



Universidad Nacional de Colombia

FACULTAD DE CIENCIAS

TERCER CASO DE ESTUDIO

Estadística Bayesiana

Docente:

Juan Camilo Sosa

Ander Steven Cristancho Sánchez (acristanchos@unal.edu.co)

Valeria del Pilar Lopez Bosiga (vlopezbo@unal.edu.co)

Noviembre 2023

1 Alcaldía de Bogotá 2023

En esta [publicación](#) de La Silla Vacía se “revisó el desempeño de cada encuestadora que hizo mediciones en las cinco ciudades principales frente a los resultados” de las elecciones regionales de Colombia de 2023. De acuerdo con este medio, Invamer fue una de las encuestadoras con *menor error total* en Bogotá. En esta [hoja de cálculo](#) se pueden consultar algunos detalles técnicos incluyendo los márgenes de error de las encuestas.

Así, Invamer S.A.S realizó una encuesta del 17 al 23 de octubre de 2023 para Noticias Caracol, Blu Radio y El Espectador, con el fin de medir la intención de voto en Bogotá, Medellín, Cali, Barranquilla, y Bucaramanga para las elecciones de alcaldes de 2023. En Bogotá se encuestaron 1200 hombres y mujeres de 18 años en adelante, de todos los niveles socio-económicos a nivel nacional, aptos para votar en las elecciones y que sean residentes de la ciudad. Se realizaron encuestas personales en el hogar de los encuestados a través de tablets y para las preguntas de intención de voto se utilizó tarjetón. La ficha técnica se puede descargar [aquí](#). Se obtuvieron los resultados que se presentan en la Tabla 1.

Intención de voto: Bogotá		
Candidato	Cantidad	Proporción
C. F. Galán	493	0.411
G. Bolivar	257	0.214
J. D. Oviedo	227	0.189
D. Molano	48	0.040
R. Lara	41	0.034
J. L. Vargas	38	0.032
J. E. Robledo	28	0.023
N. Ramos	11	0.009
R. A. Quintero	3	0.003
Voto en Blango	54	0.045
Total	1200	1.000

Tabla 1: Si las elecciones para la Alcaldía de Bogotá fueran mañana, y los candidatos fueran los que aparecen en este tarjetón, ¿por cuál de ellos votaría usted?

Aunque Invamer utilizó una clase particular de muestreo aleatorio sin reemplazo (ver Ficha Técnica), es posible considerar la muestra como una muestra aleatoria simple con reemplazo (IID), dado que el tamaño total de la muestra es muy pequeño en comparación con el tamaño del Universo. El principal interés científico y político en este problema se centra en estimar la

proporción poblacional de votos que recibirá cada candidato de acuerdo con los datos proporcionados por Invamer.

Bajo las condiciones dadas anteriormente, dado que nuestra incertidumbre acerca de las respuestas de las 1200 personas en la encuesta es intercambiable, una versión particular del Teorema de De Finetti ([Bernardo and Smith, 2000](#), pág. 176) garantiza que la única distribución muestral apropiada para datos de esta índole es la distribución multinomial.

Suponga que una población de interés tiene artículos de $k \geq 2$ tipos, y además, que la proporción de artículos de tipo j es $0 < \theta_j < 1$, para $j = 1, \dots, k$. Siendo $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, note que las componentes de $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ son tales que $\sum_{j=1}^k \theta_j = 1$. Ahora, suponga que se toma una muestra IID $\mathbf{y} = (y_1, \dots, y_n)$ de tamaño n de la población. Sea $\mathbf{n} = (n_1, \dots, n_k)$ el vector aleatorio que almacena los conteos asociados con cada tipo de artículo, así que n_j es el número de elementos en la muestra aleatoria de tipo j , para $j = 1, \dots, k$. En esta situación, se dice que \mathbf{n} sigue una distribución multinomial con parámetros n y $\boldsymbol{\theta}$, la cual se define como sigue: $\mathbf{n} \mid n, \boldsymbol{\theta} \sim \text{Multinomial}(n, \boldsymbol{\theta})$ si y solo si

$$p(\mathbf{n} \mid n, \boldsymbol{\theta}) = \frac{n!}{\prod_{j=1}^k n_j!} \prod_{j=1}^k \theta_j^{n_j} \quad (1)$$

siempre que $\sum_{j=1}^k n_j = n$ y $0 \leq n_j \leq n$ para todo $j = 1, \dots, k$.

Considere el modelo con distribución muestral $\mathbf{n} \mid n, \boldsymbol{\theta} \sim \text{Multinomial}(n, \boldsymbol{\theta})$ y distribución previa jerárquica dada por

$$\boldsymbol{\theta} \mid \alpha \sim \text{Dirichlet}(\alpha \mathbf{1}_k) \quad \text{y} \quad \alpha \sim \text{Gamma}(a, b),$$

donde $\mathbf{1}_k$ es el vector de unos de $k \times 1$ y a y b son los hiperparámetros del modelo (en el Cap. 8 de [Gelman et al. \(2013\)](#) se discuten modelos más sofisticados que consideran la estrategia de muestreo).

Preguntas

Ajustar el modelo propuesto usando un muestreador de Gibbs con $a = b = 1$ (incluir un anexo con todos los detalles). Reportar visual y tabularmente las estimaciones puntuales, los intervalos de credibilidad al 95% y los resultados oficiales de la Registraduría Nacional del Estado Civil para Galán, Bolívar y Oviedo, expresando todas las cifras en puntos porcentuales. Interpretar los resultados obtenidos (máximo 500 palabras).

