



Análisis de Elecciones Electorales en México 2018 Mediante Tablas de Contingencia

Hernández Durán Froylan,^a, Morales Luis Carmen,^a, Juárez Hernández Bulmaro,^a, Ramos Romero Mariana T.^a

^a*Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla. Av San Claudio, Cd Universitaria, Jardines de San Manuel, 72572 Puebla, Pue.*

froylan.hernandezd@alumno.buap.mx, carmen.moralesl@alumno.buap.mx, bjuarez@fcfm.buap.mx, mariana.ramos@alumno.buap.mx

Resumen.

En este trabajo se hace un análisis del comportamiento de los votantes que participaron en las elecciones del 2018, en las cuales participaron como candidatos a la presidencia de la República, los políticos: Ricardo Anaya Cortés, José Antonio Meade Kuribreña, Andrés Manuel López Obrador y Jaime Rodríguez Calderón. Si bien, el ganador de dicha contienda fue el actual presidente de la República Mexicana; aquí se retoma de nueva cuenta los datos recopilados de dichas elecciones y se analiza la preferencia de los ciudadanos para darle al candidato dicha victoria; así como también se estudia la dependencia de un estado con respecto a otro a partir de la emisión de su voto. El estudio se hace bajo el análisis categórico de datos, haciendo uso de tablas de contingencia y prueba de bondad de ajuste.

Palabras clave: datos categóricos, tablas de contingencia, elecciones electorales.

1. Introducción

En el año 2018, la sociedad mexicana exigía un cambio que a su juicio era necesario; y de repente, la preferencia electoral tuvo un cambio radical; partidos políticos que habían estado en el poder y habían sido preferidos por los mexicanos, fueron remplazados para darle la oportunidad a una alternativa diferente. Se sabe que el ganador de dicha contienda electoral fue Andrés Manuel López Obrador (actual presidente de México), con una elección histórica del 53 % de los votos. Aquí se plantean las siguientes preguntas ¿Qué movió a la sociedad para hacer dicha elección?, ¿La sociedad compartía el mismo pensamiento?, ¿Hubo dependencia entre personas, grupos o estados con respecto a otros para tomar dicha decisión? o ¿Fue una decisión independiente? Las preguntas anteriores motivaron a la realización de un análisis en el que se pretende llegar a una deducción puramente matemática a través de la estadística; apoyados, en particular, en la aplicación de tablas de contingencia. El propósito de este trabajo es hacer una comparación y contraste con los aspectos y comportamientos de cada estado de la República, basado en los resultados de las elecciones de México para candidato de la República en 2018. De esta manera nos permitirá contrastar la hipótesis que hemos planteado y que se establecerá más adelante.

2. Recopilación de datos.

Los datos que se emplean en este trabajo se obtuvieron en la página oficial del Instituto Nacional Electoral (INE), haciendo uso de la base de datos de libre consulta, disponible

en la página oficial, a través de los datos brindados es posible analizar algunos aspectos y comportamientos a nivel nacional de los votantes, con respecto a sus preferencias electorales, de manera particular en las elecciones del 2018 y sus respectivos candidatos a la presidencia [4].

El tamaño de la muestra que se empleó fue tomado a partir del total de la población que emitió su voto en todo el país y además del sector que vive en el extranjero, en este caso la participación fue de 54,955,487 individuos que realmente emitieron un voto favorable a cierto candidato y su participación se encuentra registrada en la lista nominal, en dicha elección se tuvo el 63.49 % de participación con respecto al total del padrón electoral.

En la base de datos brindada por el INE, se observó que los datos se encuentran organizados por entidad y partido, por otro lado el voto realizado en el extranjero también es clasificado de la misma manera que el nacional, así que se tuvo que reorganizar ambos votos para poder generar el voto total y además de integrar el partido con su respectiva coalición.

Para la elaboración de este trabajo únicamente se consideró a las personas cuyo voto se inclinó exclusivamente hacia algún candidato propuesto, es decir; los votos nulos no fueron tomados en cuenta, porque los votos nulos (y los blancos) se consideran generalmente como errores sin validez, por lo que se descartan del total de sufragios emitidos y no se consideran en la asignación de los cargos ni de las prerrogativas partidistas. La organización de los datos se hizo de acuerdo al partido y la coalición en la que cada candidato representó en la contienda, por lo que se tienen los siguientes datos y su respectiva coalición [5].

