

Los Estados  
en 2021

Elecciones  
en pandemia

Rosa María Mirón Lince  
Coordinadora



LA BIBLIOTECA

# LOS ESTADOS EN 2021. ELECCIONES EN PANDEMIA

*Rosa María Mirón Lince*  
*Coordinadora*



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO



LA BIBLIOTECA

México, 2023

Esta investigación, arbitrada a “doble ciego” por especialistas en la materia, se privilegia con el aval de la Facultad de Ciencias Políticas y Sociales, Universidad Nacional Autónoma de México.

Este libro fue financiado con recursos de la Dirección General de Asuntos del Personal Académico de la Universidad Nacional Autónoma de México, mediante el proyecto “Calidad de la democracia y sus factores determinantes en México 2012-2020”, a cargo de la Dra. Rosa María Mirón Lince, como parte del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) IN305418.

**Los estados en 2021.**  
**Elecciones en pandemia**  
*Rosa María Mirón Lince*  
*(coordinadora)*

Primera edición: 29 de marzo de 2023

D.R. © 2023 Universidad Nacional Autónoma de México  
Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, CDMX, Facultad de Ciencias Políticas y Sociales, Circuito Mario de la Cueva s/n, Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, CDMX.

D.R. © 2023 Ediciones La Biblioteca, S.A. de C.V.  
Azcapotzalco la Villa No. 1151  
Colonia San Bartolo Atepehuacán  
Alcaldía Gustavo A. Madero  
C.P. 07730, México, CDMX.  
Tel. 55-6235-0157 y 55-3233-6910  
Email: contacto@labiblioteca.com.mx

**ISBN UNAM: 978-607-30-7422-3**  
**ISBN Editorial: 978-607-8916-03-0**

Diseño: Fernando Bouzas Suárez  
Revisión editorial: Samuel Mendoza

Queda prohibida la reproducción parcial o total, directa o indirecta, del contenido de la presente obra, sin contar previamente con la autorización expresa y por escrito de los editores, en términos de lo así previsto por la Ley Federal del Derecho de Autor y, en su caso, por los tratados internacionales aplicables.

Impreso y encuadernado en México  
*Printed and bound in Mexico*

# EL PAPEL DE LAS ENCUESTAS EN LAS ELECCIONES FEDERALES DE MÉXICO, 2021

*Ricardo de la Peña\**

*¿Alguna vez ha tenido la sensación de que la única razón por la que tenemos elecciones es para averiguar si las encuestas eran correctas?*

ROBERT ORBEN

## Introducción

Iniciamos nuestro recorrido recuperando como epígrafe una famosa frase atribuida al mago y escritor de comedia estadounidense Robert Orben, referida según se indica en diversos sitios a personal del periódico *Ventura County Star* (2008).

Y en ello estamos al tanto de los varios significados de la palabra atribución, que no solo refiere a “aplicar, a veces sin conocimiento seguro, hechos o cualidades a alguien o algo”, sino que tiene como segunda acepción la de ser “la interpretación o explicación que se hace acerca de las causas, motivos y razones de algún suceso” (RAE, 2020).

Ello viene a cuento porque en este estudio se realizarán al menos dos procesos de atribución en este segundo sentido: por un lado, como parte sustantiva de la exposición sobre cuestiones del método utilizado en este análisis, se explicita un principio que llamaremos de equivalencia, que suele ser asumido —pero no explicitado— para justificar la comparación por mecanismos estadísticos de las estimaciones producto de encuestas con los resultados oficiales de unos comicios dados, en este caso de las elecciones para diputados federales de 2021. Y por otro, al final del texto, tras una evaluación constreñida a datos verificables, se buscará aproximar hipótesis sobre las posibles razones que pudieran estar detrás de la (in) exactitud observada en las mediciones por encuestas previas a las elecciones de marras.

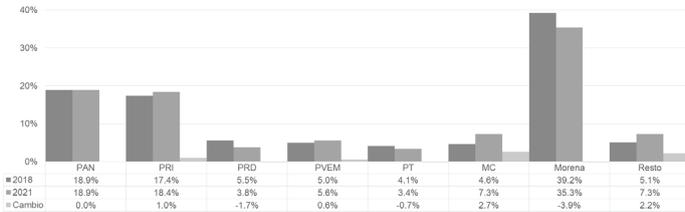
---

\* Maestro en Sociología por la Facultad de Ciencias Políticas y Sociales (FCPYS) de la Universidad Nacional Autónoma de México. Presidente Ejecutivo de ISA Investigaciones Sociales Aplicadas. Correo: ricartur@gmail.com

Debe insistirse en que los cómputos oficiales de las más recientes elecciones para diputados federales en México fueron unos comicios de mantenimiento, luego de lo que pudo ser un realineamiento electoral ocurrido tres años antes, toda vez que en esa ocasión se presentó un acen- tuado traslado en el reparto de votos, aunque su permanencia estaría por confirmarse (Bravo, 2006).

Ello sería un sustrato explicativo de la proximidad de las mediciones por encuesta respecto a los resultados por partido, sobre todo tomando en cuenta que los procedimientos convencionales para el cálculo de la exactitud de las encuestas suelen considerar una gran variedad de competidores con distinto peso relativo. En contraparte, estas variaciones acotadas en la distribución de sufragios tuvieron efectos mayores en la relación de fuer- zas entre los contendientes coaligados con la opción mayoritaria y los que se coaligaron para oponerse, vistos como bloques. Este movimiento bien pudo no ser cabalmente detectado por las encuestas (Gráfica 1).

Gráfica 1. Resultados oficiales de las elecciones para diputaciones federales en México, 2018 y 2021



Fuente: INE (2018 y 2021).

## Consideraciones preliminares

Todo esfuerzo reflexivo sobre el estado del arte de las encuestas, enten- dido como “modalidad de la investigación documental que permite el estudio del conocimiento acumulado” (Molina, 2005), y de los métodos para su evaluación y corrección, debe reconocer que las decisiones de los investigadores deben invariablemente partir del (re)conocimiento de los campos científicos en que se funda el hacer demoscópico, como la teoría de la probabilidad y la estadística, y de los desarrollos tecnológicos existentes para la aplicación de métodos y técnicas de investigación en el campo, además de cubrir los requerimientos normativos nacionales y gre- miales que regulan la realización y divulgación de este tipo de estudios.

Esto no significa que la actividad de investigación misma esté determinada estrictamente por estos condicionantes, puesto que el investigador goza de una autonomía que le permite elegir, entre diversas opciones disponibles, aquellas que considere mejores para el logro de sus objetivos conforme su propia experiencia.

Desde este punto de vista, la investigación demoscópica sobre preferencias electorales no es una mercancía, aunque así se pretenda ver mediante la imposición de criterios o modelos de “normalización”, puesto que no es ni puede ser un bien homogéneo entre los diversos productores que cruzan el proceso de selección de métodos y técnicas, aunque su empleo posterior pueda ser el de servir de insumo para estrategias partidarias o mediáticas.

En el caso de las encuestas que se publican previamente a unas elecciones, por motivos no solo de apego a la ética científica, sino también por beneficios prácticos, el investigador seleccionará los mejores métodos y procedimientos que sean accesibles y factibles (por tiempos y costos) para la disposición de estimadores sobre intenciones de sufragio de los electores o en su caso, de los eventuales concurrentes a unos comicios (mediante modelos de “votantes probables”).

