

Los votos de EPN están inflados en alrededor de 40%

Ángel Zambrano García*

Resumen

Independientemente de la acumulación de evidencias jurídicas sobre un posible fraude en la elección presidencial y del veredicto que emita el Tribunal Electoral del Poder Judicial de la Federación (TEPJF), un análisis de los resultados distritales con técnicas comunes de análisis exploratorio de datos (básicamente correlación y regresión simple entre variables) permite estimar que los votos de Enrique Peña Nieto (EPN) están inflados en 40% en promedio, con un margen de error entre 32 y 48%. En números redondos esto se traduce en 7.7 ± 1.5 millones de votos artificiales a favor de este candidato; es decir, el doble de la diferencia entre el primer y segundo lugar reconocida por el Instituto Federal Electoral (IFE). Este cálculo parte de la relación positiva y altamente significativa entre los votos distritales de este candidato y la serie de los votos anulados por distrito ($0.4, p < 0.00$, a partir de regresión), que en teoría debería ser cero, como de hecho lo es para Josefina Vázquez Mota (JVM) y Gabriel Quadri de la Torre (GQT). En el caso de Andrés Manuel López Obrador (AMLO), esta relación también es cercana a cero, pero es significativamente negativa ($-0.20, p < 0.02$) e implica que hay un faltante de votos a su favor en los datos oficiales, cercano a 3.2 millones. La punta de la madeja de estos cálculos se encuentra en los descuidados datos oficiales, aparentemente no sometidos a un proceso riguroso de validación numérica antes de declarar a un ganador, y particularmente en el peculiar patrón de votación distrital de EPN. Este patrón se caracteriza por: 1) un ascenso sistemático de su número de votos distritales en relación con la anulación y 2) una variabilidad relativa pequeña en sus votos distritales (coeficiente de variación, $CV = 24\%$) en comparación con la de los otros candidatos ($CV, 43-46\%$), casi el doble. Estas dos características de la votación de EPN son un fuerte indicador de una participación electoral tramposa por parte de la coalición PRI-PVEM y sus patrocinadores. La primera de ellas permite estimar el monto de la adición artificial de votos a favor de EPN por una o más vías ajenas al voto libre y secreto (Artículo 41 constitucional). La segunda apunta hacia un proceso electoral tan insolentemente controlado a favor de este candidato que terminó arruinando la variación aleatoria de su votación, la cual podría haberle dado un matiz de credibilidad a la suma oficial de sus votos.

**Nota sobre el autor. Mexicano con derechos plenos, incluido, supone, el de libertad de expresión. Es biólogo y maestro en ciencias por la Facultad de Ciencias de la UNAM. También doctor en biología vegetal por la Universidad Estatal de Arizona. Actualmente es investigador del Instituto Mexicano del Petróleo, institución que no tiene responsabilidad alguna en los contenidos de este documento estrictamente personal. Asimismo declara que no está afiliado a ninguna organización política. Manifiesta que en este análisis se apegó a los principios de ética científica que obligan a no mentir ni distorsionar dolosamente datos, métodos ni resultados. Acepta sin embargo que podría estar equivocado en sus interpretaciones, por lo que está abierto a todo tipo de críticas, sugerencias, aclaración de dudas, a mostrar resultados no incluidos aquí y, desde luego, a replicar cuestionamientos.*

Introducción

Se analizan los patrones de votación distrital de los candidatos presidenciales y se estima la cantidad de votos no libres ni secretos añadidos a EPN, presunto ganador de la contienda presidencial de 2012. Las Figuras 1 y 2 ilustran en lo general los tipos de patrones de votación encontrados al relacionar los votos de los candidatos con los votos anulados en los 300 distritos electorales del país. En la Figura 1 es visible el crecimiento de EPN con los niveles de anulación, mientras que AMLO disminuye y a JVM le suben y le bajan los votos sin conexión aparente con los votos nulos.

La Figura 2 despliega los 300 distritos y se observa que la línea de tendencia sólo asciende para EPN. Pero en este tipo de gráficas, esta no es la tendencia natural de un ganador, aunque así pudiera parecerle a alguien no familiarizado con estos análisis. Tampoco las tendencias planas de JVM y GQT ni la descendente (AMLO) son propias de perdedores. En realidad, como se explica más adelante, estas tendencias son planas cuando se cumple el principio de neutralidad de los votos nulos, tal como se observa en los patrones de JVM, Quadri y, en menor medida, en el de AMLO. Este patrón es independiente del total de votos obtenido, como lo confirma la semejanza entre estos tres candidatos, a pesar de que sus diferencias en votación total son de varios millones.

El patrón de votación de EPN es radicalmente distinto al de los otros candidatos. Se distingue por dos características con plena significación estadística: 1) una fuerte tendencia ascendente de su votación distrital en relación con los votos anulados, que no aparece en los otros candidatos, y 2) una variación interdistrital muy baja (24%), casi la mitad de la que presentan los demás candidatos (43-46%).

Aquí se analiza lo ilustrado en las Figuras 1 y 2 y su posible su impacto numérico en los resultados de esta elección. Los datos analizados fueron los distritales publicados por el IFE en su página de Internet el 7 de julio del 2012: (<http://computos2012.ife.org.mx/reportes/presidente/distritalPresidenteEF.html>).

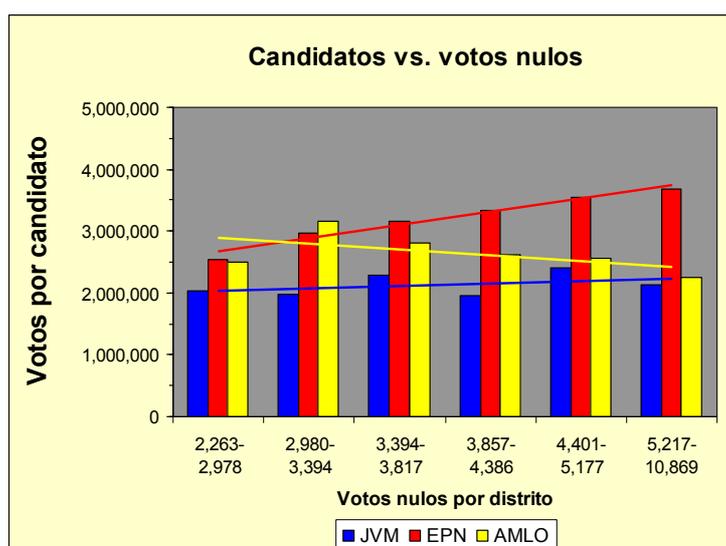


Figura 1. Variación de los votos de los principales candidatos presidenciales en relación con los niveles de anulación por distrito. Los 300 distritos están organizados en grupos de 50, de menor a mayor anulación. Los votos de los candidatos son la suma obtenida en cada grupo.