- Ricardo Anaya Cortés.
Coalición: “Por México al frente” (Partido Acción Nacional (PAN), Partido Revolucionario Democrático (PRD) y Movimiento Ciudadano (MC).
- José Antonio Meade Kuribreña
Coalición es: “Todos por México” (Partido Revolucionario Institucional (PRI)), Partido Verde Ecologista de México (PVEM) Nueva Alianza, (NA).
- Andrés Manuel López Obrador
Coalición: “Juntos Haremos Historia” (Movimiento Regeneración Nacional (MORENA), Partido del Trabajo (PT), Partido Encuentro Social (PES).
- Jaime Rodríguez Calderón (Candidato independiente.)

De manera general, en la Figura 1 se muestran los resultados electorales a nivel nacional, aquí se destaca que López Obrador ganó en todos los estados del país, a excepción de Guanajuato, en donde la elección favorable fue para el Partido Acción Nacional [3].

3. Análisis de Datos Categóricos

En 1900, Karl Pearson propuso el siguiente estadístico de prueba, que es una función de los cuadrados de las desviaciones de las cantidades observadas respecto de sus valores esperados, ponderados por los recíprocos de sus valores esperados [6]. Es posible demostrar que cuando n es grande, el estadístico tiene una distribución de probabilidad Ji cuadrada aproximada, como se verá a continuación.

3.1. Prueba de bondad Ji cuadrada. Una prueba de bondad de ajuste se emplea para decidir cuando un conjunto de datos se apega a una distribución de probabilidad dada. A continuación se da un resultado básico en la prueba de bondad de ajuste.

La Estadística de Prueba de Pearson.

$$X^2 = \sum_{i=1}^k \frac{[N_i - E(N_i)]^2}{E(N_i)} = \sum_{i=1}^k \frac{[N_i - np_i]^2}{np_i}. \quad (1)$$

donde n es el tamaño de la muestra aleatoria y cada N_i tiene distribución binomial con parámetros n y p_i , tiene



Figura 1. Resultados electorales por entidad

una distribución asintótica Ji cuadrada con $k - 1$ grados de libertad.

Se prueba a continuación este importante resultado para $k = 2$. Esto es, si $k = 2$ entonces $N_2 = n - N_1$ y $p_1 + p_2 = 1$. De forma que,

$$\begin{aligned} X^2 &= \sum_{i=1}^2 \frac{[N_i - E(N_i)]^2}{E(N_i)} = \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_2 - np_2)^2}{np_2} \\ &= \frac{(N_1 - np_1)^2}{np_1} + \frac{[(-N_1 + np_1)]^2}{n(1 - p_1)} \\ &= (N_1 - np_1)^2 \left(\frac{1}{np_1} + \frac{1}{n(1 - p_1)} \right) \\ &= \frac{(N_1 - np_1)^2}{p_1(1 - p_1)}. \end{aligned}$$

Ahora, por el Teorem Central del Límite, para n grande se tiene que:

$$\frac{N_1 - np_1}{\sqrt{np_1(1 - p_1)}}.$$

tiene aproximadamente una distribución normal estándar. Como el cuadrado de una variable aleatoria normal estándar tiene una distribución χ^2 con un grado de libertad. Entonces, para $k = 2$ y n grande, X^2 tiene una distribución χ^2 aproximada con 1 grados de libertad [6].

Si se sigue este razonamiento, puede demostrarse que para $k \geq 2$ categorías distintas, la estadística X^2 dada en 1 tiene una distribución χ^2 aproximada con $k - 1$ grados de libertad si n tiene un valor suficientemente grande. La estadística anterior recibe el nombre de estadística de prueba para bondad de ajuste Ji cuadrada de Pearson.

3.2. Tablas de Contingencia..

Definición 1. Sea n una muestra aleatoria de una población que se clasifica de acuerdo con dos o más características, cada una de las cuales contiene r y c categorías, respectivamente. Además, sea N_{ij} el número de observaciones en la categoría (i, j) de las características dadas, respectivamente, para $i = 1, 2, \dots, r$ y $j = 1, 2, \dots, c$. Entonces una tabla de contingencia es un arreglo de $r \times c$ en donde las entradas representan las realizaciones de las variables aleatorias N_{ij} [1] y [6].

