

Bayesian Methods Final Project

SESGO DE GÉNERO EN EL RENDIMIENTO ACADÉMICO EN LA MATERIA DE  
MATEMÁTICAS

Aráiztegui, Aránzazu  
Ferrara, Lorenzo  
Lucchini, Marco

04 January, 2023



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat de Matemàtiques i Estadística



UNIVERSITAT DE  
BARCELONA

# Índice

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Descripción del problema . . . . .	1
1.2	Objetivos del modelo . . . . .	1
<b>2</b>	<b>Descripción de la base de datos</b>	<b>2</b>
2.1	Definición de las variables utilizadas . . . . .	2
2.2	Análisis exploratorio de los datos . . . . .	2
<b>3</b>	<b>Análisis bayesiano</b>	<b>5</b>
3.1	Modelo 1 . . . . .	5
3.1.1	Presentacion del modelo . . . . .	5
3.1.2	Justificación e interpretación de los modelos . . . . .	6
3.1.3	Validación de los modelos . . . . .	9
3.2	Modelo 2 sin areas territoriales, con interaccion sexo:lenguas . . . . .	10
3.2.1	Presentacion de los modelos 2 . . . . .	10
3.2.2	Justificación e interpretación de los modelos 2 . . . . .	10
3.2.3	Validación de los modelos 2 . . . . .	10
3.3	Modelo 3: quito “ciudad” . . . . .	12
3.3.1	Validación de los modelos . . . . .	15
<b>4</b>	<b>Implementación del modelo</b>	<b>17</b>
<b>5</b>	<b>Conclusions</b>	<b>18</b>
<b>6</b>	<b>Apéndice</b>	<b>19</b>
6.1	Código R . . . . .	19
6.2	Tablas parámetros estimados . . . . .	19
<b>7</b>	<b>Referencias</b>	<b>20</b>

# 1 Introducció

El sesgo de género está presente en numerosos ámbitos de nuestra vida y se aprecia de forma notable en el ámbito educativo, más concretamente en las materias del ámbito STEAM.

Parece ser que aunque no hay una evidente brecha de género en el comienzo de los estudios primarios, el sesgo de género comienza a aparecer de forma clara a lo largo del proceso educativo, agudizándose en las últimas etapas de la educación obligatoria y evidenciándose de forma clara en la educación universitaria.

## 1.1 Descripción del problema

Con el objetivo de evaluar la capacidad y el nivel de competencia en las diferentes áreas del conocimiento que tiene el alumnado de Cataluña, el Departament d'Educació de la Generalitat de Cataluña realiza una prueba de competencias y conocimientos básicos en las áreas lingüísticas, matemáticas y científico-tecnológicas en los últimos cursos de la educación primaria y secundaria.

Según el Departament se trata de una evaluación de carácter formativo y orientador que pueda servir tanto a los centros como al profesorado y al propio Departament para impulsar las mejoras en el sistema educativo catalán.

## 1.2 Objetivos del modelo

Este trabajo tiene como objetivo principal ver si existe un sesgo de género en los resultados obtenidos en la competencia matemática con respecto al sexo y a las competencias humanísticas para ello intentaremos crear un modelo que relacione la puntuación obtenida en la competencia matemática con respecto al sexo, a las competencias lingüísticas e incluso con respecto al tipo de centro educativo o al tamaño de la población. De esta manera podríamos ver si en el sexo femenino no se da una diferencia entre el rendimiento en humanidades y matemáticas, personas con bajo rendimiento en humanidades también lo tendrían en matemáticas.

Y en cambio en el sexo masculino personas con bajo rendimiento en humanidades tendrían buenos resultados en matemáticas.

Si ampliamos los datos y vemos la tabla de resultados para un mismo individuo en sexto de primaria y cuarto de la podríamos establecer un segundo objetivo que sería ver si se mantienen los resultados en ambos sexos o si hay diferencias significativas en cuanto al rendimiento en el área de matemáticas al aumentar la edad en relación al sexo.

## 2 Descripción de la base de datos

La base de datos utilizada es una mezcla de los datos ofrecidos para cuarto curso de ESO y los datos que se ofrecen para sexto curso de primaria. La razón de realizar la fusión de las dos tablas es poder evaluar la evolución de una misma persona desde el final de la educación primaria hasta el final de la educación secundaria. Se puede acceder a la base de datos completa en el siguiente enlace:

Avaluació de quart d'Educació Secundària Obligatòria | Dades obertes de Catalunya

El dataset contiene los resultados obtenidos por el alumnado de cuarto curso de ESO en la evaluación de competencias básicas al final de la educación secundaria desde el año 2012.

Se incluye un código de alumno para poder hacer comparativas con los resultados obtenidos en sextos de primaria. Dado que el código solo está disponible a partir del año 2016 se utilizarán únicamente los datos del alumnado a partir de este año.

La base de datos ha sido actualizada el 20 de octubre de 2022 y contiene los datos de 46384 estudiantes.

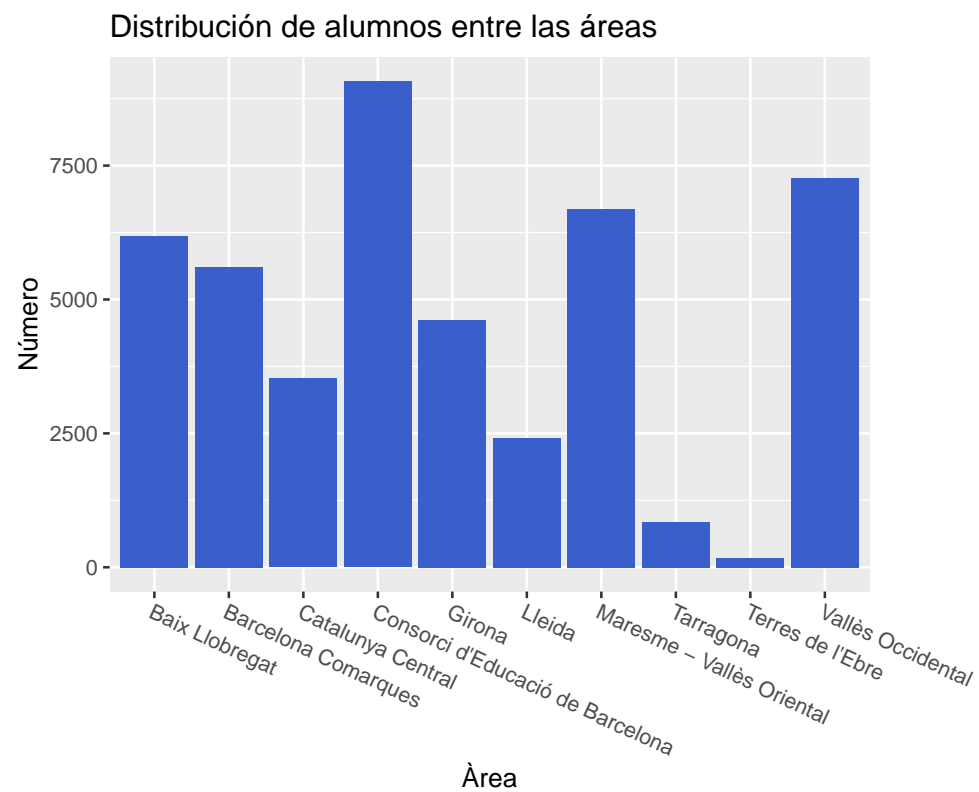
### 2.1 Definición de las variables utilizadas

