



# 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, 17 de Mayo de 2023

# Preguntas

1. Hacer el DAG de  $M_6$  incluyendo las variables auxiliares.

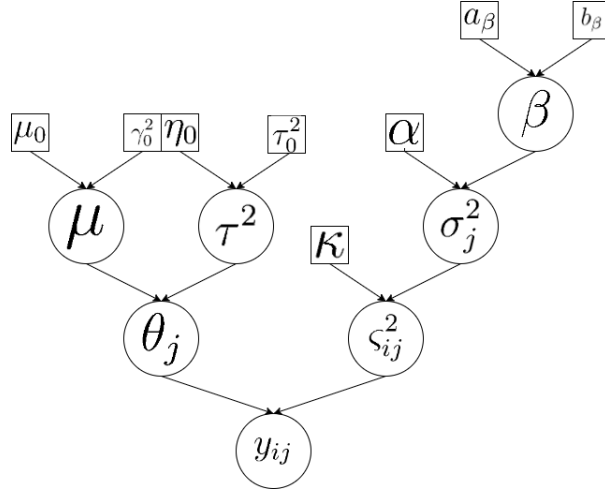


Figure 1: DAG de  $M_6$

2. Graficar la cadena de la log-verosimilitud de cada  $M_k$ , para  $k = 1, \dots, 6$ . Incluir un anexo con las distribuciones condicionales completas de cada modelo (no incluir la demostración, solo cada distribución con sus respectivos parámetros).

Nota: usar un solo panel para facilitar la comparación.

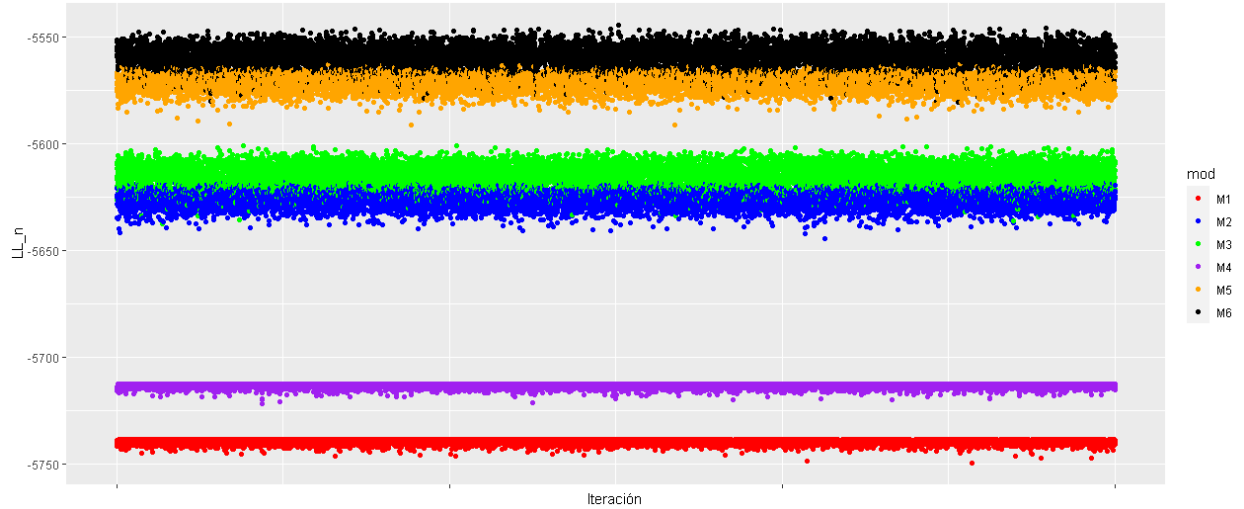


Figure 2: Cadenas de la log-verosimilitud de los 6 modelos

3. Calcular el DIC y el  $p_{\text{DIC}}$  de cada  $M_k$ , para  $k = 1, \dots, 6$ . Presentar los resultados tabularmente. Interpretar los resultados obtenidos (máximo 100 palabras).

Modelo	$p_{\text{DIC}}$	DIC
$M_1$	1.980	11480.83
$M_2$	24.242	11276.77
$M_3$	47.647	11276.59
$M_4$	1.937	11428.41
$M_5$	24.591	11168.03
$M_6$	48.206	11166.02

Gracias al DIC podemos ver que el modelos 6 es el mejor, además en la cadena de verosimilitud vemos que la cadena converge con una verosimilitud mayor. Sin embargo la diferencia entre el DIC de los modelos 5 y 6 no es tan grande para tener aproximadamente el doble de número efectivo de parámetros. También es importante ver que respecto al modelo 3, el modelo 6 es mejor con un número efectivo de parámetros muy similar, en general los modelos que usan la distribución t escalada, parecen ser mejores que sus análogos con la distribución normal

4. Calcular los ppp's para Bogotá asociados con la media, la mediana, la desviación estándar, el coeficiente de variación, el rango y el rango intercuartílico usando  $M_6$ . Presentar los resultados tabularmente. Interpretar los resultados obtenidos (máximo 100 palabras).