Candidato	Estimación	Q2.5%	Q97.5%	Resultado oficial
C. F. Galán	40.89	38.19	43.66	49.02
G. Bolívar	21.35	19.09	23.68	18.71
J. D. Oviedo	18.88	16.70	21.13	20.10

Tabla 2: Resultados en puntos porcentuales

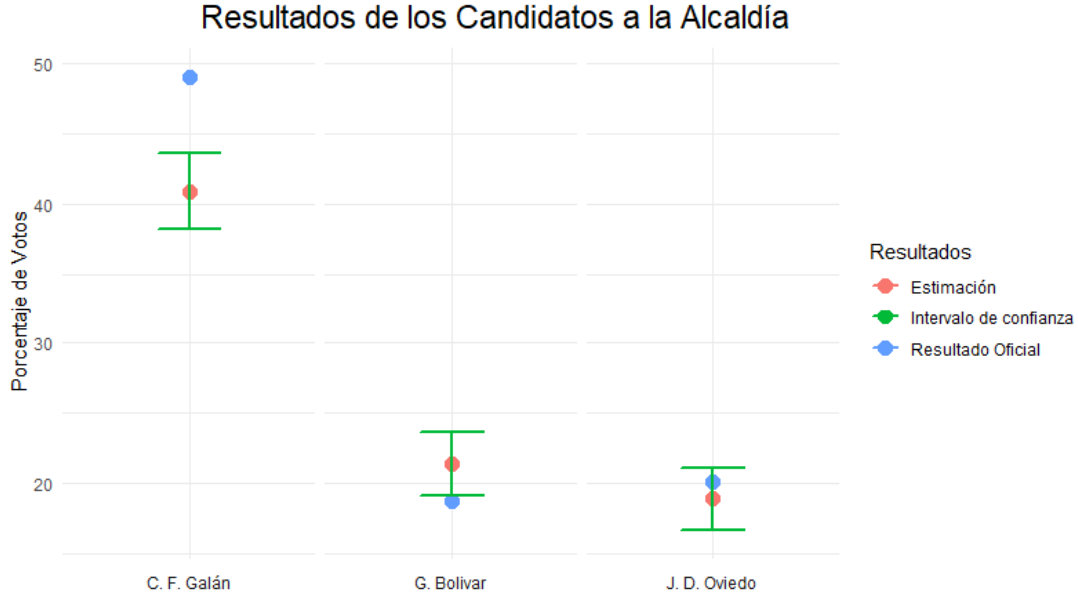


Figure 1: Resultados de los candidatos en puntos porcentuales

A partir de los resultados obtenidos en la tabla 2, se tiene que para el candidato C. Galán, la estimación se sitúa 8 puntos porcentuales por debajo del resultado oficial, el cual no está incluido en el intervalo de credibilidad. Esto sugiere que la información suministrada por la encuesta de Invamer no logró capturar de manera precisa el resultado real para este candidato. Por otro lado, la estimación para el candidato G. Bolívar está 2.6 puntos porcentuales por encima que el resultado oficial, y este último también cae fuera de los límites del intervalo de credibilidad, lo cual sugiere que, al igual que para el candidato anterior, la encuesta no brindó información precisa para capturar el resultado oficial para Gustavo Bolívar. Sin embargo, en el caso de J. D. Oviedo, la estimación está 1.22 puntos porcentuales por debajo del resultado oficial, pero se destaca que este último se encuentra dentro del intervalo de credibilidad. A pesar de la diferencia observada, la información de la encuesta de Invamer logró proporcionar una estimación relativamente buena para J. D. Oviedo.

Teniendo esto en cuenta, se puede decir que los resultados oficiales generaron sorpresa dado a que presentan diferencias importantes con respecto a lo que se había estimado previamente con la información de la encuesta. Sin embargo, es importante señalar que el candidato que se preveía como ganador efectivamente lo fue.

2 Selección de modelos

Puede ocurrir que en un análisis de regresión haya un gran número de variables independientes x , aunque puede que la mayoría de estas variables no tengan una relación sustancial con la variable dependiente y . En estas situaciones, incluir todas las variables regresoras en el modelo de regresión conduce a modelos saturados poco parsimoniosos difíciles de interpretar con un rendimiento deficiente. Por lo tanto, se recomienda considerar en el modelo final solo aquellas variables x para las que exista evidencia sustancial de una asociación con y . Esto no solo produce análisis de datos más simples, sino que también proporciona modelos con mejores propiedades estadísticas en términos de predicción y estimación.

Datos de diabetes

Considere la base de datos de diabetes dada en la Sección 9.3 de [Hoff \(2009, p. 161\)](#), que contiene datos asociados con 10 medidas basales x_1, \dots, x_{10} en un grupo de 442 pacientes diabéticos, así como una medida de progresión de la enfermedad y tomada un año después de las medidas basales. Los datos se pueden descargar de este [enlace](#) en `yX.diabetes.train` y `yX.diabetes.test`. A partir de estos datos, el objetivo es hacer un modelo predictivo para y basado en x_1, \dots, x_{10} (tanto y como las x_j se encuentran estandarizadas). Si bien un modelo de regresión con diez variables no sería abrumadoramente complejo, se sospecha que la relación entre y y las x_j puede no ser lineal, así que se recomienda considerar términos de segundo orden de la forma x_j^2 y $x_j x_k$ para potenciar la capacidad predictiva del modelo. Así, las variables regresoras incluyen diez efectos principales x_j , $\binom{10}{2} = 45$ interacciones $x_j x_k$ y nueve términos cuadráticos x_j^2 (no es necesario considerar x_2^2 en el modelo porque $x_2 = \text{sexo}$ es binaria, y por lo tanto $x_2 = x_2^2$). Esto da un total de $p = 64$ variables regresoras (no es necesario considerar el intercepto porque todas las variables se encuentran estandarizadas).

Modelamiento

Se considera un modelo de regresión de la forma $\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \mathbf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, donde \mathbf{y} es un vector de $n \times 1$ que contiene los valores de la variable respuesta, \mathbf{X} es una matriz de $n \times p$ que contiene los valores de las variables regresoras, $\boldsymbol{\beta}$ es un vector de $p \times 1$ que contiene los parámetros desconocidos, y finalmente, \mathbf{I}_n es la matriz identidad de $n \times n$.

