

---

# Ausentismo de funcionarios en las casillas electorales

## Reporte EST-46114

**Pedro Vladimir Hernández Serrano & José Alfredo Méndez Barrera** ITAM

---

26 de mayo de 2016

En el presente documento se revisan los diferentes factores que pueden influenciar el ausentismo de funcionarios de casillas en los comicios electorales, lo cual es una problemática importante en el desarrollo de la operación durante las jornadas electorales. El propósito es desarrollar un procedimiento que utilice consideraciones estocásticas para seleccionar un conjunto de regresores con ayuda de simulaciones y de esta manera encontrar modelos prometedores. Se revisará el método de Selección Estocástica de Variables, la clave para el potencial de SSVS es la simulación rápida y eficiente del sistema de muestreo de Gibbs. Posteriormente se ajusta un modelo de Regresión Lineal Generalizado a la información de CDMX, los datos descriptivos utilizados son al corte final de las elecciones del 2015 en su acumulado nacional.

### 1. Introducción

El Instituto Nacional Electoral es el encargado de organizar las elecciones en nuestro país. Éstas son la culminación de todo un proceso electoral que inicia el año anterior y concluye una vez que los resultados de las contiendas son ratificados.

A raíz de la última reforma electoral se estipuló que en las entidades donde se elijan tanto puestos fede-

rales como locales, las casillas se integrarán de seis funcionarios y en los casos donde se elijan únicamente puestos de ámbito federal serán cuatro. El conseguir que cada una de estas personas lleguen a ocupar su puesto el día de la jornada es una de las tareas donde más atención pone el Instituto. El área encargada de esto es la Dirección de Ejecutiva de Capacitación Electoral y Educación Cívica (*DECEEyEC*), por medio de sus capacitadores electorales.

Cuando un funcionario no asiste el día de la jornada, su puesto tiene que ser cubierto por alguna de las personas que se encuentran en la fila listas para votar. El problema de lo anterior es que esas personas no fueron capacitadas para el puesto y pueden poner en peligro la calidad de la elección en esa casilla. Por lo tanto, una forma en la que se puede medir la calidad en la capacitación es utilizando a la variable (o alguna transformación de la misma) determinada por el número de *funcionarios que no llegaron* a ocupar su puesto el día de la jornada.

El problema descrito anteriormente es una dificultad clave a la que se enfrenta el INE requiere estudiar este fenómeno con el fin de entender qué variables o factores pueden ayudar a identificar los casos donde existirá *ausencia de ciudadanos* que ya estaban capacitados y designados para ser funcionarios de casilla.

## 2. Caso de Estudio

Después de las elecciones electorales del 2015 surgen ecos de la práctica democrática, los consejeros electorales comienzan a analizar los resultados de la jornada. El INE comienza a explorar medidas para accionar de las personas durante las elecciones, con el fin de encontrar áres de oportunidad y fortalezas en el proceso.

Las actividades electorales requieren accionar a los ciudadanos que fungen como funcionarios de casilla. Durante el proceso electoral se pueden presentar situaciones que entorpecen el proceso electoral, por lo que se requiere realizar un esfuerzo para el calculo de probabilidades de ocurrencia. La ausencia de funcionarios de casilla es una de las situaciones más habituales y que significan mayor problemática para el INE (en ausencia de alguno, se tomará a una persona de la fila para votar para que tome las responsabilidades).

En 2015, 7.7 % de los funcionarios capacitados no se presentaron. Estas faltas ocurrieron en casi una cuarta parte de las casillas (23 %) en todo el pas. Del mismo modo y para hacer un análisis más fundamentado es necesario realizar un estudio de los factores inciden en la probabilidad de asistencia. De manera general se tomarán los rechazos (se invita a participar al proceso electoral y rechaza), notificaciones efectivas(Persona con probabilidad de fungir como funcionario) y sustituciones(Casillas en la que el capacitado declina en algún punto del adiestramiento y se sustituye por una persona en la lista de reserva) como los factores de alerta (variables explicativas).

