

Los márgenes de error en las encuestas electorales

El hecho de que el último año y medio – 2015 y la primera mitad del 2016-haya sido tan fecundo en contiendas electorales –andaluzas, catalanas, municipales, autonómicas y las dos elecciones generales- ha traído como efecto colateral inevitable el que la población española haya estado sometida durante este período a un bombardeo incesante de sondeos electorales.

No entraremos en el apasionante y controvertido tema de la mayor o menor capacidad predictiva de las encuestas preelectorales. Pero cabe constatar que además de haber demostrado históricamente ser una herramienta útil en términos generales –sus luces son muy superiores a sus sombras–, no hay un método alternativo mejor para hacer unos pronósticos que nuestra sociedad valora y requiere.

Con el loable objetivo de propiciar la transparencia metodológica de las encuestas, su publicación viene usualmente acompañada de una somera descripción metodológica de las mismas, lo que en España se conoce como ficha técnica, donde entre otros aspectos (instituto, fechas de campo, sistema de muestreo, tamaño de muestra, tipo de entrevista, etc.) se consigna la magnitud del llamado margen de error del estudio. Con este indicador se trata de trasladar al lector una valoración de la precisión esperable del sondeo.

Contenido obligatorio/recomendado de las fichas técnicas

En España y otros países de nuestro entorno se establecen por ley los requisitos a cumplir en la publicación de las encuestas electorales. En particular, en España la Ley Orgánica 5/1985 del Régimen Electoral General establece que junto a los resultados de las encuestas publicadas se deben especificar las *características técnicas del sondeo, que incluyan necesariamente los siguientes extremos: sistema de muestreo, tamaño de la muestra, margen de error de la misma, nivel de representatividad, procedimiento de selección de los encuestados y fecha de realización del trabajo de campo* así como *texto íntegro de las cuestiones planteadas y número de personas que no han contestado a cada una de ellas*. Llama la atención que el legislador hable del margen de error de una muestra –concepto inexistente que será objeto de una consideración posterior–, del nivel de representatividad –concepto vago y del que se parece solicitar una espinosa cuantificación– y que terminen por pedir, en una poco entendible pirueta final, una especie de tasa de respuesta para cada una de las preguntas. Hay que decir

que nuestras fichas técnicas habituales ignoran olímpicamente estas dos últimas especificaciones.

Quizá son más conocidas y fundadas las recomendaciones de las asociaciones profesionales en el sector de la investigación social y de mercado. Se ha pasado sucinta revista a las correspondientes a dos asociaciones americanas, una británica y las emitidas por la conjunción de dos asociaciones de carácter global:

- AAPOR (American Association for Public Opinion Research)
- National Council on Public Polls.
- British Polling Council.
- ESOMAR / WAPOR. World Association for Social, Opinion and Market Research y World Association for Public Opinion Research.

Sus recomendaciones sobre la información que debe acompañar a la publicación de los resultados de los estudios por encuesta presentan características muy similares. En general, y homogeneizando el lenguaje, se pide que las fichas técnicas contengan:

- Institución responsable de la realización del estudio e institución para la que se ha realizado el mismo (es decir, la que ha pagado por él).
- Universo al que se refieren los resultados.
- Marco de muestreo (fuente de selección de los entrevistados) y su cobertura.
- Tamaño de muestra, con diferentes desgloses si ello es relevante
- Método de muestreo.
- Metodología de recogida de la información (telefónica, personal, Web, correo, etc.).
- Procedimientos de ponderación / equilibraje para el cálculo de los pesos de las entrevistas individuales.
- Redacción / verbalización de las preguntas en el cuestionario.
- Fechas del trabajo de campo.

Se puede observar que la publicación de algún indicador del margen de error no aparece como requisito. AAPOR y NCPP sí lo mencionan y solicitan pero solo en el caso de muestras probabilísticas, lo que no es el caso de la generalidad de los sondeos electorales en nuestro país, que declaran utilizar cuotas (procedimiento no probabilístico) en alguna de las fases del muestreo. Específicamente, AAPOR dice que *para evitar confusiones es mejor evitar el término “margen de error” en los casos de muestras no probabilísticas.*

En España, AEDEMO –especialmente a través de su Comité de Ética– apoya y difunde las normas de ESOMAR y promueve su cumplimiento entre los profesionales e instituciones del sector de la investigación social¹.