Estadístico	ppp	Valor
Media	0.791	13.796
Mediana	0.477	13.934
Desv. Est.	0.909	1.174
Coef. Var	0.909	0.085
Rango	0.816	7.262
Rango Intercuart.	0.455	1.198

Los ppp son significativos para la mediana y el rango intercuartílico, posiblemente quiere decir que el modelo no está capturando los datos atípicos debidamente a pesar de estar trabajando en escala logarítmica.

5. Hacer un ranking Bayesiano de los dominios basado en los efectos promedio de  $M_6$ . Hacer una visualización que incluya simultáneamente las estimaciones puntuales y los intervalos de credibilidad al 95%. Interpretar los resultados obtenidos (máximo 100 palabras).

Nota: hacer la visualización en escala logarítmica. Además, usar la siguiente convención

de colores: rojo oscuro para efectos promedio significativamente inferiores a 13.830, negro para efectos promedio que no difieren significativamente de 13.830 y verde oscuro para efectos promedio significativamente superiores a 13.830 (observe que 13.830 corresponde a un SMLMV de 2022 en escala logarítmica).

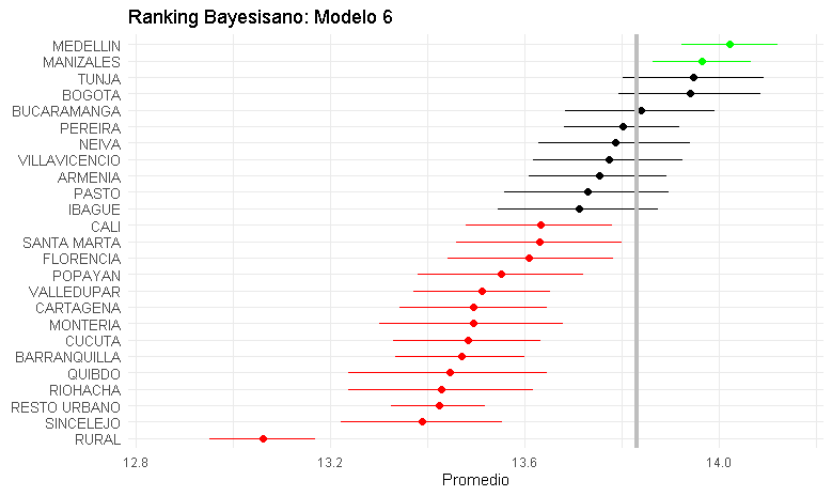


Figure 3: Ranking Bayesiano

La mayoría de los ingresos promedio de los dominios son significativamente menores que el salario mínimo, y solo dos dominios son significativaemnte mayores en cuanto a sus ingresos promedio que el salario mínimo. También es importante ver que el sector rural tiene un ingreso promedio significativamente más bajo que el resto de los dominios.

- Estimar puntualmente la media, la desviación estándar y el coeficiente de variación de los ingresos para el Top 5 del ranking usando  $M_6$ . Reportar los resultados tabularmente. Interpretar los resultados obtenidos (máximo 100 palabras).

Nota: reportar las estimaciones de la media y la desviación estándar usando la escala real en pesos y del coeficiente de variación en puntos porcentuales.

Dominio	Media estimada	Desv. Estándar Estimada	Coef. Var. Estimado
BUCARAMANGA	1023809.231	3.958	5.7%
BOGOTA	1133739.726	4.071	5.8%
TUNJA	1140630.178	4.451	6.2%
MANIZALES	1160077.145	3.335	5%
MEDELLIN	1227820.27	2.615	4%

Se logra evidenciar que la variabilidad entre los dominios puede ser relativamente baja, ya que las desviaciones estándar no muestran valores muy disitntos para los 5 dominios

con mayores efectos promedio, a la vez que el coeficiente de variación se mantiene dentro del mismo rango de valores, por lo que se refuerza lo señalado en puntos anteriores sobre considerar el modelo 5, en donde se tiene solamente una varianza general para todos los dominios. También era de esperarse que las ciudades principales del país se encontraran dentro de los puestos más altos del ranking debido a que concentran los flujos más grandes de capital.