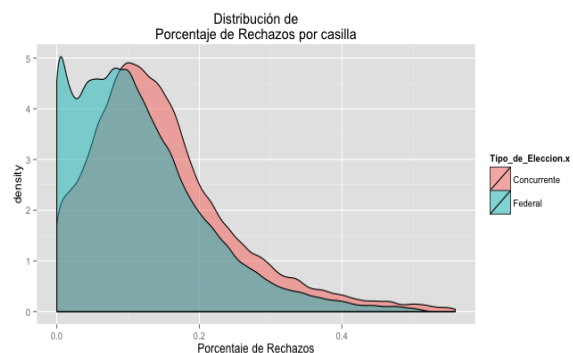
## 3. Análisis Exploratorio

### 3.1. Transformación de variables

De las más de 140,000 casillas, más de 32,000 tuvieron *Ausentismo*, lo cual representa un 23 % del total.

#### 3.1.1. Rechazos

Figura 1



La gráfica 1 el porcentaje de casillas con ausentismo para diferentes niveles de porcentaje de rechazos.

$$\% \text{ de rechazos} = \frac{\text{rechazos}}{\text{visitados}}$$

Los intervalos de porcentaje de rechazos en cada casilla se construyeron como sigue:

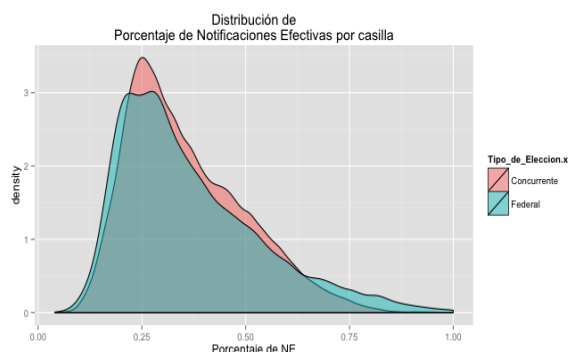
$$0 - 10 \% > 10 \% - 20 \% > 20 \% - 30 \% > 30 \% - 40 \% > 40 \%$$

- **En casillas federales y concurrentes**, a mayor nivel de rechazos hay mayor porcentaje de casillas con ausentismo. Si hay un alto nivel de rechazos, el porcentaje de casillas con ausentismo parece estabilizarse
- **En casillas federales**, la proporción de casillas con ausentismo pasa de 12 % para el nivel ms bajo de rechazos a 24 % en el nivel ms alto de esta variable

Por lo tanto, si se obtiene ms de 40 % de rechazos, la proporción de casillas con ausentismo aumenta en 100 % respecto a la proporción de casillas con ausentismo que hay con 0 a 10 % de rechazos.

### 3.1.2. Notificaciones efectivas

**Figura 2**



En la gráfica 2 se presenta el porcentaje de casillas con ausentismo para diferentes niveles de porcentaje de notificaciones efectivas.

$$\% \text{ de notificaciones efectivas} = \frac{\text{notificaciones efectivas}}{\text{visitados}}$$

Al igual que en la variable rechazos, se hicieron intervalos de porcentaje de notificaciones efectivas en cada casilla:

$$0 - 25 \% > 25 \% - 50 \% > 50 \% - 75 \% > 75 \%$$

- **En casillas federales**, la proporción de casillas con ausentismo pasa de 17 % para el nivel ms bajo de notificaciones efectivas a 9 % en el nivel ms alto de esa variable.

Por lo tanto, si se obtiene ms de 75 % de notificaciones efectivas, la proporción de casillas con ausentismo disminuye en 80 % respecto a la proporción de casillas con ausentismo que hay con 0 a 25 % de notificaciones efectivas

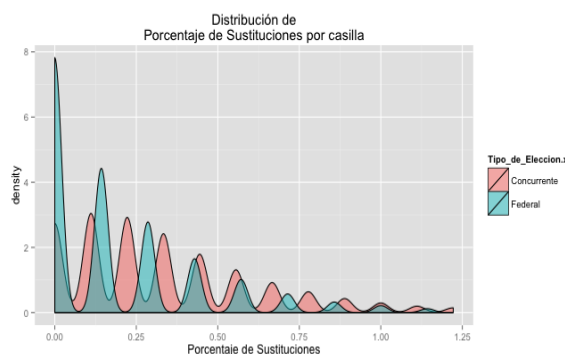
En las notificaciones efectivas en elecciones concurrentes se observa que a mayor nivel de notificaciones

efectivas por casilla, la probabilidad de que en la casilla se presentara ausentismo disminuye. La probabilidad de ausentismo pasa del 35 % para el nivel ms bajo de Notificaciones efectivas al 30 % en el nivel ms alto.

