y la inclusión de interacciones se realizan de la misma manera. Y la interpretación de los coeficientes refleja el impacto de aumentar en una unidad cada covariable sobre el logaritmo del riesgo en el momento t. En cuanto al CR, es fácil notar que bajo esta especificación, adopta la siguiente forma:

$$CR(t; x_1, x_2) = e^{\beta(x_1 - x_2)}. \tag{4.12}$$

En particular si una de las covariables indica la pertenencia de un individuo a un grupo o tratamiento, e^{β} se interpreta como el riesgo de experimentar el

evento de interés con respecto a un individuo que pertenezca al grupo complentario.

Hasta aquí se ha prestado especial atención a la función de riesgo, pero en la práctica, la función de supervivencia juega un rol más importante debido a que es más sencilla de interpretar y permite una visualización más clara del fenómeno bajo estudio. Gracias a las ecuaciones presentadas en la sección 4.1 la función de sobrevida del i-ésimo individuo es:

$$S(t; x_i) = e^{-\int_0^t h_0(u)e^{x_i\beta}du} = e^{-e^{x_i\beta}\int_0^t h_0(u)du} = S_0(t)^{e^{x_i\beta}}. \tag{4.13}$$

4.2.2. Modelo paramétrico

En el apartado anterior se presentó el caso en el que la función de riesgo de referencia se deja sin especificar. Se plantearán ahora las características correspondientes al caso de que la distribución de los tiempos de sobrevida es específica en su totalidad. Este caso merece especial atención debido a que en las ocasiones en que esta familia de modelos proporcionan un buen ajuste, las estimaciones que proporcionan suelen ser más precisas. Otra de las ventajas de estos modelos es que además de la formulación manejada hasta aquí, también admiten la parametrización de tiempo acelerado (AFT por sus siglas en inglés) la cual postula que:

$$S(t; x_i, \theta) = S_0(te^{x_i\theta}) \tag{4.14}$$

es decir, la sobrevida de un individuo en el momento t (con covariables x_i) es la misma que la de un individuo de referencia en el momento $te^{x_i\theta}$ y la cantidad $e^{x_i\theta}$ es llamada "factor de aceleración".
La representación clásica de estos modelos es mediante la siguiente ecuación:

$$\log T_i = x_i\alpha + \sigma\epsilon_i \tag{4.15}$$

donde β y σ son parámetros a estimar y ϵ_i es el error que introduce aleatoriedad al modelo, cuya distribución y parámetros se asumen conocidos. Elecciones típicas para esta distribución suelen ser la logística, que implica una distribución log-logística para T y la distribución estandar de valores extremos, la cual acarrea la distribución Weibull para T. Para describir la función de riesgo del modelo Weibull, existen diversas alternativas en la literatura, la presentada aquí corresponde al caso de definir $p = \frac{1}{\sigma}$, de esta

manera:

$$h(t; x_i, \alpha, \lambda, p) = \lambda p t^{p-1} e^{-\frac{x_i' \alpha}{p}} \tag{4.16}$$

siendo esta la representación denominada "tiempo acelerado". Sin embargo, al definir $\beta = -p\alpha$ se obtiene la representación de riesgos proporcionales:

$$h(t; x_i, \beta, \lambda, p) = \lambda p t^{p-1} e^{x_i' \beta} \tag{4.17}$$

donde $\lambda e^{\beta_0} t^{\lambda-1}$ no es más que la función de riesgo de referencia. Pese a que σ está asociado a la varianza de la distribución, la literatura hace referencia al mismo como el "parámetro de escala", mientras que se refiere a λ como un parámetro de forma. En última instancia, la función de sobrevida de T adopta la siguiente parametrización:

$$S(t; x_i, \beta, \lambda) = e^{-t^p e^{x_i \beta}}. \tag{4.18}$$

Bajo esta parametrización y análogo al significado que β tiene para el cociente de riesgos, e^{θ} representa el cociente de percentiles del tiempo de sobrevida para individuos con distinto valor en una covariable.

4.2.3. Covariables cambiantes en el tiempo

Todo lo expuesto anteriormente es válido en un marco donde las covariables se mantienen incambiadas a lo largo del período de estudio. Sin embargo, pueden existir situaciones donde las covariables cambien a lo largo del tiempo, afectando diferencialmente a la función de riesgo. Pese a que este nuevo escenario generaliza el planteo inicial de los modelos de riesgo proporcional, las modificaciones que implica sobre la formulación del modelo (y su función de verosimilitud) no son de gran complejidad. Desde un punto de vista meramente operacional, la modificación no involucra alterar el modelo sino que pasa por alterar los datos *adecuadamente* para tener en cuenta la naturaleza cambiante de las covariables y luego utilizar el modelo de riesgo proporcional sobre estos datos modificados.

A modo de ejemplo, considere un individuo con un tiempo de sobrevida de 8 unidades ($T_i = 8$ y $\delta_i = 1$) y supongamos que su función de riesgo dependiera de una covariable que hasta el momento 5 adoptara el valor 0 y que luego valiera 1, esto es: $X_i(t) = 0$ si $t \leq 5$ y $X_i(t) = 1$ si $t > 5$. Para tener en cuenta este cambio, se crean 2 versiones del individuo i de forma tal

que la primera tenga una sobrevida de 5 unidades ($T_{i1} = 5$), con una covariable constante a lo largo de ese período ($X_{i1} = 0$) y una segunda versión del mismo individuo con una sobrevida de 3 unidades ($T_{i2} = 3$), cuya covariable valga 1 a lo largo de ese período($X_{i2} = 1$). Adicionalmente, como la adición de ambas versiones es igual al individuo original, se debe tener en cuenta que la segunda versión es la continuación de la primera, por lo tanto la primera versión corresponderá a una censura ($\delta_{i1} = 0$) mientras que la segunda será la que experimente el evento ($\delta_{i1} = 1$).

4.3. Inferencia

En cuanto a la estimación de estos modelos, se debe hacer la distinción entre el caso paramétrico y el semiparamétrico debido a que en este último, la dependencia de la verosimilitud de la función de riesgo, plantea un inconveniente importante. Es por esto que Cox planteó la función de verosimilitud parcial (Cox, 1975) como una manera de eludir este obstáculo y estimar los parámetros que modifican la función de riesgo de referencia, sin tener que estimar la función propiamente dicha. En el caso del modelo paramétrico este problema no se presenta debido a que la función de riesgo queda totalmente especificada una vez que se escoje una distribución para los datos. Por lo tanto, los parámetros de estos modelos se estiman mediante técnicas convencionales de máxima verosimilitud.

4.3.1. Estimación

En el caso del modelo de Cox, la idea detrás del método de verosimilitud parcial (véase anexo A.3) está en plantear la probabilidad de la ocurrencia de un evento en el momento t_i dado que alguno de los individuos experimenta dicho evento. Esto es:

$$\mathcal{PL}(\beta) = \prod_{i=1}^{n} \frac{h_0(t_i)e^{x_i\beta}}{\sum_{j\in R(t_i)} h_0(t_j)e^{x_j\beta}} = \prod_{i=1}^{n} \frac{e^{x_i\beta}}{\sum_j e^{x_j\beta}}. \tag{4.19}$$

Donde se debe tener en cuenta que la suma planteada en el denominador abarca a los $R(t_i)$ individuos cuya sobrevida supera el momento t_i. Finalmente, el vector β es estimado mediante técnicas numéricas de optimización. En el caso del modelo paramétrico (más concretamente el modelo Weibull)

la función de verosimilitud es la siguiente:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} \underbrace{(\lambda e^{\beta_0} t_i^{\lambda-1} e^{-x_i\beta})}_{h_0(t_i)}^{\delta_i} \underbrace{e^{-t^\lambda e^{\lambda x_i \beta}}}_{S(t_i)} \tag{4.20}$$

que también es maximizada numéricamente.
En ambos casos, la varianza de los estimadores es aproximada utilizando la diagonal de la inversa de la matriz de información de Fisher, denotada por $I(\beta)$ (negativa de la matriz Hessiana de la verosimilitud (parcial)).

4.3.2. Inferencia sobre β

A la hora de realizar pruebas de hipótesis o intervalos de confianza sobre las estimaciones de los coeficientes de regresión, se pueden utilizar las mismas herramientas presentadas en el apartado de inferencia de la sección de datos longitudinales. Por lo cual, para llevar a cabo la siguiente prueba de hipótesis:

$$\begin{array}{ll} H_0) & \beta = \underline{\beta_0} \\ H_1) & \beta \neq \underline{\beta_0} \end{array}$$

siendo $\underline{\beta}$ un elemento o un conjunto de elementos del vector β, es válido utilizar estadísticos de Wald, de cociente de verosimilitud o verosimilitud perfil.

4.3.3. Estimación de la función de riesgo de referencia

En el ámbito del modelo de Cox, una vez estimados los coeficientes de regresión, puede surgir interés en estimar la función de riesgo de referencia $h_0(t)$. A partir de dicha estimación es que se construyen curvas de sobrevida, posiblemente ajustadas por los valores de las covariables. En este sentido, la literatura describe dos alternativas, el estimador de Breslow (1972) y el propuesto por Kalbfleisch and Prentice (1973).

El estimador de Breslow surge de maximizar la verosimilitud completa con respecto a $h_0(t)$ remplazando β por su estimación $\hat{\beta}$. De esta manera el estimador es:

$$\hat{h}_{0,B}(t_i) = \frac{1}{\sum\limits_{j \in R_{(t_i)}} exp(x_j^{'}\hat{\beta})} \tag{4.21}$$

Dicho estimador lleva a la siguiente expresión de la función de sobrevida de referencia:

$$\hat{S}_{0,B}(t_i) = \prod_{i|t_i<t} \left[1 - \frac{1}{\sum_{j \in R_{(t_i)}} exp(x_j^{'} \hat{\beta})} \right]. \tag{4.22}$$

La gran desventaja de este estimador es que puede llegar a adoptar valores negativos. En dichos casos se suele sustituir su valor por cero.

El estimador propuesto por Kalbfleisch y Prentice parte del supuesto de que $S_0(t)$ es una función decreciente que presenta saltos en los tiempos observados $t_1 < t_2 <, \ldots, < t_n$. Al sustituir $S(t|x)$ por $S_0(t)^{x\beta}$ dentro de la verosimilitud y maximizarla con respecto de dicha cantidad se obtiene el siguiente estimador:

$$\hat{S}_{0,KP}(t_i) = \prod_{i|t_i<t} \left[1 - \frac{x_i{}' exp(\hat{\beta})}{\sum_{j \in R_{(t_i)}} exp(x_j^{'} \hat{\beta})} \right]^{exp(-x_i{}' \hat{\beta})}. \tag{4.23}$$

Este estimador tiene la desventaja de que no se puede generalizar al caso de que haya tiempos iguales.

Capítulo 5

Datos Faltantes

Los estudios longitudinales suelen encontrarse con el problema de datos faltantes (*missing data*), esto significa que las variables de interés pueden no ser registradas en los individuos en algunas de las ocasiones planificadas según el diseño del estudio. Esto puede suceder por diversos motivos, ya sea porque los individuos no son encontrados en alguna ocasión en particular, por eventos adversos (como ser enfermedades), por razones administrativas, por falta de cooperación o simplemente porque los sujetos abandonan el estudio a partir de un cierto momento. Este es un caso especial de datos faltantes, al que la literatura de análisis de datos longitudinales le ha prestado especial atención y es denominado deserción (*drop-out*). Verbeke and Molenberghs (2000) detallan cuatro razones por las cuales se debe recalcar el análisis de la deserción en este tipo de estudios.

1. La clasificación del proceso generador de la falta de datos tiene una interpretación mucho más sencilla que en otras formas de pérdida de datos.

2. Formular modelos en el contexto de la deserción es más sencillo.

3. Gran parte de la literatura referente a datos faltantes en el entorno del análisis de datos longitudinales se restringe a este caso particular.

4. La deserción es, por amplio margen, el caso de dato faltante más común en estudios longitudinales.

5.1. Métodos ad-hoc

El objetivo de cualquier análisis suele ser obtener estimaciones insesgadas de ciertos parámetros que describan alguna característica de la población bajo estudio. Una vez obtenidas estas estimaciones, el investigador suele calcular algún indicador de la precisión de los estimadores. En última instancia se realizan pruebas de hipótesis, intervalos de confianza, con la mayor potencia posible manteniendo el nivel de significación nominal. Los datos faltantes son capaces de alterar todas estas etapas debido a que pueden presentar los siguientes problemas:

- sesgar las estimaciones,
- disminuir la precisión de las estimaciones,
- disminuir la potencia de los test y
- alterar el nivel de significación.

Tratando de solucionar estos problemas, se han propuesto diversos métodos que si bien solucionan alguno de estos inconvenientes, por lo general agravan otros. Una primera aproximación al análisis, incorporando al análisis los datos faltantes, podría ser mediante alguno de los siguientes métodos:

- análisis de datos completos (ADC)
- imputación simple (IS)
- imputación condicional (IC)
- acarreo del último dato (AUD).

Debido a que la descripción de estos métodos excede al propósito de esta tesis, a continuación se presenta una muy breve descripción de los mismos. Pese a que todos estos métodos proveen soluciones sencillas al problema bajo estudio, se basan en supuestos que por lo general son poco realistas o proveen inferencias sesgadas.

En el caso del ADC, el análisis basado únicamente en individuos que completaron el estudio implica en primer lugar una gran pérdida de información, mientras que los resultados sólo se refieren a una parte de la población por

lo cual se introducen severos problemas de sesgo.

En los casos de métodos de imputación, IS o IC, se utiliza algún procedimiento para sustituir los valores faltantes, sin embargo Dempster and Rubin (1987) aclaran que estos métodos pueden ser *seductores y peligrosos* debido a que pueden crear la ilusión de que, luego del proceso de imputación, los datos están completos y que tanto las observaciones originales como las imputadas proveen la misma información.

Por último el AUD, es un caso especial de imputación en el que el último dato disponible de un individuo perdido por seguimiento se utiliza para imputar las mediciones correspondientes a las siguientes instancias hasta el fin del estudio. Claramente este supuesto es poco realista en tanto supone que el perfil de aquellos individuos que desertan del estudio se mantiene incambiado en el tiempo.

Estos métodos pueden ser categorizados como procedimientos *ad-hoc*. Como se mencionó antes, todos ellos se basan en supuestos fuertes y poco realistas y su principal propósito es el de lidiar con la situación de datos faltantes más que de aprovechar la posible información extra que puedan suministrar. Es por esto que la idea detrás de las metodologías desarrolladas recientemente es formular modelos que tengan en cuenta el proceso generador de la deserción y su relación (o la falta de esta) con la variable analizada longitudinalmente.

