

Cómo se mide la exactitud de las encuestas electorales

Peña, Ricardo de la

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Peña, R. d. I. (2015). Cómo se mide la exactitud de las encuestas electorales. *Política y Cultura*, 44, 217-247. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-458895>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

Cómo se mide la exactitud de las encuestas electorales*

*Ricardo de la Peña**

Resumen

La intención de este texto es caracterizar y definir de manera operativa los conceptos que las encuestas intentan medir y evaluar los estimadores existentes para calcular la exactitud de las encuestas electorales, como prelude a la presentación en un ensayo posterior de un estimador alternativo para medir dicha exactitud y diversos estimadores complementarios, con lo que se busca amortiguar problemas detectados en las opciones disponibles.

Palabras clave: medición, precisión, exactitud, encuestas, elecciones.

Abstract

The intention of this paper is to characterize and define operatively the concepts that surveys attempt to measure and evaluate the existing estimators to calculate the accuracy of polls, as a prelude to the presentation in a later article an alternative estimate of the accuracy and various complementary estimators, thereby cushioning problems detected in the measurement options available.

Key words: measurement, precision, accuracy, polls, elections.

Artículo recibido el 20-02-15

Artículo aceptado el 21-09-15

* El autor del presente artículo desea explicitar su agradecimiento al doctor Edmundo Berumen por una revisión formal del contenido de este documento sin la que hubiera sido imposible su realización. Sin embargo, el contenido final del texto es responsabilidad exclusiva de quien lo firma.

** Indagaciones y Soluciones Avanzadas, SC. [ricartur@prodigy.net.mx; ricartur@gmail.com].

Después de más de cincuenta años de encuestas electorales [...] no ha sido adoptada por la comunidad de encuestadores una métrica estándar para medir la exactitud de las encuestas.

Warren Mitofsky

INTRODUCCIÓN

El alcance de este ensayo, por sí sólo, es limitado: de interés fundamentalmente para quienes hacen y utilizan la investigación demoscópica, su cabal sentido será adquirido cuando se complemente con la propuesta de un estimador alternativo que corrija y supere los errores detectados.

En el desarrollo de este texto se busca que su contenido pueda ser entendido por cualquier persona con conocimientos medios de matemática, pero expuesto con el rigor requerido para eliminar cualquier ambigüedad. Por ello, se explicita y formaliza cada paso dado, pues más tarde se trataría no sólo de proponer un nuevo estimador, sino de generar el algoritmo que lo calcule apegado al criterio que se define para su estimación.

En lo posible, en este texto se siguen las normas convencionales de tipografía y notación matemática, privilegiando además el empleo de una única letra para etiquetar un símbolo, usualmente cursiva, para evitar confusiones.¹

¹ Rein Taagepera, *Making Social Sciences More Scientific*, Nueva York, Oxford University Press, 2008, p. 56.

CONSIDERACIONES PRELIMINARES

Una encuesta es un estudio científico que no modifica el entorno, cuyo objetivo es producir información que permita estimar la ocurrencia de eventos del pasado reciente, dando periodos específicos de referencia, percepciones vigentes sobre temas actuales o métricas específicas sobre temas factuales imperantes en el periodo de medición. Regularmente, ello se consigue a partir del llenado 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.

El principal fundamento detrás del término científico aplicado a 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, no necesariamente igualitaria en la fase de selección, y que otorga para fines de estimación un peso equitativo a cada ciudadano, calculado como el inverso de su probabilidad de selección. De cumplirse esto, los datos que se obtengan permitirán hacer inferencias sobre los valores esperables para un reactivo dentro la población considerada, en caso de que en lugar de haber medido a una muestra se hubiera realizado un censo bajo las mismas condiciones en que se realizó la encuesta.

Luego, el diseño no debiera juzgarse a partir de un único ejercicio de medición, sino a partir de la totalidad de posibles resultados que pudieran tenerse al aplicar ese diseño: la distribución muestral, mediante la cual se puede estimar la precisión o el error esperado con un tamaño de muestra determinado, conforme al paradigma de la inferencia basada en el diseño muestral.²

Debemos recordar que, en estadística, precisión y exactitud no son conceptos equivalentes,³ aunque estén relacionados. 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. Luego, refiere a la magnitud escalar con la que se mide la proximidad de los

² Jerzy Neyman, "On the Two Different Aspects of the Representative Methods: The Method of Stratified Sampling and the Method of Purposive Selection", *Journal of the Royal Statistical Society*, Series A 97, 1934, pp. 558-625.

³ Joint Committee for Guides in Metrology, *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*, 2000, 2012, 2008 (version with minor corrections), pp. 21-22.

resultados de una muestra para reproducir los resultados que se obtendrían de un recuento completo llevado a cabo usando las mismas técnicas. El diseño de una buena encuesta probabilística busca la generación de estimadores con la menor dispersión posible, dadas las condiciones concretas para su realización.

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, y cuando la diferencia es distinta de cero mide el 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. El diseño de la encuesta puede privilegiar la adopción de estimadores teóricamente insesgados o bien utilizar estimadores sesgados, pero con alta precisión, según lo que se considere óptimo en el diseño.

Es muy probable que el estadístico elegido para la medición, aun cuando sea teóricamente insesgado y eficiente, o sesgado pero de mayor precisión, no estime exactamente el parámetro poblacional con la muestra particular que resulte seleccionada por el diseño utilizado. Lo relevante es la confianza que se tiene en que réplicas del procedimiento se abultarán en valores cercanos al que se desea estimar y que la bondad de los diseños probabilísticos permitirán con los datos de la única muestra disponible estimar también, al nivel de confianza que se desee, el nivel de precisión de cualquiera de las estimaciones que con los datos de la encuesta se calculen.

Por ello, lo pertinente al evaluar la calidad de un estudio por encuesta en general, y uno electoral en particular, es lo riguroso de los procedimientos utilizados, no la adecuación que hayan podido tener con aquellos que se hubieran obtenido en caso de un censo, bajo los mismos procedimientos de la encuesta, en lugar de la estimación muestral.

Cabe precisar que existe acuerdo entre los especialistas del campo sobre los datos mínimos que deben difundirse respecto de todo estudio por encuesta que se publicite: quién realizó el estudio, quién lo patrocinó y difundió, cuándo se hizo, a qué población se observó, cuál fue la unidad primaria de muestreo, su cantidad y método de selección, el tamaño de la muestra final, el error genérico de las estimaciones a un nivel de confianza dado, el método de aproximación, la tasa de rechazo a la entrevista y, claro, los resultados obtenidos en las preguntas relevantes y si ellos corresponden a frecuencias observadas o a un modelo de estimación. La disposición de estos datos, junto con los detalles del diseño adoptado, permitiría en principio evaluar la calidad de un estudio.

Para el grueso del público interesado, sin embargo, lo relevante para evaluar la calidad de las encuestas electorales es su proximidad con los resultados, entendiendo las estimaciones por muestreo bajo un concepto radicalmente diferente al formal: el público interesado entiende los resultados de las encuestas en el terreno electoral como “pronósticos” de eventos futuros, a diferencia de lo que hasta ahora hacen la mayoría de los encuestadores; estimar eventos pasados o vigentes en el momento de la entrevista. Buscan luego disponer de un único medidor del éxito logrado por las mediciones, cualquiera que sea la medida de este éxito.

Y hablar de un único medidor implica asumir que lo que se demanda es la búsqueda de un estimador agregado de la exactitud que de cuenta de este fenómeno en un único número.