Proposición: El valor esperado y estimado de la frecuencia por celda N_{ij} para una tabla de contingencia es igual al producto de sus respectivos totales de renglón y columna divididos entre el tamaño muestral total. Esto es

$$E(\widehat{N}_{ij}) = \frac{r_i c_j}{n}.$$

Grados de libertad Los grados de libertad asociados con una tabla de contingencia que tenga r renglones y c columnas siempre será igual a $(r - 1)(c - 1)$.

4. Resultados

La Tabla 1, se construyó a partir de la clasificación de resultados electorales por entidad y los distintos partidos por coalición, así mismo el total de votos por estado y partido, para que de esta manera se pudiera analizar la tabla de contingencia.

Tabla 1. Conteo de votos por entidad y frecuencia relativa y esperada por entidad.

No.	Estado	PAN	PRI	MORENA	CAND. IND.	TOTAL
1	Ags.	178,988 (125,127.85)	103,639 (92,194.33)	222,528 (298,748.20)	40,299 (29,383.61)	545,454
2	B. C.	275,503 (323,109.41)	124,225 (238,067.38)	918,939 (771,437.83)	89,823 (75,875.37)	1,408,490
3	B. C. S.	56,794 (67,811.93)	28,202 (49,963.91)	193,842 (161,903.96)	16,766 (15,924.19)	295,604
4	Camp.	54,417 (100,353.19)	96,584 (73,940.34)	275,262 (239,597.64)	11,194 (23,565.81)	437,457
5	Coah.	307,590 (308,838.81)	358,279 (227,552.78)	609,362 (737,366.16)	71,051 (72,524.23)	1,346,282
6	Col.	56,428 (76,046.75)	62,004 (56,031.33)	197,316 (181,564.94)	15,753 (17,857.96)	331,501
7	Chis.	198,107 (524,474.63)	562,863 (386,433.50)	1,485,699 (1,252,206.13)	39,607 (123,161.72)	2,286,276
8	Chih.	425,019 (330,481.90)	240,725 (243,499.44)	642,652 (789,039.99)	132,232 (77,606.65)	1,440,628
9	Cd. Mx.	1,292,576 (1,210,384.92)	652,061 (891,812.98)	3,108,385 (2,889,846.91)	223,256 (284,233.18)	5,276,278
10	Dgo.	187,947 (164,268.75)	141,291 (121,033.40)	340,829 (392,198.82)	46,009 (38,575.02)	716,076
11	Gto.	940,133 (516,671.55)	381,692 (380,684.18)	707,222 (1,233,575.92)	223,214 (121,329.33)	2,252,261
12	Gro.	217,838 (354,730.309)	285,799 (261,365.69)	1,018,163 (846,934.12)	24,531 (83,300.87)	1,546,331
13	Hgo.	188,028 (311,083.96)	57,548 (229,207.01)	850,863 (742,726.561)	59,630 (73,051.45)	1,356,069
14	Jal.	1,179,160 (779,124.51)	509,112 (574,059.82)	1,461,170 (1,860,193.82)	246,897 (182,960.83)	3,396,339
15	Méx.	1,549,824 (1,802,047.294)	1,548,662 (1,327,750.489)	4,373,267 (4,302,466.695)	383,684 (423,172.522)	7,855,437
16	Mich.	443,805 (434,321.31)	335,854 (320,008.435)	991,154 (1,036,961.12)	122,469 (101,991.13)	1,893,282
17	Mor.	142,553 (215,827.83)	99,506 (159,022.19)	638,689 (515,298.39)	60,083 (50,682.58)	940,831
18	Nay.	79,818 (108,383.15)	66,445 (79,856.83)	315,816 (258,769.527)	10,382 (25,451.48)	472,461
19	N.L.	701,702 (486,345.62)	314,333 (358,340.005)	745,219 (1,161,171.435)	358,811 (114,207.93)	2,120,065
20	Oax.	221,686 (427,460.83)	342,108 (314,953.62)	1,260,562 (1,020,581.437)	39,020 (100,380.0962)	1,863,376
21	Pue.	618,397 (682,971.8817)	490,737 (503,214.4565)	1,630,625.2255 (1,630,625.2255)	113,463 (160,381.4365)	2,977,193
22	Qro.	347,664 (228,405.21)	150,927 (168,289.22)	424,162 (545,327.44)	72,905 (53,636.11)	995,658
23	Q. R.	116,031 (137,027.23)	76,758 (120,118.64)	488,434 (389,234.63)	29,441 (38,283.48)	710,664
24	S. L. P.	334,763 (276,537.73)	260,211 (203,753.31)	527,546 (660,245.93)	82,956 (64,939.0122)	1,205,476
25	Sin.	163,956 (628,9400.26)	234,416 (213,230.44)	834,001 (690,955.7863)	29,173 (67,959.50403)	1,261,546
26	Son.	167,273 (244,068.73)	181,059 (179,830.12)	651,806 (582,724.78)	63,800 (57,314.35)	1,063,938
27	Tab.	91,342 (268,477.26)	107,538 (197,814.35)	961,710 (641,001.20)	9,749 (63,046.18)	1,170,339
28	Tams.	475,201 (367,074.85)	228,386 (270,461.166)	786,310 (876,407.25)	110,246 (86,199.73)	1,600,143
29	Tlax.	66,729 (137,764.87)	74,744 (101,505.31)	433,127 (328,919.658)	25,941 (32,351.15)	600,541
30	Ver.	1,050,599 (851,963.77)	471,313 (627,727.87)	2,059,209 (2,034,100.75)	132,737 (200,065.59)	3,713,858
31	Yuc.	320,144 (261,179.32)	324,055 (192,437.21)	455,216 (623,577.04)	39,111 (61,332.41)	1,138,526
32	Zac.	156,844 (169,093.28)	177,672 (124,588.12)	366,371 (403,717.61)	36,220 (39,707.96)	737,107
TOTAL		12,606,859	9,288,748	30,099,427	2,960,453	54,955,487

