# Proyecto final GLM Ausentismo de funcionarios en las casillas electorales

Pedro Vladimir Hernández Serrano José Alfredo Méndez Barrera Eliza Hernández

#### 1. Introducción

Durante el presente documento se hará mención de algunas herramientas de Estadística computacional las cuales nos ayudarán durante el estudio, el cual se llevó acabo en lenguaje R, utilizando las bondades de la paquetería rjags cuya función principal es la de simular valores de una distribución posterior vía cadenas de markov a partir de distribuciones de probabilidad propuestas con la técnica de muestreo de Gibs. Encontraremos 3 fases en el proceso, un vistazo general del problema propuesto y los datos utilizados, después se comparan y ajustan los métodos estudiados, y finamente se harán comentarios pertinentes del experimento.

#### 2. Caso de Estudio

Después de las elecciones electorales del 2015 comienzan a surgir ecos de dicha práctica democrática, los consejeros electorales en particular comienzan a analizar los resultados de la jornada, y más allá de conocer quienes resultaron electos o no, al Instituto Nacional Electoral le interesa saber el accionar de las personas durante las elecciones, esto claro con miras a encontrar áres de oportunidad y fortalezas en el proceso.

Un suceso en particular que es de suma importancia para que las actividades electoreras sigan su marcha es la de el pertinente accionar de los ciudadanos que fungen como funcionarios de casilla, es completamente comprensible que nada resulte perfecto, y debido a al importancia de hacer cumplir el deber ciudadano siendo funcionario, la ausencia de dichas personas se convierte en una problemática a la cual poner atención, ya que

en ausencia de alguno, se tomará a una persona de la fila para votar para que tome las responsabilidades.

Para el ejercicio se construye una tabla resumen la cual es por si misma una consulta hecha a la base de datos de capacitación electoral .<sup>el</sup>ecciones 2015 procesado" de la bodega de datos en la que considera todas las variables referentes a la selección de ciudadanos para fungir como funcionarios, a los procesos de capacitación de los mismos y el registro de actividades de los funcionarios el día de las elecciones.

El estudio se puede aterrizar en tratar de identificar los factores que inciden en el ausentismo de funcionarios de casilla el día de la elección a pesar de que dichas personas se capacitaron.

En 2015, 7.7% de los funcionarios capacitados no se presentaron. Estas faltas ocurrieron en casi una cuarta parte de las casillas (23%) dispersadas en todo el pas

Del mismo modo y para hacer un análisis más fundamentado se analizan qu factores inciden en la probabilidad de asistencia. De manera general podemos decir que el comportamiento de las variables 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) son datos de alerta, Revisemos más a fondo lo que sucede

#### 3. Análisis Exploratorio

#### Rechazos:

Se grafica el porcentaje de casillas con ausentismo para diferentes niveles de porcentaje de rechazos. % de rechazos = rechazos/visitados

Se hicieron intervalos de porcentaje de rechazos en cada casilla: 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 proporcin 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 m<br/>s de  $40\,\%$  de rechazos, la proporcin de casillas con ausentismo aumenta en  $100\,\%$  respecto a la proporcin de casillas con ausentismo que hay con 0 a  $10\,\%$  de rechazos

Notificaciones efectivas:

Se grafica el porcentaje de casillas con ausentismo para diferentes niveles de porcentaje de notificaciones efectivas. % de notificaciones efectivas = notificaciones efectivas/visitados

Como con rechazos, se hicieron intervalos de porcentaje de notificaciones efectivas en cada casilla: 0 - 25 %  $\stackrel{.}{,}25$  % - 50 %  $\stackrel{.}{,}50$  % - 75 %  $\stackrel{.}{,}75$  % En casillas federales:

La proporcin 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 m<br/>s de 75 % de notificaciones efectivas, la proporcin de casillas con ausentismo disminuye en 80 % respecto a la proporcin de casillas con ausentismo que hay con 0 a 25 % de notificaciones efectivas

Sustituciones:

Se grafica el porcentaje de casillas con ausentismo para cada nmero de sustituciones que hubo por casilla

En casillas federales y concurrentes, a mayor nmero de sustituciones por casilla aumenta la proporcin de casillas con ausentismo.