Por cierto, es posible sustentar el carácter no pronosticador de las estimaciones por encuesta no solamente en las condiciones y límites implícitos a su generación y el alcance explícitamente reconocido por sus realizadores, sino en la existencia de opciones para el tratamiento de la información que sí constituyen propuestas expresamente orientadas al pronóstico y que están diferenciadas de las encuestas, aunque puedan partir total o parcialmente de éstas. Dos ejemplos de estos instrumentos de previsión son las proyecciones con base en la detección de tendencias en colecciones de mediciones a partir de modelos de series de tiempo, que permiten determinar lo esperable en el momento de la elección con base en el registro histórico dado por las propias encuestas. Esto, sin embargo, supone asumir criterios adicionales a los derivados directamente del acto de medición por encuesta –sea que se tome la tendencia agregada o que se separe en componentes detectables para su tratamiento particular–, entre los cuales está la preservación del sentido y magnitud del cambio detectado por la serie hasta el día de la elección, lo que no siempre resulta corroborado en la realidad.

Existen asimismo los propiamente llamados modelos de pronóstico electoral, desarrollados y publicitados en diversas naciones desde hace más de tres décadas.⁴ La racionalidad detrás de estos modelos es identificar los factores que normalmente influyen en el voto, revisar los datos disponibles para medirlos y detectar cómo han afectado históricamente el voto. Regularmente, estos modelos adoptan como componente relevante las encuestas electorales, pero se alimentan de otras fuentes de información. Estos ejercicios son también imperfectos, debido a la limitada disposición de indicadores adecuados y homogéneos, referidos a periodos suficientemente prolongados, y a que cada

⁴ James E. Campbell, “Forecasting the 2012 American National Elections”, *Political Science & Politics*, American Political Science Association, vol. 45, núm. 4, 2012, pp. 610-613.

campaña tiene su propia idiosincrasia, afectando el voto en forma imprevisible. Una asunción implícita de estos modelos, como en la proyección de series temporales, es que no se presentarán eventos peculiares entre el momento de elaboración del pronóstico y la elección que provoquen que el resultado difiera del esperado a partir de los factores incorporados en el modelo.

Si una medición es un procedimiento para asignar un número a un evento conforme reglas definidas, de manera tal que permita graduar una propiedad, como lo propone Duncan,⁵ 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 oficialmente 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.

El empleo de este procedimiento no solamente responde a esta lógica, sino a razones empíricas, dada la recurrente coincidencia entre lo medido y el resultado oficial, que alimenta la expectativa depositada en el instrumento como supuesto anticipador del reparto de votos. Es primordialmente por eso que los medios las difunden y que el público las atiende, aunque ello no sea científicamente sustentable. La importancia otorgada a estos contrastes tiene que ver, además, con la relevancia pública de los estudios, su eventual, pero indemostrado, impacto en el electorado y la supuesta disposición de un parámetro contra el cual contrastar.

En términos estrictos, tiene sentido hablar de margen de error muestral sólo cuando se trata de estimaciones sustentadas en diseños y procesos de selección de muestras probabilísticas, debidamente observados en las fases de campo y estimación. Sólo entonces se pueden calcular los márgenes de error y reportarlos para cualquiera de las estimaciones realizadas, y sólo entonces se puede identificar cuál de ellas presenta el margen de error máximo y a cuál parámetro estimado se refiere.

Al respecto, existen limitaciones para el análisis secundario del error estadístico de los datos generados por los estudios publicados, pues aunque algunas encuestadoras reportan el cálculo preciso y particular del mismo para cada opción de respuesta en el reactivo utilizado para detectar las preferencias, como debiera ser, en otros casos este dato se omite, reportándose sólo el margen de error máximo implícito para un diseño muestral, cual si fuera un muestreo aleatorio simple. Ello es práctica usual en todo el mundo. Esto hace inviable generar un estadístico para el cálculo homogeneizado de la

⁵ Robert F. DeVellis, *Scale Development: Theory and Applications*, Applied Social Research Methods Series, vol. 26, Londres, SAGE Publications, 2003, p. 5.

proximidad de las mediciones respecto al resultado que considere el error estadístico observado para cada componente de la estimación de preferencias, tomando en cuenta el diseño particular y la varianza observada en cada medición de un estudio.

Debe entonces optarse por construir algún método de aproximación que elimine las objeciones planteadas a los métodos de cálculo existentes, usando la información disponible para todos los estudios a comparar, después de excluir todos aquellos para los que claramente no tiene sentido alguno calcular niveles de precisión, por no estar sustentados en diseños probabilísticos.

Esto, sin embargo, puede hacerse relajando la demanda de cumplimiento estricto de un procedimiento de selección aleatorio en la última etapa, pues adoptar un criterio riguroso, aunque pudiera ser lo correcto, implicaría desechar una parte importante de las mediciones, lo que impediría cumplir con el objetivo de cotejo de los estudios disponibles, resultando un criterio estrecho en la práctica.

Ello tiene que ver también con el hecho de que en diversas naciones se continúa con la lógica de identificar a la muestra más como una “miniatura de la población”, acorde con lo originalmente postulado a fines del siglo antepasado por Kiaer,⁶ que como un proceso de selección objetivo utilizado para obtener estimaciones, lo que deriva en la adopción de mecanismos intencionales para el balanceo de casos en muestra, sobre todo en la última etapa de selección, conforme características consideradas relevantes.

Asimismo, tiene que ver con la adopción de procedimientos similares, que buscan disponer de una cantidad previamente definida de casos por punto de muestreo, permitiendo la continuación de recorrido en puntos de entrevista u otras estrategias de sustitución de casos ausentes por problemas para el encuentro de residentes, o bien por no existir quién cumpla condiciones para ser incluido en la muestra, rechazos o entrevistas incompletas, lo que aunque rompe formalmente con el esquema de aleatoriedad, suele ser un criterio de selección validado en algunos estudios, dado el efecto marginal de la selección en la última etapa para el cálculo de la varianza. Además, en muchos casos es imposible saber cuál fue el método de selección utilizado en esa última etapa, por no detallarse dicho procedimiento en los reportes hechos públicos.

Otra objeción a la exclusión indiscriminada de todo estudio que no adopte un muestreo estrictamente probabilístico es la imperfección entre el diseño y

⁶ William Kruskal y Frederick Mosteller, “Representative Sampling IV: the History of the Concept in Statistics, 1895-1939”, *International Statistical Review*, vol. 48, pp. 169-195.

lo realizado en los estudios que sí lo adoptan, pues éstos enfrentan el riesgo real de reducción o supresión de la aleatoriedad debido a limitaciones para el encuentro o la cooperación de los individuos que debieran ser observados y a imprecisiones en el trabajo de campo, lo que afecta desde el proceso de toma de datos y sorteo de residentes en la vivienda –por negativas de los informantes a proporcionar a desconocidos información detallada–, hasta los procedimientos de nueva visita posterior para el encuentro, considerando la alteración constatada en los patrones de respuesta que genera el conocimiento de la intención de entrevistar y la insistencia en el logro de una respuesta por personas poco proclives a participar. Aunque todos estos aspectos suelen ser reparados, es inevitable la posible presencia de sesgos por imputación, pues “estrictamente hablando, una muestra probabilística obtenida incompletamente deja de ser probabilística, aunque se continúe llamándola así”.⁷

El método de cálculo buscado resultaría entonces de construir un estadístico que permita calcular la relación entre el error respecto al resultado de un censo al momento de la medición, contra la diferencia entre éste y el oficial, para cada proporción estimada con el tamaño de la muestra utilizada, aunque tomándola como si fuera una muestra aleatoria simple para fines de estandarización, lo que supone excluir del cálculo el efecto del diseño.

Como medidor de exactitud de las encuestas electorales en general, el contraste entre las estimaciones de preferencias previas a la jornada electoral y los resultados electorales es equívoco e injusto, salvo tal vez en el caso de las encuestas de salida. Las encuestas preelectorales están sujetas a muy distintas fuentes de divergencias con el supuesto parámetro que a futuro las juzgará, más que la mayoría de ejercicios demoscópicos, debido al menos a la diferencia entre el universo medido, que es el posible de observar en cada encuesta preelectoral, contra el subuniverso de éste que efectivamente acude a votar el día de la jornada electoral, que hace incorrecto su contraste directo.

