

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

ESCUELA DE POSGRADO



Estimación de la probabilidad de anemia infantil usando un modelo de
regresión skew-probit

Tesis para optar por el grado académico de Maestro en Estadística
que presenta:

Freddy Rolando Mestas Ventocilla

Asesora:

Zaida Jesús Quiroz Cornejo


Lima, 2024

Informe de Similitud

Yo Zaida Jesús Quiroz Cornejo docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesora de la tesis titulada *Estimación de la probabilidad de anemia infantil usando un modelo de regresión skew-probit*, del autor Freddy Rolando Mestas Ventocilla, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 20%. Así lo consigna el reporte de similitud emitido por el software Turnitin el 12/08/2024.
- He revisado con detalle dicho reporte y confirmo que cada una de las coincidencias detectadas no constituyen plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lima, 12 de agosto de 2024

Apellidos y nombres de la asesora: Quiroz Cornejo Zaida Jesús	
DNI: 43704124	Firma: 
ORCID: https://orcid.org/0000-0003-3821-0815	

Resumen

La desnutrición crónica infantil en menores de cinco años de edad es un problema de salud pública en el Perú. Modelos estadísticos apropiados pueden ayudar a identificar variables o factores que permitan estimar la probabilidad de anemia infantil. Los modelos estadísticos para este tipo de datos binarios más conocidos son el modelo de regresión logística y probit. En esta tesis se aplican estos modelos y el modelo skew-probit, una extensión del modelo probit cuya función de enlace es asimétrica, en particular usando una versión estandarizada de la distribución skew normal. La inferencia se realiza a través del enfoque bayesiano, específicamente a través de la aproximación de Laplace integrada y anidada (INLA) debido a su eficiencia computacional. Cabe resaltar que se usa una distribución a priori penalizada compleja (PC prior) para el parámetro de sesgo de la skew normal, de esta forma se “cuantifica” la elección del modelo skew-probit respecto al modelo probit. Los resultados obtenidos para la estimar la probabilidad de anemia en niños menores de cinco años justifican la elección del modelo skew-probit.

Palabras-clave: anemia infantil, INLA, PC prior, regresión logística, regresión probit, regresión skew probit.

Abstract

Chronic childhood malnutrition in children under five years of age is a public health problem in Peru. Appropriate statistical models can help to identify variables or factors that allow estimating the probability of childhood anemia. The best-known statistical models for this type of binary data are the logistic regression and probit models. In this thesis, these models and the skew-probit model, an extension of the probit model whose link function is asymmetric, are applied, in particular using a standardized version of the skew normal distribution. Inference is performed through the Bayesian approach, specifically through integrated nested Laplace approximation (INLA) due to its computational efficiency. It should be noted that a complex penalized prior distribution is also used for the skew parameter, through this approach the choice of the skew-probit model is “quantified” with respect to the probit model. The results obtained to estimate the probability of anemia in children under five years of age justify the choice of the skew-probit model.

Keywords: childhood anemia, INLA, logistic regression, PC prior, probit regression, skew-probit regression.

Índice general

Índice de figuras	VII
Índice de cuadros	ix
1. Introducción	1
1.1. Planteamiento y justificación del tema	1
1.2. Objetivos de la tesis	3
1.3. Organización del trabajo	3
2. Conceptos y modelos	5
2.1. Distribución normal	5
2.2. Distribución skew-normal	7
2.3. Modelo lineal generalizado (MLG)	9
2.3.1. Componentes de un MLG	9
2.3.2. Regresión binaria	11
2.4. Inferencia bayesiana	12
2.4.1. Monte Carlo vía cadenas de Markov (MCMC)	14
2.4.2. Aproximación de Lapalace integrada anidada (INLA)	15
3. Modelo de regresión skew probit	18
3.1. Estandarización de la función enlace	18
3.2. Intercepto cuantil e identificación de los parámetros	20
3.3. PC-prior para el parámetro de sesgo	21
3.3.1. Principio 1 - Occam's razor	21
3.3.2. Principio 2 - Medida de complejidad	22
3.3.3. Principio 3 - Ratio de penalización constante	22
3.3.4. Principio 4 - Escala definida por el usuario	23
3.3.5. Reparametrización del parámetro de sesgo	24

3.4. Inferencia bayesiana	24
4. Estudio de Simulación	27
4.1. Generación de datos	27
4.2. Estudio de simulación 1: número de ensayos N_i grande	27
4.2.1. Resultados	28
4.3. Estudio de simulación 2: número de ensayos N_i pequeño	29
4.3.1. Resultados	29
4.4. Estudio de simulación 3: Efecto de intercepto cuantil	31
4.4.1. Resultados	31
4.5. Estudio de simulación 4: Modelo Probit	32
4.5.1. Resultados	32
4.6. Estudio de simulación 5: Modelo Bernoulli skew probit	32
4.6.1. Resultados	32
5. Aplicación	34
5.1. Situación de la anemia en la población infantil en el Perú	34
5.2. Procesamiento y análisis exploratorio de los datos	36
5.3. Modelos propuestos	39
5.4. Resultados	40
5.4.1. Ajuste del modelo en datos de entrenamiento	40
5.4.2. Ajuste del modelo en datos de prueba	41
6. Conclusiones	44
6.1. Conclusiones	44
6.2. Sugerencia para trabajos futuros	44
Bibliografía	47

Índice de figuras

2.1. Función de densidad de variables aleatorias con distribución normal con media μ y varianza σ^2	6
2.2. Función de distribución acumulada de variables aleatorias con distribución normal con media μ y varianza σ^2	6
2.3. Función de densidad de variables aleatorias skew normal con $\xi = 0$ y $\omega = 1$ y diferentes valores para el parámetro de sesgo α	7
2.4. Función de distribución acumulada de variables aleatorias skew normal con cuando $\xi = 0$ y $\omega = 1$ y diferentes valores para el parámetro de sesgo α	8
2.5. Distribución condicional de respuesta en modelo de regresión con distribución normal y función de enlace identidad. Fuente: Wicklin (2015).	10
2.6. Distribución condicional de respuesta en modelo de regresión con distribución normal y función de enlace logarítmica. Fuente: Wicklin (2015).	11
2.7. Comparación de regresión lineal simple, logit y probit para datos binarios. Fuente: Jaramillo (2024)	12
2.8. Función de verosimilitud (línea azul), distribución a priori (inicial, línea roja) y distribución a posteriori (línea verde).	13
3.1. PC prior de α para $\theta = 5$. Fuente: Van Niekerk y Rue (2021).	23
3.2. PC prior de γ_1 para $\theta = 5$. Fuente: Van Niekerk y Rue (2021).	24
5.1. Porcentaje de población con anemia infantil (línea naranja) y desnutrición crónica (línea azul) del año 2010 al año 2022 en el Perú. Fuente: Tavera Salazar et al. (2023).	34
5.2. Prevalencia de anemia en niñas y niños de 6 a 36 meses según departamentos en el Perú en el año 2022. Fuente: INEI (2023).	35

5.3. Porcentaje de población de niñas y niños de 6 a 35 meses de edad que consumieron suplemento de hierro en los último 7 días, del año 2009 al 2022. Fuente: Tavera Salazar et al. (2023). 36

5.4. Diagrama de cajas de variables numéricas asociadas al niño: Edad en meses (HW1), Peso en kg. (HW2), Altura en cm. (HW3), Nivel de hemoglobina (HW53), Nivel de hemoglobina globina ajustado por altitud (HW56), Talla/Edad (HW70), Peso/Edad (HW71), Peso/Talla (HW72), desviación estándar del IMC (HW73), según anemia (1) o no anemia (0). 37

5.5. Diagrama de cajas de variables numéricas para el modelamiento: Peso del niño (V437), altura del niño (V438), IMC del niño (V445), Índice de Rohrer (V446), Edad de madre (V447A), nivel de hemoglobina de madre (V453), nivel de hemoglobina de madre ajustado por la altitud (V456), Meses de lactancia (M5), total de personas en el hogar (HV009), según anemia (1) o no anemia (0). 37

5.6. Diagrama de cajas de variables numéricas para el modelamiento: Número de miembros de residentes habituales (HV012), edad de madre en años (HA1), peso de madre en kg (HA2), talla de madre en cm (HA3), peso/talla de madre (HA12B), IMC de madre (HA40), Índice de rohrer de madre (HA41), nivel de hemoglobina de madre (HA53), nivel de hemoglobina de madre ajustado por altitud (HA56), según anemia (1) o no anemia (0). 38

5.7. Correlaciones de posibles covariables numéricas. 39

5.8. Función de densidad marginal a posteriori de los coeficientes de regresión según el modelo skew probit. 42

5.9. Función de densidad marginal a posteriori del parámetro q (izquierda) y γ_1 (derecha) según el modelo skew probit. 43

Índice de cuadros

4.1. Tabla de escenarios para N_i grande.	28
4.2. Tabla de resultado de simulaciones para N_i grande - a priori (PC).	28
4.3. Tabla de resultado de simulaciones para N_i grande - a priori gaussiana.	29
4.4. Tabla de escenarios para N_i pequeño	30
4.5. Tabla de resultado de simulaciones para N_i pequeño - a priori (PC)	30
4.6. Tabla de resultado de simulaciones para N_i pequeño - a priori gaussiana.	31
4.7. Tabla de simulaciones con cálculo clásico de intercepto	31
4.8. Tabla de resultados de simulaciones con estimación clásica del intercepto.	31
4.9. Tabla de simulaciones con parámetro de sesgo $\alpha = 0$	32
4.10. Tabla de resultados de simulaciones con PC prior cuando $\alpha = 0$	32
4.11. Tabla de simulaciones - Bernoulli skew probit.	33
4.12. Tabla de resultados - Bernoulli Skew Probit	33
5.1. Criterios de selección de los modelos ajustados.	40
5.2. Resúmenes a posteriori del Modelo Skew Probit. Media a posteriori, desviación estándar a posteriori, Intervalo de credibilidad del 95 % ($Q_{2.5}$, $Q_{97.5}$) y mediana (Q_{50}).	41
5.3. Evaluación de la capacidad predictiva de los modelos ajustados en los datos de prueba.	43
6.1. Tabla 1 de definición de covariables	45
6.2. Tabla 2 de definición de covariables	46

Capítulo 1

Introducción

1.1. Planteamiento y justificación del tema

Se entiende como anemia a la enfermedad cuando un cuerpo no tiene suficientes glóbulos rojos sanos o hemoglobina para transportar oxígeno a los tejidos del cuerpo. Teniendo como principales consecuencias en los niños cansancio y palidez, incremento de sueño, bajo rendimiento escolar, aumento de riesgo de infecciones, inadecuado desarrollo cerebral, entre otros. Sabiendo que la principal causa de esta enfermedad es la deficiencia de hierro en el organismo (Miniero et al., 2018).

En base a la información publicada por el Ministerio de Salud, basada en los datos proporcionados anualmente por la Encuesta Nacional de Demografía y Salud (ENDES), se ha notado un estancamiento en la reducción de la desnutrición crónica infantil en menores de cinco años de edad en el Perú (Tavera Salazar et al., 2023). Ante lo descrito anteriormente, es importante proponer modelos estadísticos que permitan identificar variables o factores que expliquen la presencia de anemia infantil para así poder disminuir la tasa anémica infantil.

En este contexto, los modelos estadísticos buscan predecir el comportamiento de una variable en base a otras con el fin de poder saber cómo se comportaría esta variable objetivo en diferentes situaciones, de acuerdo con las variables explicativas. Uno de los modelos estadísticos más utilizados para esta necesidad es la regresión lineal, la cual conlleva diferentes supuestos, entre los más importantes es el de asumir una distribución normal para la variable objetivo. Debido a ello, cuando aparecen variables que no cumplen este supuesto de normalidad, generalmente se acude a las transformaciones de estas para poder llevarlos a una distribución normal. Sin embargo, al realizar esto se pierde interpretabilidad. Otra manera de sobrellevar este problema es el de utilizar los modelos lineales generalizados (MLG), los cuales son muy utilizados cuando la variable objetivo tiene una distribución perteneciente a

la familia exponencial, por ejemplo, distribuciones Bernoulli y Binomial, que son de interés para modelar la presencia de anemia. Estas distribuciones son de gran utilidad cuando se presentan datos de tipo binario o proporciones.

Tres componentes identifican básicamente un MLG: Un componente aleatorio, que caracteriza a la variable respuesta Y y su distribución de probabilidad; un componente sistemático, especificado por las variables explicativas o covariables que sirven para explicar el comportamiento de la variable respuesta y en particular se usan con el objetivo de predicción de la variable respuesta; y la función de enlace que adecua la media de la variable respuesta, $E(Y)$, en el rango especificado del predictor lineal (McCullagh, 1989).

Un ejemplo muy claro donde se aplica los modelos lineales generalizados es cuando se tiene una variable dicotómica como variable respuesta, es decir un comportamiento binario o cuando se tiene proporciones. En estos casos se aplica, dentro de la familia de modelos lineales generalizados, el modelo de regresión logística el cual, por ejemplo en el caso binario, es capaz de predecir si se tiene uno de los dos valores dicotómicos. Otra de los posibles modelos a utilizar ante una situación de variable respuesta de estos tipos es el modelo probit el cual asume como función de enlace la inversa de la función de distribución acumulada de una normal. Este modelo asume igualdad en la velocidad de los ratios de respuesta para las dos alternativas binarias. Sin embargo, a veces podemos encontrar escenarios donde la velocidad de las ratios de respuesta para las respuestas binarias no son las mismas y como consecuencia la función de enlace (función de distribución acumulada normal) no es la óptima en estas situaciones.

Al emplear funciones de enlaces simétricas en modelos de regresión logit o probit puede generar sobre o sub estimación de la media de la probabilidad respectiva de éxito y estimación errónea de la desviación estándar de los parámetros en cuestión. Debido a ello, se puede usar un modelo asimétrico como función de enlace, el cual pueda discriminar las tasas de respuesta dicotómicas (Chen y Qi-Man Shao, 1999). Este modelo asimétrico se puede ver como una modificación de un modelo simétrico junto con un parámetro adicional el cual mide la asimetría (α). En particular, la regresión skew-probit es una de las opciones de regresión con función de enlace de distribución normal asimétrica, esta es una generalización de la regresión probit, la función de enlace es la inversa de la función de distribución acumulada (fda) de la distribución asimétrica skew-normal en lugar de la función de distribución acumulada de una normal estándar.

En la regresión skew-probit, la mayoría de propuestas usan la fda de una skew-normal, con parámetros de locación $\xi = 0$ y escala $\omega = 1$, para que la función de enlace sea definida de

forma similar a la de una normal estándar. La identificabilidad del parámetro de asimetría α , de la skew normal suele ser un problema. Es por ello que, desde un punto de vista bayesiano, se debe definir una distribución a priori apropiada para este parámetro que permita su estimación. Para la identificabilidad del parámetro de asimetría (α) se han propuesto una serie de distribuciones a priori como la distribución a priori, de Jeffrey (Liseo, 1990), a priori gaussiana truncada (Arellano-Valle et al., 2007), a priori t-student y aproximación de a priori de Jeffrey (Bayes y Branco, 2007), a priori uniforme (Azevedo et al., 2011), a priori coincidencia de probabilidad (Cabras et al., 2012), etc. Sin embargo, en general estas distribuciones a priori no son invariantes bajo una reparametrización del parámetro de asimetría.

En este contexto, Van Niekerk y Rue (2021) recientemente han propuesto un modelo de regresión skew-probit, donde desde un punto de vista inferencial el modelo skew normal se “reduce” hacia un modelo gaussiano. Para lo mencionado anteriormente, se abordaron problemáticas como estandarización de la función de enlace, identificabilidad, parámetro relacionado con la asimetría, y finalmente se construyó una distribución a priori penalizada compleja para el parámetro de asimetría (Simpson et al., 2017). Esta nueva distribución a priori se basa en la divergencia de Kullback-Leibler y garantiza que sea invariante a reparametrizaciones.

Dado que la distribución a posteriori del modelo de regresión skew-probit no tiene la forma de una distribución conocida, la inferencia bayesiana podría realizarse usando métodos de Monte Carlo vía cadenas de Markov (MCMC) o el método de aproximación de integración anidada de Laplace (INLA). Este último método se aplicará en la presente tesis, ya que presenta ventajas sobre el MCMC sobre todo en términos del costo computacional.

1.2. Objetivos de la tesis

En esta tesis se propone aplicar el modelo de regresión skew-probit, propuesto por Van Niekerk y Rue (2021) a datos de salud, en particular para estimar la presencia o ausencia de anemia en niños menores de cinco años en el Perú, en el año 2022. Para estimar los parámetros del modelo se usará inferencia bayesiana a través del INLA.