5.2. Procesos generadores de datos faltantes

En el apartado anterior se mencionó que una de las principales debilidades de los procedimientos allí descritos es que no aprovechan la información contenida en los datos faltantes y su posible relación con la variable bajo estudio. Con el fin de ahondar en esta relación resulta necesario introducir los tres mecanismos de datos faltantes mencionados por Rubin (1976). Sin embargo, en primer lugar se mencionarán algunas definiciones que permitirán allanar el camino hacia la clasificación de Rubin.

Sea r_i un vector cuyas componentes (r_{ij}) adoptan el valor 1 cuando el individuo i es observado en la ocasión j y 0 cuando no es observado en dicha

ocasión. La definición de este vector es necesaria para particionar el vector de observaciones y_i de la forma (y_i^O, y_i^M), siendo y_i^O el sub-vector asociado a las componentes de r_i donde se registran valores iguales a uno (datos observados), mientras que y_i^M es el sub-vector correspondiente a los ceros de r_i (datos faltantes).
Se puntualizó que en estudios longitudinales es común que los datos faltantes se den a través de la deserción de los individuos, por esto, también suele definirse la variable d_i que indica el número de mediciones efectivamente observadas del sujeto i.

Para definir la clasificación de Rubin es importante aclarar que la misma refiere a la siguiente factorización de la distribución conjunta de r_i y y_i:

$$p(r_i, y_i|x_i, \theta) = p(y_i|x_i, \theta)p(r_i|y_i, x_i, \theta). \tag{5.1}$$

Los mecanismos generadores de datos faltantes se refieren al modelo probabilístico detrás del vínculo entre el vector de datos faltantes r_i y la variable de respuesta y_i. La idea de la clasificación de estos mecanismos es la factorización de la distribución condicional de r_i dado el vector (y_i^O, y_i^M).

5.2.1. Datos Perdidos Completamente al Azar (*Missing Completely at Random (MCAR)*)

En este caso la factorización de la ecuación (5.1) es la siguiente:

$$p(r_i|y_i^O, y_i^M, \theta) = p(r_i|\theta) \tag{5.2}$$

Este mecanismo asume que la distribución de r_i no guarda relación alguna con la variable de respuesta y_i, definiendo así, una situación de independencia entre los vectores r_i y y_i. La característica más importante de $MCAR$ es que el sub-vector y_i^O puede ser considerado como una muestra aleatoria de los datos completos Y_i, lo cual implica que los datos observados pueden ser considerados como una muestra de la población bajo estudio. De esta manera, el resultado de asumir $MCAR$, es que las inferencias realizadas utilizando los datos disponibles son válidas para toda la población sin tener en cuenta en ningún momento el proceso generador de datos faltantes.

5.2.2. Datos Perdidos al Azar (*Missing at Random (MAR)*)

En este caso la factorización de la ecuación (5.1) es la siguiente:

$$p(r_i|y_i^O, y_i^M, \theta) = p(r_i|y_i^O, \theta) \tag{5.3}$$

y r_i se asume independiente de las mediciones no observadas y_i^M dada la información contenida en y_i^O. En esta factorización, al permitir que r_i dependa de los valores observados y_i^O se está incurriendo en un supuesto menos restrictivo que el postulado bajo $MCAR$.

Dado que la única diferencia entre los mecanismos $MCAR$ y MAR es la dependencia de r_i en los valores observados, uno podría poner a prueba la hipótesis de que suponer $MCAR$ es razonable contra la alternativa de que MAR sea el enfoque adecuado (véase Hedeker and Gibbons (2006)). Sin embargo, debido al hecho de que el mecanismo de datos faltantes depende de y_i^O, la distribución de y_i^O no es la misma que la de Y_i por lo cual, los datos observados no pueden considerarse una muestra aleatoria de los datos completos y por ende tampoco conforman un muestra aleatoria de la población como sí es el caso de $MCAR$.

No obstante, la distribución de los valores faltantes y_i^M condicionada a los valores observados y_i^O concide con su contraparte poblacional. De esta manera, los valores faltantes pueden ser predichos de una manera válida utilizando los datos observados asumiendo un modelo correctamente especificado para el vector (y_i^O, y_i^M). De este modo surge que los análisis llevados a cabo en un contexto MAR pueden proveer inferencias válidas aún si se ignora la contribución a la verosimilitud de r_i. Para aclarar esto último, considérese la siguiente factorización de la verosimilitud:

$$\begin{aligned} \mathrm{L}(\theta) &= \int p(y_i, r_i \,|\theta) dy_i^M \\ &= \int p(y_i^O, y_i^M \,|\theta) p(r_i \,|y_i^O, y_i^M, \theta) dy_i^M \\ &= \int p(y_i^O, y_i^M \,|\theta) p(r_i \,|y_i^O, \theta) dy_i^M \\ &= p(y_i^O \,|\theta_y) p(r_i \,|y_i^O, \theta_r) \\ &= L(\theta_y) L(\theta_r) \end{aligned} \tag{5.4}$$

Donde θ_y y θ_r son los parámetros asociados a los modelos longitudinales y de pérdida respectivamente. Más aún, en este caso se da que el espacio

paramétrico del vector θ equivale al producto cartesiano de los espacios paramétricos de θ_y y θ_r respectivamente. Inferencias correspondientes a θ_y pueden realizarse sobre la distribución marginal observada $p(y_i^O\,|\theta)$ ignorando la verosimilitud del proceso generador de los datos faltantes. Esta propiedad, mediante la cual las inferencias basadas en la verosimilitud construida bajo MAR, son válidas, es conocida como *ignorabilidad*.

En el caso de los modelos mixtos, utilizados para modelar las trayectorias de los individuos en estudios longitudinales, el análisis correspondiente es realizado bajo MAR siempre y cuando el modelo longitudinal esté correctamente especificado, es decir, que se hayan incluído todas las covaraibles correctas y que se haya modelado la estructura de covarianza de la variable de respuesta adecuadamente.

5.2.3. Datos Perdidos no al Azar (*Missing not at Random (MNAR)*)

Finalmente, cuando la distribución condicional de r_i también incluye a y_i^M (o al menos alguno de sus elementos) el mecanismo es llamado $MNAR$. Al igual que en el caso MAR, bajo $MNAR$ los datos observados no constituyen una muestra de la población objetivo. Por otro lado, al contrario de MAR, la distribución predictiva de y_i^M condicional a y_i^O no coincide con la poblacional ya que adicionalmente depende de $p(y_i\,|r_i)$. Por este motivo, la correcta especificación del proceso de pérdida de datos es crucial y debe ser incluída en la verosimilitud. De esta manera resulta claro que el mecanismo de pérdida más complejo para trabajar es $MNAR$, sin embargo no plantea supuestos restrictivos ni poco reales.

Cuando los datos longitudinales surgen de un mecanismo $MNAR$, la validez del proceso inferencial está ligada a que se modeló adecuadamente la distribución conjunta de y_i y r_i. En la literatura se pueden distinguir tres tipos de modelos: modelos de selección, modelos de mezcla de patrones y modelos de parámetros compartidos.

1. **Modelos de selección.**
   Esta familia de modelos se caracteriza por realizar la siguiente factorización de (5.1):

$$p(y_i, r_i|\theta) = p(y_i|\theta_y)p(r_i|y_i, \theta_r). \tag{5.5}$$

Estos modelos fueron introducidos por Heckman (1976) en la literatura econométrica y su nomenclatura se basa en que mediante la distribución de r_i condicional a la trayectoria descrita por y_i se puede pensar que cada individuo *selecciona* probabilísticamente si deserta del estudio o si continúa en el mismo. Su uso en el análisis de datos longitudinales se debe principalmente, al trabajo de Diggle and Kenward (1994).

2. **Modelos de mezcla de patrones.**
   En estos modelos propuestos por Little (1993), la distribución conjunta de y_i y r_i adopta la siguiente forma:

$$p(y_i, r_i|\theta) = p(y_i|r_i, \theta_y)p(r_i|\theta_r). \tag{5.6}$$

   Como puede verse, estos modelos se basan en la factorización opuesta a la de los "modelos de selección". Tal como lo indica su nombre, estos modelos permiten modelar la distribución de y_i de distinta manera en cada *patrón* de datos faltantes. En el caso de que los datos faltantes correspondan únicamente a deserciones, cada patrón indicaría el momento de la deserción de cada individuo. De esta manera, la distribución marginal de y_i corresponde a una mezcla probabilística con pesos dados por la distribución marginal de cada *patrón*.

3. **Modelos de parámetros compartidos.**
   Por último, los "modelos de parámetros compartidos" (véase Wu and Bailey, 1988) se basan en la idea de que existe un proceso latente (descrito a través de efectos aleatorios) que dictamina los valores observados en y_i y r_i. La ecuación correspondiente a este modelo es:

$$p(y_i, r_i|\theta) = \int p(y_i|b_i, \theta_y)p(r_i|b_i, \theta_r)p(b_i\theta_b)db_i. \tag{5.7}$$

   Así, para un valor dado de los efectos aleatorios, los procesos de pérdida y de medición se consideran independientes.

Capítulo 6

Análisis conjunto de datos longitudinales y de sobrevida

En el capítulo Antecedentes se mencionaron varias investigaciones donde se generaron tanto datos longitudinales (mediante la repetición de ciertas mediciones) como datos de tiempo hasta cierto evento. La gran mayoría del trabajo mencionado en dicho apartado se basa en casos particulares donde metodologías específicas se desarrollaron para cada caso debido a que cada uno de ellos perseguía objetivos ligeramente diferentes. Sin embargo Henderson et al. (2000) resumen los objetivos de estos y otros estudios en tres categorías:

1. Ajustar las inferencias referentes al proceso longitudinal de manera de permitir una posible dependencia del proceso de sobrevida.
2. Ajustar la distribución hasta el tiempo de falla en función de las variaciones del proceso longitudinal.
3. Caracterizar la evolución conjunta de ambos procesos.

La metodología que se presenta en este apartado trata de cumplir con estos tres objetivos partiendo de la base de la maximización de la verosimilitud conjunta entre los dos procesos involucrados. Esta estrategia de estudio resulta mucho más eficiente a la hora de llevar a cabo los análisis ya que los datos del proceso longitudinal se usan simultáneamente con los del proceso de sobrevida. En este sentido es de esperar que se obtengan estimaciones más precisas de la intensidad de la asociación entre ambos procesos.

En el apartado referente al modelo se podrá observar que esta asociación se basará en la estructura de los efectos aleatorios incluídos en el modelo longitudinal. También se presentará una característica muy apropiada de estos modelos donde en los casos en los que no exista asociación entre las dos variables, el análisis será el equivalente a llevar a cabo un análisis de sobrevida por un lado y un análisis de datos longitudinales por otro, de manera independiente.

6.1. Formulación del modelo

Tal como se mencionó anteriormente, la idea principal detrás de los modelos conjuntos, es la de acoplar el modelo de sobrevida con el modelo longitudinal. El objetivo específico de esta familia de modelos es medir la asociación entre el nivel de la variable registrada longitudinalmente (libre de ruido) con el riesgo de experimentar el evento de interés en cada momento del tiempo, teniendo en cuenta a su vez, la posible influencia de un conjunto de covariables. Para una mejor descripción matemática, Rizopoulos (2012b) propone utilizar la siguiente denominación del modelo longitudinal:

$$y(t_{ij}) = m(t_{ij}) + \epsilon(t_{ij}) \tag{6.1}$$

donde $m(t_{ij})$ es la estimación del nivel de la variable longitudinal en el momento t_{ij} para el i-ésimo individuo y $\epsilon(t_{ij})$ es un componente de ruido que se desea remover de la medición $y(t_{ij})$.

El punto de partida del modelo conjunto serán los modelos presentados en los capítulos 3 y 4, por lo cual este capítulo utilizará la misma notación. Se denotará por T_i y δ_i al tiempo de sobrevida del individuo i así como al indicador de que dicho tiempo sea una observación o una censura respectivamente. En el caso del proceso longitudinal se denotará por $y(t)$ al valor observado en el momento t. Debe hacerse la aclaración de que $y(t)$ no se observa en todos los posibles valores de t sino en un sub-conjunto $\{t_{ij}\}$, ocasiones que no necesariamente serán para cada sujeto del estudio. De esta manera el conjunto de valores longitudinales de cada sujeto será $Y_i = \{y(t_{i1}), y(t_{i2}), \ldots, y(t_{in_i})\}$.

La información longitudinal es obtenida intermitentemente y con un componente de error. Es debido a esto, que la tarea del sub-modelo longitudinal

consiste en estimar el valor del proceso libre de ruido $m(t)$ para cualquier valor de t de manera de poder reconstruir toda la historia del proceso longitudinal. Con este fin se especifican los componentes del modelo longitudinal (de componentes normales) presentado en (3.3) presentado en el capítulo 3:

$$\begin{cases} y(t) = m(t) + \epsilon(t) \\ m(t) = x_L(t)\beta_L + z(t)b \\ b \sim N\left(0, D(\gamma)\right) \\ \epsilon(t) \sim N\left(0, \sigma^2\right) \end{cases} \tag{6.2}$$

donde una vez estimados los efectos fijos β y predichos los efectos aleatorios b, se podrá contar con la predicción del componente $m(t)$ para cualquier momento. Este sencillo modelo longitudinal es capaz de separar el ruido incluido en las mediciones del verdadero valor el proceso y reconstruir completamente la evolución de $m(t)$ gracias a la parametrización temporal impuesta en las covariables $x_L(t)$ y $z(t)$.

Por otro lado, la parametrización del sub-modelo de sobrevida (de riesgo proporcional) incluirá el historial del proceso $m(t)$ de la siguiente manera:

$$h(t|m(t), x_S(t)) = h_0(t)exp\left\{x(t)\beta_S + \alpha m(t)\right\}, \tag{6.3}$$

Nótese cómo se optó por diferenciar las covariables x_S y x_L. Las primeras refieren a las involucradas en el sub-modelo longitudinal mientras que las segundas son las intervinientes en el sub-modelo de sobrevida. La misma convención se utilizó en los vectores de efectos fijos β_S y β_L.