Así, la realización de una encuesta que no sea meramente un instrumento de propaganda, sino de conocimiento, tendrá la intención de reducir la incertidumbre producto del carácter muestral del ejercicio e incrementar su posible exactitud. Para ello, partirá de revisar y asumir en lo fundamental la parte dura del “canon científico” en la materia, pero también empleará un modelo “típico-ideal”, en el sentido de instrumento conceptual propuesto por Weber (Donato, 2007), de lo que deben ser los objetivos, contenidos y procedimientos de una medición por encuesta para la estimación de intenciones de voto, aproximando estos referentes a las condiciones específicas y recursos disponibles para su realización.

La intención de este texto es, a partir de una caracterización y definición operativa de los conceptos que las encuestas intentan medir, evaluar los estimadores existentes para calcular la exactitud de las encuestas electorales publicadas previamente a las elecciones para diputados federales de 2021 y su capacidad de previsión de los niveles de competitividad que se presenten. En lo posible, se seguirán normas convencionales de tipografía y notación matemática, privilegiando el empleo de una única letra para etiquetar un símbolo para evitar confusiones (Taagepera, 2008).

Partamos de definir lo que es una encuesta (De la Peña, 2015): es un estudio científico cuyo objetivo es producir información que permita estimar la ocurrencia de eventos actuales o no, por lo general a través

de instrumentos que permiten un tratamiento normalizado de los datos. Teóricamente, las estimaciones que se obtienen deben estar cerca de un valor real que se pretende conocer, denominado parámetro.

Hay que reiterar que el principal fundamento de la científicidad de una encuesta es la disposición de una muestra probabilística de la población objetivo, que en el caso de las electorales están diseñadas para que cada ciudadano tenga una probabilidad conocida y distinta de cero de participar y obtenga un peso equitativo en la estimación. De cumplirse esto, los datos que se obtengan permitirán hacer inferencias sobre los valores esperables para un reactivo dentro la población considerada.

Recuérdese que en estadística, precisión y exactitud no son conceptos equivalentes, aunque estén relacionados (Joint Committee for Guides in Metrology, 2012). La precisión es una medida de la dispersión del conjunto de valores obtenibles en repetidas mediciones de una magnitud determinada: a menor dispersión, mayor precisión. Refiere entonces a la magnitud escalar con la que se mide la proximidad de los resultados de una muestra para reproducir los resultados que se obtendrían de un recuento completo llevado a cabo usando las mismas técnicas.

A diferencia, la exactitud es una medida de la proximidad entre los valores medidos y los reales, aquellos resultantes si se hubiera realizado un censo bajo los mismos procedimientos que la encuesta, por lo que mide el eventual sesgo de una estimación: a menor sesgo, mayor exactitud. Luego, refiere a la magnitud escalar con la que se mide la proximidad del estimador producto de una muestra respecto del parámetro objetivo.

Existen siempre imperfecciones entre el diseño y lo realizado en la práctica de encuestar, lo que genera constantes riesgos de afectación de la aleatoriedad debido a limitaciones intrínsecas a la cobertura de la población objetivo, derivadas de la selección del modo de aproximación a los informantes, las tasas de encuentro o cooperación de los individuos e imprecisiones en el trabajo de campo.

Aunque todos estos aspectos pueden ser tentativamente reparados, es inevitable la posible presencia de sesgos. Las fuentes generadoras de estos sesgos son significativamente más reducidas en el caso mexicano actual en encuestas que recurren a tradicionales procedimientos de entrevista personal directa en vivienda, sea o no asistida por computadora, donde se logran alcanzar coberturas cercanas al total de la población objetivo, que cuando se realizan mediante aproximación por vía telefónica, que permite alcanzar solamente tres cuartas partes de la población y que presentan elevados niveles de rechazo a la entrevista, o que en ejercicios mediante la compilación de datos de usuarios de redes sociales, cuyo alcance se limita

a la mitad de la población en el mejor de los casos y donde a las reducidas proporciones de respuesta, se suma la incapacidad de control autónomo de la selección de informantes por parte del investigador.

Y si deformaciones en la distribución de la población que se observe respecto al universo de estudio pueden ser corregidas mediante post-estratificaciones para variables demográficas e incluso socioeconómicas, resulta sumamente complejo e incierto adecuar la distribución de la muestra conforme parámetros psicométricos, cuyas características son prácticamente conocidas y dudosamente observadas en la medición.

Es por ello que las encuestas mediante entrevistas personales en vivienda son ejercicios que efectivamente brindan hoy día la posibilidad de estimar el reparto de intenciones de voto del grueso de la ciudadanía en México. Y en consecuencia, es sobre la colección de encuestas con estas características que fueron hechas públicas antes de la jornada electoral, que se realiza el ejercicio de comparación contra los resultados oficiales.

## Para medir la exactitud de las encuestas

El procedimiento científico pertinente para estimar la exactitud de encuestas electorales será el contraste entre lo esperado conforme la medición y lo observado en la realidad. Así, a pesar de que en términos estrictos es fundamental observar lo riguroso de los procedimientos utilizados y no necesariamente su adecuación con un supuesto parámetro con el que pudiera diferir por múltiples razones, no puede verse como algo impertinente la insistencia colectiva y mediática de comparar los estimadores producidos en encuestas previas contra los resultados oficiales de una elección.

El rechazo de este cotejo, más allá del argumento de que lo importante es el método y no la exactitud de una medición, suele ser expresión de una negativa al examen *a posteriori* que pudiera arrojar luz para identificar metodologías más o menos eficientes y para contrastar hipótesis pretendidamente explicativas de las divergencias observadas, lo que permitiría detectar factores incidentes en la capacidad de medición y anticipación por encuesta de los resultados.

Luego, si una medición es un procedimiento para asignar un número a un evento conforme reglas definidas, de manera tal que permita graduar una propiedad (De Villis, 2003), para medir la exactitud de las encuestas electorales respecto de los resultados suele usarse como escala la diferencia entre las proporciones arrojadas por un estudio y las proporciones computadas en

una elección. Ello supone tomar como parámetro los resultados y calcular el error absoluto, como se hace entre cualquier medición y un valor considerado como el verdadero.

Esto obliga a echar mano de procedimientos para el cálculo de la proximidad entre conjuntos de datos, usando como medida negativa la distancia para descubrir la eficacia de los métodos utilizados. Con base en lo que genera este ejercicio de comparación, se puede tener capacidad para sugerir potenciales mejoras para alcanzar mayor exactitud en futuros ejercicios. Pero, ¿cómo realizar este ejercicio de comparación?

Partamos de lo más elemental: en una elección cada persona que forma parte del universo de electores manifiesta mediante un procedimiento consensuado una decisión que ha tomado a lo largo del tiempo, la cual puede diferir entre el momento de toma de datos y la fecha de la elección. Es por ello que lo idóneo en un cotejo es tomar solamente estimaciones que se hayan realizado poco tiempo antes de dicha fecha y que resulten ser las últimas de cada encuestadora incluida en una colección a evaluar para fines comparativos.