7. Conformar un arreglo que contenga las estimaciones puntuales de los efectos promedio y las desviaciones estándar de todos los dominios usando  $M_6$ . Usando este arreglo como insumo, segmentar los dominios por medio de agrupación jerárquica con cuatro grupos. Presentar los resultados visualmente. Interpretar los resultados obtenidos (máximo 100 palabras).

Nota: estandarizar las columnas del arreglo antes de hacer el proceso de segmentación.

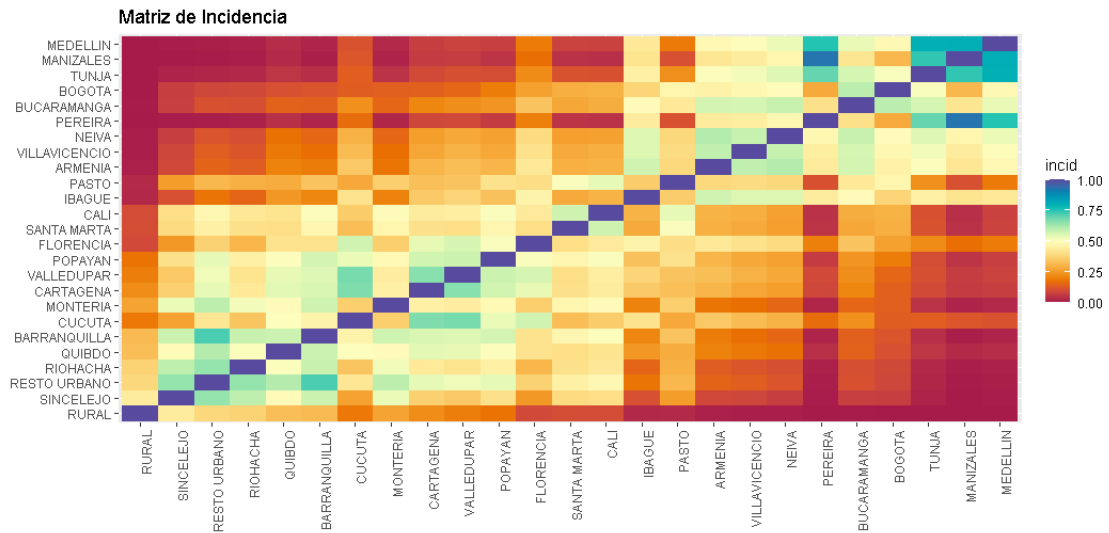
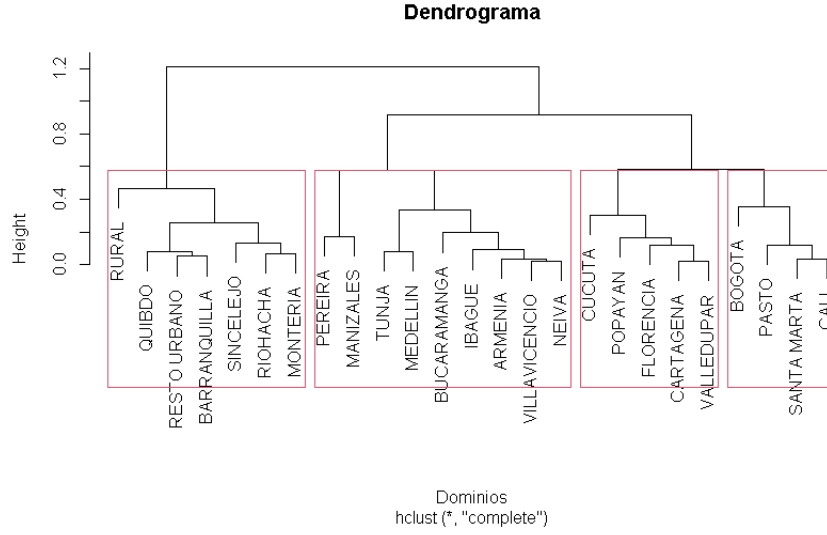