Para evaluar los modelos de regresión, se dividieron aleatoriamente a los 442 individuos con diabetes en 342 individuos de entrenamiento y 100 individuos de prueba, lo que provee un conjunto de datos de entrenamiento $(\mathbf{y}_{\text{train}}, \mathbf{X}_{\text{train}})$ y un conjunto de datos de prueba $(\mathbf{y}_{\text{test}}, \mathbf{X}_{\text{test}})$. Así, se ajustan los modelos usando $(\mathbf{y}_{\text{train}}, \mathbf{X}_{\text{train}})$, y luego, usando los coeficientes de regresión estimados $\hat{\boldsymbol{\beta}} = \mathbf{E}(\boldsymbol{\beta} \mid \mathbf{y}_{\text{train}})$, se genera $\hat{\mathbf{y}}_{\text{test}} = \mathbf{X}_{\text{test}}\hat{\boldsymbol{\beta}}$. Luego, se evalúa el rendimiento predictivo del modelo comparando $\hat{\mathbf{y}}_{\text{test}}$ con \mathbf{y}_{test} por medio de una métrica apropiada.

Modelo 1: Regresión clásica previa unitaria

Distribución previa: [Previa unitaria](#) (*unit information prior*; Kass y Wasserman, 1995).

Modelo 2: Regresión clásica previa g

Distribución previa: [Previa \$g\$](#) (*g-prior*; Zellner, 1986).

Modelo 3: Regresión rígida

Distribución previa:

$$p(\boldsymbol{\beta}, \sigma^2, \lambda) = \mathbf{N}(\boldsymbol{\beta} \mid \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p) \cdot \text{Gl}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \text{G}(\lambda \mid a_\lambda, b_\lambda),$$

con $\nu_0 = 1$, $\sigma_0^2 = \hat{\sigma}_{\text{OLS}}^2$, $a_\lambda = 1$ y $b_\lambda = 2$.

Modelo 4: Regresión con errores correlacionados

Distribución muestral:

$$\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2, \rho \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{C}_\rho),$$

donde \mathbf{C}_ρ es una matriz con estructura autoregresiva de primer orden de la forma

$$\mathbf{C}_\rho = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}.$$

Distribución previa:

$$p(\boldsymbol{\beta}, \sigma^2, \rho) = \prod_{j=1}^p \mathcal{N}(\beta_j \mid 0, \tau_0^2) \cdot \text{Gl}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \text{U}(\rho \mid a_\rho, b_\rho)$$

con $\tau_0^2 = 50$, $\nu_0 = 1$, $\sigma_0^2 = \hat{\sigma}_{\text{OLS}}^2$, $a_\rho = 0$ y $b_\rho = 1$.

Preguntas

Ajustar cada modelo utilizando los datos de entrenamiento $(\mathbf{y}_{\text{train}}, \mathbf{X}_{\text{train}})$ (incluir un anexo con todos los detalles).

1. Para cada modelo, generar $\hat{\mathbf{y}}_{\text{test}} = \mathbf{X}_{\text{test}}\hat{\boldsymbol{\beta}}$ usando los coeficientes de regresión estimados $\hat{\boldsymbol{\beta}} = \mathbf{E}(\boldsymbol{\beta} \mid \mathbf{y}_{\text{train}})$. Graficar \hat{y}_{test} frente y_{test} y calcular el error absoluto medio $\frac{1}{n} \sum_i |y_{\text{test},i} - \hat{y}_{\text{test},i}|$ correspondiente.

Nota:

- Incluir todos los gráficos en una sola figura con 2×2 paneles (todos los paneles deben tener la misma escala en los ejes).
- En cada gráfico, superponer la recta $\hat{y}_{\text{test}} = y_{\text{test}}$ como punto de referencia (ver el segundo panel de la Figura 9.7 de [Hoff 2009](#), p. 170).
- En el encabezado de cada gráfica, incluir el error absoluto medio usando tres (3) cifras decimales.

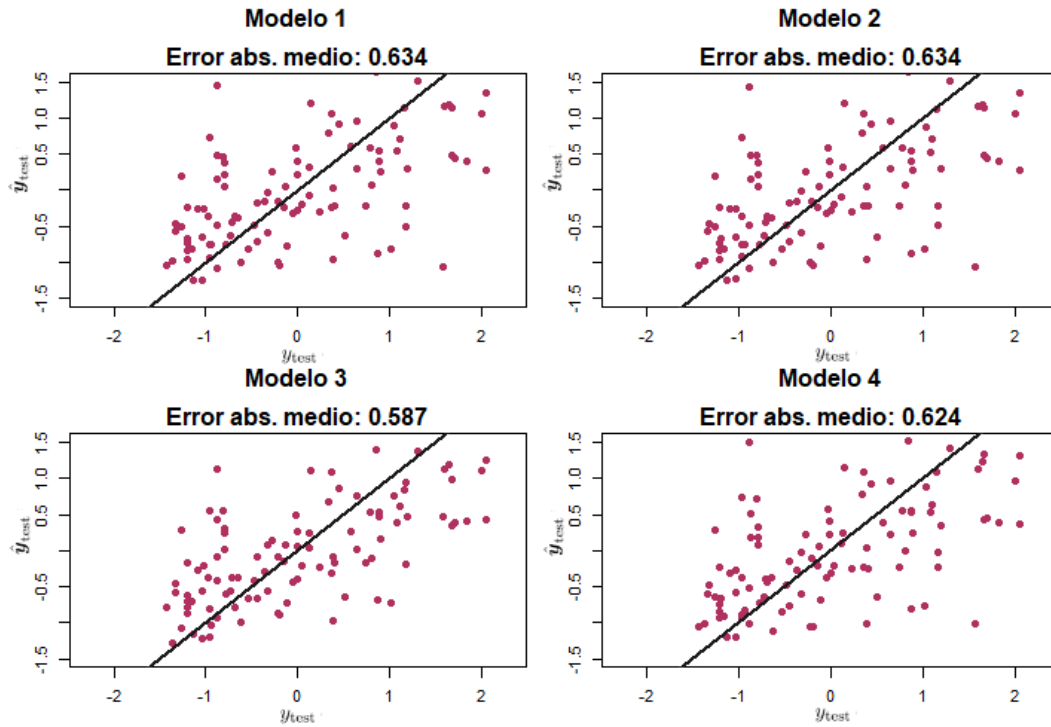


Figure 2: \hat{y}_{test} frente y_{test} con error absoluto medio

- Para cada modelo, chequear la bondad de ajuste usando la media como estadístico de prueba. Graficar la distribución predictiva posterior por medio de un histograma.

