



# Universidad Nacional de Colombia

FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA

## CASO DE ESTUDIO II

*Estadística Bayesiana*  
*Prof. Juan Camilo Sosa*

### **Autores**

Mateo Santiago Cardona Ayala - Estadística -  
mcardonaay@unal.edu.co  
Santiago Arias Garzón - Estadística - saarias@unal.edu.co

Bogotá, 16 de Junio de 2023

# Preguntas

Ajustar cada modelo utilizando los datos de entrenamiento  $(\mathbf{y}_{\text{train}}, \mathbf{X}_{\text{train}})$ .

Nota: Incluir un anexo describiendo los detalles del ajuste de cada modelo.

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  $\hat{\mathbf{y}}_{\text{test}}$  frente  $\mathbf{y}_{\text{test}}$  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  $2 \times 2$  paneles (todos los paneles deben tener la misma escala en los ejes).
- En cada gráfico, superponer la recta  $\hat{\mathbf{y}}_{\text{test}} = \mathbf{y}_{\text{test}}$  como punto de referencia (ver el segundo panel de la Figura 9.7 de ?, p. 170).
- En el encabezado de cada gráfica, incluir el error absoluto medio usando tres (3) cifras decimales.

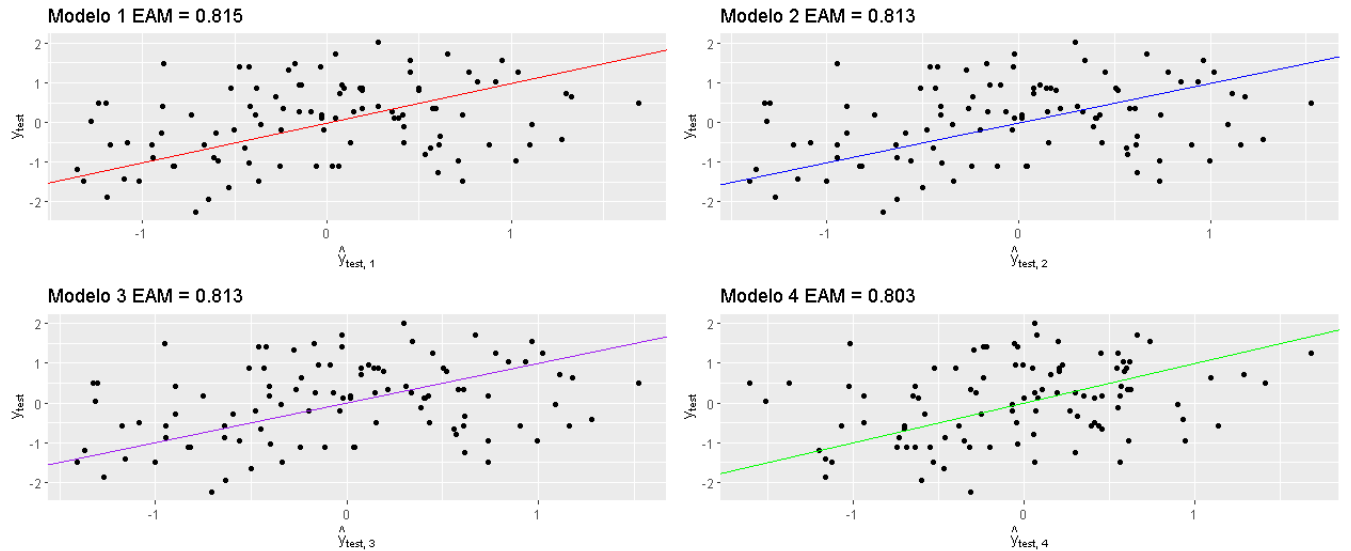


Figure 1: Gráfico de valores observados contra ajustados en el conjunto de prueba

2. Para cada modelo, chequear la bondad de ajuste usando como estadísticos de prueba de la media y la desviación estándar. Graficar la distribución predictiva posterior de

ambos estadísticos simultáneamente por medio de un dispersograma con las muestras correspondientes.