En las notificaciones efectivas en elecciones federales se muestra que a mayor nivel de notificaciones efectivas por casilla, la probabilidad de que en la casilla se presentara ausentismo disminuye. La probabilidad de ausentismo pasa del 17 % para el nivel ms bajo de Notificaciones efectivas al 9 % en el nivel ms alto. Por lo tanto, si se obtienen ms del 75 % de notificaciones efectivas, el ausentismo disminuirá su probabilidad de aparición en un 80 % con respecto a la ms alta.

### 3.1.3. Sustituciones

**Figura 3**



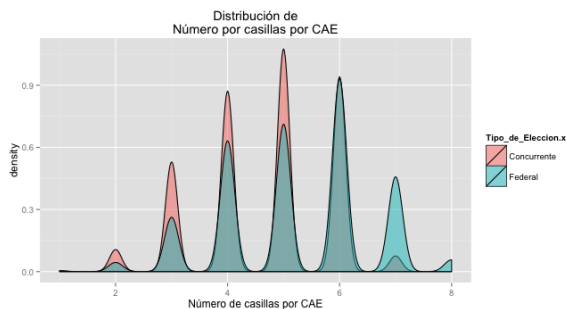
En la gráfica 3 se muestra el porcentaje de casillas con ausentismo para cada número de sustituciones que hubo por casilla

- **En casillas federales y concurrentes**, a mayor número de sustituciones por casilla aumenta la proporción de casillas con ausentismo.
- **En casillas federales**, la proporción de casillas con ausentismo pasa de 10 % cuando no hay sustituciones a 27 % cuando hay 8 ms sustituciones por casilla.

De las variables que se presentan en el estudio, sustituciones es la que tiene la relación más clara, casi lineal, con porcentaje de casillas con ausentismo. Si hay 8 más sustituciones en una casilla, la proporción de casillas con ausentismo aumenta en 170 %

### 3.1.4. Número de Casillas por Capacitador

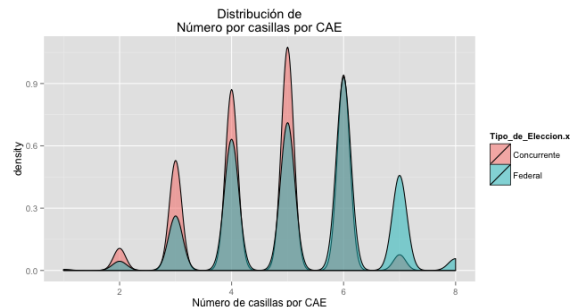
Figura 4



Cada *Capacitador Electoral (CAE)* tuvo a su cargo determinado número de casillas para integrarlas por medio de la capacitación y designación de ciudadanos que fungirán como funcionarios de las mismas. EL *INE* estipula promedios de casillas por *CAE* diferenciados entre casos con elecciones concurrentes y federales; en promedio los números para los primeros casos son menores que para los segundos. Para problemas con la escala, esta variable se divide entre el máximo de casillas por *CAE* que hubo en el ámbito respectivo. Es decir, para las casillas concurrentes, se dividió entre 9 y para las federales fue entre 10.

### 3.1.5. Porcentaje de Mujeres Designadas

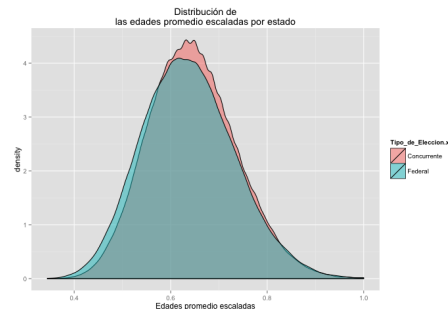
Figura 5



El porcentaje de mujeres es en general mayor para elecciones concurrentes. Por otra parte se distingue que si el *CAE* tiene más de 6 casillas, el porcentaje de mujeres disminuye considerablemente.

### 3.1.6. Edad Promedio

Figura 6



La edad de los ciudadanos designados tanto en elecciones concurrentes como en Federales va de 20 a 64 años y el promedio en ambos casos es muy parecido.