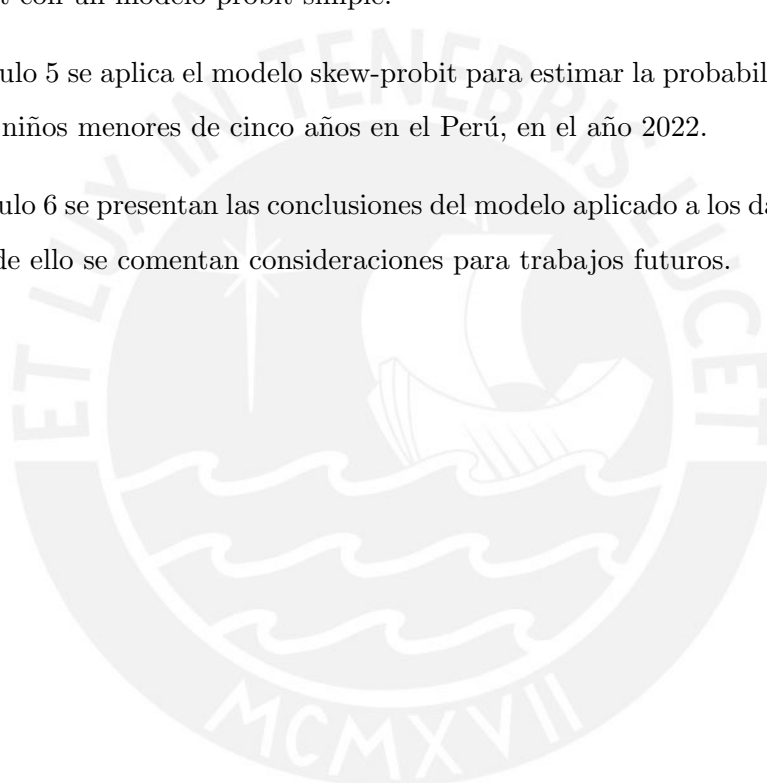
1.3. Organización del trabajo

La presente tesis está organizada de la siguiente manera:

- En el capítulo 2 se presenta un marco teórico de los conceptos fundamentales a tratar

en la tesis. Este va desde lo más básico como la definición de una distribución normal, un modelo lineal generalizado, hasta inferencia bayesiana utilizando la metodología de aproximación integrada anidada de Laplace (INLA).

- En el capítulo 3 se presenta el modelo a aplicar en la presente tesis, y se busca explicar las principales características del mismo.
- En el capítulo 4 se realiza un estudio de simulación de datos provenientes de una distribuciones binomiales, donde se busca analizar el comportamiento del modelo bajo diferentes escenarios de los parámetros; y además comparar el enfoque del modelo skew-probit con un modelo probit simple.
- En el capítulo 5 se aplica el modelo skew-probit para estimar la probabilidad de anemia infantil en niños menores de cinco años en el Perú, en el año 2022.
- En el capítulo 6 se presentan las conclusiones del modelo aplicado a los datos de anemia, y además de ello se comentan consideraciones para trabajos futuros.



Capítulo 2

Conceptos y modelos

En este capítulo se presentan algunos conceptos y consideraciones que serán usados en el desarrollo de esta tesis. La explicación se realizará desde conceptos más generales hasta conceptos con más detalle en base a lo que se requiere en la tesis. Una revisión de estos conceptos puede encontrarse en diferentes fuentes tal como Fahrmeir L. (2013).

2.1. Distribución normal

Una de las distribuciones de probabilidad continuas más importante en todo el campo de la estadística es la distribución normal. La gráfica de su función de densidad, denominada curva normal, es la curva con forma de campana, la cual describe de manera aproximada muchos fenómenos que ocurren en la naturaleza; es decir, se dice que es la distribución de uso más generalizado para la distribución de una variable aleatoria (Montgomery y Runger, 2003).

Una variable aleatoria X sigue una distribución normal con media μ y varianza σ^2 se denota por $X \sim N(\mu, \sigma^2)$, y tiene función densidad dada por:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0. \quad (2.1)$$

donde $\pi = 3.14159$ y $e = 2.71828$. La función de densidad de una variable aleatoria normal definida en la ecuación (2.1) tiene una forma acampanada y simétrica.

La media y la varianza de una variable aleatoria $X \sim N(\mu, \sigma^2)$, son características de la distribución, y son expresadas por $E(X) = \mu$ y $V(X) = \sigma^2$. Una vez establecidos estos parámetros, generan la forma de campana de la función de densidad de probabilidad (fdp) de $X \sim N(\mu, \sigma^2)$, se ve en la Figura 2.1. La media μ hace que la curva se traslade a lo largo del eje de las abscisas (valores de X), y la varianza hace que la curva sea más angosta si σ^2

es pequeña, o más ancha si la varianza σ^2 es más grande.

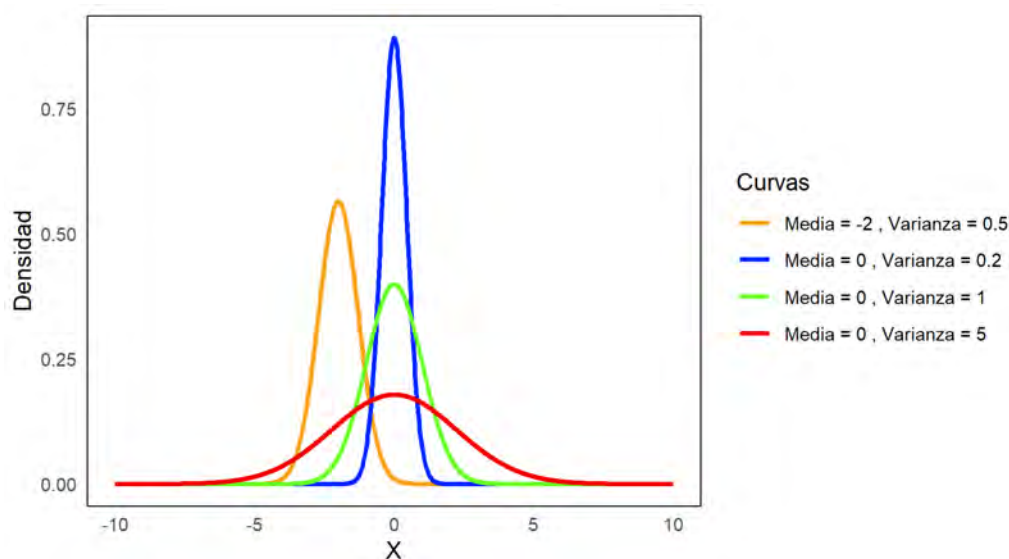


Figura 2.1: Función de densidad de variables aleatorias con distribución normal con media μ y varianza σ^2 .

La función de distribución acumulada (fda) de $X \sim N(\mu, \sigma^2)$ es:

$$\Phi_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx. \quad (2.2)$$

La fda anteriormente descrita en la ecuación (2.2) se representa en la Figura 2.2, tomando en cuenta diferentes valores de μ y σ . Se puede observar que para el máximo valor de X la probabilidad acumulada tiende a 1.

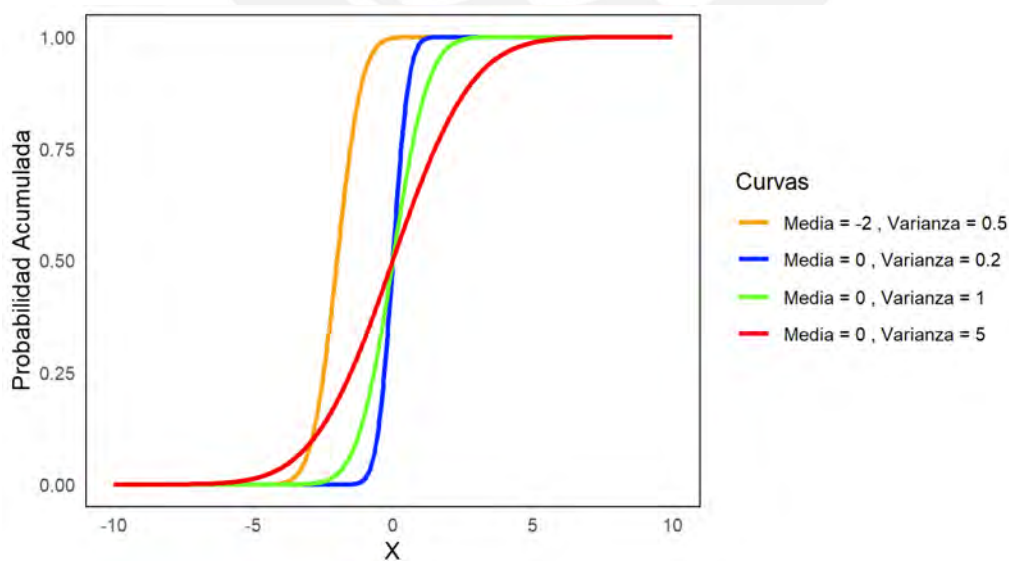


Figura 2.2: Función de distribución acumulada de variables aleatorias con distribución normal con media μ y varianza σ^2 .

2.2. Distribución skew-normal

El término skew-normal (normal asimétrica) o SN se refiere a una clase paramétrica de distribuciones de probabilidad que incluye la normal estándar como caso especial. Esta distribución generaliza a la distribución normal, permitiendo que el sesgo sea distinto de cero.

Sea X una variable aleatoria que sigue una distribución skew-normal, denotada por $X \sim SN(\xi, \omega, \alpha)$, donde ξ es un parámetro de ubicación, ω un parámetro de escala y α es un parámetro de forma o sesgo (Azzalini, 1985). La función de densidad de probabilidad para $X \sim SN(\xi, \omega, \alpha)$ es:

$$f_X(x) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\alpha \left(\frac{x - \xi}{\omega}\right)\right), \quad (2.3)$$

donde ϕ y Φ son las funciones de densidad y de distribución acumulada de una distribución normal estándar, respectivamente.

El componente α es denominado parámetro de forma (shape parameter) porque este regula la forma de la función de densidad. Por ejemplo, en la Figura 2.3, muestra las funciones de densidad en la ecuación (2.3), donde se puede observar el efecto de variar el parámetro de sesgo α , cuando $\xi = 0$ y $\omega = 1$. Se observa que cuando el sesgo es cero, la función de densidad es simétrica, cuando el sesgo es negativo la función de densidad es asimétrica hacia la izquierda, mientras que cuando el sesgo es positivo la la función de densidad es asimétrica hacia la derecha.

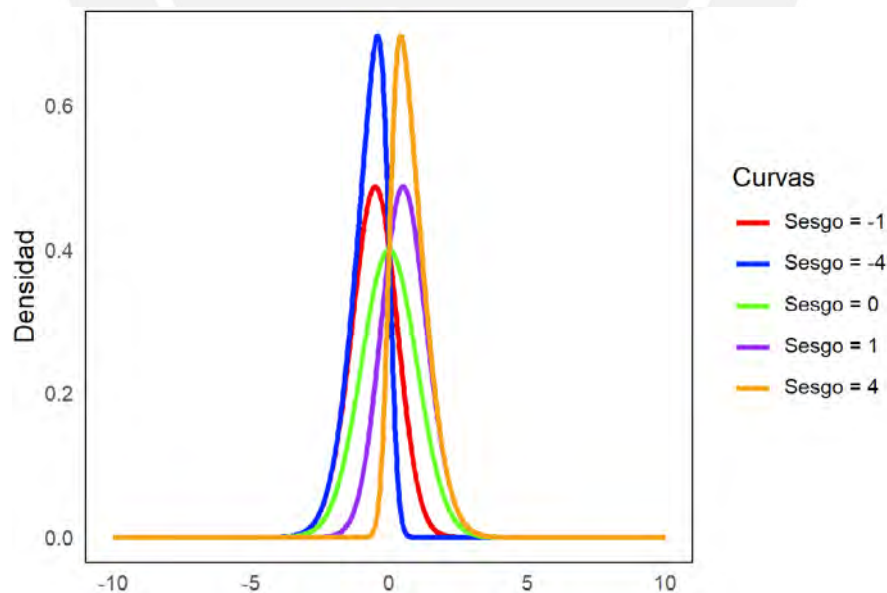


Figura 2.3: Función de densidad de variables aleatorias skew normal con $\xi = 0$ y $\omega = 1$ y diferentes valores para el parámetro de sesgo α .

Específicamente el parámetro α de la función de densidad de una skew-normal presenta varias propiedades interesantes como las siguientes:

- Cuando $\alpha = 0$, la asimetría desaparece y obtenemos la densidad normal estándar.
- A medida que aumenta α (en valor absoluto), aumenta la asimetría de la distribución.
- Cuando $\alpha \rightarrow \infty$, la densidad converge a la llamada función de densidad seminormal (o plegada normal).
- Si el signo de α cambia, la densidad se refleja en el lado opuesto del eje vertical.

La función de distribución acumulada de $X \sim SN(\xi, \omega, \alpha)$ está dada por:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

La función anteriormente descrita se representa en la Figura 2.4, tomando en cuenta diferentes valores de α . Se puede observar que para el máximo valor de X la probabilidad acumulada tiende a 1.

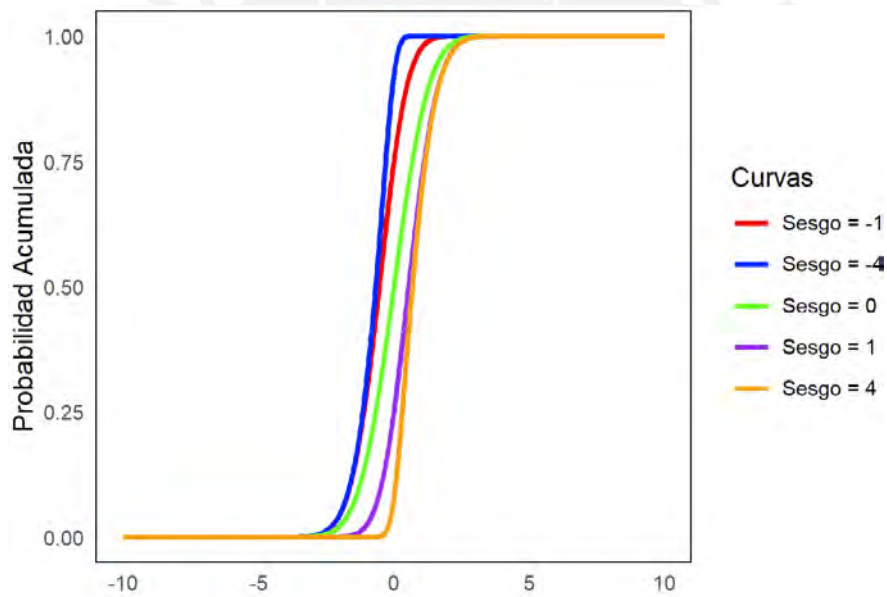


Figura 2.4: Función de distribución acumulada de variables aleatorias skew normal con cuando $\xi = 0$ y $\omega = 1$ y diferentes valores para el parámetro de sesgo α .

La media y la varianza para $X \sim SN(\xi, \omega, \alpha)$ son:

$$E(x) = \xi + \omega \delta \sqrt{\frac{2}{\pi}},$$

$$V(x) = \omega^2 \left(1 - \frac{2\delta^2}{\pi} \right),$$

donde $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$.

2.3. Modelo lineal generalizado (MLG)

Según (Faraway, 2004) el modelo lineal generalizado es aplicado para evaluar y cuantificar la relación entre un grupo de variables regresoras y una variable respuesta, pero este modelo se diferencia del modelo de regresión ordinaria en dos aspectos importantes:

- La distribución de la variable respuesta pertenece a la familia exponencial. Por lo que no necesariamente la variable respuesta tiene que tener una distribución normal.
- La transformación de la media de la respuesta se relaciona linealmente con las variables explicativas.

2.3.1. Componentes de un MLG

Un modelo lineal generalizado tiene tres componentes básicos:

a) Componentes aleatorios

Consiste en variables aleatorias Y_i independientes las cuales siguen una distribución perteneciente a la familia exponencial,

$$Y_i \stackrel{ind}{\sim} FE(\mu_i, \phi), \quad i = 1, \dots, n.$$

donde $E(Y_i) = \mu_i$ es la media de Y_i , y ϕ es un parámetro de escala o dispersión, y la variable aleatoria Y_i pertenece a la familia exponencial si su fdp tiene la siguiente forma:

$$f(y_i) = c(y_i, \phi) \exp \left\{ \frac{y_i \theta - a(\theta)}{\phi} \right\},$$

donde $c(\cdot)$ y $a(\cdot)$ son funciones conocidas.

b) Componente sistemático

Especifica las variables explicativas que se incorporan en forma de efectos fijos en la ecuación de predictor lineal, tal que:

$$\eta_i = \sum_j \beta_j x_{ij}, \quad j = 1, \dots, p.$$

donde x_{ij} es el valor de la j -ésima covariable del i -ésimo individuo y β_j el coeficientes de regresión asociado a la j -ésima covariable.

c) **Función de enlace**

Se define el valor esperado de la variable aleatoria Y_i como $\mu_i = E(Y_i)$, entonces la función de enlace especifica una función $g(\cdot)$ que relaciona el componente aleatorio y sistemático, de tal modo que:

$$\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij}.$$

Estos modelos generalizan la regresión ordinaria de dos modos: permitiendo que Y_i tenga distribuciones diferentes a la normal y, por otro lado, incluyendo distintas funciones de enlace para la media. Los modelos de regresión lineal normal para respuestas continuas son un caso particular de los GLM.