Bajo esta nueva especificación, el parámetro α cuantifica el efecto del verdadero valor (no observado) del proceso longitudinal sobre el riesgo de experimentar el evento en el momento t. La interpretación tanto de α como de β_S es la misma que la explicada en la sección 4 donde e^{α} y e^{β_S} representan el cambio en el cociente de riesgos por cada unidad de incremento en $x_S(t)$ y $m(t)$ respectivamente. Debe notarse como bajo esta parametrización, el riesgo de experimentar el evento sólo depende del valor actual de $m(t)$. Sin embargo esto no se cumple para la función de sobrevida, véase que en dicho caso se observa que:

$$S(t|m(t), x_S(t)) = exp\left(\int_0^t h_0(s)exp\left\{x_S(s)\beta_S + \alpha m(s)\right\} ds\right) \tag{6.4}$$

lo que implica que la función de sobrevida depende de toda la historia de $m(t)$. Bajo esta especificación, sólo la función de riesgo depende del valor de $m(t)$ en el instante t, pero este supuesto puede no ser muy realista. Tenga en cuenta el caso en el que el riesgo del evento depende del valor puntual de $m(t)$ y de la tendencia experimentada por este último o por el historial completo de las fluctuaciones experimentadas por este. En dichos casos, es sensato pensar en la siguiente parametrización del sub-modelo de sobrevida:

$$\begin{array}{lcl} h(t|m(t), x_S(t)) & = & h_0(t)exp\left\{x_S(t)\beta_S + \alpha_0 m(t) + \alpha_1 f_1(m(t)) + \alpha_2 f_2(m(t))\right\} \\ f_1(m(t)) & = & m^{'}(t) \\ f_2(m(t)) & = & \int_0^t m(s)ds \end{array} \tag{6.5}$$

donde $f_1(m(t))$ y $f_2(m(t))$ no son más que la derivada evaluada en t y la integral hasta t de la estimación de $m(t)$. De incluir estos términos, los coeficientes α_1 y α_2 miden la intensidad de la asociación del riesgo con la pendiente y el historial de la verdadera trayectoria longitudinal.

Finalmente se debe agregar que en el contexto del modelo conjunto, no es conveniente dejar sin especificar la función de riesgo de base. Esto se debe a que, como notó Hsieh et al. (2006), esta elección puede acarrear una subestimación de los errores estándar de las estimaciones, que puede llevar a inferencias equívocas. Por este motivo, en el caso de que el sub-modelo de sobrevida no tenga una distribución específica (por ejemplo Weibull) es preferible especificar algún tipo de parametrización alternativa para $h_0(t)$. La más comunmente utilizada es la siguiente función de riesgo en escalera (RE):

$$h_0(t) = \sum_{q=1}^{Q} \xi_q I\left(v_{q-1} < t_q\right) \tag{6.6}$$

donde $0 = v_0 < v_1 < v_2 < \ldots, v_Q$ es un conjunto de nodos (típicamente especificados por el investigador o equiespaciados entre el menor y el mayor valor de los eventos) que dividen la escala temporal y ξ_q indica el valor de la función de riesgo en el intervalo $(v_{q-1}, v_q]$. Nótese cómo al incrementar el número de nodos, aumenta la flexibilidad de esta función.

6.2. Estimación

A continuación se detallan los pasos consistentes en la estimación de los parámetros del modelo conjunto mediante el método de máxima verosimilitud[1]. Este método consiste en la búsqueda del modo de la log-verosimilitud de la distribución conjunta de $\{Y_i, T_i\}$. Para definir esta distribución conjunta Rizopoulos (2012b) asume que el vector de efectos aleatorios b_i está presente tanto en el proceso longitudinal como en el de sobrevida. Por este motivo, los efectos aleatorios son los responsables de la correlación entre las variables aleatorias involucradas. De este modo, al asumir que los dos procesos involucrados se acoplan mediante los efectos aleatorios, se postula el supuesto de independencia de las variables Y_i y T_i condicional al valor de b_i. Así, la distribución conjunta se factoriza de la siguiente manera:

$$\begin{aligned} \mathrm{p}(\mathrm{T}_i, Y_i|b_i;\theta) &= p(T_i,|b_i;\theta)p(Y_i|b_i;\theta) \\ &= p(T_i,|b_i;\theta)\prod_{j=1}^{n_i} p(y_i(t_{ij})|b_i;\theta) \end{aligned} \tag{6.7}$$

donde θ contiene los parámetros del sub-modelo longitudinal (fijos y aleatorios) y los del sub-modelo de sobrevida. Bajo la formulación (6.7) se asume que condicional a los datos, la censura y los tiempos de visita son independientes de los tiempos observados y de las futuras mediciones del proceso longitudinal. A efectos prácticos esto significa que el hecho de que un sujeto abandone el estudio depende sólo de su "historial" en el estudio.

La contribución del i-ésimo sujeto a la log-verosimilitud será:

$$\begin{aligned} \log p(T_i, Y_i;\theta) &= \log \int p(T_i,|b_i;\theta)p(Y_i|b_i;\theta)db_i \\ &= \log \int p(T_i,|b_i;\theta_t)\left[\prod_{j=1}^{n_i} p(y_i(t_{ij})|b_i;\theta_y)\right] p(b_i|\theta_b)db_i \end{aligned} \tag{6.8}$$

donde θ_t son los parámetros de sobrevida, θ_y son los parámetros de efectos fijos del sub-modelo longitudinal y θ_b son los parámetros de covarianza

[1]Debido a que el modelo conjunto contiene un sub-modelo de riesgo proporcional, el método puede ser semiparamétrico (en el caso del modelo de Cox) o paramétrico (en el caso Weibull)

del sub-modelo longitudinal. En última instancia se define cada uno de los componentes de la ecuación (6.8) de la siguiente manera:

$$\begin{aligned} p(T_i, |b_i; \theta_t) &= [h_0(T_i) exp\{x_S(T_i)\beta_S + \alpha m_i(T_i)\}]^{\delta_i} \\ &\quad \times \exp\left(\int_0^{T_i} h_0(s) \exp\{x_S(T_i)\beta_S + \alpha m_i(s)\}\, ds\right) \\ \prod_{j=1}^{n_i} p(y(t_{ij})|b_i; \theta_y) &= (2\pi\sigma^2)^{-\frac{n_i}{2}} \exp\left\{-\frac{\sum_j (y(t_{ij}) - x_L(t_{ij})\beta_L - z(t_{ij})b_i)^2}{2\sigma^2}\right\} \\ p(b_i|\theta_b) &= (2\pi)^{-\frac{q_b}{2}} |D(\gamma)|^{-\frac{1}{2}} \exp\left\{-b_i' D(\gamma)^{-1} b_i\right\} \end{aligned} \tag{6.9}$$

En la ecuación (6.9) $h_0(t)$ denota una función de riesgo de base que puede ser RE o la perteneciente a cualquier modelo paramétrico, $x_S(t)$ y $x_L(t)$ son las matrices de efectos fijos de los sub-modelos de sobrevida y longitudinales respectivamente, $z_i(t)$ es la matriz correspondiente a los efectos aleatorios y $D(\gamma)$ es la matriz de covarianza de los efectos aleatorios.

La maximización de la función de log-verosimilitud se realiza mediante métodos numéricos, los usados más frecuentemente son el algoritmo EM Dempster et al. (1977) o variantes del método de Newton-Raphson, Nocedal and Wright (2006). Sin embargo, el primero es el más utilizado (véase anexo A.5), dado que convenientemente trata a los efectos aleatorios como "datos faltantes" y provee predicciones de los mismos como un resultado adicional del proceso de estimación y que en el paso M, algunos de los estimadores tienen una expresión cerrada.

Antes de finalizar este apartado debe hacerse una breve precisión respecto de las integrales presentes en la verosimilitud. Tanto en la definición de la función de riesgo como en el *score* (utilizado en el paso M de algoritmo EM) de la log-verosimilitud aparecen integrales, la primera se hace respecto al tiempo y la segunda respecto de los efectos aleatorios. Ambas integrales no tienen soluciones analíticas salvo algunos casos puntuales por lo cual, estas integrales se aproximan numéricamente para cada individuo, haciendo que la estimación de estos modelos sea una tarea sumamente exigente desde un punto de vista computacional.

De las dos integrales involucradas la que suele ser más compleja es la integral

respecto de los efectos aleatorios debido a que suele ser multidimensional, la integral respecto del tiempo en la definición de la función de sobrevida siempre es unidimensional por lo que puede ser aproximada fácilmente mediante reglas de Gauss-Kronrod (Hazewinkel, 2001). La integral multidimensional puede aproximarse por cuadratura gaussiana o métodos de Monte Carlo, sin embargo Ye et al. (2008) proponen utilizar la transformada de Laplace en el ámbito del modelo conjunto debido a que constituye un aumento en la eficiencia del algoritmo. De todos modos, el tiempo de cómputo a la hora es estimar estos modelos es un aspecto a tener en cuenta. Una idea propuesta por Rizopoulos (2012a) consiste en estimar el modelo longitudinal antes del modelo conjunto y utilizar las predicciones de los efectos aleatorios como insumo para escalar las integrales involucradas en el *score* de la verosimilitud del modelo conjunto.

6.3. Inferencia

6.3.1. Pruebas de hipótesis

Una vez estimado el modelo mediante máxima verosimilitud, los mismos procedimientos descritos en las secciones 3.3 y 4.3 se pueden aplicar del mismo modo, tanto el estadístico LRT como las pruebas de Wald ya sea re-estimando el modelo bajo las restricciones que se deseen poner a prueba y comparando los valores de la verosimilitud (en el caso LRT) o comparando las estimaciones de los coeficientes de regresión con sus errores estandar (en el contexto del estadístico de Wald). El único parámetro extra que aparece en estos modelos es α. Tal vez la hipótesis más importante a contrastar sea:

$$\begin{array}{ll} H_0) & \alpha = 0 \\ H_1) & \alpha \neq 0 \end{array}$$

donde la hipótesis nula implica la ausencia de asociación entre los procesos longitudinal y de sobrevida.
El caso de los elementos de la matriz D (parámetros correspondientes a los efectos aleatorios), los estadísticos LRT se utilizan con los mismos cuidados descritos en la sección 3.3.

6.3.2. Intervalos de confianza

Los intervalos de confianza de los parámetros del modelo conjunto (β_S y β_L) son calculados mediante la aproximación asintótica de los estimadores a la distribución normal en la que se basa el estadístico de Wald. De esta manera, los intervalos se construyen utilizando las estimaciones y los errores estándar ($\hat{\theta} \pm z_{1-\frac{\alpha}{2}} s\hat{e}.(\hat{\theta})$). Del mismo modo se pueden construir intervalos para la media del proceso longitudinal, es decir:

$$X\hat{\beta} \pm z_{1-\frac{\alpha}{2}} \left\{ diag \left[X\widehat{\mathbb{V}ar}(\hat{\beta})X' \right] \right\}^{\frac{1}{2}} \tag{6.10}$$

6.3.3. Predicción de efectos aleatorios

Hasta este punto nos hemos enfocado en la realización de inferencias sobre los efectos fijos del modelo conjunto. Pese a que los efectos aleatorios b_i fueron introducidos al modelo con el propósito de tener en cuenta la heterogeneidad entre los perfiles individuales de los individuos (y para modelar la asociación entre los procesos involucrados en el modelo conjunto), puede que el énfasis del estudio esté sobre la predicción individual de futuras mediciones del proceso longitudinal, para lo cual la predicción de los efectos aleatorios es de vital importancia. Debido al caracter aleatorio de estas cantidades, lo más natural es predecirlas bajo un enfoque bayesiano (Rizopoulos, 2012b, cap. 4.5). De esta manera, la distribución a posteriori de los efectos aleatorios será:

$$P(b_i|T_i, Y_i; \theta) = \frac{p(T_i|b_i; \theta)p(Y_i|b_i; \theta)p(b_i; \theta)}{p(T_i, Y_i; \theta)}. \tag{6.11}$$

Contrario a lo que sucede en el caso de modelos mixtos gaussianos, la distribución en (6.11) no pertenece a la familia gaussiana y ni siquiera posee una "forma cerrada" por lo cual, su estudio es realizado mediante métodos numéricos. Sin embargo debe realizarse la aclaración que en el caso de que el número de mediciones por individuo (n_i) tienda a infinito, esta distribución a posteriori tenderá a una distribución normal. Típicamente, suele caracterizarse esta distribución o bien mediante su media o mediante su modo.

$$\begin{aligned} \bar{b_i} &= \int b_i p(b_i|T_i, Y_i; \theta) db_i \\ \hat{b_i} &= \arg\max_b \log p(b_i|T_i, Y_i; \theta) \ . \end{aligned} \tag{6.12}$$

La elección del modo frente a la media responde a lo indicado anteriormente, debido a que la distribución a posteriori de los efectos aleatorios no necesa-

riamente es normal, es posible que presente asimetrías, por lo cual, la media puede no ser la mejor medida de resumen. En cuanto a la variabilidad de estas mediciones, es común aproximar su varianza mediante las fórmulas indicadas en la ecuación (6.13).

$$\begin{aligned} \widehat{\mathbb{V}ar}(b_i) &= \int (b_i - \bar{b_i})^2 p(b_i|T_i, Y_i; \theta) db_i \\ H_i &= \left\{ -\frac{\partial^2 \log p(b_i|T_i,Y_i;\theta)}{\partial b' \partial b}|_{b=\hat{b_i}} \right\}^{-1} . \end{aligned} \tag{6.13}$$

Tanto para el cálculo de las cantidades en (6.12) como en (6.13) se sustituye a θ por $\hat{\theta}$.

6.4. Utilidad en el caso de datos faltantes

Dada la conexión entre los dos procesos involucrados en el modelado conjunto de los datos, el supuesto del investigador es que la ocurrencia de un evento implica la discontinuación (si el evento es el fallecimiento) o al menos un cambio en la distribución de la variable longitudinal. Debido a este enunciado es que se pueden conectar las ideas sobre mecanismos generadores de datos faltantes de la sección 5.2 con la teoría detrás del modelo conjunto.

Es importante notar que dada la formulación del modelo conjunto, el submodelo mixto asume una distribución para todas las posibles mediciones de la variable medida repetidamente, esto significa que (al menos en principio) la distribución del vector completo Y_i es válida tanto para los valores observados como para los faltantes. En la sección 5.2 se descompuso al vector Y_i en dos componentes, y_i^O y y_i^M. En este caso en particular $y_i^O = \{y_i(t_{ij}) : t_{ij} < T_i^*, j = 1, 2, \ldots, n_i\}$ representa las mediciones efectivamente observadas del individuo i, y $y_i^M = \{y_i(t_{ij}) : t_{ij} \geq T_i^*, j = 1, \ldots, n_i\}$ contiene las mediciones que se habrían observado del mismo individuo si no hubiese experimentado el evento. Bajo estas definiciones se puede explicitar el mecanisno de pérdida de datos, como la distribución de la variable T^* condicional a los elementos y_i^O y y_i^M.

$$\begin{aligned} p(T_i^*|y_i^O, y_i^M; \theta) &= \int p(T_i^*, b_i|y_i^O, y_i^M; \theta) db_i \\ &= \int p(T_i^*|b_i, y_i^O, y_i^M; \theta) p(b_i|y_i^O, y_i^M; \theta) db_i . \\ &= \int p(T_i^*|b_i; \theta) p(b_i|y_i^O, y_i^M; \theta) db_i \end{aligned} \tag{6.14}$$

La simplificación del último paso se debe al supuesto de independencia condicional entre T y Y (véase el apartado de estimación). Finalmente, se puede

observar cómo el tiempo transcurrido hasta la ocurrencia del evento depende tanto de y_i^O como de y_i^M a través de la distribución a posteriori de los efectos aleatorios. Por este motivo, el modelo conjunto pertenece a la familia de modelos $MNAR$.