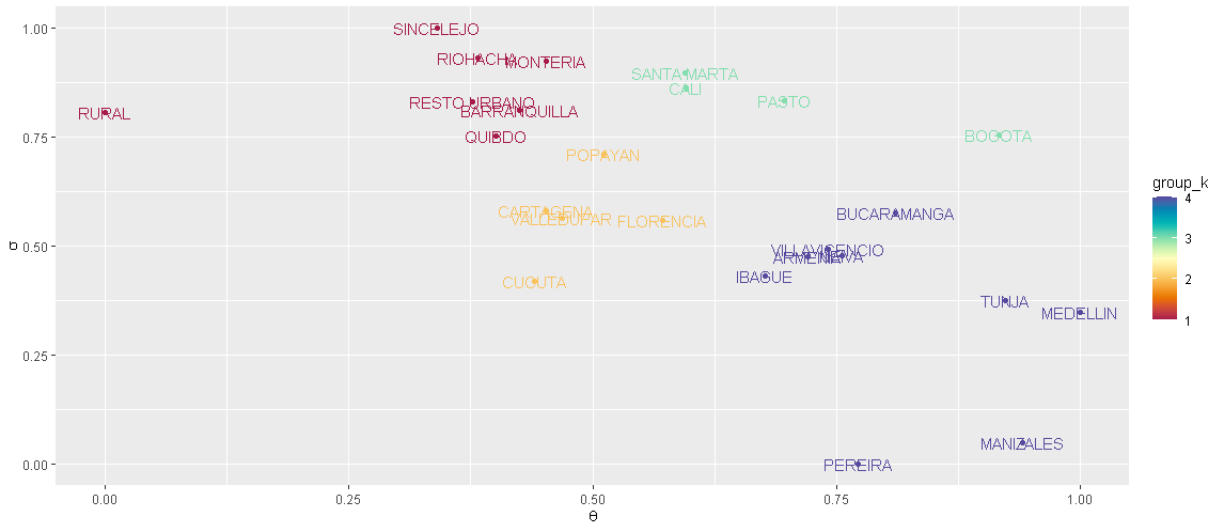
Figure 4: Matriz de Incidencia

En la matriz de incidencia podemos ver que principalmente los dominios se dividen en dos grupos, los primeros del ranking y los últimos del ranking, además dentro de cada grupo podemos ver dos subgrupos, también de acuerdo al ranking, por lo que podría decirse que entre más cerca estén dos dominios dentro del ranking bayesiano es más probable que estén dentro del mismo grupo.

Además se realizó el clustering con los estimadores bayesianos, usando la distancia euclidiana y el método completo llegando a los resultados vistos en 5 y 6 donde podemos apreciar que los cuatro grupos se diferencian respecto a media grandes, desviaciones



*Figure 5: Dendrograma*



*Figure 6: Gráfico de Clusters*

estándar pequeñas (Azul), medias grandes pero desviaciones estándar también (Verde), medias y desviaciones estándar no tan grandes (Amarillo) y por último medias pequeñas y desviaciones estándar grandes (Rojo).

8. (Bono) Demostrar que la distribución  $t$  se puede expresar como una mezcla de distribuciones Normales ponderadas por una distribución Gamma-Inversa. Es decir, demostrar que la distribución muestral  $y_i \mid \theta, \sigma^2 \stackrel{iid}{\sim} t_\kappa(\theta, \sigma^2)$ , para  $i = 1, \dots, n$ , es equivalente a la

distribución jerárquica dada por

$$y_i \mid \theta, V_i \stackrel{\text{ind}}{\sim} \mathbf{N}(\theta, V_i), \quad V_i \mid \sigma^2 \stackrel{\text{iid}}{\sim} \text{Gl} \left( \frac{\kappa}{2}, \frac{\kappa\sigma^2}{2} \right).$$

Sugerencia: obtener la función de densidad de la distribución  $\mathbf{t}_\kappa(\theta, \sigma^2)$ , para  $i = 1, \dots, n$ , a partir de

$$p(y_i \mid \theta, \sigma^2) = \int_0^\infty p(y_i, V_i \mid \theta, \sigma^2) dV_i.$$

$$\begin{aligned} p(y_i \mid \theta, \sigma^2) &= \int_0^\infty p(y_i, V_i \mid \theta, \sigma^2) dV_i \\ &= \int_0^\infty p(y_i \mid V_i, \theta) \cdot p(V_i \mid \sigma^2) dV_i \\ &= \int_0^\infty \mathbf{N}(y_i \mid \theta; V_i) \cdot \text{Gl} \left( V_i \mid \frac{\kappa}{2}; \frac{\kappa\sigma^2}{2} \right) dV_i \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi V_i}} \exp \left\{ -\frac{1}{2V_i} (y_i - \theta)^2 \right\} \frac{\left( \frac{\kappa\sigma^2}{2} \right)^{\frac{\kappa}{2}}}{\Gamma \left( \frac{\kappa}{2} \right)} (V_i)^{-\left( \frac{\kappa}{2} + 1 \right)} \exp \left\{ -\frac{\kappa\sigma^2}{2V_i} \right\} dV_i \\ &= \frac{\left( \frac{\kappa\sigma^2}{2} \right)^{\frac{\kappa}{2}}}{\sqrt{2\pi} \Gamma \left( \frac{\kappa}{2} \right)} \int_0^\infty (V_i)^{-\left( \frac{\kappa+1}{2} + 1 \right)} \exp \left\{ -\frac{1}{2V_i} (y_i - \theta)^2 - \frac{\kappa\sigma^2}{2V_i} \right\} dV_i \end{aligned}$$