Esto, dejando de lado otras fuentes de diferencia práctica, como el hecho de que el ejercicio de recuperación de las intenciones de voto que se realiza en una encuesta parte de condiciones distintas a las que incidirán en la decisión final de los electores que luego efectivamente se convertirán en votantes, principalmente por el hecho de que en las encuestas se acude al domicilio del entrevistado, lo que es una condición artificial y en ocasiones incómoda, o se le contacta a través de un sistema de comunicación que el entrevistado

⁷ Alan Stuart, “Sample Surveys II: Nonprobability Sampling”, en David I. Sills (ed.), *International Encyclopaedia of the Social Sciences*, vol. 13, Nueva York, Macmillan and Free Press, 1968, pp. 612-616.

tiene inmediatamente a la mano, mientras que para registrar su decisión final al votar, el elector debe tomar la decisión de salir de su domicilio y acudir al lugar que le corresponda para ejercer su derecho a votar.

En rigor, por consideraciones prácticas, presupuestales, de tiempo y por sentido común no conocemos ni podemos conocer ninguno de los parámetros a estimar en los cambiantes universos de cada medición preelectoral, mucho menos el futuro parámetro frente al cual luego se pretende contrastar las preferencias de la totalidad de los electores en el momento de realizar la medición para con ello evaluar la calidad de cada encuesta. Lo que es más, “algunos científicos creen que los valores verdaderos no existen separados del proceso de medida a utilizar”.⁸ De asumirse tal lógica, el estimador pertinente del verdadero parámetro poblacional sería el promedio de los estimadores centrales de las repetidas mediciones que se realicen. Luego, la exactitud de una estimación particular debiera calcularse por la diferencia contra la media del conjunto de estimaciones disponibles, la llamada “encuesta de encuestas”.

Aunque ello permite identificar los datos que resultan significativamente diferentes a otros dentro de la colección, estimaciones atípicas o inconsistentes respecto al grueso de la evidencia, no dilucida si tales mediciones son producto de errores o si son casos periféricos dentro de distribuciones con elevada curtosis o pesadas colas, lo que de ser el caso advertiría de la inconveniencia de presuponer una distribución normal para la propia colección de mediciones y eventualmente la pertinencia de tomar un estimador alternativo de la tendencia central tal como la mediana, dado su carácter robusto.

Sin embargo, adoptar la lógica de medir la exactitud a partir de la tendencia central del conjunto de estimaciones realizadas implica reducir las propiedades que se demanda de una medición a su fiabilidad, o grado de consistencia y estabilidad de las puntuaciones obtenidas a lo largo de sucesivos procesos de medición con un mismo instrumento. Al asumirse como “verdadero” el valor medio de las estimaciones, se descarta la posible existencia de errores sistemáticos y persistentes que afecten a todos o la mayoría de los elementos del conjunto de estimaciones realizadas y provoquen un sesgo en la tendencia central. Ello supone una petición de principio que, si bien es sustentable desde un punto de vista estrictamente estadístico, evade el asunto clave de la validez de los ejercicios: la determinación de la capacidad que tiene el propio instrumento de medición para cuantificar de manera significativa y adecuada el rasgo para cuya medición fue supuestamente diseñado (en este

⁸ Frederick Mosteller, “Errors I: Nonsampling Errors”, en David I. Sills (ed.), *International Encyclopaedia of the Social Sciences*, vol. 5, Nueva York, Macmillan and Free Press, 1968, p. 114.

caso, la medición de preferencias electorales de una población determinada) mediante su contrastación con evidencias empíricas externas al propio acto de medición, lo que llevaría a tener que definir un criterio distinto de validación. En el caso de las encuestas sobre preferencias electorales, el único criterio de validación externo que resulta pertinente sería en todo caso, como suele hacerse, el resultado oficial de la contienda.

Ahora bien: solamente partiendo del reconocimiento de las limitaciones particulares, adicionales a las propias a cualquier ejercicio de medición por encuesta, un cotejo entre estimaciones y futuros resultados oficiales puede servir para especular de manera razonada posibles eventos propiciadores de una magnificación de las divergencias de por sí esperadas, por la sencilla razón de referirse a universos y tiempos diferentes, las cuales pueden ser endógenas a la investigación, poniendo en evidencia errores de diseño o problemas de aplicación, o exógenas, como las antes referidas y las posibles equivocaciones involuntarias o interesadas en la contabilidad de votos, entre otros factores potencialmente presentes.

Desde esta perspectiva, no puede verse como impertinente la insistencia colectiva y mediática de comparar los estimadores producidos en encuestas previas contra los resultados oficiales de una elección. El rechazo categórico del cotejo representaría la negativa a toda forma de examen *a posteriori* que pudiera arrojar luz para identificar metodologías más o menos eficientes bajo esta óptica y contrastar hipótesis pretendidamente explicativas de divergencias observadas, que permitan la detección de factores incidentes en la capacidad misma de medición y anticipación por encuesta de los resultados.

Todo ello, sin soslayar el hecho ya mencionado de que las encuestas no tienen en sí mismas una finalidad predictiva y de que “ni siquiera es evidente que las predicciones se deben buscar siempre, o siempre son deseables”.⁹

Al respecto, debe apuntarse que si bien algunas etapas del proceso productivo de una encuesta están sujetas a un rigor científico que puede ser controlado a cabalidad, como el diseño y selección de la muestra, o el proceso de estimación apegado a diseño, otras etapas no tienen este carácter a plenitud, aunque pueden amortiguarse mediante pruebas piloto. Procesos como la elaboración de instrumentos para la recuperación de datos, en particular lo relacionado con decisiones sobre ubicación, redacción y mecánica de aplicación del reactivo electoral; el acopio y control de la toma de información

⁹ Audun Øfsti y Dan Østerberg, “Self-defeating Predictions and the Fixed-point Theorem: A Refutation”, *Inquiry: An Interdisciplinary Journal of Philosophy*, vol. 25, núm. 3, 1982, pp. 331-352.

en campo, problemas de cobertura por razones de seguridad del personal de campo o del informante, efecto de los criterios adoptados para la selección particular de los informantes y la potencial sistematicidad en los patrones de autoselección del respondedor; así como los procesos posteriores tales como la validación, codificación y digitalización, están sujetos a errores no descartables, aun y cuando se sigan protocolos científicos o regulaciones de calidad que deben regir la actividad. El sometimiento a las reglas del “bien hacer” simplemente tenderá a minimizar la ocurrencia de errores, pero no los eliminará. Por ello, buscar las fórmulas para la detección y estimación de la presencia y magnitud de errores es fundamental para el mejoramiento del quehacer demoscópico.

Pero hay un abuso común en el cotejo de las estimaciones de las encuestas: la comparación entre diversas mediciones, y su ordenamiento según la proximidad entre las estimaciones públicas y los resultados oficiales.

Ello suele realizarse para varios tipos de agregados, pudiendo ser dos de ellos valiosos, nuevamente sólo después de desechar a todos aquellos ejercicios que no tiene sentido alguno calcular sus niveles de precisión, por la simple y sencilla razón de no estar sustentados en diseños probabilísticos: el cálculo para conjuntos de estimaciones previas de las distintas encuestadoras en una misma elección, cotejando luego la exactitud alcanzada en una elección con la obtenida en otras ocasiones; y el cálculo para conjuntos de estimaciones de una misma encuestadora en diversas elecciones semejantes, comparando la exactitud media lograda por una encuestadora contra la de otras dentro de una ventana temporal definida.

Empero, en la mayoría de las ocasiones lo que se hace es simplemente determinar el ordenamiento entre encuestadoras conforme al error observado en una única elección. Aunque ello es posible, contraviene el principio conforme al cual no debe juzgarse la calidad de los diseños de una encuestadora a partir juicios emitidos en un único ejercicio de medición, que responde eminentemente a lo fortuito, sino en todo caso juzgarla por su desempeño en diversas mediciones, posibilitando que los errores de carácter estocástico tiendan a tomar un nivel normal, lo que dejaría como excedente un error de carácter sistemático, cuyo reconocimiento pudiera ser de utilidad práctica.