## 4. Metodología y Modelo

Bajo el contexto de construir un modelo de regresión lineal generalizado, se considera un punto de vista Bayesiano para la selección de variables. Dada una variable  $Y$  y un conjunto  $p$  de posibles regresores  $X_1, \dots, X_p$ , el problema se reduce en encontrar el mejor modelo de la forma  $Y = X_1\beta_1 + \dots + X_q\beta_q + \epsilon$  donde  $X_1, \dots, X_q$  es el subgrupo seleccionado de

$X_1, \dots, X_p$ . La identificación de ese subgrupo óptimo es un problema muy importante en el análisis estadístico que ha sido causa de mucha labor mental de los estadísticos que se han enfocado en solucionarlo.

Una primera alternativa que se piensa es la de evaluar todos y cada uno de los modelos que son posibles generar con las covariables con las que se cuenta. El problema de hacer tomar este camino es que conforme  $p$  crece, la cantidad de modelos por evaluar crece exponencialmente ( $2^p$  modelos).

Es por lo anterior que se recurre a alguna de las técnicas de selección de variables que algunas de esas grandes mentes han generado. En este caso de estudio se recurrió a uno de los métodos de la Selección de variables por Búsqueda Estocástica (SSVS por sus siglas en inglés), más particularmente, el que considera efecto aleatorio y que fue introducido por *Meuwissen & Goddard*.

Este procedimiento conlleva la especificación de una prior de tipo jerárquica que usa los datos para asignar mayor probabilidad posterior a los modelos ms prometedores. Para evitar la carga abrumadora de calcular las probabilidades a posteriori de todos los modelos  $2^p$ , SSVS utiliza el muestreo de Gibbs para simular una muestra de la distribución posterior. Dado que los modelos de alta probabilidad son más propensos a aparecer rápidamente, el muestreo de Gibbs a veces puede identificar esos modelos con recorridos relativamente cortos.

En este enfoque, la clave reside en suponer una distribución estrecha concentrada alrededor de cero. Sea  $\theta_j = \beta_j$  y donde la variable indicadora afecta a la distribución a-priori de  $\beta_j$ , es decir,  $P(\gamma_j, \beta_j) = P(\beta_j | I_j)P(I_j)$ . Podemos entonces definir una distribución mixta para  $\beta$

$$P(\beta_j | \gamma_j) = (1 - \gamma_j)N(0, \tau_2) + \gamma_j N(0, g\tau^2) \quad (1)$$

Donde la primera densidad está centrado en cero y tiene una pequeña diferencia, pero a fin de obtener la convergencia del algoritmo requiere la especificación de los parámetros  $\tau^2$  y  $g\tau^2$  por lo que se busca que el ajuste en la precisión no es fácil, ya que  $P(\beta_j | \gamma_j = 0)$  tiene que ser muy pequeña, pero al mismo tiempo no demasiado restringida que impida que las variables

puedan cambiar su valor de 0 a 1.

El hecho de que todas las  $\beta_j$  impliquen de manera inicial, por medio de su a-prioris, que sus respectivas covariables son no significativas para el modelo y que sean los propios datos los que vayan generando las correcciones en dichas distribuciones y les vayan dando importancia a cada covariable, dependiendo de la participación real que van teniendo en cada una de las observaciones, es lo que hizo a este método un buen candidato para aplicar la selección de variables. Otra razón es que nos permitía hacer la precisión de las  $\beta_j$  variantes y en términos de las probabilidades de la inclusión de las covariables de la base de datos. Por esto, entre otras razones, fue por lo que se eligió a este método para este ejercicio.

Por lo anterior, es necesario mencionar también una característica que se consideró en las a-prioris de este modelo. Para este método es necesario determinar cuál es la probabilidad inicial de que cada una de las  $\gamma_j$  tome el valor de 1, ó, lo que es lo mismo, cuál es la probabilidad inicial de que cada una de las  $X_j$  sea incluida en el modelo. Durante la aplicación se generaron dos corridas sobresalientes donde cada una contó con una probabilidad inicial para las  $\gamma_j$  diferente. Esto se mencionará con mayor profundidad más adelante.

El modelo de SSVS se implementa en BUGS, y con esta idea, se tiene que hacer incapié en la importancia de elegir las funciones prior utilizadas y como son implementadas.