Nota:

- Incluir todos los gráficos en una sola figura con  $2 \times 2$  paneles (todos los paneles deben tener la misma escala en los ejes).
- En cada gráfico, superponer el valor observado de la media y de la desviación estándar como punto de referencia. Ver la Sección 3.4 en <https://rpubs.com/jstats1702/1017076>.
- En el encabezado de cada gráfica, incluir los valores  $p$  predictivos posteriores correspondientes usando tres (3) cifras decimales.

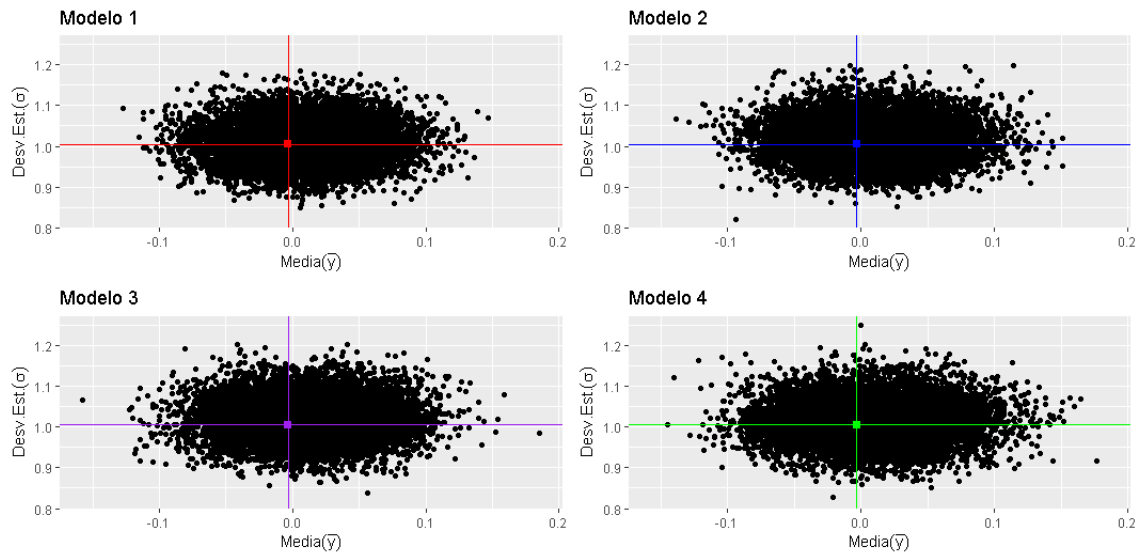


Figure 2: Gráfico de las distribuciones predictivas posteriores de la media y la desviación estándar

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

| Modelo    | 1       | 2       | 3       | 4       |
|-----------|---------|---------|---------|---------|
| $p_{DIC}$ | 10.473  | 11.008  | 10.931  | 9.927   |
| DIC       | 748.521 | 748.746 | 748.611 | 797.778 |