Nota:

- Incluir todos los gráficos en una sola figura con 2×2 paneles (todos los paneles deben tener la misma escala en los ejes).
- En cada gráfico, superponer el valor observado de la media como punto de referencia.
- En el encabezado de cada gráfica, incluir el valor p predictivo posterior correspondiente usando tres (3) cifras decimales.

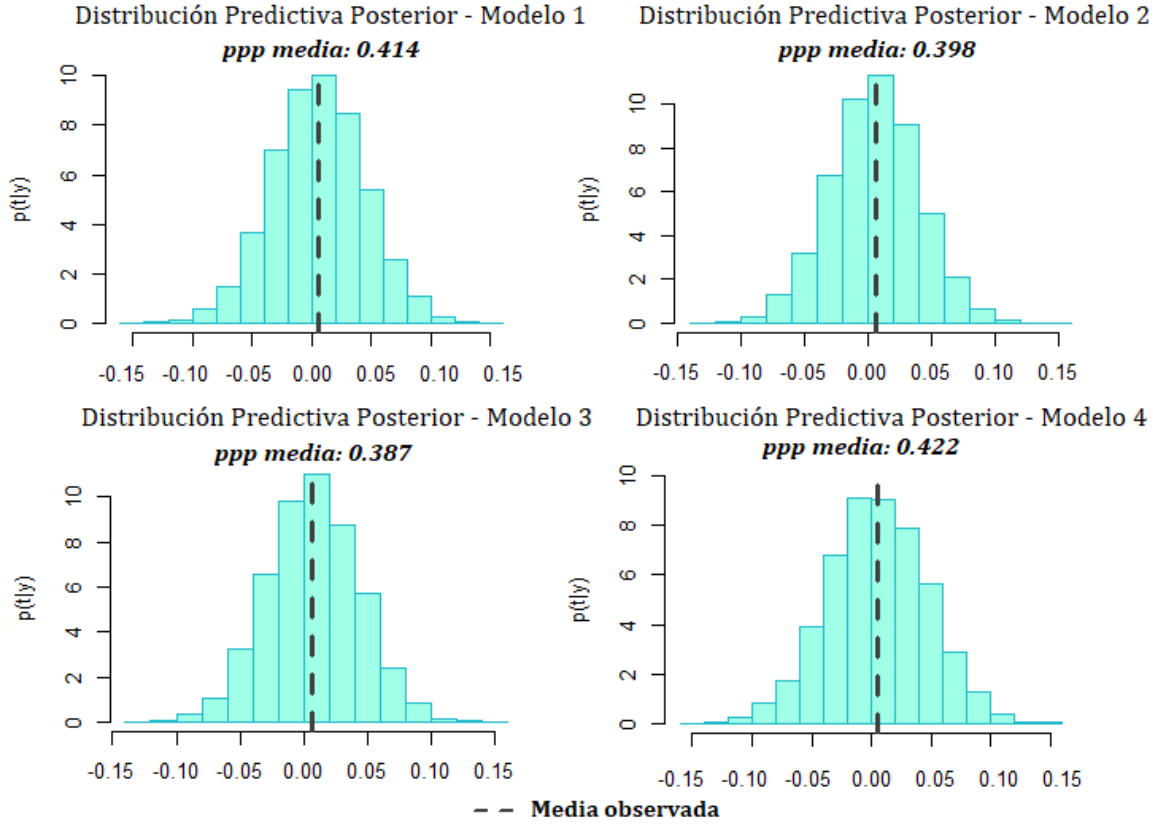


Figure 3: Histograma de la distribución predictiva posterior por modelo

- Para cada modelo, calcular el DIC. Presentar los resultados tabularmente usando tres (3) cifras decimales.

Modelo	P_{DIC}	DIC
1	64.644	784.988
2	64.679	777.804
3	48.942	761.536
4	63.215	782.025

Tabla 3: Comparación entre modelos por DIC

- Interpretar los resultados obtenidos en los numerales anteriores (máximo 500 palabras).

Considerando los resultados obtenidos en los numerales anteriores y tomando como referencia la figura 2, se observa que los errores absolutos medios de los modelos son relativamente similares; sin embargo, se destaca el modelo 3 al exhibir el error absoluto medio

más bajo, posicionándolo como el más preciso en la predicción de la medida de progresión de la diabetes, según esta métrica.

Al evaluar la bondad de ajuste mediante el uso de la media como estadístico de prueba, se obtienen valores de ppp cercanos a 0.4, indicando que los modelos son capaces de capturar adecuadamente la tendencia media en la medida de progresión de la diabetes de manera aceptable. En este contexto, resalta el modelo 4, que presenta el ppp más cercano a 0.5.

En relación con la precisión predictiva, evaluada a través del cálculo del DIC (ver tabla 3), se observa que el modelo 3, correspondiente a la regresión rígida, presenta el DIC más bajo. A pesar de que este modelo tiene un menor número efectivo de parámetros, demuestra una mejor precisión predictiva según este criterio.

En resumen, los resultados indican que el modelo 3, se presenta como la elección más acertada para la predicción de la medida de progresión de la diabetes.