Base de dades

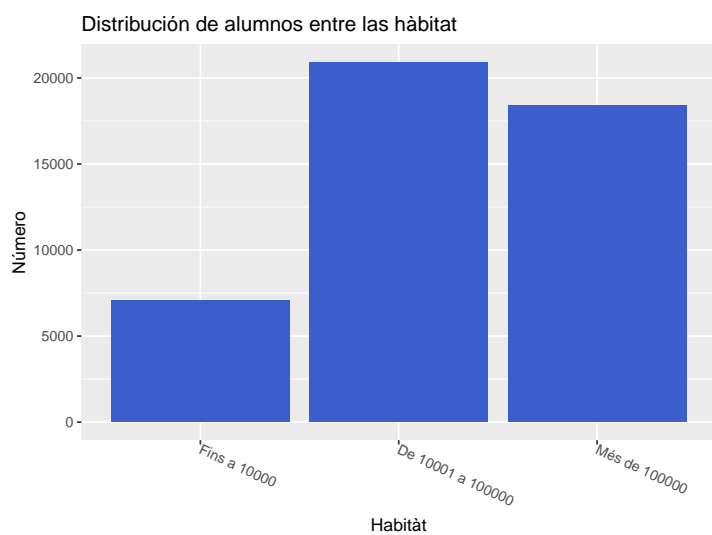
Nom de columna	Descripció	Tipus
PMAT_4	Puntuació global ponderada de competència matemàtica en el examen de Quart	Nombre
PMAT_6	Puntuació global ponderada de competència matemàtica en el examen de Sisè	Nombre
PLENG_4	Puntuació global ponderada de la competència lingüística en llengua catalana y castellana en el examen de Quart	Nombre
PLENG_6	Puntuació global ponderada de la competència lingüística en llengua catalana y castellana en el examen de Sisè	Nombre
PANG_4	Puntuació global ponderada de la competència lingüística en llengua anglesa en el examen de Quart	Nombre
PANG_6	Puntuació global ponderada de la competència lingüística en llengua anglesa en el examen de Sisè	Nombre
GENERE	Gènere de l'alumne/a que es presenta a l'avaluació	Text Pla
AREA_TERRITORIAL	Regió on es troba el centre de l'alumne/a que es presenta a l'avaluació	Text Pla
NATURALESA	Determina si el centre de l'alumne/a és públic, privat o concertat	Text Pla
HÀBITAT	Municipis per trams de població	Text Pla

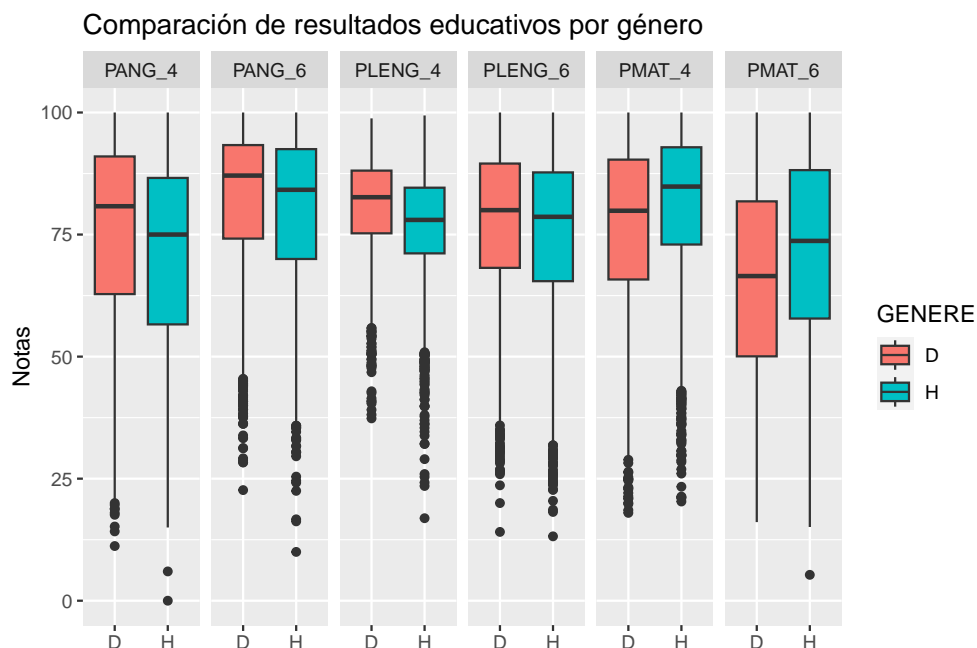
### 2.2 Análisis exploratorio de los datos

En el análisis inicial de los datos podemos ver que la distribución de hombres y mujeres es uniforme con una proporción de 50,1% niños y 49,9% niñas. La distribución de alumnos por áreas es la siguiente.

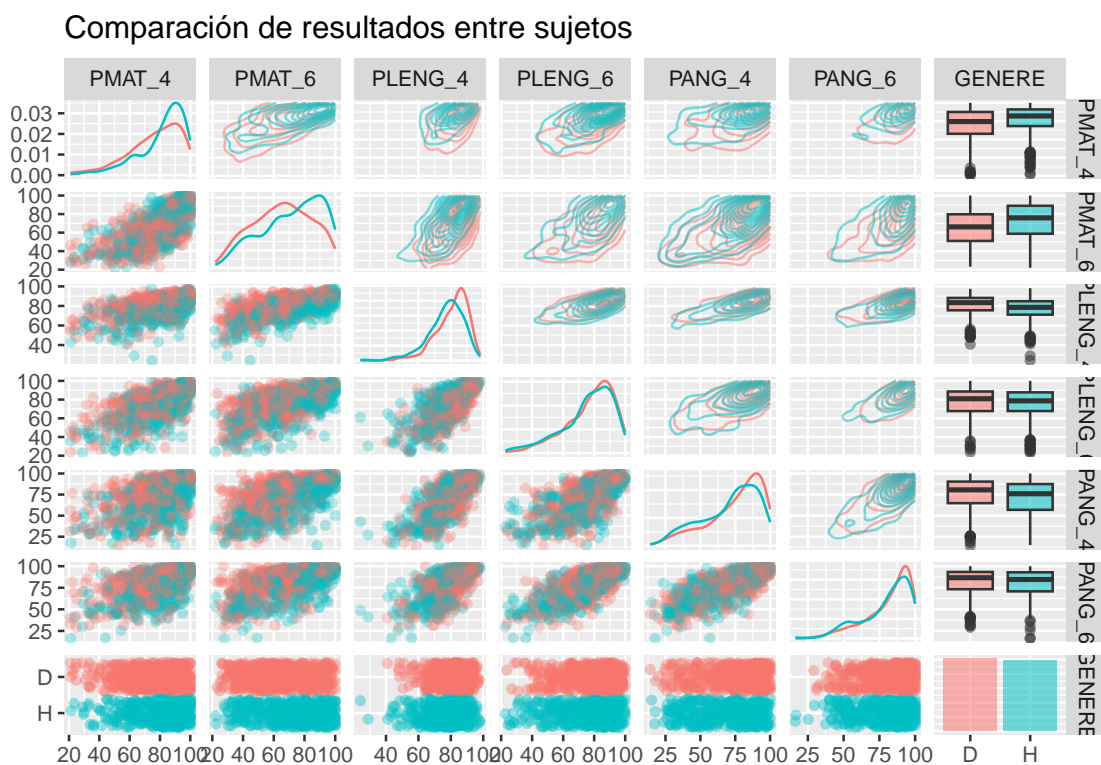


Según el tamaño de la población los datos de los que disponemos se concentran en individuos de ciudades de tamaño medio o grande.