Recordemos que en el marco bayesiano, el problema de selección del modelo se transforma en una estimación de parámetros, y en lugar de buscar el modelo óptimo individual, se buscará estimar la probabilidad posterior de todos los modelos dentro de la clase considerada de modelos, comenzamos describiendo un modelo general que constituye la base de la caracterización del método. En primer lugar, el modelo lineal normal estándar se utiliza para describir la relación entre la variable dependiente observada y el conjunto de todos los predictores potenciales  $X_1, \dots, X_p$  sea.

$$f(Y | \beta, \sigma) = N_n(X\beta, \sigma^2 I) \quad (2)$$

donde  $Y$  es  $n \times 1$ ,  $X = [X_1, \dots, X_p]$  es una matriz

de  $n \times p$ ,  $\beta$  es un vector de  $p \times 1$  de regresores desconocidos, y  $\tau^2 > 0$ .

Una de las motivaciones de SSVS se produce cuando hay algún subconjunto predictores con coeficientes de regresión tan pequeños que sería preferible ignorarlos, sean los regresores.

$$\gamma = (\gamma_1, \dots, \gamma_p)' \quad (3)$$

donde  $\gamma_i = 0$  o  $1$  si  $\beta_i$  es pequeña o grande respectivamente, el tamaño de el conjunto se denota como  $q_\gamma = \gamma'1$  dado que el valor apropiado de  $\gamma$  es desconocido modelamos estocásticamente una prior  $\pi(\beta, \tau^2, \gamma) = \pi(\beta|\tau^2, \gamma)\pi(\tau^2|\gamma)\pi(\gamma)$  y de manera condicional se especifica como sigue:

$$\pi(\beta|\tau^2, \gamma) = N_p(0, \gamma\tau^2, \gamma) \quad (4)$$

Con respecto al modelo de regresión, es válido tomar una distribución con base en el conocimiento que se tenga de los mismos. En caso de no tener información al respecto suele proponerse alguna distribución *no informativa* que permita al modelo abarcar todos (o amplia cantidad) los casos posibles donde “vivan” cada una de las verdaderas variables. es decir *inicial a-priori*.

Uno de los objetivos más importantes al utilizar un modelo *Bayesiano* es poder conocer la distribución *final* o *a-posteriori* de nuestros parámetros una vez observados y analizada la información de la muestra. Es decir, la información recabada de la muestra actualiza la información que se conocía de cada una de las variables y se tiene una nueva caracterización de cada una de las mismas y, como consecuencia, los modelos se vuelven más precisos al momento de estimar o predecir valores de nuestra variable dependiente.

## 4.1. Consideraciones a-priori

Una de las motivaciones para elegir SSVS es por que justamente se prefiere que la configuración de arranque de las simulaciones sea con todos los coeficientes no significativos; se toman en cuenta 26 variables.

Para cada conjunto de datos, se utilizaron dos conjuntos de a-prioris. El primer set fue elegido para ser informativa, y el segundo fue elegida para ser vaga. En

el primero se tomaron las probabilidades de que cada una de las covariables fueran incluidas en el modelo como  $P(\gamma_j = 1) = \frac{1}{5}$ , ya que esa es la proporción de ausentismo que se manifestó en todo el país. Sin embargo, se volvió a correr el algoritmo con una segunda a-priori sin sesgo de considerar que un caso es ms o menos probable que el otro, por lo que en este caso se tomó a  $P(\gamma_j = 1) = \frac{1}{2}$ . Los resultados que se presentan son de este segundo caso.

## 4.2. Subconjunto de Regresores

En el presente trabajo se propuso un tipo de *función liga* así como un escenario de distribuciones *iniciales no informativas*. considerando que tenga una combinación de simplicidad para interpretar.

- Establecer una distribución para nuestra variable dependiente.
- Definir la *función liga* a utilizar.
- Proponer ditribuciones iniciales (a-priori) para los coeficientes de regresión.

La variable de interés es dicotómica, se hay ausentismo o no del funcionario de casilla el día de las elecciones, la cual se busca explicar con las variables mencionadas anteriormente.

Consideramos una distribución Bernoulli para explicar la variable aleatoria donde el parámetro de probabilidad  $p$  se refieren a la probabilidad de ausencia.

Lo anterior hace sentido ya que se trata de una variable aleatoria con dominio  $\{0,1\}$  lo cual va acorde con el la probabilidad de clasificación en el problema.