El cual es el núcleo de una distribución  $\text{Gl} \left( \frac{\kappa+1}{2}; \frac{(y_i - \theta)^2 + \kappa\sigma^2}{2} \right)$

$$\begin{aligned} &= \frac{\left( \frac{\kappa\sigma^2}{2} \right)^{\frac{\kappa}{2}}}{\sqrt{2\pi} \Gamma \left( \frac{\kappa}{2} \right)} \cdot \frac{\Gamma \left( \frac{\kappa+1}{2} \right)}{\left( \frac{(y_i - \theta)^2 + \kappa\sigma^2}{2} \right)^{\frac{\kappa+1}{2}}} \\ &= \frac{\left( \frac{\kappa\sigma^2}{2} \right)^{\frac{\kappa}{2}}}{\sqrt{2\pi} \Gamma \left( \frac{\kappa}{2} \right)} \cdot \frac{\Gamma \left( \frac{\kappa+1}{2} \right)}{\left( \frac{\kappa\sigma^2}{2} \right)^{\frac{\kappa+1}{2}} \left( 1 + \frac{(y_i - \theta)^2}{\kappa\sigma^2} \right)^{\frac{\kappa+1}{2}}} \\ &= \frac{1}{\sqrt{\pi \kappa \sigma^2}} \frac{\Gamma \left( \frac{\kappa+1}{2} \right)}{\Gamma \left( \frac{\kappa}{2} \right)} \left( 1 + \frac{(y_i - \theta)^2}{\kappa\sigma^2} \right)^{-\frac{\kappa+1}{2}} \end{aligned}$$

La cual resulta ser la función de densidad de una distribución  $\mathbf{t}_\kappa(\theta; \sigma^2)$  por lo que se ha demostrado que puede expresar como una mezcla de distribuciones Normales ponderadas por una distribución Gamma-Inversa.

## Anexo

En este anexo se incluyen las distribuciones condicionales completas y los DAG's de los modelos.

### M<sub>1</sub>: Modelo Normal

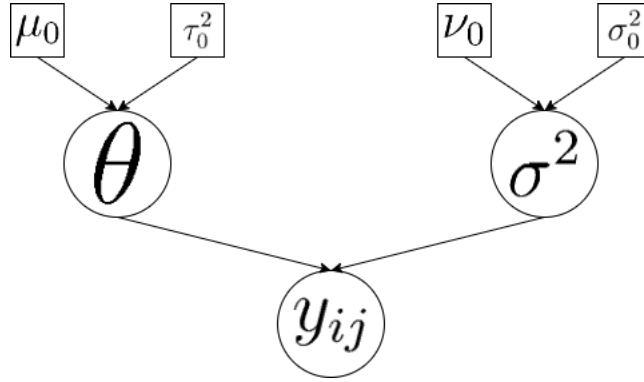


Figure 7: DAG de M<sub>1</sub>

$$\begin{aligned}\theta \mid \text{resto} &\sim \text{N} \left( \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}; \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \right) \\ \sigma^2 \mid \text{resto} &\sim \text{Gl} \left( \frac{\nu_0 + n}{2}; \frac{\nu_0 \sigma_0^2 + \sum_{i=1}^n (y_i - \theta)^2}{2} \right)\end{aligned}$$

### M<sub>2</sub>: Modelo Normal con medias específicas

$$\begin{aligned}\mu \mid \text{resto} &\sim \text{N} \left( \frac{\frac{\mu_0}{\gamma_0^2} + \frac{m\bar{\theta}}{\tau^2}}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}}; \frac{1}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}} \right) \\ \tau^2 \mid \text{resto} &\sim \text{Gl} \left( \frac{\eta_0 + m}{2}; \frac{\eta_0 \tau_0^2 + \sum_{j=1}^m (\theta_j - \mu)^2}{2} \right) \\ \theta_j \mid \text{resto} &\sim \text{N} \left( \frac{\frac{\mu}{\tau^2} + \frac{n_j \bar{y}_j}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n_j}{\sigma^2}}; \frac{1}{\frac{1}{\tau^2} + \frac{n_j}{\sigma^2}} \right)\end{aligned}$$



