

Resumen

Estudios demuestran que las características subyacentes de un paciente tienen fuerte influencia sobre su condición de infectado por COVID-19. Es un fenómeno epidemiológico que trajo consigo altas tasas de mortalidad y aumento hospitalario. En este trabajo se evalúan las características médicas de un paciente adulto con infección por SARS-CoV-2 ingresado a hospitalización en Georgia Atlanta, para determinar la influencia de éstas sobre la mortalidad, dado un perfil clínico. A partir de cuatro posibles modelos logísticos Bayesianos, se selecciona el más apropiado. Se utilizan distribuciones a priori poco informativas y métodos MCMC para aproximar la distribución posterior de los parámetros del modelo. En el modelo final resultaron con influencia significativa las afecciones de tipo: pulmonar (2), cardiovascular(3), hematológico (2), inflamatorio (2) y otros (4).

Introducción

El COVID-19 es “una enfermedad infecciosa causada por el SARS-coV-2”, proveniente de Wuhan, China en 2019 (WHO, 2020). Los síntomas más frecuentes presentados son los respiratorios, y es una enfermedad que puede atacar no sólo a los pulmones, sino también diferentes partes del cuerpo.

Los primeros casos de enfermedad por coronavirus en Georgia, inició el 2 de marzo de 2020 donde se confirmaron los dos primeros casos, el 7 de marzo se confirma la primera muerte en el hospital WellStar Kenesstone; agregaron que el paciente tenía afecciones médicas subyacentes (Morales,2021).

Los datos de este estudio provienen de un paquete de software de vigilancia de atención médica (*Clinical Looking Glass [CLG]*), involucra datos de registros médicos primarios y variables que comprometen la salud de un paciente hospitalizado infectado por COVID-19 en el año 2020 (Kaggle, 2020). La base de datos depurada consta de 2455 registros y 67 variables. Se construye un modelo de regresión logístico con enfoque Bayesiano con el fin de predecir la mortalidad de un paciente a partir de sus características médicas.

Metodología

1. Obtención y depuración de la base de datos.
2. Construcción y elección del modelo bayesiano mediante factores de Bayes.
3. Validación del modelo.
4. Análisis de los resultados obtenidos.

Modelo de regresión logístico.

$$Y_i = \begin{cases} 1 & \text{El paciente } i \text{ fallece} \\ 0 & \text{El paciente } i \text{ sobrevive} \end{cases}$$

$$Y_i \sim \text{Bernoulli}(\theta_i), i = 1, 2, \dots, 2,455$$

$$P(Y_i = 1) = \theta_i, P(Y_i = 0) = 1 - \theta_i.$$

$$\theta_i = \frac{e^{\beta^T X_i}}{1 + e^{\beta^T X_i}}, \quad \beta_j \sim N(0, 10^{12}), \quad j = 0, 1, \dots, 15$$

Finalmente, la distribución posterior está dada por:

$$f(\beta|Y, X) = \prod_{i=1}^n \left[\left(\frac{e^{\beta^T X_i}}{1 + e^{\beta^T X_i}} \right)^{Y_i} \left(\frac{1}{1 + e^{\beta^T X_i}} \right)^{1-Y_i} \right] * f(\beta)$$

con $f(\beta)$ la distribución a priori conjunta del vector β .

El modelo se ajustó usando *Stan* en el software R (R, 2021). Se utilizan tres cadenas con 200000 iteraciones y un quemado de 20000.

Referencias

- Morales, J (2021). Cronología de la pandemia en Georgia: Esto ha pasado a un año de detectar los primeros casos de coronavirus. Retrieved 25 June 2022, from <https://www.univision.com/local/atlanta-wuvg/esto-ha-pasado-a-un-ano-de-detectar-los-primeros-casos-de-coronavirus-georgia>
- Kaggle.com. 2022. Predict Severity and Deaths of COVID - 19 Patients. [online] Available at: <https://www.kaggle.com/code/imperiop/predict-severity-and-deaths-of-covid-19-patients> [Accessed 27 August 2022].
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.<https://www.R-project.org/>
- WHO (2020). Información básica sobre la COVID-19. World Health Organization.<https://www.who.int/es/news-room/questions-and-answers/item/coronavirus-disease-covid-19>