Así, como resultado de una elección, el universo finito y numerable de electores ( $E$ ) se divide en tres grupos: los votantes, referido a aquellos que votaron por cada uno de los contendientes ( $V_i$ ); los abstencionistas activos, que son quienes votaron por alguien que no competía o cuyo voto fue invalidado de manera voluntaria o incidental ( $V_0$ ); y los abstencionistas pasivos, que son quienes de hecho no acudieron a votar ( $V'$ ). Entonces:

$$E = \left( \sum_{i=1}^m V_i \right) + V_0 + V'$$

El resultado oficial de una elección es, luego, la suma de los votos contabilizados por cada uno de los contendientes que compiten en dicha elección, más los votos contados que se emitieron por alguien no registrado como contendiente en la elección o que fueron anulados. A partir de estos datos, una vez que fueron eliminados los votos no válidos, es posible obtener las proporciones de voto por cada contendiente registrado respecto al total de votos válidos por:

$$p_i = \frac{V_i}{\sum_{i=1}^m V_i}$$

Ahora bien, como resultado directo de una encuesta, el universo finito y contable de electores que fue observado una vez ajustado por los

factores de expansión permitentes ( $\hat{E}$ ) se divide en tres grupos (2.6): quienes respondieron que votarían por alguno de los diversos contendientes ( $\hat{V}_i$ ), quienes declararon que votarían por un candidato no registrado o que anularían su voto o quienes así lo hicieron de facto cuando la medición fue vía simulacro de boleta ( $\hat{V}_0$ ) y quienes no respondieron que votarían por alguno de los contendientes ( $\hat{V}'$ ), sea porque no saben por quién votarían, porque no piensan votar por alguno de los contendientes o porque por razones solo por ellos conocidas no respondieron al reactivo aplicado para conocer la preferencia, quienes no necesariamente se convertirán en no votantes. Así:

$$\hat{E} = \left( \sum_{i=1}^m \hat{V}_i \right) + \hat{V}_0 + \hat{V}'$$

Y por ende, una vez que se eliminan los casos correspondientes a electores entrevistados que no se definieron por alguno de los contendientes registrados, es posible obtener las proporciones estimadas de intenciones de voto por contendiente registrado respecto al total de casos validados por haberse definido por algún contendiente, a partir de:

$$\hat{p}_i = \frac{\hat{V}_i}{\sum_{i=1}^m \hat{V}_i}$$

Y aunque los dos conjuntos disponibles de repartos del voto, el estimado y el observado por contendiente, pueden no ser idénticos por múltiples motivos, para el cotejo comparativo se parte de una petición de principio por la que se obvian las diferencias y se asume una equivalencia entre el segmento de electores que votaron por algún contendiente registrado y el segmento de entrevistados que declararon que votarían por alguno de los contendientes registrados, excluyendo segmentos que no se pronunciaron por algún contendiente en uno y otro de los grupos. Así, al garantizarse que cada lado de la ecuación sea equivalente e igual a la unidad:

$$\sum_{i=1}^m \hat{p}_i = 1 = \sum_{i=1}^m p_i$$

Se tiene que:

$$\frac{V_i}{\sum_{i=1}^m V_i} = \frac{\hat{V}_i}{\sum_{i=1}^m \hat{V}_i}$$

Cabe aclarar que este procedimiento no supone una imputación o asignación de preferencias para los casos indefinidos, sino que asume su exclusión para fines de estimación a costa de perder precisión, al reducir el tamaño de muestra usado para la estimación. Esta exclusión es requerida dado que resulta impropio calcular el error de una estimación a partir del cotejo entre los votos emitidos, debiendo considerarse solamente los votos válidos, o de proporciones de casos observados, debiendo hacerse a partir de casos efectivos, lo que permite establecer la debida igualdad en la suma de las proporciones de lo estimado con lo oficial.

Es posible luego definir la diferencia entre la estimación para un componente dado en cualquier encuesta respecto del resultado oficial ( $e_i$ ) como

$$e_i = \hat{p}_i - p_i$$

Que produce una magnitud escalar cuyo valor varía de uno a menos uno, por lo que aporta no solamente una distancia, sino un sentido de la diferencia entre la estimación particular de un contendiente en una encuesta y el resultado oficial.

Dado que la distancia entre dos puntos del espacio euclídeo equivale a la longitud del segmento de la recta que los une, expresado numéricamente, la estimación de  $e_i$  arroja como resultado un estimador de distancia. Sin embargo, a partir de este dato no es posible el cálculo de la distancia para el conjunto de contendientes, pues su suma siempre será cero, dado que los errores en un sentido se compensan con los errores en sentido contrario.

Una forma de enfrentar la imposibilidad de obtener un estimador del error a partir de la suma de los errores observados es simplemente eliminar el sentido de la medición, tomando como error la diferencia absoluta:

$$|e_i| = |\hat{p}_i - p_i|$$

Estimador que corresponde estrictamente a la definición convencional del error absoluto de una medición. Este estadístico, aplicado a los dos partidos mayores, aporta un dato sobre la diferencia entre lo estimado y lo real del margen de victoria, que corresponde a un estimador de la serie de posibles métodos para mesurar el error de las estimaciones por encuesta respecto a los resultados de una elección propuestos y analizados por Mosteller (1949), luego del desencuentro entre mediciones y resultados en la elección presidencial de 1948 en Estados Unidos:

$$M5 = |(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)|$$

Sin embargo, cuando el número de contendientes considerados es mayor de dos, la reducción del resultado de una encuesta y la elección a la brecha entre los dos contendientes mayores es una solución incompleta al problema de medición del error. Lo anterior es más relevante si se considera que el propio reactivo aplicado en una encuesta en realidades multipartidarias suele considerar como opciones expresas de respuesta a cada uno de los contendientes registrados para la elección, bien sea que los refiera en el fraseo de la pregunta, bien que los incluya en el símil de boleta que se utilice para la recuperación de preferencias de los entrevistados.

Por ende, a partir de la estimación de la diferencia absoluta entre lo estimado y lo real por cada contendiente considerado, se puede efectuar una suma y sacar un error promedio que aporte un valor distinto de cero y que indique la distancia entre lo observado y el parámetro:

$$M3 = \frac{\sum_{i=1}^m |\hat{p}_i - p_i|}{m}$$

Otra solución, que parecería preferible, es recurrir al cálculo de la distancia euclidiana ( $d$ ), que permite estimar la magnitud de la diferencia, aunque perdiendo la posibilidad de definir un sentido como magnitud escalar, pues éste no es unívoco cuando se trabaja en un espacio de dimensión superior a dos:

$$d = \sqrt{\sum_{i=1}^m (\hat{p}_i - p_i)^2}$$

Tomando en cuenta el acotamiento a la unidad del total de las proporciones consideradas para el cálculo, este estimador arroja valores que se ubican entre cero y la raíz cuadrada de dos, puesto que lo que es saliente en un caso es entrante en otro. Esto puede corregirse si tomamos la distancia euclidiana normalizada ( $D$ ), que sería el resultado de dividir entre dos la suma de los cuadrados de los errores observados, eliminando así el efecto de doble contabilidad:

$$D = \sqrt{\frac{\sum_{i=1}^m (\hat{p}_i - p_i)^2}{2}}$$

Ergo, se estaría contando solamente la magnitud de las salientes o de las entrantes, no de ambas. Este estimador de la distancia entre una encuesta y un resultado oficial se ubicará luego necesariamente entre cero y uno: cero, cuando hay perfecta concordancia entre lo estimado y lo real, y uno, cuando la diferencia es la máxima posible. Este dato, convertido en porcentaje, puede ser indicador de una distancia relativa entre lo estimado y lo real.

