

DETERMINACIÓN DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE DE MEJOR AJUSTE PARA LA PREDICCIÓN DEL ABSTENCIONISMO ELECTORAL EN GUATEMALA.

Determination of the best-fit multiple linear regression model for the prediction of electoral abstentionism in Guatemala

Carlos Ottoniel Guevara Carranza

Mtro. en Estadística Aplicada
guevaracarlos.c88@gmail.com

Mayra Virginia Carvajal Castillo

Mtra. en Estadística Aplicada
mayracarvajalcastillo91@gmail.com

Recibido: 15 de marzo de 2022.

Revisado: 3 de junio de 2022.

Aprobado: 7 de octubre de 2022.

RESUMEN

Este estudio se orienta a determinar un modelo de regresión, por medio del análisis de las variables del padrón electoral, para describir el fenómeno del abstencionismo en Guatemala. Se analizan las variables del padrón electoral más actualizadas hasta el 2019 así como los resultados de los comicios del 2015 en la primera elección; se estudia la correlación que tienen estos entre sí y variables socioeconómicas como el Índice de desarrollo humano. Para el abstencionismo, se analiza la proporción de votos para calcular un modelo de regresión, agotando los supuestos de linealidad, hasta encontrar el mejor ajuste.

Entre los modelos que se evalúan, el modelo *Extreme gradient boosting* es el mejor estimado, dentro de una data en donde las variables como el rango etario, el sexo, el alfabetismo, no se relacionan entre sí, ni con variables socioeconómicas, cuando de abstencionismo se trata, en virtud de que cada una de ellas se ve afectada por este fenómeno, sin distinción alguna.

ABSTRACT

This study is oriented to determine a regression model, through the analysis of the variables of the electoral roll, to describe the abstentionism phenomenon in Guatemala. The updated electoral registration variables, as of 2020, were analyzed as well as the result of the 2015 first elections; the correlation between these two and the socioeconomic variables such as the human development index. For abstentionism, the proportion of votes was analyzed to calculate a regression model, exhausting the assumptions of linearity, until the best fit was found.

Within the evaluated models, the Extreme gradient boosting, is better adjusted inside a database in which the variables like age, range gender, and literacy, are not related to each other, or to socioeconomic variables, when it comes to abstentionism, in virtue of each of them being affected by this phenomenon with no distinction whatsoever.

PALABRAS CLAVE

Bondad de ajuste, correlación, normalidad, abstencionismo electoral.

KEYWORDS

Goodness of fit, correlation, normality, electoral abstentionism.

INTRODUCCIÓN

Este estudio analiza el fenómeno del abstencionismo en Guatemala, con base en la data proporcionada por el Tribunal Supremo Electoral, la cual contiene los votos emitidos durante la primera elección del año dos mil quince, y el padrón actualizado al año dos mil diecinueve, con la pretensión de encontrar un modelo de regresión que permita inferir el comportamiento de dicho fenómeno de la manera mejor ajustada posible. Para tal efecto se analizan los estados civiles y el alfabetismo de conformidad con la clasificación del padrón, estableciendo el nivel de asociación de cada uno de los componentes, así como el ajuste del mejor modelo de regresión para describir el comportamiento del abstencionismo electoral en Guatemala.

Las campañas electorales difundidas por la institución competente, tienen como propósito incentivar a la población a emitir su sufragio mediante un evento electoral, sin embargo, aunque la cantidad del padrón electoral crece a un ritmo más acelerado que el de la población, no significa que las tendencias de participación sean altas; por el contrario, la apatía a los eventos antes mencionados crece, por lo que es importante definir las proporciones que participan en dicho evento, para efectos de establecer los segmentos de la población que afectan el fenómeno anteriormente relacionado.