En las Figuras 2.5 y 2.6 podemos observar la diferencia de realizar un modelo lineal típico asumiendo distribución normal para la variable respuesta Y_i , y ajustar un modelo GLM para los mismos datos, el cual se adecua más a la forma de los datos usando una función de enlace logarítmica.

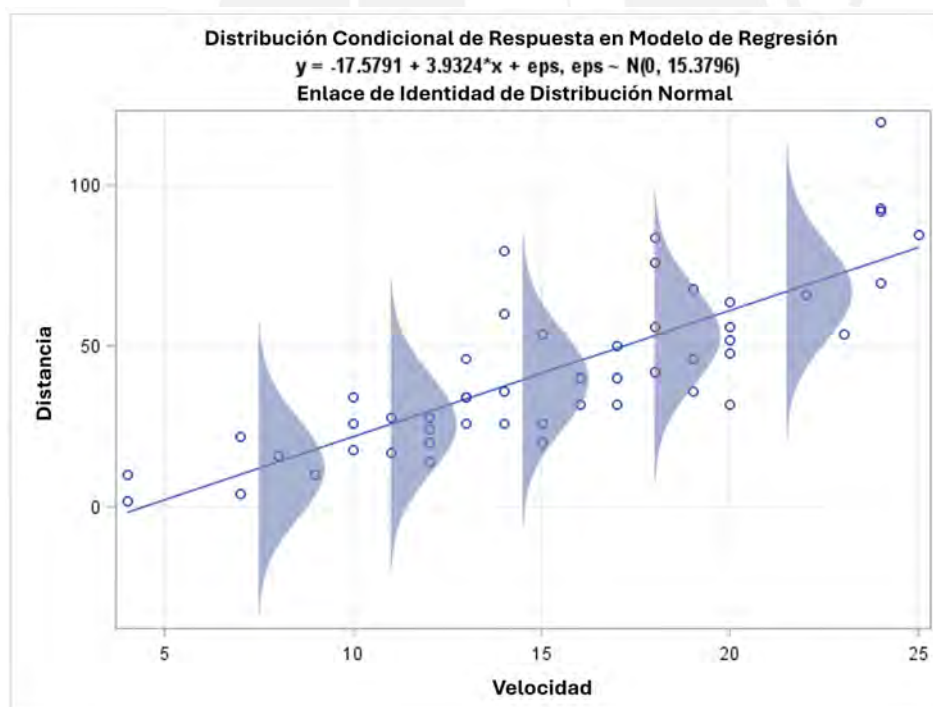


Figura 2.5: Distribución condicional de respuesta en modelo de regresión con distribución normal y función de enlace identidad. Fuente: Wicklin (2015).

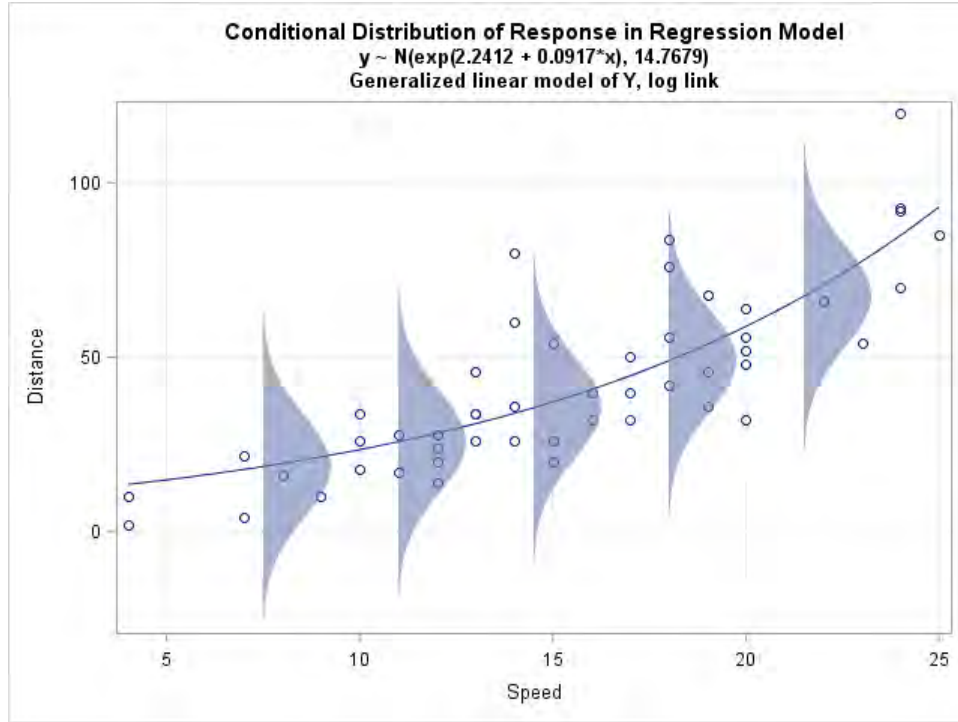


Figura 2.6: Distribución condicional de respuesta en modelo de regresión con distribución normal y función de enlace logarítmica. Fuente: Wicklin (2015).

2.3.2. Regresión binaria

En muchos casos la variable respuesta Y_i responde solo a dos valores si o no, verdadero o falso, etc. Es decir Y_i sigue una distribución Bernoulli, caso particular de una binomial denotada por $Y_i \sim Bin(1, \pi_i)$. Su media y varianza son:

$$E(Y_i) = \mu_i = \pi_i \quad \text{y} \quad V(Y_i) = \pi_i(1 - \pi_i).$$

Al aplicar un modelo lineal generalizado se tiene lo siguiente:

$$Y_i \stackrel{ind}{\sim} B(1, \pi_i), \quad g(\pi_i) = x_i^\top \beta,$$

donde g es la función de enlace del MLG. Para esta función podemos tomar distintas alternativas, como son la función logit y probit entre las más conocidas (Faraway, 2004).

Regresión logística

En el modelo de regresión logística, se usa como función de enlace la función logit:

$$g(\mu_i) = \text{logit}(\mu_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = x_i^\top \beta \quad \Rightarrow \quad \pi_i = \frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}}.$$

La función de enlace logit de π_i asegura que no exista ningún problema estructural respec-

to al rango de valores de la probabilidad π_i , y además el parámetro (o vector de parámetros) β determina el rango y la velocidad de incremento y decremento de la curva (Faraway, 2004).

Regresión probit

Otro modelo alternativo para definir una regresión binaria utilizando MLG es el modelo de regresión probit. Este utiliza como función de enlace la inversa de la función de distribución acumulada de una distribución normal (Faraway, 2004), como se presenta a continuación:

$$g(\mu_i) = \Phi^{-1}(\pi_i) = x_i^T \beta \quad \Rightarrow \quad \pi_i = \Phi(x_i^T \beta).$$

En la Figura 2.6 se aprecia las curvas de regresión lineal, logit y probit frente a un conjunto de datos de respuesta binaria (puntos negros). Se puede observar que los modelos de regresión probit y logit se ajustan de mejor manera a la variable respuesta Y_i .

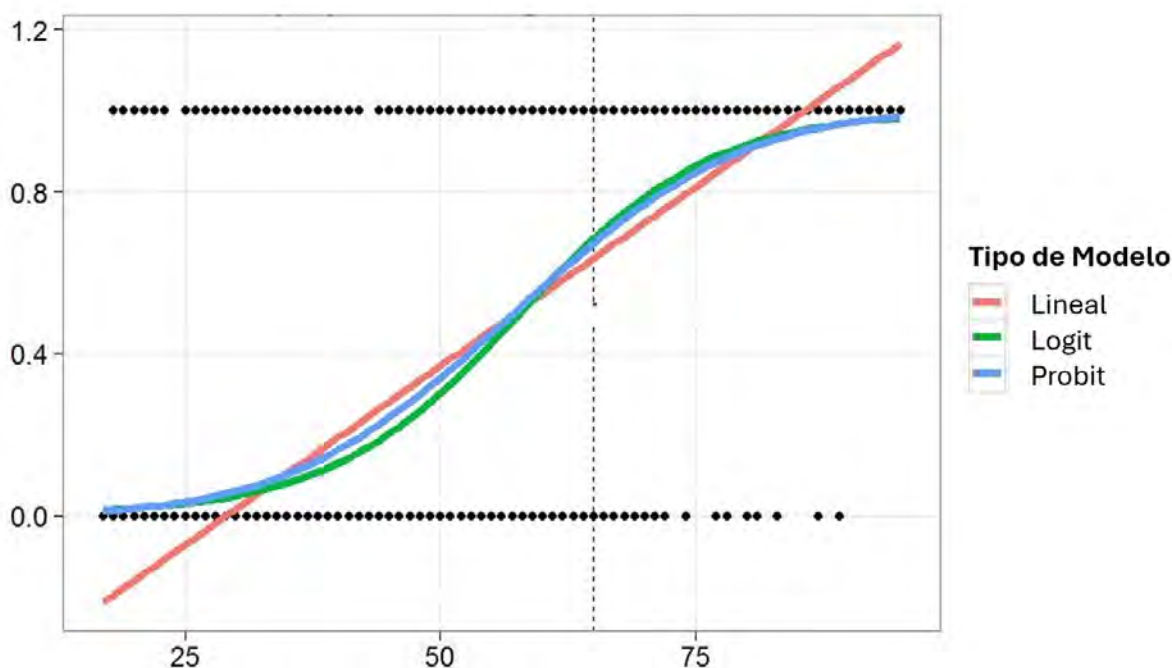


Figura 2.7: Comparación de regresión lineal simple, logit y probit para datos binarios.

Fuente: Jaramillo (2024)

2.4. Inferencia bayesiana

La inferencia bayesiana asume una distribución de probabilidad sobre los parámetros del modelo. Específicamente, para realizar inferencia bayesiana necesitamos tres componentes principales que son los siguientes:

- **Función de verosimilitud:** Resume la información que traen los datos sobre el parámetro $\theta \in \Theta$. Es definida por $\pi(\mathbf{y} | \theta)$. En caso Y_i sean condicionalmente indepen-

dientes dado θ , se tiene que:

$$\pi(\mathbf{y} | \theta) = \pi(y_1, \dots, y_n | \theta) = \prod_{i=1}^n \pi(y_i | \theta).$$

- **Distribución a priori:** Incorpora la información inicial que tiene el investigador sobre el parámetro (o vector de parámetros) θ . La distribución a priori de θ se denota como $\pi(\theta)$.
- **Distribución a posteriori:**

Es obtenida luego de observar los datos. Es una actualización de la distribución a priori. Simplemente condicionando θ a los datos Y , utilizando la propiedad básica de probabilidad condicional conocida como el Teorema de Bayes, se obtiene la función de probabilidad o densidad a posteriori:

$$\pi(\theta | \mathbf{y}) = \frac{\pi(\theta, \mathbf{y})}{\pi(\mathbf{y})} = \frac{\pi(\theta)\pi(\mathbf{y} | \theta)}{\pi(\mathbf{y})}.$$

La ecuación anterior se puede aproximar de la siguiente manera:

$$\pi(\theta | \mathbf{y}) \propto p(\theta)p(\mathbf{y} | \theta).$$

La Figura 2.8 muestra gráficamente un ejemplo de distribución a priori, función de verosimilitud y la distribución a posteriori correspondiente.

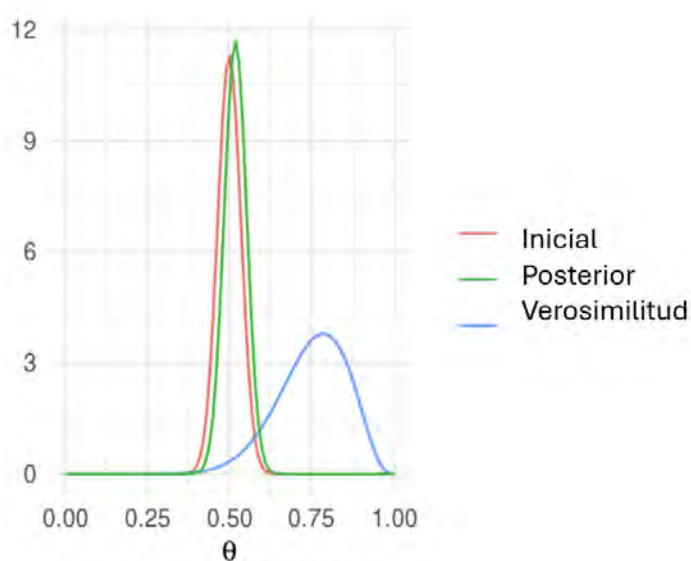


Figura 2.8: Función de verosimilitud (línea azul), distribución a priori (inicial, línea roja) y distribución a posteriori (línea verde).

2.4.1. Monte Carlo vía cadenas de Markov (MCMC)

En muchas ocasiones el análisis bayesiano requiere calcular integraciones de alta dimensión, ya sea para realizar inferencias sobre los parámetros de un modelo o para realizar predicciones. En el pasado se resolvía este problema a través de métodos de integración numérica. Sin embargo este problema se agrava cuando se trata de resolver integraciones en mayor dimensión para modelos complejos. Una solución es considerar un procedimiento Monte Carlo iterativo o Monte Carlo vía Cadenas de Markov (MCMC, Gamerman y Lopes (2006)).

Los métodos MCMC son algoritmos iterativos que se utilizan cuando el muestreo directo de la distribución de interés $\pi(\boldsymbol{\theta} \mid \mathbf{y})$ no es factible, donde $\boldsymbol{\theta}$ es un vector de parámetros. La metodología MCMC es una herramienta de gran alcance para la modelación estadística y se ha vuelto muy popular en la computación bayesiana en modelos estadísticos de gran complejidad. Se simula una cadena de Markov hasta llegar a la distribución estacionaria dada por la distribución a posteriori $\pi(\boldsymbol{\theta} \mid \mathbf{y})$. Las características (media, varianza, etc) de la distribución a posteriori son obtenidas encontrando promedios ergódicos.

Se han desarrollado soluciones generales utilizando MCMC en software como WinBUGS, OpenBUGS. Sin embargo, el MCMC puede ser lento, no escalar bien y, en algunos casos, pueden no converger las cadenas de Markov sobre todo en modelos complejos (Brooks, 2011). Para abordar estos desafíos, se ha desarrollado algoritmos más eficientes como los implementados en JAGS y Stan (Kruschke, 2014).

Muestreador de Gibbs

El muestreador de Gibbs o Gibbs sampling es un método de simulación que permite obtener muestras de la distribución a posteriori conjunta. Este método requiere del muestreo de las distribuciones condicionales completas, por ello estas deben ser distribuciones conocidas (Gelfand, 2000).

Específicamente, para obtener una muestra de la distribución conjunta

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}) = \pi(\theta_1, \dots, \theta_d \mid \mathbf{y}),$$

el muestreador Gibbs obtiene muestras de cada parámetro usando su distribución condicional completa, y realiza el siguiente muestreo de forma iterativa hasta obtener el número deseado de muestras, tal que se haya llegado a la convergencia de todas las cadenas. Los pasos del Gibbs sampling se resumen a continuación, sea $t = 0$, entonces:

- Muestree $\theta_1^{(t+1)}$ de $\pi(\theta_1 \mid \theta_2^{(t)}, \dots, \theta_d^{(t)})$

- Muestree $\theta_2^{(t+1)}$ de $\pi\left(\theta_2 \mid \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_d^{(t)}\right)$
- \vdots
- Muestree $\theta_d^{(t+1)}$ de $\pi\left(\theta_d \mid \theta_1^{(t+1)}, \dots, \theta_{d-1}^{(t+1)}\right)$.
- Realice los pasos en los items anteriores para $t = t + 1$, hasta llegar a la convergencia de todas las cadenas.

2.4.2. Aproximación de Laplace integrada anidada (INLA)

Los modelos jerárquicos son comúnmente utilizados en inferencia bayesiana para representar estructuras de dependencia complejas en los datos. La incertidumbre asociada a los parámetros del modelo y a las variables o procesos latentes puede ser capturada mediante la especificación de distribuciones a priori apropiadas dentro de un enfoque bayesiano. Sin embargo debido a la complejidad de los modelos jerárquicos la inferencia es un reto. En general, se requieren métodos de estimación más complejos como el MCMC o el INLA. A diferencia de MCMC, que es un método asintóticamente exacto, INLA es una aproximación. Aunque puede generar preguntas sobre la precisión de los resultados, estudios de simulación y experiencias prácticas han demostrado que, en muchos casos, el error asociado al INLA es comparable al de MCMC (Muff et al., 2015).