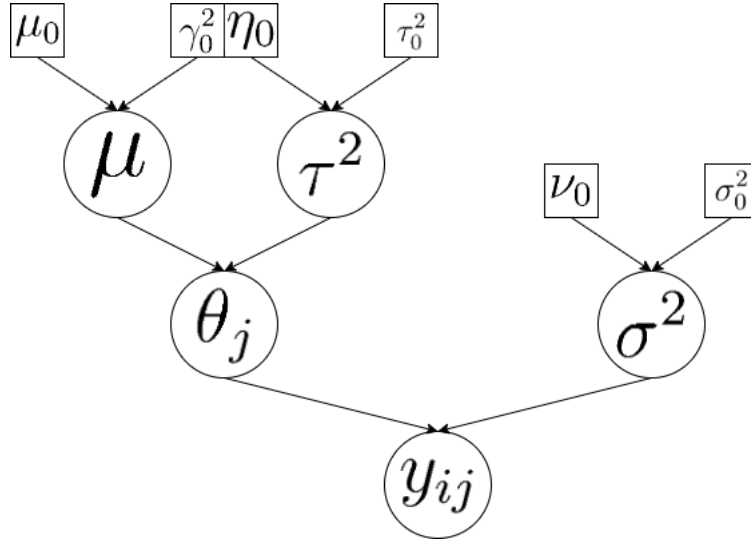


Figure 8: DAG de  $M_2$

$$\sigma^2 \mid \text{resto} \sim \text{Gl} \left( \frac{\nu_0 + n}{2}; \frac{\nu_0 \sigma_0^2 + \sum_{j=1}^m \sum_{i=1}^{n_j} (y_{ij} - \theta_j)^2}{2} \right)$$

### $M_3$ : Modelo Normal con medias y varianzas específicas

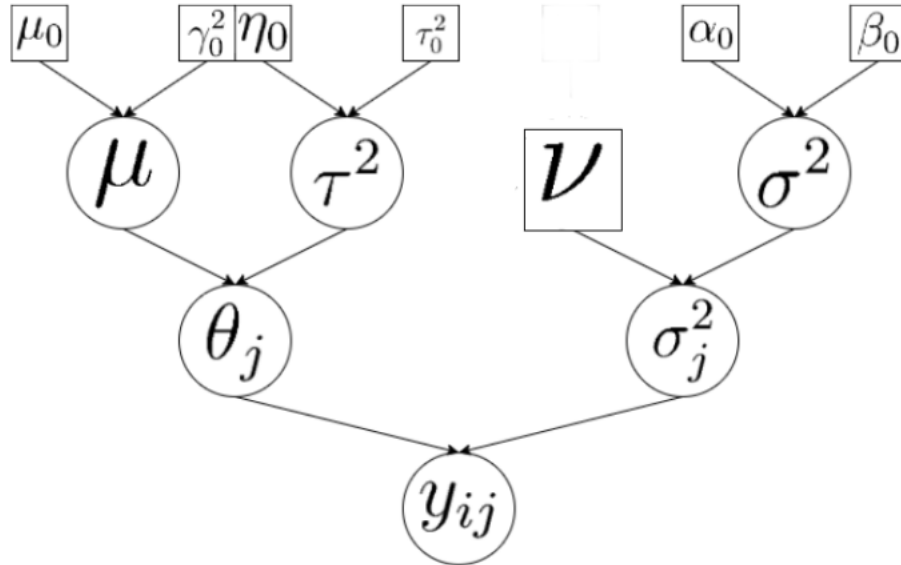


Figure 9: DAG de  $M_3$

$$\begin{aligned}
\mu \mid \text{resto} &\sim \text{N} \left( \frac{\frac{\mu_0}{\gamma_0^2} + \frac{m\bar{\theta}}{\tau^2}}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}}; \frac{1}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}} \right) \\
\tau^2 \mid \text{resto} &\sim \text{GI} \left( \frac{\eta_0 + m}{2}; \frac{\eta_0 \tau_0^2 + \sum_{j=1}^m (\theta_j - \mu)^2}{2} \right) \\
\theta_j \mid \text{resto} &\sim \text{N} \left( \frac{\frac{\mu}{\tau^2} + \frac{n_j \bar{y}_j}{\sigma_j^2}}{\frac{1}{\tau^2} + \frac{n_j}{\sigma_j^2}}; \frac{1}{\frac{1}{\tau^2} + \frac{n_j}{\sigma_j^2}} \right) \\
\sigma^2 \mid \text{resto} &\sim \text{Gamma} \left( \frac{\alpha_0 + m\nu}{2}; \frac{\beta_0}{2} + \frac{\nu}{2} \sum_{j=1}^m \frac{1}{\sigma_j^2} \right) \\
\sigma_j^2 \mid \text{resto} &\sim \text{GI} \left( \frac{\nu + n_j}{2}; \frac{\nu \sigma^2 + \sum_{i=1}^{n_j} (y_{ij} - \theta_j)^2}{2} \right)
\end{aligned}$$