DESARROLLO DEL ESTUDIO

El estudio es de tipo cuantitativo, alcance descriptivo y diseño no experimental. La población en estudio es la registrada en el padrón electoral de la República de Guatemala, la cual se encuentra dividida en subpoblaciones descritas en el artículo 5 y 5 bis del Reglamento de la Ley Electoral y de Partidos Políticos, la cual establece la residencia electoral y de conformidad a la ley la circunscripción electoral más pequeña para efectos del cambio de alguna papeleta es la vecindad, por lo que los datos provenientes de los municipios se estudian en su totalidad.

El análisis de datos se inicia agotando los supuestos de normalidad, como base para construir una matriz

de correlación. Posteriormente, se aplica el test de Friedman para identificar las circunscripciones domiciliarias del padrón electoral, a efectos de establecer si la diferencia entre la cantidad de empadronados de la misma, ostenta una diferencia significativa. Finalmente, se evalúan diferentes modelos de regresión como lo sugiere Pons (2019) y Riba (2000), sobre cada una de las variables del padrón electoral, teniendo como variable respuesta la proporción de votos emitidos, inferida a partir de los rangos etarios, el sexo y el alfabetismo, evaluando cada variable por separado y de manera global.

RESULTADOS OBTENIDOS

Luego de analizar los residuos del modelo se infiere la proporción de votos, con base en la cantidad de sufragios de cada zona municipal; se determina que dicho modelo lineal múltiple no presenta un comportamiento normal a pesar de las transformaciones realizadas. Por consiguiente, se correlaciona cada una de las variables del padrón como lo realiza Hernández (1991), y se relaciona con variables socioeconómicas como lo muestra la figura 1, de conformidad a lo referido por Fornos , Power y Garand (2004), conjuntamente con Chavarría (2014), que sugieren la asociación antes mencionada.

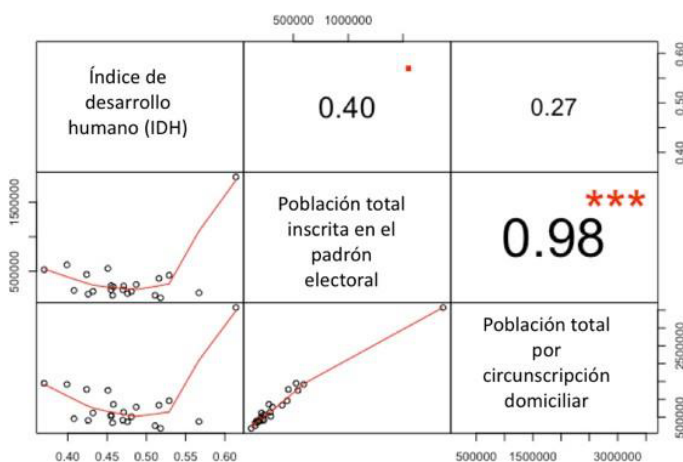


Figura 1. Correlación de Spearman.

Fuente: elaboración propia.

Para determinar si existen poblaciones con similar comportamiento electoral se aplica el test de Friedman

sobre la data de empadronados, la cual demuestra que cada circunscripción departamental tiene diferencias significativas, por lo que no existen dos departamentos homogéneos, desde el punto de vista estadístico, con respecto al padrón electoral.

El ajuste de los modelos se evalúa con base en el coeficiente de determinación, el error cuadrado medio y la media del error absoluto. Como resultado de interés, se determina que el mejor ajuste se obtiene para el modelo *Extreme gradient boosting*, cuando se utilizan todas las variables del padrón electoral.

DISCUSIÓN DE RESULTADOS

Cada una de las circunscripciones electorales es tan dispersa que todas ellas tienen diferencias significativas entre sí, por lo que cada una tiene una participación diferenciada dentro del padrón electoral, haciendo imposible inferir una con respecto de otra. Luego del análisis de mejor ajuste, el modelo *Extreme gradient boosting* es el que mejor estimación presenta con respecto de otros, vale la pena mencionar que la población más joven es la más dispersa en su comportamiento.