El método INLA fue propuesto por Rue et al. (2009), ofrece una alternativa rápida a MCMC específicamente para modelos gaussianos latentes (LGM) y puede aplicarse a una amplia gama de modelos estadísticos, incluyendo modelos lineales generalizados (MLG), modelos aditivos generalizados (GAM), series temporales, modelos espaciales, entre otros. Para poder aplicar INLA, tenemos que: (i) ajustar un modelo gaussiano Latente (LGM), (ii) tener en el segundo nivel un campo aleatorio Markoviano gaussiano (CAMG) y (iii) usar la aproximación de Laplace.

Modelo gaussiano Latente (LGM)

Un LGM es un modelo jerárquico compuesto por los siguientes niveles:

- Vector aleatorio (\mathbf{Y}): Sea $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ donde las v.a.'s Y_i condicionadas a $\mathbf{X} = \mathbf{x}$, $\boldsymbol{\theta} = \boldsymbol{\theta}$ son independientes. Luego, la fdp conjunta de \mathbf{Y} es

$$\pi(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^n \pi(y_i \mid \mathbf{x}, \boldsymbol{\theta}).$$

- Campo Gaussiano Latente (\mathbf{X}): Si $\mathbf{X} = (X_1, \dots, X_J)$ condicionado a $\boldsymbol{\theta}$ tiene una dis-

tribución $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. Si de forma general:

$$g(E(\mathbf{Y})) = \beta_0 \mathbf{1} + \mathbf{A}\boldsymbol{\beta} + \mathbf{f};$$

luego $\mathbf{X} = \{\beta_0, \boldsymbol{\beta}, \mathbf{f}\}$;

- Efecto "fijo": intercepto β_0
 - \mathbf{A} : covariables, Efectos "fijos" $\boldsymbol{\beta}$,
 - Efectos estructurados: \mathbf{f} , por ejemplo estructuras espaciales, temporales, etc.
- Hiperparámetros ($\boldsymbol{\theta}$): Sea el vector de hiperparámetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p) \in \Theta$, tal que la fdp de $\boldsymbol{\theta}$ es:

$$\pi(\boldsymbol{\theta}),$$

donde $\boldsymbol{\theta}$ es compuesto por los parámetros desconocidos restantes. Por ejemplo:

- Parámetros de escala de la distribución de \mathbf{Y} .
- Parámetros de los efectos estructurados o no estructurados, generalmente precisiones.

Campos aleatorios markovianos gaussianos (CAMG)

Son un tipo específico de modelo estadístico muy utilizado para representar la dependencia espacial o temporal en datos. Se basan en la distribución gaussiana multivariante y en la propiedad de Markov para modelar la dependencia condicional entre las variables aleatorias.

Decimos que \mathbf{x} es un CAMG con media $\boldsymbol{\mu}$ y matriz de precisión (covarianza inversa) \mathbf{Q} si su densidad tiene la forma

$$\pi(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{Q}|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

tal que la matriz de precisión está llena de ceros. Específicamente, un CAMG es un vector aleatorio gaussiano $\mathbf{x} = (x_1, \dots, x_n)$ con propiedad de Markov, tal que x_i y x_j son condicionalmente independientes al resto de variables x_{-ij} , para $i \neq j$. Debido a esta propiedad la matriz \mathbf{Q} : $Q_{ij} = 0$ si y solo si x_i and x_j son condicionalmente independientes. El patrón de elementos cero y distintos de cero en dicha matriz se denomina estructura de dispersión. Esto último describe la fuerza y la dirección de la dependencia condicional entre las variables.

En particular, el INLA requiere que el segundo nivel del modelo gaussiano latente descrito en la sección anterior sea un CAMG.

Aproximación de Laplace

INLA calcula de forma rápida y precisa las aproximaciones de las distribuciones marginales a posteriori de las componentes del campo gaussiano latente, definidas por:

$$\pi(x_i | \mathbf{y}) = \int \pi(x_i | \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad i = 1, \dots, J, \quad (2.4)$$

así como la distribución marginal para los hiperparámetros del modelo gaussiano latente, definidas por:

$$\pi(\theta_j | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j}, \quad j = 1, \dots, \dim(\boldsymbol{\theta}), \quad (2.5)$$

donde $\boldsymbol{\theta}_{-j}$ representa el conjunto de los parámetros en $\boldsymbol{\theta}$ sin θ_j .

La función de densidad marginal de los hiperparámetros es aproximada usando una aproximación gaussiana de la distribución condicional completa del campo latente, $\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$, evaluada en la moda $x^*(\boldsymbol{\theta}) = \arg \max_x \pi_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$, es decir,

$$\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=x^*(\boldsymbol{\theta})},$$

esta aproximación es una aproximación de Laplace.

La función de densidad marginal de cada hiperparámetro, $\pi(\theta_j | \mathbf{y})$ definida en la ecuación (2.5), es calculada usando la aproximación de Laplace de $\pi(\boldsymbol{\theta} | \mathbf{y})$ e integrando en $\boldsymbol{\theta}_{-j}$, esto es,

$$\tilde{\pi}(\theta_j | \mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j},$$

luego esta integral se resuelve numéricamente.

La aproximación $\pi(x_i | \mathbf{y})$ puede ser obtenida utilizando una aproximación gaussiana, de Laplace o una aproximación simplificada de Laplace. En algunas situaciones, la aproximación gaussiana produce errores en la localización y falla para detectar un comportamiento asimétrico. La aproximación de Laplace es preferible a la gaussiana, pero es más costosa. La aproximación de Laplace simplificada (que es la que viene implementada por defecto en el paquete R-INLA) tiene un costo más pequeño y remedia los problemas que surgían con la aproximación gaussiana referente a la asimetría y la localización.

Después INLA aproxima la ecuación (2.4) construyendo la siguiente aproximación anidada:

$$\tilde{\pi}(x_i | \mathbf{y}) = \int \tilde{\pi}(x_i | \boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta},$$

finalmente, esta integral aproximada puede ser obtenida numéricamente.

Capítulo 3

Modelo de regresión skew probit

En base a los estudios e investigaciones de Van Niekerk y Rue (2021), ellos proponen la regresión probit asimétrica desde un punto de vista diferente a muchos otros autores de literatura estadística (Bazán et al., 2010; Lee y Sinha, 2019). A través de su investigación se enfrentan a tres principales observaciones y problemas, los cuales se detallarán a continuación, así como la solución planteada.

3.1. Estandarización de la función enlace

Parametrización usual

En primer lugar, se define y explica la regresión probit asimétrica con las consideraciones que muchos otros autores plantean (Lee y Sinha, 2019). Se define una variable aleatoria $Y_i \sim \text{Binomial}(N_i, \pi_i)$, con la siguiente función de enlace:

$$G^{-1}(\pi_i) = \eta_i,$$

donde $G(\cdot)$ es la fda de una variable aleatoria (v.a.) con distribución skew normal.

A continuación se define la fda de una skew normal para este modelo. De forma general, se asume que $X \sim SN(\xi, \omega, \alpha)$ es una v.a. con distribución skew normal con parámetros (ξ, ω, α) , los cuales hacen referencia a localización, escala y sesgo, respectivamente. Luego, la fda de X es:

$$G(x | \alpha) = \int_{-\infty}^x g(x | \alpha) dx, \quad (3.1)$$

y la función de densidad de X es:

$$g(x | \alpha) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left[\alpha \left(\frac{x - \xi}{\omega}\right)\right], \quad (3.2)$$

donde $(\alpha, \xi) \in \mathfrak{R}$, $\omega > 0$. Además $\phi(\cdot)$ y $\Phi(\cdot)$ son las funciones de densidad y función de distribución acumulada (fda) de la distribución estándar gaussiana.

La media de X es $E(X) = \xi + \omega\delta\sqrt{\frac{2}{\pi}}$ y la varianza de X es $V(X) = \omega^2(1 - \frac{2\delta^2}{\pi})$, donde $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$. La parametrización usual para ajustar el modelo skew-probit asume que el parámetro de localización $\xi = 0$ y de escala $\omega = 1$. Sin embargo, bajo este supuesto la media y la varianza de la variable X dependen del parámetro α (Lee y Sinha, 2019). Así si α cambia de valor entonces también cambian la media y varianza.

Nueva parametrización

Van Niekerk y Rue (2021) postulan una nueva propuesta, indicando que la media y varianza de la función de enlace deben ser ajustadas para ser definidas como en el modelo probit, en el sentido de que se defina una función de enlace a partir de la función de distribución acumulada de una normal asimétrica estandarizada, cuya media sea cero y varianza sea uno, no los parámetros de locación y escala, pues en el caso de la skew-normal no son la media y varianza.

Se define una variable aleatoria $Y_i \sim \text{Binomial}(N_i, \pi_i)$, con la siguiente función de enlace:

$$F^{-1}(\pi_i) = \eta_i,$$

donde $F(\cdot)$ es la fda de una variable aleatoria (v.a.) con distribución skew normal estandarizada.

A continuación se define la fda de una skew normal estandarizada. De forma general se asume que $X \sim \text{SNE}(\xi, \omega, \alpha)$ es una v.a. con distribución skew normal estandarizada si X tiene media cero y varianza 1, y se define $F(y|\alpha)$ como su función de distribución acumulada (fda), para todos los valores de α , es decir:

$$F(x|\alpha) = \int_{-\infty}^x f(x|\alpha)dx,$$

donde

$$f(x|\alpha) = \frac{2}{\omega(\alpha)}\phi\left(\frac{x - \xi(\alpha)}{\omega(\alpha)}\right)\Phi\left[\alpha\left(\frac{x - \xi(\alpha)}{\omega(\alpha)}\right)\right], \quad (3.3)$$

$$\xi(\alpha) = -\omega(\alpha)\alpha\sqrt{\frac{2}{\pi(1+\alpha^2)}},$$

$$\omega(\alpha) = \sqrt{\left(1 - \frac{2\alpha^2}{\pi(1+\alpha^2)}\right)^{-1}}.$$

De esta manera la media y la varianza son fijas y los parámetros ξ y ω dependen de α .

Explicada la primera problemática y la forma como lo maneja Van Niekerk y Rue (2021), se puede explicar de manera general cómo se aborda este modelo regresión skew probit. En base a las reglas básicas de inferencia bayesiana y teniendo claro las consideraciones explicadas en el primer punto de este capítulo, podemos definir la distribución Binomial para Y_i de la siguiente forma:

$$Y_i \mid \beta_0, \boldsymbol{\beta}, \alpha \sim \text{Bin}(N_i, \pi_i)$$

$$\pi_i = F(\eta_i)$$

$$\eta_i = \beta_0 + X_i^\top \boldsymbol{\beta},$$

donde X_i es un vector de covariables, β_0 es el efecto fijo llamado usualmente intercepto y $\boldsymbol{\beta}$ son los otros efectos fijos o coeficientes de regresión. Como podemos observar la variable aleatoria Y_i depende de los parámetros β_0 , $\boldsymbol{\beta}$ y α . Para estas últimas se tienen que definir sus distribuciones a priori respectivas. Para el vector $\boldsymbol{\beta}$, que contiene los coeficientes de regresión de la función sistemática del modelo lineal generalizado, podemos asignarle una distribución a priori de distribución normal con media 0 y una varianza amplia. Sin embargo, para el β_0 de intercepto se realizara más adelante en este capítulo un enfoque diferente explicado por Van Niekerk y Rue (2021). Por otro lado, para el parámetro α se asumirá una distribución a priori especial y además de ello en este capítulo se explicará la reparametrización que se aplicará a α . Explicado todo lo anterior podemos obtener la distribución a posteriori de los parámetros y así poder tener un modelo de regresión asimétrico en base a la propuesta de Van Niekerk y Rue (2021).

3.2. Intercepto cuantil e identificación de los parámetros

Como segunda problemática, se tiene la identificabilidad del parámetro de intercepto β_0 y α , ya que en particular computacionalmente cada vez que se estima el valor que se obtiene es muy diferente. Lo que planteó Van Niekerk y Rue (2021) es rectificar la formulación del intercepto introduciendo el cuantil de este mismo parámetro.

Para explicar lo anterior, primero nos basaremos en una regresión probit con un intercepto y una covariable X_i , donde la probabilidad de acierto se define de la siguiente manera:

$$\pi_i = P(Y_i = 1) = \Phi(\beta_0 + \beta_1 X_i).$$

Si $\beta_1 X = 0$ entonces $q = P(Y_i = 1) = \Phi(\beta_0) = P(Z < \beta_0)$, luego β_0 es el q -ésimo cuantil de una distribución normal estándar Z y hay una relación uno a uno entre β_0 y q . Se observa

además que no se ve afectado el intercepto β_0 porque sólo depende de la fda Φ normal estándar que no depende de otro parámetro. Esto significa que β_0 es desinformativo para los demás coeficientes de regresión.

La casuística anterior no ocurre cuando se tiene una regresión skewed probit estandarizada, donde la probabilidad de acierto es la siguiente:

$$\pi_i = P(Y_i = 1) = F(\beta_0 + \beta_1 X_i | \alpha).$$

De lo anterior no podemos afirmar que β_0 es desinformativo para los demás coeficientes de regresión ya que la fda F depende de α . A diferencia del modelo de regresión probit, si $\beta_1 \neq 0$ la probabilidad de acierto cambia, afectando a β_0 , porque la distribución F depende del parámetro de sesgo α . Para lidiar con esto Van Niekerk y Rue (2021) afirman que β_0 es desinformativo para los demás coeficientes de regresión si

$$q = P(Y_i = 1) = F(\beta_0 | \alpha), \quad (3.4)$$

es constante para cualquier α , y eso se cumple si β_0 es definido como el q -ésimo cuantil de una distribución skew-normal estandarizada X con fda F . Luego se reformula β_0 de la siguiente manera:

$$\beta_0(q, \alpha) = F^{-1}(q | \alpha).$$

Finalmente, el nivel de cuantil q es el parámetro de intercepto no conocido a estimar en vez de β_0 .

3.3. PC-prior para el parámetro de sesgo

La a priori penalizada compleja (“PC prior”) fue propuesta por Simpson et al. (2017) y es muy utilizada cuando se quiere cuantificar el hecho de contraer un modelo complejo a un modelo simple. Además de ello, esta es muy útil en casos cuando las distribuciones a priori son seleccionadas por convención o por conveniencia matemática.

A continuación se describirán los principios para poder construir la distribución a priori del parámetro de sesgo α .

3.3.1. Principio 1 - Occam’s razor

Muchos modelos complejos (flexibles) se pueden reducir a un modelo base, es decir el usuario puede definir una “versión simple” del modelo a tratar. Este principio se basa en que

existen muchos modelos complejos (modelo flexible) anidados en un modelo simple (modelo base). Bajo estas definiciones se desea comparar los modelos skew-probit y probit, que es equivalente a comparar las fdp de skew-probit y probit. De lo anterior se plantea el modelo flexible con fdp $f(x | \alpha)$ la cual es definida por la distribución de $X \sim \text{SNE}(\xi, \omega, \alpha)$. Y por otro lado se plantea el modelo base con fdp $f(x | \alpha = 0)$ donde el sesgo es cero, y es definida por la distribución $X \sim \text{SNE}(\xi, w, \alpha = 0)$.

3.3.2. Principio 2 - Medida de complejidad

Se usa la medida de divergencia Kullback-Leibler (KLD) con el objetivo de comparar las densidades de distintas distribuciones del modelo base y modelo flexible. El KLD mide la cantidad de información que se pierde al utilizar el modelo base, ya que el objetivo principal es aproximar un modelo complejo a otro más simple. El KLD de las fdp asociadas al modelo flexible y base es dado por:

$$\begin{aligned} \text{KLD}(\alpha) &= \int f(x | \alpha) \log \frac{f(x | \alpha)}{f(x | \alpha = 0)} dx \\ &= \frac{\pi^2 + 16 - 8\pi}{6\pi^3} \alpha^6 - \frac{144\pi + 3\pi^3 - 38\pi^2 - 168}{6\pi^4} \alpha^8 \\ &\quad + \frac{-42240\pi - 2560\pi^3 + 16176\pi^2 + 129\pi^4 + 39936}{120\pi^5} \alpha^{10} + \mathcal{O}(\alpha^{12}) \\ &\approx c_1 \alpha^6 + c_2 \alpha^8 + c_3 \alpha^{10}, \end{aligned}$$

donde las fdp $f(\cdot)$ se definen a partir de la ecuación (3.3).