De las variables que se presentan aqu, sustituciones es la que tiene la relacin ms clara, casi lineal, con porcentaje de casillas con ausentismo En casillas federales:

La proporcin de casillas con ausentismo pasa de  $10\,\%$  cuando no hay sustituciones a  $27\,\%$  cuando hay 8 ms sustituciones por casilla.

Por lo tanto, si hay  $8\,$  ms sustituciones en una casilla, la proporcin de casillas con ausentismo aumenta en  $170\,\%$  respecto a la proporcin de casillas con ausentismo que hay cuando no hay sustituciones en una casilla

- 12. Cambio en la clasificación de casillas
- 16. Modalidad de capacitacin y asistencia a simulacros La mayor parte de los funcionarios designados  $(45\,\%)$  tuvo capacitacin individual y particip en al menos un simulacro.

De todos los funcionarios que no asistieron a la casilla el da de la jornada, 42% tuvo capacitacin individual y no asisti a ningn simulacro. Esa fue la modalidad de capacitacin con la que hubo el mayor nmero de ausencias.

La modalidad de capacitacin con la que hubo el menor nmero de ausencias a la casilla fue la capacitacin grupal y asistencia a al menos un simulacro (12%).

18. Asistencia a simulacros

–grafico de rechazos La mayora de los datos est<br/>n para porcentajes entre 0% y 40%. –grafico de Notificaciones Efectivas La mayora de los datos est<br/>n para porcentajes entre 20% y 50%. Sin embargo, tambin hay una buena<br/> cantidad de informacin repartida en el resto del intervalo [0,100%].

-Probabilidad de ausentismo graficas

rechazos, elecciones concurrentes

Observamos que a mayor nivel de rechazos por casilla, la probabilidad de que en la casilla se presentara ausentismo aumenta

La probabilidad de ausentismo pasa del  $26\,\%$  para el nivel m<br/>s bajo de Rechazos al  $39\,\%$  en el nivel m<br/>s alto.

Por lo tanto, si se obtienen m<br/>s del  $40\,\%$  de rechazos, el ausentismo aumentara su probabilidad de aparicin en un  $50\,\%$  con respecto a la m<br/>s baja rechazos elecciones federales

Observamos que a mayor nivel de rechazos por casilla, la probabilidad de que en la casilla se presentara ausentismo aumenta

La probabilidad de ausentismo pasa del 12.48 % para el nivel m<br/>s bajo de Rechazos al 23.92 % en el nivel m<br/>s alto.

Por lo tanto, si se obtienen m<br/>s del  $40\,\%$  de rechazos, el ausentismo duplicara su probabilidad con respecto a la m<br/>s baja.

notificaciones efectivas, elecciones concurrentes

Observamos 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.

Por lo tanto, si se obtienen m<br/>s del 75 % de notificaciones efectivas, el ausentismo disminuira su probabilidad de aparici<br/>n en un 14 % con respecto a la m<br/>s alta.

notificaciones efectivas, elecciones federales

Observamos 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 m<br/>s del 75 % de notificaciones efectivas, el ausentismo disminuira su probabilidad de aparici<br/>n en un 80 % con respecto

a la ms alta.

### 4. Modelo y Metodología

Se eligiío un modelo de regresión (lineal sobre los parametros) para el análisis, ya que permite mayor intuitividad en la interpretación de los resultados.

Para poder realizar un enfoque bayesiano de regresión es necesario definir las distribuciones de las variables que forman funciones de verosimilitud, así como las distribuciones de conocimiento previo o a priori, y de este modo el producto de ambas será proporcional a una distribución posterior.

Necesitamos también definir una función liga que relacione los estimadores de la regresión con los parámetros de la distribución de la variable respuesta o de interés, Es decir, una componente sistemática y otra aleatoria para el modelo.