6.5. Análisis de sensibilidad

Uno de los problemas más importantes a la hora de evaluar los resultados del modelo conjunto reside en el hecho de que no es posible poner a prueba el supuesto MAR respecto de $MNAR$. De hecho, Molenberghs et al. (2008) demostraron que, dado un conjunto de datos, cualquier modelo $MNAR$ puede llegar al mismo valor de la verosimilitud que su equivalente MAR pero las inferencias provistas por dichos modelos, pueden ser sustancialmente diferentes. En este sentido, ya que los datos no pueden diferenciar entre ambos mecanismos de pérdida, la única manera práctica de evaluar los supuestos en este ámbito es a través de un análisis de sensibilidad (Diggle et al., 2007).

En primer lugar, debe hacerse mención de los distintos tipos de análisis de sensibilidad que pueden llevarse a cabo. En el análisis de sensibilidad global se investiga una amplia clase de modelos donde la idea es determinar cuánto hay que alejarse de los supuestos del modelo bajo estudio para que las inferencias cambien. Por otro lado, el análisis de sensibilidad local se basa en evaluar el cambio en las inferencias en un "vecindario" cercano a los supuestos del modelo bajo estudio (Daniels, 2008).

Por lo general, el primer paso a la hora de llevar a cabo este análisis es comparar las estimaciones del modelo $MNAR$ con su equivalente MAR (sensibilidad global), que en este contexto implica comparar los resultados del modelo conjunto con el modelo mixto. Luego, la propuesta consiste en evaluar la estimación del vector β_L mediante un indicador de sensibilidad (local) al supuesto de ignorabilidad, este indicador es el $ISNI$ por su sigla en inglés (*Index of local sensitivity to non-ignorability*). Pese a que este estimador fue propuesto por Troxel et al. (2004) en el contexto de los modelos de selección, su adaptación al caso de los modelos conjuntos es simple. El propósito de este índice es cuantificar que tanto varían las estimaciones del sub-modelo longitudinal al alejarse de la hipótesis de ignorabilidad del mecanismo de pérdida. La idea detrás del $ISNI$ es medir el cambio en los parámetros al

alejarse del supuesto de MAR ($\alpha = 0$), esto es:

$$ISNI = \frac{\partial}{\partial \alpha} \hat{\beta(\alpha)} \mid_{\alpha=0} . \quad (6.15)$$

En Rizopoulos (2012b) se aclara que al no poder calcularse analíticamente el valor exacto del $ISNI$, se propone una aproximación de segundo orden de la log-verosimilitud del modelo conjunto (véase anexo A.2). A partir de la misma se deriva el siguiente estimador:

$$ISNI \approx - \left\{ \frac{\partial^2}{\partial \beta^T \partial \beta} \ell(\theta)|_{\theta=\theta^{(0)}} \right\}^{-1} \left\{ \frac{\partial^2}{\partial \beta^T \partial \alpha} \ell(\theta)|_{\theta=\theta^{(0)}} \right\} \quad (6.16)$$

siendo $\theta^{(0)}$ el vector que incluye a todos los parámetros del modelo conjunto estimado bajo el supuesto de MAR. De esta manera, el cálculo indicado en (6.16) requiere la Hessiana[2] del modelo bajo la restricción $\alpha = 0$.

Una vez calculado el indicador, algunos autores (Ma et al., 2004; Viviani, 2012), proponen normalizarlo con respecto a la magnitud de la estimación o con respecto a la estimación del desvío del parámetro correspondiente. De esta manera se logran $ISNI$s relativos que permiten evaluar sensibilidad sin tener en cuenta la unidad de medida de las variables. Troxel considera que valores del $INSI$ (relativo al desvío estándar) mayores que uno, indican cierta sensibilidad al mecanismo de pérdida. Otra alternativa, propuesta por Viviani considera que valores mayores a 0.5 indican cierta sensibilidad en el parámetro.

[2]En el caso de que el modelo incluya tanto un término con α_1 como otro con α_2, se calculó la Hessiana bajo la restricción $\alpha_1 = \alpha_2 = 0$

Capítulo 7

Selección de modelos

En los apartados de inferencia (tanto de análisis longitudinal como de sobrevida) se propusieron diversas alternativas para seleccionar entre distintas estructuras para los modelos, donde se hizo especial hincapié en el estadístico LRT. El mismo permite llevar a cabo una estrategia para seleccionar entre distintos modelos que estén "anidados", es decir, seleccionar entre dos modelos donde uno de ellos sea un caso especial (restringido) del otro. Sin embargo, en algunos casos es importante seleccionar entre modelos que no presentan esta jerarquía. Esto es sencillo de ver en el caso que se quiera comparar un modelo longitudinal con un modelo conjunto, ya que ninguno es un caso restringido del otro, por lo tanto el estadístico LRT no puede ser utilizado. Es por esto que en la literatura se han propuesto diversos indicadores de bondad de ajuste de modelos, siendo los más comunes el AIC propuesto por Akaike (1974) y el BIC propuesto por Schwarz (1978).

- AIC
  La definición del indicador es la siguiente:

  $$AIC = -2(\mathcal{L}(\theta) - p) \tag{7.1}$$

  Siendo p el número de parámetros en el modelo. Básicamente, este indicador penaliza el valor de la verosimilitud del modelo con el número de parámetros estimados, intentando de esta manera, seleccionar el modelo de mejor ajuste (mayor valor de la log-verosimilitud) pero que a la vez sea lo más sencillo posible (menor número de parámetros).

- BIC

El caso del BIC es similar al anterior, pero con una penalización diferente:

$$BIC = -2(\mathcal{L}(\theta) - p\log(N)) \tag{7.2}$$

La idea detrás de este indicador es penalizar de manera más severa la inclusión de más parámetros en el modelo. Sin embargo debe aclararse que en el caso de modelos mixtos la definición de N no es sencilla ya que uno de los supuestos de partida es que los datos se jerarquizan en (al menos) dos niveles, habiendo de esta manera n individuos con n_i observaciones cada uno. Pese a que no existe un acuerdo global en este asunto se suele definir a N como el número de individuos utilizados en el ajuste del modelo. En el caso de que el método de estimación haya sido $REML$, N será igual al número de individuos menos la cantidad de contrastes utilizados para remover la influencia de los parámetros de la media.

A modo de comentario final es importante notar que los modelos a comparar deben haber sido ajustados sobre el mismo conjunto de datos, este punto también debe ser tenido en cuenta al utilizar estos indicadores. Por último, es necesario enfatizar que estos indicadores no deben ser utilizados como una prueba formal de significancia de un modelo frente a otro, sino que simplemente son reglas prácticas para seleccionar modelos.

Capítulo 8

Aplicación a un estudio longitudinal

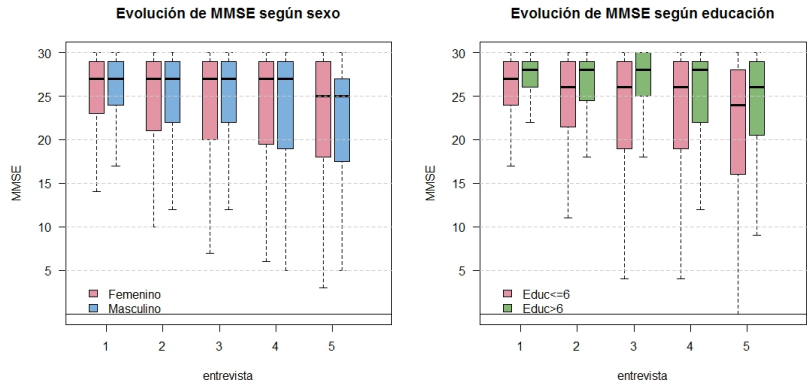
En esta tesis se presenta una aplicación sobre los datos del estudio "*Origins of Variance in the Oldest-Old: Octogenarian Twins*" ($OCTO - Twin$). El mismo se compone de 351 parejas de mellizos de 80 años o más, seleccionados del registro sueco de mellizos. El estudio tenía como propósito investigar las causas de las diferencias individuales en este conjunto de personas ancianas en características que iban desde salud, capacidad funcional, funcionamiento cognitivo, hasta bienestar psicológico entre otras.

El estudio comenzó en 1991 y la cohorte involucrada fue estudiada hasta 1999. En dicho período se realizaron entrevistas cada 2 años aproximadamente. En esta tesis se presentan datos donde cada participante presenta información parcial o completa sobre su estado cognitivo. Esta información fue relacionada con el sexo, la escolaridad y la edad de los individuos, con el fin de determinar diferentes patrones de deterioro cognitvo.

8.1. Análisis exploratorio inicial

Al inicio del estudio, la muestra consistió de 702 individuos (234 hombres y 468 mujeres) cuyas edades variaban entre los 79 y 98 años con una edad mediana de 86 años. En cuanto a la educación de los mismos, se detectó que la gran mayoría contaba con 6 años de educación, razón por la cual se optó por dividir la muestra en 2 grupos; "6 años de educación o menos" y "más de 6

años". Como parte del análisis descriptivo se investigó la posible asociación del deterioro cognitivo de estas personas con el sexo y el nivel educativo. Con este fin se construyeron gráficos de caja para cada visita, diferenciando cada subpoblación. Esto se expone en las Figuras 8.1a y 8.1b.



(a) Según sexo (b) Según educación

Figura 8.1: Distribución del $MMSE$

En la Figura 8.2 se pueden apreciar algunos de los perfiles de evolución del $MMSE$ a través del tiempo. En cuanto a la posible asociación con el nivel educativo, la Figura 8.2 muestra algunos patrones pero, en primera instancia no se logra observar ningún comportamiento evidente, por lo cual, esto fue estudiado en la etapa de modelación. En cuanto al mecanismo de pérdida de datos, en la Figura 8.3 se puede observar como el paso del tiempo (medido en el número de visitas por sujeto) tuvo un impacto importante en la cantidad de información disponible. Se puede observar que el 85 % de los individuos completaron la segunda visita, 62 % alcanzaron la tercera, 45 % la cuarta y tan sólo el 32 % de la muestra cumplió con las cinco visitas. Vale mencionar que de las 660 pérdidas que se produjeron, 4 correspondieron a abandonos y el resto fueron fallecimientos. Al final del período de seguimiento sobrevivieron 42 individuos. Para finalizar la descripción inicial de los datos se presenta el cuadro 8.1, el cual contiene el promedio de $MMSE$ de los individuos agrupados según el número total de visitas (en las columnas) en cada una de las evaluaciones que se les realizaron (presentado en las filas). Asimismo se expone la edad promedio de cada grupo al inicio del estudio para facilitar la comparación. En primera instancia se puede ver que hay una leve

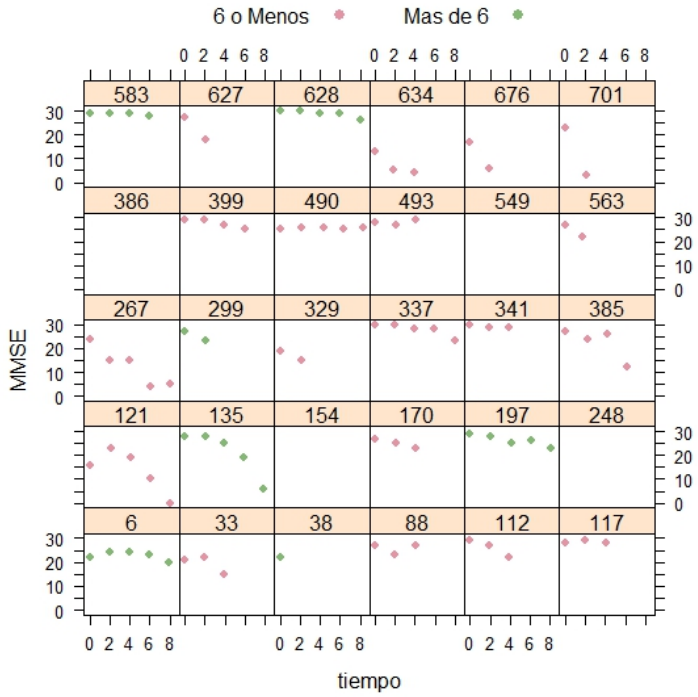


Figura 8.2: Evolución del $MMSE$ según educación

Cuadro 8.1: $MMSE$ promedio según número total de visitas y momento de evaluación

	Total de visitas				
Evaluación	1	2	3	4	5
primera	19.98	23.80	24.60	24.29	27.10
segunda	–	19.36	22.36	23.51	26.79
tercera	–	–	18.92	20.95	26.35
cuarta	–	–	–	17.10	24.95
quinta	–	–	–	–	22.00
Edad	84.62	85.03	85.60	86.66	86.82
n	102	166	119	93	222

tendencia que indica que las personas mayores son las que permanecieron más tiempo en el estudio. Pese a que esto parezca contradictorio, más adelante se verá que se debe a las condiciones en las que estas personas llegan al estudio. En este cuadro se puede ver que si bien las personas que lograron ser evaluadas una sola vez eran las más jóvenes, su $MMSE$ era el más bajo,

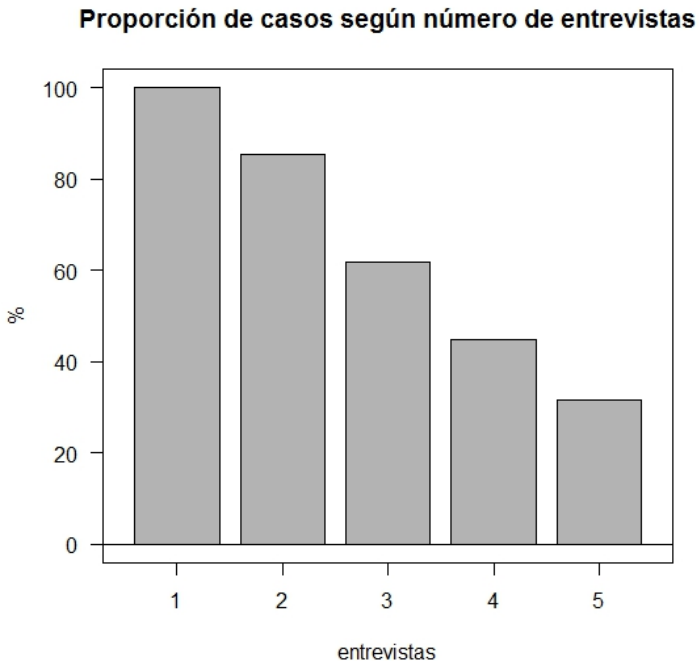


Figura 8.3: Disminución del tamaño muestral

por lo cual ya presentaban un deterioro avanzado. Por otro lado, las personas que completaron las cinco visitas, pese a ser las de mayor edad, no mostraban señales de deterioro al principio del estudio (figura 8.4).

