

Reflexiones con relación al muestreo de las encuestas sobre la intención de voto 2011

Nave Herrera O.

Unidad de Biométrica, Instituto de Investigaciones Químicas y Biológicas
Facultad de Ciencias Químicas y Farmacia, Universidad de San Carlos de Guatemala

Introducción

Al revisar las encuestas publicadas en los medios de comunicación sobre la intención de voto de la población ante las elecciones generales de nuestro país, llama la atención una serie de elementos estadísticos que deben ser motivo de análisis y reflexión. Máxime cuando públicamente las encuestas son cuestionadas sobre su fiabilidad y que los medios de comunicación las defienden con argumentos poco claros para la mayoría, con la indicación que se realizaron cumpliendo los estándares estadísticos más rigurosos. Es por ello que en el presente artículo, se analizan algunos de los aspectos relacionados con las encuestas de opinión para medir la intención de voto que se publicaron previo a las Elecciones Generales de Guatemala realizadas el 11 de septiembre de 2011, para que el lector conozca los lineamientos estadísticos que deben regir en este tipo de investigaciones y se forme un criterio al momento de interpretarlas.

El estudio concluyente descriptivo

Inicio estas reflexiones refiriéndome al tipo de estudio que se indica en algunas de las fichas técnicas de las encuestas publicadas, en las que se menciona que se trata de investigaciones de tipo concluyente descriptivo. Este término "concluyente" es utilizado casi exclusivamente en investigación de mercados, definiendo a la investigación cuyo objetivo básico es proporcionar evidencia para contrastar hipótesis, evaluar y seleccionar alternativas de acción y establecer relaciones entre las variables de interés. Si se trata de un estudio descriptivo, su objetivo básico es cuantificar las variables de interés describiendo los fenómenos de mercado, midiendo la frecuencia con la que se presentan (Pedret, Sagnier & Camp, 2000; Malhorta, 2004). En términos estadísticos comunes, una investigación concluyente descriptiva, es en esencia una estimación inferencial, es decir, con los resultados se pretende predecir el comportamiento de las variables en la población a

partir de una muestra. La estadística inferencial o inductiva sirve para deducir, inferir, generalizar o "sacar conclusiones" de una población a partir de los datos de una muestra (Mendenhall, Beaver & Beaver, 2005). De allí la interpretación que se le debe dar al término concluyente que se aplica en estos casos.

¿Investigación social o investigación de mercado?

Las encuestas sobre la intención de voto, son estrictamente un asunto social o sociológico, sirviendo como un instrumento de investigación política para conocer las preferencias de la población. Este tipo de investigación ha sido objeto de desarrollo como investigación de mercados, debido a la invasión de marketing que ha sufrido el campo político y puede servir de referencia a los partidos en la toma de decisiones, elaboración de planes estratégicos y estrategias de campaña (Alonso & Adell, 2011). Es por eso que en muchos países del mundo, incluyendo Guatemala, la investigación de la preferencia electoral o intención de voto, ha sido llevada a cabo por compañías que realizan investigaciones de mercado, con la terminología y la metodología propias de este campo. Si la encuesta la patrocina un partido, definitivamente es dentro del campo del marketing y se considera un estudio de mercado; pero si es patrocinada por un medio de comunicación o una institución independiente, supuestamente debería ser tratada como una investigación social. Pero la investigación de mercado no se aleja de lo que estadísticamente, puede hacerse en investigaciones de otra naturaleza, los estudios de estimación inferenciales y la metodología para el cálculo del tamaño de la muestra y diseño muestral es (y debe ser) un lenguaje común. Sin embargo, algunos elementos no se aplican tan estrictamente como lo requiere una investigación estadística, social o económica, y es básicamente con relación al cálculo del tamaño de la muestra y al diseño muestral.