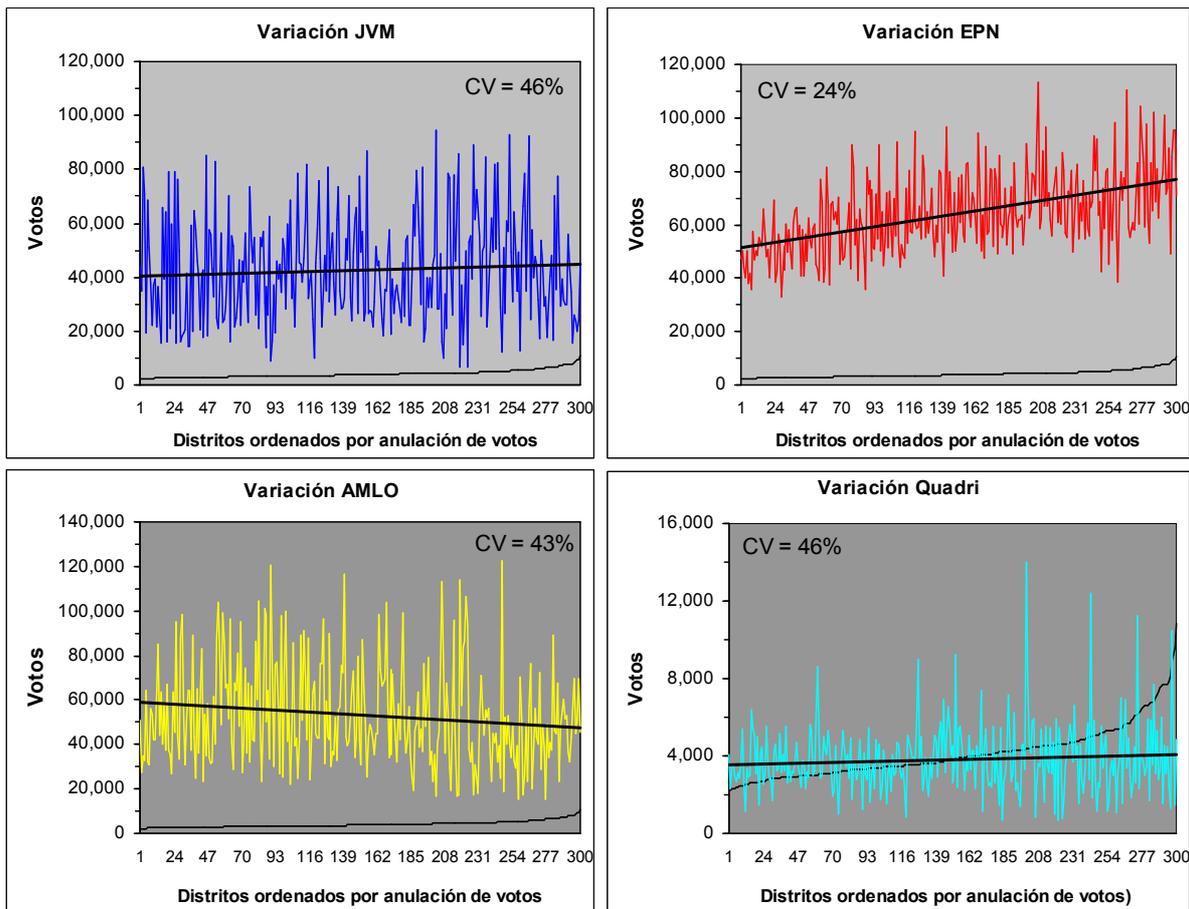


Figura 2. Variación distrital de los votos por candidato (líneas en color) vs. anulación de votos. Los distritos están ordenados de menor a mayor número de votos anulados. Las líneas negras gruesas representan la tendencia distrital promedio de los candidatos. La línea negra delgada en la base de las figuras es la cantidad de votos anulados. Como los votos se anulan por razones técnicas, según el artículo 274 del COFIPE, en principio no deben favorecer ni perjudicar a ningún candidato. Si esto se cumple, la votación promedio de cada candidato en este tipo de gráficas debe ser constante; es decir, independiente del nivel de anulación. Así se observa en los casos de JVM y Quadri, cuyas líneas de tendencia son planas. Pero no fue el caso de AMLO, a quien los votos le disminuyeron con la anulación. Su línea de tendencia comienza con un promedio cercano a los 60 mil votos y termina con alrededor de 47 mil. En cambio, la tendencia de EPN empieza con un promedio cercano a 52 mil votos y termina con alrededor de 77 mil, una diferencia sustancial, aproximadamente de 25 mil votos entre los distritos con más y con menos anulación; Así, EPN fue el único candidato a quien le crecieron significativamente los votos con la anulación. Un segundo aspecto a destacar en estos patrones de votación es la baja variación de los votos de EPN alrededor de su promedio, con un coeficiente de variación (CV, inserto en las gráficas) que es casi de la mitad que el de los otros candidatos (24% vs. 43-46%).

El razonamiento

Según el COFIPE, Artículo 274, un voto se anula cuando la boleta se depositó en blanco o fue marcada incorrectamente por el elector. Se trata de dos criterios puramente técnicos que se pueden asumir como políticamente neutrales. Por lo tanto, el número de votos anulados en las casillas o en compartimentos electorales superiores no debe guardar relación significativa, ni positiva ni negativa, con la votación de los candidatos. Si no fuese así, se puede afirmar que la anulación de votos no ha sido neutral o que algún otro factor relacionado con ella terminó beneficiando o perjudicando a algún

candidato. Esto se puede verificar con una sencilla prueba hipótesis aplicada a los datos oficiales de cada candidato.

Una forma práctica de poner a prueba esta hipótesis de neutralidad de los votos nulos es evaluar si los votos de los candidatos no tienen alguna relación estadística significativa con ellos. Según esta hipótesis, los votos de cada candidato pueden ser altos o muy altos en algunos distritos o bajos y muy bajos en otros, pero sin conexión con los niveles de anulación. La correlación resultante debe ser cero o no significativamente diferente de cero. Alternativamente, si se observa que los votos de algún candidato aumentan o disminuyen en función de la anulación, se someten a una prueba estadística para determinar si hay evidencia suficiente para rechazar la hipótesis de neutralidad. Se utilizó la correlación de Pearson para detectar posibles asociaciones entre votos de candidatos y nulos; posteriormente, regresión lineal simple para determinar la significancia, sentido y fuerza de las relaciones encontradas. Por razones de disponibilidad de datos en forma procesable y de tamaño de muestra, esta hipótesis se puso a prueba con los datos distritales del IFE.

Los cálculos y figuras que ilustran este trabajo se hicieron con Excel y Statistica, versión 7.

¿Por qué es importante analizar los votos anulados?

Entre las variables de la elección proporcionadas por el IFE junto con los votos de los candidatos, como son los tipos y números de casillas instaladas, votos totales emitidos, votos por candidatos no registrados y anulados, éstos últimos tienen características especiales. Los votos totales permiten calcular, por ejemplo, el nivel de participación ciudadana respecto a los votantes potenciales (padrón), pero los votos anulados contienen información adicional sobre el electorado, los partidos contendientes y la jornada electoral. Por ejemplo, estos votos ponen números a la capacidad de los electores de una casilla, sección, distrito o país para expresar adecuadamente sus preferencias en la boleta. Son también una expresión numérica, al parecer todavía no cuantificada, de un sector de la población que deliberadamente anula su voto emitiéndolo en blanco o escribiendo leyendas en la boleta.

El número y la variación de los votos anulados también puede depender de otros factores, como la aplicación correcta o incorrecta por parte de los funcionarios de casilla y los representantes partidarios de los criterios previstos al respecto en el COFIPE. Los partidos pueden deliberadamente maximizar la anulación de votos del candidato contrario y minimizar la propia durante los conteos al desplegar las boletas y tomar decisiones sobre si un voto es válido o debe ser anulado. Sus representantes de casilla pueden perjudicar eficientemente a los candidatos contrarios exigiendo obsesiva y firmemente la anulación de toda boleta marcada claramente a su favor por el más pequeño indicio de marcado incorrecto, como podría ser algún ligero toque de crayón al logotipo de otro partido. Con la misma vehemencia también pueden ser muy eficientes en evitar que se le anulen votos a su candidato, aunque las boletas estén marcadas incorrectamente. Estas y otras acciones se pueden repetir miles y miles de veces en todo el país durante el conteo en las casillas y llegar a afectar a favor o en contra a uno o más candidatos. Por consiguiente, la serie de los votos anulados son una representación viva, a escala, de la jornada electoral. Las acciones extracasilla de los partidos antes, durante y quizá posteriores a la jornada electoral pueden quedar involuntariamente grabadas en

la *variación* de estos votos, aunque no necesariamente en sus montos totales o subtotales.

En esta elección se anularon 2.5% de los votos presidenciales a nivel nacional, una cantidad menor a la diferencia entre el primer y segundo lugar oficiales. Pero este trabajo no pretende, desde luego, determinar si unos cuantos de esos votos eran legítimos para uno u otro candidato. Importa analizar si esos votos significaron algo más en el proceso electoral y el resultado oficial. Recientemente Héctor Aguilar Camín se refirió de manera cruda y pedagógica a este tipo de votos en el contexto de la discusión sobre el llamado voto nulo (Milenio, 24 de abril 2012). Dijo: *“En el fondo es una forma indirecta de votar por el ganador, al que el abstencionismo y los votos anulados le mejoran automáticamente el valor porcentual de los votos que sí consigue. Los votos efectivamente conseguidos por el ganador valen más si el total de los votos válidos es menor. Diez votos efectivos de cien, valen 10 por ciento. Los mismos diez votos efectivos de 50, valen 20 por ciento. Los mismos 10 votos efectivos de 20, valen 50%.”* Existe, pues, plena conciencia del potencial de la anulación de votos para ganar o perder una elección. Pero la variación de estos votos puede revelar, además de los movimientos porcentuales mencionados por ese escritor, cambios en los votos absolutos que vale la pena averiguar.