8.2. Estrategia de Análisis y Estimación

8.2.1. Estrategia de Análisis

La estimación del modelo conjunto, requiere de tres insumos fundamentales que son: la especificación de un modelo longitudinal que describa la trayectoria de la medida bajo estudio (en este caso el $MMSE$), un modelo de sobrevida que describa la probabilidad de deserción del estudio de cada individuo (en este caso debido al fallecimiento) y un proceso que vincule ambos modelos. Para la elaboración de este modelo conjunto, se optó por seguir los siguientes pasos:

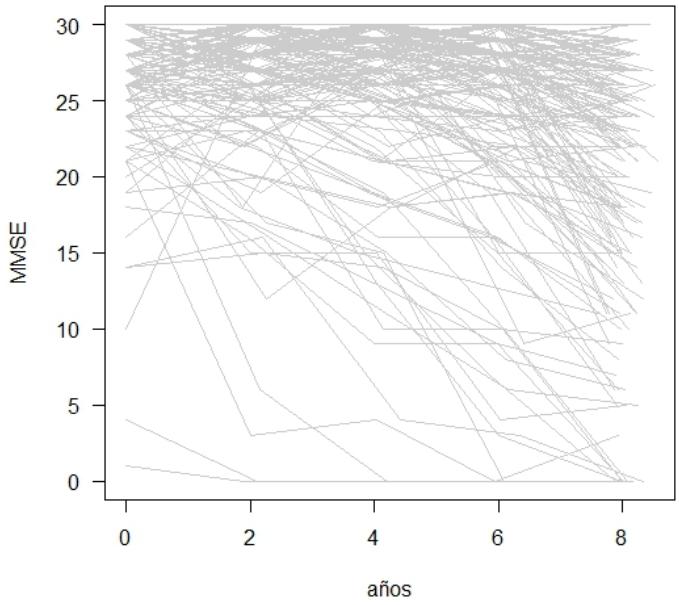


Figura 8.4: Evolución del MMSE de los sujetos que completaron el estudio

1. Construcción del modelo longitudinal.
   En esta etapa se planteó la construcción de un modelo que describa el crecimiento (o decrecimiento) del $MMSE$ a través de una trayectoria lineal o cuadrática. Dicha trayectoria se midió a partir del comienzo del estudio y se ajustó por la edad de los individuos en dicho momento. La significación de los coeficientes del modelo fue puesta a prueba mediante pruebas de cociente de verosimilitud (LRT).

   Asimismo se trató de determinar si factores como el sexo y la educación inciden en dicha trayectoria. La elección de estos factores responde a que las mujeres tienden a presentar una mayor sobrevida que los hombres (Waldron and Johnston, 1976) por lo cual, sobre el final del estudio es probable que existan más mujeres que hombres y probablemente, dada su longevidad, su estado cognitivo sea peor. En cuanto

a la educación, varios estudios sugieren que individuos con mayor grado de educación pueden compensar los daños sufridos en el cerebro mediante lo que se conoce como *reserva cognitiva* (Stern, 2012). Simultáneamente se exploraron diferentes especificaciones en los efectos aleatorios del modelo. En este caso, las pruebas LRT fueron corregidas según el procedimiento descrito en Verbeke and Molenberghs (2000).

2. Construcción de un modelo de sobrevida.
   En esta etapa se elaboró un modelo para los tiempos de permanencia en el estudio de los individuos. El evento a modelar fue el fallecimiento, considerando como "censurados" a aquellos casos correspondientes a los individuos que sobrevivieron todo el período de seguimiento. En una etapa preliminar se construyeron curvas de sobrevida utilizando el estimador Kaplan and Meier (1958) y se investigó si existían diferencias en subpoblaciones determinadas por sexo y educación. Luego se ajustó el modelo de riesgos proporcionales (Cox, 1972) y el modelo paramétrico de Weibull. Para facilitar la comparación con el anterior, se formuló este último en su forma de riesgos proporcionales.

3. Construcción del modelo conjunto.
   La última instancia fue la correspondiente a la estimación conjunta, que permitió establecer los determinantes de la velocidad del deterioro cognitivo, teniendo en cuenta los factores que incidían sobre la deserción de los individuos.

   Para vincular los modelos longitudinales y de sobrevida se plantearon diversas especificaciones, cada una de ellas se describe en Rizopoulos (2012b). Finalmente se llevó a cabo un análisis de sensibilidad de la variación de los parámetros ante los supuestos de *Missing at Random* (supuesto detrás del modelo longitudinal) y *Missing not at Random* (supuesto detrás del modelo conjunto). En esta última instancia se utilizó el ISNI (*index of local sensitivity to non-ignorability*) como medida de sensibilidad de los parámetros del proceso longitudinal.

En la Figura 8.5 se puede ver un breve resumen de la estrategia de análisis planteada.

Antes de finalizar el apartado se deben realizar dos consideraciones. La primera es que el tiempo se midió en décadas a partir del inicio de estudio, ya que de esta forma las estimaciones de las varianzas de los efectos aleatorios se

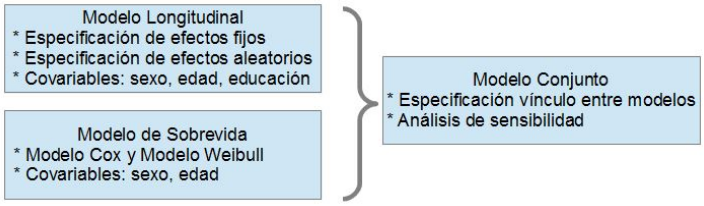


Figura 8.5: Estrategia de análisis

alejaron del borde del espacio paramétrico otorgando así forma mayor validez a las inferencias realizadas sobre las mismas. Por lo tanto, se debe tener en cuenta que para medir el cambio en el $MMSE$ por año a partir del comienzo del estudio, se debe multiplicar el valor de los coeficientes por 10. La segunda consideración, es que dado que debido a que en algunas instancias no se contaba con el dato pertinente al $MMSE$ o a la educación de los participantes, se debió reducir el tamaño muestral. Finalmente se trabajó con 2125 ocasiones relevadas sobre 688 individuos.

El análisis de los datos fue llevado a cabo en el software de uso libre R (R Core Team, 2013).

8.2.2. Estimación

8.2.2.1. Estimación del sub-modelo longitudinal

El modelo de partida utilizado para modelar la variación temporal del $MMSE$ fue el siguiente:

$$\begin{aligned} MMSE_{ij} &= \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}t_{ij}^2 + \epsilon_{ij} \qquad j = 1, \ldots, n_i \qquad i = 1, \ldots, n \\ \beta_{0i} &= \beta_{00} + \beta_{01}Sexo_i + \beta_{02}(Edad_i - 82) + \beta_{03}Educ_i + b_{0i} \\ \beta_{1i} &= \beta_{10} + \beta_{11}Sexo_i + \beta_{12}(Edad_i - 82) + \beta_{13}Educ_i + b_{1i} \\ \beta_{2i} &= \beta_{20} + \beta_{21}Sexo_i + \beta_{22}(Edad_i - 82) + \beta_{23}Educ_i + b_{2i} \\ b_i &\sim N_3(0, D) \\ \epsilon_{ij} &\sim N(0, \sigma^2) \end{aligned} \tag{8.1}$$

siendo $(Edad - 82)$ es el valor centrado de la covariable que indica la edad del individuo al momento del inicio del estudio. De esta manera β_{00} indica el valor de $MMSE$ de una mujer de 82 años con 6 años de educación o menos.

A partir de este modelo de base (Modelo 1) se ensayaron distintas alternativas. La estimación de las mismas se llevó a cabo a través de la librería *nlme* (Pinheiro et al., 2013). En el cuadro 8.2 se presentan los resultados del modelo de base (Modelo1), con efectos aleatorios independientes (Modelo 2), del modelo sin la incidencia del sexo (Modelo 3) y de un modelo sin efectos fijos en el término cuadrático (Modelo 4). Entre paréntesis figuran los p-valores asociados al estadístico de Wald.

El análisis de los efectos aleatorios fue llevado a cabo de la siguente manera. En el Modelo 1 se especificó una estructura general, con varianzas distintas para cada efecto (constante, efecto lineal y efecto cuadrático) y covarianzas libres. Luego se testeó la posibilidad de que las covarianzas fueran iguales a cero. Para esto se compararon las verosimilitudes de los Modelos 1 y 2 mediante el estadístico LRT comprobándose que el supuesto de efectos aleatorios independientes era demasiado restrictivo. Luego se puso a prueba si el efecto aleatorio correspondiente al término cuadrático era significativo, rechazando la hipótesis de nulidad de varianza. Para esta prueba se comparó el valor del estadístico LRT con el cuantil 95 de una distribución $\chi^2_{2:3}$[1].

Finalmente, se optó por trabajar con el Modelo 4 ya que el mismo poseía una estructura de efectos aleatorios adecuada, efectos fijos significativos y menores valores de BIC y AIC.

La interpretación de sus coeficientes indican que al comenzar el estudio, una persona de 82 años (sin importar su sexo) con menos de 6 años de educación tiene un valor promedio de $MMSE$ de 25.47 puntos, mientras que con más de 6 años tendría 26.75 puntos. Personas mayores de 82 años comienzan el estudio con 0.24 puntos menos por cada año de edad que supere los 82 y 0.24 más por cada año de edad menos.
Al comparar personas de baja y alta educación (sin tener en cuenta el sexo ni la edad inicial) se observa que el $MMSE$ de los individuos de baja educación desciende 3.58 puntos por década, mientras que en los de mayor educación, este puntaje se decrementa 0.69 unidades. En cuanto a la edad al inicio del estudio, se nota que cuanto mayor es la persona, menor es su puntaje inicial de $MMSE$ y más pronunciada es la pendiente que indica la velocidad de su

[1]Se denota como $\chi^2_{2:3}$ a la mezcla de dos distribuciones χ^2 con 2 y 3 grados de libertad respectivamente

Cuadro 8.2: Estimaciones e indicadores de modelos longitudinales

		Modelo 1	Modelo 2	Modelo 3	Modelo 4
Efecto	Parámetro	EMV (p-valor)	EMV (p-valor)	EMV (p-valor)	EMV (p-valor)
Constante					
Constante	β_{00}	25.27 (<0.001)	25.56 (<0.001)	25.52 (<0.001)	25.47 (<0.001)
Sexo Masculino	β_{01}	-0.15 (0.703)	-0.16 (0.721)		
(Edad - 82)	β_{02}	-0.26 (<0.001)	-0.26 (<0.001)	-0.26 (<0.001)	-0.24 (<0.001)
Educ>6	β_{03}	1.28 (<0.001)	1.30 (0.002)	1.28 (<0.001)	1.28 (<0.001)
Tiempo					
Constante	β_{10}	-6.41 (0.004)	-5.81 (0.003)	-6.26 (0.002)	-3.58 (0.012)
Sexo Masculino	β_{11}	0.37 (0.870)	1.39 (0.494)		
(Edad - 82)	β_{12}	-0.61 (0.051)	-0.44 (0.146)	-0.60 (0.052)	-0.75 (<0.001)
Educ>6	β_{13}	3.72 (0.078)	3.51 (0.065)	3.67 (0.083)	2.89 (0.020)
Tiempo^2					
Constante	β_{20}	9.12 (0.011)	7.27 (0.052)	7.34 (0.032)	
Sexo Masculino	β_{21}	-4.02 (0.146)	-4.70 (0.056)		
Edad - 82	β_{22}	-0.65 (0.570)	-0.88 (0.006)	-0.60 (0.077)	
Educ>6	β_{23}	-1.40 (0.577)	-0.61 (0.782)	-1.24 (0.620)	
Covarianza de b_i					
var(b_{i1})	d_{11}	17.04	20.93	17.04	16.99
var(b_{i2})	d_{22}	297.45	181.55	297.17	297.89
var(b_{i3})	d_{33}	205.05	24.40	206.76	200.32
cor(b_{i1},b_{i2})	$\frac{d_{12}}{\sqrt{d_{11}d_{22}}}$	0.62	0	0.62	0.61
cor(b_{i1},b_{i3})	$\frac{d_{13}}{\sqrt{d_{11}d_{33}}}$	-0.52	0	-0.53	-0.50
cor(b_{i2},b_{i3})	$\frac{d_{23}}{\sqrt{d_{11}d_{33}}}$	-0.75	0	-0.75	-0.74
Varianza residual					
var(ϵ_{ij})	σ^2	6.81	7.13	7.18	7.23
$L(y\|\theta)$		-6269.28	-6324.0	-6271.80	-6274.28
AIC		12576.56	12680.00	12575.61	12574.55
BIC		12684.15	12770.60	12666.22	12648.17

deterioro (parámetro asociado β_{12}).

En cuanto a los elementos de la matriz D se puede ver que cuando el efecto aleatorio de la constante es elevado, la caída es más pronunciada (correlaciones negativas con los efectos asociados a la pendiente y al término cuadráti-

co). En la Figura 8.6 se presentan las trayectorias esperadas del $MMSE$ de personas con diferentes edades y años de educación. Se puede observar que

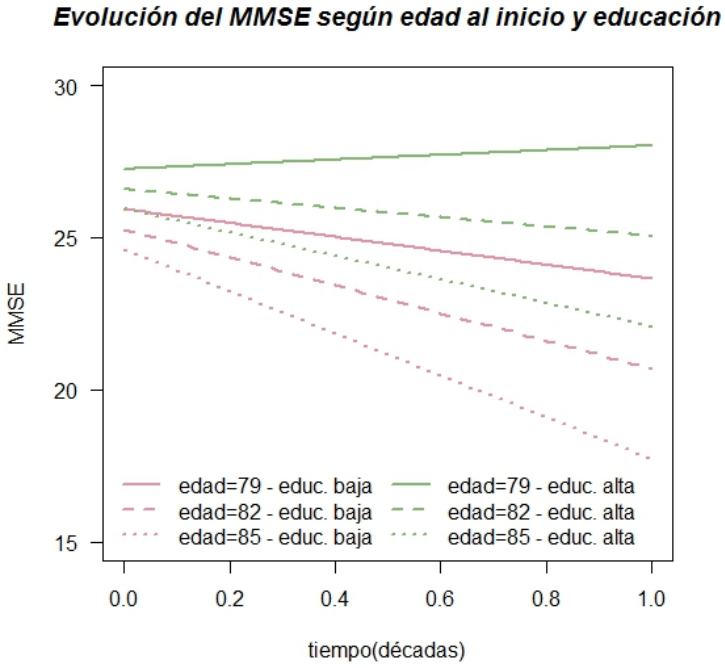


Figura 8.6: Evolución del $MMSE$ según modelo 4

la recta correspondiente a la trayectoria esperada para los individuos de 79 años con alta educación presenta una tendencia ascendente, lo cual no es compatible con una situación de deterioro. Para determinar si esta recta es creciente o no, se llevó a cabo el test F indicado en (3.9) donde se contrastó la siguiente hipótesis:

$$\begin{array}{ll} H_0) & (76-82)\beta_{12}+\beta_{13}=0 \\ H_1) & (76-82)\beta_{12}+\beta_{13}>0 \end{array} \tag{8.2}$$

El valor del estadístico F fue de 6.934, cuyo $p-valor$ es menor al 1 %. De esta manera se puede ver que el modelo estimado bajo el supuesto MAR (que no tiene en cuenta la deserción de los individuos) puede brindar resultados con sesgos muy importantes.