Table 1: Tabla con los DIC y  $p_{DIC}$  de los modelos

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

En términos predictivos el modelo 4 tiene una capacidad algo mayor, sin embargo no es muy distinta de la de los demás modelos, los cuales solo difieren en EAM (Error Absoluto Medio) por 0.01 unidades, es decir, los cuatro modelos hacen predicciones bastante acertadas por lo que en este caso sería mejor optar por los modelos más simples como el 1 y el 2, ya que la estructura de correlación contemplada en el modelo 3 no parece brindar una mejora significativa y si representa un costo computacional mayor, al igual que la selección automática de variables del modelo 4 a pesar de que tiene una calidad predictiva mayor no es considerable para el costo computacional que representa. Ahora en cuanto a la bondad de ajuste primero se debe mencionar la gran similitud entre las distribuciones predictivas de la media y la desviación estandar lo que apoya lo dicho anteriormente, e incluso se debe hacer mención a la presencia de posibles datos atípicos en los modelos 3 y 4, además en cuanto a DIC el modelo 1 es el mejor con una diferencia no muy grande con respecto a los modelos 2 y 3, lo cual es bastante consistente con las observaciones que se han hecho previamente, mientras que algo que vale la pena ver es que el modelo 4 a pesar de tener un número efectivo de parámetros menor y un mayor poder predictivo tiene un DIC considerablemente mayor a los demás, por lo que es posible que aunque los efectos de las variables sean muy cercanos a cero sigan siendo significativos, ya que las variables están estandarizadas, y el modelo 4 no se ajuste de la mejor forma por eso. En conclusión, los resultados son muy similares en cuanto a predicción y ajuste por lo que es mejor optar por los modelos más simples.

## Anexo

### Modelo 1: Regresión Rígida

Distribución muestral:

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

Distribución previa:

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

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

Se procede a encontrar las distribuciones condicionales completas del modelo, las cuales son;

$$p(\boldsymbol{\beta} \mid \text{resto}) = \mathbf{N}_p((\mathbf{X}^t \mathbf{X} + \lambda \mathbf{I}_p)^{-1} \mathbf{X}^t \mathbf{y}; \sigma^2 (\mathbf{X}^t \mathbf{X} + \lambda \mathbf{I}_p)^{-1})$$

$$\begin{aligned}
p(\sigma^2 \mid \text{resto}) &= \text{Gl} \left( \frac{\nu_0+n+p}{2}; \frac{1}{2}(\nu_0\sigma_0^2 + \text{SSR}(\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^t\boldsymbol{\beta}) \right) \\
p(\lambda \mid \text{resto}) &= \text{Gamma}(a_\lambda + \frac{p}{2}; b_\lambda + \frac{1}{2\sigma^2}\boldsymbol{\beta}^t\boldsymbol{\beta})
\end{aligned}$$

Con las cuales se contruye un muestreador de Gibbs con los siguientes pasos:  
Desde un estado actual  $(\boldsymbol{\beta}^{(b)}, (\sigma^2)^{(b)}, \lambda^{(b)})$  se genera el siguiente estado como sigue;

- (a) Muestrear  $\boldsymbol{\beta}^{(b+1)} \sim \text{N}(\boldsymbol{\beta} \mid (\sigma^2)^{(b)}, \lambda^{(b)})$
- (b) Muestrear  $(\sigma^2)^{(b+1)} \sim \text{Gl}(\sigma^2 \mid \lambda^{(b)}, \boldsymbol{\beta}^{(b+1)})$
- (c) Muestrear  $\lambda^{(b+1)} \sim \text{Gamma}(\lambda \mid (\sigma^2)^{(b+1)}, \boldsymbol{\beta}^{(b+1)})$
- (d) Establecer al estado actual como  $(\boldsymbol{\beta}^{(b+1)}, (\sigma^2)^{(b+1)}, \lambda^{(b+1)})$
- (e) Repetir (a) – (d) hasta convergencia.

En este cas se hicieron 10000 iteraciones con un burn de 1000 iteraciones para hacer la convergencia y además se tomaron los estados iniciales como  $(\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}_{\text{OLS}}, (\sigma^2)^{(0)} = \sigma_{\text{OLS}}^2, \lambda^{(0)} = 1)$ .

## Modelo 2: Regresión Lasso

Distribución muestral:

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

Distribución previa:

$$p(\boldsymbol{\beta}, \sigma^2) = \prod_{j=1}^p \text{DE}(\beta_j \mid 0, \tau_0) \cdot \text{Gl}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2/2),$$

con  $\tau_0 = 5$ ,  $\nu_0 = 1$  y  $\sigma_0^2 = \hat{\sigma}_{\text{OLS}}^2$ , donde  $\text{DE}(\mu, \sigma)$  es la distribución Doble Exponencial (distribución de Laplace) con media  $\mu$  y varianza  $2\sigma^2$ .