El cálculo del tamaño de la muestra

La ficha técnica de una investigación como las encuestas, además del tamaño de la muestra, comúnmente indica la confiabilidad y el margen de error. La confiabilidad, que generalmente es del 95%, representa la confianza con la cual los valores muestrales (estadísticos) estimarán a los verdaderos valores poblacionales (parámetros); esto implica necesariamente que en el análisis se deben calcular los intervalos de confianza para las principales variables del estudio; en este caso específico, se debería reportar el intervalo de confianza de la intención de voto para cada candidato, lo cual haría que la investigación en efecto fuera concluyente, es decir, una estimación inferencial. Es un error tratar de generalizar la intención de voto de toda la población con los porcentajes puntuales arrojados por la muestra, siendo ésta una posible explicación a las variaciones que se dan entre los resultados de encuestas realizadas en momentos diferentes o por diferentes agencias, dentro del mismo proceso electoral. Pero el cálculo de intervalos de confianza para cada candidato, representa un error estadístico si la muestra se ha calculado sobre una base simplista de una variable binomial, como estoy seguro que se hizo en estas encuestas, ya que en muchos de los textos de investigación de mercados se presenta esta forma de cálculo como la llave para casi cualquier investigación (se asume que “p” es la frecuencia de ocurrencia de la variable de interés y “q” su complemento y en la fórmula de muestreo se calcula la variabilidad suponiendo que $p=q=0.50$). Pero resulta que no son solo dos los candidatos, en el caso de las presentes elecciones, fueron diez y por lo tanto se trata de una variable multinomial y estadísticamente estaríamos obligados a calcular una muestra para estimar parámetros multinomiales, lo cual ya de por sí es complicado y pocos trabajos estadísticos se hacen de esa forma (Medina, 2002; Rudnykh, 2005). Por eso, asumamos que a pesar de ese error, el cálculo del tamaño de la muestra suponiendo solo dos posibles resultados podría ser aceptable. La premisa sería que a pesar que al encuestado le proporcionan una boleta con diez alternativas, la persona solamente tiene opción de elegir entre “su” candidato y los “otros” (de esta forma se dicotomiza lo que inicialmente está disperso en 10 opciones). En caso de una segunda vuelta electoral para la elección de Presidente y Vicepresidente de la República, sí correspondería realizar un planteamiento binomial.

Es curioso ver que casi todas las encuestas reportan un valor alrededor de $\pm 3\%$ como margen de error y números de muestra similares, lo cual se observa tanto en las encuestas actuales como en las publicadas en elecciones anteriores. La explicación es muy sencilla, si a la fórmula general de muestreo para una variable binomial, se le introducen los valores de confiabilidad (95%), la variabilidad antes indicada y un 2.8% de error, el número de muestra calculado es de 1,225 personas, para una población aproximada de siete millones de empadronados. A esto hay que agregar que la European Society for Opinion and Market Research (ESOMAR), recomienda el empleo de muestras de 1,000 a 2,000 casos para encuestas electorales en las que se desee representar globalmente a universos (poblaciones) mayores a 50,000 personas (ESOMAR, 2009) y se observa que esta recomendación se sigue fielmente, por lo que los números de muestra que indican las encuestas raras veces rebasan los 1,500 encuestados. Esta cantidad recomendada tiene su explicación en función del propio cálculo del tamaño de la muestra, si la población es mayor de 50,000 personas, el número de muestra tiende a “estabilizarse” y cambia muy poco aunque la población tienda a infinito.

El diseño muestral

Hasta aquí entonces, los valores de muestra presentados en las encuestas podrían ser tomados como correctos, pero hay un factor adicional que aún no se ha analizado: el diseño muestral. La fórmula general de muestreo sobre la que se basa el cálculo antes mencionado, que considera únicamente el nivel de confiabilidad, variabilidad y error, es útil para lo que se denomina un muestreo totalmente al azar o irrestricto aleatorio, pero las agencias que realizan las encuestas no usan este tipo de muestreo. Generalmente, se trabaja con muestreos polietápicos, estratificados por diferentes condiciones, básicamente para la medición de la intención de voto, se estratifica por departamento, municipio o aldea, con base supuestamente en el padrón electoral, que sería el marco muestral. Si a esto se le suma la división según área urbana y rural, así como cuotas por sexo (en algunas encuestas en países como España, fijan cuotas hasta para rangos de edad), ya podemos darnos una idea de lo complicado que resulta el diseño. Con relación a fijar cuotas de personas encuestadas no estoy de acuerdo ya que afecta la aleatoriedad, aunque es una práctica común en este tipo de estudios.