Los sondeos electorales publicados en nuestro paísⁱⁱ siguen en general estas líneas de actuación. En algunos puntos, de forma quizás excesivamente esquemática y concisa, explicable probablemente por el reducido espacio disponible. Pero llama la atención la ausencia de información sobre dos puntos significativos:

- El marco de muestreo utilizado. Esta ausencia es especialmente relevante en los estudios telefónicos, que son la gran mayoría, dado que la utilización exclusiva de los directorios telefónicos disponibles deja fuera de la selección muestral a ese aproximadamente 25% de la población que son usuarios exclusivos de teléfono móvil. Se echa en falta alguna consideración metodológica al respecto y es sorprendente que cuando se produce la combinación de teléfonos fijos y móviles para eliminar ese problema de cobertura (combinación que se encuentra en al menos un caso) no se ponga en valor de forma fehaciente.
- Procesos de ponderación / equilibraje aplicados. Procedimientos que se utilizan de forma regular para ajustar la muestra a las proporciones del universo. Y, en el caso específico de los sondeos electorales, el que parece ser usual equilibraje por el recuerdo del voto.

Sobre las peculiaridades y diferencias acerca de la forma de declarar el margen de error estadístico en España volveremos más adelante.

Hablemos del margen de error

La palabra *error* en este contexto es un concepto estadístico consustancial a la incertidumbre derivada del proceso de muestreo aunque sugiera, al lector no profesional, falsas ideas de “equivocación”.

Quizá los profesionales de la investigación por muestreo deberíamos entonar un *mea culpa* por no habernos esforzado lo suficiente –no haber sabido hacerlo– en trasladar a los lectores y usuarios el mensaje fundamental de que no existe tal cosa como “el margen de error de un estudio”. La idea de que con un único indicador se pueda evaluar de forma global la precisión de un estudio no deja de ser atractiva pero, como en otras muchas cosas, la realidad es más compleja y no responde a nuestros deseos simplificadoros. Ciertamente, los resultados en un estudio por muestreo están inevitablemente afectados por algún grado de imprecisión estadística pero un estudio produce muchas estimaciones y cada una de ellas tiene su margen de error específico. Por ello, es lícito preguntarse por el margen de error de una de las estimaciones concretas pero no buscar el inexistente e imposible indicador de precisión global. En relación a un estudio globalmente considerado y por un mínimo sentido del rigor, solo cabe pluralizar el término y hablar de márgenes de error.

Para dar una solución al problema expuesto dentro del limitado espacio que tradicionalmente se permite ocupar a la *ficha técnica*, se ha optado históricamente por informar del margen de error exclusivamente para un caso hipotético muy específico: el de la proporción del 50%. No es que el 50% sea un caso muy representativo en los sondeos electorales (con proporciones estimadas de un orden de magnitud generalmente menor), pero tiene la característica de ser la proporción con el mayor margen de error absoluto para un tamaño de muestra dado dentro de un sistema de muestreo aleatorio simple. Y por tanto, dentro del marco expresado, puede tener una cierta connotación de “error máximo”. Es necesario señalar que si se toman errores relativos –a cuyo uso se tiende por ser de comprensión más fácil e intuitiva y de mayor utilidad práctica– la proporción del 50% ya no tiene esta cualidad de error máximo, sino que el error va aumentando cuando las estimaciones de las proporciones van haciéndose menores.

Cualquier margen de error está indisolublemente unido a un determinado “nivel de confianza”, es decir, el nivel de confianza elegido determina el cálculo del margen de error. En otras palabras, y para no extenderme en la explicación precisa de la ligazón de estos dos conceptos estadísticos asociada a la construcción de intervalos de confianza, señalaré de forma gráfica que las siguientes expresiones de precisión: “margen de error de $\pm 2,6$ puntos para un nivel de confianza del 80%”, “margen de error de $\pm 3,9$ puntos para un nivel de confianza del 95%” y “margen de error de $\pm 5,2$ puntos para un nivel de confianza del 99%” son totalmente equivalentes. En la misma línea, y aún a riesgo de pecar de redundante, cualquier estimación de un estudio puede tener un margen de error de $\pm 1,0$ puntos o de cualquier valor que el investigador decida: basta seleccionar el nivel de confianza apropiado.