Nota: Si  $\theta \mid \phi^2 \sim \text{N}(\mu, \phi^2)$  y  $\phi^2 \sim \text{G}(1, \frac{1}{2\tau^2})$ , entonces  $\theta \sim \text{DE}(\mu, \tau)$ .

Veamos las distribuciones condicionales completas del modelo;

$$\begin{aligned}
p(\boldsymbol{\beta} \mid \text{resto}) &= \text{N}_p \left( (\mathbf{X}^t\mathbf{X} + \sigma^2\mathbf{D}_{\phi^2}^{-1})^{-1}\mathbf{X}^t\mathbf{y}; \sigma^2(\mathbf{X}^t\mathbf{X} + \sigma^2\mathbf{D}_{\phi^2}^{-1})^{-1} \right) \\
p(\sigma^2 \mid \text{resto}) &= \text{Gl} \left( \frac{\nu_0+n}{2}; \frac{1}{2}(\nu_0\sigma_0^2 + \text{SSR}(\boldsymbol{\beta})) \right) \\
p(\frac{1}{\phi_j^2} \mid \text{resto}) &= \text{IG} \left( \frac{1}{\sqrt{\beta_j^2\tau_0^2}}; \frac{1}{\tau_0^2} \right)
\end{aligned}$$

Donde  $\mathbf{D}_{\phi^2} = \mathbf{diag}(\phi_j^2)$

Con las cuales se contruye un muestreador de Gibbs con los siguientes pasos:

Desde un estado actual  $(\boldsymbol{\beta}^{(b)}, (\sigma^2)^{(b)}, (\phi^2)^{(b)})$  se genera el siguiente estado como sigue;

- (a) Muestrear  $\boldsymbol{\beta}^{(b+1)} \sim \mathbf{N}(\boldsymbol{\beta} \mid (\sigma^2)^{(b)}, (\phi^2)^{(b)})$
- (b) Muestrear  $\frac{1}{(\phi^2)^{(b+1)}} \sim \text{IG}(\frac{1}{\phi^2} \mid \boldsymbol{\beta}^{(b+1)})$
- (c) Muestrear  $(\sigma^2)^{(b+1)} \sim \text{GI}(\sigma^2 \mid \boldsymbol{\beta}^{(b+1)})$
- (d) Establecer al estado actual como  $(\boldsymbol{\beta}^{(b+1)}, (\sigma^2)^{(b+1)}, (\phi^2)^{(b+1)})$
- (e) Repetir (a) – (d) hasta convergencia.

En este cas se hicieron 10000 iteraciones con un burn de 1000 iteraciones para hacer la convergencia y además se tomaron los estados iniciales como  $(\boldsymbol{\beta}^{(0)} = \boldsymbol{\beta}_{\text{OLS}}, (\sigma^2)^{(0)} = \sigma_{\text{OLS}}^2, (\phi^2)^{(0)} = \boldsymbol{\beta}_{\text{OLS}}^t \boldsymbol{\beta}_{\text{OLS}})$ .

### Modelo 3: Regresión con errores correlacionados

Distribución muestral:

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

donde  $\mathbf{C}_\rho$  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:

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

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

Veamos las distribuciones condicionales completas del modelo;

$$\begin{aligned} p(\boldsymbol{\beta} \mid \text{resto}) &= \mathbf{N}_p \left( (\mathbf{X}^t \mathbf{C}_\rho^{-1} \mathbf{X} + \frac{\sigma^2}{\tau_0^2} \mathbf{I}_p)^{-1} \mathbf{X}^t \mathbf{C}_\rho^{-1} \mathbf{y}; \sigma^2 (\mathbf{X}^t \mathbf{C}_\rho^{-1} \mathbf{X} + \frac{\sigma^2}{\tau_0^2} \mathbf{I}_p)^{-1} \right) \\ p(\sigma^2 \mid \text{resto}) &= \text{GI} \left( \frac{\nu_0 + n}{2}; \frac{1}{2} (\nu_0 \sigma_0^2 + \text{SSR}(\boldsymbol{\beta})) \right) \end{aligned}$$