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 parmetro 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:

$$Y_i|p_i \ Bin(p_i)$$

Se utilizaron dos tipos de ligas para el modelo bernoulli, la logística y la probit, las cuales van muy bien con dicha distribución, más aún, la logística es la liga canónica.

$$\begin{array}{l} \text{logit}(p_i) = \frac{p_i}{1-p_i} = \bar{\beta} * \bar{X} \\ \Phi(p_i)^-1 = \bar{\beta} * \bar{X} \end{array}$$

Se utilizaron también en las funciones a priori, una opción vaga o no informativa (se construye con varianza grande), como sigue.

$$\beta_i N(0,0,001)$$

La varianza se ve como  $1/\tau$  donde  $\tau$  Gamma(0,001,0,001)

### 5. Cómputo y resultados

Una manera útil de manipular los modelo de tipo *bugs* es escribirlas como funciones en R, y así poder cambiar muy fácil entre las variaciones del modelo con solo cambiar el nombre de la función.

Para simular probabilidades posteriores, utilizamos la función jags la cual está cargada en la paquetería rjags cuyo método de simulación es un proceso recursivo de muestreo de gibs, para el presente experimento se llevaron a cabo corridas con el listado de parámetros previamente mencionados, 1 cadena, 50 mil iteraciones y un proceso de calentamiento de 5 mil iteraciones, en caso de querer replicar la información se utiliz la semilla 159549.

Tenemos los siguientes resultados

|       | Modelo 1 | Modelo 2    |  |
|-------|----------|-------------|--|
| Liga  | Logit    | Probit      |  |
| Prior | Vaga     | Informativa |  |
| DIC   | 8.54     | 6.86        |  |

Cuadro 1: Variaciones en el Modelo Bernoulli

opcion binomial-logit-vaga

```
\begin{split} p &= exp(alpha + beta*LD50)/(1 + exp(alpha + beta*LD50))\\ liga_p &< -log(0.5/(1-0.5)) \text{ como es cero entonces}\\ eta &= beta0 + beta1*LD50 = 0 \text{ de aqui que } LD50 < --beta0/beta1\\ LD50 &= -0.1157384 \end{split}
```

Considerando el criterio de informacin de la devianza (DIC).

DIC = Dbar + pD = Dhat + 2pD la cual mide la bondad de ajuste, y dado que el modelo con el menor DIC es que mejor se ajusta al conjunto de datos. De aquique debido al criterio elegimos la variación 4 para utilizarse en las estimaciones como sigue.

|             | Mean  | SD        |
|-------------|-------|-----------|
| $\hat{y_1}$ | 0.025 | 0.1685331 |
| $\hat{y_2}$ | 0.961 | 1.0404622 |
| $\hat{y_3}$ | 3.077 | 1.2832450 |
| $\hat{y_4}$ | 4.959 | 0.2671898 |

Cuadro 2: Tabla de Estimaciones

|           | Mean  | 2.25%  | 97.5% |
|-----------|-------|--------|-------|
| $\beta_0$ | 0.585 | -0.359 | 1.700 |
| $\beta_1$ | 5.223 | 1.844  | 10.07 |

Cuadro 3: Tabla de Regresores

Continuando con la misma temática en las siguientes gráficas observamos la estabilización y convergencia de los parámetros monitoreados, es claro observar que depués de la primera cuarta parte de las iteraciones, los valores arrojados por la cadena oscilan al rededor de la media, esto gracias al teorema ergódico de convergencia

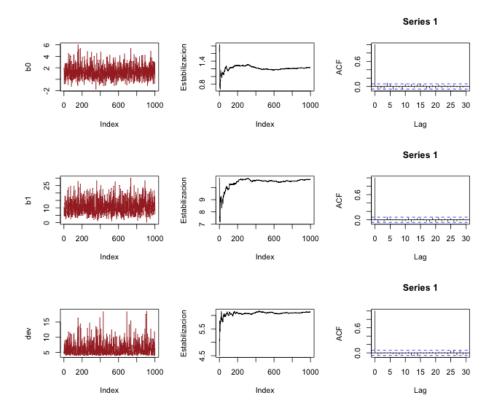


Figura 1: Convergencia

## 6. Discusión

El problema anterior tiene mucho material para extenderse, as como para continuar con nuevas pruebas y modelos, se eligieron los anteriores modelos debido a su naturaleza intuitiva con respecto a la utilizacin de las variables y a su algortmo matem atico