Aunque a lo largo de este texto se recurrirá a analizar las diferencias entre encuestas y elecciones para diputados federales en 2021 mediante los diversos estimadores expuestos, se privilegiará la distancia euclidiana normalizada, dado que este medidor se corresponde con la definición básica de la distancia entre dos puntos en un espacio multidimensional, manteniendo una sencilla y clara lectura, cercana a lo intuitivo.

Esto no invalida ni excluye de manera alguna el posible recurso a métodos más complejos para la medición de distancias, como los que expone Márquez (2021), quien también reflexiona sobre las desventajas advertidas en las métricas elementales y advierte de los problemas de tomar pruebas de bondad de ajuste como alternativa, como la  $\chi^2$  (listada como  $M7$  por Mosteller), pues en ellos la hipótesis nula suele ser falsamente rechazada.

Márquez apuesta por el uso de medidas de distancias vectoriales entre estimaciones por encuesta de la preferencia electoral respecto al resultado electoral, que son independientes de escala, de permutaciones y de perturbaciones, a la vez que no suponen un comportamiento totalmente lineal. Por ello, estos medidores de distancia son una interesante sugerencia, aunque presentan el problema de ser difíciles de interpretar y lejanos a lo intuitivo.

## Los indicadores de competitividad

Nuestro camino para avanzar en la lectura de la proximidad de estimaciones por encuesta frente a resultados de elecciones es ir más allá de la lectura de un medidor de la distancia entre lo estimado y lo observado a partir de las estimaciones puntuales por contendiente, como se ha hecho de manera tradicional, sino recurrir además al cálculo de estimadores agregados relativos a la competitividad de las contiendas y lo advertido por las encuestas, para estimar su separación.

Como acertadamente apuntan Taagepera y Shugart (1989) en su ya clásica obra sobre conversión de votos en asientos, los estudios de los sistemas electorales pueden aportar una *Piedra Roseta* para otros campos

de la ciencia política. Así, partiendo de cantidades fácilmente definibles y mensurables, se puede arribar a nociones más complejas, con un significado preciso. Es por ello que desde hace décadas este potencial tratamiento matemático del campo electoral ha propiciado la postulación y el empleo de indicadores agregados que dan cuenta de los eventos electorales.

Los índices son herramientas con un elevado poder comunicativo al poder describir “en una sola cifra (...) una posición relativa o una tendencia” (Tarriba, 2014). En el ámbito electoral se cuenta así con salidas unitarias que permiten un tratamiento sencillo, diáfano e inequívoco de datos para la construcción de índices. De manera destacada, los datos de votos y de asientos por partido han podido colapsarse en indicadores simples que pretenden dar cuenta en un único dato de las distribuciones o diferencias observadas.

En el caso particular de los indicadores que buscan medir la competitividad electoral a través de la estimación de los niveles de concentración-fragmentación del voto, Dunleavy y Boucek (2003) han avanzado en la demostración de que no existe un único índice que dé cuenta del número de competidores en un sistema y que a la vez esté exento de problemas.

Es por ello que pareciera pertinente avanzar, más que en la ruta por intentar el encuentro de un mágico indicador perfecto, en la búsqueda por definir “familias” de indicadores con un par de miembros representativos que puedan dar cuenta más plena del fenómeno, a la vez que se compensen entre sí para amortiguar el impacto de las imperfecciones de cada cual.

El problema del cómputo de componentes efectivos en un conjunto de contendientes no es algo que se reduzca al campo de lo electoral y, de hecho, es un problema que ha sido centro de atención de los economistas antes que de los politólogos. Y precisamente del ámbito de la economía es de donde han surgido los fundamentos que luego han sido retomados y transformados por la ciencia política para intentar resolver el problema del cómputo: del número de participantes efectivos en un mercado se pasó al cálculo del número de competidores en la arena electoral, contados de manera tal que corresponda a aquellos que sean relevantes conforme a algún criterio claro y pertinente.

De hecho, el problema del cómputo en cuestión no es un asunto de cuál escala que se emplee será la “verdadera” o “correcta”. En muchos asuntos de la ciencia no existe un problema de “verdad” involucrado. Cuando se trata de escalas, simplemente se está designando una relación matemática que supone un factor de conversión para representar la realidad en una ficción simplificadora, que permita reducir información a un medidor unitario que se presumiría universalmente útil.

En las comunidades científicas, se llega a acuerdos sobre las escalas para describir la naturaleza de la información contenida en un conjunto de números asignados a objetos. Y claro, son varias las operaciones posibles dependiendo de qué se quiere medir y qué se quiere lograr con la medición, entendida ésta como la asignación de números a conjuntos de objetos o sucesos siguiendo ciertas reglas.

Respecto a la medición de la competitividad electoral, se dispone de diversas propuestas de indicadores que buscan agrupar en un único valor la distribución de sufragios entre partidos contendientes en una elección. El estimador más comúnmente empleado corresponde al indicador del número de componentes en cualquier mercado y que, en el caso del ámbito electoral, ha tomado el nombre de “número efectivo de partidos” ( $N$ ), adjudicado por los autores que sugirieron este índice: Laakso y Taagepera (1979). Este indicador es igual al inverso de la sumatoria de los cuadrados de las proporciones de votación por los diversos partidos, es decir:

$$N_v = \frac{1}{\sum_{i=1}^n v_i^2}$$

Este índice representa la cantidad de partidos de igual tamaño que genera el mismo efecto de concentración-fragmentación de los componentes, medido por el índice de concentración de Herfindahl y Hirschman ( $H$ ), definido como:

$$H = \sum_{i=1}^n v_i^2$$

O bien, por su complemento, el índice de fragmentación de Rae ( $F$ ), definido como:

$$F = 1 - H = 1 - \sum_{i=1}^n v_i^2$$

N resulta, en este sentido, ser un aporte muy importante a la medición convencional del número de componentes, en la medida en que adopta una presentación que otorga mayor claridad, al resultar menos abstracta que las formulaciones anteriores. Es de mencionar que, al igual que en el caso de otros indicadores agregados de competitividad,

puede establecerse un número efectivo de partidos para la votación y otro para la distribución de asientos.

Sin embargo, este índice muestra una imperfección, dado que adquiere valores superiores a lo intuitivo en casos de hegemonía de un partido. Debido a esta limitación, planteamos una potencial aplicabilidad al análisis electoral del índice de dominancia (P), desarrollado por García Alba (1994) como aporte para disponer de un indicador que dé cuenta de las condiciones efectivas de competencia económica, definido como:

$$P = \sum_{i=1}^n \left( \frac{v_i^2}{H} \right)^2 = \sum_{i=1}^n \left( \frac{v_i^2}{\sum_{i=1}^n v_i^2} \right)^2$$

Que resulta ser un indicador del nivel de concentración de la concentración del voto y, por ende, un promedio de las participaciones de cada partido en la concentración del voto, medida a través de H.