3.3.3. Principio 3 - Ratio de penalización constante

El objetivo principal es definir una a priori para la distancia entre el modelo flexible y modelo base. Para ello se define esta distancia de la siguiente manera:

$$\begin{aligned} d(\alpha) &= \sqrt{2\text{KLD}(\alpha)} \\ &\approx \sqrt{2(c_1 \alpha^6 + c_2 \alpha^8 + c_3 \alpha^{10})}. \end{aligned}$$

Para construir la distribución PC prior para el parámetro de sesgo α se asigna una distribución a priori exponencial con parámetro θ a la distancia $d(\alpha)$. La distribución exponencial tiene una tasa de decaimiento constante $r = \exp(-\theta)$, por lo cual permite penalizar de la misma forma cada porción adicional de distancia en el espacio paramétrico de $d(\alpha)$.

A partir de la distribución exponencial de $d(\alpha)$ se puede hallar la distribución a priori

PC prior para α , definida como:

$$\begin{aligned}
 \pi(\alpha) &= \pi_d(d(\alpha)) \left| \frac{\partial(d(\alpha))}{\partial(\alpha)} \right| \\
 &= \frac{1}{2} \theta \exp[-\theta d(\alpha)] \left| \frac{\partial d(\alpha)}{\partial \alpha} \right| \\
 &\approx \frac{\theta}{2\sqrt{2(c_1\alpha^6 + c_2\alpha^8 + c_3\alpha^{10})}} |2(6c_1\alpha^5 + 8c_2\alpha^7 + 10c_3\alpha^9)| \\
 &\quad \times \exp\left[-\theta |\alpha^3| \sqrt{2(c_1 + c_2\alpha^2 + c_3\alpha^4)}\right],
 \end{aligned} \tag{3.5}$$

para valores pequeños de α .

3.3.4. Principio 4 - Escala definida por el usuario

El parámetro definido por el usuario θ se usa para gobernar la reducción hacia el modelo de regresión probit. Para una probabilidad $p_U > 0$, y a partir de la distribución exponencial de la distancia $d(\alpha) \sim \exp(\theta)$ se tiene que

$$P(d(\alpha) > U) = p_U = 1 - (1 - \exp(-\theta U)) = \exp(-\theta U),$$

de la cual se tiene que $\theta = -\log p_U / U$. El usuario puede definir la probabilidad p_U y U para definir su subjetividad para obtener el valor de θ en la PC prior de la ecuación 3.5. Un ejemplo de la PC prior para α cuando $\theta = 5$ se muestra en la Figura 3.1.

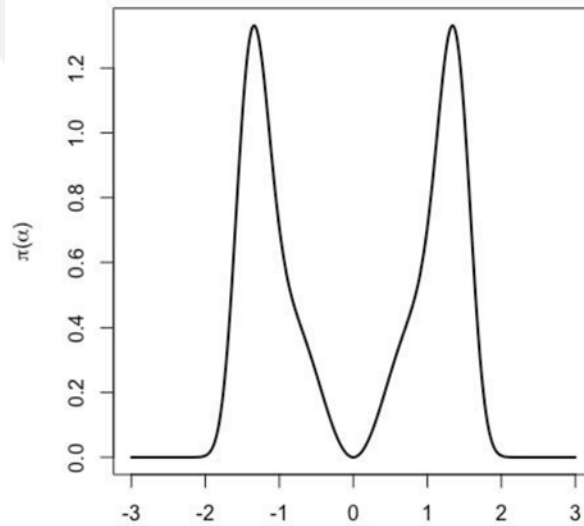


Figura 3.1: PC prior de α para $\theta = 5$. Fuente: Van Niekerk y Rue (2021).

3.3.5. Reparametrización del parámetro de sesgo

Ya definida la distribución a priori del parámetro de sesgo α se sabe que este parámetro tiene problemas en la estimación cuando toma valores alrededor de 0, debido al comportamiento de la fdp de α alrededor de cero como podemos ver en la Figura 3.1. Debido a ello se plantea reparametrizar el parámetro α a través del coeficiente de asimetría (estandarizado) de la distribución skew-normal, γ_1 , definido por:

$$\gamma_1 = \frac{(4 - \pi) \left(\sqrt{\frac{2\delta^2}{\pi}} \right)^3}{2 \left(1 - \frac{2\delta^2}{\pi} \right)^{\frac{3}{2}}},$$

donde $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$. Cabe resaltar que $-0.99527 < \gamma_1 < 0.99527$. Una ventaja de la PC prior es que es invariante bajo reparametrizaciones, así la información acerca de α dados los datos es la misma para α que para γ_1 . Por lo tanto proponer las PC priors para α o γ_1 son equivalentes en el sentido de que producen la misma inferencia a posteriori. En particular la fdp de la PC prior para γ_1 tiene un compartamiento asintótico y es una función monótona. Podemos ver su comportamiento en la Figura 3.2.

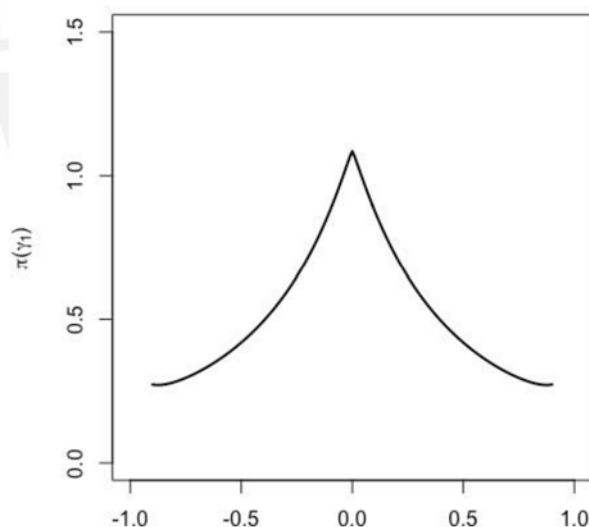


Figura 3.2: PC prior de γ_1 para $\theta = 5$. Fuente: Van Niekerk y Rue (2021).

3.4. Inferencia bayesiana

En base a la teoría explicada en el capítulo 2, se aplicará la metodología INLA para el caso del modelo de regresión skew probit, el cual se puede escribir como un modelo gaussiano latente:

- Vector aleatorio (\mathbf{Y}): Sea $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ donde las v.a.'s Y_i condicionadas a $\boldsymbol{\psi}, \boldsymbol{\theta}$ son independientes, específicamente $Y_i | \boldsymbol{\psi}, \boldsymbol{\theta} \stackrel{ind}{\sim} \text{Bin}(N_i, \pi_i)$. Luego, la fdp conjunta de \mathbf{Y} es

$$\pi(\mathbf{y} | \boldsymbol{\psi}, \boldsymbol{\theta}) = \prod_{i=1}^n \pi(y_i | \boldsymbol{\psi}, \boldsymbol{\theta}),$$

donde

$$F^{-1}(\pi_i) = \eta_i = \beta_0 + X_i^\top \boldsymbol{\beta},$$

donde F es la fda de una v.a. skew normal estandarizada que depende del parámetro γ_1 .

- Campo gaussiano latente: Se define por $\boldsymbol{\psi} = \{\beta_0, \boldsymbol{\beta}\} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\beta)$.

Cabe resaltar, que como se indicó en la sección 3.2 se asigna una distribución a priori para el nivel de cuantil q . Esto se hace a partir de la distribución normal asignada a β_0 , y a partir de la ecuación (3.4) se deriva la distribución a priori para q .

- Hiperparámetros: Es definido por $\boldsymbol{\theta} = \gamma_1$, tal que γ_1 sigue una distribución PC prior $\pi(\gamma_1)$ como se explicó en la sección 3.3.5.

Al identificar el modelo gaussiano latente, podemos definir la distribución conjunta a posteriori de la siguiente manera:

$$\pi(\boldsymbol{\psi}, \gamma_1 | \mathbf{y}) \propto \pi(\gamma_1) \pi(\boldsymbol{\psi} | \gamma_1) \prod_{i=1}^n \pi(y_i | \boldsymbol{\psi}, \gamma_1).$$

Las marginales a posteriori de cada elemento de x_i del campo latente \mathbf{x} son definidas por:

$$\pi(\psi_j | \mathbf{y}) = \int \pi(\psi_j | \gamma_1, \mathbf{y}) \pi(\gamma_1 | \mathbf{y}) d\gamma_1. \quad (3.6)$$

La densidad marginal del hiperparámetro γ_1 es aproximada usando una aproximación gaussiana de la distribución condicional completa del campo latente, $\tilde{\pi}_G(\boldsymbol{\psi} | \gamma_1, \mathbf{y})$, evaluada en la moda $\boldsymbol{\psi}^*(\gamma_1) = \arg \max_{\boldsymbol{\psi}} \pi_G(\boldsymbol{\psi} | \gamma_1, \mathbf{y})$, es decir,

$$\tilde{\pi}(\gamma_1 | \mathbf{y}) \propto \left. \frac{\pi(\boldsymbol{\psi}, \gamma_1, \mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\psi} | \gamma_1, \mathbf{y})} \right|_{\boldsymbol{\psi}=\boldsymbol{\psi}^*(\gamma_1)},$$

esta aproximación es una aproximación de Laplace. A partir de esta aproximación se puede definir un grid γ_{1k} de los valores más probables de γ_1 .

La aproximación $\pi(\psi_i | \mathbf{y})$ puede ser obtenida utilizando una aproximación gaussiana, una de Laplace o una aproximación simplificada de Laplace. Después INLA aproxima la

ecuación 3.6 construyendo la siguiente aproximación anidada:

$$\tilde{\pi}(\psi_i | \mathbf{y}) = \int \tilde{\pi}(\psi_i | \gamma_1, \mathbf{y}) \tilde{\pi}(\gamma_1 | \mathbf{y}) d\gamma_1.$$

Finalmente, estas aproximaciones pueden ser integradas numéricamente respecto a γ_1

$$\tilde{\pi}(\psi_j | \mathbf{y}) = \sum_k \tilde{\pi}(x_j | \gamma_{1k}, \mathbf{y}) \tilde{\pi}(\gamma_{1k} | \mathbf{y}) \times \Delta_k,$$

donde Δ_k indica el peso del área correspondiente a γ_{1k} .



Capítulo 4

Estudio de Simulación

En este capítulo, se presenta un estudio de recuperación de parámetros a través del método de estimación INLA (utilizando la librería INLA en R). Para lo anterior, en la siguiente sección se explicará los diferentes parámetros modificados y objetivos de las pruebas.

4.1. Generación de datos

En esta sección se presentan diferentes configuraciones de parámetros que representan a distintos escenarios para las simulaciones. Para simular datos del modelo skew-probit definido en el capítulo 3, se definen valores para los parámetros β_1 , γ_1 (asociado a α) y para el nivel de cuantil q asociado al intercepto β_0 . Además se simulan datos de la covariable $x_i \sim N(0, 0.5)$ para $i = 1, \dots, n$. A partir de estas definiciones se simulan datos del predictor lineal $\eta_i = \beta_0(\alpha, q) + \beta_1 x_i$, donde $\beta_0(\alpha, q) = F^{-1}(q|\alpha)$ es el q -ésimo cuantil de $F(\cdot)$, la función de distribución acumulada (fda) de la distribución Skew-Normal estandarizada con media cero y varianza uno. Luego, como la probabilidad de éxito se define como $\pi_i = F(\eta_i | \alpha)$, se simulan dichas probabilidades y a partir de ellas se simulan datos de la variable respuesta Y_i , donde $Y_i \sim \text{Bin}(N_i, \pi_i)$, donde N_i es el i -ésimo número de ensayos.

4.2. Estudio de simulación 1: número de ensayos N_i grande

Para las simulaciones se considera $n = 200$, y se cambian los valores de los parámetros cuantil del intercepto q , parámetro de asimetría γ_1 , y coeficiente de regresión de la variable predictora β_1 , según los escenarios presentados en el cuadro 4.1. Se muestra el detalle de los escenarios simulados. Se tiene el escenario 1.0 como el escenario principal o de anclaje para las comparaciones. Para los escenarios 2.1, 2.2 y 2.3 se modifica el intercepto; para los escenarios 3.1, 3.2 y 3.3 se modifica el parámetro de asimetría; y para los escenarios 4.1, 4.2 y 4.3 se

modifica el coeficiente de regresión de la variable predictora. Bajo los distintos escenarios se realizaron simulaciones cuando se tiene grandes cantidades de ensayos $N_i = 200$.

Cuadro 4.1: Tabla de escenarios para N_i grande.

Escenario	n	N_i	β_1	γ_1	q
1.0	200	200	1	0.50	0.75
2.1	200	200	1	0.50	0.85
2.2	200	200	1	0.50	0.50
2.3	200	200	1	0.50	0.25
3.1	200	200	1	0.90	0.75
3.2	200	200	1	0.75	0.75
3.3	200	200	1	0.25	0.75
4.1	200	200	-1	0.50	0.75
4.2	200	200	-10	0.50	0.75
4.3	200	200	-50	0.50	0.75

Para la estimación se tiene en cuenta la distribución a priori del parámetro de sesgo γ_1 como una a priori compleja penalizada (PC) asumiendo $\theta = 10$, en base a lo explicado en el capítulo 3. Además de ello, se asume una a priori no informativa normal para β_1 . Además si se asume una a priori normal para β_0 cuando $\alpha = 0$ ($\gamma_1 = 0$), la distribución a priori para q es derivada de esta a priori.

Esta configuración de los parámetros será comparada con una a distribución a priori gaussiana con media 0 y varianza 100 (precisión de 0.01) para el coeficiente de asimetría γ_1 .

4.2.1. Resultados

En el cuadro 4.2 se observa los resultados de las medias a posteriori e intervalos de credibilidad al 95% ($Q_{2.5}, Q_{97.5}$) de los parámetros de interés bajo los escenarios con un número de ensayos grande.

Cuadro 4.2: Tabla de resultado de simulaciones para N_i grande - a priori (PC).

Esc.	β_1			γ_1			q		
	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$
1.0	0.962	0.901	1.036	0.336	0.030	0.613	0.749	0.744	0.754
2.1	0.982	0.881	1.107	0.644	0.235	0.910	0.851	0.847	0.855
2.2	1.001	0.974	1.028	0.660	0.523	0.772	0.508	0.502	0.514
2.3	0.972	0.905	1.055	0.487	0.257	0.657	0.252	0.246	0.258
3.1	0.927	0.847	1.006	0.700	0.355	0.923	0.753	0.749	0.758
3.2	0.960	0.875	1.056	0.603	0.229	0.885	0.746	0.742	0.751
3.3	0.972	0.913	1.038	0.192	-0.088	0.472	0.748	0.743	0.754
4.1	-0.975	-1.066	-0.897	0.593	0.254	0.859	0.751	0.747	0.756
4.2	-9.712	-9.965	-9.460	0.550	0.427	0.661	0.748	0.738	0.757
4.3	-28.758	-29.759	-27.761	0.937	0.831	0.983	0.663	0.647	0.679

Conjuntamente con los cuadros 4.1 y 4.2 podemos observar que para el escenario 1.0 la media de las distribuciones a posteriori de los parámetros de interés están muy cercanos al verdadero valor de los parámetros o al menos dentro de los intervalos de credibilidad; es decir, cerca a 1 para β_1 , cerca a 0.5 para el parámetro de asimetría γ_1 , y cerca a 0.75 para el cuantil q del intercepto β_0 . De igual manera, para los escenarios del grupo 2, donde el cuantil de intercepto es modificado, las medias a posteriori se acercan bastante al valor verdadero. Del cuadro 4.2 en general se tiene como conclusión que para número de ensayo grande se obtienen buenos resultados.

Se presenta además el cuadro 4.3 en el cual podemos observar los resultados de los escenarios presentados en el cuadro 4.1 pero teniendo en cuenta una a priori gaussiana para el coeficiente de asimetría. De lo anterior se puede observar buenos resultados al tener la media de los parámetros muy cercanos al objetivo. Finalmente comparando estos resultados con los obtenidos en el cuadro 4.2, podemos concluir que según estos escenarios la a priori compleja penalizada tiene tan buenos resultados como una a priori gaussiana.

Cuadro 4.3: Tabla de resultado de simulaciones para N_i grande - a priori gaussiana.

Esc.	β_1			γ_1			q		
	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$
1	0.983	0.914	1.093	0.565	0.154	0.912	0.748	0.743	0.754
2.1	1.019	0.826	1.178	0.536	-0.100	0.944	0.851	0.848	0.856
2.2	0.999	0.972	1.026	0.684	0.554	0.790	0.509	0.503	0.515
2.3	0.955	0.893	1.032	0.535	0.332	0.684	0.253	0.247	0.259
3.1	0.976	0.926	1.024	0.979	0.921	0.988	0.754	0.749	0.758
3.2	1.028	0.971	1.081	0.978	0.905	0.988	0.747	0.742	0.751
3.3	0.997	0.926	1.078	0.306	-0.041	0.597	0.750	0.744	0.755
4.1	-1.045	-1.093	-0.996	0.979	0.936	0.988	0.752	0.747	0.756
4.2	-9.715	-9.968	-9.464	0.569	0.448	0.678	0.748	0.739	0.757
4.3	-28.525	-29.527	-27.542	0.983	0.956	0.988	0.669	0.655	0.686

4.3. Estudio de simulación 2: número de ensayos N_i pequeño

Se presenta a continuación, en el cuadro 4.4, un resumen de los distintos escenarios que se simularon para ver los resultados cuando se tiene $N_i = 50$, es decir pequeñas cantidades de ensayos, las consideraciones para la estimación son iguales que en la sección anterior.