De este modo, la verosimilitud de los datos se ve como :

$$Y_i|p_i \text{ Bin}(p_i) \quad (5)$$

Se utilizó la liga para el modelo bernoulli de tipo logística hay que notar que en este caso la liga canónica.

$$\text{logit}(p_i) = \frac{p_i}{1 - p_i} = \alpha + \bar{\beta} * \bar{X} \quad (6)$$

Como se mencionó con anterioridad se utilizaron a-prioris vagas o no informativas (se construye con varianza grande), como  $\beta_j \sim N(0, \tau_j^2[\gamma])$  donde

$$\alpha \sim N(0, 10^{-6})$$

$$\sigma_{\beta_j} \sim \text{Unif}(0, 20)$$

$$\tau^2 = \sigma_{\beta_j}^2$$

$$\tau_j^2[1] = \tau^2$$

$$\tau_j^2[0] = g * \tau^2 = 1000 * \tau^2$$

$$\gamma \sim \text{Ber}(0,5)$$

Figura 7

	X	mean	sd
1	Ind[1]	0.059	0.23566398
2	Ind[2]	0.434	0.49566297
3	Ind[3]	0.992	0.08909908
4	Ind[4]	0.984	0.12677505
5	Ind[5]	0.187	0.38997653
6	Ind[6]	0.289	0.45352852
7	Ind[7]	0.100	0.29960498
8	Ind[8]	0.185	0.38863252
9	Ind[9]	0.495	0.50006157
10	Ind[10]	0.103	0.30400955
11	Ind[11]	0.337	0.47287852
12	Ind[12]	0.438	0.49622382
13	Ind[13]	1.000	0.00000000
14	Ind[14]	1.000	0.00000000
15	Ind[15]	1.000	0.00000000
16	Ind[16]	0.822	0.38257651
17	Ind[17]	0.563	0.49614004
18	Ind[18]	0.409	0.49166940
19	Ind[19]	0.724	0.44709130
20	Ind[20]	0.354	0.47839054
21	Ind[21]	0.510	0.49997655
22	Ind[22]	0.516	0.49982725
23	Ind[23]	0.410	0.49191530
24	Ind[24]	0.667	0.47148311
25	Ind[25]	0.352	0.47777677
26	Ind[26]	0.541	0.49839924
27	TauM[1]	243.946	143.73406840
28	TauM[2]	0.244	0.14373407
29	alpha	-2.367	2.46815457
30	beta[1]	0.010	0.05919376
31	beta[2]	-0.257	0.30813647
32	beta[3]	-0.890	0.22873374
33	beta[4]	1.667	0.53844087
34	beta[5]	0.072	0.15748504
35	beta[6]	-0.126	0.17249664
36	beta[7]	-0.011	0.14331770
37	beta[8]	0.096	0.14828052

El resultado de las corridas nos arroja la distribución de probabilidades de cada variable después de probarse las diferentes configuraciones

Por comodidad se manipularon los modelo de tipo *bugs* como funciones en R, con el fin de poder cambiar facilmente de escenario. Para simular probabilidades posteriores, se utilizó la función *jags* la cual está cargada en la paquetería *rjags* cuyo método de simulación es un proceso recursivo de muestreo de Gibbs.

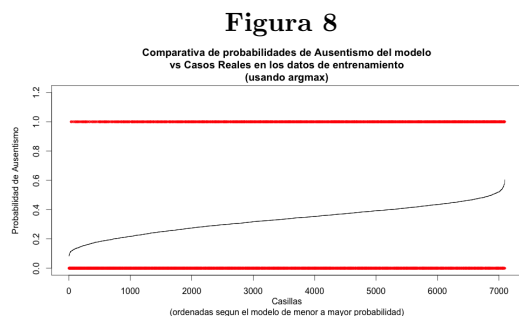
## 5.1. Resultados

Lo que primero que nos arrojan los resultado es que las variables *Notificaciones Efectivas*, *Porcentaje de Aptos*, *Porcentaje de funcionarios con primaria como máximo grado de estudios*, *Porcentaje de funcionarios con secundaria como máximo grado de estudios*, *Porcentaje de funcionarios con bachillerato como máximo grado de estudios* y *Edad mínima de los funcionarios de la casilla con respecto a las edades de la entidad* son las que resultaron con las más altas probabilidades de ser incluidas en algún modelo. De éstas todas resultaron con  $\beta'$ s significativas ya que sus intervalos de probabilidad no incluyen al cero.