Anexos

2.1 Alcaldía de Bogotá

Distribución muestral

$$\mathbf{n}|n, \theta \sim \text{Multinomial}(n, \boldsymbol{\theta})$$

Distribución previa

$$\boldsymbol{\theta}|\alpha \sim \text{Dirichlet}(\alpha \mathbf{1}_k)$$

$$\alpha \sim \text{Gamma}(a, b)$$

Distribución posterior

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{n}) &\propto p(\mathbf{n}|n, \boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha)p(\alpha) \\ &\propto \frac{n!}{\prod_{j=1}^k n_j!} \prod_{j=1}^k \theta_j^{n_j} \times \frac{\Gamma(\sum_{j=1}^k \alpha)}{\prod_{j=1}^k \Gamma(\alpha)} \prod_{j=1}^k \theta_j^{\alpha-1} \times \frac{b^a}{\Gamma(a)} \alpha^{a-1} e^{-b\alpha} \end{aligned}$$

Distribuciones condicionales completas

$$\boldsymbol{\theta}|\text{resto} \sim \text{Dirichlet}(\mathbf{n} + \alpha \mathbf{1}_k)$$

Como α no tiene una distribución condicional completa conocida, se tiene que:

$$p(\alpha|\text{resto}) \propto \frac{\Gamma(k\alpha)}{\Gamma(\alpha)^k} \prod_{j=1}^k \theta_j^{\alpha-1} \cdot \alpha^{a-1} e^{-b\alpha}$$

Para estimar los parámetros, se utilizó el siguiente muestreador de Gibbs:

1. Se inicializan los parámetros $\boldsymbol{\theta}^{(0)}, \alpha^{(0)}$.
2. Se genera un vector aleatorio $\boldsymbol{\theta}^{(b)} \sim \text{Dirichlet}(\mathbf{n} + \alpha^{(b-1)} \mathbf{1}_k)$
3. Para estimar $\alpha^{(b)}$ se hizo uso del siguiente algoritmo:
 - (a) Se inicializa $\gamma^{(0)}$ como $\log \alpha^{(0)}$.
 - (b) Se genera $\gamma^* \sim \mathbf{N}(\gamma^{(b-1)}, \delta^2)$, con $\delta^2 = 1$.
 - (c) Como $\gamma^* \in \mathbb{R}$, y $\alpha^* \in (0, \infty)$, tenemos que transformar la variable γ^* , de manera que esté en el soporte de la variable α^* . Esto lo hacemos con la transformación $\alpha^* = e^{\gamma^*}$, de manera que la función de distribución de α^* queda en términos de γ^* . Para esto, sólo cambiamos el término de las funciones de distribución por $\alpha^* = e^{\gamma}$ y multiplicamos el jacobiano, que es e^{γ} . Se calcula la tasa de aceptación en escala logarítmica para tener estabilidad numérica, quedando de la siguiente manera:

$$r = \exp\left(\log p(e^{\gamma^*}|\text{resto}) - \log p(e^{\gamma^{(b-1)}}|\text{resto}) + \gamma^* - \gamma^{(b-1)}\right)$$

- (d) Se establece el valor $\gamma^{(b)}$ de la siguiente manera:

$$\gamma^{(b)} = \begin{cases} \gamma^* & \text{con probabilidad } \min(1, r) \\ \gamma^{(b-1)} & \text{con probabilidad } 1 - \min(1, r) \end{cases}$$

- (e) Se establece el valor de $\alpha^{(b)} = e^{\gamma^{(b)}}$.
4. Se almacenan los valores de $\boldsymbol{\theta}^{(b)}$ y $\alpha^{(b)}$.
5. Se repiten 2, 3 y 4 hasta terminar las B iteraciones y tener convergencia.

Convergencia

A continuación se presenta la cadena de Log-Verosimilitud, los tamaños efectivos de muestra para cada parámetro y los coeficientes de variación de Monte Carlo para el modelo realizado:

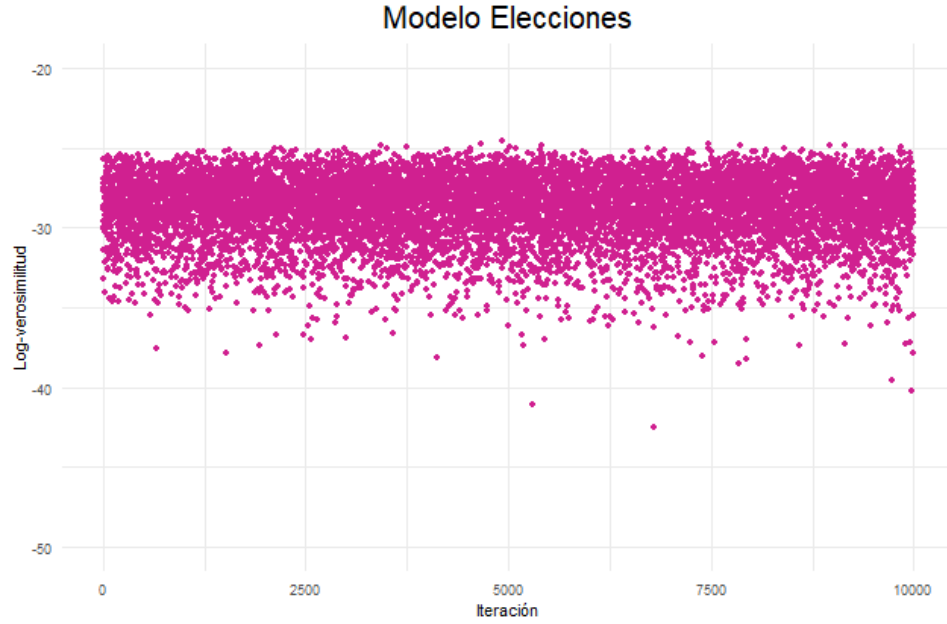


Figure 4: Cadena de Log-Verosimilitud

Parámetro	Tamaño	Parámetro	Tamaño
θ_1	10385	θ_6	10000
θ_2	10436	θ_7	10000
θ_3	9906	θ_8	10000
θ_4	10000	θ_9	10000
θ_5	10000	θ_{10}	10000
α	10000	/	/

Tabla 4: Tamaños efectivos de muestra

Parámetro	CV(%)	Parámetro	CV(%)
θ_1	0.03	θ_6	0.16
θ_2	0.05	θ_7	0.18
θ_3	0.06	θ_8	0.29
θ_4	0.14	θ_9	0.52
θ_5	0.15	θ_{10}	0.13
α	0.37	/	/

Tabla 5: Coeficientes de variación de Monte Carlo

En la figura 4 se observa la convergencia del modelo. Por otro lado, los tamaños efectivos de muestra son suficientes para garantizar la inferencia y los coeficientes de variación de Monte Carlo son muy bajos, lo que indica una mayor precisión y estabilidad en los resultados de la simulación.

