

Contents

Introducción	2
Antecedentes	2

Introducción

La estructura del modelo bayesiano permite capturar las relaciones de dependencia que existe entre los atributos de los datos que se estudien, por medio de una distribución “posterior”, por lo cual en este artículo veremos una aplicación en la rama de la salud, específicamente en casos de cáncer de mama, esto aplicando inferencia bayesiana por medio de un modelo logístico binario, para ver las probabilidades según un conjunto de datos conocidos de que X mujer (hablando desde la parte biológica, abstrayendo todo concepto de género diferente) pueda ser diagnosticada con cáncer de mama (sin discriminación por tipo de cancer)

Más adelante veremos por medio de diagnósticos (la mayoría, por no decir todos) visuales, para un mejor entendimiento (cada uno de ellos explicados posteriormente) del modelo y si este nos ayuda a predecir el problema en cuestión, además veremos un cambio en las variables de respuesta por lo que veremos otros modelos del mismo tipo para una comparación entre ellos.

Antecedentes

En base al software estadístico R versión 4.2.0, con la ayuda de los paquetes:

- tidyverse
- magrittr
- janitor
- readr
- BayesFactor
- bayesplot
- brms
- RStan

con las propiedades de RStan para la construcción del modelo logístico; cabe destacar que donde se corrieron los modelos no contaban con un gran poder computacional, por lo que los tiempos de ejecución eran muy largos.

La base de datos fue tomada del Instituto Nacional del Cáncer (INC), donde recopilaron información de 20.000 mamografías digitales y 20.000 de proyección de película realizadas entre enero de 2005 y diciembre de 2008 de mujeres incluidas en el Consorcio de Vigilancia del Cáncer de Mama en las cuales encontramos las siguientes variables:

Nombre de la Variable	Descripción	Codificación
Edad	Edad de la paciente en años en el momento de la mamografía	Numérico
Evaluación	Valoración del radiólogo basada en la Escala BI-RADS (Hallazgos de posibles masas o tumores)	0 = Necesita imágenes adicionales 1 = Negativo 2 = Hallazgo(s) Benigno(s) 3 = Probablemente Benigno 4 = Anomalía sospechosa 5 = Probablemente maligno
Desidad	Densidad mamaria en escala BI-RADS de la paciente registrada en el momento de la radiografía	1 = Casi totalmente Graso 2 = Densidades fibroglandulares dispersas 3 = Heterogéneamente denso 4 = Extremadamente denso
Antecedentes Familiares	antecedentes familiares de cáncer de mama en un familiar de primer grado	0 = no 1 = sí 9 = falta
IMC	índice de masa corporal en el momento de la mamografía	Numérico
Cáncer	<i>indicador binario de diagnóstico de cáncer en la entrega de resultados del examen ambulatorio</i>	<i>0 = Negativo 1 = Positivo</i>