DEFINICIONES BÁSICAS

Antes, es preciso comenzar por el principio: en una elección cada individuo que forma parte del universo de electores manifiesta mediante un procedimiento

consensuado una decisión que ha tomado a lo largo del tiempo. Esta decisión puede diferir entre el momento de toma de datos en las diversas encuestas y la fecha de la jornada electoral, aunque el lapso tiende a acortarse a medida que se aproxima el día de la elección, cuando los ciudadanos que votan eligen entre diversas opciones para definir quién ocupará determinado cargo de representación en disputa (o entre las diversas opciones sobre un asunto de interés público sobre el que se les consulta, posibilidad que aunque existente no resulta esencialmente distinta a la elección entre contendientes por un cargo para los fines de la argumentación).

Como resultado de una elección, el universo finito y numerable de electores (N) se divide en tres grupos: los votantes, referido a aquellos que votaron por cada uno de los contendientes; los abstencionistas activos, aquellos que votaron por alguien que no competía o cuyo voto fue invalidado, voluntariamente o incidentalmente; y los abstencionistas pasivos, quienes no acudieron a votar.

Sea:

“ V_i ” el conjunto de votantes que emitieron un voto válido a favor del i -ésimo candidato registrado en la elección ($i = 1, 2, 3, \dots, m$),

“ V_0 ” el total de votantes cuyo voto no fue validado por cualquier razón,

“ V' ” el total de electores que no acudieron a votar.

Entonces (1):

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

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.

Se parte del principio de correspondencia formal, que pudiera no ser real, entre los votos contabilizados y los emitidos por los electores. De hecho, existe evidencia de que, en prácticamente cualquier elección, ocurren eventos que derivan en errores en la contabilidad: equívocos en el registro o recuento de sufragios, errores en la captura o recuperación de los datos, correcciones realizadas conforme normas por órganos revisores, e incluso manipulaciones de los datos oficiales. Sin embargo, al final de un proceso electoral que es

legalmente válido, existe un resultado que corresponde a una cuantificación oficial de los votos, y a ella debe uno atenerse.

A partir de esta colección, una vez que fueron eliminados los votos no válidos, es posible obtener las proporciones de voto por cada contendiente registrado (p_i) respecto al total de votos válidos (2):

$$p_i = \frac{V_i}{\sum_{i=1}^m V_i} \quad (2)$$

La siguiente suma es necesariamente igual a la unidad (3):

$$\sum_{i=1}^m p_i = 1 \quad (3)$$

Tanto con la cantidad de votos como con las proporciones calculadas, es posible obtener un ordenamiento de los contendientes y con ello determinar al candidato que obtuvo la mayor cantidad de votos en la contienda.

Digamos que el candidato “ w ” es quien más votos obtuvo, entonces, si para cualquier $i \neq w$ (4):

$$V_w \geq V_i \quad \square \quad p_w \geq p_i \quad (4)$$

Puede estimarse la diferencia entre dos contendientes cualesquiera como (5):

$$p_{i-j} = \frac{V_i - V_j}{\sum_{i=1}^m V_i} \quad (5)$$

Ahora bien, como resultado directo de una encuesta, el universo finito y contable de electores que fue observado (n) se divide en tres grupos (6): quienes respondieron que votarían por alguno de los diversos contendientes (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 (v_0) y quienes no respondieron que votarían por alguno de los contendientes (v') sea porque no saben por quién votarían, porque no piensan votar por alguno de los contendientes o porque por razones sólo por ellos conocidas no respondieron al reactivo aplicado para conocer la preferencia, quienes no necesariamente se convertirán en no votantes.

$$n = \left(\sum_{i=1}^m v_i \right) + v_0 + v' \quad (6)$$

El resultado inmediato de una encuesta es luego la suma de casos que declaran tener la intención de votar por cada uno de los diversos contendientes considerados, más los casos cuya respuesta termina anulada, más los casos no definidos por algún contendiente.

Sin embargo, esta suma de casos no necesaria ni usualmente corresponde a frecuencias simples, sino que corresponde a estimadores obtenidos mediante el empleo de factores de expansión, calculados como el inverso de la probabilidad de selección de cada caso de acuerdo con la distribución final de la muestra y que, por ende, responden al diseño adoptado; y/o de ajustes para aproximar la distribución de la muestra a la poblacional para determinadas características de las que se cuenta con parámetros y que son relevantes para los fines del estudio (7):

$$\hat{N} = \left(\sum_{i=1}^m \hat{V}_i \right) + \hat{V}_0 + \hat{V}' \quad (7)$$

En todo caso, el tamaño de la muestra a considerar para fines de la posterior estimación de niveles de precisión debe ser el número real de casos observados, y no la expansión al universo total de la población bajo estudio.

A partir de la suma de casos, es posible obtener las proporciones de voto por cada contendiente respecto al total de casos observados (\hat{r}_i), como (8).

$$\hat{r}_i = \frac{\hat{V}_i}{\hat{N}} = \frac{\hat{V}_i}{\left(\sum_{i=1}^m \hat{V}_i \right) + \hat{V}_0 + \hat{V}'} \quad (8)$$

Aunque tanto con la cantidad de casos observados con preferencia definida a favor de cada contendiente como con su expresión en proporciones es posible obtener un ordenamiento de los contendientes conforme los resultados de la encuesta, ello no resulta pertinente en tanto no se estimen por igual los niveles de precisión de cada estimación respecto al valor esperado del estimador usado y los intervalos de confianza correspondientes para examinar si se traslapan. Digamos que este ordenamiento es (9):

$$\hat{r}_1 \geq \hat{r}_2 \geq \hat{r}_3 \geq \dots \geq \hat{r}_m \tag{9}$$

Donde:

- 1: es el candidato que resultó con la estimación mayor a cualquiera de los otros $m-1$ candidatos,
- 2: es el candidato que resultó con la estimación mayor a cualquiera de los otros $m-1$, *excepto el candidato "1"*,
- 3: es el candidato que resultó con la estimación mayor a cualquiera de los otros $m-1$, *excepto los candidatos "1 y 2"*,
- ...
- m : es "el candidato" que resultó con la estimación menor a cualquiera de los otros $m-1$ candidatos.

Llegado el día en que se conozcan los resultados oficiales, a partir de esta asignación de un orden a cada componente conforme la encuesta es posible construir una variable dicotómica (\square) que de cuenta de la coincidencia entre el ordenamiento arrojado por una encuesta previa y el oficial (10) y de ello derivar una "tasa de acierto al orden":

$$\begin{aligned} \square &= 1 \square p_1 \geq p_2 \geq p_3 \geq \dots \geq p_m, \\ \square &= 0 \text{ para cualquier otro orden.} \end{aligned} \tag{10}$$

El valor medio de \square para cualquier conjunto de encuestas que se desee analizar pudiera utilizarse como estimador de la proporción de encuestas que aportan un ordenamiento correcto entre los contendientes en modelos o pruebas referidas a conjuntos de encuestas.

De igual suerte, es posible construir una variable dicotómica que de cuenta de la condición de coincidencia entre el ganador estimado por cualquier encuesta previa y el oficial (w) y de ahí derivar una "tasa de acierto al ganador" en las mediciones, donde para cualquier $l \neq 1$ (11):

$$\begin{aligned} w &= 1 \square p_1 \geq p_l, \\ w &= 0 \square p_1 < p_l \end{aligned} \tag{11}$$

El valor medio de w para cualquier conjunto de estudios que se desee analizar pudiera utilizarse como estimador de la proporción de casos que aportan un ganador correcto, e incluirla como variable acotada al intervalo unitario en modelos o pruebas referidas a conjuntos de estudios.