Diferencias observadas en la declaración referente a los errores estadísticos

- Nombre del indicador. Extraña e infelizmente se sustituye en más de la mitad de los casos el término adecuado y universalmente aceptado de “margen de error” (*margin of error*, *marge d ’erreur*) por otros tales como “error de muestreo”, “error muestral”, “error de la muestra” o “error estadístico”. El término “error de muestreo” (*sampling error*) se utiliza en los textos de estadística como concepto general para expresar la variabilidad y falta de precisión en las estimaciones como consecuencia de la utilización de muestras y como contraposición a los llamados errores ajenos al muestreo. En ese sentido se puede decir que el margen de error cuantifica el error de muestreo. En los no demasiado frecuentes casos en que el “error de muestreo” se utiliza como indicador de precisión se hace como sinónimo de “error estándar”, término técnico más preciso, utilizado para denotar la raíz cuadrada de la varianza en el muestreo. Tratado como margen de error correspondería

necesariamente al nivel de confianza del 68%. ¿Por qué introducir una incorrección terminológica en una ya de por sí farragosa redacción?ⁱⁱⁱ

- Nivel de confianza. Tampoco hay una total unanimidad aunque el nivel del 95,5% es claramente mayoritario. Al menos tres institutos utilizan el 95% y hay uno (más como anécdota que como diferencia) que habla del 95,45%. Hay algún otro que no menciona el margen de confianza al que corresponde el error. Posiblemente los redactores que suprimen la mención al margen de confianza, que seguro saben de lo que hablan, han pensado que el lector de todas formas no lo va a entender y para qué complicarle más la vida; o que ya todo el mundo sabe que el nivel de confianza es del 95,5% y que no tendría sentido gastar el poco espacio del que se dispone para decir algo obvio. En relación a las diferencias entre los niveles seleccionados cabe decir que el 95% es el de uso más habitual en estadística y parece ser también el generalmente utilizado en los sondeos de opinión realizados en otros países. Pero con el 95,5% de uso mayoritario en España la fórmula del cálculo del error queda muy simple, redonda y de recuerdo fácil (volveremos sobre ello). En una elección que es esencialmente opinable sería recomendable el uso del más universal estándar del 95% que además permite el uso de la expresión quizá más gráfica de “19 de cada 20 casos”. Hay casos donde se “aclara” lo que significa el nivel de confianza del 95,5% añadiendo alegremente “dos sigma” o “para $k=2$ ”, manifestando no se sabe si una optimista confianza en el nivel de conocimiento estadístico del lector o la crueldad de complicar un poco más la ya difícil comprensión de la ficha técnica.
- Con redacciones un poco variables, la práctica totalidad de ejemplos especifican la estimación de proporción a la que se refiere el cálculo del margen de error, aunque con expresiones algo diferentes. Así, encontramos “ $p=q=50$ ”, “ $p/q=50/50$ ”, “ $P=Q$ ”, “ $p=q=0,5$ ” e incluso “ $p=q=50/50$ ” (!!!), entre otras expresiones de carácter esencialmente idéntico. La comprensión de su significado para un lector medio parece hartamente improbable. A veces la especificación del caso se acompaña con expresiones de “máxima indeterminación”, “supuesto más desfavorable” o similares. A este respecto resulta curioso el caso alemán, donde hay institutos que dan dos márgenes de error: para el caso 50% –interpretado como el potencial resultado de un partido político grande– y también para la proporción del 10% o del 15% –para un partido pequeño–, lo que supone un pequeño pero estimable esfuerzo aclaratorio.^{iv}