2.2 Datos diabetes

Modelo 1: Regresión clásica previa unitaria

Distribución muestral

$$\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$

Distribución previa

$$p(\boldsymbol{\beta}, \sigma^2) = \mathcal{N}_p(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) \cdot \text{Gl}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2/2)$$

Donde $\boldsymbol{\beta}_0 = \hat{\boldsymbol{\beta}}_{\text{ols}}$, $\boldsymbol{\Sigma}_0 = n\hat{\sigma}_{\text{ols}}^2(\mathbf{X}^T\mathbf{X})^{-1}$, $\nu_0 = 1$, $\sigma_0^2 = \hat{\sigma}_{\text{ols}}^2$.

Distribucion posterior

$$\begin{aligned} p(\boldsymbol{\theta} \mid \mathbf{y}) &\propto p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times p(\boldsymbol{\beta}, \sigma^2) \\ &\propto \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \cdot \mathcal{N}_p(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) \cdot \text{Gl}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2/2) \end{aligned}$$

Distribuciones condicionales completas

$$\boldsymbol{\beta} \mid \text{resto} \sim \mathcal{N}_p \left(\left(\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \left(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\beta}_0 + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{y} \right), \left(\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \right)$$

$$\sigma^2 | \text{resto} \sim \text{Gl} \left(\frac{n + \nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}(\boldsymbol{\beta})}{2} \right)$$

Donde $\text{SSR}(\boldsymbol{\beta})$ es la suma de cuadrados de los residuales, la cual está definida como:

$$\text{SSR}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Para estimar los parámetros, se utilizó el siguiente muestreador de Gibbs:

1. Se inicializan los parámetros $\boldsymbol{\beta}^{(0)}, \sigma^{2(0)}$.
2. Se genera un vector aleatorio de la distribución condicional completa $\boldsymbol{\beta}^{(b)} \sim \text{N}_p \left(\boldsymbol{\beta} | \sigma^{2(b-1)} \right)$.
3. Se genera un número aleatorio de la distribución condicional completa $\sigma^{2(b)} \sim \text{Gl} \left(\sigma^2 | \boldsymbol{\beta}^{(b)} \right)$.
4. Se almacenan los valores de $\boldsymbol{\beta}^{(b)}$ y $\sigma^{2(b)}$.
5. Se repiten 2, 3 y 4 hasta terminar las B iteraciones y tener convergencia.

Modelo 2: Regresión clásica previa g

Distribución muestral

$$\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \text{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$

Distribución previa

$$p(\boldsymbol{\beta}, \sigma^2) = \text{N}_p(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) \cdot \text{Gl}(\sigma^2 | \nu_0/2, \nu_0 \sigma_0^2/2)$$

Donde $\boldsymbol{\beta}_0 = \mathbf{0}_p$, $\boldsymbol{\Sigma}_0 = k(\mathbf{X}^T \mathbf{X})^{-1}$, $\nu_0 = 1$, $\sigma_0^2 = \hat{\sigma}_{\text{ols}}^2$, $k = g\hat{\sigma}_{\text{ols}}^2$, $g = n$.

Distribucion posterior

$$\begin{aligned} p(\boldsymbol{\theta} | \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times p(\boldsymbol{\beta}, \sigma^2) \\ &\propto \text{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \cdot \text{N}_p(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) \cdot \text{Gl}(\sigma^2 | \nu_0/2, \nu_0 \sigma_0^2/2) \end{aligned}$$

Distribuciones condicionales completas

$$\boldsymbol{\beta} | \text{resto} \sim \text{N}_p \left(\frac{g}{g+1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \frac{g}{g+1} \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \right)$$

$$\sigma^2 | \mathbf{X}, \mathbf{y} \sim \text{Gl} \left(\frac{n + \nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}_g}{2} \right)$$

Donde SSR_g se define como:

$$\text{SSR}_g = \mathbf{y}^T \left(\mathbf{I}_n - \frac{g}{g+1} \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{y}$$

Para estimar los parámetros, se utilizó la siguiente simulación de Montecarlo, no fue necesaria una cadena de Markov.

1. Se genera un número aleatorio de la distribución condicional completa $\sigma^{2(b)} \sim \text{Gl} \left(\frac{n+\nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \text{SSR}_g}{2} \right)$
2. Se genera un vector aleatorio de la distribución condicional completa $\boldsymbol{\beta}^{(b)} \sim \text{N}_p \left(\boldsymbol{\beta} | \sigma^{2(b)} \right)$
3. Se almacenan los valores de $\boldsymbol{\beta}^{(b)}$ y $\sigma^{2(b)}$.
4. Se repiten 1, 2, y 3 hasta terminar las B iteraciones.

Modelo 3: Regresión Ridge

Distribución muestral

$$\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim \text{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$

Distribución previa

$$p(\boldsymbol{\beta}, \sigma^2, \lambda) = \text{N}_p(\boldsymbol{\beta} | \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p) \cdot \text{Gl}(\sigma^2 | \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \text{G}(\lambda | a_\lambda, b_\lambda)$$

Distribucion posterior

$$\begin{aligned} p(\boldsymbol{\theta} | \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2) \times p(\boldsymbol{\beta}, \sigma^2, \lambda) \\ &\propto \text{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \cdot \text{N}_p(\boldsymbol{\beta} | \mathbf{0}_p, \frac{\sigma^2}{\lambda} \mathbf{I}_p) \cdot \text{Gl}(\sigma^2 | \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \text{G}(\lambda | a_\lambda, b_\lambda) \end{aligned}$$