Es claro que los casos en que \square adquiere un valor positivo son un subconjunto propio de los casos en que w adquiere un valor positivo, pues el posible error en el ordenamiento puede no estar relacionado con la detección del ganador (12):

$$\square = 1 \square w = 1 \quad (12)$$

Es posible estimar una diferencia entre las proporciones observadas hacia dos contendientes cualesquiera (13):

$$\hat{r}_{i-j} = \frac{\hat{v}_i - \hat{v}_j}{\hat{N}} = \frac{\hat{v}_i - \hat{v}_j}{\sum_{i=1}^m \hat{v}_i + \hat{v}_0 + \hat{v}'} \quad (13)$$

Sin embargo, salvo en el caso infrecuente de que la no respuesta al reactivo sobre preferencia en cualquiera de sus expresiones (votará por un contendiente no registrado, no sabe por quién votará, no irá a votar o no responde) sea igual a cero, la suma de las proporciones respecto al total de casos no será igual a la unidad y, por ende, afectará su cotejo directo contra los resultados de la elección (14).

$$\sum_{i=1}^m \hat{r}_i = 1 \square \hat{v}_0 + \hat{v}' = 0 \quad (14)$$

Ello debido a que las proporciones obtenidas por este procedimiento para los contendientes deja un remanente adjudicado a los casos no definidos (15):

$$\sum_{i=1}^m \hat{r}_i = 1 - \frac{\hat{v}_0 + \hat{v}'}{\hat{N}} \quad (15)$$

De igual suerte, no tiene por qué ser cierto y es inusual que la proporción de no respuesta a la pregunta sobre preferencias en una encuesta sea igual a la proporción de no votantes en una elección (16).

$$\frac{\hat{v}_0 + \hat{v}'}{\hat{N}} \geq 0 \leq \frac{v_0 + v'}{N} \text{ y } \frac{\sum_{i=1}^m \hat{v}_i}{\hat{N}} \geq 0 \leq \frac{\sum_{i=1}^m v_i}{N} \quad (16)$$

Cabe mencionar al respecto que empíricamente se ha constatado que la proporción de entrevistados que se definen por algún contendiente en una encuesta tiende a ser mayor que la proporción de votantes.¹⁰

Por ello, resulta pertinente efectuar un ejercicio de estimación de las proporciones de intención de voto obtenidas en una encuesta a partir de (17):

$$\hat{N}_a = \hat{N} - (\hat{V}_0 + \hat{V}') \quad (17)$$

De manera equivalente, se puede definir la muestra efectiva (\tilde{N}_a) como (18):

$$\hat{N}_a = \sum_{i=1}^m \hat{V}_i \quad (18)$$

Para fines de la posterior estimación de la precisión, es conveniente calcular el tamaño de la muestra efectiva, que es asignada a los contendientes (n_a), como (19):

$$n_a = n - (\hat{v}_0 + \hat{v}') \quad (19)$$

Sabiendo el número de casos efectivos, es posible estimar las proporciones por cada uno de los contendientes como (20).

$$\hat{p}_i = \frac{\hat{V}_i}{\hat{N}_a} = \frac{\hat{V}_i}{\sum_{i=1}^m \hat{V}_i} \quad (20)$$

A partir de lo cual se puede además calcular la diferencia entre cualesquiera dos contendientes, como (21):

$$\hat{p}_{i-j} = \frac{\hat{V}_i - \hat{V}_j}{\hat{N}_a} = \frac{V_i - V_j}{\sum_{i=1}^m V_i} \quad (21)$$

La definición de \hat{p}_i , a diferencia de \hat{r}_i , provoca necesariamente que la suma de las proporciones iguale la unidad (22).

¹⁰ Israel Waismel-Manor y Joseph Sarid (2011), "Can Overreporting in Surveys be reduced? Evidence from Israel's Municipal Elections", *International Journal of Public Opinion Research*, vol. 23, núm. 4, p. 522.

$$\sum_{i=1}^m \hat{p}_i = 1 \quad (22)$$

Y, aunque modifica los valores, mantiene inalterado el ordenamiento entre los contendientes conforme al nuevo conjunto de estimadores de la encuesta (23).

$$\hat{p}_1 \geq \hat{p}_2 \geq \hat{p}_3 \geq \dots \geq \hat{p}_m \quad (23)$$

Cumpliendo con lo anterior, es pertinente comparar las proporciones de intención de voto de cualquier encuesta contra las proporciones de votos oficiales, al garantizarse que cada lado de la ecuación sea equivalente e iguale la unidad (24).

$$\sum_{i=1}^m \hat{p}_i = 1 = \sum_{i=1}^m p_i \quad (24)$$

Contrario a lo comúnmente percibido ya afirmado, este procedimiento no supone en forma alguna una imputación o asignación de preferencias para casos no definidos, sino que asume su exclusión de la muestra para fines de estimación, pero a costa de perder formalmente 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 de las proporciones de casos observados, debiendo hacerse a partir de las proporciones de casos efectivos, considerando 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 (25):

$$e_i = \hat{p}_i - p_i \because e_i \neq \hat{r}_i - p_i \quad (25)$$

Siendo una magnitud escalar cuyo valor varía de menos uno a uno (26):

$$-1 \leq e_i \leq 1 \quad (26)$$

Por lo que aporta no solamente una distancia, sino un sentido de la diferencia entre la estimación particular bajo examen y el resultado oficial.

Cabe señalar que existe una probabilidad mayor de cero de posible concordancia entre el valor estimado por una encuesta y el valor real,¹¹ si ésta es entendida como la probabilidad de existencia de un punto fijo dentro de una brecha mínima donde puedan coincidir las proporciones y si se parte de reconocer el carácter discreto del conjunto de posibles resultados de una elección, considerando la finitud del universo de votantes (27).

$$P(\hat{p}_i \equiv p_i) \rightarrow \frac{s^s - (s-1)^s}{s^s} = 1 - \left(1 - \frac{1}{s}\right)^s = 1 - \frac{1}{e} \cong 0.632 \quad (27)$$

Donde “P” denota la probabilidad de ocurrencia de un evento, “s” el número de posibles resultados de una elección, cuyo valor es necesariamente discreto dado el número finito de electores existentes, y “e” es la base del logaritmo natural, cuyo valor aproximado es 2.718 (que en este texto se anota en redondas simplemente para evitar confusiones con el símbolo de error ya utilizado). De asumirse lo anterior, más allá de cualquier precisión posible en el cálculo de proporciones, existirán casos en que el error sea nulo, aun considerando la totalidad de cifras significativas posibles.

Ahora bien, es sencillo y posible estimar una magnitud con sentido para la distancia existente entre la diferencia estimada para dos contendientes cualesquiera y la diferencia oficial (28):

$$e_{i-j} = (\hat{p}_i - \hat{p}_j) - (p_i - p_j) \quad (28)$$

Si se supone que se trata de una elección entre dos contendientes solamente, es lícito adoptar este último estimador como el medidor del error de una encuesta más simple posible, donde el signo indica el sentido del sesgo en la estimación: si es positivo se sobreestimó, si es negativo se subestimó.

Empero, cuando el número de contendientes (*m*) es mayor de dos, el cálculo a partir de este estimador es imposible, pues su suma y su promedio (\bar{e}) para el conjunto de todos los contendientes siempre será cero, dado que los errores en un sentido se compensan con los errores en sentido contrario (29).

$$\bar{e} = \frac{\sum_{i=1}^m (\hat{p}_i - p_i)}{m} = 0 \quad (29)$$

¹¹ Karl Egil Aubert, “Spurious mathematical modeling”, *The Mathematical Intelligencer*, vol. 6, núm. 3, 1984, p. 59.