En el siguiente gráfico podemos ver las diferencias entre las calificaciones en matemáticas, inglés y lengua de cada sexo dependiendo del nivel educativo. Si comparamos la misma asignatura en los dos cursos los datos muestran una distribución casi lineal y con poca variabilidad, lo cual nos hace pensar que el alumnado que obtiene buenos resultados en sexto de primaria también obtienen buenos resultados al acabar la secundaria, como cabría esperar. También cabe destacar la distribución de las frecuencias por género, en especial en el caso de matemáticas donde parece observarse una diferencia más evidente entre chicos y chicas.



## 3 Análisis bayesiano

### 3.1 Modelo 1

#### 3.1.1 Presentacion del modelo

In order to identify result which differ from the average behaviour in the result of the students in the Math exam of Quart we use a covariates the difference of the grades from the mean of each exam.

So  $\$ MAT\_4 = MAT\_4 - \text{mean}(MAT\_4)$   $\$$  and the same for the other variables

We build the following model

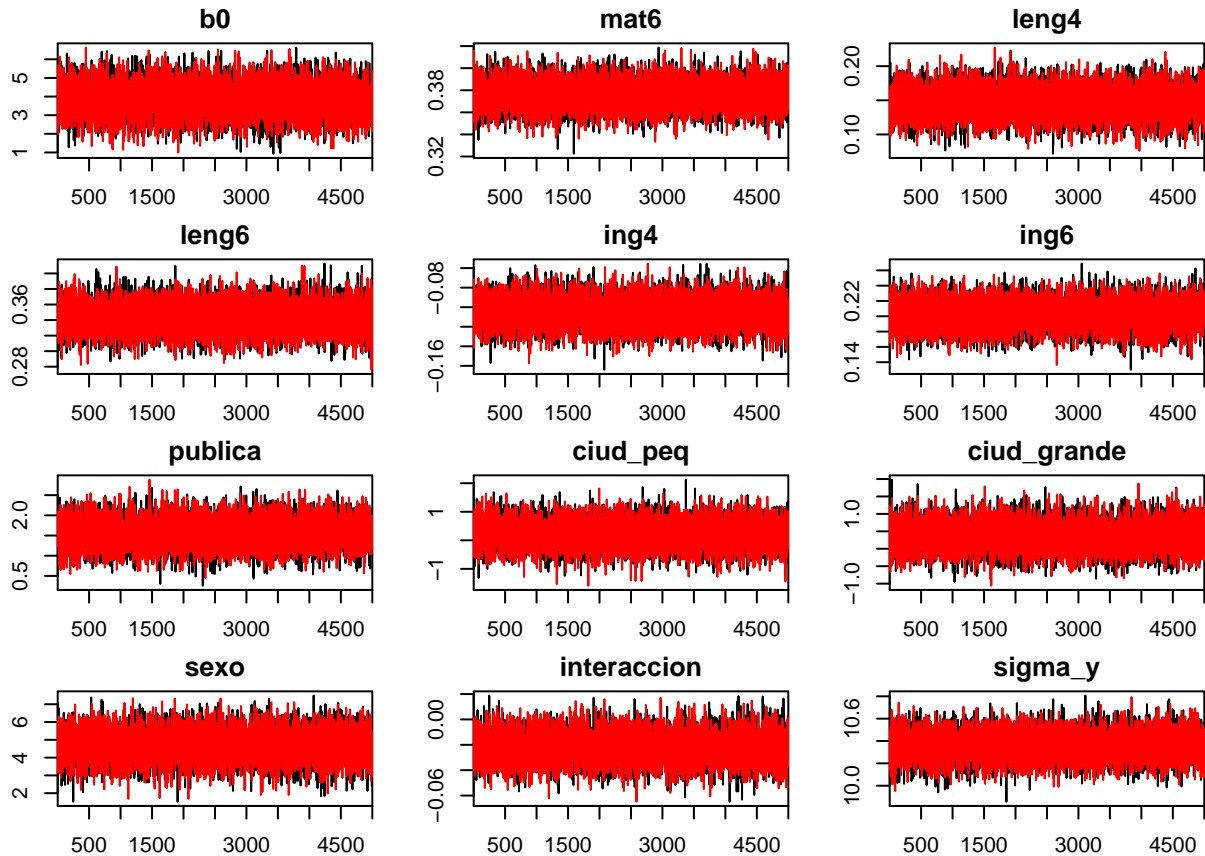
$$y_i \sim N(b_0 + mat_4 * x_i^1 + leng_4 * x_i^2 + leng_6 * x_i^3 + ing_4 * x_i^4 + ing_6 * x_i^5 + publica * x_i^6 + ciudPeq * x_i^7 + ciudGrande * x_i^8 + sexo * x_i^9 + interaccion * x_i^3 * x_i^9 + G[area[i]], tau_y)$$

We suppose the coefficient having the following distribution. We choose Standard normal distribution since the variables are bias from the mean of each exam.