La disposición simultánea de H y P permite ver la relación entre la estructura de un mercado con el poder de un competidor de fijar precios, o en un sistema electoral, la relación entre el reparto de los votos y el poder de un partido mayoritario derivado de detentar una condición hegemónica.

García Alba postula este índice como parte de una familia de la que H es un elemento, siendo P el elemento de grado menor en un subconjunto dentro de esta familia que cumple con la condición de aumentar cuando se fusionan componentes relativamente grandes y disminuir cuando se fusionan componentes relativamente pequeños.

Así, el valor de este índice no aumenta con cualquier fusión o alianza, sino solo con las que involucran componentes mayores, asumiendo que cada partido ejercerá mayor capacidad de influir en un sistema mientras mayor sea su respaldo electoral relativo al de los demás partidos. De hecho, para evaluar la posibilidad de que una distribución de votos propicie prácticas hegemónicas, debe considerarse no solamente la concentración en cada componente, sino la relación con las concentraciones entre los demás componentes. Ello, toda vez que el componente mayor podrá ejercer un poder hegemónico no solamente cuando supere cierta marca de votación relativa, sino entre menos concentrado esté el voto en los restantes componentes.

Este índice no constituye un medidor numérico del número de componentes de una distribución, pero sí su inverso, repitiendo el proceso de

conversión usado antes por Laakso-Taagepera con  $H$ . El inverso de este índice, que entonces pudiera llamarse “número de componentes autónomos” puede definirse como (De la Peña, 2005):

$$A = \frac{1}{P} = \frac{1}{\sum_{i=1}^n (Nv_i^2)^2} = \frac{1}{\sum_{i=1}^n \left(\frac{v_i^2}{H}\right)^2} = \frac{1}{\sum_{i=1}^n \left(\frac{v_i^2}{\sum_{i=1}^n v_i^2}\right)^2}$$

Que constituiría un estimador alternativo para el número de competidores en un sistema electoral, y que supone que a mayor nivel de dominancia de uno o más partidos en un sistema, menor número de partidos autónomos y viceversa. Las características de este índice serían: el número de autonomías nunca es mayor al número efectivo de componentes (siendo igual cuando las proporciones de los componentes son iguales); cualquier transferencia o fusión hacia el componente mayor disminuye su valor; si un componente representa más de la mitad del voto, el índice será menor que dos; el índice disminuye cuando se fusionan o alían dos componentes cuya participación es mayor a la que resultaría de la fusión de cualesquiera otros dos componentes; si la participación conjunta de dos componentes distintos al mayor supera la mitad, el índice es mayor a dos; el índice aumenta ante cualquier fusión que no involucre al componente mayor, si la participación de éste es mayor a la mitad; si la fusión de dos componentes disminuye el índice, lo mismo sucede de fusionarse dos componentes de mayor tamaño, y si lo aumenta, lo mismo sucede de hacerlo dos componentes menores.

Cuando se compara el comportamiento de  $A$  con  $N$  se encuentra que cuando un componente es hegemónico, si el resto del sufragio se concentra en un componente único,  $N$  alcanza su valor mínimo, mientras que, en el mismo caso,  $A$  llega al valor máximo, indicando una elevada competitividad.

Así, mientras que para  $N$  el voto disperso en una multiplicidad de partidos frente a uno hegemónico sería muestra de mayor competitividad que cuando se concentra en un solo oponente, conforme  $A$  se logra una mayor competitividad precisamente cuando una fuerza opositora logra concentrar el voto restante al acaparado por un partido que goza de una condición hegemónica.

Lo anterior lleva a postular la conveniencia de considerar este número de autonomías como estimador pertinente del formato de competencia en

un sistema electoral, que refleja más eficientemente que  $N$  el número de competidores reales en un sistema dado, sobre todo uno multipartidista, y que es preferible si de lo que se trata es de determinar a partir de un indicador único el potencial de formación de coaliciones opositoras que logren superar al partido mayoritario.

Es por estos motivos que, para fines de este análisis, se recurrirá a comparar los índices de competitividad  $N$  y  $\hat{A}$  derivados del reparto de votos estimado por las encuestas con los niveles de competitividad que arrojan los resultados electorales. Esto se logra mediante el simple cálculo de la diferencia entre el índice estimado por encuesta y el observado en la realidad. Para el caso del número efectivo de componentes esto sería:

$$e_N = \widehat{N}_v - N_V$$

Mientras que para el número de contendientes autónomos sería:

$$e_{\hat{A}} = \widehat{\hat{A}}_v - \hat{A}_v$$

Con la finalidad de evaluar las distancias entre estimadores por encuesta y resultados electorales y las diferencias en número de competidores que se derivan de estos repartos, se recurre a dos fuentes: por un lado, los datos oficiales sobre los resultados de las elecciones puestos a disposición pública por el Instituto Nacional Electoral, en 2018 y 2021; y por el lado de las encuestas, la colección de mediciones nacionales mediante entrevistas personales en vivienda que fueron compiladas por Márquez (2021).

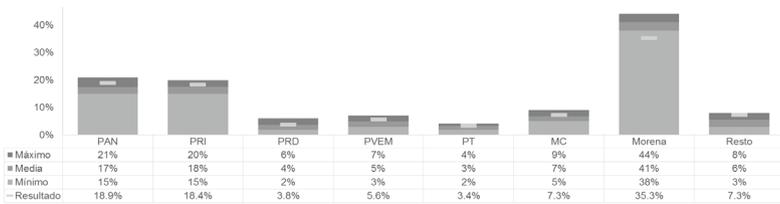
Pasemos entonces a la evaluación de la exactitud de las encuestas previas a las elecciones para diputados federales de 2021, haciendo uso de los estimadores de distancia y de las diferencias en el cálculo de componentes que hemos elegido. Como anexo a este texto se presentan tablas donde se recuperan los datos de las estimaciones puntuales por casa encuestadora a nivel de partido y de bloque, sus comparativos contra los resultados oficiales y los estimadores de distancia utilizados en este análisis.

## Estimaciones por partido

Cuando se observan las estimaciones por encuestas previas a las elecciones para diputaciones federales de 2021 para cada partido político (Gráfica 2) se descubre que el conjunto de encuestas públicas disponibles ubicaron los márgenes para seis de los competidores y para el grupo de nuevos

partidos concurrentes. El único caso en que ello no ocurrió fue el de MORENA, que tuvo un mínimo estimado de 38%, cuando en realidad apenas rebasó los 35 puntos.

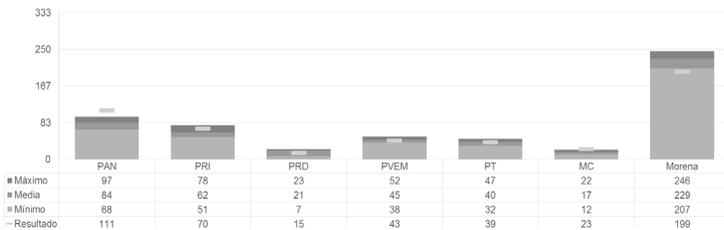
Gráfica 2. Estimación promedio de la votación por partido de las encuestas nacionales en viviendas y resultado de las elecciones para diputaciones federales en México, 2021



Fuente: INE (2021) y Oraculus (2021).