**M<sub>4</sub>: Modelo t**

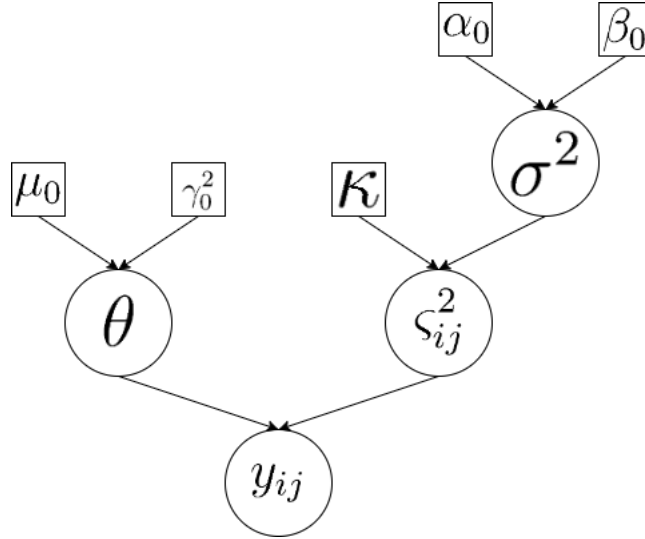


Figure 10: DAG de  $M_4$

$$\theta \mid \text{resto} \sim \text{N} \left( \frac{\frac{\mu_0}{\gamma_0^2} + \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{y_{ij}}{\varsigma_{ij}^2}}{\frac{1}{\gamma_0^2} + \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2}}; \frac{1}{\frac{1}{\gamma_0^2} + \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2}} \right)$$

$$\sigma^2 \mid \text{resto} \sim \text{Gamma} \left( \frac{\alpha_0 + n\kappa}{2}; \frac{\beta_0}{2} + \frac{\kappa}{2} \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2} \right)$$

$$\varsigma_{ij}^2 \mid \text{resto} \sim \text{Gl} \left( \frac{\kappa + 1}{2}; \frac{\kappa\sigma^2 + (y_{ij} - \theta)^2}{2} \right)$$

### M<sub>5</sub>: Modelo t con medias específicas

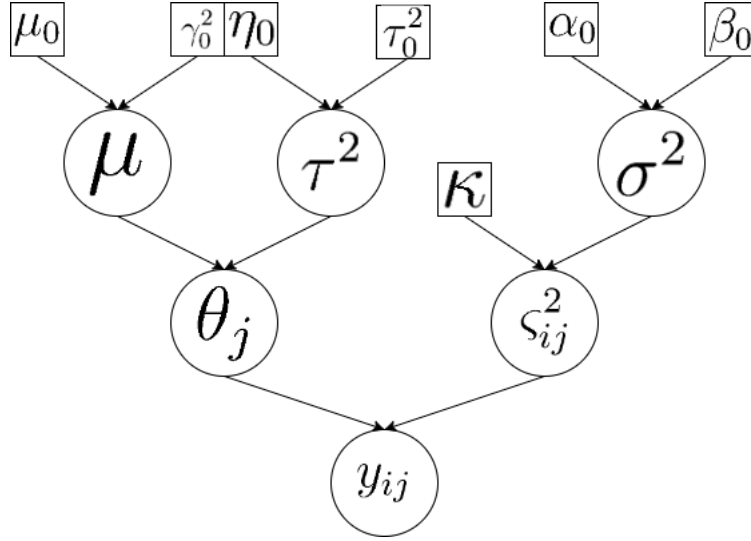


Figure 11: DAG de M<sub>5</sub>

$$\mu \mid \text{resto} \sim \text{N} \left( \frac{\frac{\mu_0}{\gamma_0^2} + \frac{m\bar{\theta}}{\tau^2}}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}}; \frac{1}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}} \right)$$

$$\tau^2 \mid \text{resto} \sim \text{Gl} \left( \frac{\eta_0 + m}{2}; \frac{\eta_0\tau_0^2 + \sum_{j=1}^m (\theta_j - \mu)^2}{2} \right)$$

$$\theta_j \mid \text{resto} \sim \text{N} \left( \frac{\frac{\mu}{\tau^2} + \sum_{i=1}^{n_j} \frac{y_{ij}}{\varsigma_{ij}^2}}{\frac{1}{\tau^2} + \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2}}; \frac{1}{\frac{1}{\tau^2} + \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2}} \right)$$

$$\sigma^2 \mid \text{resto} \sim \text{Gamma} \left( \frac{\alpha_0 + n\kappa}{2}; \frac{\beta_0}{2} + \frac{\kappa}{2} \sum_{j=1}^m \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2} \right)$$

$$\varsigma_{ij}^2 \mid \text{resto} \sim \text{Gl} \left( \frac{\kappa + 1}{2}; \frac{\kappa\sigma^2 + (y_{ij} - \theta_j)^2}{2} \right)$$