¿Hubo relación entre los votos anulados y los resultados de esta elección?

La respuesta es sí, independientemente de su cuantía.

Como primera aproximación a este asunto, la Figura 3 ilustra que EPN amplió su ventaja en votos absolutos sobre AMLO a nivel estatal cuando aumentaron los votos anulados. Cada punto en esta figura representa uno de los 32 estados del país. En el eje Y hay números positivos y negativos. Los primeros representan los estados donde EPN tuvo más votos que AMLO; los negativos son los estados donde AMLO aventajó. El punto amarillo y grande en la parte baja del gráfico corresponde al DF, un caso completamente atípico en el contexto nacional por el contundente apoyo recibido por AMLO, quien aquí obtuvo más de 1.3 millones de votos sobre EPN. El punto negro arriba a la derecha de la gráfica, donde termina la línea roja de tendencia, es el Estado de México. Ahí EPN tuvo casi 627 mil votos más que AMLO. Como una curiosidad estatal, la mayor diferencia en votos absolutos de EPN sobre AMLO se dio en Guanajuato (636 mil votos), no en el Estado de México como podría haberse esperado. El tercer lugar en este aspecto fue Jalisco, con poco más de 596 mil. La línea roja en esta figura representa la tendencia favorable a EPN. La línea se trazó sin incluir al DF por ser un dato muy alejado de la tendencia general. La inclinación o pendiente de esta recta es el coeficiente de 3.9 en la ecuación inserta en la figura; significa que por cada voto anulado a nivel estatal hubo 3.9 votos a favor de EPN sobre AMLO. El número 4,333.8 es la llamada ordenada al origen y es irrelevante en este caso. El valor de R^2 significa que 34% de la variación de la diferencia de votos entre estos candidatos se explica por la variabilidad de los votos nulos.

Por varios detalles, esta gráfica no calificaría como ejemplo de libro de texto. No se profundizó en ella hasta conseguir un mejor ajuste porque el número de estados es bajo y hay enormes diferencias poblacionales entre ellos, lo que produce un gran error estadístico. No obstante, su propósito es introducir algunos conceptos básicos de este análisis.

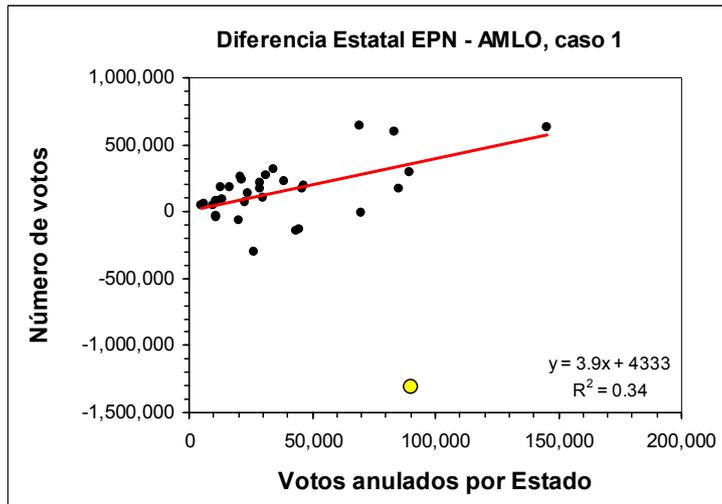


Figura 3. Tendencia estatal de la ventaja de EPN sobre AMLO respecto a los votos anulados, expresada como la diferencia de votos entre EPN y AMLO. La línea roja es un ajuste por mínimos cuadrados sin tomar en cuenta al DF (punto amarillo) por ser este un dato muy extremo y alejado de la tendencia general.

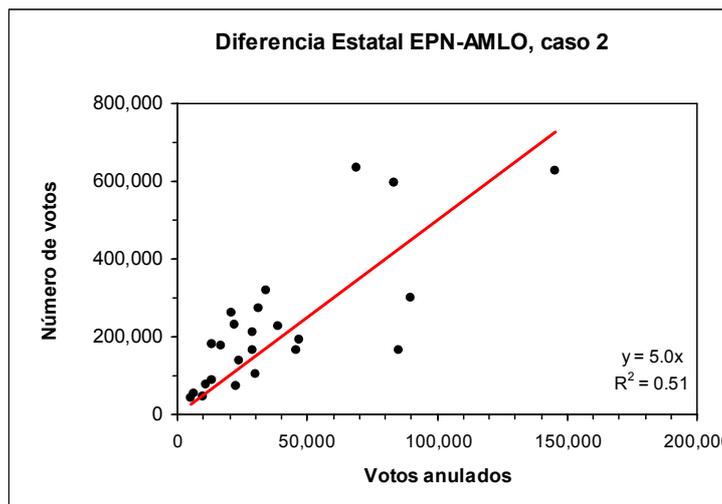


Figura 4. Mismos datos de la Figura 3 pero sólo con estados donde EPN superó a AMLO. La curva fue ajustada a través del origen.

La Figura 4 es una sección de la Figura 3 que se construyó únicamente con los estados donde la votación de EPN fue mayor que la de AMLO. Básicamente se intensifica la tendencia favorable a EPN observada en la Figura 3. La pendiente en esta figura indica que en los estados ganados por este candidato hubo cinco votos a su favor por cada voto anulado y que la variación de la diferencia EPN-AMLO explicada por la anulación aumentó de 34 a 51%. La dispersión de los puntos alrededor de la línea de tendencia permite hipotetizar que la operación a favor de EPN fue controlada con más o menos éxito a nivel estatal, dependiendo quizá también de la densidad de la población y la fuerza de AMLO en esas entidades.

Estadísticas de los votos nulos

La Tabla 1 resume los votos anulados por estados y distritos. Se anuló 2.5% de la votación nacional. Hubo un aumento de cuatro décimas porcentuales respecto a los anulados en la elección del 2000 (2.1%) y de tres décimas en comparación con las del 2006 (2.2%), según datos de Nacif (2009). Las diferencias en los niveles de anulación promedio y mediana entre estados y distritos simplemente reflejan las diferencias en tamaño de poblacional y número de electores entre estas subdivisiones geográfico-políticas. En ambos casos la anulación promedio fue mayor que la mediana, lo cual indica que la distribución de frecuencias de votos anulados es asimétrica y sesgada hacia la derecha. En estos casos, el promedio geométrico representa mejor los datos que el promedio común. El mínimo y máximo de anulación distrital se presentaron en Tijuana (Distrito BC5) y Tamazunchale (distrito SLP7), respectivamente. La diferencia entre estos distritos fue de 8,606 votos nulos, un intervalo relativamente amplio para poder contrastar los patrones de votación de los candidatos en relación con esta variable. La variabilidad estatal de los nulos, expresada como coeficiente de variación (CV), fue 1.5 veces mayor que la distrital. Esto implica que hubo más homogeneidad en los datos distritales que en los estatales. Si se recuerdan los coeficientes de variación de los candidatos en la Figura 2, la variación de 33% en los distritos se acerca a un promedio entre el bajo CV de EPN (24%) y los de los otros candidatos (43 y 46%). Así, la baja variación de EPN tuvo un efecto cuantitativo en toda la elección.

Tabla 1. Comparación de votos anulados entre Estados y distritos en la elección presidencial de julio 2012*.

	Estados (32)	Distritos (300)
Total	1,241,154	1,241,154
Promedio	38,786	4,137
Promedio geométrico	28,352	3,952
Mediana	29,082	3,837
Mínimo	5,180, BCS	2,263, BC5, Tijuana
Máximo	145,605, EdoMex	10,869, SLP7, Tamazunchale
CV (%)	83	33

*Cifras y cálculos a partir de base de datos del IFE obtenida de su página de Internet el 7 de julio, 2012.

Dado que el total de votos nulos en esta elección es considerablemente menor que la diferencia oficial al 7 de julio entre EPN y AMLO (3.33 millones de votos), es necesario subrayar que no es propósito de este análisis determinar si alguna porción de esos votos corresponden en realidad a algún candidato, sino encontrar signos cuantitativos de una operación electoral a gran escala reflejados en ellos. Todo parece indicar que así fue y que es posible hacer una estimación razonable del monto total de esa operación en votos absolutos. La cantidad resultante, como se verá en la sección de regresión, es mucho mayor que la de votos anulados en sí y que la diferencia entre EPN y AMLO.