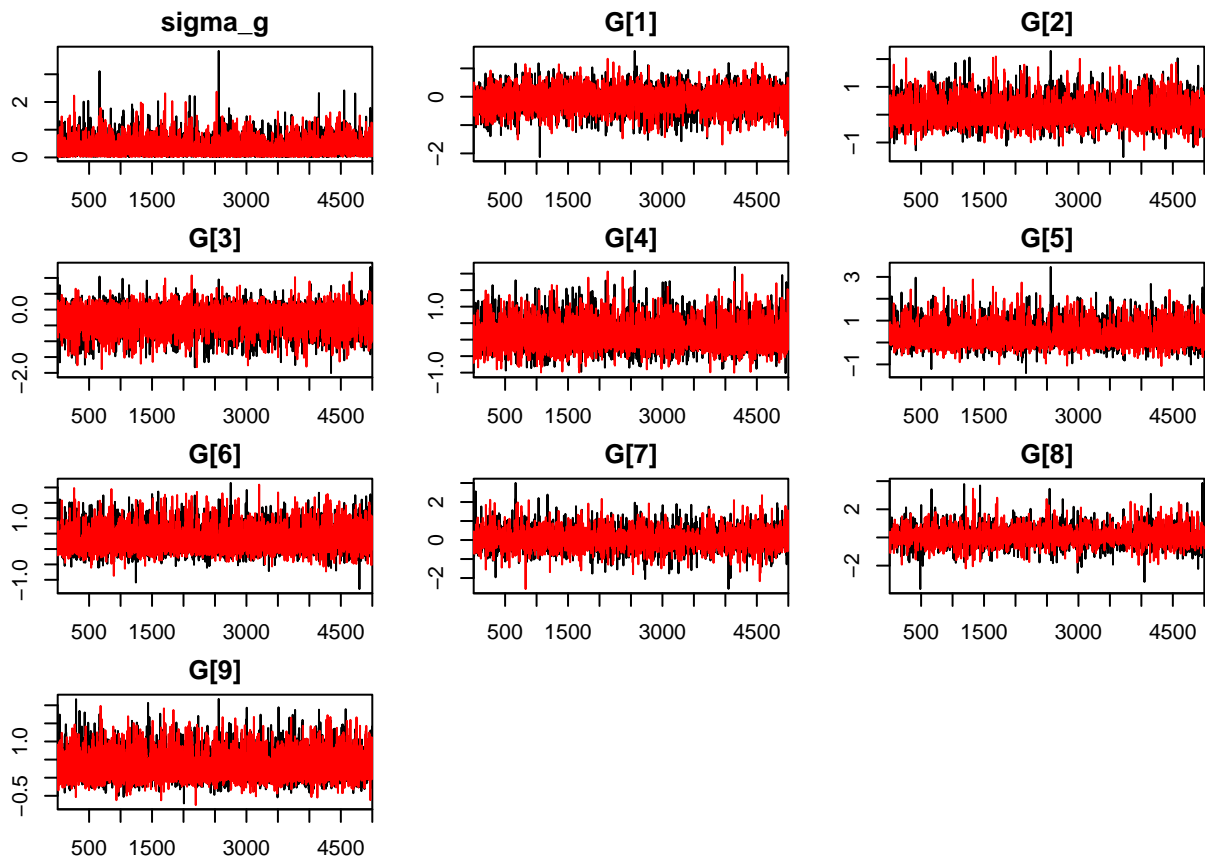
$$\begin{aligned} b_0 &\sim N(0, 1) \\ mat_4 &\sim N(0, 1) \\ leng_4 &\sim N(0, 1) \\ leng_6 &\sim N(0, 1) \\ ing_4 &\sim N(0, 1) \\ ing_6 &\sim N(0, 1) \\ publica &\sim N(0, 1) \\ ciudPeq &\sim N(0, 1) \\ ciudGrande &\sim N(0, 1) \\ sexo &\sim N(0, 1) \\ interaccion &\sim N(0, 1) \\ tau_y &\sim \Gamma(0.001, 0.001) \\ sigma_y &= \frac{1}{\sqrt{tau_y}} \\ G_i &\sim N(0, tau_g) \\ tau_g &\sim \Gamma(0.001, 0.001) \\ sigma_g &= \frac{1}{\sqrt{tau_g}} \end{aligned}$$

### 3.1.2 Justificación e interpretación de los modelos

```
## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 4000
##   Unobserved stochastic nodes: 22
##   Total graph size: 87242
##
## Initializing model
```







```
## Inference for Bugs model at "4", fit using jags,
## 2 chains, each with 5400 iterations (first 400 discarded)
## n.sims = 10000 iterations saved
##
```

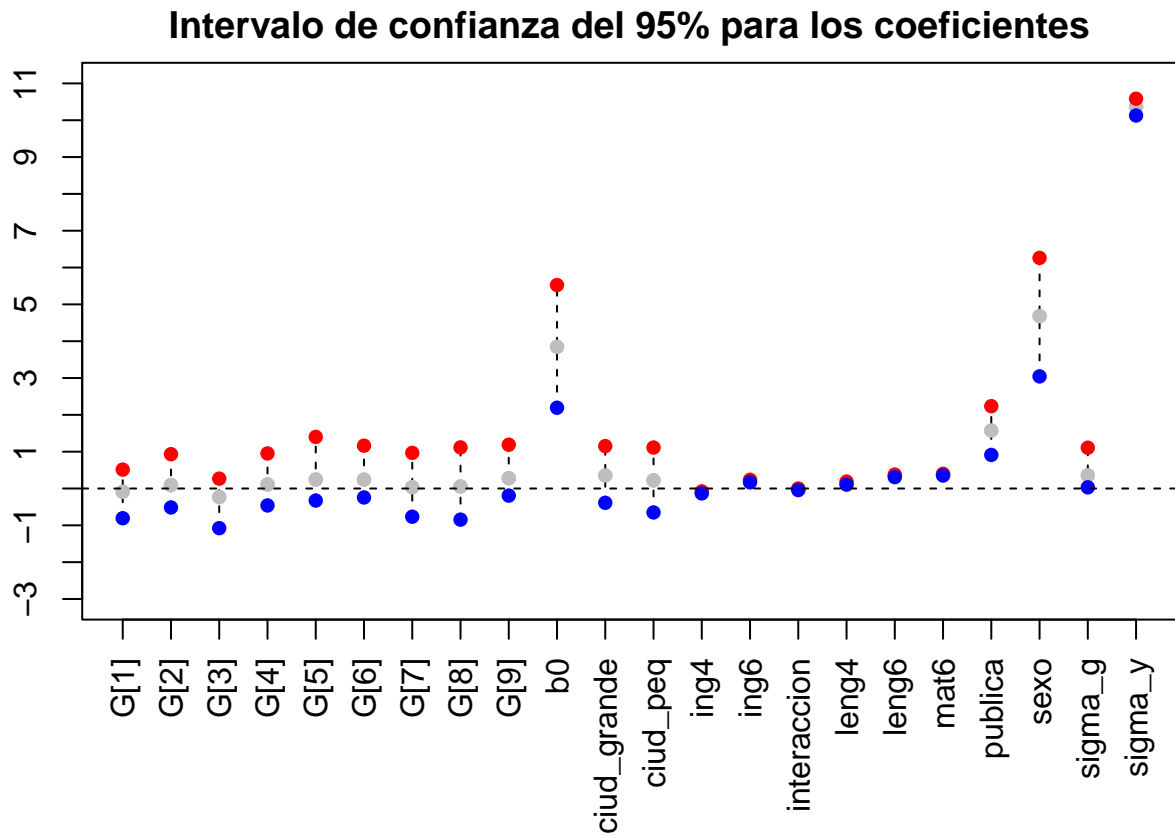
	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%	Rhat
## G[1]	-0.09	0.31	-0.81	-0.23	-0.04	0.06	0.51	1
## G[2]	0.09	0.34	-0.52	-0.07	0.03	0.22	0.93	1
## G[3]	-0.23	0.34	-1.08	-0.41	-0.14	0.00	0.27	1
## G[4]	0.11	0.33	-0.46	-0.06	0.04	0.24	0.95	1
## G[5]	0.24	0.43	-0.33	-0.02	0.10	0.42	1.40	1
## G[6]	0.24	0.36	-0.24	0.00	0.13	0.42	1.16	1
## G[7]	0.04	0.39	-0.77	-0.12	0.01	0.17	0.97	1
## G[8]	0.05	0.45	-0.85	-0.11	0.01	0.19	1.12	1
## G[9]	0.28	0.36	-0.19	0.02	0.17	0.47	1.19	1
## b0	3.85	0.84	2.19	3.28	3.84	4.41	5.52	1
## ciud_grande	0.35	0.39	-0.39	0.09	0.34	0.61	1.15	1
## ciud_peq	0.22	0.45	-0.65	-0.08	0.23	0.53	1.11	1
## ing4	-0.11	0.01	-0.13	-0.12	-0.11	-0.10	-0.08	1
## ing6	0.20	0.02	0.17	0.19	0.20	0.21	0.24	1
## interaccion	-0.02	0.01	-0.04	-0.03	-0.02	-0.01	0.00	1
## leng4	0.15	0.02	0.11	0.13	0.15	0.16	0.19	1
## leng6	0.34	0.02	0.31	0.33	0.34	0.36	0.38	1
## mat6	0.38	0.01	0.35	0.37	0.38	0.38	0.40	1
## publica	1.57	0.34	0.91	1.35	1.57	1.80	2.23	1
## sexo	4.68	0.82	3.04	4.14	4.69	5.23	6.26	1
## sigma_g	0.36	0.30	0.03	0.12	0.28	0.52	1.11	1
## sigma_y	10.36	0.12	10.13	10.28	10.36	10.43	10.58	1

```

## deviance      30052.62   10.33 30034.13 30045.37 30052.14 30059.21 30074.14   1
##              n.eff
## G[1]          1700
## G[2]          3900
## G[3]         10000
## G[4]         10000
## G[5]         10000
## G[6]         10000
## G[7]         10000
## G[8]          5600
## G[9]          1700
## b0           10000
## ciud_grande  10000
## ciud_peq    10000
## ing4        10000
## ing6         5500
## interaccion 10000
## leng4       10000
## leng6       10000
## mat6        10000
## publica     1500
## sexo        10000
## sigma_g     1100
## sigma_y     4100
## deviance    10000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 53.3 and DIC = 30106.0
## DIC is an estimate of expected predictive error (lower deviance is better).

##          G[1]          G[2]          G[3]          G[4]          G[5]          G[6]
##          FALSE          FALSE          FALSE          FALSE          FALSE          FALSE
##          G[7]          G[8]          G[9]          b0 ciud_grande  ciud_peq
##          FALSE          FALSE          FALSE          TRUE          FALSE          FALSE
## deviance          ing4          ing6 interaccion          leng4          leng6
##          TRUE          TRUE          TRUE          TRUE          TRUE          TRUE
##          mat6      publica      sexo      sigma_g      sigma_y
##          TRUE          TRUE          TRUE          TRUE          TRUE