Se pretende probar el siguiente juego de hipótesis.

H_0 : ¿Los datos dan suficiente evidencia para indicar una independencia entre cada estado de la República y la emisión de su voto hacia cierto partido político?

Contra la hipótesis alternativa.

H_a : ¿Los datos determinan dependencia entre cada estado de la República y la emisión de su voto hacia cierto

partido político?

Se hizo el contraste de hipótesis, utilizando un nivel de significancia $\alpha = 0.05$. A partir de la organización de los datos, el siguiente paso fue determinar el valor esperado de la frecuencia relativa y esperada por celda, como se puede apreciar en la Tabla 1, resaltada en negritas. Con esta tabla, es posible calcular el valor del estadístico de prueba X^2 , como sigue:

$$X^2 = \frac{(178,988 - 125,127.85)^2}{125,127.852} + \dots + \frac{(36,220 - 39,707.9663)^2}{39,707.9663} = 23,183.6189 + \dots + 306.3846 = 4,408,912.04.$$

(Los resultados obtenidos del estadístico de prueba X^2 fueron calculados en el Software Excel.)

Por lo tanto el estadístico de prueba es igual a $X^2 = 4,408,912.045$.

Para poder determinar los grados de libertad, se calcula a partir del número de renglones, en este caso $r = 32$ y el número de columnas es $c = 4$, con estos datos se aplicó la siguiente regla:

$$(r - 1)(c - 1) = (4 - 1)(32 - 1) = (3)(31) = 93$$

Con lo anterior, se tiene que la prueba Chi-Cuadrada para la hipótesis nula es $\chi^2_{(0.05,93)}$ con un nivel de significancia $\alpha = 0.05$ y 93 grados de libertad.

Entonces el cuantil de cola derecha para $\chi^2_{(0.05,93)}$ de la distribución del estadístico de prueba bajo la hipótesis nula es, $\chi^2_{(0.05,93)} = 71.760$.

(El valor de $\chi^2_{(0.05,93)}$ fue calculado con el Software Excel.)

Contrastando el valor del estadístico de prueba X^2 con $\chi^2_{(0.05,93)}$, rechazaremos la hipótesis nula cuando $X^2 > 71.760$ y efectivamente el valor de $X^2 = 4,408,912.04 > 71.760$.

Entonces se rechaza la hipótesis nula en favor de la hipótesis alternativa y se puede concluir que, los datos emitidos de acuerdo a las elecciones presidenciales del 2018 no presentan suficiente evidencia para establecer dependencia entre cada estado de la República y la emisión de su voto.