## 5.2. Modelos

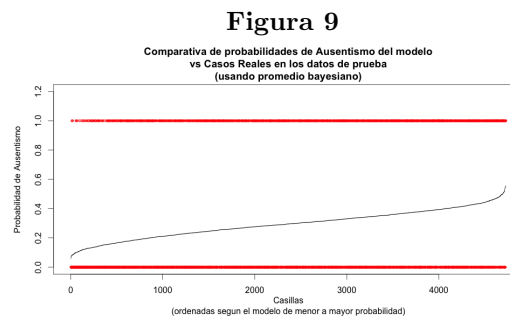
### 5.2.1. Modelo utilizando la moda de las configuraciones

Para este caso se tomó la configuración con mayor probabilidad arrojada entre las simulaciones generadas, es decir con mayor frecuencia. Posteriormente se calculan las probabilidades de ausentismo para cada casilla a la inversa de la función logit y observamos que, aunque hay cosas que aún no logra explicar el modelo, su forma creciente va alineada con más casos reales de ausentismo.



### 5.2.2. Modelo utilizando promedio bayesiano

Este modelo se ajusta vía promedio bayesiano tomando en cuenta todas las configuraciones, ponderando por la probabilidad de cada una y multiplicando por las medias de las distribuciones finales de las  $\beta'$ s. En particular, la estimación de ausentismo en esta Entidad se dificultó, ya que su capacitación electoral fue muy buena y por lo tanto el porcentaje de casillas que presentaron ausentismo fue muy bajo, lo anterior complica la estimación de las casillas con ausentismo. A primera vista parece que este modelo es que el mejor se adapta de las tres opciones que analizamos.



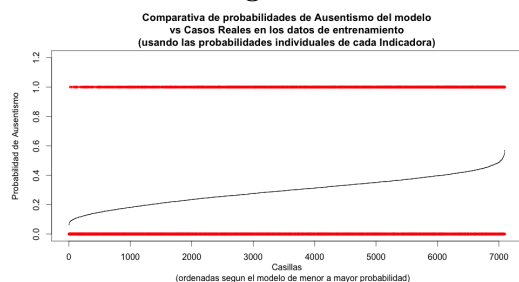
### 5.2.3. Modelo utilizando la distribución posterior generada (probabilidades individuales de cada indicadora)

En este caso, tanto para los datos de entrenamiento como los de prueba, se generó un modelo que consideraba las medias de la distribución final de las variables  $\gamma_j$  y se multiplicaban por las medias de las  $\beta_j$  que resultaron significativas con respecto a sus intervalos de probabilidad.

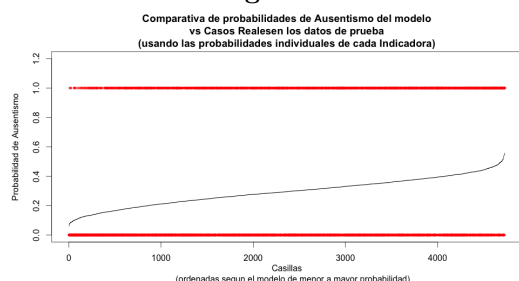
Esta configuración arrojó pocas cosas interesantes pero vale la pena ilustrarlas para conocimiento del lector.



**Figura 10**



**Figura 11**



Para tener una mayor certeza de lo último comentado, podemos hacer una evaluación de la efectividad de dichos modelos, esto con ayuda de una matriz de confusión en la que comparamos los valores reales contra los predichos, tanto los positivos como negativos

**Figura 12**

```
> matriz_confusion(data_comparacion_entrena)
      Positivos predichos Negativos predichos
Positivos Reales      117          2650
Negativos reales       91          4237

> matriz_confusion(data_comparacion_prueba)
      Positivos predichos Negativos predichos
Positivos Reales       22          1815
Negativos reales       14          2878

> matriz_confusion(data_comparacion_entrenamiento2)
      Positivos predichos Negativos predichos
Positivos Reales       35          2732
Negativos reales       33          4295

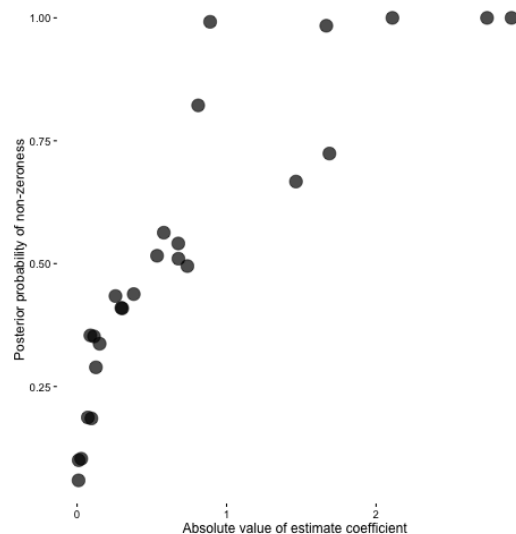
> matriz_confusion(data_comparacion_prueba2)
      Positivos predichos Negativos predichos
Positivos Reales       22          1815
Negativos reales       14          2878
```