```

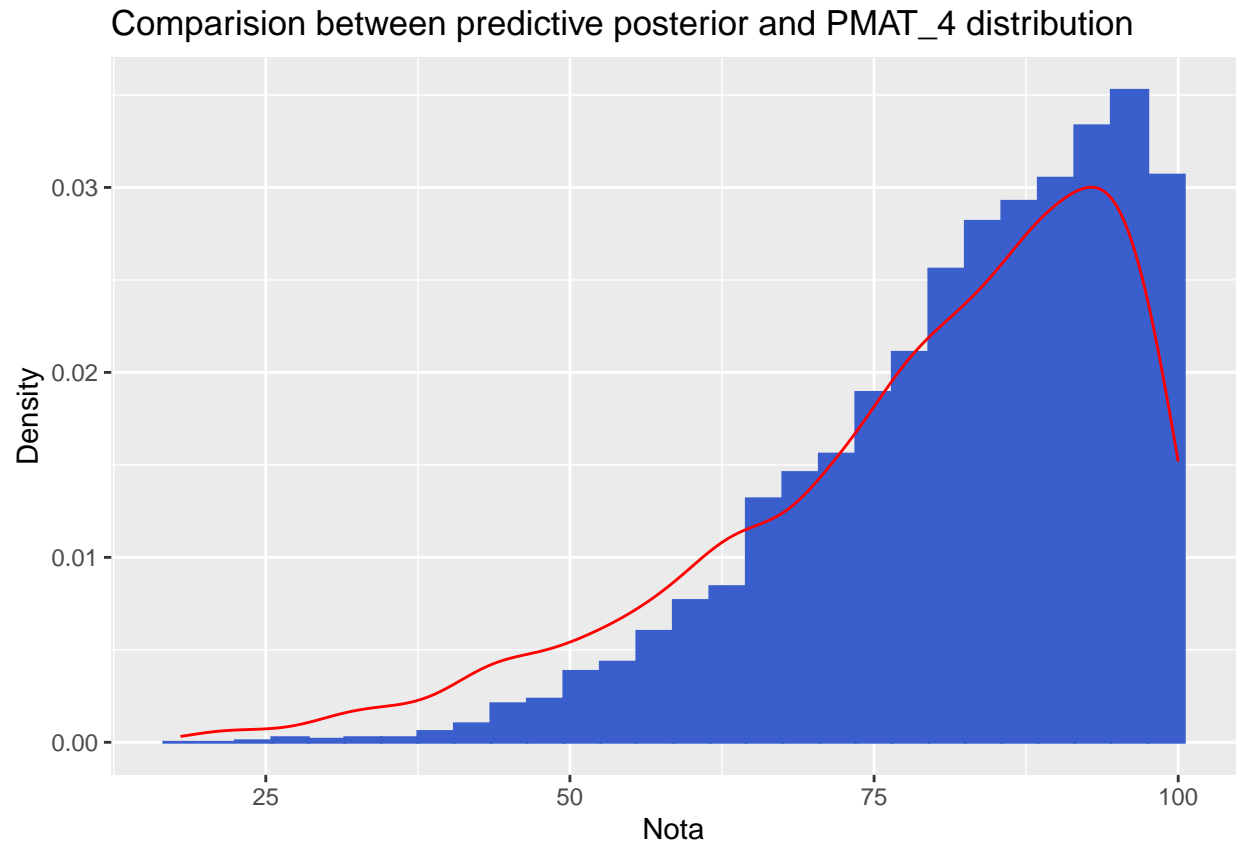


### 3.1.3 Validación de los modelos

We now predict the values of PMAT\_4 for the students in the test set and compare the results.

```
##          5%          95%
## 56.72235 98.58773
```