5. Análisis y Conclusión.

El resultado tiene una explicación social, y se debe al comportamiento de la sociedad ante la toma de alguna decisión electoral; esto puede ser explicado a través de

diferentes teorías, en donde se destaca que la decisión deviene de una construcción social [2]; los individuos toman decisiones en función de presiones y condiciones sociales en las que viven; los resultados de las votaciones obedecen a la situación de clase, a la ocupación y al estatus socio-económico de los electores; es decir, los obreros suelen votar por partidos obreros, mientras que los empresarios, comerciantes e individuos que pertenecen a los grupos de mayores ingresos, votan por los partidos de la derecha o afines a su condición de clase, esto implica un bajo nivel de movilidad de partidos y candidatos; por lo que en este sentido se puede plantear que las campañas políticas ejercen pocos efectos en la decisión de los electores. Paul Lazarsfeld, un investigador de origen austriaco que llevó a cabo una investigación en la Universidad de Columbia, se centraron en investigar la influencia de los medios (prensa y radio) en la toma de decisión de los electores; de cuyas investigaciones se desprenden los siguientes planteamientos de la Escuela de Columbia: El primero, es el esquema de la comunicación filtrada, que describe la influencia que recibimos de los medios de comunicación a partir de nuestros intereses, usos y gratificaciones que de ellos obtenemos. Somos receptivos a los mensajes que se asemejan a lo que pensamos y rechazamos lo que son en tipo y forma diferentes a lo que estamos acostumbrados (Lazarsfeld, 1960). El segundo esquema plantea que la comunicación suele arrojar mejores resultados cuando llega a nosotros mediada por un líder de opinión que nos influye. La Escuela de Columbia era un tanto escéptica en cuanto a los efectos de los medios masivos de comunicación para lograr cambios efectivos en la conducta de los electores, ya que consideraba que únicamente se logra motivar a los inadaptados y marginales que no se encuentran integrados a sus grupos de referencia; por lo tanto, las campañas político-electorales y el uso de los medios masivos de comunicación, no cambian el comportamiento de los electores, pues éstos deciden su voto en función de su grupo de pertenencia. Es decir, se concluye entonces que, para el año 2018, la sociedad mexicana se encontraba bajo cierta presión, problemas a causa del gobierno de Enrique Peña Nieto, así como la crisis económica; esto obligó a la sociedad a exigir un cambio y eliminar la preferencia por partidos como el PRI y el PAN. Andrés Manuel López Obrador, realizó una campaña electoral, sin embargo, poco a poco fue tomado como un líder de opinión que influyó en las personas, así como tener propuestas que concordaba con el pensamiento y necesidades de la sociedad; por lo que prácticamente, casi todo el país aún cuando la mayoría de los medios de comunicación formales estaban en contra, llevo a que la sociedad tuviera una gran preferencia al candidato, llevándolo así a la victoria.

Lo anterior lleva a concluir que, como lo indica el resultado obtenido en la prueba de independencia, los estados tuvieron influencia entre ellos (dependencia) lo cual fue un factor que se reflejo en el amplio margen de triunfo del entonces candidato Andrés Manuel López Obrador.

Referencias

- [1] Canavos, George C. *Probabilidad y Estadística. Aplicaciones y otros métodos*. (primera edición en español) México: McGraw-Hill/Interamericana de México S.A de C.V., 1988.
- [2] Kuschick Ramos, Murilo. *Teorías del comportamiento electoral y algunas de sus aplicaciones*. *Revistas UNAM. Revista Mexicana de Ciencias Políticas y Sociales*, junio, 2004.
- [3] Ramírez, E., Soria Cruz, J., Cruz Valenzuela, R. y Castillo Zúñiga, I. *Análisis de los resultados en las Elecciones Federales Mexicanas 2018 mediante técnicas de minería de datos*. *Revista de Análisis Cuantitativo y Estadístico*, Vol. 5 No.16 6-10, ECORFAN, Septiembre 2018.
- [4] Sistema de Consulta de la Estadística del Proceso Electoral de 2017-2018. *Página oficial del INE: <https://www.ine.mx/voto-y-elecciones/resultados-electorales/>*
- [5] Sonnleitner, W. *La diversidad del voto nulo en México*. Mayo 10, 2021 de Universidad Nacional Autónoma de México - Instituto de Investigaciones Sociales, diciembre 2019.
- [6] Wackerly, Dennis D., Mendenhall III, W. & L. Scheaffer, R. *Estadística matemática con aplicaciones*. Séptima Edición. México, D.F.: Cengage Learning, 2008.