Análisis de correlación

El examen de los datos distritales empezó con una revisión de las distribuciones de frecuencia y la dispersión de las variables proporcionadas por el IFE. Se utilizaron gráficas y pruebas de rutina para evaluar la normalidad de los datos (Kolmogorov-Smirnov y de Lilliefors), se calcularon sus parámetros de centralidad y dispersión y se aplicaron técnicas gráficas para visualizar las correlaciones entre los tipos de votos, entre los candidatos y los votos. Se produjeron más materiales de estas pruebas, pero se presentan las más relevantes. Están disponibles para el interesado en revisar la veracidad de lo aquí expuesto o la corrección o incorrección de las técnicas aplicadas.

Las Figuras 5, 6 y 7 dan cuenta de lo anterior. La Figura 5 muestra las correlaciones de los tipos de votos entre sí (totales, válidos y nulos). En los diagramas de barras de esta figura se aprecia los votos totales y los válidos tienen una distribución aproximadamente normal, que se expresa en su forma de campana. Los votos nulos se distribuyeron de una manera diferente: la mayoría de los distritos presentaron niveles bajos de anulación (entre 2 y 4 mil votos) y unos pocos con anulación alta (8 y 9 mil) o muy alta (>10 mil). La distribución lognormal se aproxima a la de estos votos. Las gráficas de puntos en esta figura son de correlación y los números insertos en ellas son el índice de correlación de Pearson (r). Todas estas correlaciones fueron significativas ($p < 0.05$). La alta correlación entre válidos y totales ($r = 0.999$) no sorprende; significa que la cantidad de votos válidos por distrito dependió directamente de los votos emitidos. Sin embargo, esta relación casi perfecta disminuyó dramáticamente al correlacionar los totales y los válidos con los anulados, $r = 0.21$ y 0.16 , respectivamente. Esta disminución de r se visualiza en las muy dispersas nubes de puntos azules en las gráficas de los nulos. Este es un primer indicio de que la variación de estos votos tuvo características muy diferentes a los otros votos. Este fue el hilo seguido para adentrarse en el posible papel de la variabilidad de estos votos en los resultados electorales. El siguiente paso fue analizar la correlación de estos votos con cada candidato.

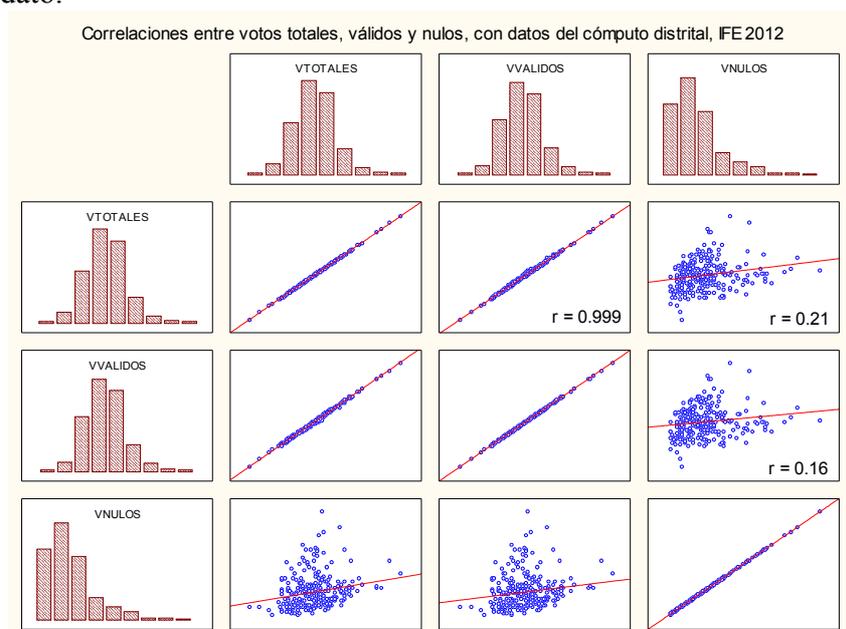


Figura 5. Correlaciones distritales entre votos totales emitidos (VTOTALES); votos válidos VVALIDOS, que se calcularon como la diferencia entre totales y nulos (VNULOS). Los números en las gráficas son el índice de correlación de Pearson. Todos son significativos ($p < 0.05$).

En la Figura 6 se muestran las correlaciones entre los candidatos y entre ellos con los votos nulos. La primera característica notable en esta figura es que la distribución de los votos de los candidatos no es normal y que es más o menos similar a la de los nulos. Más adelante se retoma este punto porque obligó a una transformación de los datos originales al aplicar un modelo de regresión. El segundo aspecto es la forma de las nubes de puntos en cada correlación y el índice que las resume (r de Pearson). Si la nube es circular, como aproximadamente se observa en la relación entre JVM con los nulos (indica que no hay correlación, que $r = 0$ o no significativamente diferente de cero). En este caso particular, r fue igual a -0.03 . Si la nube es alargada y consistentemente ascendente hacia la derecha, la correlación es positiva); si es alargada descendente, negativa. Las líneas rojas en medio de las nubes de puntos representan la tendencia promedio, son una versión simplificada de estas nubes y facilitan visualizar si las correlaciones son de un tipo u otro, o si no existen. Todo es expresado por el índice de correlación de Pearson, que resume en un solo número la fuerza y sentido de cada correlación. Como se sabe, este índice varía entre -1 (relación completamente negativa) y $+1$ (correlación completamente positiva). No obstante, siempre es recomendable visualizar las relaciones entre variables con gráficas de este tipo para una mejor interpretación del índice r .

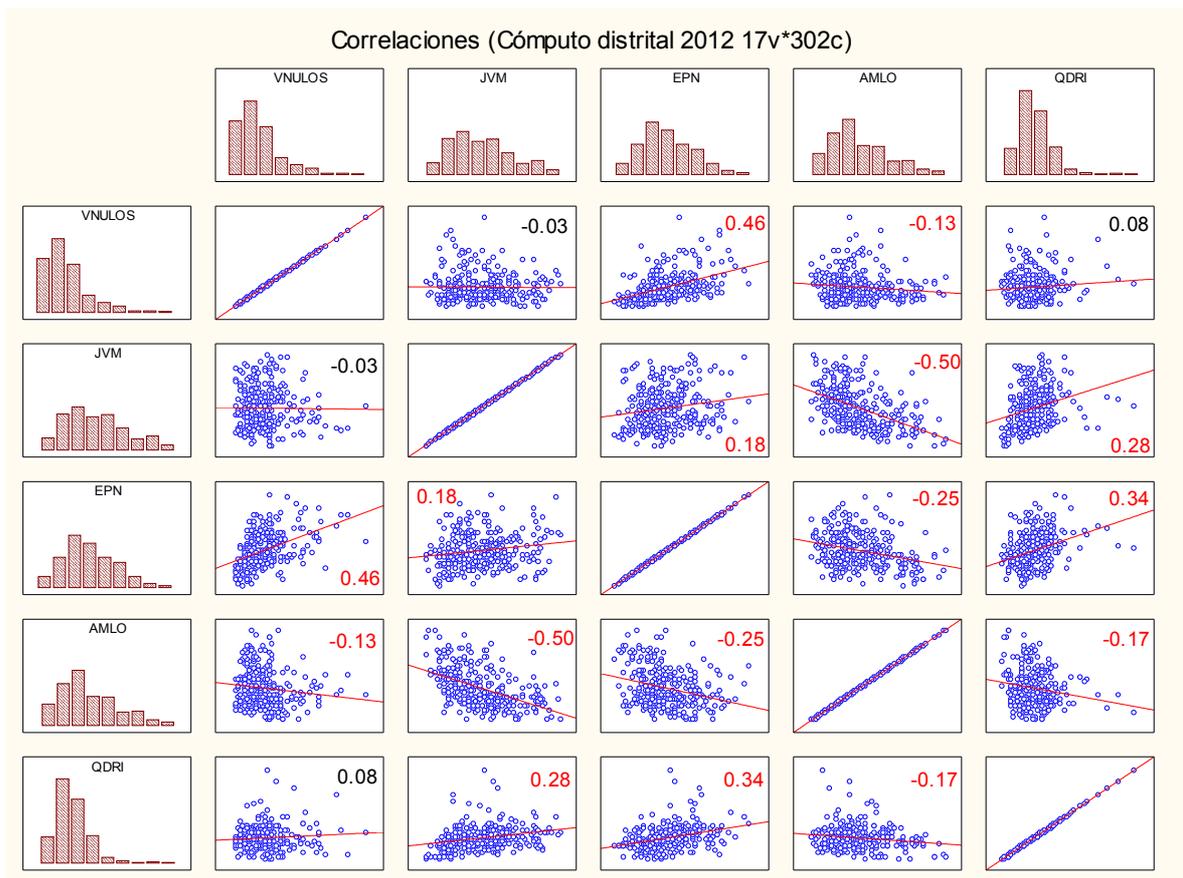


Figura 6. Correlaciones distritales entre candidatos y votos nulos. Los números dentro de las gráficas son el índice de correlación de Pearson (en rojo, significativos, $p < 0.05$; en negro, no significativos).