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

Debido a que  $\rho$  no tiene una distribución condicional completa cerrada vamos a implementar en el muestreador de Gibbs un algoritmo de MonteCarlo Hamiltoniano para  $\rho$ . Se puede mostrar que para  $0 \leq \rho \leq 1$  se cumplen las siguientes propiedades;

$$\begin{aligned} \left| \mathbf{C}_\rho \right| &= (1 - \rho^2)^{n-1} \\ \mathbf{C}_\rho^{-1} &= \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} \\ \frac{d}{d\rho} \mathbf{C}_\rho^{-1} = \mathbf{C}_\rho'^{-1} &= \frac{1}{(1 - \rho^2)^2} \begin{bmatrix} 2\rho & -(1 + \rho^2) & 0 & \cdots & 0 & 0 \\ -(1 + \rho^2) & 4\rho & -(1 + \rho^2) & \cdots & 0 & 0 \\ 0 & -(1 + \rho^2) & 4\rho & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 4\rho & -(1 + \rho^2) \\ 0 & 0 & 0 & \cdots & -(1 + \rho^2) & 2\rho \end{bmatrix} \end{aligned}$$

De donde podemos obtener;

$$\begin{aligned} p(\rho \mid \text{resto}) &= (2\pi)^{-\frac{n}{2}} \left| \mathbf{C}_\rho \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^t \mathbf{C}_\rho^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ p(\rho \mid \text{resto}) &= (2\pi)^{-\frac{n}{2}} (1 - \rho^2)^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^t \mathbf{C}_\rho^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ \log p(\rho \mid \text{resto}) &= -\frac{n}{2} \ln(2\pi) - \frac{n-1}{2} \ln(1 - \rho^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^t \mathbf{C}_\rho^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial}{\partial \rho} \log p(\rho \mid \text{resto}) &= -\frac{(n-1)\rho}{1 - \rho^2} - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^t \mathbf{C}_\rho'^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

Por lo que se procede a realizar el muestreador de Gibbs con los siguientes pasos;  
Desde un estado actual  $(\boldsymbol{\beta}^{(b)}, (\sigma^2)^{(b)}, (\phi^2)^{(b)})$  se genera el siguiente estado como sigue;

- (a) Muestrear  $\boldsymbol{\beta}^{(b+1)} \sim \mathbf{N}(\boldsymbol{\beta} \mid (\sigma^2)^{(b)}, \rho^{(b)})$
- (b) Muestrear  $(\sigma^2)^{(b+1)} \sim \text{Gl}(\sigma^2 \mid \boldsymbol{\beta}^{(b+1)}, \rho^{(b)})$

(c) Muestrar  $\rho^{(b+1)} \sim p(\rho \mid \text{resto})$  con los siguientes pasos;

i. Simular  $\varphi \sim \mathbf{N}(0, \frac{1}{3})$

ii. Actualizar  $(\rho^{(l)}, \varphi^{(l)})$  mediante 20 pasos como sigue;

A. Actualizar  $\varphi$ :

$$\varphi^* \leftarrow \varphi + \frac{1}{40} \frac{\partial}{\partial \rho} \log p(\rho \mid \text{resto})$$