El análisis demuestra el bajo nivel de asociación que tiene el padrón electoral y las variables socioeconómicas, en virtud que cada una de ellas no se correlacionan entre sí. Al contrario de la multicolinealidad, cada una de las variables en la que se divide el padrón aporta al modelo final del abstencionismo indicado, que dicho fenómeno afecta a toda la población indistintamente de su sexo, alfabetismo o edad. Por lo que el fenómeno del abstencionismo se extiende a todo el padrón electoral, indistintamente de los estados civiles e independientemente de las variables o condiciones socioeconómicas en las que se encuentre la población.

CONCLUSIONES

1. La asociación entre el Índice de desarrollo humano, las variables del padrón electoral como el sexo y el alfabetismo, correlacionadas

a nivel domiciliario, no presentan ningún nivel de asociación relevante, caso contrario de la vinculación estadística entre el total de la población, el padrón electoral y sus variables civiles, y la ausencia o presencia de analfabetismo.

2. De acuerdo con los modelos de regresión, se determina que todas las variables y sus permutas, entre sexo, abstencionismo y rangos etarios son influyentes dentro del modelo, debido a que la ausencia de una de ellas conlleva a la pérdida de información de tal forma, que la estima del modelo se ve afectada.
3. Luego de la modelización de los datos, el modelo mejor evaluado basado en los estimadores es el de regresión con base en el modelo *Extreme gradient boosting* inferido con todas las variables independientes se establece como el modelo de mejor ajuste.

RECOMENDACIONES

1. Es relevante que el Tribunal Supremo Electoral, por medio del registro de ciudadanos y el departamento de informática amplíen el registro de los votantes, precisando la data obtenida, ya que con el tiempo se necesita mayor precisión para realizar análisis más profundos con respecto del comportamiento de las variables.
2. Se recomienda dar seguimiento y segmentar la data de conformidad a los períodos que la ley establece, es decir, los rangos etarios pueden establecerse cada cuatro años en virtud de que las elecciones generales se efectúan en ese mismo período de tiempo.

REFERENCIAS BIBLIOGRÁFICAS

- Chavarría Mora, E. (2014). Relación entre desarrollo humano y abstencionismo a nivel local en Costa Rica para las elecciones del 2010. *Revista de Derecho Electoral*. (18). Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=5607359>

- Fornos, C., Power, T. y Garand, J. (2004). Explaining Voter Turnout in Latin America, 1980 to 2000. *Comparative Political Studies*. 8 (37) 09-40. doi: 10.1177/0010414004267981
- Hernández, O. (1991). Universidad de Costa Rica. Anuario de Estudios Centroamericanos. Portal de Revistas Académicas. *Análisis del abstencionismo en las elecciones presidenciales de Costa Rica en el periodo 1953-1986*. 16. (16). Recuperado de <https://revistas.ucr.ac.cr/index.php/anuario/article/view/2734>
- Pons Sarlot, Ll. (2019). *L'abstenció diferencial i el vot dual a les Illes Balears*. Recuperado de <http://hdl.handle.net/10609/91466>
- Riba, C. (2000). Voto dual y abstención diferencial. Un estudio sobre el comportamiento electoral en Cataluña. *Reis: Revista Española de Investigaciones Sociológicas*. 91, 59-88 doi: 10.2307/40184275

INFORMACIÓN DE LOS AUTORES

Ingeniero en Industrias Agropecuarias y Forestales, Carlos Ottoniel Guevara Carranza, Facultad de Agronomía, Universidad de San Carlos de Guatemala, 2019.

Maestro en Estadística Aplicada egresado de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería, Universidad de San Carlos de Guatemala, 2021.

Afiliación laboral: Congreso de la República de Guatemala.

Mayra Virginia Carvajal Castillo, Ingeniera Industrial, egresada de la Facultad de Ingeniería, USAC, 2016.

Maestra en Estadística Aplicada, egresada de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería, USAC, 2018.

Afiliación laboral: Facultad de Ingeniería. USAC.