

Abstract

El estudio de aplicaciones específicas en diversos campos de la investigación científica se ha asociado con el desarrollo de múltiples enfoques y modelos estadísticos. En particular, analizar y estudiar mediciones repetidas y desbalanceadas a lo largo del tiempo sigue siendo un desafío para el modelado estadístico. Por esta razón, el siguiente manuscrito se ha dedicado a proponer nuevos modelos estadísticos aún más desafiantes para el análisis de datos longitudinales, tratando de capturar la esencia de un conjunto de datos estudiado y analizado en numerosas propuestas científicas, motivando la nuestra. Se trata de un conjunto de 173 mujeres embarazadas para las que se registraron los niveles de la hormona gonadotropina coriónica humana (β -HCG), considerado un biomarcador determinante en los primeros días de gestación. Este conjunto de datos se dividió en dos grupos, un grupo normal que consta de 124 mujeres que llevaron a término su embarazo y un grupo anormal, compuesto por las 49 mujeres restantes que lamentablemente sufrieron algún problema ocasionando la pérdida del feto. Luego de una breve introducción y algunos resultados preliminares (Capítulo 2) que enmarcan nuestra investigación, dividimos esta tesis en tres partes principales:

En la primera parte de este manuscrito, proponemos analizar y estimar un modelo semi-paramétrico usando P-splines, partiendo de un modelo mixto no lineal y agregando una parte no paramétrica usando funciones B-splines con la idea de agregar mayor flexibilidad al modelo. Los modelos generados fueron aplicados utilizando un conjunto de datos de mujeres embarazadas, lo que permitió hacer clasificación supervisada en dos grupos, “embarazos normales” y “embarazos anormales”, utilizando los modelos propuestos en este capítulo. Algunas mujeres pierden el feto durante el estudio, por lo que los embarazos que no llegaron a término se clasifican como anormales.

La segunda y tercera parte de esta tesis están dedicadas a analizar y estimar modelos conjuntos con la idea de capturar mayor información en el conjunto de datos disponible. En el Capítulo 4, propusimos un modelo mixto no lineal para el análisis de mediciones repetidas con un modelo lineal generalizado de respuesta primaria que modela la probabilidad de que una mujer tenga un aborto. Este modelo pudo estimarse siguiendo un enfoque frecuentista, y considerando el conjunto de datos de 173 gestantes, realizando la clasificación en dos grupos. Finalmente, en el Capítulo 5, nuevamente proponemos un modelo conjunto considerando el modelo mixto no lineal para datos longitudinales con un modelo de supervivencia en tiempo

discreto, donde el evento es la pérdida del feto de cada mujer. Todos los modelos se estimaron utilizando el algoritmo SAEM, una versión estocástica del algoritmo EM que permite estimar modelos mixtos complejos sin recurrir a métodos de linealización que han mostrado generar estimaciones inconsistentes y problemas de convergencia en algunas situaciones (pocas observaciones por individuo, muchos efectos aleatorios, datos sparse, entre otras).