La apariencia complicada de las correlaciones en la Figura 6 se puede resumir en las siguientes observaciones:

- Ni la votación de JVM ($r = -0.03$) ni la de GQT ($r = 0.08$) tuvieron relación significativa con los votos anulados. Esto se manifiesta en sus líneas de tendencia planas e índices de correlación con los votos nulos cercanos a cero. En otras palabras, la anulación no les afectó ni para bien ni para mal y mantuvieron su preferencia general entre el electorado independientemente de la anulación distrital. Son dos casos interesantes, especialmente el de JVM, porque remiten a la racionalidad y pertinencia de poner a prueba la hipótesis de neutralidad de los votos anulados.
- El caso de AMLO es parecido al de JVM y GQT, aunque su línea de tendencia tiene una caída significativa con el incremento de la anulación ($r = -0.13$, $p < 0.05$). Esto significa que la anulación no fue neutral con AMLO y que este candidato perdió votos con la anulación. Este punto se retoma en la siguiente sección para determinar su posible monto.
- Las correlaciones muestran que EPN fue el único de los candidatos con una relación positiva relativamente alta con los niveles de anulación ($r = 0.46$; $p < 0.05$). La neutralidad de votos los nulos también falló con él, pero en el sentido de favorecerlo. Este punto también se retoma en la siguiente sección para determinar su posible monto.
- Es llamativo en la Figura 6 que todas las correlaciones entre JVM, EPN y Quadri fueron, con más o menos fuerza, positivas y significativas. Implica que sus votos, en general, crecieron y disminuyeron juntos. Pero todos ellos tuvieron correlaciones negativas con AMLO. Cuando a este candidato le aumentó la votación, a ellos les disminuyó y viceversa. La correlación más negativa entre dos candidatos fue entre JVM y AMLO ($r = -0.50$; $p < 0.05$), un indicador de que las posturas políticas de sus electores están bien definidas y son muy contrapuestas entre sí.

La Figura 7 es una representación más sencilla de las relaciones entre los candidatos y votos nulos. Esta especie de árbol genealógico se elaboró con la conocida técnica de análisis de grupos o cúmulos (*Cluster analysis*) y con los índices de correlación de Pearson como criterio de agrupación. AMLO, quien como se acaba de ver tuvo índices de correlación negativos con todos los candidatos y con los votos anulados, aparece aquí como una rama separada de los otros candidatos. Esto es consistente con el planteamiento que él mencionó en repetidas ocasiones en su campaña: que en estas elecciones había sólo dos opciones diferentes, la representada por él y la de los demás candidatos. Quadri también llegó a externar esta idea.

En esta misma figura se destaca con un círculo rojo la estrecha asociación entre EPN y los votos nulos. Esta relación de EPN fue más fuerte que con los otros candidatos y con los votos totales. Este es un indicio claro de que la coalición PRI-PVEM jugó la carta de la anulación como una de las vías para ganar la elección. Pero lo hizo con tal fuerza y desparpajo con otros mecanismos de agenciarse votos que su accionar quedó irremediabilmente grabado en la variación de los votos nulos. En otras palabras, la variación distrital de los votos nulos es una representación a escala de la operación del PRI-PVEM a nivel de distritos. En aquellos donde hubo más votos anulados, operó con más fuerza la maquinaria de esta coalición e incrementó su promedio de votos. Donde hubo menos votos anulados, EPN tuvo sus niveles de votación más bajos. Ninguno de los otros contendientes parece haber actuado así, al menos con tal intensidad.

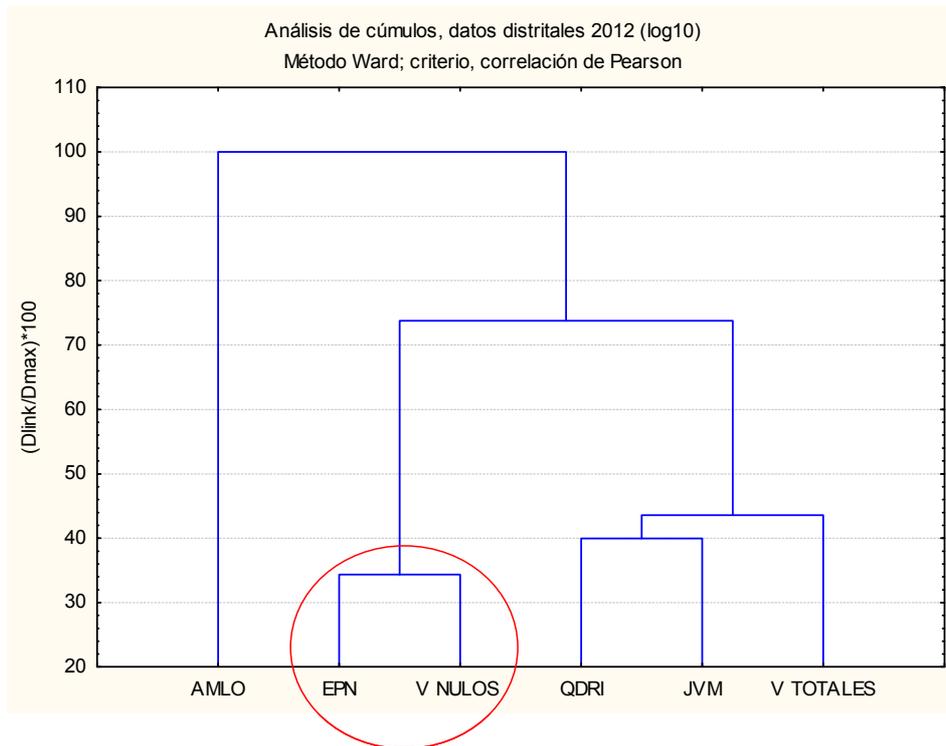


Figura 7. Relaciones entre candidatos y tipos principales de votos. Árbol hecho a partir de sus correlaciones con los datos logarítmicos y análisis de grupos.

Análisis de regresión

Se hizo un ajuste lineal mediante regresión simple a los votos de los candidatos como función de los votos anulados. Esta técnica de análisis define con más exactitud y precisión la correlación entre dos variables, posibilita estimar el error de sus dos parámetros principales y permite estimar votos adicionados o descontados a partir de la ecuación resultante. Se utilizó el modelo clásico de regresión simple basado en la ecuación de la línea recta. Intentos preliminares de aplicación de este modelo a los datos originales (votos sin transformar) o transformando sólo una de las variables produjeron resultados inaceptables por no cumplir con los supuestos del modelo, básicamente la normalidad de datos y de los residuos del ajuste, así como la homogeneidad de la varianza. El modelo se aplicó finalmente a los datos transformados a escala logarítmica, lo que produjo ajustes lineales más satisfactorios. El modelo se sintetiza así:

$$Y = \beta_0 + \beta_1 x$$

Donde Y representa los votos de cada candidato, β_0 es la ordenada al origen o el número de votos obtenido cuando los nulos son cero (lo cual no ocurrió a nivel distrital) o uno en este caso, por la transformación logarítmica. En condiciones de neutralidad absoluta de los votos nulos la ordenada correspondería al promedio de votación distrital de los candidatos. Como se verá más adelante, los votos de JVM y los de Quadri se aproximaron a esto. El parámetro β_1 es la pendiente de la recta, que define el incremento de votos de los candidatos en función de los nulos. Este es el parámetro

clave para decidir si la hipótesis de neutralidad de los votos nulos se cumple o no, así como para determinar el ritmo de una posible ganancia o pérdida de votos. Esta hipótesis establece que los votos de los candidatos no aumentan ni disminuyen significativamente en relación con los votos nulos. En un sentido más formal, la hipótesis puesta a prueba fue:

$$\beta_{1JVM} = \beta_{1EPN} = \beta_{1AMLO} = \beta_{1QDRI} = 0$$

Los resultados principales de este análisis se resumen en las Tablas 3 y 4. En la primera de estas se presentan las pruebas de las sumas de cuadrados de este análisis por candidato y se añaden los parámetros β_0 y β_1 de las curvas de ajuste de cada candidato con sus respectivos intervalos de confianza del 95%. El modelo fue significativo sólo para EPN y AMLO ($p = 0.00$ y 0.02 , respectivamente); es decir, $\beta_1 \neq 0$ (Figura 8). Por lo tanto, para estos candidatos se rechaza la hipótesis de neutralidad de los votos nulos. En cambio, la hipótesis se sostiene para JVM y Quadri, a quienes la anulación no los afectó ni para bien ni para mal.