- La muy debida aclaración de que el cálculo del error se hace “bajo los supuestos del muestreo aleatorio simple” aparece en un grupo reducido de casos. Parece necesaria y ética la aclaración porque el margen de error no se puede calcular cuando el muestreo es no probabilístico donde seguramente se engloban la totalidad de ejemplos, ya que la selección por cuotas está siempre presente (aunque este hecho no siempre se declare explícitamente) en algunas de las fases del muestreo. Pero la advertencia resulta realmente sorprendente y contradictoria en los no pocos casos donde antes se ha descrito el procedimiento de muestreo como estratificado con afijación no proporcional.
- En el caso más general, el algoritmo con el que se calcula el margen de error es $\frac{100}{\sqrt{n}}$, que implica la selección de un nivel de confianza del 95,5%. En los pocos casos en que se utiliza el 95%, la fórmula se modifica ligeramente a $\frac{98}{\sqrt{n}}$. Vemos que el margen de error es solamente función del tamaño de muestra “n”. En ningún caso parece que se introduzca ningún ajuste por el efecto reductor de la precisión derivado de los habituales procesos de ponderación/equilibrio. Y es verdaderamente llamativo que en los casos de repartos de muestra no proporcionales por provincia la fórmula de cálculo siga muchas veces siendo la misma infraestimando el margen de error al obviar que la desproporcionalidad por estratos declarada en el diseño pediría cambiar el algoritmo de cálculo a $\frac{100 \sqrt{\sum \frac{N_h^2}{n_h}}}{\sum N_h}$ (sobre el supuesto extendido de $p_h = q_h = 50$ en todas y cada uno de las provincias). Es justo señalar que en algún caso no se comete esta evidente incorrección. Uno de los sondeos incluso sigue el modelo tan anglosajón de calcular y declarar la eficiencia muestral y la muestra efectiva. Publicar la muestra efectiva (aquel tamaño muestral que tendría análogo margen de error bajo el supuesto de muestreo aleatorio simple) es, indudablemente, una forma elegante y comprensible (comprensible dentro de un orden) de trasladar al lector las consecuencias de la desproporcionalidad de la muestra en la precisión de las estimaciones. Fuera del Reino Unido, el recurso de acudir al concepto de muestra efectiva es muy poco frecuente.
- La estricta ligazón o subordinación que hay entre margen de error y nivel de confianza no siempre se expresa con la necesaria corrección conceptual. Y no falta, por desgracia, el caso extremo donde las dos magnitudes se presentan de forma totalmente independiente, algo que no es de recibo porque el nivel de confianza “per se” no significa absolutamente nada.

- Uno de las interpretaciones incorrectas más habituales que se hace del margen de error que figura en la ficha técnica es tomarlo como error relativo por parte de periodistas, políticos, tertulianos, etc. Cuando el margen de error declarado es del 3%, interpretan que ese 3% se puede aplicar, por ejemplo, al número de diputados del PP, de Ciudadanos y del PNV. En un sondeo que pronosticase 119 escaños para el PP, 40 para Ciudadanos y 6 para el PNV, las respectivas horquillas serían 115-123, 39-41 y 6-6. Esta falsa interpretación se favorece al escribir “±3%” y se suaviza ligeramente plasmando “±3 puntos”. Aunque la inmensa mayoría de las fichas utilizan el símbolo del porcentaje, hay un instituto que, muy acertadamente, se inclina por utilizar “puntos” en línea con lo que se hace por parte de muchos institutos de otros países, por ejemplo el Reino Unido, con una alta tradición demoscópica.

Un poco a modo de resumen voy a crear una declaración tipo que sea razonablemente precisa y, *grosso modo*, representativa de las prácticas habituales en nuestro país. No corresponde a ningún instituto específico sino a un *collage* artificial a partir de las fichas reales analizadas.

Partiendo de los criterios del muestreo aleatorio simple, para un nivel de confianza del 95,5% (dos sigmas) y para el caso de máxima indeterminación (p=q=50), el margen de error estadístico para los datos globales es (despreciando el factor de corrección para poblaciones finitas) de ±3,2 puntos.

Con este lenguaje oscuro y pretendidamente científico, tal redacción podrá ser valorada como mínimamente rigurosa pero, a todas luces, es ininteligible para el lector medio^v, que tiende a retener simplemente el dato del ±3,2 sin una clara idea de su significado y ámbito de aplicación. Aunque esto pretende hacerse en aras de la transparencia metodológica, es muy dudoso que un párrafo de este tipo contribuya a la transparencia que hipotéticamente se persigue.

Hacia una salida razonable

La cultura actual imperante en este terreno procede de un pasado (muy reciente) dominada por los medios impresos, siempre con limitaciones de espacio, para los que el interés principal está en los resultados de la encuesta y no en las consideraciones metodológicas, tributo legal necesario pero carente de interés salvo para un reducido grupo de lectores. Tributo que intentan dejar de pagar: no se atreven todavía a eliminar la ficha técnica de la edición impresa pero en casi todos los casos dejan de consignarla en la edición electrónica (en aquellos medios que distribuyen ambos tipos de ediciones).