8.2.2.2. Estimación del sub-modelo de sobrevida

En esta etapa se analizó el componente que describe el tiempo transcurrido por cada individuo hasta que se dejan de observar sus valores de $MMSE$, ya sea debido al fallecimeinto o a la censura (finalización del período de seguimiento). Como se mencionó en la sección 8.2.1, la primera etapa de este análisis consistió en la construcción de curvas de sobrevida mediante el estimador de Kaplan and Meier (1958). La Figura 8.7 muestra la sobrevida general de todos los inidividuos sin discriminar, por sexo, edad ni educación. Tanto esta figura, como todos los análisis llevados a cabo en este apartado fueron realizados utilizando la librería *survival* (Therneau, 2012). Luego se

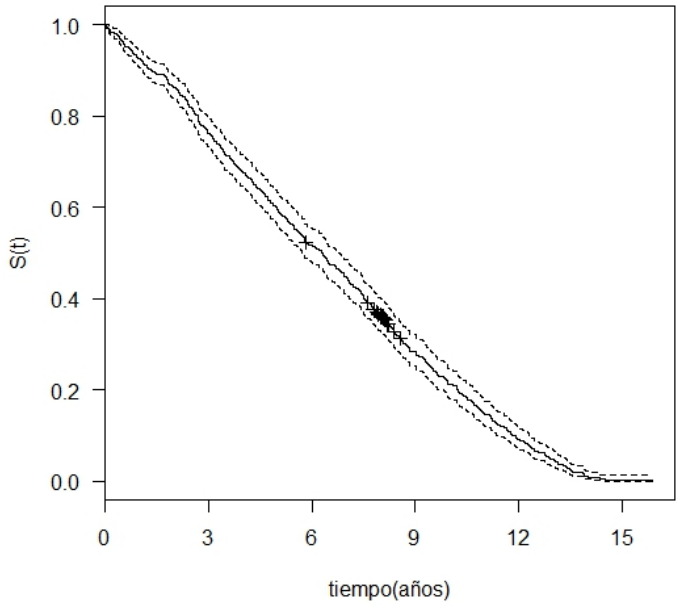


Figura 8.7: Sobrevida de los individuos

procedió a determinar si existían diferentes comportamientos en cuanto a la sobrevida entre los grupos generados por sexo y educación. Se llegó a la conclusión de que hombres y mujeres presentan diferencias en cuanto al tiempo hasta el fallecimiento sin encontrarse diferencias entre los dos grupos educativos (véase cuadro 8.3).

Cuadro 8.3: Exploración de diferencias en $S(t)$ mediante la rueba del rango logarítmico

Grupos	χ^2	g.l.	p-valor
Sexo	11.7	1	<0.001
Educación	0.2	1	0.622

Luego se estimaron distintos modelos de sobrevida que permitieran cuantificar las diferencias entre los distintos grupos. En esta etapa se consideró además, la inclusión de la edad al inicio del estudio. De esta manera se estimó el modelo de Cox y el modelo paramétrico con distribución de Weibull. En ambos casos se adicionó la variable educación con el fin de comprobar si la misma incidía en la sobrevida, al controlar por la edad al inicio del estudio. A continuación se detalla la función de riesgo de los modelos estimados.

$$h_i(t) = h_0(t) \exp\{\gamma_1 Sexo_i + \gamma_2(Edad_i - 82) + \gamma_3 Educ_i\}$$

En el caso del modelo de Cox, la función riesgo de referencia $h_0(t)$ (en este caso siendo la referencia mujeres de 82 años con baja educación) se dejó sin especificar, mientras que en el modelo Weibull se especificó mediante un parámetro de escala y otro de forma. Finalmente, se descartó incluir la educación en los modelos y se obtuvieron las estimaciones que se detallan en el cuadro 8.4.

Los resultados indican que los hombres tienen un riesgo menor de falleci-

Cuadro 8.4: Modelos de regresión

	Parámetro	Cox	Weibull[2]
Sexo Masculino	γ_1	-0.35 (<0.001)	-0.26 (0.001)
(Edad - 82)	γ_2	0.09 (<0.001)	0.07 (<0.001)
Escala	p	-	0.62 (<0.001)
Forma	λ	-	0.66 (<0.001)

miento (30 % menos según el modelo de Cox y 24 % menos según el modelo paramétrico) mientras que por cada año que el individuo excede los 82 años de edad al comienzo del estudio, su riesgo de fallecimiento aumenta un 9 %

[2]Los parámetros de este modelo están su versión PH y no AFT.

según el modelo semiparamétrico y un 7.6 % según el modelo paramétrico. A continuación se puso a prueba el supuesto de riesgos proporcionales del modelo de Cox sin rechazarce la hipótesis nula. Por último se presentan curvas de sobrevida bajo el modelo Weibull, para individuos de ambos sexos con distintas edades al inicio (Figura 8.8).

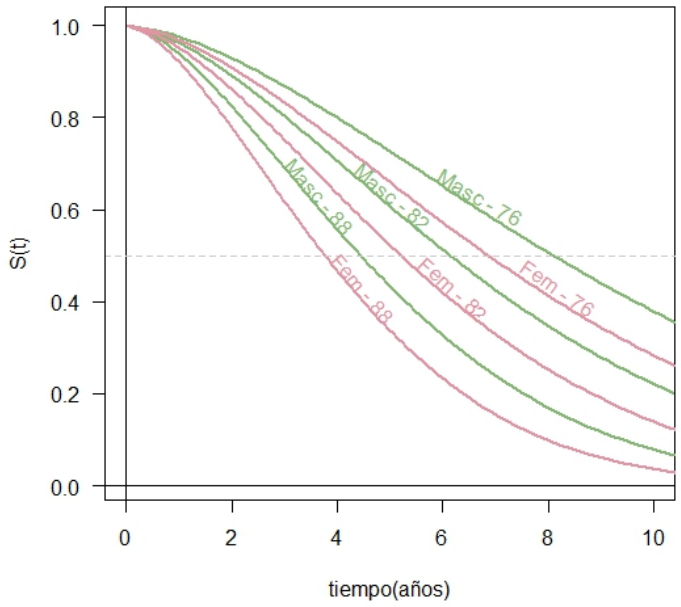


Figura 8.8: Curvas ajustadas según sexo y edad al inicio del estudio

8.2.2.3. Estimación del Modelo Conjunto

En esta etapa se investigó si las estimaciones del modelo longitudinal podían verse "alteradas" al adicionar al análisis la información contenida en el proceso de sobrevida. Para ello se ligaron los modelos finales de las secciones 8.2.2.1 y 8.2.2.2 de la siguiente manera:

$$
\begin{aligned}
h_i(t) &= h_0(t)\exp\{\gamma_1 Sexo_i + \gamma_2(Edad_i - 82) + \alpha_1 f_1(m_i(t)) + \alpha_2 f_2(m_i(t))\} \\
MMSE_{ij} &= m_i(t_j) + \epsilon_{ij} \qquad j = 1,\ldots,n_i \qquad i = 1,\ldots,n \\
m_i(t_j) &= \beta_{0i} + \beta_{1i}t_j + \beta_{2i}t_j^2 \\
\beta_{0i} &= \beta_{00} + \beta_{01}(Edad_i - 82) + \beta_{02}Educ_i + b_{0i} \\
\beta_{1i} &= \beta_{10} + \beta_{11}(Edad_i - 82) + \beta_{12}Educ_i + b_{1i} \\
\beta_{2i} &= b_{2i} \\
b_i &\sim N_3(0, D) \\
\epsilon_{ij} &\sim N(0, \sigma^2)
\end{aligned}
$$

Es claro cómo, bajo esta formulación la relación entre ambos sub-modelos se da a través de los efectos aleatorios b_i. Los parámetros α_1 y α_2 se encargan de "transportar" la información contenida en el proceso longitudinal al de sobrevida y viceversa. Es más, a través de la estimación y significación de los mismos se puede hacer una primera aproximación a la sensibilidad del modelo al supuesto de $MNAR$. Por último vale recordar que comunmente se especifica a $f_1(m(t))$ como la identidad, mientras que algunas posibles alternativas para la función $f_2(m(t))$ son:

1. Efecto rezagado: $f_2(m(t)) = m(\text{máx}(0, t-c))$.
   En este caso el valor de la función de riesgo se ve afectado por valores rezagados del proceso longitudinal "c" períodos de tiempo.

2. Efecto de la pendiente: $f_2(m(t)) = m'(t)$.
   Aquí se asume que la función de riesgo se ve alterada según el crecimiento (decrecimiento) del proceso longitudinal.

3. Efecto acumulado: $f_2(m(t)) = \int_0^t m(s)ds$.
   En esta especificación se supone que todos los valores pasados del proceso longitudinal tienen un efecto sobre el riesgo de fallecimiento.

A partir de la formulación presentada anteriormente se estimaron distintas alternativas de modelos conjuntos. Todas las estimaciones concernientes a esta etapa fueron llevadas a cabo utilizando la librería *JM* (Rizopoulos, 2010). Se comprobó que el vínculo entre ambos modelos (medido a través de los parámetros α_i) fue significativo en todas las especificaciones mientras que de las tres alternativas planteadas para $f_2(m(t))$, la que logró mejores resultados fue la correspondiente a la inclusión de la derivada de la función $m_i(t)$.

En cada una de las situaciones se llevó a cabo el análisis tanto para el sub-modelo de sobrevida con función de riesgo de referencia en escalera, como para el caso Weibull.

Para cada especificación se comprobó que el modelo RE arrojó un BIC menor al modelo Weibull. En el cuadro 8.5 se presentan los resultados del modelo longitudinal (Modelo 1), el modelo conjunto (Modelo 2), el modelo que incluye el efecto de la pendiente (Modelo 3) y el modelo que incluye el efecto acumulado (Modelo 4).

Cuadro 8.5: Estimaciones e indicadores de modelos longitudinales

Efecto	Parámetro	EMV IC 95 %	EMV IC 95 %	EMV IC 95 %	EMV IC 95 %
Sub-modelo longitudinal		Modelo 1	Modelo 2	Modelo 3	Modelo 4
Constante					
Constante	β_{00}	25.27 (24.69 ; 25.85)	24.95 (23.96 ; 25.94)	25.46 (24.47 ; 26.45)	25.44 (25.07 ; 25.81)
(Edad - 82)	β_{02}	-0.22 (-0.35 ; -0.09)	-0.21 (-0.34 ; -0.07)	-0.27 (-0.4 ; -0.14)	-0.27 (-0.33 ; -0.2)
Educ>6	β_{03}	1.32 (0.53 ; 2.11)	1.6 (0.54 ; 2.67)	1.51 (0.45 ; 2.57)	1.54 (1.06 ; 2.02)
Tiempo					
Constante	β_{10}	-4.59 (-7.58 ; -1.6)	-5.74 (-9.6 ; -1.87)	-2.25 (-6.12 ; 1.61)	-2.15 (-4.41 ; 0.11)
(Edad - 82)	β_{12}	-0.77 (-1.01 ; -0.53)	-0.75 (-0.97 ; -0.53)	-0.99 (-1.21 ; -0.77)	-0.99 (-1.20 ; -0.78)
Educ>6	β_{13}	3.05 (0.44 ; 5.67)	1.31 (-1.26 ; 3.88)	4.28 (1.71 ; 6.84)	4.39 (2.11 ; 6.67)
Sub-modelo de sobrevida					
Sexo	γ_1		-0.39 (-0.57 ; -0.22)	-0.37 (-0.54 ; -0.2)	-0.39 (-0.56 ; -0.22)
(Edad - 82)	γ_2		0.04 (0.01 ; 0.06)	0.03 (0.01 ; 0.06)	0.04 (0.01 ; 0.07)
$m_i(t)$	α_1		-0.04 (-0.06 ; -0.03)	-0.03 (-0.04 ; -0.02)	-0.08 (-0.09 ; -0.06)
$m_i^{'}(t)$	α_2			-0.01 (-0.39 ; 0.36)	
$\int^t m_i(t)$	α_2				0.09 (0.06 ; 0.12)

A través de estas opciones se pretende estudiar la sensibilidad de las estimaciones al proceso de pérdida de los datos. Puede notarse como el modelado conjunto no parece afectar la constante, sin embargo altera el efecto de la edad y la educación tanto en el nivel de base como en el efecto del tiempo. La Figura 8.9 presenta la distribución de los estimadores bajo cada una de las 4 alternativas.

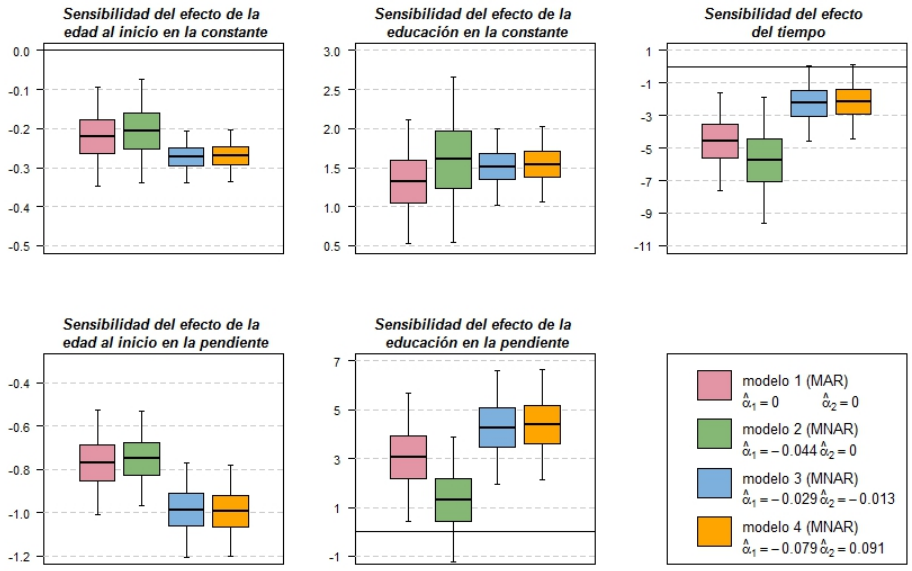


Figura 8.9: Sensibilidad de los parámetros a los distintos mecanismos de pérdida de datos

Puede verse como al suponer que la probabilidad de fallecimiento se ve afectada sólo por el valor del $MMSE$ (Modelo 2), el efecto de la educación sobre la pendiente se ve atenuado con respecto al estimado por el modelo MAR. También puede verse que en los modelos 3 y 4 este efecto se acentúa. En el caso de la edad al inicio, los modelos 3 y 4 decrementan su efecto sobre el cambio en la pendiente.