El modelo utilizado se ajustó mejor a los datos de EPN ($R^2 = 0.25$) que a los de AMLO ($R^2 = 0.02$). El precio estadístico de esta situación se paga con un intervalo de confianza de la pendiente de AMLO del doble del tamaño de la de EPN (Figura 8). Llama la atención también que la varianza no explicada por el modelo (SC residuos) es de 4.5 a seis veces menor en EPN en comparación con sus contendientes. Esta diferencia es mucho más grande que la expresada por los coeficientes de variación obtenidos directamente de los datos reales (ver Figura 2). Esto ilustra la mejor resolución numérica de la técnica de regresión. La escasa variación interdistrital de EPN sugiere un que hubo un control sobre su serie de votos en la base de datos publicada por el IFE. Sin embargo, determinar en cuál o cuáles pasos del proceso electoral pudieron haber operado esos controles va mucho más allá de los alcances de este análisis.

Tabla 3. Resumen de las pruebas de la suma de cuadrados (SC) del modelo vs. la SC de los residuos de la regresión de los votos distritales de los candidatos vs. los votos anulados. Este análisis se hizo a los datos transformados logarítmicamente (\log_{10}). Se añaden aquí la ordenada (β_0) y la pendiente (β_1) de las curvas de ajuste \pm sus intervalos de confianza del 95%. GL, grados de libertad.

CANDIDATO	GL	SC modelo	SC residuos	F	<i>p</i>	R ² múltiple	β_0	β_1
JVM	298	0.005	0.048	0.10	0.75	0.00	4.47±0.706	0.03±0.196
EPN	298	0.768	0.008	97.04	0.00	0.25	3.37±0.285	0.40±0.079
AMLO	298	0.192	0.036	5.39	0.02	0.02	5.40±0.605	-0.20±0.168
QDRI	298	0.004	0.043	0.09	0.77	0.00	3.44±0.664	0.03±0.184

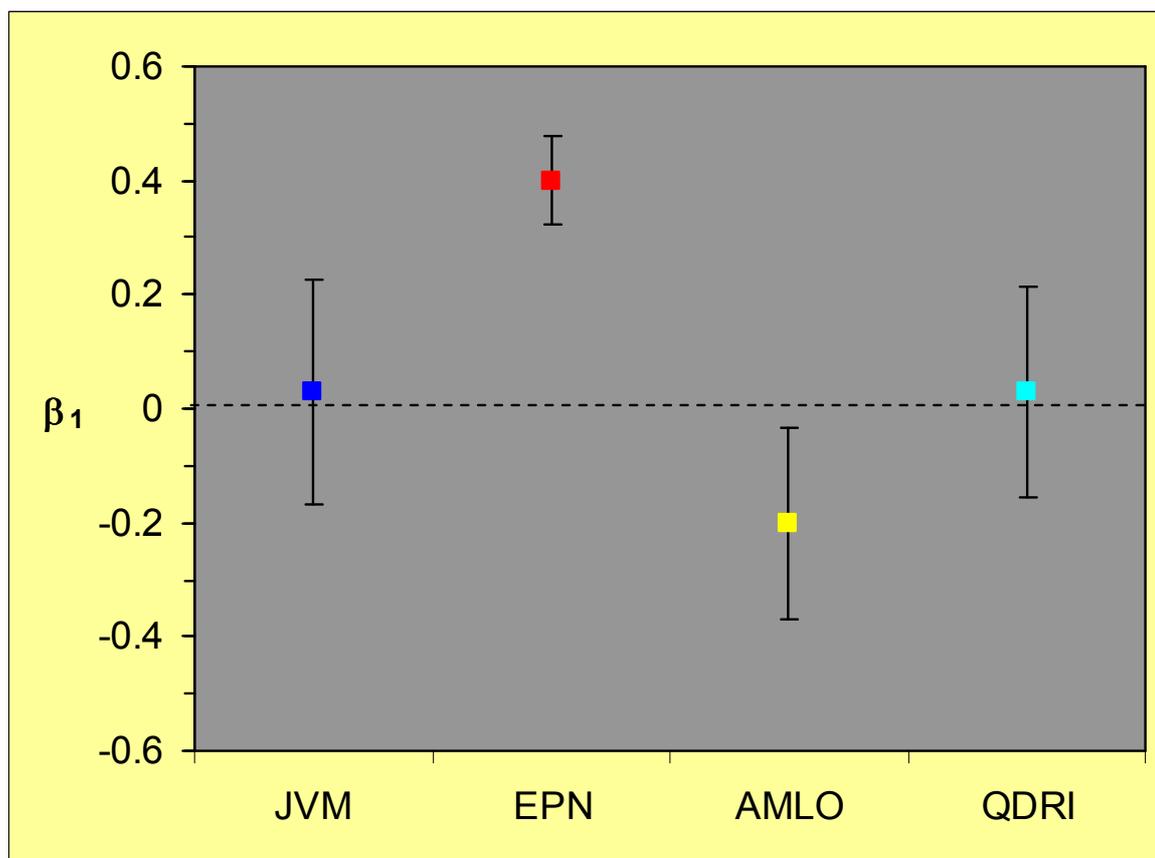


Figura 8. Valor e intervalo de confianza (IC95%) del parámetro β_1 (pendiente) de las rectas de ajuste de los votos distritales de los candidatos vs. los votos nulos. Las pendientes cuyo IC95% incluye el valor cero (JVM y QDRI), no son significativamente diferentes de este valor. Esto significa que la anulación no los afectó. Para los otros dos candidatos, hay elementos para rechazar la hipótesis de neutralidad de los votos. A EPN, con pendiente positiva (0.40) la anulación lo benefició; a AMLO, con pendiente negativa (-0.20), lo perjudicó. De manera interesante, el IC95% de EPN es la mitad del de los otros candidatos; otra expresión del control sistemático ejercido sobre su votación.

La Tabla 4 muestra las ecuaciones producidas por el análisis de regresión en escala logarítmica y en la escala original de los datos (votos). Se incluyen las ecuaciones de JVM y Quadri, aunque no fueron significativas ($\beta_1 = 0$). La pendiente positiva de EPN (0.40) y la negativa de AMLO (-0.20) son la expresión cuantitativa del efecto de la anulación, y posiblemente de otros factores relacionados con ésta, en los votos distritales de estos candidatos. El tamaño de la pendiente de EPN es la medida de la adición artificial de votos a su favor: $40 \pm 8\%$; mientras que a AMLO le significó una sustracción de votos del orden de $20 \pm 17\%$. El significado en votos se puede calcular aplicando estos porcentajes a las sumas totales reportadas por el IFE (07-07-2012): EPN, 19,226,784 votos; AMLO, 15,896,999. Redondeando, EPN recibió 7.7 ± 1.5 millones de votos no libres ni (tan) secretos. A AMLO le faltan 3.2 ± 2.7 millones de votos. Como se observa, este margen de error es muy grande, va de 500 mil a 5.9 millones de votos y se debe a la alta variabilidad de la votación distrital de este candidato. Estos cálculos implican que la suma de votos limpios de EPN estaría entre 10 y 13 millones, con 11.5 millones en promedio, cifra menor a la de JVM (12.8 millones

de votos). La de AMLO estaría, añadiendo el faltante entre 16.4 y 21.8 millones de votos, con un promedio de 19.1 millones de votos, cifra muy similar a la reconocida por el IFE a EPN.