B. Actualizar  $\rho$ :

$$\rho^* \leftarrow \rho + \frac{\varphi}{60}$$

C. Repetir 19 veces

D. Actualizar  $\varphi$  para completar el último paso

iii. Calcular la tasa de aceptación;

$$r = \frac{p(\rho^* \mid \text{resto})p(\varphi^*)}{p(\rho^{(l)} \mid \text{resto})p(\varphi^{(l)})}$$

iv. Establecer  $\rho^{(b+1)} = \rho^*$  con probabilidad  $\min\{1, r\}$  o  $\rho^{(b+1)} = \rho^{(b)}$  con probabilidad  $1 - \min\{1, r\}$

(d) Establecer al estado actual como  $(\beta^{(b+1)}, (\sigma^2)^{(b+1)}, \rho^{(b+1)})$

(e) Repetir (a) – (d) hasta convergencia.

En este caso se hicieron 10000 iteraciones con un burn de 1000 iteraciones para asegurar la convergencia, también se usaron 20 iteraciones para el algoritmo de MonteCarlo Hamiltoniano para obtener una tasa de aceptación de aproximadamente el 82% y además se tomaron los estados iniciales como  $(\beta^{(0)} = \beta_{\text{OLS}}, (\sigma^2)^{(0)} = \sigma_{\text{OLS}}^2, \rho^{(0)} = \frac{1}{2})$ .

## Modelo 4: Regresión con selección automática de covariables

Distribución muestral:

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

donde  $\mathbf{b} = (b_1, \dots, b_p) \in \mathbb{R}^p$ ,  $\mathbf{z} = (z_1, \dots, z_p) \in \{0, 1\}^p$  y  $\boldsymbol{\beta} = (b_1 \cdot z_1, \dots, b_p \cdot z_p)$ .

Distribución previa:

$$p(\mathbf{b}, \mathbf{z}, \sigma^2) = \mathbf{N}(\mathbf{b}_z \mid \mathbf{0}_{p_z}, n \sigma^2 (\mathbf{X}_z^\top \mathbf{X}_z)^{-1}) \cdot \text{Gl}(\sigma^2 \mid \nu_0/2, \nu_0 \sigma_0^2/2) \cdot \prod_{j=1}^p \text{Ber}(z_j \mid p_0),$$

con  $g = n$ ,  $\nu_0 = 1$ ,  $\sigma_0^2 = \hat{\sigma}_{\text{OLS}}^2$  y  $p_0 = 0.5$ , donde  $p_z = \sum_{j=1}^p z_j$ ,  $\mathbf{b}_z$  es un vector de  $p_z \times 1$  conformado por las entradas de  $\boldsymbol{\beta}$  tales que  $z_j = 1$  y  $\mathbf{X}_z$  es una matriz de  $n \times p_z$



correspondiente a la matriz de diseño asociada con las entradas de  $\boldsymbol{\beta}$  tales que  $z_j = 1$ . Además se define la matriz  $\mathbf{C}_z$  como  $\mathbf{diag}(\mathbf{z})$  sin las columnas de ceros, por lo que su tamaño es  $p \times p_z$ . Se puede ver que  $\mathbf{X}_z = \mathbf{X}\mathbf{C}_z$  y  $\boldsymbol{\beta} = \mathbf{C}_z\mathbf{b}_z$ .

Primero vamos a obtener  $p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}, \sigma^2)$ ;