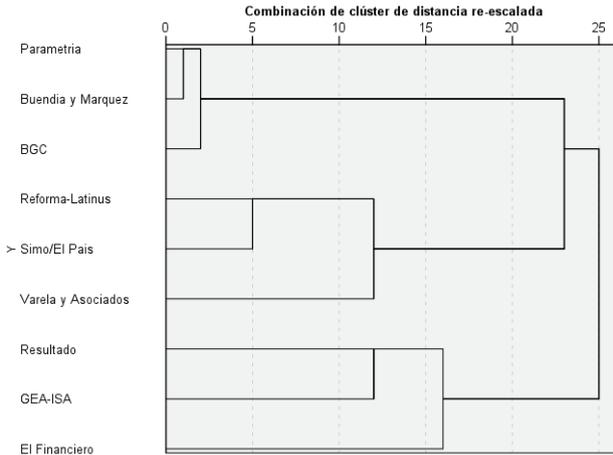
Debido a que las estimaciones sobre el reparto de curules con base en encuestas tienen una fuente de imprecisión adicional a las propias mediciones del voto potencial, por los efectos de las reglas de conversión, suelen ser más sensibles y por ende menos exactas que las estimaciones puntuales de las proporciones de voto por partido. Esto ocurrió en el caso del reparto esperado de asientos producto de las encuestas nacionales en vivienda que efectuó Oraculus, por lo que no solamente MORENA estuvo fuera de los rangos previos calculados, sino que esto también ocurrió en los casos del PAN y de Movimiento Ciudadano (MC) (Gráfica 3).

Gráfica 3. Estimación de curules por partido derivada de las encuestas nacionales en vivienda y reparto oficial de curules resultado de las elecciones para diputaciones federales en México, 2021



Fuente: INE (2021) y Oraculus (2021).

Gráfica 4. Dendrograma entre estimaciones de votación por partido de las encuestas nacionales en vivienda y resultado de las elecciones para diputaciones federales en México, 2021

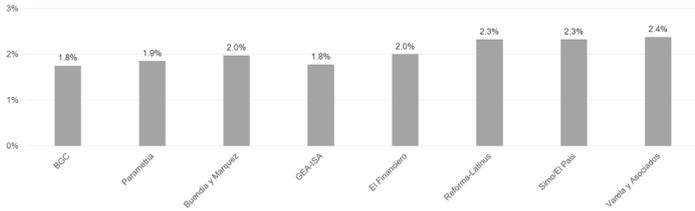


Fuente: INE (2021) y Oraculus (2021).

Una manera alternativa de visualizar la proximidad entre estimaciones por encuesta y los resultados es trazar el diagrama de árbol que muestra precisamente dicha cercanía, pero refleje a la vez los agrupamientos existentes entre las distintas mediciones, lo que se presenta en la Gráfica 4. Este diagrama refleja no solamente la mayor proximidad de las mediciones en primer lugar de GEA-ISA y en segundo lugar de *El Financiero* con el resultado, sino la existencia de otros dos grupos más distantes de mediciones, uno formado por casas encuestadoras incorporadas al organismo gremial llamado CEDE y otras por encuestadoras mediáticas y ajenas a este organismo.

Pasando a los estimadores numéricos de la distancia entre estimaciones por encuestas y resultados a nivel de partido, en la Gráfica 5 se muestra que, a pesar del agrupamiento anteriormente observado, fueron dos las casas encuestadoras con menor distancia promedio ( $M3$ ) con el resultado: GEA-ISA y BGC, aunque Parametría se situaría muy cercana a los niveles de inexactitud observadas por las dos encuestadoras previamente referidas.

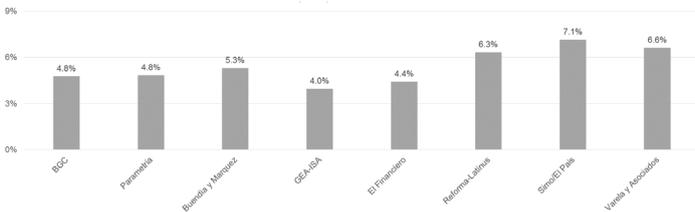
Gráfica 5. Diferencia promedio entre las estimaciones por partido ( $ME_p$ ) de las encuestas nacionales en vivienda y el resultado de las elecciones para diputaciones federales en México, 2021



Fuente: INE (2021) y Oraculus (2021).

Este ordenamiento se altera cuando lo que se toma en consideración es la distancia euclidiana normalizada ( $D$ ), que se presentan en la Gráfica 6, pues a partir de este estimador de exactitud GEA-ISA sería la encuestadora más próxima al resultado, mientras que *El Financiero* ocuparía el segundo lugar.

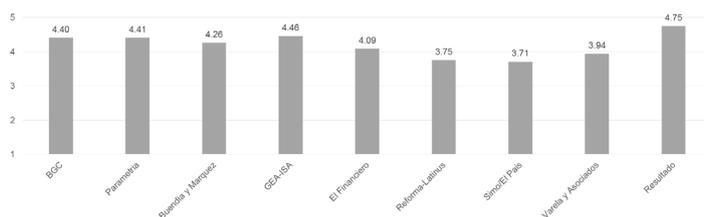
Gráfica 6. Distancia euclidiana normalizada entre las estimaciones por partido ( $D_p$ ) de las encuestas nacionales en vivienda y el resultado de las elecciones para diputaciones federales en México, 2021



Fuente: INE (2021) y Oraculus (2021).

El número efectivo de partidos que se registró como resultado de las elecciones para diputaciones federales de 2021 fue de 4.75. Como se muestra en la Gráfica 7, todas las encuestadoras subestimaron el nivel de competencia medido por este indicador, aunque hubo una elevada varianza entre publicaciones: de nuevo GEA-ISA sería la más próxima al resultado, seguida cercanamente por BGC y Parametría, mientras que *Reforma-Latinus* y *Simo-El País* habrían reportado los menores niveles de fragmentación del voto.

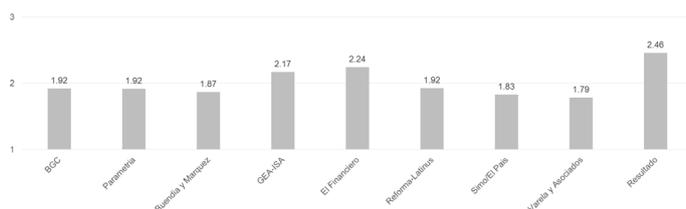
Gráfica 7. Número electivo de partidos ( $N_p$ ) estimados por las encuestas nacionales en vivienda y resultado de las elecciones para diputaciones federales en México, 2021



Fuente: INE (2021) y Oraculus (2021).

El número de contendientes autónomos en las elecciones para diputados federales de 2021 fue de 2.46 partidos, como se muestra en la Gráfica 8. Y en este indicador, nuevamente las encuestas reflejaron un potencial de competencia inferior al que se registró en la realidad, siendo los más cercanos a ella *El Financiero* y GEA-ISA.