Análisis Preliminares

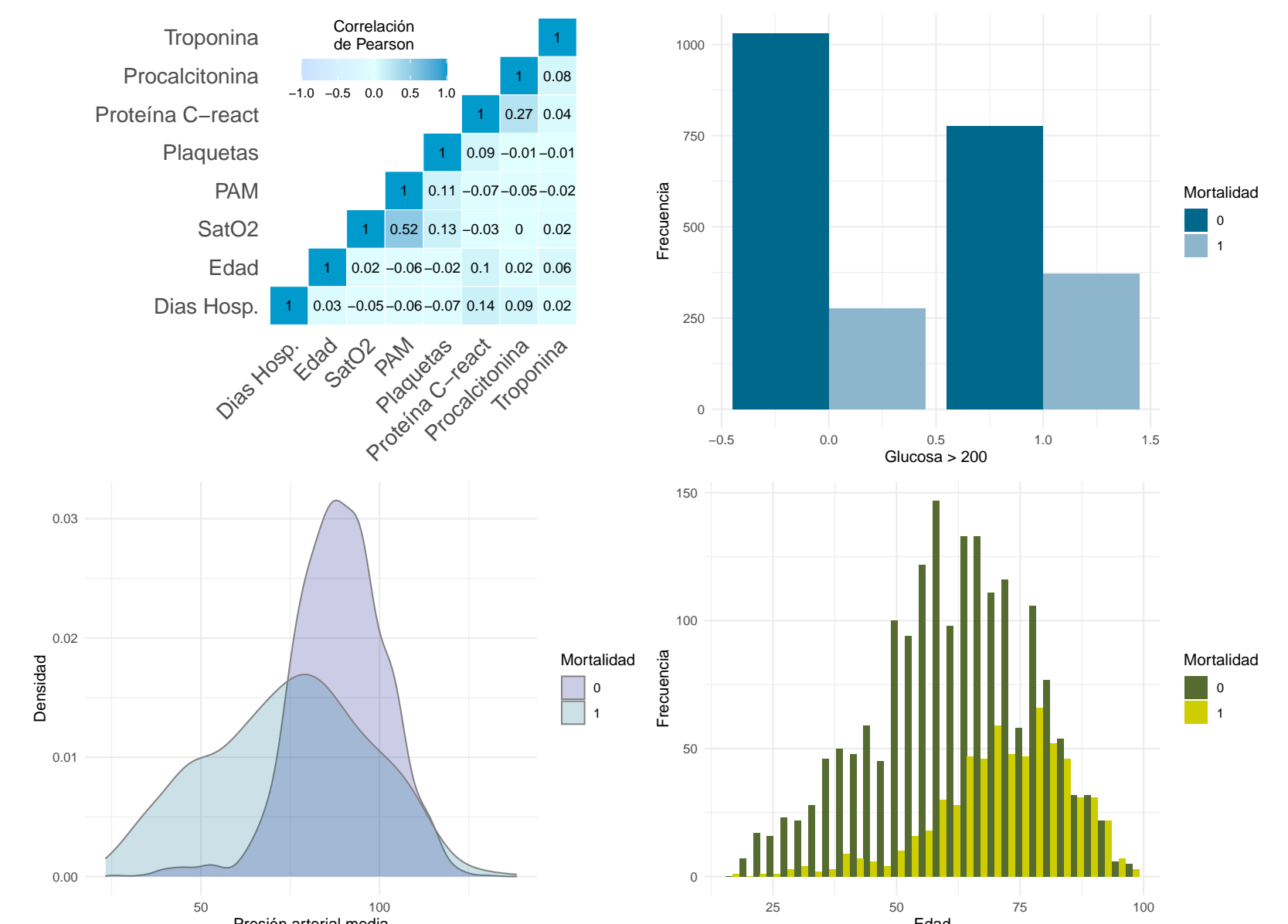


Figura 1: Gráficos descriptivos

Resultados

$$\text{Modelo Ajustado: } \hat{P}(Y_i = 1) = \frac{e^{\hat{\beta}^T X_i}}{1 + e^{\hat{\beta}^T X_i}}$$

Variable	$e^{\hat{\beta}_i}$	HDI	Variable	$e^{\hat{\beta}_i}$	HDI
Intercepto	7.39	(1.2,13.5)	Procalcitonina	1.04	(1.02,1.03)
Días Hospitalización	1.02	(1.01,1.03)	Troponina	2.20	(1.30,1.77)
Derrame Cerebral	4.85	(1.93,7.38)	Ferritina <300	1.35	(1.03,1.22)
Edad	1.04	(1.03,1.05)	Glucosa >200	1.75	(1.40,1.63)
Saturación de Oxígeno	0.97	(0.95,0.99)	Dimero D	0.47	(0.32,0.41)
Presión Arterial Media	0.96	(0.95,0.96)	Severidad 2	1.79	(1.23,1.57)
Plaquetas	0.99	(0.98,0.99)	Severidad 3	2.89	(1.84,2.46)
Proteína C-reactiva	1.02	(1.01,1.04)			