Si las encuestas de esta naturaleza se hacen por

medio de visita domiciliar, como se indica en las publicaciones de prensa, uno se pregunta entonces ¿Cuál es el marco muestral? ¿Cuáles son las unidades de muestreo, de análisis e informativas? Lo ideal, sería que el marco muestral fuera el padrón electoral y que las unidades de muestreo, de análisis e informativas sean iguales, es decir la persona seleccionada de manera totalmente aleatoria dentro del padrón será quien responda a la entrevista, pero en un muestreo polietápico, se dan diferencias entre estas unidades, por lo que el cálculo y selección de la muestra debería tomarlo en cuenta. Un dato más, en la ficha técnica de algunas encuestas se indica que son de aplicación individual ¿Qué quiere decir esto? Nuevamente surgen las dudas sobre las unidades de muestreo.

En una encuesta domiciliar, en el proceso de selección de la muestra (diseño muestral), debería considerarse el domicilio como unidad muestral y seleccionar dentro de cada casa aquellas personas que cumplan con el criterio de ser mayores de 18 años y estar empadronadas. Pero para llegar a cada domicilio, el proceso de estratificación o etapas de la encuesta debe ser muy riguroso y tener como marco la cantidad de domicilios en cada nivel de estratificación (incluso en lo que se refiere a la división entre urbano y rural), con todo un marco de referencia de los respectivos sectores cartográficos (Feres y Medina, 2001), tal y como el Instituto Nacional de Estadística (INE) realiza sus encuestas. En estos casos, las muestras estadísticamente representativas del INE no bajan de 10,000 viviendas (sobre un número total aproximado de dos millones de viviendas) y el margen de error estimado en estos casos oscila entre el 10 y 12%, lo cual es considerado aceptable al hacer proyecciones hacia la población (INE, 2002, 2006). Al comparar estos datos con el número de muestra entre 1,200 a 1,500 personas sobre una población de más de siete millones de empadronados y con un margen de error de 3%, salta a la vista que algo no debe estar bien: ¿Cómo es posible que con un número poblacional mayor, se reporte un error mucho menor con una muestra también muchísimo menor? Además, resulta muy curioso que todas las encuestas nacionales reportan aproximadamente el mismo número de muestra y margen de error, pero sobre todo que las encuestas a nivel local municipal tienen los mismos datos en su ficha técnica (aproximadamente 1,200 encuestados con un margen de error alrededor del 3%), esto confirma mi sospecha que el cálculo de muestra se hace indiscriminadamente sin tomar en cuenta tamaño ni estructura poblacional, menos pensando en un diseño

polietápico.

En una entrevista televisiva, un ejecutivo de una compañía que realizó una encuesta, indicó que una muestra pequeña puede reflejar adecuadamente a toda la población haciendo una analogía con una muestra de sangre que con muy poco volumen se puede conocer el estado de toda la sangre, pero lo que no mencionó y obviamente porque desvirtuaría su comentario, es que esto funciona muy bien para poblaciones homogéneas (como se puede suponer que es la sangre que circula en todo el organismo), no así en poblaciones heterogéneas, en las cuales se requiere más muestra, tal el caso de las encuestas que tratan de medir la intención del voto, donde al indicar que el muestreo es polietápico, se reconoce tácitamente la heterogeneidad de la población y por ello la necesidad de estratificarla en varios niveles y como corolario, la necesidad también de muestras más grandes.