Gráfica 8. Número de partidos autónomos ( $A_p$ ) estimados por las encuestas nacionales en vivienda y resultado de las elecciones para diputaciones federales en México, 2021



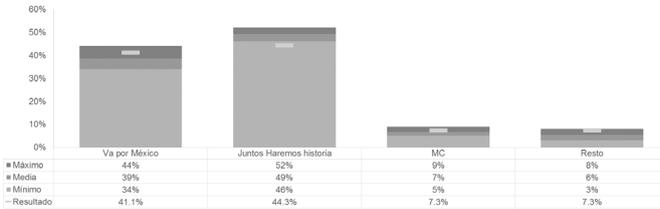
Fuente: INE (2021) y Oraculus (2021).

## Estimaciones por coalición

Si en vez de observar el reparto por partido de la votación estimada a través de encuestas previas a los comicios para diputados federales de 2021, lo que se revisa son las estimaciones que corresponderían a las coaliciones participantes definidas como bloques de partidos, que fue reportada por varias de las casas encuestadoras y que se presentan en la Gráfica 9, se observa que se registró una sobrestimación de la proporción de voto para el bloque gobernante, que en muchos casos

reportaba una intención de voto que representaba por sí sola el logro de una mayoría absoluta de los sufragios.

**Gráfica 9. Estimación promedio de la votación por coalición de las encuestas nacionales en vivienda y resultado de las elecciones para diputaciones federales en México, 2021**



Fuente: INE (2021) y Oraculus (2021).

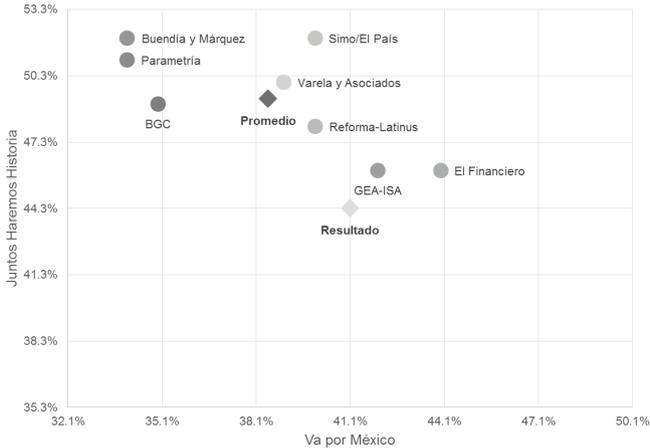
**Gráfica 10. Estimación de curules por coalición derivada de las encuestas nacionales en vivienda y reparto oficial de curules resultado de las elecciones para diputaciones federales en México, 2021**



Fuente: INE (2021) y Oraculus (2021).

Este fenómeno, sin embargo, se matiza cuando lo que se ve es el reparto de curules estimado por Oraculus a partir de las mediciones por encuesta, que se muestra en la Gráfica 10, donde se ve que aunque tirado hacia abajo, la proporción de curules esperada para la alianza *Va por México* estuvo dentro de los rangos estimados y que, igualmente, aunque ubicada hacia arriba, la participación de la coalición *Juntos Haremos Historia* se mantuvo dentro de los rangos de las mediciones.

Gráfica 11. Mapa de las estimaciones de votación por coalición de las encuestas nacionales en vivienda, promedio y resultado de las elecciones para diputaciones federales en México, 2021

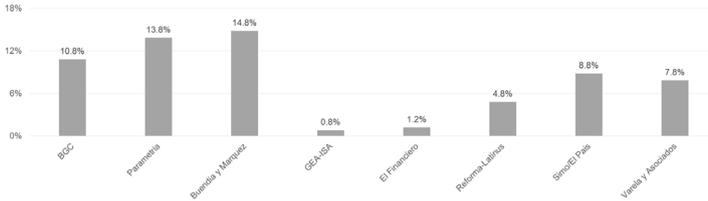


Fuente: INE (2021) y Oraculus (2021).

En la Gráfica 11, se presenta un mapa que muestra en ejes cartesianos la distancia observada entre las mediciones por encuesta y su promedio contra los resultados de las elecciones para diputaciones federales de 2021 para las dos coaliciones contendientes. Este mapa da cuenta simultáneamente de la distancia de las encuestas y el sentido del sesgo en la medición, mostrando claramente la inclinación a sobrerrepresentar la porción del voto esperada para la alianza *Juntos Haremos Historia*, en detrimento de la coalición del PAN-PRI-PRD.

Dado que en el caso del voto por bloques sí resulta relevante la distancia esperada y la observada entre las dos principales fuerzas, resulta de interés ver los datos que se presentan en la Gráfica 12, donde se muestra que si bien hubo dos casas encuestadoras que reflejaron muy certeramente la reducida ventaja de la coalición *Juntos Haremos Historia*, la gran mayoría de las encuestas tendió a sobrestimar este margen, destacando tres casos en que se sobredimensionó la distancia en más de diez puntos, lo que propició que parte de quienes atendieron a los datos producto de encuestas esperaran una victoria significativamente más holgada de la coalición gobernante y alimentó la expectativa de que incluso llegara a alcanzar la mayoría calificada al convertir votos en curules, lo que estuvo lejos de poder ocurrir.

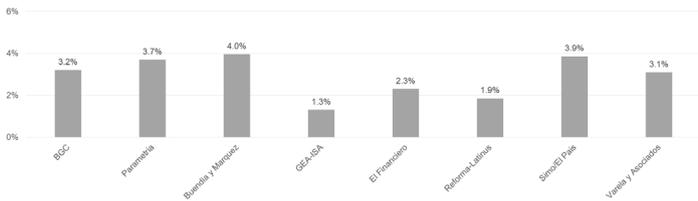
**Gráfica 12. Diferencia entre el margen de victoria entre coaliciones ( $M5_c$ ) estimado por las encuestas nacionales en vivienda y el resultado de las elecciones para diputaciones federales en México, 2021**



Fuente: INE (2021) y Oraculus (2021).

La Gráfica 13 presenta las diferencias promedio ( $M3$ ) entre las estimaciones por coalición y los resultados oficiales. En ella se observa que fueron dos las casas encuestadoras que mostraron una distancia reducida con el resultado: GEA-ISA y *Reforma-Latinus*.

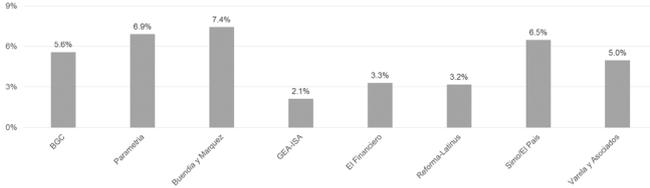
**Gráfica 13. Diferencia promedio entre las estimaciones por coalición ( $M3_c$ ) de las encuestas nacionales en vivienda y el resultado de las elecciones para diputaciones federales en México, 2021**



Fuente: INE (2021) y Oraculus (2021).

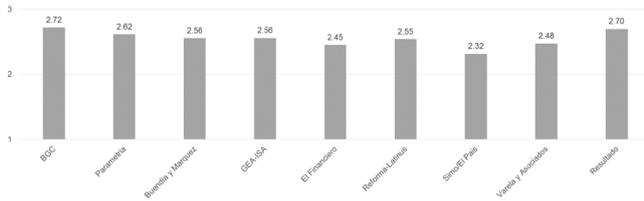
En la Gráfica 14, se muestra la distancia euclidiana normalizada ( $D$ ) entre las estimaciones por encuesta a nivel de coalición y los resultados oficiales de los comicios. De nueva cuenta, conforme a este estimador de distancia, serían GEA-ISA, *Reforma-Latinus* y *El Financiero* las tres encuestadoras más próximas a la realidad.