Frente a ello, una solución obvia es recurrir al cálculo de la distancia euclidiana (d_e), que permite estimar la magnitud de la diferencia (30), pero perdiendo la posibilidad de definir un sentido como magnitud escalar, pues este no es unívoco cuando se trabaja en un espacio de dimensión superior a dos:

$$d_e = \sqrt{\sum_{i=1}^m e_i^2} \quad (30)$$

Sin embargo, considerando 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 (31):

$$0 \leq d_e \leq \sqrt{2} \quad (31)$$

Esto puede corregirse si tomamos la distancia 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 (32).

$$d = \frac{d_e}{\sqrt{2}} = \sqrt{\frac{1}{2} \sum_{i=1}^m e_i^2} \quad (32)$$

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 (33). Así, este dato convertido en porcentaje pudiera ser un indicador de una distancia relativa entre lo estimado y lo real:

$$0 \leq d \leq 1 \quad (33)$$

Otra 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 (34), tomando luego como error la diferencia absoluta ($|e_i|$):

$$|e_i| = |\hat{p}_i - p_i| \quad (34)$$

Estimador que corresponde estrictamente a la definición convencional del error absoluto de una medición.

Puede calcularse asimismo la distancia absoluta y, por ende, sin sentido, entre la diferencia estimada entre dos contendientes cualesquiera y la diferencia real (35):

$$|e_{i-j}| = |(\hat{p}_i - \hat{p}_j) - (p_i - p_j)| \quad (35)$$

Aun cuando este estadístico, aplicado a los dos partidos mayores, aporta un dato sobre la diferencia entre lo estimado y lo real del margen de victoria, cuando el número de contendientes considerados en el reactivo electoral es mayor de dos, la reducción del resultado de una encuesta a esta 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 de carácter no espontáneo 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.

A partir de la estimación de la diferencia absoluta entre lo estimado y lo real por cada contendiente considerado sí puede efectuarse una suma y sacar el promedio de error (e) que aporte un valor distinto de cero y que indique la distancia entre lo observado y el parámetro (36):

$$e = \frac{\sum_{i=1}^m |e_i|}{m} = \frac{\sum_{i=1}^m |\hat{p}_i - p_i|}{m} \quad (36)$$

Sin embargo, resulta importante observar que el rango en el que se mueve el valor de e no es la unidad, pues va de un mínimo de cero a un máximo igual al cociente de dos entre el número de contendientes considerado (37):

$$0 \leq e \leq \frac{2}{m} \quad (37)$$

Luego, el valor máximo de este estimador del error será igual a uno, y por ende a d , solamente cuando sean dos los contendientes, pues cuando haya más contendientes el valor de será menor (38):

$$\begin{aligned}
 m = 2 \quad & \square \quad d = e \\
 m > 2 \quad & \square \quad d > e
 \end{aligned}
 \tag{38}$$

Una alternativa para evitar este acotamiento superior es obtener el error absoluto total entre los datos observados y los reales ($|e|$), sumando las diferencias absolutas de las estimaciones con los resultados y dividiendo el resultado entre dos (39), para evitar nuevamente una doble contabilidad.

$$|e| = \frac{1}{2} \sum_{i=1}^m |e_i| = \frac{1}{2} \sum_{i=1}^m |\hat{p}_i - p_i|
 \tag{39}$$

El valor de $|e|$, al igual que el de d , se ubica luego en el intervalo entre cero y uno (40):

$$0 \leq |e| \leq 1
 \tag{40}$$

Sin embargo, cuando el número de contendientes es mayor a dos, los valores que toman ambos estimadores suelen no ser iguales.

Adicionalmente, puede definirse el error relativo de una estimación dada (ϵ_i) como (41):

$$\epsilon_i = \frac{|\hat{p}_i - p_i|}{p_i}
 \tag{41}$$

Cuyo valor va de cero, cuando la proporción estimada es idéntica a la real, hasta cerca de infinito, límite inalcanzable en términos prácticos dada la finitud de la muestra observada (42):

$$0 \leq \epsilon_i < \infty
 \tag{42}$$

Como colofón a este punto, se puede precisar qué debe entenderse y cuál es la utilidad de una “encuesta de encuestas”, que no es otra cosa que un meta-análisis de encuestas que combina los resultados de diversos estudios, haciendo tabla rasa de las diferencias entre sus protocolos de investigación.

Este ejercicio corresponde a la estimación de la media de cada componente, medido por una colección de encuestas para un mismo evento electoral dentro de una ventana temporal arbitraria. Definamos así (\bar{p}_i) como (43):

$$\bar{p}_i = \frac{\sum_{i=1}^k \hat{p}_i}{k} \quad (43)$$

Donde “ k ” son los elementos de la colección de encuestas considerada.

Como apuntamos previamente, este cálculo de la media de las estimaciones es considerado por algunos como el único estimador pertinente del verdadero parámetro poblacional que se pretende estimar, lo que supone la inexistencia por definición de un posible sesgo en el conjunto de estimaciones y, por ende, asume como implícita la capacidad del instrumento para cuantificar adecuadamente la propiedad que se pretende medir, al margen de su eventual contraste con toda evidencia empírica.

A partir del cálculo de este valor de tendencia central para un conjunto dado de estimaciones, es posible medir la diferencia entre una estimación particular y la media del conjunto (h_i), simplemente como (44):

$$h_i = \hat{p}_i - \bar{p}_i \quad (44)$$

Y obtener el valor absoluto ($|h_i|$) correspondiente (45):

$$|h_i| = |\hat{p}_i - \bar{p}_i| \quad (45)$$

Luego, puede calcularse el promedio de las diferencias entre las estimaciones por componente de un estudio específico y las estimaciones medias por componente en una colección de estudios (h_r , siendo $r = 1, 2, 3... h$, como (46):

$$h_r = \frac{\sum_{i=1}^m |\hat{p}_i - \bar{p}_i|}{m} \quad (46)$$

Lo anterior viene al caso dado que es conveniente precisar que existen claras diferencias entre obtener la media del error medio de un conjunto de encuestas (\bar{e}), que corresponde al promedio del error registrado en cada uno de los ejercicios, medido por el estimador del error medio por componente (47):

$$\bar{e} = \frac{\sum_{j=1}^k \frac{\sum_{i=1}^m |\hat{p}_{ij} - p_{ij}|}{m}}{k} \quad (47)$$

Y obtener el error medio del promedio de las estimaciones de un conjunto de encuestas ($e_{\bar{p}}$), lo que corresponde al error medio por componente de la llamada “encuesta de encuestas”, no al error registrado por las propias mediciones (48):

$$e_{\bar{p}} = \frac{\sum_{i=1}^m |\bar{p}_i - p_i|}{m} = \frac{\sum_{i=1}^m \left| \frac{\sum_{j=1}^k \hat{p}_{ij}}{k} - p_i \right|}{m} \quad (48)$$

Esto es relevante cuando el cálculo del error se efectúa para un conjunto de mediciones, pues ambos ejercicios suelen realizarse y sus resultados no son idénticos ni comparables.

Por demás, agregar las mediciones producto de distintas encuestas, si bien reduce potencialmente la volatilidad entre las estimaciones, no mejora la exactitud de manera necesaria, puesto que no siempre es cierto que el posible sesgo en las encuestas tienda a cancelarse entre sí, dado que para una elección el sesgo de diversas mediciones puede y suele presentarse en la misma dirección,¹² aspecto cuyo análisis y explicación debiera ser materia de otro ensayo.

ESTIMADORES DEL ERROR EXISTENTES

Son muy diversos los posibles estimadores para medir la proximidad entre un grupo de datos y otro. En el caso del cotejo entre estimadores producto de encuestas y resultados de una elección son muchas las propuestas presentadas, aunque pocas las realmente utilizadas.

Idealmente, un estimador de la exactitud de una encuesta o de la diferencia entre lo estimado y el parámetro, debiera cumplir algún conjunto de criterios convenidos *a priori* con un mínimo de sentido común. A manera de ilustración,

¹² Mark M. Blumenthal, “Toward an Open-Source Methodology: What We Can Learn from the Blogosphere”, *Public Opinion Quarterly*, vol. 69, núm. 5, 2005, pp. 655-669.

sin que exista necesariamente un consenso al respecto, se someten a examen las siguientes condiciones:

Primero: debiera posibilitar la comparabilidad entre mediciones y de éstas con un valor teórico esperado, bien para un conjunto de encuestas en una elección, para diversos ejercicios de una misma firma o para una colección de encuestas en elecciones diversas, permitiendo un análisis agregado, a la vez que un tratamiento particularizado que determine las relaciones entre distintas mediciones.