Lo que debe tenerse en cuenta en el cálculo del tamaño de la muestra son los niveles de estratificación, el número de conglomerados o estratos que componen el marco muestral y su peso correspondiente, además se deben introducir en el cálculo dos elementos fundamentales adicionales: la tasa de no respuesta (para ajustar la muestra cuando las personas seleccionadas se niegan a ser entrevistadas) y el efecto de diseño. Este último es fundamental en un diseño polietápico ya que las estimaciones que se pretenden realizar dependen de lo que se denomina el Error Estándar, que es la variabilidad de los valores estimados entre posibles muestras de igual tamaño; en un diseño polietápico el error estándar es más complejo que el error estándar basado en el muestreo aleatorio simple y tiende a ser mayor, de allí la importancia de introducir una corrección en la fórmula de muestreo (el "efecto de diseño"), que como mínimo podría incrementar la muestra en un 50% más que para un muestreo aleatorio simple (INDEC, 2001).

Interpretación de los resultados

Hasta aquí en términos muy generales, se puede concluir que una muestra que pretende estimar un parámetro poblacional (intención de voto), sobre un padrón electoral que ronda los siete millones, debe ser mayor a la que se presenta en las encuestas que se publicaron en los diferentes medios de comunicación, sean éstas patrocinadas por los propios medios o los partidos políticos. Los resultados que ofrecen estas encuestas, podrían reflejar tendencias y ser indicadores de cuál o cuáles candidatos tendrían

más opción de ganar, pero en ningún momento los porcentajes reportados pueden ser tomados como la tendencia real del electorado, ni mucho menos que reflejen los posibles resultados de la elección, pero las publicaciones no lo indican así, por el contrario todas apostaron a que “su” encuesta sería la más certera en reflejar los resultados finales. Al tener los resultados finales de la elección, las reacciones fueron similares en los medios ¿Qué pasó con las encuestas? Los márgenes de error oscilaron entre el 6 y 12%. No faltaron publicaciones donde se “analizaron” los factores que incidieron en la falta de acierto de todas las encuestas de todos los medios de comunicación (cuatro medios escritos estuvieron patrocinando y publicando encuestas). Como dato curioso cabe resaltar que dos medios tenían encuestas con el mismo margen de error y confiabilidad, pero diferente número muestral, el error reportado en ambos casos fue de 2.8%, con menos muestra que otro medio que reportó un error mayor, contradiciendo el principio estadístico que a mayor muestra, menos error. Sin embargo, las explicaciones estadísticas como ésta y otras más profundas que deberían haberse hecho, brillaron por su ausencia en todos los casos, abundaron explicaciones de otro tipo y con el común denominador que se extrañaban que con los cálculos de muestra hechos no se esperaba una diferencia tan grande entre los porcentajes publicados antes de las elecciones y el resultado final de éstas. Se menciona por ejemplo, que la campaña fue irregular por no tener certeza de la participación de dos de los posibles candidatos, que probablemente las personas entrevistadas estaban en proceso de decisión, que no se tuvo acceso al padrón electoral ¿Cuál fue su marco muestral entonces?, que todas estaban técnicamente bien diseñadas pero que fallaron por razones de tipo social, al cambio de decisión de las personas, cambios en los ofrecimientos de campaña en los últimos días cuando ya no se podía realizar otra encuesta y hasta no faltaron los señalamientos al sistema electoral guatemalteco que permitiría la compra de votos y voluntades, uso de cédulas falsas y hasta culpan al Tribunal Supremo Electoral por errores en el cómputo de votos.