Using Truncated Noemal 0-100 Blue predicted, Red Dataset



This is not the right way to validate, will fix this part in the next version

Seem to be good but we remove the area effect

## 3.2 Modelo 2 sin areas territoriales, con interaccion sexo:lenguas

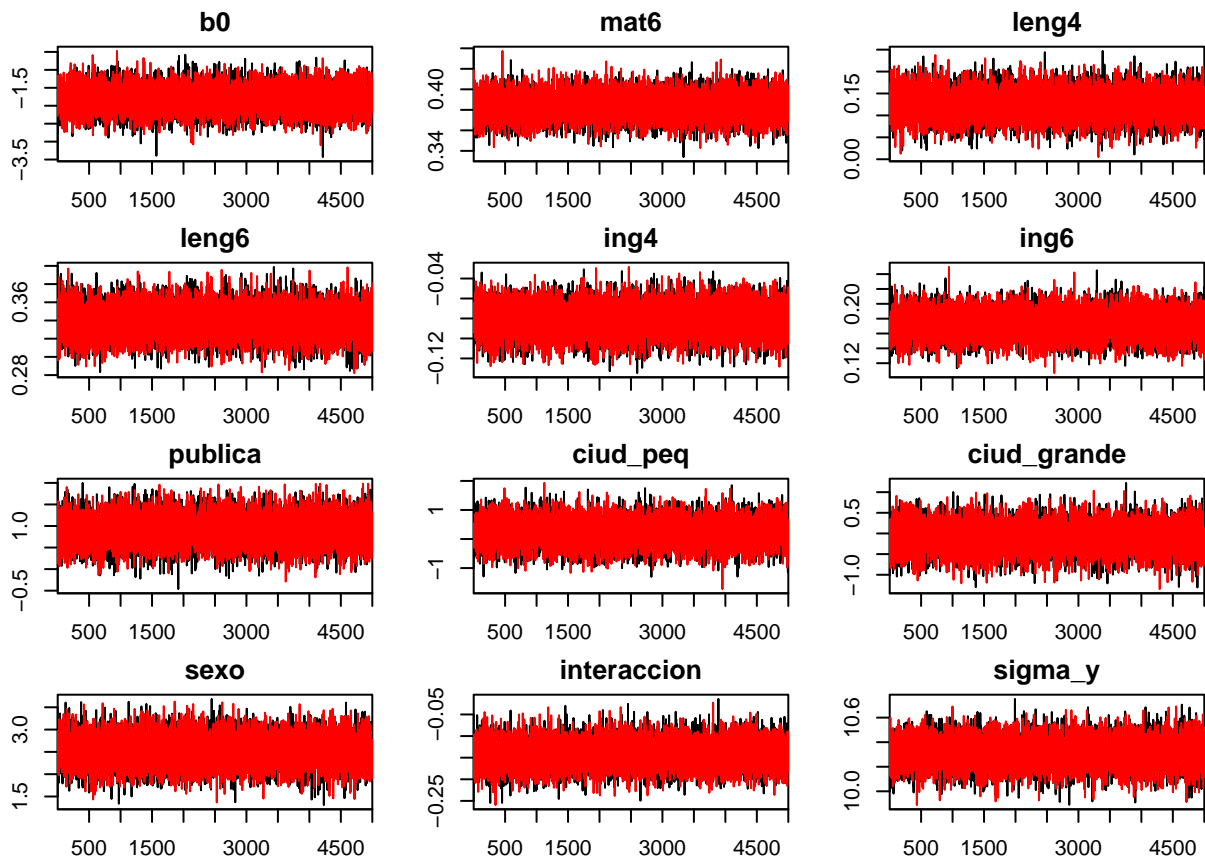
### 3.2.1 Presentacion de los modelos 2

### 3.2.2 Justificación e interpretación de los modelos 2

### 3.2.3 Validación de los modelos 2

```
## module glm loaded

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 4000
##   Unobserved stochastic nodes: 12
##   Total graph size: 48002
##
## Initializing model
```



```
## Inference for Bugs model at "4", fit using jags,
## 2 chains, each with 5400 iterations (first 400 discarded)
## n.sims = 10000 iterations saved
##
```

	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%	Rhat
## b0	-1.82	0.35	-2.49	-2.05	-1.82	-1.59	-1.15	1
## ciud_grande	-0.12	0.33	-0.77	-0.34	-0.13	0.11	0.53	1
## ciud_peq	0.25	0.44	-0.64	-0.05	0.26	0.55	1.10	1
## ing4	-0.08	0.01	-0.11	-0.09	-0.08	-0.07	-0.05	1
## ing6	0.17	0.02	0.14	0.16	0.17	0.18	0.21	1
## interaccion	-0.14	0.03	-0.20	-0.16	-0.14	-0.12	-0.08	1
## leng4	0.13	0.03	0.06	0.10	0.13	0.15	0.19	1
## leng6	0.34	0.02	0.31	0.33	0.34	0.35	0.37	1
## mat6	0.38	0.01	0.36	0.38	0.38	0.39	0.41	1
## publica	0.90	0.33	0.25	0.68	0.90	1.13	1.56	1
## sexo	2.53	0.33	1.88	2.30	2.52	2.75	3.18	1
## sigma_y	10.31	0.11	10.09	10.23	10.31	10.39	10.53	1
## deviance	30017.91	5.36	30009.50	30014.03	30017.17	30021.10	30030.12	1

```
##
```

	n.eff
## b0	8500
## ciud_grande	10000
## ciud_peq	10000
## ing4	7600
## ing6	10000
## interaccion	4100
## leng4	10000
## leng6	10000

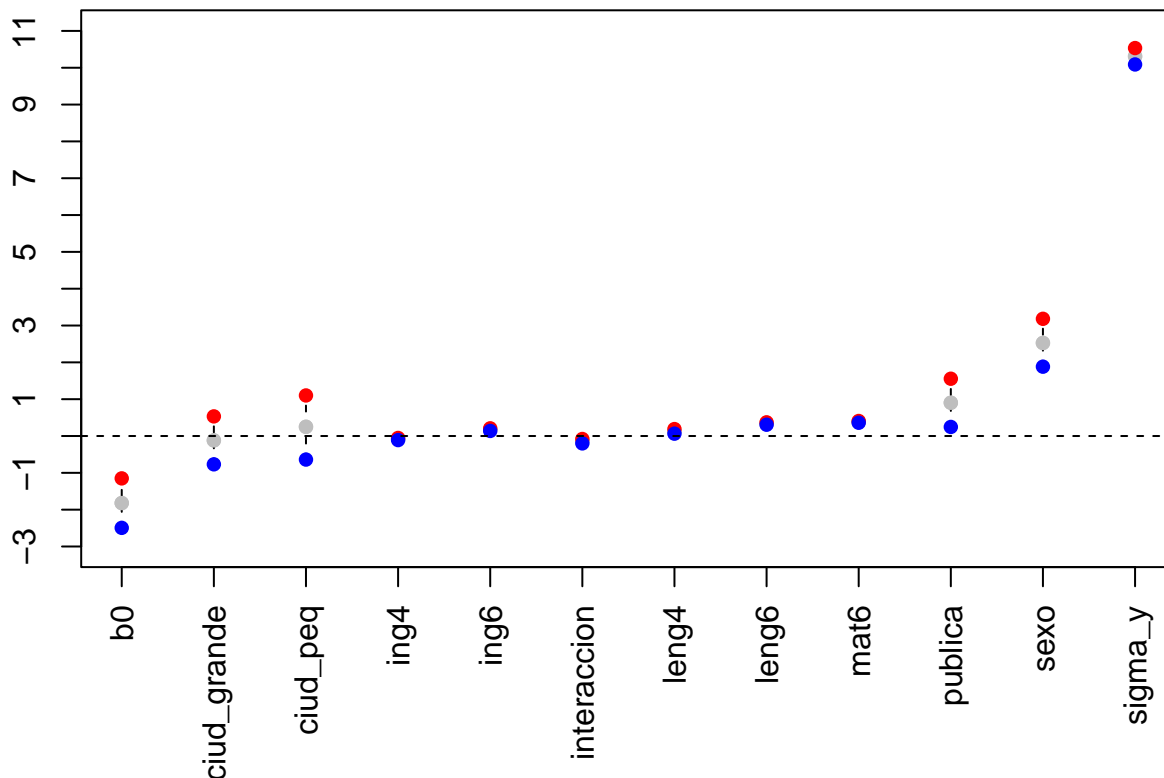
```

## mat6          2800
## publica       7700
## sexo         10000
## sigma_y       10000
## deviance      10000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 14.4 and DIC = 30032.3
## DIC is an estimate of expected predictive error (lower deviance is better).

##          b0 ciud_grande   ciud_peq   deviance      ing4      ing6
##          TRUE      FALSE      FALSE      TRUE      TRUE      TRUE
## interaccion    leng4    leng6    mat6    publica    sexo
##          TRUE      TRUE      TRUE      TRUE      TRUE      TRUE
##      sigma_y
##          TRUE