**Gráfica 14. Distancia euclidiana normalizada entre las estimaciones por coalición ( $D_c$ ) de las encuestas nacionales en vivienda y el resultado de las elecciones para diputaciones federales en México, 2021**



Fuente: INE (2021) y Oraculus (2021).

**Gráfica 15. Número efectivo de coaliciones ( $N_c$ ) estimados por las encuestas nacionales en vivienda y resultado de las elecciones para diputaciones federales en México, 2021**



Fuente: INE (2021) y Oraculus (2021).

El número efectivo de coaliciones resultante de las elecciones para diputados federales de 2021 sería de 2.7 (Gráfica 15). Todas las encuestadoras, salvo BGC, subestimarían este nivel de competitividad.

**Gráfica 16. Número estimado de coaliciones autónomas (AC) por las encuestas nacionales en vivienda y resultado de las elecciones para diputaciones federales en México, 2021**



Fuente: INE (2021) y Oraculus (2021).

A pesar de lo anterior, cuando lo que se estima es la distancia entre las intenciones de voto por coalición medidas por las encuestas y el resultado en el número de contendientes autónomos que produjo la elección, se observa que fueron tres casas encuestadoras las que aproximaron relativamente mejor este indicador de competitividad: GEA-ISA y *El Financiero*, que advierten de la potencialidad de que exista más de una autonomía conforme los resultados de la votación, y Reforma-Latinus, que apenas queda por debajo de la marca de dos autonomías como producto del reparto estimado de votos en estas elecciones.

## Consideraciones finales

Siempre es conveniente recordar que las personas consideran como “normal”, y por ende como lo esperable en un evento por ocurrir, no necesariamente las medidas de tendencia central de un fenómeno, sino más bien una corrección que buscaría un equilibrio entre el valor esperado científicamente y los ideales de las propias personas, producto de sus valores, anhelos y percepciones subjetivas.

Es por ello que, a pesar de la relativamente elevada exactitud que lograron las encuestas previas a las elecciones para diputados federales de 2021 respecto de los resultados a nivel promedio, no pueda soslayarse el hecho de que los sesgos que presentaron la mayoría de las encuestas alimentaron la creencia de que se estaría ante una elección donde la coalición gobernante superaría ampliamente al bloque opositor, sin poderse descartar la posibilidad de que incluso MORENA y sus aliados alcanzaran directamente la mayoría calificada en la Cámara baja del Congreso de la Unión.

Pareciera que nuevamente la evaluación de las encuestas en México, vistas como conjunto, presenta claroscuros, pues se dio un éxito desde lo superficial, que encubre la reiteración del pecado más típico: la sobreestimación al partido en el gobierno. El recuento histórico de la investigación por encuestas para la estimación de preferencias electorales ha tenido sesgos sistemáticos de manera regular, como el acaecido en 2021, que favoreció a MORENA, y por lo general se ha favorecido al partido gobernante, como muestran diversos ensayos publicados en este libro.

Mal harán los encuestadores si se congratulan con lo logrado, puesto que lo esperable sería la revisión de sus ejercicios para determinar las razones de este patrón de sesgo para buscar y eventualmente encontrar cómo prevenirlo y que no se presente en ocasiones por venir.

Y eso cabe para todas las encuestadoras serias, dejando por el momento de lado la debida crítica a las encuestadoras improvisadas o que

aplican metodologías carentes de rigor o pertinencia, a las que ya habrá oportunidad de evaluar.

## Fuentes de consulta

- BRAVO, M. (2006). El realineamiento electoral en México: elementos para su estudio. *Estudios políticos*, núm. 8, pp. 219-242. Disponible en: <http://www.scielo.org.mx/pdf/ep/n8/0185-1616-ep-08-219.pdf>.
- DE la Peña, R. (2005). El número de autonomías y la competitividad electoral. *Política y Cultura*, núm. 24, pp. 233-255. Disponible en: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-459419>.
- DE la Peña, R. (2015). Cómo se mide la exactitud de las encuestas electorales. *Política y Cultura*, núm. 44, pp. 217-247. Disponible en: <https://www.ssoar.info/ssoar/handle/document/59902>.
- DE Villis, R. (2003). *Scale Development: Theory and Applications*. Applied Social Research Methods Series. London: SAGE Publications.
- DONATO, J. (2007). El carácter de los tipos ideales weberianos y su relación con las ciencias naturales. *Diánoia*, núm. 59 (noviembre), pp. 151-177.
- DUNLEAVY, P. y Boucek, F. (2003). Constructing the Number of Parties. *Party Politics*, núm. 9 (mayo), pp. 291-315.
- GARCÍA, P. (1998). Un índice de dominación para el análisis de la estructura de los mercados. *El Trimestre Económico*, núm. 243. (julio-septiembre), pp. 499-524.
- INSTITUTO Nacional Electoral (2018). *Sistema de Consulta de la Estadística de las Elecciones del Proceso Electoral 2017-2018*. México. Disponible en: <https://siceen.ine.mx:3000/#/tablas-resultados>.
- INSTITUTO Nacional Electoral (2021). *Cómputos Distritales 2021: Elecciones Federales*. México. Disponible en: <https://computos2021.ine.mx/votos-districto/mapa>.
- JOINT Committee for Guides in Metrology (2012). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*. 3rd edition.
- LAAKSO, M. y Taagepera, R. (1979). Effective Number of Parties: A Measure with Application to West Europe. *Comparative Political Studies*, núm. 12, pp. 3-27.
- MÁRQUEZ, J. (2021). *Elección para la Cámara de Diputados 2021*. México. Disponible en: <https://oraculus.mx/diputados2021/>.
- MÁRQUEZ, J. (2021). Precisión de las encuestas. *Voz y Voto*, núm. 341 (julio).

- MOLINA, N. (2005). ¿Qué es el estado del arte? *Ciencia y Tecnología para la Salud Visual y Ocular*, núm. 5, pp. 73-75. Disponible en: <https://ciencia.lasalle.edu.co/svo/vol3/iss5/10/>.
- MOSTELLER, F. (1949). Measuring the error. En M. F. Herbert *et al.* *The Pre-election Polls of 1948, Report of the Committee on Analysis of Pre-election polls and forecasts*. Bulletin 60. New York: Social Science Research Council.
- REAL Academia Española (2020). *Diccionario de la lengua española. Edición del Tricentenario*. Disponible en: <https://dle.rae.es/>.
- TAAGEPERA, R. (2008). *Making Social Sciences More Scientific*. New York: Oxford University Press.
- TAAGEPERA, R. y Shugart, M. (1989). *Seats & Votes. The Effects & Determinants of Electoral Systems*. New Haven: Yale University Press.\*
- TARRIBA, G. (2014). Los índices: hacer concreto lo abstracto. *Este País*, núm. 282.
- VENTURA County Star (2008). *Famous lasting words about elections - From Will Rogers to Yogi Berra, immortal quotations for Tuesday*. Citado en: <https://citas.in/autores/robert-orben/>