A manera de conclusión

Termino estas reflexiones indicando que fueron motivadas para tratar de aclarar desde el punto de vista estadístico, algunos aspectos relacionados con el cálculo del tamaño de la muestra y diseño muestral para un estudio como las encuestas para medir la

intención de voto, no pretendo desprestigiar ni invalidar los datos que se presentan, sino que el lector se forme un criterio a la luz de lo que la teoría estadística de muestreo nos ofrece y no dar por sentado lo que se publica simplemente porque se presenta una ficha técnica con terminologías a veces poco comprensibles para la mayoría. Soy de la opinión que una investigación puede llevarse a cabo con un número de muestra reducido o no calculado estadísticamente, incluso el diseño muestral puede ser no probabilístico, pero al publicarse los resultados esto debe ser lo primero que debe aclarar el investigador y tener presente que no le es permitido hacer inferencias o llegar a conclusiones sobre las poblaciones bajo esas circunstancias. Lo que no estoy de acuerdo es que se haga un estudio estadísticamente poco confiable y que se trate de enmascarar esta situación, debe haber congruencia entre la forma de calcular la muestra y el diseño muestral, lo que en estas líneas creo haber demostrado ampliamente que no fue así. Y también, como corolario, para que de alguna forma los que nos dedicamos a la docencia y asesoría estadística, no seamos cuestionados el día de mañana que lo que enseñamos no funciona o no es práctico, ya que por este tipo de tergiversaciones que se hacen públicas al difundirse por los medios de comunicación, se puede crear una falsa imagen de lo que en realidad es la estadística y perjudicar de alguna manera a los que nos dedicamos a ella.

Agradecimientos

El autor deja patente su agradecimiento al Doctor Tomás José Barricento Monzón, especialista en Muestreo Estadístico, con amplia experiencia docente y en encuestas nacionales por muestreo (USPADA 1994-95, INE 1998-2009), por sus opiniones y comentarios al presente artículo.

Referencias

- Alonso, M. A. & Adell A. (2011). Marketing político 2.0. Barcelona, España: Gestión 2000.
- ESOMAR. (2009). ESOMAR/WAPOR guide to opinion polls and published surveys. Disponible en: http://www.esomar.org/uploads/professional_standards_guidelines/WAPOR-ESOMAR_Guidelines.pdf. Fecha de acceso: octubre 01, 2011.

Feres, J. C. & Medina F. (2001). Hacia un sistema integrado de encuestas de hogares en los países de América Latina. Santiago de Chile: Naciones Unidas, CEPAL.

Instituto Nacional de Estadística (INE). (2002). Guatemala: Revisión del diseño muestral de distintas encuestas de hogares. Programa para el mejoramiento de las encuestas y la medición de las condiciones de vida en América Latina y El Caribe. CEPAL, BID, BIRF. Disponible en: <http://www.eclac.cl/deype/mecovi/docs/TALLER10/25.pdf>. Fecha de acceso: octubre 01, 2011.

Instituto Nacional de Estadística (INE). (2006). Encuesta nacional de condiciones de vida -ENCOVI 2006-, diseño de la muestra. Guatemala: INE.

Instituto Nacional de Estadística y Censos de la República Argentina (INDEC). (2001). Teoría y práctica del muestreo. 5o. curso regional del programa para el mejoramiento de las encuestas y la medición de las condiciones de vida en América Latina y El Caribe. CEPAL, BID, BIRF. Buenos Aires, Argentina.

Malhorta, NK. (2004). Investigación de mercados. 4a. Ed. México: Pearson Educación.

Medina, F. (2002). Determinación del tamaño de muestra en encuestas complejas de propósitos múltiples. Programa para el mejoramiento de las encuestas y la medición de las condiciones de vida en América Latina y El Caribe. CEPAL, BID, BIRF. Disponible en: <http://www.eclac.cl/deype/mecovi/docs/TALLER10/4.pdf>. Fecha de acceso: octubre 01, 2011.

Mendenhall, W., Beaver, R. J. & Beaver, B. M. (2005). Introducción a la probabilidad y estadística. 11a. Ed. México: Thomson.

Pedret, R., Sagnier, L & Camp, F. (2000). La investigación comercial como soporte del marketing. Bilbao, España: Ediciones Deusto, S. A.

Rudnykh, S. I. (2005). Tamaño de muestra para poblaciones multinomiales en muestreo bietápico. Dugandia, Ciencias Básicas, 1(1), 52-58.