Segundo: debiera generarse por un procedimiento homogéneo para todos y cada uno de los contendientes, cualquiera que sea su número, siendo aplicable a resultados tanto en sistemas bipartidistas como multipartidistas; y considerar la totalidad de las opciones de respuesta predeterminadas y dadas a conocer al encuestado de respuesta al reactivo sobre preferencias usado para fines de estimación, así como aquellas que se hubieran dado de manera espontánea, adicionales a las opciones acotadas por el entrevistador.

Las propuestas más conocidas y usadas datan de hace más de medio siglo y en todos los casos reclaman la disposición de dos grupos de datos: las proporciones por contendiente observadas en los estudios y las proporciones oficiales obtenidas por los mismos contendientes.

Fue la experiencia de desencuentro entre las estimaciones por encuesta y los resultados oficiales de la elección presidencial de 1948 en Estados Unidos la que llevó a una revisión profunda de los ejercicios, que incluyó un análisis, a cargo de Frederick Mosteller, sobre los posibles métodos para mesurar el error de las estimaciones. En este texto, el autor lista varios métodos que considera factibles y eventualmente pertinentes.¹³

Es preciso mencionar que un problema que Mosteller no atiende en su momento, debido a que aún no había surgido como tal, y que ha sido fuente de divergencias y discusiones en las evaluaciones posteriores respecto al error en encuestas, es el relativo a qué datos considerar para fines de análisis. De nuevo, es un debate en torno a la supuesta inadecuación entre encuestas y resultados, ocurrido medio siglo más tarde, en las elecciones presidenciales de Estados Unidos en 1996, lo que coloca en la palestra la discusión sobre el problema de los casos no definidos.¹⁴

¹³ Frederick Mosteller, "Measuring the error", en Frederick Mosteller, Herbert Hyman, Philip J. McCarthy, Eli S. Marks y David B. Truman, *The Pre-election Polls of 1948, Report of the Committee on Analysis of Pre-election polls and forecasts*, Bulletin 60, Nueva York, Social Science Research Council, 1949, capítulo V, p. 55.

¹⁴ Warren J. Mitofsky, "Review: Was 1996 a Worse Year for Polls Than 1948?", *The Public Opinion Quarterly*, vol. 62, núm. 2, 1998, p. 233.

Los reportes de resultados de las encuestas electorales pueden referir a diversas salidas de información: una primera dicotomía es la existente entre proporciones observadas, incluyendo como un componente diferenciado los casos no definidos por algún contendiente, contra las proporciones efectivas, que excluyen los casos no definidos o suponen su asignación. Otra dicotomía es entre los datos directamente obtenidos, con las ponderaciones acordes al diseño adoptado, y los datos ajustados conforme a la probabilidad estimada de sufragio, producto de los llamados modelos de “votantes probables” u otras opciones para el cálculo de proporciones.

En el caso de Estados Unidos, las proporciones de casos no definidos suelen situarse en niveles relativamente reducidos,¹⁵ por lo que las diferencias en los cálculos de error entre los datos observados y los resultantes de la exclusión o asignación proporcional de no definidos no es tan relevante. Ello tiene como trasfondo una visión de los profesionales estadounidenses, para los cuales el tratamiento pulcro de los datos debiera privilegiar el reporte de lo observado tal cual fue recuperado, sin manipulación alguna, pero que a la vez concibe los niveles de no definición en la pregunta electoral como reflejo de problemas en el diseño de instrumentos o de carencias del debido rigor en los procedimientos operativos, lo que sería al menos cuestionable; de hecho, suele considerarse pertinente que el investigador deba realizar esfuerzos para lograr niveles de respuesta elevados.

Empero, en otras democracias, la proporción de casos no definidos en una encuesta suele ubicarse en niveles más elevados y considerarse reflejo de una actitud real de indefinición del elector y no como un problema de técnicas de acopio de datos, por lo que no efectuar una exclusión o asignación proporcional lleva a estimaciones poco diáfanas de los niveles de error registrados. Es por ello que, antes de un tratamiento para el cálculo de los errores en las estimaciones, suele efectuarse una homologación de los datos, aplicando procedimientos de cálculo que excluyen o asignan los casos no definidos mediante algún procedimiento uniforme (o desproporcional si así lo determina la experiencia y creatividad del analista).

No existe hasta ahora un consenso respecto al tratamiento pertinente de los casos no definidos, aunque sí han surgido propuestas de estimadores que buscan expresamente resolver de manera práctica este asunto.

¹⁵ Véanse como ejemplo los compendios de encuestas nacionales presidenciales del National Council on Public Polls [<http://www.ncpp.org/?q=node/137>], que indican una proporción de indefinidos en torno al 6%, contra los niveles superiores al 10% registrados en países como México, según puede consultarse en [http://www.ine.mx/documentos/proceso_2011-2012/EncuestasConteosRapidos/inicio.html].

De los diversos métodos para el cálculo del error de las encuestas resumidos por Mosteller, siete de ellos resulta de interés recuperar, por basarse en la relación entre las proporciones de voto estimadas y observadas (un octavo refiere a la relación entre estimación y reparto real de votos electorales, lo que es propio de un sistema de elección indirecta, como Estados Unidos).

La mayoría de éstos no dan respuesta alguna sobre el posible tratamiento de los casos no definidos. Mosteller pareciera asumir que se debieran tomar los datos tal cual se reportan, sin recálculo alguno para excluir la proporción no definida en la muestra. En análisis posteriores, sin embargo, es usual que se recurra a procesos de asignación o exclusión, no siempre idénticos.

Hagamos un recuento de los métodos propuestos por Mosteller, proponiendo una notación que los precise y señalando sus alcances y limitaciones, siguiendo al autor y a otros analistas posteriores, particularmente Mitofsky.

El recorrido lo realizaremos no siguiendo el orden originalmente dado en el texto por Mosteller, sino en razón al número de componentes considerados para la estimación del error, de menos a más, para clarificar la exposición.

Además, dada su relevancia como aspecto generador de ambigüedad y divergencias entre cálculos del error para idénticos conjuntos de encuestas, se recobrará la notación relativa a las dos opciones fundamentales de estimación: con los datos directamente obtenidos, incluyendo un segmento de casos con una intención de voto no definida; y con los datos relativos exclusivamente al segmento de casos definidos por algún contendiente.

M1. La diferencia en puntos porcentuales entre la proporción prevista para el ganador y la proporción oficial obtenida respecto al total de votos emitidos, que pudiera expresarse como (49):

$$M1 = \begin{cases} |\hat{r}_1 - p_1|, \\ |\hat{p}_1 - p_1| \end{cases} \quad (49)$$

Este método resulta el más simple posible y ocupa solamente dos datos para fines de estimación del error, lo que lo hace intuitivo para cualquiera. Empero, como Mitofsky advierte, este método resulta artificial, sobre todo cuando el líder cuenta con menos de la mitad de los votos por existir terceros contendientes o al incluirse al segmento de indefinidos.

M2. La diferencia en puntos porcentuales entre las proporciones predichas y reales de votos recibidos por los dos mayores contendientes, que pudiera expresarse como (50):

$$M2 = \left| \left(\frac{\hat{r}_1 - \hat{r}_2}{\sum_{i=1}^2 \hat{r}_i} \right) - \left(\frac{p_1 - p_2}{\sum_{i=1}^2 p_i} \right) \right| = \left| \left(\frac{\hat{p}_1 - \hat{p}_2}{\sum_{i=1}^2 \hat{p}_i} \right) - \left(\frac{p_1 - p_2}{\sum_{i=1}^2 p_i} \right) \right| \quad (50)$$

Este método, sencillo de calcular y de fácil entendimiento, cuenta con la clara ventaja de ser el único de los propuestos por Mosteller que no se afecta por la inclusión o no de los casos no definidos. Ello se logra eliminando del cálculo las proporciones correspondientes a opciones menores regularmente reportadas por separado.