```

### Intervalo de confianza del 95% para los coeficientes



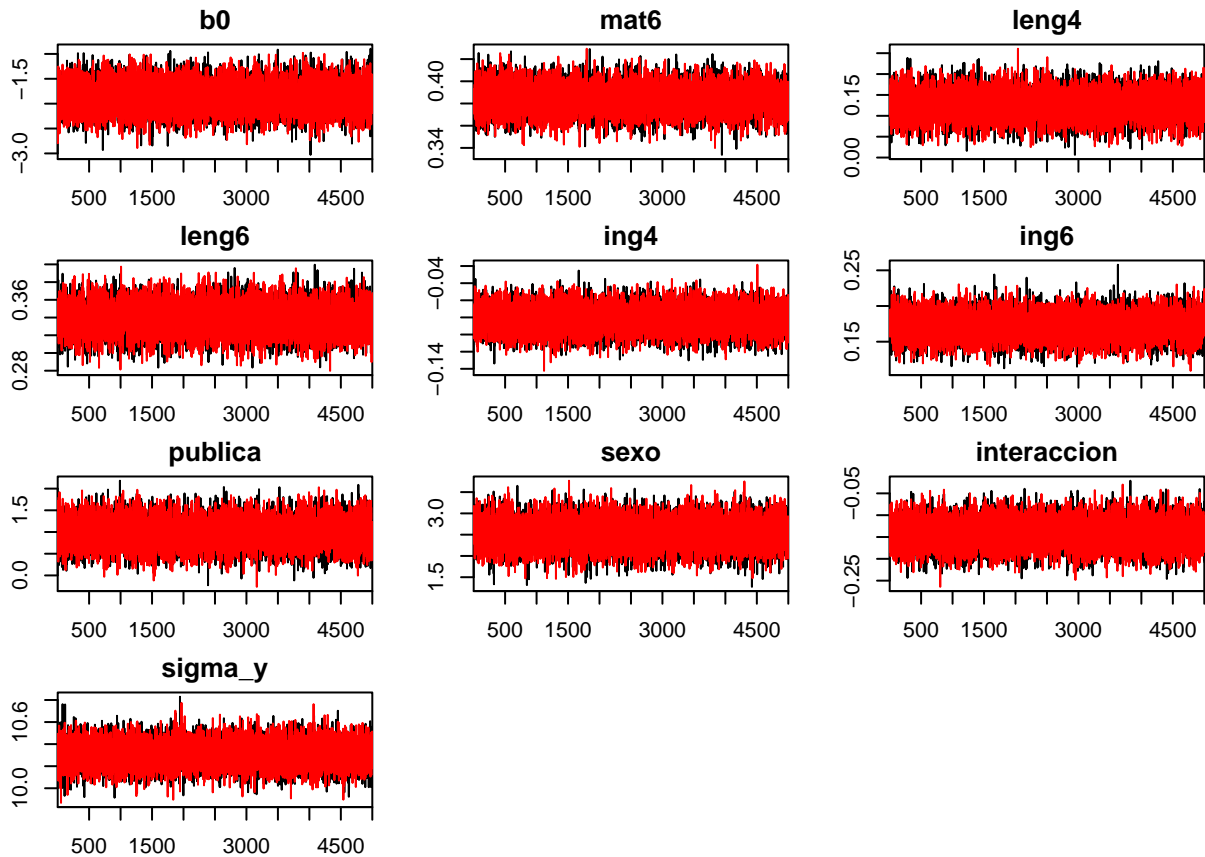
Missing validation

Remove non significative covariates

### 3.3 Modelo 3: quito “ciudad”

```
## Compiling model graph
```

```
## Resolving undeclared variables
## Allocating nodes
## Graph information:
## Observed stochastic nodes: 4000
## Unobserved stochastic nodes: 10
## Total graph size: 39996
##
## Initializing model
```



```
## Inference for Bugs model at "5", fit using jags,
## 2 chains, each with 5400 iterations (first 400 discarded)
## n.sims = 10000 iterations saved
##
```

	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%	Rhat
b0	-1.87	0.29	-2.44	-2.07	-1.86	-1.67	-1.29	1
ing4	-0.08	0.01	-0.11	-0.09	-0.08	-0.07	-0.05	1
ing6	0.17	0.02	0.14	0.16	0.17	0.19	0.21	1
interaccion	-0.14	0.03	-0.20	-0.16	-0.14	-0.12	-0.08	1
leng4	0.13	0.03	0.06	0.10	0.13	0.15	0.19	1
leng6	0.34	0.02	0.31	0.33	0.34	0.35	0.37	1
mat6	0.38	0.01	0.36	0.38	0.38	0.39	0.41	1
publica	0.97	0.32	0.35	0.75	0.97	1.19	1.62	1
sexo	2.52	0.33	1.88	2.30	2.53	2.74	3.17	1
sigma_y	10.31	0.11	10.09	10.23	10.31	10.39	10.54	1
deviance	30016.65	5.09	30008.80	30012.94	30015.92	30019.58	30028.70	1
n.eff								

```
##
```

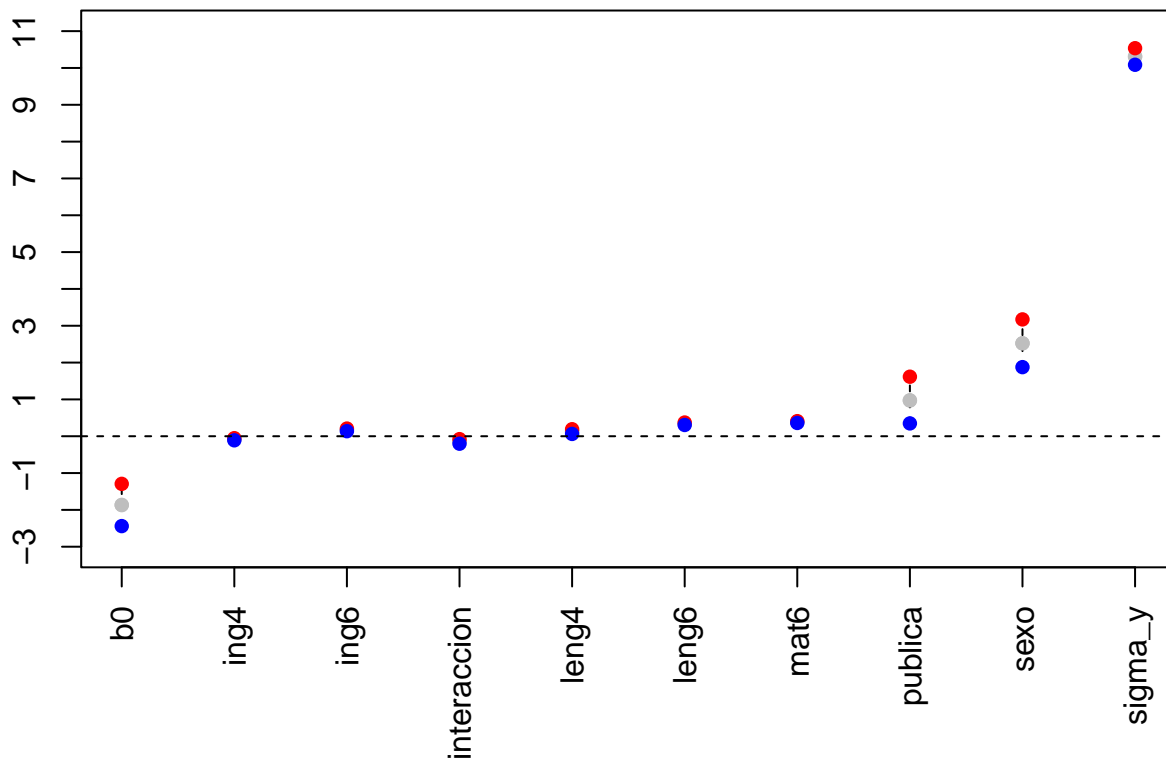
```

## b0          3900
## ing4        9500
## ing6       10000
## interaccion 10000
## leng4       10000
## leng6       10000
## mat6        6200
## publica    10000
## sexo        8500
## sigma_y     10000
## deviance    10000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 13.0 and DIC = 30029.6
## DIC is an estimate of expected predictive error (lower deviance is better).

##          b0      deviance      ing4      ing6 interaccion      leng4
##        TRUE      TRUE      TRUE      TRUE      TRUE      TRUE
##      leng6      mat6      publica      sexo      sigma_y
##        TRUE      TRUE      TRUE      TRUE      TRUE