## 6. Conclusiones

Tomando como criterio la información que nos arrojan las matrices de confusión se decide proponer al *INE* la implementación del modelo que utiliza la configuración basada en promedios bayesianos ya que

la proporción de verdaderos positivos con respecto a los positivos predichos tuvo su valor más alto en ese modelo. Así también se propone éste pues considera mucha más información que los modelos de las medias finales o los valores de la configuración-modal, es decir, contiene mayor información y eso propociona mayor certeza y tranquilidad para confiar en las estimaciones/predicciones que se necesiten.

**Figura 13**



En el gráfico anterior observamos lo anteriormente comentado, tomando los valores absolutos de las betas estimadas, lo comparamos con las probabilidades encontradas de las configuraciones diferentes de cero, de tal manera que se puede ver que mientras mayor es el valor de las betas, es decir más significativo la probabilidad de que dicha variable ocupe un lugar en el modelo óptimo se acercan a 1.

Por otro lado, el modelo de regresión adecuado para el DF fue el logit. se presentan los efectos que tiene el aumento en la probabilidad por un cambio en la variable en cuestión.

Por lo tanto, una primera recomendación para realizar es generar modelos por entidad, donde en cada una se consideren las variables respectivas que resulten significativas y puedan ser evidentes los patrones que generan las variables explicativas.

Consecuentemente, el siguiente paso es el de realizar la selección y análisis para cada una de las entidades, para ir identificando las variables significativas para

el *Ausentismo* en cada una, pero ya de una manera automática.

Luego entonces, durante los siguientes procesos electorales, conforme, para cada casilla, se vaya generando la información del proceso de capacitación, ésta se podrá ingresar al modelo de la entidad correspondiente y saber cuál es la probabilidad estimada de que en esa casilla se presente *Ausentismo*.

Un siguiente paso en la continuidad de este estudio es la de realizar un análisis de sensibilidad más detallado de manera que la calibración de los parámetros de las distribuciones iniciales y los valores iniciales sean más adecuados a los datos y los modelos que se generen sea más certeros.

As también, se propone iniciar la incorporación de un modelo espacial para poder considerar las geo-referencias de cada casilla, pues se tiene la hipótesis que casillas cercanas se comportan de manera similar y si e pueden extraer datos de las zonas de las casillas pueden irse agregando al estudio y ayudar a explicar mejor el fenómeno.

Cabe señalar que dependiendo el tipo de regresores la efectividad del método de selección de variables se verá afectada, se corroboró que SSVS es computacionalmente exhaustivo. Por lo tanto, en el paso de detección de variables se tiene que modular sobre la correcta elección de la función a-priori, así como la configuración inicial. El muestreo de Gibbs también requiere la selección de los parámetros anteriores. En este estudio de simulación, cambiando el prior de los valores  $\gamma$  podemos ver diferencia significativa en el poder de SSVS y en este caso se debió a que los mismos datos fueron tomando un protagonismo importantísimo a tal grado que corrigieron sin ningún problema los valores de las distribuciones iniciales que se habían asignado. Fue por ello que cuando se cambió por una distribución sin el sesgo hacia un lado o hacia el otro, fue mucho más fácil que convergiera hacia el resultado final.

history strategy - a cross-cultural study. *Human Nature*, 20:317–330.

[R.B. OHara and M. J. Sillanpaa, 2009] A Review of Bayesian Variable Selection Methods: What, How and Which *Bayesian Analysis* , 85-118.

[Edward I George; Robert E. McCulloch, 1993] Variable Selection Via Gibbs Sampling *Journal of the American Statistical Association* , 881-889.

[Jian Huang, Patrick Breheny and Shuangge Ma, 2012] A Selective Review of Group Selection in High-Dimensional Models *Institute of Mathematical Statistics*.

## Referencias

[Figueredo and Wolf, 2009] Figueredo, A. J. and Wolf, P. S. A. (2009). Assortative pairing and life