4.3.1. Resultados

Para el grupo de simulaciones con escenarios con número de ensayos pequeño podemos observar los resultados de las estimaciones a posteriori en el cuadro 4.5.

Cuadro 4.4: Tabla de escenarios para N_i pequeño

Escenario	n	N_i	β_1	γ_1	q
5.0	200	50	1	0.50	0.75
6.1	200	50	1	0.50	0.85
6.2	200	50	1	0.50	0.50
6.3	200	50	1	0.50	0.25
7.1	200	50	1	0.90	0.75
7.2	200	50	1	0.75	0.75
7.3	200	50	1	0.25	0.75
8.1	200	50	-1	0.50	0.75
8.2	200	50	-10	0.5	0.75
8.3	200	50	-50	0.5	0.75

Cuadro 4.5: Tabla de resultado de simulaciones para N_i pequeño - a priori (PC)

Esc.	β_1			γ_1			q		
	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$
5	0.854	0.753	0.954	0.017	-0.308	0.379	0.743	0.734	0.753
6.1	0.853	0.715	1.013	0.056	-0.223	0.368	0.845	0.835	0.852
6.2	0.995	0.937	1.052	0.146	-0.150	0.531	0.487	0.476	0.497
6.3	1.042	0.902	1.168	0.359	-0.027	0.686	0.250	0.240	0.260
7.1	0.784	0.689	0.893	0.030	-0.397	0.508	0.742	0.732	0.751
7.2	0.829	0.721	0.951	0.074	-0.248	0.428	0.745	0.736	0.755
7.3	0.902	0.798	0.989	-0.022	-0.261	0.200	0.745	0.736	0.754
8.1	-0.912	-1.032	-0.817	0.143	-0.281	0.599	0.743	0.734	0.754
8.2	-9.113	-9.577	-8.653	0.389	0.144	0.610	0.729	0.709	0.749
8.3	-19.153	-20.267	-19.153	-0.707	-0.924	-0.386	0.610	0.573	0.642

En general se observa que no se tienen buenos resultados para el coeficiente de asimetría γ_1 ya que en su mayoría las medias a posteriori están muy lejanas del verdadero valor del parámetro y en algunos casos inclusive no se encuentran dentro de los intervalos de credibilidad. Usando la a priori PC, la media a posteriori del coeficiente de asimetría γ_1 tiene a cero en muchos casos. Los mejores resultados se obtuvieron para el escenario 6.3 y escenario 8.2. Se concluye que para un número de ensayos pequeño, al no tener suficiente información para explicar la variable respuesta, el modelo no consigue estimar bien el coeficiente de asimetría.

Además de ello se presenta el cuadro 4.6 las estimaciones a posteriori bajo los mismos escenarios pero con la a priori gaussiana para el coeficiente de asimetría, y de igual manera no se obtienen buenos resultados para el coeficiente de asimetría, excepto para algunos casos puntuales como son el escenario 6.3, escenario 6.2, escenario 8.1 y escenario 8.2.

Cuadro 4.6: Tabla de resultado de simulaciones para N_i pequeño - a priori gaussiana.

Esc.	β_1			γ_1			q		
	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$
5	0.875	0.708	1.067	0.222	-0.536	0.900	0.746	0.737	0.758
6.1	1.053	0.695	1.184	0.984	0.964	0.988	0.852	0.845	0.860
6.2	0.982	0.923	1.040	0.265	-0.225	0.641	0.491	0.478	0.504
6.3	0.971	0.861	1.115	0.507	0.116	0.759	0.253	0.241	0.264
7.1	0.851	0.660	0.998	0.276	-0.542	0.870	0.744	0.733	0.755
7.2	0.930	0.749	1.074	0.799	0.073	0.988	0.752	0.744	0.756
7.3	0.879	0.730	1.030	-0.093	-0.639	0.534	0.744	0.734	0.755
8.1	-0.996	-1.136	-0.834	0.487	-0.187	0.923	0.746	0.736	0.757
8.2	-9.130	-9.594	-8.671	0.462	0.216	0.674	0.732	0.712	0.751
8.3	-19.696	-20.783	-18.611	-0.365	-0.503	-0.207	0.584	0.548	0.627

4.4. Estudio de simulación 3: Efecto de intercepto cuantil

Se presenta a continuación, en el cuadro 4.7, un resumen de los distintos escenarios que se simularon para comparar los resultados cuando se calcula la distribución a posteriori del intercepto de forma clásica con respecto al cálculo del intercepto mediante el cuantil.

Cuadro 4.7: Tabla de simulaciones con cálculo clásico de intercepto

Escenario	n	N_i	β_1	γ_1	q
2.1	200	200	1	0.50	0.85
2.2	200	200	1	0.50	0.50
2.3	200	200	1	0.50	0.25

4.4.1. Resultados

Las estimaciones a posteriori del nivel de cuantil a partir de la estimación clásica del intercepto β_0 se presentan en el cuadro 4.8.

Cuadro 4.8: Tabla de resultados de simulaciones con estimación clásica del intercepto.

Esc.	β_1			γ_1			q		
	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$
2.1	0.962	0.901	1.036	0.336	0.030	0.613	0.749	0.744	0.754
2.2	0.982	0.881	1.107	0.644	0.235	0.910	0.851	0.847	0.855
2.3	1.001	0.974	1.028	0.660	0.523	0.772	0.508	0.502	0.514

En base a los cuadros 4.1 y 4.8 podemos observar que de la forma clásica no se obtienen mejores resultados respecto al cálculo de la distribución a posteriori del intercepto mediante el cuantil. Se observa que en la mayoría de los casos, de la forma clásica, la media a posteriori del coeficiente de simetría γ_1 es distinto del valor verdadero. Es por ello que se concluye que la metodología de estimar el intercepto mediante cuantil es la más adecuada.

4.5. Estudio de simulación 4: Modelo Probit

Se presenta a continuación, en el cuadro 4.9, un resumen de los distintos escenarios que se simularon para analizar los resultados cuando se tiene el parámetro de sesgo $\alpha = 0$ ($\gamma_1 = 0$) bajo diferentes tamaños de muestra n . Se realiza la prueba anterior con el objetivo de ver el comportamiento a posteriori de los parámetros al tratar de acercarse a un modelo probit.

Cuadro 4.9: Tabla de simulaciones con parámetro de sesgo $\alpha = 0$.

Escenario	n	N_i	β_1	$\gamma_1 = 0$	q
9.1	200	200	1	0.00	0.75
9.2	1000	200	1	0.00	0.75
9.3	10000	200	1	0.00	0.75

4.5.1. Resultados

Las estimaciones a posteriori de los parámetros se presentan en el cuadro 4.10. Podemos observar que en general estimaciones a posteriori de los parámetros tienen buenos resultados, al tener valores muy cercanos a los verdaderos. Además de ello, mientras mayor sea el tamaño de muestra n , mejoran los resultados. Se puede observar que el coeficiente de asimetría es cercano a cero y que el intervalo de credibilidad contiene al cero.

Cuadro 4.10: Tabla de resultados de simulaciones con PC prior cuando $\alpha = 0$.

Esc.	β_1			γ_1			q		
	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$
9.1	1.012	0.955	1.073	0.145	-0.107	0.403	0.753	0.747	0.759
9.2	1.002	0.967	1.026	-0.030	-0.167	0.088	0.749	0.747	0.752
9.3	1.003	0.995	1.011	-0.005	-0.054	0.040	0.750	0.750	0.751

4.6. Estudio de simulación 5: Modelo Bernoulli skew probit

Se presenta a continuación, en el cuadro 4.11, un resumen de los distintos escenarios que se simularon para ver los resultados cuando se tiene el $N_i=1$, es decir se usa una variable respuesta con distribución Bernoulli(p_i). Se realiza la prueba anterior con el objetivo de ver el comportamiento de la estimación a posteriori de los parámetros cuando se aumenta el tamaño de muestra n .

4.6.1. Resultados

Las estimaciones a posteriori de los parámetros se presentan en el cuadro 4.12. Podemos observar que cuando el tamaño de muestra no es tan grande los resultados no son buenos para

Cuadro 4.11: Tabla de simulaciones - Bernoulli skew probit.

Escenario	n	N_i	β_1	γ_1	q
10.1	1000	1	1	0.50	0.75
10.2	10000	1	1	0.50	0.75
10.3	100000	1	1	0.50	0.75

el parámetro de asimetría γ_1 ya que la media a posteriori no esta cerca al valor verdadero. Sin embargo, aumentando el tamaño de muestra n a números muy elevados se obtienen mejores resultados.

Cuadro 4.12: Tabla de resultados - Bernoulli Skew Probit

Esc.	β_1			γ_1			q		
	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$	Media	$Q_{2.5}$	$Q_{97.5}$
10.1	1.013	0.803	1.227	-0.01	-0.250	0.222	0.752	0.723	0.780
10.2	0.856	0.753	0.950	-0.032	-0.407	0.361	0.744	0.734	0.753
10.3	0.996	0.962	1.035	0.437	0.266	0.609	0.747	0.744	0.750

Capítulo 5

Aplicación

En el presente capítulo se aplica el modelo skew probit descrito en el capítulo 3 para analizar la anemia infantil en niños de entre 6 y 36 meses de edad en el Perú, en el año 2022. Además de ello, se realiza comparaciones estadísticas con modelos competencia presentados en el capítulo 2. El objetivo principal de la aplicación es estimar la probabilidad de un infante tener anemia en base a un grupo de covariables apropiadas. Con lo anteriormente mencionado se busca poder identificar variables con mayor peso predictivo y así poder explicar las causas de la anemia infantil, con la finalidad de realizar un plan a nivel nacional para reducir la población anémica infantil del país.

5.1. Situación de la anemia en la población infantil en el Perú

En base a la información publicada por el Ministerio de Salud, según los datos proporcionados anualmente por la Encuesta Nacional de Demografía y Salud (ENDES), la anemia infantil en niños de entre 6 y 36 meses de edad se ha mantenido a un nivel muy similar desde el año 2011 como puede verse en la Figura 5.1.



Figura 5.1: Porcentaje de población con anemia infantil (línea naranja) y desnutrición crónica (línea azul) del año 2010 al año 2022 en el Perú. Fuente: Tavera Salazar et al. (2023).

También se ha notado una ligera reducción de la desnutrición crónica infantil en menores de cinco años de edad (Figura 5.1). La población infantil más afectada por la anemia reside especialmente en poblaciones más rurales del país como indica la Figura 5.2.

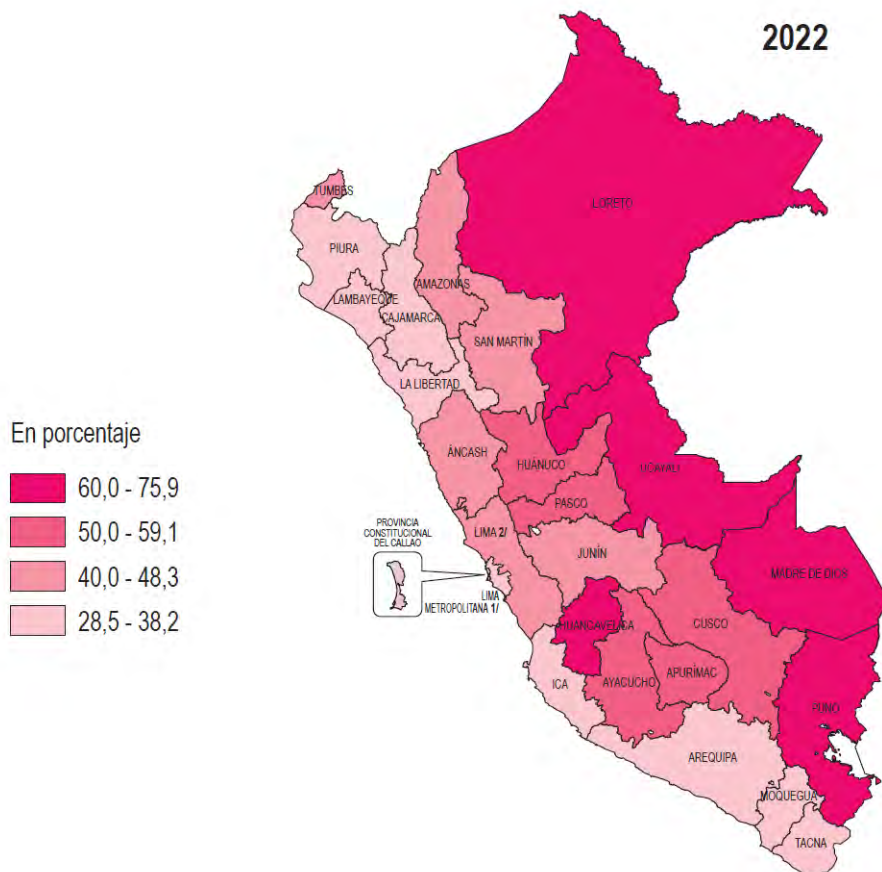


Figura 5.2: Prevalencia de anemia en niñas y niños de 6 a 36 meses según departamentos en el Perú en el año 2022. Fuente: INEI (2023).

Por ello, uno de los principales objetivos del plan de gobierno es contrarrestar el déficit de hierro en los niños, especialmente en poblaciones más rurales del país, donde la pobreza es extrema y la falta de educación nutricional es común en estas zonas. Desafortunadamente encontramos que poco efecto se está teniendo con el plan nutricional por parte del estado ya que los niveles de anemia infantil siguen en los mismos niveles. Considerando que la población infantil sí está consumiendo los suplementos de hierro que el gobierno está ofreciendo como solución ante esta problemática según se observa en la Figura 5.3. Ante esta problemática es necesario analizar otros factores para poder disminuir la tasa anémica infantil.



Figura 5.3: Porcentaje de población de niñas y niños de 6 a 35 meses de edad que consumieron suplemento de hierro en los último 7 días, del año 2009 al 2022. Fuente: Tavera Salazar et al. (2023).

5.2. Procesamiento y análisis exploratorio de los datos

Los datos a usar en esta tesis han sido recolectados por el Instituto Nacional de Estadística e Informática (INEI) a través de la Encuesta Nacional de Demografía y Salud (ENDES). Para analizar la anemia infantil se tiene información acerca de 27184 niños entre 6 y 36 meses de edad en el año 2022, donde se evaluó si cada niño es anémico (1) o no es anémico (0).

Para recolectar las covariables se utilizaron de igual manera otras bases de datos de la ENDES relacionadas a características del hogar, lactancia, características de anemia del niño, y salud en el mismo año. En conjunto se obtuvieron 72 posibles covariables entre numéricas y categóricas luego de un procesamiento de datos. Los cuadros 6.1 y 6.2 del apéndice presentan la definición de todas las posibles covariables consideradas en el análisis. De estas variables se decide eliminar todos los casos donde se encuentren datos perdidos, con ello se tiene $n = 26407$ observaciones. Para las variables numéricas se decide ver el comportamiento de estas covariables respecto a la presencia de anemia (1) o no (0) a través de diagramas de cajas (ver Figuras 5.4, 5.5 y 5.6), con el objetivo de evaluar si son covariables candidatas adecuadas para el modelamiento.

Según estos resultados se puede observar que las variables: Edad en meses (HW1), Peso en kg. (HW2), Altura en cm. (HW3), Nivel de hemoglobina del niño (HW53), Nivel de hemoglobina del niño globina ajustado por altitud (HW56), Meses de lactancia (M5), y nivel de hemoglobina de madre ajustado por altitud (HA56), son las que discriminan mejor entre niños anémicos y no anémicos, por lo cual pueden ser buenas candidatas para el modelo.