```

### Intervalo de confianza del 95% para los coeficientes



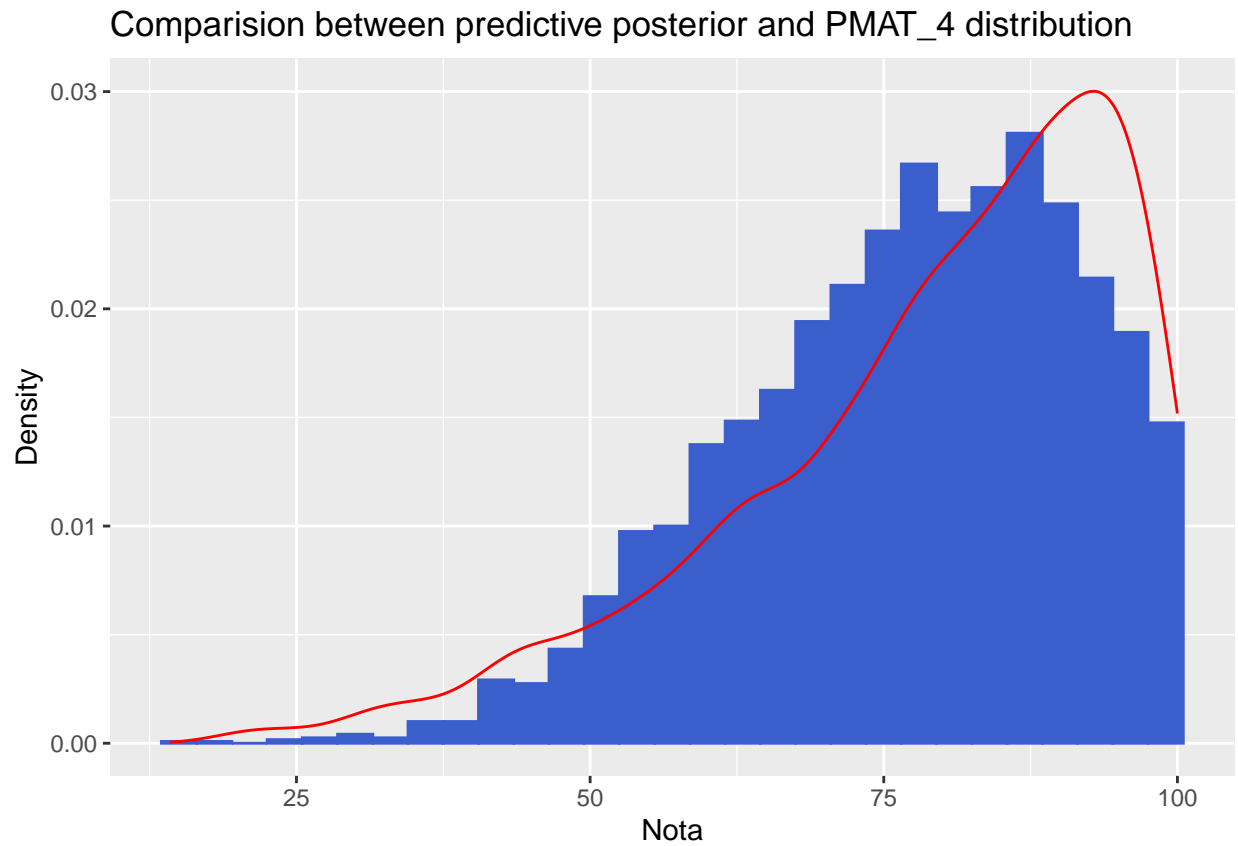


### 3.3.1 Validación de los modelos

We now predict the values of PMAT\_4 for the students in the test set and compare the results.

```
##          5%          95%  
## 51.25326 97.31034
```

Using Truncated Noemal 0-100 Blue predicted, Red Dataset



Quick explanation for us:

```
## Inference for Bugs model at "5", fit using jags,
## 2 chains, each with 5400 iterations (first 400 discarded)
## n.sims = 10000 iterations saved
##      mu.vect sd.vect      2.5%      25%      50%      75%      97.5% Rhat
## b0      -1.87   0.29     -2.44     -2.07     -1.86     -1.67     -1.29    1
## ing4     -0.08   0.01     -0.11     -0.09     -0.08     -0.07     -0.05    1
## ing6      0.17   0.02      0.14      0.16      0.17      0.19      0.21    1
## interaccion -0.14   0.03     -0.20     -0.16     -0.14     -0.12     -0.08    1
## leng4      0.13   0.03      0.06      0.10      0.13      0.15      0.19    1
## leng6      0.34   0.02      0.31      0.33      0.34      0.35      0.37    1
## mat6      0.38   0.01      0.36      0.38      0.38      0.39      0.41    1
## publica     0.97   0.32      0.35      0.75      0.97      1.19      1.62    1
## sexo       2.52   0.33      1.88      2.30      2.53      2.74      3.17    1
## sigma_y    10.31  0.11     10.09     10.23     10.31     10.39     10.54    1
## deviance 30016.65  5.09 30008.80 30012.94 30015.92 30019.58 30028.70  1
##      n.eff
## b0      3900
## ing4     9500
## ing6    10000
## interaccion 10000
## leng4    10000
## leng6    10000
## mat6      6200
## publica   10000
## sexo      8500
## sigma_y   10000
## deviance  10000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 13.0 and DIC = 30029.6
## DIC is an estimate of expected predictive error (lower deviance is better).
```

The last model is the one that is chosen. From this model we can answer to our objective of the project:

**All this conclusion should be explained in terms of difference with the mean result since our covariates refer to the difference with the mean**

- 1 Is gender a relevant factor in the result of math?

Yes it is, we have to do some test to confirm this but the coeff of sexo show that boys should get on average 2.52 point more than girls in the exam

- 2 Is it true that girls are coherent in the grades of leng and math while boys tend to have the opposit behaviour?

Yes, the coeff of leng4 show that for girls the result is the same for both leng and math. While if for boys the interaction reduce this coefficient and gets negative ( $0.13 - 0.14 = -0.01$ ) so for boys the result of math is inverse corelated with the result of leng.

## 4 Implementación del modelo

## 5 Conclusions

## 6 Apéndice

### 6.1 Código R

### 6.2 Tablas parámetros estimados

## 7 Referencias