Distribuciones condicionales completas

$$\begin{aligned} \boldsymbol{\beta} | \text{resto} &\sim \text{N}_n \left(\left(\frac{\lambda \mathbf{I}_p}{\sigma^2} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y}, \left(\frac{\lambda \mathbf{I}_p}{\sigma^2} + \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} \right)^{-1} \right) \\ \sigma^2 | \text{resto} &\sim \text{Gl} \left(\frac{n + p + \nu_0}{2}, \frac{\nu_0 \sigma_0^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{\beta} + \text{SSR}(\boldsymbol{\beta})}{2} \right) \end{aligned}$$

$$\lambda | \text{resto} \sim \text{G} \left(a_\lambda + \frac{p}{2}, b_\lambda + \frac{\boldsymbol{\beta}^T \boldsymbol{\beta}}{2\sigma^2} \right)$$

Para estimar los parámetros, se utilizó el siguiente muestreador de Gibbs:

1. Se inicializan los parámetros $\boldsymbol{\beta}^{(0)}$, $\sigma^{2(0)}$ y $\lambda^{(0)}$.
2. Se genera un número aleatorio de la distribución condicional completa $\lambda^{(b)} \sim \text{G} \left(\lambda | \boldsymbol{\beta}^{(b-1)}, \sigma^{2(b-1)} \right)$
3. Se genera un vector aleatorio de la distribución condicional completa $\boldsymbol{\beta}^{(b)} \sim \text{N}_p \left(\boldsymbol{\beta} | \sigma^{2(b-1)}, \lambda^{(b)} \right)$.
4. Se genera un número aleatorio de la distribución condicional completa $\sigma^{2(b)} \sim \text{Gl} \left(\sigma^2 | \boldsymbol{\beta}^{(b)}, \lambda^{(b)} \right)$
5. Se almacenan los valores de $\boldsymbol{\beta}^{(b)}$, $\sigma^{2(b)}$ y $\lambda^{(b)}$.
6. Se repiten 2, 3, 4 y 5 hasta terminar las B iteraciones y tener convergencia.

Modelo 4: Regresión con errores correlacionados

Distribución muestral

$$\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2, \rho \sim \text{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{C}_\rho)$$

Distribución previa

$$p(\boldsymbol{\beta}, \sigma^2, \rho) = \prod_{j=1}^p \text{N}(\beta_j | 0, \tau_0^2) \cdot \text{Gl}(\sigma^2 | \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \text{U}(\rho | a_\rho, b_\rho)$$

Distribucion posterior

$$\begin{aligned} p(\boldsymbol{\theta} | \mathbf{y}) &\propto p(\mathbf{y} | \mathbf{X}, \boldsymbol{\beta}, \sigma^2, \rho) \times p(\boldsymbol{\beta}, \sigma^2, \rho) \\ &\propto \text{N}_n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{C}_\rho) \cdot \prod_{j=1}^p \text{N}(\beta_j | 0, \tau_0^2) \cdot \text{Gl}(\sigma^2 | \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \text{U}(\rho | a_\rho, b_\rho) \end{aligned}$$

Distribuciones condicionales completas

$$\begin{aligned} \boldsymbol{\beta} | \text{resto} &\sim \text{N}_p \left(\left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{X} + \frac{1}{\tau_0^2} \mathbf{I}_p \right)^{-1} \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{y}, \left(\frac{1}{\sigma^2} \mathbf{X}^T \mathbf{C}_\rho^{-1} \mathbf{X} + \frac{1}{\tau_0^2} \mathbf{I}_p \right)^{-1} \right) \\ \sigma^2 | \text{resto} &\sim \text{Gl} \left(\frac{n + \nu_0}{2}, \frac{\nu_0 \sigma_0^2 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{C}_\rho^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2} \right) \end{aligned}$$

Dado a que ρ no tiene una distribución condicional completa conocida, se tiene que:

$$p(\rho|\text{resto}) \propto |\mathbf{C}_\rho^{-1}| \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{C}_\rho^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

Para estimar los parámetros, se utilizó el siguiente muestreador de Gibbs:

1. Se inicializan los parámetros $\boldsymbol{\beta}^{(0)}$, $\sigma^{2(0)}$ y $\rho^{(0)}$.
2. Se genera un vector aleatorio de la distribución condicional completa $\boldsymbol{\beta}^{(b)} \sim \mathbf{N}_p\left(\boldsymbol{\beta}|\sigma^{2(b-1)}, \rho^{(b-1)}\right)$.
3. Se genera un número aleatorio de la distribución condicional completa $\sigma^{2(b)} \sim \text{Gl}\left(\sigma^2|\boldsymbol{\beta}^{(b)}, \rho^{(b-1)}\right)$.
4. Para estimar $\rho^{(b)}$ se hizo uso del siguiente algoritmo:
 - (a) Se inicializa $\rho^{(0)}$.
 - (b) Se genera $\rho^* \sim \text{U}(\rho^{(b-1)} - \delta, \rho^{(b-1)} + \delta)$, con $\delta = 0.3$.
 - (c) Como ρ^* puede ser menor que 0 o mayor que 1 dependiendo de $\rho^{(b-1)}$, calculamos $|\rho^*|$ y luego simetrizamos ρ^* , hallando $\min(\rho^*, 2 - \rho^*)$, de manera que $\rho^* \in (0, 1)$. Se calcula la tasa de aceptación en escala logarítmica para tener estabilidad numérica, quedando de la siguiente manera:

$$r = \exp(\log p(\rho^*|\text{resto}) - \log p(\rho^{(b-1)}|\text{resto}))$$

- (d) Se establece el valor de $\rho^{(b)}$ de la siguiente manera:

$$\rho^{(b)} = \begin{cases} \rho^* & \text{con probabilidad } \min(1, r) \\ \rho^{(b-1)} & \text{con probabilidad } 1 - \min(1, r) \end{cases}$$

5. Se almacenan los valores de $\boldsymbol{\beta}^{(b)}$, $\sigma^{2(b)}$ y $\rho^{(b)}$.
6. Se repiten 2, 3, 4 y 5 hasta terminar las B iteraciones y tener convergencia.