No parece descabellado pensar que pudieran convivir de forma simultánea dos niveles de información de amplitud y profundidad diferente.

- Una *ficha técnica resumen* con un contenido muy corto: instituto, tamaño de muestra, metodología de entrevista y fechas del trabajo de campo (una o dos líneas de texto exclusivamente) y donde adicional y necesariamente se consigne un enlace del tipo <http://www.instituto.es/encuesta/metodología>. Este texto corto debería cubrir las necesidades mínimas de información metodológica y sería la parte que necesariamente se debe publicar conjuntamente con los resultados de la encuesta. Y no debería contener ningún tipo de información sintética sobre los márgenes de error. La indicación del tamaño de muestra es, en una primera aproximación, una expresión suficiente de la posible precisión estadística del sondeo.
- En la dirección a la que remite el enlace, el instituto puede dirigirse a través de una amplia *“Descripción metodológica”* al colectivo de lectores más interesados por los aspectos técnicos de la encuesta, que también suelen tener mayores conocimientos estadísticos. Ya sin la servidumbre de la limitación de espacio, los institutos pueden volcar en este documento electrónico toda su voluntad de transparencia y toda su capacidad pedagógica, ampliando la información técnica a aspectos que hoy no se cubren. La creación de este tipo de documento electrónico accesible por Internet es muy habitual en estos momentos para los sondeos electorales realizados en el Reino Unido y no es desconocido su uso en España. Se puede ampliar con la redacción de las preguntas más relevantes del cuestionario y con tablas de resultados secundarias que se pongan a disposición del público general.
- En relación a los márgenes de error esperables se puede incluir en la página web una pequeña aplicación que calcule la magnitud de los errores estadísticos –tanto absolutos como relativos– para cualquier tipo de proporción y segmentación geográfica, superando la insostenible autolimitación actual del “50% para los datos globales”. Otra alternativa sería la inclusión de tablas que cubrieran el mismo objetivo. Sería deseable que las fórmulas de cálculo añadiesen los efectos de la variabilidad de los coeficientes de elevación derivados de la desproporcionalidad geográfica de la muestra y/o de otros ajustes producidos por procesos de equilibraje. Y se debería aclarar que, en todo caso, estos márgenes de error son solamente válidos, en sentido estricto, para la intención directa de voto, porque no cubren los riesgos de sesgos en la asignación de voto para los indecisos, ni las posibles imperfecciones en la asignación de escaños, ni las elaboraciones posteriores que se engloban bajo el término coloquial de “cocina”. Por último, creo que no estaría de más que se declarara la existencia de otras fuentes de incertidumbre posibles (sesgos muestrales de los que los sistemas de cuotas no están libres y que tienen la

maligna propiedad de que no se reducen al aumentar el tamaño de la muestra, bajas tasas de respuesta, sesgos declarativos, la “cocina”, etc.) que no están reflejadas en la optimista valoración resultante de considerar exclusivamente los errores estadísticos.

Carlos Lamas Alonso
Ex Director Adjunto de AIMC
Septiembre 2016

Notas

ⁱ José Vicente del Barco. El Comité de ética responde: La publicación de los resultados de las encuestas. Revista Investigación y Marketing número 67. Junio 2000.

ⁱⁱ Se han analizado buena parte de las fichas técnicas de los sondeos efectuados en el 2015 y la primera mitad de 2016. La lista de institutos y organismos con alguna ficha técnica analizada es lo suficientemente amplia para poder dar una visión general: GAD3, Celeste Tel, TNS Demoscopia, MyWord, NC Report, GESOP, Metroscopia, Sigma Dos, Instituto DYM, Opinometre, Sondaxe, Feedback, CIS, CEO y CAPDEA.

ⁱⁱⁱ Como referencia para ver el uso correcto del indicador “error de muestreo” (como sinónimo de “error estándar”) se puede ver, como ejemplo, http://www.eustat.eus/document/datos/calculo_errores_pra_c.pdf

^{iv} Ver, por ejemplo, <http://www.wahlrecht.de/lexikon/wahlumfragen.html>

^v Ver artículo del Defensor del Lector en El País del 4-1-1998 con el título “La verdad de las encuestas” en http://elpais.com/diario/1998/01/04/opinion/883868404_850215.html