Tabla 4. Ecuaciones del modelo de regresión aplicado con los datos en escala logarítmica. Las ecuaciones a escala real se obtuvieron por conversión antilogarítmica.

Escala logarítmica	Escala real
$\text{Log}_{10}(\text{votos JVM}) = 0.03[\text{log}_{10}(\text{votos nulos})] + 4.47$	$\text{JVM} = \text{votos nulos}^{0.03} / 0.000034$
$\text{Log}_{10}(\text{votos EPN}) = 0.40[\text{log}_{10}(\text{votos nulos})] + 3.37$	$\text{EPN} = \text{votos nulos}^{0.40} / 0.00043$
$\text{Log}_{10}(\text{votos AMLO}) = -0.20[\text{log}_{10}(\text{votos nulos})] + 5.40$	$\text{AMLO} = \text{votos nulos}^{-0.20} / 0.000004$
$\text{Log}_{10}(\text{votos QDRI}) = 0.03[\text{log}_{10}(\text{votos nulos})] + 3.44$	$\text{QDRI} = \text{votos nulos}^{0.03} / 0.00036$

En un ensayo previo con el modelo de regresión lineal aplicado directamente a los datos (votos) en bruto, los resultados fueron muy parecidos y más fáciles de interpretar, pero el ajuste fue ligeramente menos satisfactorio en cuanto a las suposiciones del mismo (Tabla 5). Las pendientes de JVM y GQT tampoco difieren de cero y se concluye que no fueron afectados por la anulación. Las pendientes de EPN y AMLO también fueron diferentes de cero y se concluye que sus votos distritales fueron sensibles a la anulación. De acuerdo con estos resultados, a EPN, con $\beta_1 = 5.05$, se le suman cinco votos por cada voto anulado, mientras que a AMLO ($\beta_1 = -2.1$) se le descuentan 2.1 votos por cada anulado. Esto conduce a un incremento artificial de 32.6% (6,627,555 votos) a la cuenta total de EPN. Por consiguiente, sus votos limpios en total serían 12,959,229. Por el lado de AMLO, hay un descuento de 16.5% (2,618,721 votos) respecto al total que le reconoce el IFE, por lo que su total debiera ser de 18,515,720 votos. Aunque las cifras son menores, las conclusiones que se extraen de este ejercicio son las mismas que con el modelo logarítmico. En votos limpios, AMLO aventajó con 5,556,491 a EPN.

Tabla 5. Parámetros de las líneas de regresión simple aplicadas directamente a los datos distritales del IFE.

Candidato	β_0			β_1			Promedio distrital IFE
	-95%CI	Promedio	+95%CI	-95%CI	Promedio	+95%CI	
JVM	36,213	43,309	50,406	-1.80	-0.17	1.46	42,622
EPN	38,318	43,212	48,105	3.92	5.05	6.17	64,089
AMLO	53,546	61,737	69,927	-3.99	-2.11	-0.23	52,990
QDRI	2,744	3,385	4,027	-0.038	0.11	0.256	3,836

Es interesante observar en esta tabla que cuando se cumple la hipótesis de neutralidad de los votos nulos para un candidato, la ordenada al origen de su curva es muy similar su promedio distrital; por ejemplo, la ordenada para JVM es de 43,309 votos y su promedio distrital fue 42,622. En el caso de Quadri, la ordenada fue de 3,385 y su promedio distrital fue 3,836. Estas diferencias se deben a que sus pendientes no fueron exactamente iguales a cero. No obstante, los promedios reales de estos candidatos se encuentran dentro del intervalo de confianza cuantificado por la regresión. En contraste,

cuando no hubo neutralidad de la anulación, la ordenada y el promedio de los candidatos son muy distintos y su diferencia expresa el sentido del efecto: el promedio distrital de EPN a partir de los datos oficiales (64,089 votos) es muy superior a su ordenada (43,212 votos, curiosamente casi igual a la de JVM) y se encuentra fuera y muy por arriba de su intervalo de confianza. En el caso de AMLO, su promedio distrital (52,990 votos) es inferior a su ordenada (61,737), y también queda fuera y debajo del intervalo de confianza de la ordenada. Estos son indicios estadísticos fuertes e irrefutables de que la votación presidencial fue distorsionada a favor de EPN y en contra de AMLO.

Como un ejercicio adicional, si se asume que la ordenada de la ecuación de EPN representa su promedio real de votos limpios, se puede calcular su total de votos: $43,212 \times 300$ distritos = 12,963,300 votos. Para AMLO: $61,737 \times 300 = 18,521,100$ votos totales. Si el ejercicio se hace más conservadoramente con los límites superior de EPN e inferior de AMLO, el resultado es 14,431,500 votos para EPN y 16,063,800 para AMLO; es decir, este candidato volvería a ganarle a EPN con poco más de 1.6 millones de votos.

La Figura 9 muestra dos representaciones gráficas de los modelos ajustados a los datos electorales de los dos candidatos afectados por la anulación. Se destacan las tendencias positiva de EPN y negativa de AMLO con sus respectivos intervalos de confianza. De acuerdo con ellas, la derrota de EPN y el triunfo de AMLO parecen estadísticamente imposibles, como si hubieran sido fijados de antemano. Sin embargo, los resultados de este trabajo muestran que ambas tendencias contienen fuertes indicios de haber sido inducidas artificialmente, a fuerza de adicionar a EPN votos ajenos a las definiciones del artículo 41 constitucional y de restarle votos a AMLO. La adición artificial de votos a EPN fue de tal magnitud que terminó por deformar la naturalidad de su variación. Esto dejó en la base oficial de datos un rastro clave para cuantificarla.

Sobre la base de estos análisis se puede considerar que los patrones de variación de JVM, AMLO y Quadri derivaron básicamente del proselitismo de sus campañas. Esto se puede afirmar a pesar de que localizadamente alguno de ellos haya tenido prácticas desleales. Por ejemplo, se observa que Quadri tuvo niveles de votación extremadamente alta con respecto a su tendencia general en distritos de Aguascalientes, Chiapas, Chihuahua, Hidalgo, Nuevo León y Tamaulipas, entre otros. No obstante, estos eventos no lograron impactar significativamente su patrón de votación. En el caso de EPN se trata de un operativo de adquisición de votos a gran escala, abarcando todo el país aunque con mayor intensidad en algunos distritos identificables.

Estos resultados muestran que es posible estimar la cantidad de votos añadidos artificialmente mediante una prueba de hipótesis, la cual reveló aspectos de la estructura de los datos oficiales que no son fácilmente distinguidos ni entendidos por el ciudadano común al que se debe servir con honestidad. El rechazo de la hipótesis de neutralidad de los votos nulos para los dos principales contendientes significa que el proceso de elección no transcurrió de acuerdo con las especificaciones del COFIPE en cuanto a los criterios para anular votos y que esta falta afectó significativamente el resultado oficial al intercambiar los lugares de los dos principales candidatos. Este tipo de pruebas pueden ser instrumento importante de validación de votos en procesos electorales futuros. La otra gran ventaja de esta prueba es que permite no sólo identificar el sentido

del efecto de la anulación (a favor o en contra) sino cuantificarlo en votos. Esto no implica que la anulación de votos sea la única causa, ni mucho menos la principal, de lo que parecen votos en exceso para el candidato del PRI. Varios otros factores como los que se han mencionado en los medios pueden haber contribuido a ello. En cierto sentido, todos estos quedaron reflejados en la variación de los votos nulos: a más anulados en un distrito, más intervención del PRI-PVEM en sus diferentes manifestaciones y más votos para EPN; a menos anulados, quizá por más vigilancia de los contrincantes, menos votos para el presunto ganador. Este fue un rasgo principal de esta elección.