Adicionalmente a estos resultados, se llevó a cabo el mismo contraste de la sección 8.2.2.1 para determinar si, bajo un contexto $MNAR$, la pendiente de los individuos de 79 años con alta educación es o no positiva. En este caso el valor del estadístico F fue de 3.359, cuyo $p - valor$ fue de 0.067, indicando que no habría suficiente evidencia para pensar que dicha pendiente fuera positiva.

8.3. Análisis de Sensibilidad

En última instancia se trató de cuantificar el efecto del modelado conjunto en el cambio de los parámetros del modelo longitudinal. Para esto se utilizó el $ISNI$ (*index of local sensitivity to non-ignorability*) propuesto por Troxel et al. (2004). A continuación se presenta el valor del $ISNI$ y dos modificaciones del mismo para los tres modelos $MNAR$ considerados previamente. La primera alternativa consiste en dividir el valor del indicador entre el error estándar de cada parámetro (estimado bajo MAR). Rizopoulos (2012b) considera que valores mayores a uno indican gran sensibilidad. La segunda alternativa también es un índice relativo, pero en este caso, relativo al tamaño del coeficiente estimado bajo MAR.

Cuadro 8.6: Estimaciones del ISNI y sus modificaciones

	Parámetro	ISNI	$\frac{ISNI}{s.e.\beta_j}$	$\frac{ISNI}{\beta_j}$
Modelo 2				
Constante	β_{00}	26.43	89.10	1.05
(Edad - 82)	β_{01}	-0.01	-0.22	0.06
Educ6	β_{02}	-28.44	-70.68	-21.57
Tiempo	β_{10}	54.26	35.51	-11.82
(Edad - 82) (tiempo)	β_{11}	1.10	8.95	-1.43
Educ6 (tiempo)	β_{12}	-71.50	-53.66	-23.41
Modelo 3				
Constante	β_{00}	116.41	392.48	4.61
(Edad - 82)	β_{01}	10.88	168.13	-49.27
Educ6	β_{02}	-136.76	-339.81	-103.71
Tiempo	β_{10}	65.13	42.64	-14.19
(Edad - 82) (tiempo)	β_{11}	16.08	130.47	-20.91
Educ6 (tiempo)	β_{12}	-267.85	-201.01	-87.70
Modelo 4				
Constante	β_{00}	17.34	58.46	0.69
(Edad - 82)	β_{01}	4.55	70.32	-20.61
Educ6	β_{02}	-22.42	-55.72	-17.01
Tiempo	β_{10}	4.49	2.94	-0.98
(Edad - 82) (tiempo)	β_{11}	3.00	24.34	-3.90
Educ6 (tiempo)	β_{12}	-34.67	-26.02	-11.35

El cuadro 8.6 presenta los valores calculados para cada parámetro bajo cada especificación de $f_2(m(t))$ en el modelo con función de riesgo RE. En dicho cuadro se puede apreciar que el Modelo 2 ($\alpha_2 = 0$) presenta coeficientes muy sensibles salvo el correspondiente al efecto de la edad al inicio en la constante. Esta situación cambia al considerar modelos donde el riesgo de fallecimiento se ve afectado por el efecto de la pendiente de la evolución del $MMSE$ (Modelo 3) o el efecto acumulado del $MMSE$ (Modelo 4). En estos casos, todas las variables presentan una alta sensibilidad al supuesto MAR, por lo cual se concluye que el acoplamiento de ambos modelos es capaz de producir alteraciones en las estimaciones que hacen que el modelo reproduzca una situación más cercana a la realidad.

Capítulo 9

Conclusiones y Trabajos a futuro

9.1. Conclusiones

Como conclusiones globales se puede mencionar que la técnica utilizada permite modelar el proceso de envejecimiento bajo un supuesto más realista que el de MAR. También debe recalcarse que desde el punto de vista meramente estadístico, esta metodología puede utilizarse en diversos ámbitos de investigación. Finalmente, debe mencionarse que existe la oportunidad de extender estos modelos de manera de identificar subgrupos de individuos cuyas trayectorias sean similares (Muthen, 2004; Proust-Lima, 2014), ya que el modelo de efectos mixtos asume que todos los individuos siguen una trayectoria común. Los modelos que asumen la existencia de "clases" permiten relajar ese supuesto.

En cuanto a los resultados obtenidos en el estudio $OCTO - Twin$, se llegó a la conclusión de que la evolución del $MMSE$ se ve afectada tanto por la edad como por la educación de las personas. Al incluir el análisis de sobrevida como parte del proceso de deterioro cognitivo, se "corrigieron" los valores de los coeficientes estimados en el modelo especificado bajo MAR. Se observó que la sobrevida de los hombres es inferior a la de las mujeres y (como era de esperar) que a mayor edad al inicio, mayor es el riesgo de fallecimiento. Adicionalmente se observó que valores bajos de $MMSE$ incrementan el riesgo de fallecimiento. Lo mismo sucede con la pendiente en la evolución

del $MMSE$, al disminuir la misma (avance del deterioro) el riesgo de fallecimiento aumenta aún más. En cuanto al modelo longitudinal se observó que la trayectoria del $MMSE$, entre los individuos con educación baja, decrece más lentamente que lo indicado en principio por el modelo MAR, mientras que la educación tiene un efecto mayor al pensado originalmente.

9.2. Trabajos a futuro

Algunas de las posibles líneas a considerar para continuar este proyecto serían las siguientes:

1. Considerar modelos alternativos como *Pattern Mixture Models* y *Selection Models* para comparar los resultados bajo distintas parametrizaciones.
2. Complementar el análisis de sensibilidad mediante un proceso inferencial de los distintos $INSI$ considerados en este documento y con otros indicadores.
3. Introducir efectos de *ceiling* y *floor* en el modelo a través de técnicas de regresión para datos truncados.

Bibliografía

Abbatecola, A., Lattanzio, F., Spazzafumo, L., Molinari, A., Cioffi, M., Canonico, R., Dicioccio, L., and Paolisso, G. (2010). Adiposity predicts cognitive decline in older persons with diabetes: a 2-year follow-up. *PloS one*, 5(4):7.

Akaike, H. (1974). A new look at the statistical model identification. *EEE Transactions on Automatic Control*, 19(6):716–723.

Altham, P. (1984). Improving the precision of estimation by fitting a model. *Journal of the Royal Statistical Society*, 46:118–119.

Ansari, T. and Derakshan, N. (2011). The neural correlates of cognitive effort in anxiety: Effects on processing efficiency. *Biological Psychology*, 86:337–348.

Atti, A., Forlani, C., De Ronchi, D., Palmer, K., Casadio, P., Dalmonte, E., and Fratiglioni, L. (2010). Cognitive impairment after age 60: Clinical and social correlates in the "faenza project". *Journal of Alzheimer's Disease*, 21(4):1325–1324.

Bellew, K., Pigeon, J., Stang, P., Fleischman, W., Gardner, R., and Baker, W. (2004). Hypertension and the rate of cognitive decline in patients with dementia of the alzheimer type. *Alzheimer Disease and Associated Disorders*, 18(4).

Biessels, G., Staekenborg, S., Brunner, E., and Brayne, C. Scheltens, P. (2006). Risk of dementia in diabetes mellitus: a systematic review. *The Lancet. Neurology*, 5(1):64–74.

Breslow, N. (1972). Discussion following *"Regression models and life tables"* by d. r. cox. *Journal of the Royal Statistical Society*, B(34):187–220.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30:89–99.

Cabella, W. and Pellegrino, A. (2009). *La seguridad social en el Uruguay. Contribuciones a su historia*, chapter El envejecimiento de la población uruguaya y la transicion estructural de las edades, pages 89–114. AFAP-BROU , Montevideo.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*, 34(2):187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62:269–276.

Daniels, M. anda Hogan, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman and Hall/CRC.

DeGruttola, V. and Tu, X. (1994). Modeling progression of cd-4 lymphocyte count and its relationship to survival time. *Biometrics*, 50:1003–1014.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38.

Dempster, A. and Rubin, D. B. (1987). Incomplete data in sample surveys. *Theory and Anotated Bibliography*, 2.

Diggle, P., Farewell, D., and Henderson, R. (2007). Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. *Journal of the Royal Statistical Society, Series C*, 56:499–550.

Diggle, P. and Kenward, M. (1994). Informative dropout in longitudinal data analysis. *ournal of the Royal Statistical Society, Series C*, 43:49–93.

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72:557–565.

Faucett, C. and Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685.

Folstein, M., Folstein, S., and McHugh, P. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198.

Ghisletta, P. (2008). Application of a joint multivariate longitudinal-survival analysis to examine the terminal decline hypothesis in the swiss interdisciplinary longitudinal study on the oldest old. *The Journals of Gerontology*, 63(3):185–192.

Hazewinkel, M. (2001). *Encyclopedia of Mathematics*. Springer.

He, B. and Luo, S. (2013). Joint modeling of multivariate longitudinal measurements and survival data with applications to parkinson's disease. *Statistical Methods in Medical Research*.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5:475 – 492.

Hedeker, D. and Gibbons, R. (2006). *Longitudinal Data Analysis*. Wiley Series in Probability and Statistics. Wiley.

Henderson, H., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480.

Henderson, H., Diggle, P., and Dobson, A. (2002). Identification and efficacy of longitudinal markers for survival. *Identification and efficacy of longitudinal markers for survival*, 3(1):33–50.

Hogervorst, E., Matthews, F., and Brayne, C. (2010). Are optimal levels of testosterone associated with better cognitive function in healthy older women and men? *Biochimica et Biophysica Acta (BBA)*, 1800(10):1145–1152.

Hsieh, F., Tseng, Y.-K., and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, 62:1037–1043.

Hughes, S., Gibbs, J., Dunlop, D., Edelman, P., Singer, R., and Chang, R. (1997). Predictors of decline in manual performance in older adults. *Journal of the American Geriatrics Society*, 45(8):905–910.

Kalbfleisch, J. and Prentice, R. (1973). Marginal likelihoods based on cox's regression and life model. *Biometrika*, 60:267–278.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Kenward, M. and Roger, J. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53:983–997.

Kerwin, D., Gaussoin, S., Chlebowski, R., Kuller, L., Vitolins, M., Coker, L., Kotchen, J., Nicklas, B., Wassertheil-Smoller, S., Hoffmann, R., and Espeland, M. (2011). Interaction between body mass index and central adiposity and risk of incident cognitive impairment and dementia: results from the women's health initiative memory study. *Journal of the American Geriatrics Society*, 59(1):107–112.

Lenehan, M., Summers, M., Saunders, N., Summers, J., and Vickers, J. (2015). Relationship between education and age-related cognitive decline: a review of recent research. *Psychogeriatrics*, 15(2):154–162.

Li, Z., Tosteson, T., and Bakitas, M. (2013). Joint modeling quality of life and survival using a terminal decline model in palliative care studies. *Statistics in Medicine*, 32(8):1394–1406.

Little, R. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88:125–134.

Ma, G., Troxel, A., and Heitjan, F. (2004). An index of local sensitivity to non ignorable drop-out in longitudinal modelling. *Statistics in Medicine*, 24:2129–2150.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50(3):163–170.

Molenberghs, G., Beunckens, C., Sotto, C., , and Kenward, M. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, 70:371–388.

Musoro, J., Geskus, R., and A., Z. (2014). A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant. *Biometrical Journal*.

Muthen, B. (2004). *Handbook of quantitative methodology for the social sciences*, chapter 19 - Latent Variable Analysis. Growth Mixture Modeling and Related Techniques for Longitudinal Data, pages 345–368. Kaplan.

Newberg, A., Wintering, N., Khalsa, D., Roggenkamp, H., and Waldman, M. (2010). Meditation effects on cognitive function and cerebral blood flow in subjects with memory loss: a preliminary study. *Journal of Alzheimer's Disease*, 20:517–526.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York, 2nd edition.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-108.

Proust-Lima, C. (2014). Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research*, 1:74–90.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Richardson, A. and Welsh, A. (2008). Asymptotic properties of restricted maximum likelihood (reml) estimates for hierarichal mixed linear models. *Australian Journal of Statistics*, 36(1):31–43.

Rizopoulos, D. (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33.

Rizopoulos, D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56:491–501.

Rizopoulos, D. (2012b). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman & Hall/CRC Biostatistics Series. Chapman and Hall/CRC, 1 edition.

Rubin, D. (1976). Inference and missing data. *Biometrika*, 63:581–592.

Ryan, J., Stanczyk, F., Dennerstein, L., Mack, W., Clark, M., Szoeke, C., Kildea, D., and Henderson, V. (2012). Hormone levels and cognitive function in postmenopausal midlife women. *Neurobiology of Aging*, 33(3):617–639.

Satterthwaite, F. (1946). Approximate distribution of estimates of variance components. *Biometrics*, 2:110–114.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Self, S. and Pawitan, Y. (1992). *Modeling a marker of disease progression and onset of disease*. Methodological Issues. Birkhäuser Boston.

Stern, Y. (2012). Cognitive reserve in ageing and alzheimer's disease. *The Lancet. Neurology*, 11(11):1006–1012.

Stram, D. and Lee, J. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50(4):1171–1177.

Swan, G. Carmelli, D. and Larue, A. (1998). Systolic blood pressure tracking over 25 to 30 years and cognitive performance in older adults. *Stroke; a journal of cerebral circulation.*, 29(11):2334–2340.

Taylor, J., Park, Y., Ankerst, D., C., P.-L., Williams, W., Kestin, K., Bae, K., Pickles, T., and Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–219.

Terrera, G., Piccinin, A., Johansson, B., Matthews, F., and Hofer, S. (2011). Joint modeling of longitudinal change and survival: An investigation of the association between change in memory scores and death. *GeroPsych*, 24(4):177–185.