$$\begin{aligned}
p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}, \sigma^2) &= \int_{-\infty}^{\infty} p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}, \sigma^2, \mathbf{b}_z) \cdot p(\mathbf{b}_z \mid \mathbf{X}, \mathbf{z}, \sigma^2) d\mathbf{b}_z \\
&= \int_{-\infty}^{\infty} \mathbf{N}_n(\mathbf{y} \mid \mathbf{X}\mathbf{C}_z\mathbf{b}_z, \sigma^2\mathbf{I}_n) \cdot \mathbf{N}(\mathbf{b}_z \mid \mathbf{0}_{p_z}, n\sigma^2(\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z)^{-1}) d\mathbf{b}_z \\
&= \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\mathbf{C}_z\mathbf{b}_z)^t(\mathbf{y} - \mathbf{X}\mathbf{C}_z\mathbf{b}_z)\right\} \\
&\quad \times (2\pi n\sigma^2)^{-\frac{p_z}{2}} \left|\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2n\sigma^2}\mathbf{b}_z^t\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z\mathbf{b}_z\right\} d\mathbf{b}_z \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^t\left(\mathbf{I}_n + \frac{n}{n+1}\mathbf{X}\mathbf{C}_z(\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z)^{-1}\mathbf{C}_z^t\mathbf{X}^t\right)\mathbf{y}\right\} \\
&\quad \times \int_{-\infty}^{\infty} \mathbf{N}\left(\mathbf{b}_z \mid \frac{n}{n+1}(\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z)^{-1}\mathbf{C}_z^t\mathbf{X}^t\mathbf{y}; \frac{n\sigma^2}{n+1}(\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z)^{-1}\right) d\mathbf{b}_z \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} (n+1)^{-\frac{p_z}{2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^t\left(\mathbf{I}_n + \frac{n}{n+1}\mathbf{X}\mathbf{C}_z(\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z)^{-1}\mathbf{C}_z^t\mathbf{X}^t\right)\mathbf{y}\right\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} (n+1)^{-\frac{p_z}{2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{SSR}_n\right\}
\end{aligned}$$

Y ahora vamos a obtener  $p(\sigma^2 \mid \mathbf{X}, \mathbf{z}, \mathbf{y})$  de la siguiente forma;

$$\begin{aligned}
p(\sigma^2 \mid \mathbf{X}, \mathbf{z}, \mathbf{y}) &\propto p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}, \sigma^2) \times p(\sigma^2) \\
&\propto (2\pi\sigma^2)^{-\frac{n}{2}} (n+1)^{-\frac{p_z}{2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{SSR}_n\right\} \times (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\} \\
&\propto (\sigma^2)^{-\left(\frac{\nu_0+n}{2}+1\right)} \exp\left\{-\frac{1}{\sigma^2}\left(\frac{\nu_0\sigma_0^2 + \mathbf{SSR}_n}{2}\right)\right\} \\
&= \text{Gl}\left(\sigma^2 \mid \frac{\nu_0+n}{2}; \frac{\nu_0\sigma_0^2 + \mathbf{SSR}_n}{2}\right)
\end{aligned}$$

Además obteniendo la distribución condicional completa de  $\mathbf{b}_z$ , a saber,

$$\mathbf{b}_z \mid \mathbf{X}, \mathbf{z}, \sigma^2, \mathbf{y} \sim \mathbf{N}\left(\frac{n}{n+1}(\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z)^{-1}\mathbf{C}_z^t\mathbf{X}^t\mathbf{y}; \frac{n\sigma^2}{n+1}(\mathbf{C}_z^t\mathbf{X}^t\mathbf{X}\mathbf{C}_z)^{-1}\right)$$

Ahora podemos obtener muestras de la distribución  $p(\mathbf{b}_z, \sigma^2 \mid \mathbf{X}, \mathbf{z}, \mathbf{y})$  con MonteCarlo en vez de usar un muestreador de Gibbs.

Ahora procedemos a hallar  $p(\mathbf{y} \mid \mathbf{X}, \mathbf{z})$ ;

$$p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}) = \int_0^\infty p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}, \sigma^2) \times p(\sigma^2) d\sigma^2$$