La Figura 5.7 muestra las correlaciones de las covariables numéricas obtenidas con el fin de evitar multicolinealidad en el modelo. Según este resultado, podemos observar una alta correlación (mayor a 0.8) entre las covariables Edad del niño en meses (HW1), Peso del niño en kg. (HW2), Altura del niño en cm. (HW3), es por ello que para el modelamiento solo

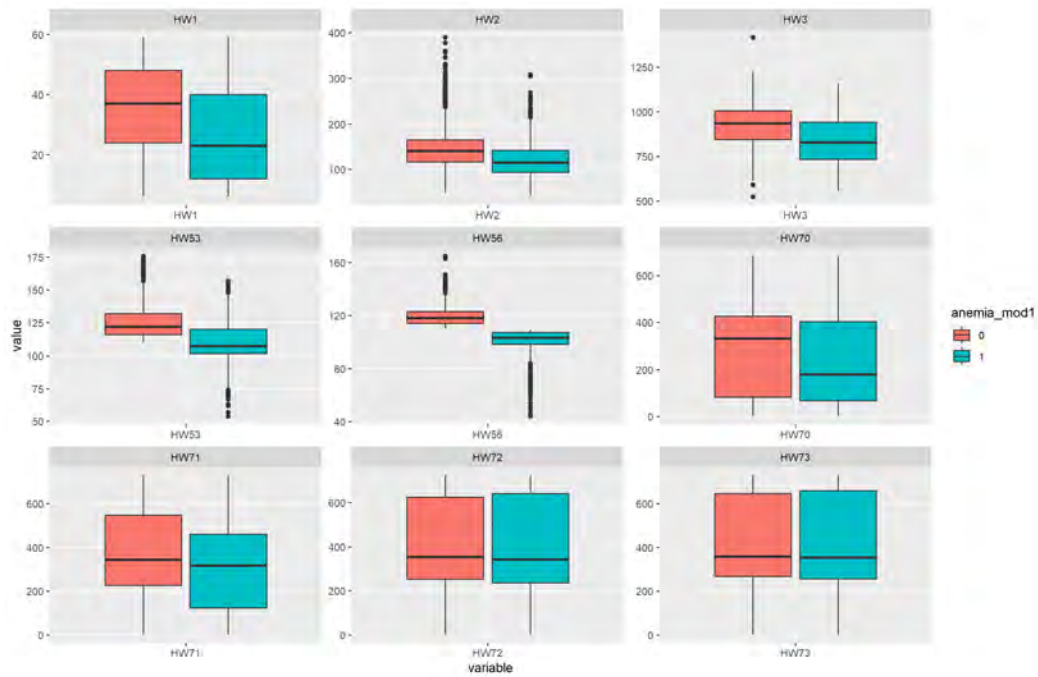


Figura 5.4: Diagrama de cajas de variables numéricas asociadas al niño: Edad en meses (HW1), Peso en kg. (HW2), Altura en cm. (HW3), Nivel de hemoglobina (HW53), Nivel de hemoglobina globina ajustado por altitud (HW56), Talla/Edad (HW70), Peso/Edad (HW71), Peso/Talla (HW72), desviación estándar del IMC (HW73), según anemia (1) o no anemia (0).

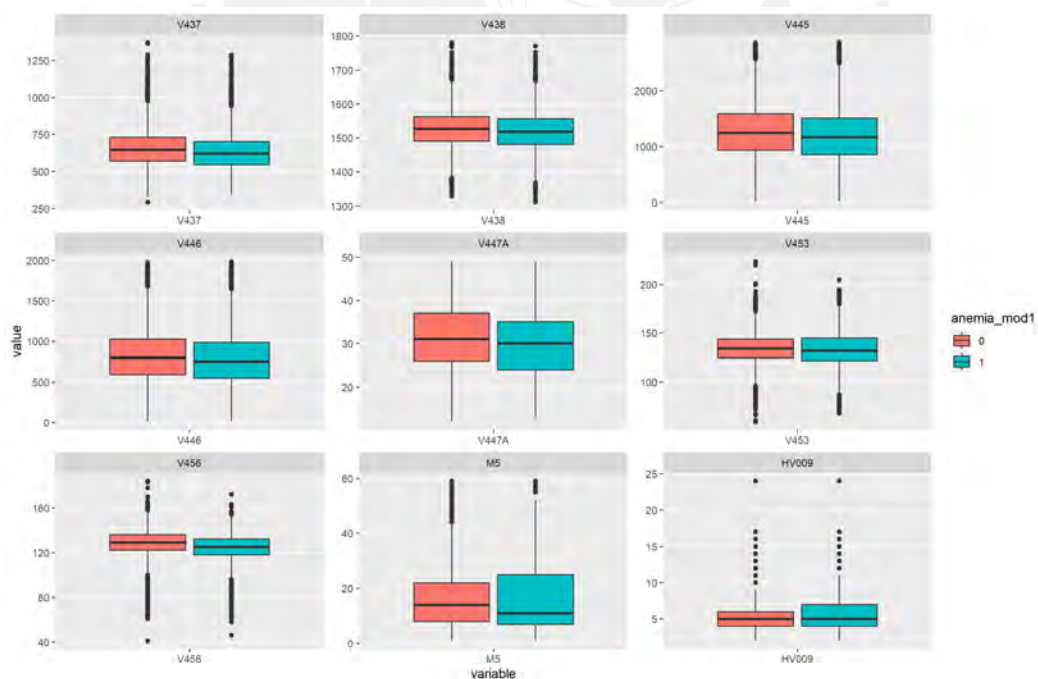


Figura 5.5: Diagrama de cajas de variables numéricas para el modelamiento: Peso del niño (V437), altura del niño (V438), IMC del niño (V445), Índice de Rohrer (V446), Edad de madre (V447A), nivel de hemoglobina de madre (V453), nivel de hemoglobina de madre ajustado por la altitud (V456), Meses de lactancia (M5), total de personas en el hogar (HV009), según anemia (1) o no anemia (0).

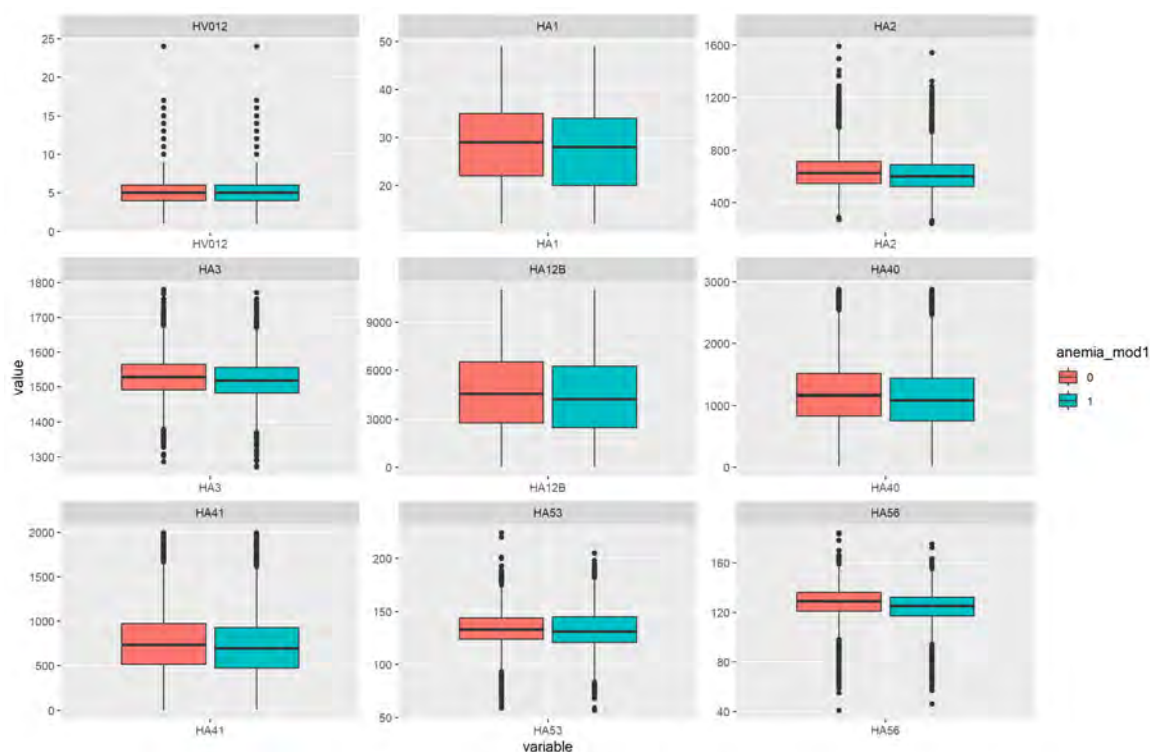


Figura 5.6: Diagrama de cajas de variables numéricas para el modelamiento: Número de miembros de residentes habituales (HV012), edad de madre en años (HA1), peso de madre en kg (HA2), talla de madre en cm (HA3), peso/talla de madre (HA12B), IMC de madre (HA40), Índice de rohrer de madre (HA41), nivel de hemoglobina de madre (HA53), nivel de hemoglobina de madre ajustado por altitud (HA56), según anemia (1) o no anemia (0).

consideramos una de ellas, en particular se selecciona la edad del niño en meses (HW1). Entre las covariables Nivel de hemoglobina del niño (HW53), Nivel de hemoglobina del niño ajustado por altitud (HW56) y nivel de hemoglobina de madre ajustado por altitud (HA56), también se debe escoger solo una covariable, se selecciona el Nivel de hemoglobina de la madre ajustado por altitud (HA56).

Según estos resultados, se escogieron como covariables numéricas: edad del niño en meses (HW1), meses de lactancia (M5), nivel de hemoglobina de la madre ajustado por altitud (HA56), y covariables categóricas: Actualmente amamantando (V404) con categorías si o no, Actualmente amenorreica (V405) con categorías si o no, esta patología puede producirse en madres con anorexia o bulimia, y que tienen un IMC inferior a 19, madre menor de 18 años (V452A) con categorías si o no, y Lugar de residencia (HV026) con categorías capital gran ciudad, pequeña ciudad, pueblo o campo.

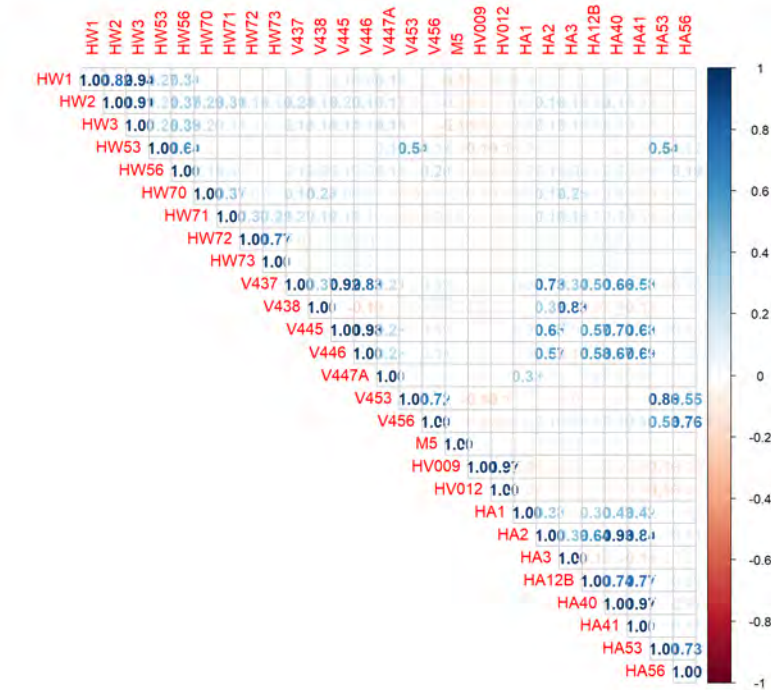


Figura 5.7: Correlaciones de posibles covariables numéricas.

5.3. Modelos propuestos

En esta sección se plantea ajustar un modelo skew probit para la predicción de niños anémicos en el Perú en base a algunas covariables descritas anteriormente, y además compararlo contra otros modelos competencia.

Para definir el modelo, se asume que Y_i es una v.a. que representa la presencia de anemia o no en el i -ésimo niño, luego $Y_i = 0$ si el niño i -ésimo no es anémico e $Y_i = 1$ si el niño es anémico, donde $i = 1, \dots, n = 26407$. El modelo a ajustar es definido por:

$$Y_i | \beta_0, \beta, \gamma_1 \stackrel{ind}{\sim} \text{Bin}(N_i = 1, \pi_i); \quad i = 1, \dots, n$$

$$g(\pi_i) = \eta_i$$

$$\eta_i = \beta_0 + X_i^T \beta,$$

donde π_i es la probabilidad del i -ésimo niño ser anémico, X_i es un vector de las covariables descritas previamente, β_0 es el efecto fijo llamado usualmente intercepto y β son los otros efectos fijos o coeficientes de regresión. Con el objetivo de poder retar al modelo skew probit se consideró ajustar otros modelos de regresión (logit y probit) considerando las mismas covariables. Luego la función de enlace $g()$ se define de acuerdo al modelo ajustado: i) Modelo skew probit: $g() = F^{-1}$, donde F es la fda de una distribución skew normal estandarizada.

ii) Modelo probit: $g() = \Phi^{-1}$, donde Φ es la fda de una distribución normal estándar. iii) Modelo logit: $g() = \text{logit}$ definida por a partir de la fda de la distribución logística.

La inferencia bayesiana para el modelo skew probit se realiza a través del INLA, y se detalló en el capítulo 3. Para los demás modelos se procede de forma similar.

5.4. Resultados

Para el modelamiento y validación de los modelos se plantea dividir los datos en grupos de entrenamiento y prueba, con una representación de 80 % y 20 %, respectivamente.

5.4.1. Ajuste del modelo en datos de entrenamiento

A continuación se presentan en el cuadro 5.1 los siguientes indicadores para evaluar la bondad de ajuste de los modelos: criterio de información ampliamente aplicable (WAIC), criterio de información de devianza (DIC), el logaritmo de la pseudo-verosimilitud marginal (LPML). El modelo skew probit tiene un menor valor para los indicadores WAIC y DIC, mientras para el indicador LPML tiene un valor mayor. Podemos concluir mediante estos indicadores que el modelo skew probit tiene mejor bondad de ajuste en comparación a los otros modelos.

Cuadro 5.1: Criterios de selección de los modelos ajustados.

Modelo	WAIC	DIC	LPML
Skew Probit	24953.34	24953.79	-12476.67
Probit	24968.26	24968.34	-12484.13
Logit	24958.34	24958.47	-12479.17

Las estimaciones a posteriori de los parámetros del mejor modelo ajustado, es decir del modelo skew probit se presentan en el cuadro 5.2, donde se puede observar que excepto la covariable meses de lactancia (M5), el resto de las covariables son significativas pues los intervalos de credibilidad al 95 % no contienen al cero. También se puede observar que el valor de la media a posteriori del coeficiente de asimetría γ_1 es igual a -0.568, y su intervalo de credibilidad no contiene al cero. Este resultado implica que este coeficiente aporta significativamente al modelo con una función de enlace asimétrica, y por lo tanto es una evidencia de que el modelo skew probit se ajusta mejor a los datos que un modelo probit.

La Figura 5.8 presenta las fdp marginales a posteriori de los coeficientes de regresión del modelo skew probit. Mientras que la Figura 5.9 presenta las fdp marginales a posteriori de los parámetros q y γ_1 del modelo skew probit.

Cuadro 5.2: Resúmenes a posteriori del Modelo Skew Probit. Media a posteriori, desviación estándar a posteriori, Intervalo de credibilidad del 95 % ($Q_{2.5}$, $Q_{97.5}$) y mediana (Q_{50}).

Covariable	Media	Desv. Estandar	$Q_{2.5}$	Q_{50}	$Q_{97.5}$
Cuantil del nivel intercepto (q)	1.000	0.003	0.999	1.000	1.000
HW1	-0.021	0.001	-0.023	-0.021	-0.020
M5	-0.001	0.001	-0.003	-0.001	0.000
V404Sí	0.155	0.024	0.108	0.155	0.202
V405Sí	0.174	0.025	0.125	0.174	0.223
V452A18 años o más	-0.431	0.064	-0.557	-0.430	-0.305
HV026Pequeña ciudad	0.198	0.035	0.130	0.197	0.266
HV026Pueblo	0.162	0.035	0.094	0.162	0.231
HV026Campo	0.446	0.034	0.379	0.445	0.512
HA56	-0.017	0.001	-0.019	-0.017	-0.016
Coef. de asimetría (γ_1)	-0.568	0.092	-0.746	-0.569	-0.389

5.4.2. Ajuste del modelo en datos de prueba

Se aplican los modelos propuestos en la sección anterior a la data de prueba para estudiar la capacidad predictiva del modelo, simulando la eficiencia de los modelos en una situación real. Para ello se calcularon como criterios de predicción: i) la raíz del error cuadrático medio de predicción (RMSP), ii) precisión que es la proporción de predicciones correctas, iii) sensibilidad que es la proporción de verdaderos positivos, anémicos correctamente identificados por el modelo; iv) especificidad, la proporción de falsos positivos, es decir de no anémicos que fueron correctamente identificados por el modelo, y v) el área bajo la curva de ROC (AUC), donde la curva de característica operativa del receptor (ROC) grafica la tasa de verdaderos positivos (TPR, también conocida como sensibilidad) frente a la tasa de falsos positivos (FPR, o $1 - \text{especificidad}$) según varias configuraciones de un umbral que define la variable respuesta según su probabilidad. Por ello el AUC es un área entre 0 y 1 que sirve para medir desempeño de un clasificador en todos los umbrales de decisión posibles. Mayor AUC mejor es la capacidad de predicción del modelo.