Thabut, G., Christie, J., Mal, H. Fournier, M., Brugiere, O., Leseche, G., Castier, Y., and Rizopoulos, D. (2013). Survival benefit of lung transplant for cystic fibrosis since lung allocation score implementation. *American Journal of Respiratory and Critical Care Medicine*, 187(12):1335–1340.

Therneau, T. (2012). *A Package for Survival Analysis in S*. R package version 2.37-2.

Troxel, A., Ma, G., and Heitjan, F. (2004). An index of local sensitivity to nonignorability. *Statistica Sinica*, 14:1221–1237.

van Elderen, S., de Roos, A., de Craen, A., Westendorp, R., Blauw, G., Jukema, J., Bollen, E., Middelkoop, H., van Buchem, M., and van der Grond, J. (2010). Progression of brain atrophy and cognitive decline in diabetes mellitus: a 3-year follow-up. *Neurobiology*, 75(11):997–1002.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. Springer, 1 edition.

Viviani, S. (2012). *Mixed effect joint models for longitudinal responses with dropout: estimation and sensitivity issues*. PhD thesis, Sapienza, Università di Roma.

Waldron, I. and Johnston, S. (1976). Why do women live longer than men? *Journal of human stress*, 2(2):19–30.

Walker, H. (1943). Degrees of freedom. *Journal of Educational Psychology*, 31(4):253–269.

Wang, H., Karp, A., Winblad, B., and Fratiglioni, L. (2002). Late-life engagement in social and leisure activities is associated with a decreased risk of dementia: a longitudinal study from the kungsholmen project. *Americal journal of epidemiology*, 155(12):1081–1087.

Waseem, S. (2007). *Kenward-Roger Approximate F Test for Fixed Effects in Mixed Linear Models*. PhD thesis, Oregon State University.

Welch, B. (1947). The generalization of "student's" problem when several different population variances are involved. *Biometrika*, 34:28–45.

Wilks, S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9:60–62.

Wu, M. and Bailey, K. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 5:337–346.

Wu, M. and Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44:175–188.

Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330–339.

Ye, W., Lin, X., , and Taylor, J. (2008). A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and Its Interface*, 1:33–45.

Apéndice A

Anexo Estadístico

A.1. REML

La idea principal detrás del método $REML$ consiste en "separar" la parte de los datos utilizada para la estimación de los parámetros que afectan a la media de la parte involucrada en la estimación de los parámetros de covarianza (Elementos de la matriz D y σ^2). Básicamente se trata de eliminar a β de la verosimilitud de modo que los únicos parámetros involucrados en esta sean los encargados de modelar la variación entre e intra individuos.

A modo de ejemplo se presenta el caso de la estimación de la varianza de una muestra aleatoria simple de una distribución normal. Sean $X_1, X_2, \ldots, X_n$ con distribución $N(\mu, \sigma^2)$. Los estimadores máximo verosímiles de estos parámetros son bien conocidos pero se sabe que:

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n}$$

es un estimador sesgado. Por este motivo, el estimador $REML$ propone eliminar a μ de la verosimilitud definiendo las siguientes $n-1$ variables aleatorias: $Y_2 = X_2 - X_1, Y_3 = X_3 - X_2, \ldots, Y_n = X_n - X_{n-1}$. Al estimar la varianza de este nuevo conjunto de variables aleatorias centradas mediante el mismo estimador, se puede ver que el sesgo es cero. Generalizando este método, la idea de este algoritmo es utilizar una matriz de contrastes A que centre los datos (eliminando los parámetros intervinientes en la media) y maximizar la distribución de los datos AX en vez de X. De esta manera, contrario a la

estimación por máxima verosimilitud, se obtienen estimadores insesgados de los parámetros de covarianza. Contrario a lo que sucede con el estimador de β, los estimadores de las varianzas no poseen forma cerrada por lo que deben ser estimados numéricamente.

A.2. ISNI

La elección de éste índice como medida de sensibilidad del supuesto de MAR responde que según Troxel et al. (2004) es un método sumamente simple de análisis (requiere aproximar únicamente un gradiente y una hessiana) que permite evaluar el potencial impacto de la ignorabilidad del supuesto MAR. En la sección de análsis de sensibilidad se comentó que la idea del $ISNI$ es medir la velocidad de cambio de las estimaciones del modelo longitudinal al alejarse de la situación de MAR. Para ello se utiliza el valor de la derivada de los coeficientes $\beta(\alpha)$, cuando $\alpha = 0$. Con el fin de estimar esta cantidad, Troxel propone la siguiente aproximación de segundo grado a la verosimilitud del modelo $MNAR$, utilizando como $\theta^{(0)}$ el vector de estimaciones máximo verosímiles bajo el supuesto MAR:

$$\mathcal{L}(\theta) = \mathcal{L}(\theta^{(0)}) + (\theta - \theta^{(0)})' \left\{ \frac{\partial}{\partial \theta'} \mathcal{L}(\theta)|_{\theta=\theta^{(0)}} \right\} + \frac{1}{2}(\theta - \theta^{(0)})' \left\{ \frac{\partial}{\partial \theta' \partial \theta} \mathcal{L}(\theta)|_{\theta=\theta^{(0)}} \right\} (\theta - \theta^{(0)}).$$

Teniendo en cuenta que la derivada de la verosimilitud en $\theta =^{(0)}$ debe ser nula ya que estos últimos son los estimadores máximo verosímiles y que el valor del ISNI se podría aproximar mediante un cociente incremental, con un poco de álgebra se llega a la siguente expresión del indicador:

$$ISNI = \frac{\partial}{\partial \alpha} \beta(\alpha) \Big|_{\alpha=0} \approx - \left\{ \frac{\partial^2}{\partial \beta^T \partial \beta} \ell(\theta)|_{\theta=\theta^{(0)}} \right\}^{-1} \left\{ \frac{\partial^2}{\partial \beta^T \partial \alpha} \ell(\theta)|_{\theta=\theta^{(0)}} \right\}.$$

A.3. Verosimilitud parcial

En su trabajo seminal, Cox (1972) introdujo el concepto de verosimilitud parcial como una alternativa a la máxima verosimilitud para los casos donde se busca la estimación de un conjunto de parámetros sin tener en cuenta una gran cantidad de "parámetros de ruido". En este sentido, para el caso de análisis de sobrevida, hay dos alternativas para llegar a la verosimilitud parcial. La primera es pensar en una verosimilitud donde la contribución de cada individuo a la verosimilitud se da condicional a que alguno de los individuos de la muestra presente el evento de interés, esto es:

$$\mathcal{PL}(\theta) = \prod_{i=1}^{n} P(\text{individuo i fallece en } t_i | \text{ hay 1 fallecimiento en } t_i) =$$
$$\prod_{i=1}^{n} \frac{P(\text{individuo i fallece en } t_i | \text{sobrevive hasta } t_i)}{P(1 \text{ fallececimiento en } t_i | \text{sobreviven hasta } t_i)} =$$
$$\prod_{i=1}^{n} \left[\frac{h_0(t_i) e^{x_i\theta}}{\sum_{j \in R(t_i)} h_0(t_j) e^{x_j\theta}} \right]^{\delta_i} = \prod_{i=1}^{n} \left[\frac{e^{x_i\theta}}{\sum_{j \in R(t_i)} e^{x_j\theta}} \right]^{\delta_i}$$

donde cada uno de los términos involucra $R(t_i)$ individuos, siendo esta una cantidad que decrece a medida que los individuos experimentan el evento o se censuran. La segunda manera de llegar a este resultado parte de la verosimilitud completa de la muestra. Para ello es importante recordar la relación $h(t) = \frac{f(t)}{S(t)}$ y que $\delta_i = 1$ indica que el tiempo de sobrevida ha sido observado mientras que $\delta_i = 0$ indica la censura de dicho individuo. Entonces la verosimilitud completa estaría dada por la expresión:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_i(t_i)^{\delta_i} S_i(t_i)^{1-\delta_i} = \prod_{i=1}^{n} h_i(t_i)^{\delta_i} S_i(t_i) =$$
$$\prod_{i=1}^{n} \left[\frac{h_i(t_i)}{\sum_{j \in R(t_i)} h_j(t_j)^{\delta_j}} \right]^{\delta_i} \left[\sum_{j \in R(t_i)} h_j(t_j)^{\delta_j} \right]^{\delta_i} S_i(t_i).$$

Finalmente se "deshechan" el segundo y el tercer miembros de cada término del producto para llegar a la verosimilitud parcial:

$$\mathcal{PL}(\theta) = \prod_{i=1}^{n} \left[\frac{h_i(t_i)}{\sum_{j \in R(t_i)} h_j(t_j)^{\delta_j}} \right]^{\delta_i} = \prod_{i=1}^{n} \left[\frac{exp(x_i\theta)}{\sum_{j \in R(t_i)} exp(x_j\theta)^{\delta_j}} \right]^{\delta_i}$$

De esta manera, se elimina la función de riesgo de referencia (dejada sin especificar) y la maximización se concentra en los parámetros de regresión. Debe notarse además que los casos censurados sólo contribuyen al conjunto de individuos "bajo riesgo de experimentar el evento" (suma presente en el denominador).

Cox mostró que esta metodología puede utilizarse del mismo modo que máxima verosimilitud por lo que las inferencias devenidas de este método son igual de válidas que las obtenidas mediante máxima verosimilitud. Sin embargo es importante hacer la precisión de que el método planteado de esta manera sólo atañe al caso donde no hay "empates" (tiempos iguales entre individuos). Ese problema es solucionado al modificar levemente el denominador de cada término del producto o bien con la modificación de Breslow (1974) o con la modificación de Efron (1977).

A.4. Tópicos de sobrevida

A.4.1. Estimador de Kaplan-Meier

En el análisis de datos de sobrevida, la manera más usual de resumir los datos es a través de la función de sobrevida muestral. El propósito del estimador Kaplan-Meier (KM) es aproximar esta función de una manera no paramétrica que fácilmente puede utilzarse en casos donde existen tiempos censurados, donde se supone la independencia e igualdad de distribución entre los datos.

El estimador *producto límite* tiene la siguiente forma:

$$\hat{S(t)} = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j} \tag{A.1}$$

donde t_j son los tiempos observados de sobrevida, n_j es el número de individuos "en riesgo" de experimentar el evento en el momento anterior a t_j y d_j representa la cantidad de individuos que efectivamente experimentan el evento en ese momento (no incluye a los casos censurados). En el caso de que no se registren censuras, el estimador KM adopta el valor cero en el último dato, o sea: $\hat{S}(\text{máx}\,\{t_j\}) = 0$.

A.4.2. Prueba del rango-logarítmico (*log-rank test*)

Existen ocasiones donde interesa comparar la sobrevida de dos o más grupos. Pese a que esta comparación se podría llevar a cabo para algún valor de t, realizar este procedimiento sobre un conjunto de valores puede inflar el error de tipo I, por este motivo, la prueba se realiza de manera global, comparando las estimaciones de la función de sobrevida en los distintos grupos. La prueba

consta de la comparación de las funciones de sobrevida poblacionales de G grupos:

$$\begin{array}{ll} H_0) & S_1(t) = S_2(t) = \ldots = S_G(t) \\ H_1) & \text{no } H_0 \end{array}$$

El estadístico de prueba (véase Mantel (1966)) compara las fracciones de eventos observados en cada momento para todos los grupos, con las esperadas bajo el cumplimiento de la hipótesis nula (donde todos los datos provienen de la misma población). La distribución límite del estadístico es χ^2_{G-1}.

A.5. Algoritmo EM

El algoritmo EM es un método general de estimación (iterativa) en presencia de datos faltantes. La idea detrás de este método es que la maximización de la verosimilitud "completa" (compuesta por datos observados y datos faltantes) suele ser más fácil que la de la verosimilitud "observada". El algoritmo procede iterativamente en dos etapas, en la primera etapa (llamada paso E) se "completan" los datos con el valor esperado de los valores faltantes, condicional a los valores observados. De esta manera se construye una verosimilitud esperada ($Q(\theta|\theta^{(t)})$). En la segunda etapa (paso M) se maximiza esta verosimilitud respecto de los parámetros de interés.
De esta manera, en el paso E, la verosimilitud esperada es:

$$Q(\theta|\theta^{(t)}) = E\left\{\log p(Y;\theta)|Y^o;\theta^{(t)}\right\}$$

mientras que el paso M es:

$$\theta^{(t+1)} = \arg\operatorname*{máx}_{\theta} \quad Q(\theta|\theta^{(t)})$$

Siendo $\theta^{(t)}$ el valor del vector de parámetros en la iteración t, $\log p(Y;\theta)$ la log-verosimilitud completa y Y^o los datos observados. Pese a su simplicidad, este algoritmo suele ser muy lento y presenta una gran dependencia de los valores iniciales del vector de parámetros. Sin embargo, es un algoritmo numéricamente estable dado que en cada iteración garantiza un aumento de la log-verosimilitud y que dada una dirección de búsqueda suele no excederse ni limitarse demasiado con respecto al máximo.

A.6. Librerías de R utilizadas

A lo largo de este trabajo se utilizaron los siguientes librerías de R.

- *nlme*
  La librería fue utilizada para estimar los modelos de efectos mixtos correspondientes al análisis de datos longitudinales.

- *survival*
  La librería se utilizó para llevara a cabo la estimación del modelo de regresión de Cox y el modelo Weibull intervinientes en el apartado de análisis de datos de sobrevida.

- *JM*
  La librería se utilizó para estimar los modelos conjuntos así como para llevar a cabo las inferencias correspondientes a cada sub-modelo.

Apéndice B

Anexo Metodológico

B.1. MMSE

A continuación se presenta una versión (en inglés) del test, indicando entre paréntesis el máximo puntaje de cada pregunta.

(5pt) "What is the year? Season? Date? Day? Month?"

(5pt) "Where are we now? State? County? Town/city? Hospital? Floor?"

(3pt) The examiner names three unrelated objects clearly and slowly, then the instructor asks the patient to name all three of them. The

patient's response is used for scoring. The examiner repeats them until patient learns all of them, if possible.

(5pt) "I would like you to count backward from 100 by sevens." (100, 93, 86, 79, ...). Alternative: "Spell WORLD backwards." (D-L-R-O-W)

(3pt) "Earlier I told you the names of three things. Can you tell me what those were?"

(2pt) Show the patient two simple objects, such as a wristwatch and a pencil, and ask the patient to name them.

(1pt) "Repeat the phrase: "No ifs, ands, or buts""

(3pt) "Take the paper in your right hand, fold it in half, and put it on the floor." (The examiner gives the patient a piece of blank paper.)

(1pt) "Please read this and do what it says." (Written instruction is "Close your eyes.")

(1pt) "Make up and write a sentence about anything." (This sentence must contain a noun and a verb.)

(1pt) "Please copy this picture." (The examiner gives the patient a blank piece of paper and asks him/her to draw the symbol below. All 10 angles must be present and two must intersect.)