Mosteller privilegia los dos primeros métodos sobre otros, debido a problemas que surgen al aplicar las opciones restantes. Pero ello es consecuencia más que de una regla generalizable a todo sistema electoral, del carácter propiamente bipartidista de la realidad que atiende este autor: las elecciones en Estados Unidos.

M5. La diferencia absoluta entre lo previsto y lo real para los dos mayores contendientes respecto al total de votos emitidos, expresable como (51):

$$M5 = \begin{cases} |(\hat{r}_1 - \hat{r}_2) - (p_1 - p_2)|, \\ |(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)| \end{cases} \quad (51)$$

En contiendas bipartidistas, M2 y M5 resultan ser equivalentes, si se asume la inexistencia o la asignación proporcional de los casos indefinidos. M5 corresponde por demás con lo regularmente resaltado en los reportes mediáticos: el margen de ventaja del líder. Por lo anterior, ha sido usado regularmente en los análisis sobre el tema, incluso por el ámbito académico estadounidense.

Sin embargo, al considerar las proporciones para los dos primeros lugares respecto al total de todos los componentes, pierde la virtud de arrojar un estimador que no se afecte por el problema de los casos no definidos.

M3. La desviación media en puntos porcentuales entre lo previsto y lo real para todos los contendientes, sin tener en cuenta el signo, que pudiera expresarse como (52):

$$M3 = \begin{cases} \frac{\sum_{i=1}^m |\hat{r}_i - p_i|}{m}, \\ \frac{\sum_{i=1}^m |\hat{p}_i - p_i|}{m} \end{cases} \quad (52)$$

El tercer método, si bien toma en cuenta todos los componentes para generar un único estimador correspondiente a lo intuitivo y sencillo de calcular, no permite una comparación diáfana cuando el número de partidos es variable e impide reconocer el sesgo en la medición. Asimismo, el grado de error no se encuentra relacionado con las proporciones de voto obtenidos, sino que asume un valor absoluto, por lo que puede derivar en cálculos reducidos de error, al incluir a diversos componentes menores que poco contribuyen al voto y cuya estimación suele presentar divergencias absolutas pequeñas, por lo que promediarlos con los componentes mayores propicia una estimación inconveniente de la magnitud del error registrado.

Alternativamente, se ha adoptado M3 como método para el cálculo del error en las encuestas preelectorales, asignando los indefinidos según las proporciones para los definidos, evitando un incremento artificial del error en los estudios en que no se realiza el ejercicio de asignación y permitiendo un tratamiento homogéneo de los datos relativos a las diversas encuestas públicas. Empero, no suele incluirse el error correspondiente al remanente de “otros candidatos”, sino solamente de aquellos contendientes que son reportados por separado por las encuestadoras en cada elección.

La solución de definir cada ocasión el número de contendientes a considerar o asumir un umbral determinado para su inclusión en el cálculo, y luego promediar los errores observados, además de ser un mecanismo arbitrario, mantendría problemas de homogeneidad bien en la cantidad de contendientes considerados, bien en la proporción del voto incluida en el cálculo, bien en ambos aspectos.

No deja de ser paradójico y relevante que las razones que sustentan la decisión de Mosteller de privilegiar para el análisis métodos determinados para el cálculo de las desviaciones de las encuestas respecto a los resultados sean también los motivos para el empleo de métodos alternativos en situaciones distintas. El objetivo primordial de Mosteller, que a su parecer debiera guiar todo análisis, es la construcción de una base de datos homogénea. En el caso del sistema estadounidense, ello se logra cuando se consideran solamente dos componentes, excluyendo o colapsando en lo posible terceras opciones que regularmente no están presentes o cuyo peso electoral es absolutamente marginal.

En sistemas multipartidarios no hay razones de peso para privilegiar métodos de agrupamiento o reducción de componentes únicamente a los dos mayores. De hecho, ni el margen de victoria, medido respecto a los votos por los dos partidos mayores o respecto al total de votos emitidos, ni mucho menos el sólo voto por el ganador resultan procedimientos completos y correctos para calcular los niveles de desviación entre lo estimado y el resultado.

En muchos análisis para sistemas multipartidarios se ha tendido entonces a privilegiar el empleo de M3 como el estimador más adecuado y completo para determinar la desviación de las encuestas respecto a los resultados, aun y cuando suele acompañarse de M5, reconociendo que es la diferencia entre los dos primeros contendientes lo que suele privilegiarse por medios y público atento como dato relevante producto de una encuesta, aunque ello sea equívoco.

Y es equívoco al menos por dos razones: no es un estimador directamente producido por la encuesta, sino el resultado de un cotejo de dos estimadores primarios (la preferencia por cada uno de los dos partidos mayores) y, por ende, sujeto a un doble error, dada la desviación esperada para cada componente que se compara; y es parcial, al reducir el fenómeno de concordancia entre encuestas y resultado a la distancia medida entre dos componentes, excluyendo el peso observado en otros componentes, que si bien menores, pudieron haber sido correcta o incorrectamente medidos al margen de la corrección del margen de victoria estimado.

M4. La media de las desviaciones de la razón entre la proporción prevista y la real para todos los candidatos, lo que se expresaría como (53):

$$M4 = \left\{ \begin{array}{l} \frac{\sum_{i=1}^m |(\hat{r}_i/p_i) - 1|}{m} \\ \frac{\sum_{i=1}^m |(\hat{p}_i/p_i) - 1|}{m} \end{array} \right., \quad (53)$$

Con mayor complejidad para su cálculo, el cuarto método presenta el problema inverso a M3: tiende a darle un peso muy elevado a las desviaciones en componentes menores, dado que pequeñas diferencias tienen un alto peso al compararse con la proporción reducida de voto logrado por dichos componentes.

M6. La diferencia máxima observada entre lo previsto y lo real para cualquiera de los contendientes, que pudiera expresarse como (54):

$$M6 = \left\{ \begin{array}{l} \max |\hat{r}_i - p_i| \\ \max |\hat{p}_i - p_i| \end{array} \right. \quad (54)$$

El sexto método presenta problemas técnicos y puede propiciar confusión como estimador del rendimiento de los estudios, según advierte el propio

Mosteller. Adicionalmente, podemos señalar que presenta serias limitaciones, no tanto en la comparabilidad entre estudios individuales, pero sí para toda pretensión de tratamiento agregado, pues no resulta adecuado para un cálculo de medias y el cálculo de máximos comunes ubica el error conjunto en un punto arbitrario.

M7. Correspondiente a la prueba estadística chi-cuadrado (χ^2), útil para evaluar la congruencia entre la distribución estimada y la oficial de los votos, cuya expresión sería (55):

$$M7 = \sum_{i=1}^m \frac{|\hat{p}_i - p_i|^2}{p_i} \quad (55)$$

El caso de este séptimo método, además de ser desestimado por Mosteller dada su complejidad y poca claridad para el público, enfrenta serias limitaciones para el tratamiento agregado de estimaciones diversas. Parte del problema se resuelve recurriendo a cálculos a partir de las frecuencias de casos y no de proporciones, pero esto hace más compleja su estimación. Por ello, la complejidad y problemas técnicos propios de un método como la prueba chi cuadrado o de opciones mejoradas de la misma, hace poco adecuado considerarlo como la opción pertinente.

CONCLUSIÓN PROVISIONAL

La revisión efectuada lleva a una conclusión obvia: no existe actualmente un estimador de la exactitud de las encuestas exento de problemas que limiten su uso y el entendimiento del fenómeno bajo observación. En una posterior colaboración este autor hará la exposición de un estimador alternativo que busque responder a este reto y resolver los diversos problemas detectados. La extensión descriptiva del problema alcanza una extensión suficiente para ser un artículo en sí y la exposición de la propuesta alterna otro, dadas las normas vigentes en las publicaciones científicas, lo que sería tema de una reflexión distinta y fuera de nuestro objetivo.