En el cuadro 5.3 se presenta los indicadores de los modelos a competidores. Se puede observar el modelo skew probit tiene menor RMSP, y mayor precisión, mayor sensibilidad y mayor AUC. En particular, se consigue predecir correctamente el 70 % (precisión) de casos de niños anémicos o no anémicos. El 40 % (sensibilidad) de los niños anémicos fueron correctamente identificados por el modelo skew probit mejorando la sensibilidad de los modelos probit y logit. Mientras que según la especificidad el 86 % de niños no anémicos fueron correctamente identificados por el modelo, ligeramente peor que los modelos probit y logit. Como el AUC es 0.634, se puede decir que la capacidad predictiva del modelo es aceptable. En base a todos los criterios, el modelo skew probit es el que tiene mejor capacidad predic-

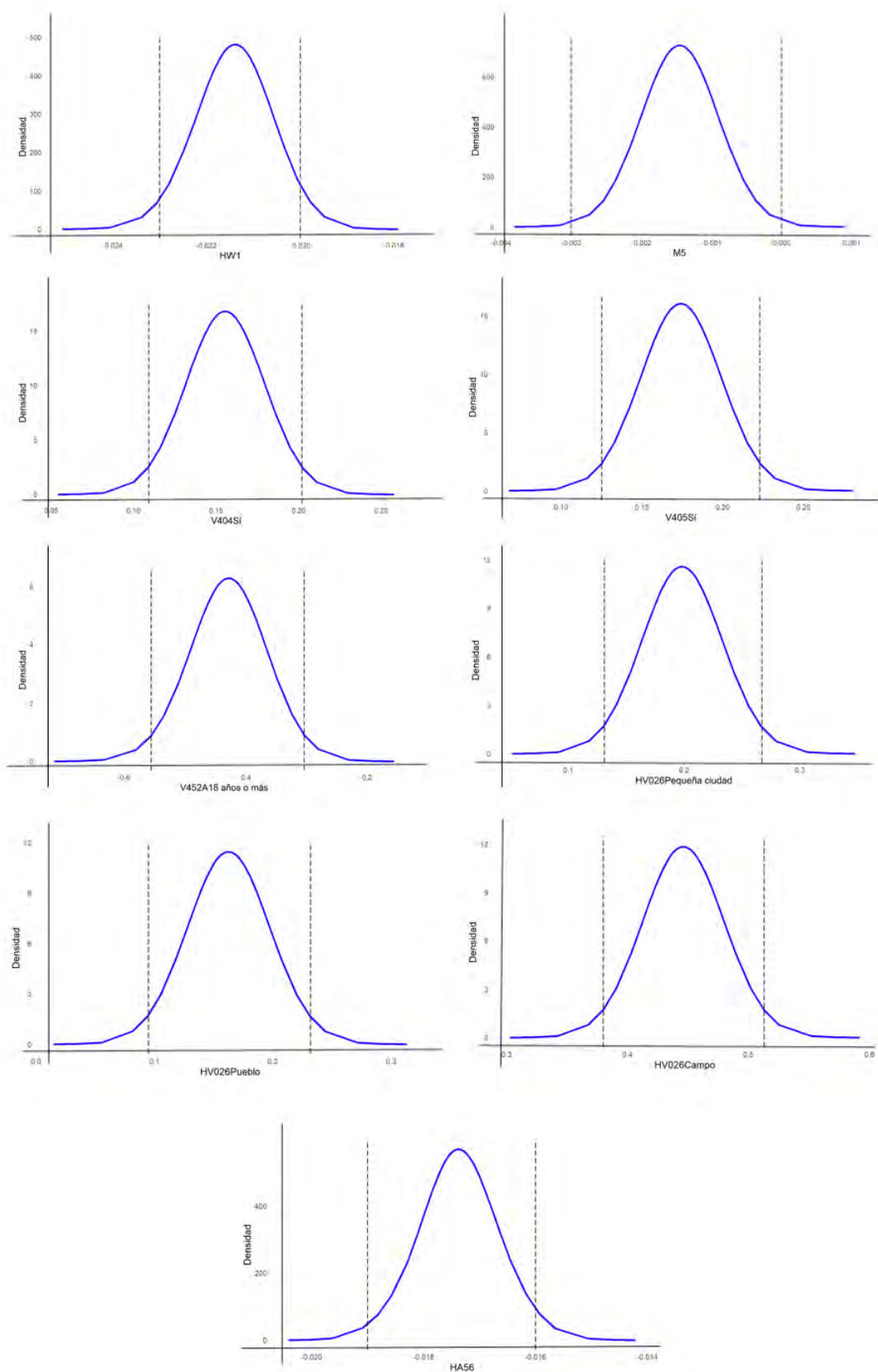


Figura 5.8: Función de densidad marginal a posteriori de los coeficientes de regresión según el modelo skew probit.

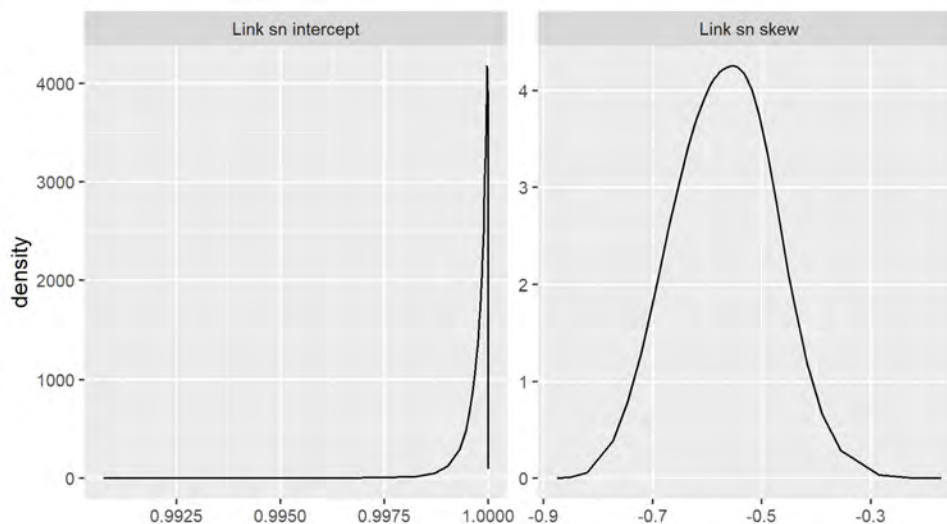


Figura 5.9: Función de densidad marginal a posteriori del parámetro q (izquierda) y γ_1 (derecha) según el modelo skew probit.

tiva. Como conclusión en base a estos resultados podemos ver un aceptable rendimiento del modelo propuesto skew probit frente a los demás modelos.

Cuadro 5.3: Evaluación de la capacidad predictiva de los modelos ajustados en los datos de prueba.

Modelo	AUC	RMSP	Precisión	Sensibilidad	Especificidad
Skew Probit	0.634	0.550	0.6973	0.4073	0.8606
Probit	0.622	0.555	0.6922	0.3720	0.8724
Logit	0.625	0.554	0.6933	0.3789	0.8704

Capítulo 6

Conclusiones

6.1. Conclusiones

- En base a las pruebas que se realizaron sobre la base de datos de anemia infantil con diferentes tipos de modelos para datos binarios, podemos observar un desempeño aceptable por parte del modelo skew-probit. Este último probablemente tiene un mejor comportamiento para aquellos datos que están en las colas respecto a la curva de distribución de probabilidad; es decir, para aquellas observaciones con mayor o menor probabilidad de anemia.
- El desempeño del modelo skew probit indica que las covariables utilizadas son muy buenos indicadores de anemia infantil. Esto último sería de gran utilidad al gobierno peruano para dar más enfoque a estas variables muy relacionadas a la alimentación del niño en los primeros días de nacido, cuando no solo debe ser a mayor edad como se está enfocando actualmente el estado.

6.2. Sugerencia para trabajos futuros

- Para estimaciones futuras se recomienda una mejor búsqueda de definición de las covariables ya que se puede mejorar los modelos actuales agregando covariables nuevas.
- Se puede extender el modelo a modelos longitudinales, para así tener una mejor de predicción.

Apéndice

Cuadro 6.1: Tabla 1 de definición de covariables

Variable	Tipo	Definición
ANEMIA MOD1	Categorico	Identificador de anemia infantil
CASEID	Numérico	Código ID único de cada observación
HW1	Numérico	Edad en meses
HW2	Numérico	Peso en kilogramos (1 dec.)
HW3	Numérico	Altura en centímetros(1 dec.)
HW53	Categorico	Nivel de hemoglobina
HW56	Numérico	Nivel de hemoglobina ajustado por altitud (g/dl - 1)
HW70	Numérico	Talla/Edad de la Desviación Estándar de la mediana de referencia
HW71	Numérico	Peso/Edad de la Desviación Estándar de la mediana de referencia
HW72	Numérico	Peso/Talla Desviación Estándar de la mediana de referencia
HW73	Numérico	Desviación Estándar del IMC
M5	Numérico	Meses de lactancia
M55A	Categorico	Los primeros 3 días, con leche que no sea materna
M55G	Categorico	Los primeros 3 días, se les administra fórmula infantil
V404	Categorico	Actualmente amamantando
V405	Categorico	Actualmente amenorreica
V406	Categorico	Actualmente absteniéndose
V415	Categorico	Bebí de biberón con tetina
V416	Categorico	He oído hablar de la rehidratación oral
V437	Numérico	Peso del encuestado (kilos-1d)
V438	Numérico	Altura del encuestado (cms-1d)
V445	Numérico	Índice de masa corporal para la MEF
V446	Numérico	Índice de Rohrer para la MEF
V447A	Numérico	Edad de mujer en años del reporte de hogar
V452A	Categorico	Menores de 18 años (informe HH)
V453	Numérico	Nivel de hemoglobina (g/dl - 1 decimal)
V454	Categorico	Actualmente embarazada (informe del hogar)
V456	Numérico	Nivel de hemoglobina ajustado por altitud (g/dl - 1 decimal)
V457	Categorico	Nivel de anemia
V463A	Categorico	Fuma cigarrillos
V463Z	Categorico	No fuma

Cuadro 6.2: Tabla 2 de definición de covariables

Variable	Tipo	Definición
V465	Categorico	Eliminación de las heces de su niña(o) cuando no usa el baño
V466	Categorico	Ud. puede decidir de llevarlo a tratamiento
V467C	Categorico	Conseguir el dinero para el tratamiento
V467D	Categorico	Lejanía de servicios de salud
V467E	Categorico	Obtener ayuda médica para uno mismo: conseguir transporte
V467H	Categorico	No haya personal de salud
V467I	Categorico	No haya medicamentos
V474	Categorico	Escuchó sobre la Tuberculosis o TBC
V481D	Categorico	Tipo de seguro de salud: privado
V481E	Categorico	Tipo de seguro de salud: ESSALUD/IPSS
V481F	Categorico	Tipo de seguro de salud: fuerzas armadas/policiales
V481G	Categorico	Tipo de seguro de salud: Seguro Integral de Salud
V481H	Categorico	Tipo de seguro de salud: entidad prestadora de salud
HV009	Numérico	Total de personas en el hogar
HV012	Numérico	Número de miembros de jure (residentes habituales)
HV023	Categorico	Dominio
HV025	Categorico	Área de residencia
HV026	Categorico	Lugar de residencia
HA1	Numérico	Edad de la mujer en años
HA2	Numérico	Peso en kilogramos (1 decimal)
HA3	Numérico	Talla en centímetros (1 decimal)
HA12B	Numérico	Peso/Talla Porcentaje respecto a la mediana (OMS)
HA35	Categorico	En las últimas 24 horas cuantos cigarrillos fumó Ud.
HA40	Numérico	Índice de Masa Corporal para la entrevistada
HA41	Numérico	Índice de Rohrer para la entrevistada
HA53	Numérico	Nivel de hemoglobina (g/dl - 1 decimal)
HA54	Categorico	Actualmente embarazada
HA55	Categorico	Resultado de la medición (hemoglobina)
HA56	Numérico	Nivel de hemoglobina ajustado por la altitud (g/dl - 1 decimal)
HA57	Categorico	Nivel de anemia
HA66	Categorico	El nivel educativo más alto aprobado de la mujer
HA67	Categorico	El año más alto de educación de la mujer
HA68	Categorico	Nivel educativo más alto
SH11A	Categorico	Seguro de salud: ESSALUD/IPSS
SH11B	Categorico	Seguro de salud: Militar
SH11C	Categorico	Seguro de salud: Integral
SH11D	Categorico	Seguro de salud: Compañía de seguros
SH11E	Categorico	Seguro de salud: Seguro privado
SH11Y	Categorico	Seguro de salud: No sabe
SH11Z	Categorico	Seguro de salud: No tienen
SH13	Categorico	Actividad de la ultima semana
SH15N	Categorico	Nivel de educación aprobado
SH15Y	Categorico	Nivel de educación en años

Bibliografía

- Arellano-Valle, R., Bolfarine, H. y Lachos, V. (2007). Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics* **34**(6): 663–682.
- Azevedo, C. L., Bolfarine, H. y Andrade, D. F. (2011). Bayesian inference for a skew-normal irt model under the centred parameterization, *Computational Statistics Data Analysis* **55**(1): 353–365.
- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics* **12**(2): 171–178.
- Bayes, C. L. y Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution, *Brazilian Journal of Probability and Statistics* **21**(2): 141–163.
- Bazán, J. L., Bolfarine, H. y D., B. M. (2010). A framework for skew-probit links in binary regression, *Communications in Statistics - Theory and Methods* **39**(4): 678–697.
URL: <https://doi.org/10.1080/03610920902783849>
- Brooks, S., G. A. J. G. M. X.-L. (2011). *Handbook of Markov Chain Monte Carlo*, Chapman and Hall/CRC.
- Cabras, S., Racugno, W., Castellanos, M. E. y Ventura, L. (2012). A matching prior for the shape parameter of the skew-normal distribution, *Scandinavian Journal of Statistics* **39**(2): 236–247.
- Chen, M.-H. y Qi-Man Shao, D. K. D. (1999). A new skewed link model for dichotomous quantal response data, *American Statistical Association* **94**(448): 1172–1186.
- Fahrmeir L., Kneib T., L. S. y M. B. (2013). *Regression : models methods and applications*, 1st edn, Springer, Berlin Heidelberg.
- Faraway, J. (2004). *Linear Models with R*, 2nd edn, Chapman and Hall/CRC, New York.

- Gamerman, D. y Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC.
- Gelfand, A. E. (2000). Gibbs sampling, *Journal of the American Statistical Association* **95**(452): 1300–1304.
URL: <http://www.jstor.org/stable/2669775>
- INEI (2023). Encuesta demográfica y de salud familiar 2022, *INEI* p. 377.
- Jaramillo, E. D. (2024). *Modelos Lineal, Logístico y Probit*. <https://rpubs.com/Edimer/540368>.
- Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, edition 2. edn, Elsevier Science Technology, San Diego.
- Lee, D. y Sinha, S. (2019). Identifiability and bias reduction in the skew-probit model for a binary response, *Journal of Statistical Computation and Simulation* **89**(9): 1621–1648.
- Liseo, B. (1990). The skew-normal class of densities: inferential aspects from a Bayesian viewpoint, *Biometrika* **50**: 59–70.
- McCullagh, P. (1989). *Generalized Linear Models*, Chapman and Hall/CRC.
- Miniero, R., Talarico, V., Galati, M. C., Giancotti, L., Saracco, P. y Raiola, G. (2018). Iron deficiency and iron deficiency anemia in children, in L. Rodrigo (ed.), *Iron Deficiency Anemia*, IntechOpen, Rijeka, chapter 2.
URL: <https://doi.org/10.5772/intechopen.79790>
- Montgomery, D. C. y Runger, G. C. (2003). *Probabilidad y estadística aplicadas a la ingeniería*, 2nd edn, Limusa Wiley, México.
- Muff, S., Riebler, A., Held, L., Rue, H. y Saner, P. (2015). Bayesian analysis of measurement error models using integrated nested laplace approximations, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **64**(2): 231–252.
URL: <http://www.jstor.org/stable/24771892>
- Rue, H., Martino, S. y Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society: Series B (Methodological)* **71**(2): 319–392.

Simpson, D., Rue, H., Riebler, A., Martins, T. G. y Sørbye, S. H. (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors, *Statistical Science* **32**(1): 1 – 28.

URL: <https://doi.org/10.1214/16-STS576>

Tavera Salazar, M., Lazo Gonzales, O. y Villar Montesinos, E. (2023). La problemática de la anemia infantil en el Perú: situación y retos, desde una nueva perspectiva, *Technical Report 1-35*, Colegio Médico del Perú.

Van Niekerk, J. y Rue, H. (2021). Skewed probit regression–identifiability, contraction and reformulation, *REVSTAT – Statistical Journal* **19**(1): 1–20.

Wicklin, R. (2015). *Error distributions and exponential regression models*. <https://blogs.sas.com/content/iml/2015/09/16/plot-distrib-exp.html>.