Cuadro 1: Resumen del modelo

Capacidad predictiva

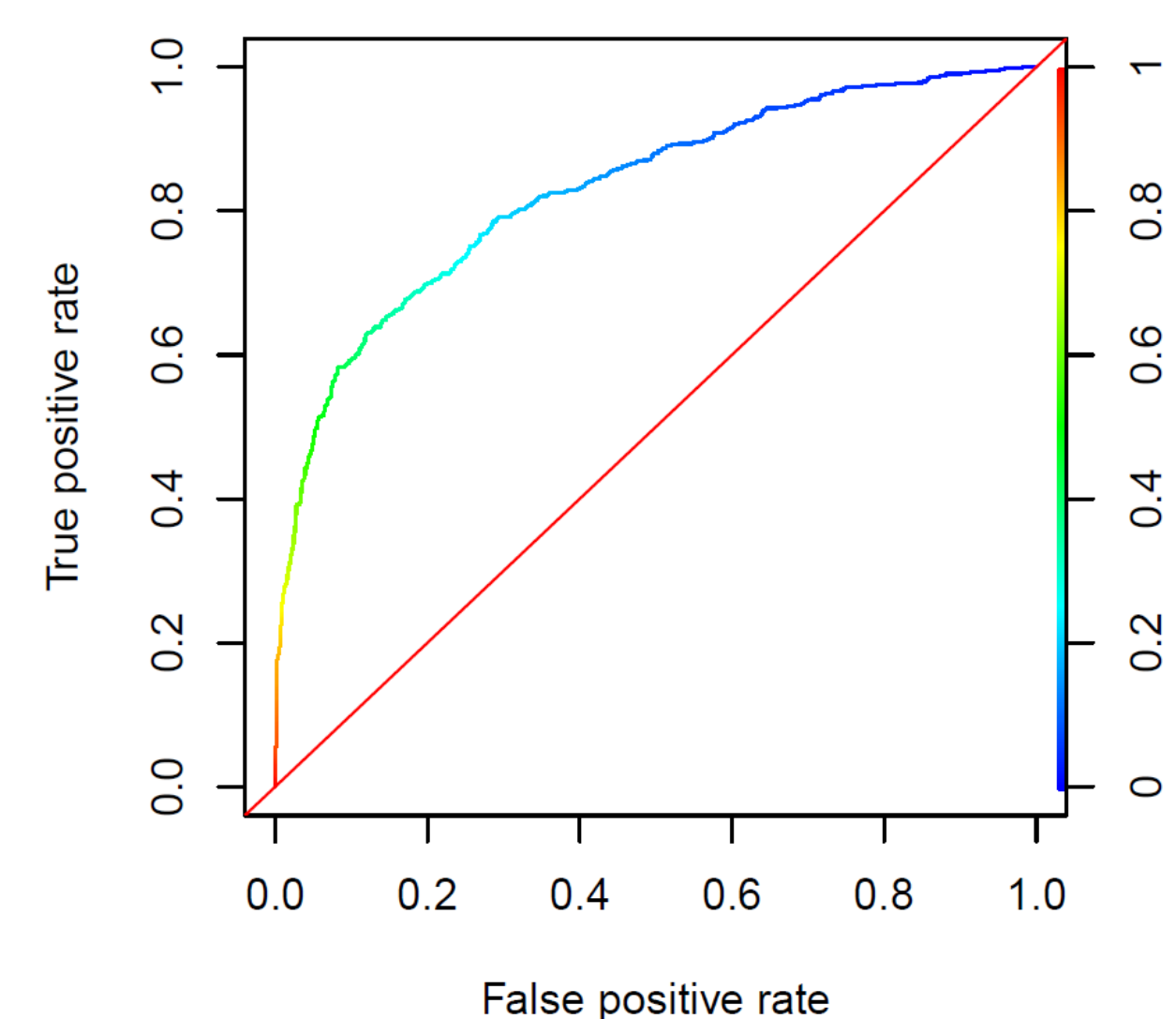


Figura 2: Curva ROC, Área bajo la curva:0.82

Conclusiones

- El modelo ajustado con el conjunto de entrenamiento clasifica nuevas observaciones con una efectividad del 84.92 %, es decir, de cada 100 observaciones, el modelo asigna acertadamente 85 de ellos en cada una de las categorías correspondientes.
- Las personas infectadas por COVID-19, tienen el riesgo de morir más alto con la presencia de características médicas subyacentes clasificadas como cerebral, pulmonares, cardiovasculares, inflamatorias y hematológicas.
- El modelo permite, en términos aplicativos, dar prioridad a pacientes clasificados en los grupos resultantes, debido a la fuerza de las variables en el marco predictivo.