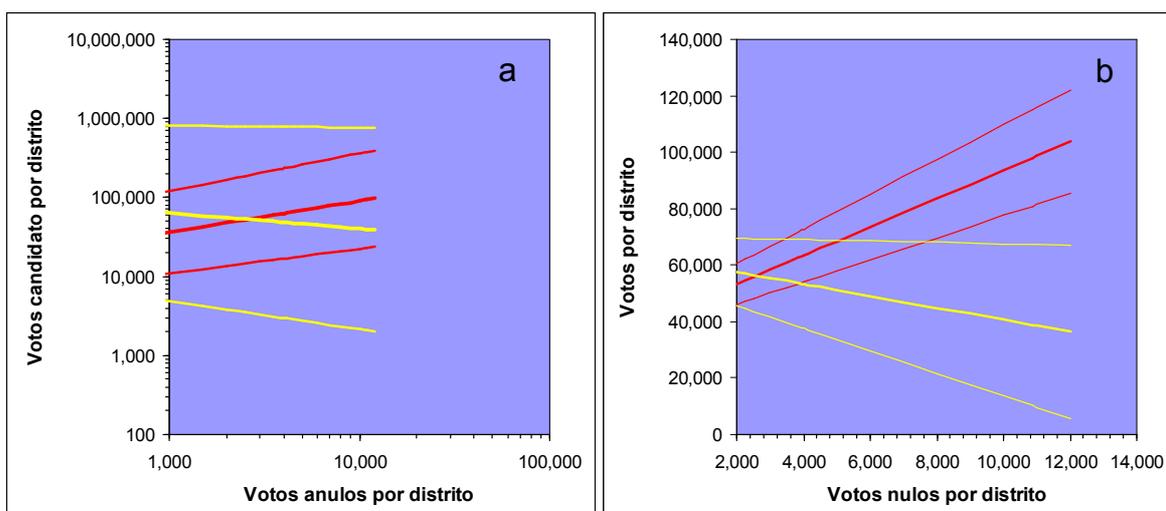


Figura 9. Representación del modelo de regresión aplicado a los datos distritales de EPN y AMLO: a) versión logarítmica; b) versión con datos sin transformar.

Es lamentable que el IFE, con todos los recursos humanos y materiales de que dispone, no haya hecho, como parece, un análisis numérico a fondo de los datos distritales antes de declarar un presunto ganador. Pasó como buenas sumas de votos a partir de datos estructuralmente cuestionables, especialmente los de EPN. México se ahorraría inmensos problemas de falta de transparencia electoral, gastos multimillonarios innecesarios y tensión política si se institucionalizaran prácticas de validación de datos. Con las herramientas modernas de cómputo y algo de claridad sobre cómo proceder, sería posible hacer estos análisis en cuestión de minutos, antes de sorprender al país los primeros domingos de Julio de cada seis años por la noche con resultados increíbles.

Conclusiones

1. La votación de EPN es la única con indicadores significativos de distorsión a su favor.
2. La suma oficial de votos de EPN está potencialmente inflada en alrededor de 40%, casi 7.7 millones de votos con error que va de 6.2 a 9.2 millones.

3. El IFE avaló la votación de EPN sin una revisión a fondo de sus características numéricas.
4. Hay evidencia numérica de que a AMLO le desaparecieron alrededor de 3.2 millones de votos. Es enteramente posible señalar cuáles fueron los distritos más probables donde se hizo esta sustracción.
5. El próximo fallo sobre la elección por parte del Tribunal Electoral del Poder Judicial de la Federación (TEPJF) no debería sustentarse en datos estadísticamente dudosos.
6. La demanda de anulación de toda la elección presidencial es exagerada, porque sólo un candidato tiene el patrón de votación anómalo, EPN. Por consiguiente, la demanda y el veredicto justos serían la descalificación del presunto ganador.
7. En contraste, la votación de los demás candidatos, aunque no se pasa por alto que hubo distritos con votación extrañamente alta para algunos de ellos, fue en términos generales consistentemente aleatoria; es decir, con una variación propia de votos obtenidos por proselitismo partidario y que refleja genuinamente las preferencias del electorado.
8. Estos resultados se añaden a las evidencias documentales y argumentos que ponen en duda el resultado de la elección presidencial. Sobre esta base numérica y las previstas en la ley electoral y la constitución, es enteramente razonable exigir que la autoridad resuelva a favor de una revisión minuciosa de todo el proceso electoral tocante a la elección presidencial.

Referencias

Nacif H. B. 2009. Haz valer tu voto, no lo anules. http://pac2009.ife.org.mx/boletin07_n1.html

COFIPE

Artículo 274

1. El escrutinio y cómputo es el procedimiento por el cual los integrantes de cada una de las mesas directivas de casilla, determinan:

- a) El número de electores que votó en la casilla;
- b) El número de votos emitidos en favor de cada uno de los partidos políticos o candidatos;
- c) El número de votos **nulos**; y
- d) El número de boletas sobrantes de cada elección.

2. *Son votos nulos:*

a) *Aquel expresado por un elector en una boleta que depositó en la urna, sin haber marcado ningún cuadro que contenga el emblema de un partido político; y*

b) *Cuando el elector marque dos o más cuadros sin existir coalición entre los partidos cuyos emblemas hayan sido marcados;*

3. *Cuando el elector marque en la boleta dos o más cuadros y exista coalición entre los partidos cuyos emblemas hayan sido marcados, el voto contará para el candidato de la coalición y se registrará por separado en el espacio correspondiente del acta de escrutinio y cómputo de casilla.*

4. *Se entiende por boletas sobrantes aquellas que habiendo sido entregadas a la mesa directiva de casilla no fueron utilizadas por los electores.*

Fe de erratas y correcciones a la primera versión de este documento.

Se hicieron las siguientes correcciones (sin alterar contenidos de la primera versión):

Página 2. Decía: *de menor a menor* (en el pie de la Figura 1), se cambió a *de menor a mayor*.

Página 7. Se agregó la frase: *En estos casos, el promedio geométrico representa mejor los datos que el promedio común.*

Página 7: Decía: Si se recuerdan los coeficientes de variación de los candidatos en la *Figura 1*; la referencia correcta es a la *Figura 2*.

Página 8: Decía (en pie de la Figura 5): Todos significativos ($p < 0.05$); se cambió a *Todos son significativos ($p < 0,05$).*

Página 9: se editó la Figura 6 para mejorar visibilidad. En la versión anterior la *r* de Quadri en la figura de la esquina inferior izquierda aparecía equivocadamente en rojo (significativo); se cambió a negro (no significativa). Al pie de figura se le agregó la aclaración: *(en rojo, significativos, $p < 0.05$; en negro, no significativos).*

Página 10: Decía: Esto significa la anulación no fue neutral con AMLO; se cambió a: Esto significa *que* la anulación no fue neutral con AMLO.

Página 10: Decía: Esto es consistente con el planteamiento él mencionó...; se cambió por: Esto es consistente con el planteamiento *que* él mencionó...

Página 10: Decía: Pero lo hizo con tal fuerza y desparpajo con otros mecanismos de agenciarse votos su accionar...; Se cambió por: Pero lo hizo con tal fuerza y desparpajo con otros mecanismos de agenciarse votos *que* su accionar...(uff, por lo visto esta fue la página de los ques faltantes).

Página 15: Decía: ...es inferior a su ordenada (61,737), que también queda fuera y debajo del intervalo de confianza de la ordenada. Se cambió por: ...es inferior a su ordenada (61,737), y también queda fuera y debajo del intervalo de confianza de ésta.

Página 15: Decía: La otra ran ventaja...; se cambió por: La otra gran ventaja...

Página 15: Decía: ...la cual revelo aspectos... Se cambió por: ... la cual *reveló* aspectos...

Con una disculpa del autor
Ángel Zambrano

Erratas adicionales encontradas en la versión B. Aparecen corregidas en la versión C y última. Tampoco modifican cálculos, interpretaciones ni conclusiones de la versión anterior.

Página 9: Decía: En este caso particular, r fue igual a -0.01 . Se cambió por “En este caso particular, r fue igual a -0.03 “. Lo cual la hace coincidir con lo que dice la r de JVM en la Figura 6, que es el valor correcto (esto no cambia el significado). Esto mismo también se corrigió en el primer punto de la página 10.

Página 11: Decía: ...permite estimar votos adicionados a partir de...Se cambió por: ...permite estimar votos adicionados *o descontados* a partir de...

Página 13: En el pie de la Figura 8 decía: A EPN, con pendiente positiva (0.04); se cambió por: A EPN, con pendiente positiva (0.40). Este es el valor correcto que aparece en todo el texto y las tablas.

Página 14: En la Tabla 5, los encabezados β_1 y β_0 estaban invertidos, ya se corrigió.

Página 15: Decía: Se destacan las tendencias positiva de EPN y AMLO. Se cambió por: Se destacan las tendencias positiva de EPN y *negativa de* AMLO.

Con una nueva (y última) disculpa

Ángel Zambrano