Y para estimar ρ se hizo uso del siguiente algoritmo:

Convergencia

Se presentan a continuación las cadenas de Log-verosimilitud para cada modelo:

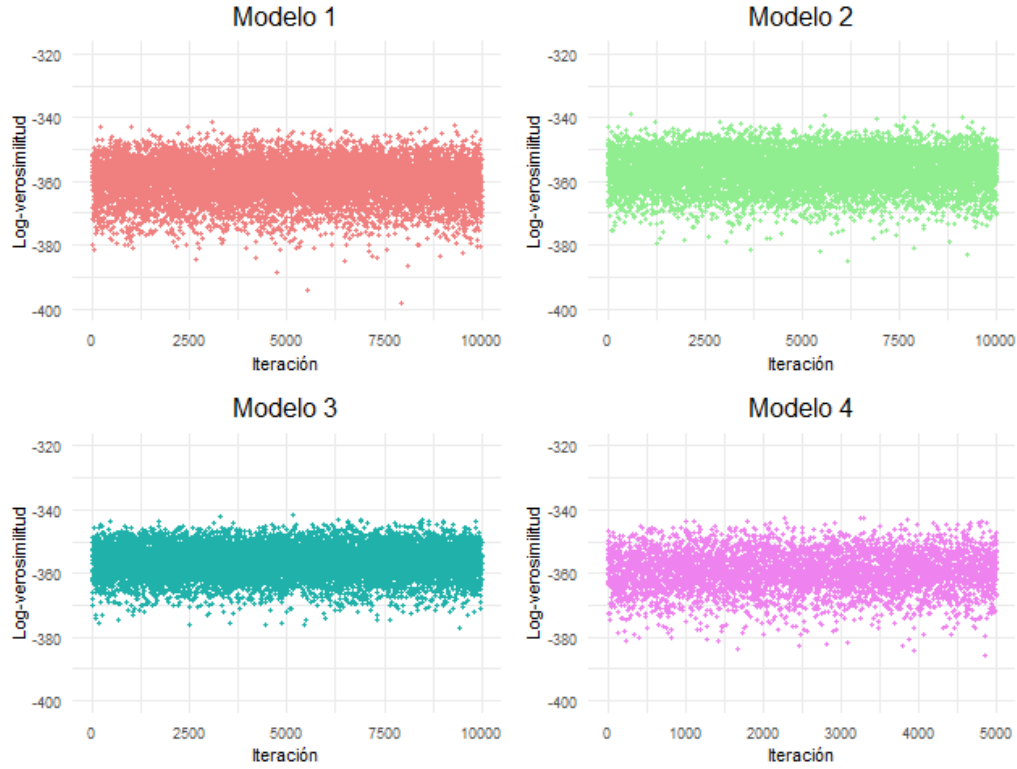


Figure 5: Cadenas de Log-Verosimilitud

Se evidencia la convergencia de las cadenas de Log-versomilitud para cada modelo.

Los resúmenes de los tamaños efectivos de muestra de β en cada modelo, y para los demás parámetros se muestra en las siguientes tablas:

Modelo	Mínimo	Q 25%	Mediana	Media	Q 75%	Máximo
1	9037	9990	10000	9947	10000	11366
2	9074	10000	10000	10044	10000	11584
3	9385	10000	10000	10081	10000	11246
4	4463	5000	5000	4989	5000	6024

Tabla 6: Resumen de tamaños efectivos de muestra para β_j

Modelo	Parámetros	N. Efectivo de muestra
1	σ^2	10000
2	σ^2	10000
3	σ^2	10000
	λ	10000
4	σ^2	5000
	ρ	4568

Tabla 7: Tamaños efectivos de muestra para σ^2, λ y ρ

Cabe resaltar que para todos los modelos se hicieron 1000 iteraciones de calentamiento y 100000 iteraciones más para la inferencia, con un muestreo sistemático de amplitud 10, excepto en el último modelo, en el que se realizaron 1000 iteraciones de calentamiento y 50000 iteraciones para la inferencia, con el mismo muestreo sistemático. De ahí, todos los modelos convergieron y los tamaños efectivos de muestra son suficientes.

Los resúmenes de los coeficientes de variación de Monte Carlo en porcentaje, para β en cada modelo, y para los demás parámetros se muestran en las siguientes tablas:

Modelo	Mínimo	Q 25%	Mediana	Media	Q 75%	Máximo
1	0.225	0.965	1.277	2.848	2.165	48.941
2	0.202	0.863	1.145	2.280	1.942	22.445
3	0.188	1.020	1.772	7.661	2.489	309.044
4	0.309	1.237	1.888	3.378	2.823	33.422

Tabla 8: Resumen de coeficientes de variación de Monte Carlo en porcentaje para β_j

Modelo	Parámetros	CV (%)
1	σ^2	0.085
2	σ^2	0.076
3	σ^2	0.077
	λ	0.201
4	σ^2	0.121
	ρ	0.944

Tabla 9: Coeficientes de variación de Monte Carlo para σ^2, λ y ρ

Se evidencia que los coeficientes de variación de Monte Carlo para los primeros tres cuantiles no son muy altos, además, las medias de los coeficientes de variación de Monte Carlo no son superiores al 5% salvo en el tercer modelo, lo que indica que en general los coeficientes de variación de Monte Carlo son pequeños. Para los coeficientes de variación que son muy grandes, esto se debe principalmente a que los β_j relacionados con esos coeficientes son muy cercanos a 0, por lo que en el cálculo, este coeficiente se dispara, y por lo mismo, es una forma de decir que la variable relacionada a ese β_j no es significativa, como es el caso de el modelo 3, en el que β_{14} tiene un coeficiente de variación de Monte Carlo del 309%, debido a que β_{14} es muy cercano a 0 y su error de Monte Carlo también es muy cercano a 0.

Referencias

- Bernardo, J. M. and Smith, A. F. (2000). *Bayesian theory*. John Wiley & Sons.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. CRC Press.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.