$$\begin{aligned}
&= \int_0^\infty (2\pi)^{-\frac{n}{2}} (n+1)^{-\frac{p_z}{2}} \frac{\left(\frac{\nu_0 \sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} (\sigma^2)^{-\left(\frac{\nu_0+n}{2}+1\right)} \exp\left\{-\frac{1}{\sigma^2} \left(\frac{\nu_0 \sigma_0^2 + \text{SSR}_n}{2}\right)\right\} d\sigma^2 \\
&= \pi^{-\frac{n}{2}} (n+1)^{-\frac{p_z}{2}} \frac{\Gamma\left(\frac{\nu_0+n}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)} \frac{(\nu_0 \sigma_0^2)^{\frac{\nu_0}{2}}}{(\nu_0 \sigma_0^2 + \text{SSR}_n)^{\frac{\nu_0+n}{2}}} \int_0^\infty \text{Gl}\left(\sigma^2 \mid \frac{\nu_0+n}{2}; \frac{\nu_0 \sigma_0^2 + \text{SSR}_n}{2}\right) d\sigma^2 \\
&= \pi^{-\frac{n}{2}} (n+1)^{-\frac{p_z}{2}} \frac{\Gamma\left(\frac{\nu_0+n}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)} \frac{(\nu_0 \sigma_0^2)^{\frac{\nu_0}{2}}}{(\nu_0 \sigma_0^2 + \text{SSR}_n)^{\frac{\nu_0+n}{2}}}
\end{aligned}$$

Cabe aclarar que  $\sigma_0^2$  es la varianza estimada por el método de mínimos cuadrados ordinarios con  $\mathbf{X}_z$  por lo que debe ser actualizado si  $z$  se actualiza también.

Por lo que la distribución condicional completa de  $\mathbf{z}_j$  es;

$$p(\mathbf{z}_j \mid \mathbf{X}, \mathbf{y}, \mathbf{z}_{-j}) = \text{Ber}\left(\frac{o_j}{1+o_j}\right)$$

Donde

$$o_j = \frac{p(\mathbf{z}_j = 1 \mid \mathbf{X}, \mathbf{y}, \mathbf{z}_{-j})}{p(\mathbf{z}_j = 0 \mid \mathbf{X}, \mathbf{y}, \mathbf{z}_{-j})} = \frac{p(z_j = 1)}{p(z_j = 0)} \times \frac{p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}_{-j}, z_j = 1)}{p(\mathbf{y} \mid \mathbf{X}, \mathbf{z}_{-j}, z_j = 0)}$$

Donde  $\mathbf{z}_{-j}$  es el vector  $z$  sin la entrada  $j$ -ésima.

Por lo que ahora se tiene todo lo necesario para hacer un Muestreador de Gibbs de la siguiente forma partiendo desde un estado  $(\mathbf{b}_z^{(s)}, (\sigma^2)^{(s)}, \mathbf{z}^{(s)})$

- (a) Muestrear cada  $\mathbf{z}_j$  de  $p(\mathbf{z}_j \mid \mathbf{X}, \mathbf{y}, \mathbf{z}_{-j})$  con  $1 \leq j \leq p$  en un orden aleatorio y hacer un vector que será  $\mathbf{z}^{(b+1)}$ .
- (b) Muestrear  $(\sigma^2)^{(s+1)} \sim \text{Gl}(\sigma^2 \mid \mathbf{z}^{(s+1)})$
- (c) Muestrear  $\mathbf{b}_z^{(s+1)} \sim \text{N}(\mathbf{b}_z \mid (\sigma^2)^{(s+1)}, \mathbf{z}^{(s+1)})$
- (d) Actualizar el estado actual a  $(\mathbf{b}_z^{(s+1)}, (\sigma^2)^{(s+1)}, \mathbf{z}^{(s+1)})$

A pesar de que no se están tomando las muestras de las distribuciones condicionales completas no hay ningún problema ya que se converge en distribución a la distribución posterior. En este caso se hicieron 10000 iteraciones con un burn de 1000 iteraciones para hacer la convergencia y se tomaron como estados iniciales  $(\mathbf{b}_z^{(0)} = \boldsymbol{\beta}_{\text{OLS}}, (\sigma^2)^{(0)} = \sigma_{\text{OLS}}^2, \mathbf{z}^{(0)} = \mathbf{1}_{10})$

## Referencias