**M<sub>6</sub>: Modelo t con medias y varianzas específicas**

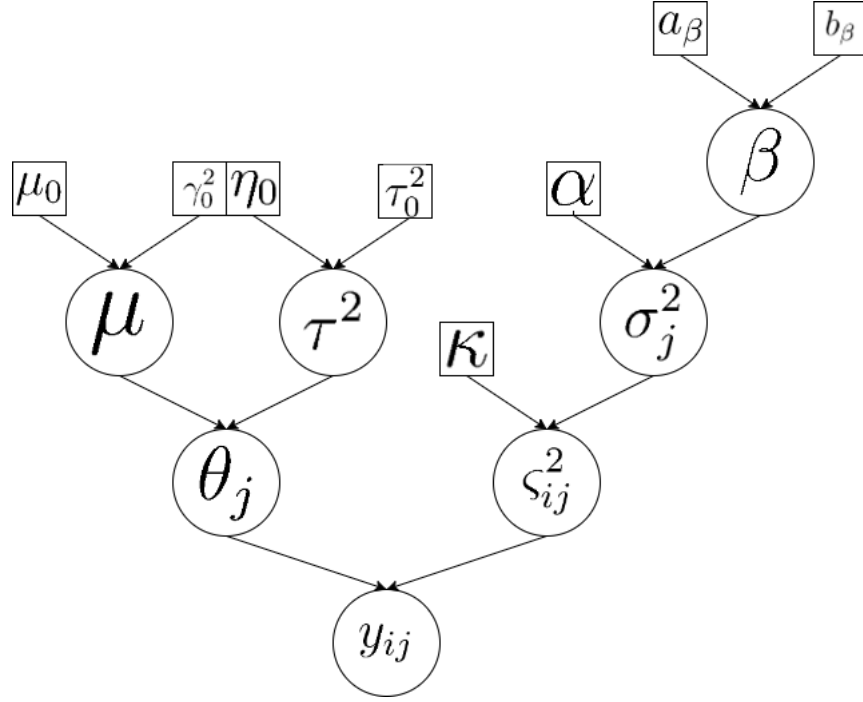


Figure 12: DAG de  $M_6$

$$\begin{aligned}
 \mu \mid \text{resto} &\sim \text{N} \left( \frac{\frac{\mu_0}{\gamma_0^2} + \frac{m\bar{\theta}}{\tau^2}}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}}; \frac{1}{\frac{1}{\gamma_0^2} + \frac{m}{\tau^2}} \right) \\
 \tau^2 \mid \text{resto} &\sim \text{Gl} \left( \frac{\eta_0 + m}{2}; \frac{\eta_0 \tau_0^2 + \sum_{j=1}^m (\theta_j - \mu)^2}{2} \right) \\
 \theta_j \mid \text{resto} &\sim \text{N} \left( \frac{\frac{\mu}{\tau^2} + \sum_{i=1}^{n_j} \frac{y_{ij}}{\varsigma_{ij}^2}}{\frac{1}{\tau^2} + \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2}}; \frac{1}{\frac{1}{\tau^2} + \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2}} \right) \\
 \beta \mid \text{resto} &\sim \text{Gamma} \left( \frac{a_\beta + m\alpha}{2}; \frac{b_\beta}{2} + \frac{1}{2} \sum_{j=1}^m \sigma_j^2 \right) \\
 \sigma_j^2 \mid \text{resto} &\sim \text{Gamma} \left( \frac{\alpha + n_j \kappa}{2}; \frac{\beta}{2} + \frac{\kappa}{2} \sum_{i=1}^{n_j} \frac{1}{\varsigma_{ij}^2} \right) \\
 \varsigma_{ij}^2 \mid \text{resto} &\sim \text{Gl} \left( \frac{\kappa + 1}{2}; \frac{\kappa \sigma_j^2 + (y_{ij} - \theta_j)^2}{2} \right)
 \end{aligned}$$

## Referencias