

## Manual del candidato deseoso de encuestas favorables

Gonzalo Flores

Ya que muchos dirigentes políticos han decidido encargar encuestas para que unos supuestos especialistas les digan cuáles son los problemas del país, en qué deberían consistir sus propuestas políticas y qué lugar ocupan en la carrera hacia la Presidencia de la República, y que varias de las encuestas publicadas han mostrado sus enormes deficiencias, he decidido hacer una lista de los errores que no se debería cometer en la planificación, ejecución e interpretación de encuestas, a ver si así los políticos aguzan su razonamiento y no se dejan tomar el pelo por charlatanes de la estadística.

Uno de los errores más frecuentes es creer que todo se puede averiguar por encuesta. No es así. Por definición, la encuesta es la aplicación sistemática y pareja de un conjunto de preguntas a un grupo de individuos, sobre cosas que ellos conocen bien. Una encuesta de viviendas, por ejemplo, extrae información sobre la calidad y materiales de las viviendas, que el informante conoce y el encuestador puede verificar. Una encuesta de salud pregunta por prácticas habituales de salud y exposición a riesgos. Pero una encuesta es absolutamente inútil si pregunta por temas que el informante apenas conoce. Por ejemplo, cuáles serían las mejores medidas para asegurar la sanidad vegetal del país o para restablecer el abastecimiento regular de combustibles.

Generalmente, los clientes creen que los “especialistas” pueden medir todo y se equivocan completamente. Muy pocas cosas pueden ser medidas con números cardinales. Algunas sólo pueden ser “medidas” con números ordinales, y la mayoría, ni eso. Generalmente sólo podemos llegar a establecer jerarquías de preferencia, pero que no pueden ser tratadas con números cardinales. Las distribuciones estadísticas que se usa para variables aleatorias discretas no son las mismas que las que se usa para variables continuas.

Muy frecuentemente se olvida verificar la información. Algunas respuestas pueden ser verificadas por el encuestador, como en el ejemplo de las viviendas. Otras, difícilmente pueden ser verificadas. La misma pregunta, formulada de otra manera, puede llevar a otros resultados, no necesariamente opuestos, pero sí lo suficientemente distintos como para dudar si se ha estado preguntando por lo mismo.

Hacer una encuesta fuera de tiempo es otro error muy frecuente. No tiene gran sentido preguntar “¿cómo cree que irá la cosecha de granos este año?” antes de haber realizado las siembras. Del mismo modo, no se debe preguntar “¿por quién votará usted?” si no se sabe quiénes serán los candidatos. Y sería peor mezclar pre candidatos con candidatos y con personas que no pueden ser candidatos. Hay que esperar que las entidades estén bien definidas como para que la gente pueda anticipar sus opciones.

Mucha gente cree que una muestra sirve para todo. Vemos con frecuencia encuestas realizadas en unos pocos municipios que dicen representar a todo el “eje central” o a “todo el país”. Esa es simplemente una patraña. En realidad, mucho –casi todo- depende del universo al que se quiere extrapolar. Si se quiere extrapolar a sólo esos pocos municipios, eso es posible y correcto. Pero si a partir de ellos se quiere extrapolar al país, se incurre en un error fenomenal. Los tamaños de muestra (y su composición) se calculan de manera diferente dependiendo si lo que se quiere estudiar son medias o proporciones (o porcentajes). Por ejemplo, el tamaño de muestra para medias no está determinado tanto por el tamaño del universo, como por el valor hipotético de la media, un valor alternativo

específico de la media, el nivel de significación, la probabilidad de error tipo II (aceptar una hipótesis nula) y el valor de la desviación estándar de la población. Lo que muy pocos saben es que, si no hay estimaciones de los parámetros a investigar, los tamaños de muestra necesarios se hacen más grandes, y no se puede extrapolar del municipio al departamento y de éste al país, sino que cada ámbito requiere su propia muestra.

Se argumenta, en contra, que en Internet se puede encontrar calculadoras de muestras, y que éstas, para una población de siete millones de votantes, requieren sólo algo menos de cuatrocientos casos. Lo que Internet no informa es que esas calculadoras se hacen para el supuesto caso de una sola pregunta, cuyas respuestas son sólo del tipo SI/NO. En esos casos se puede aplicar la teoría de pequeñas muestras, y trabajar con muestras de sólo 40 casos. Aun más, sobre la base de la técnica del LQAS (Lot Quality Assurance Sampling), se puede trabajar con muestras de sólo diecinueve casos, suponiendo que se dispone de una buena estimación del parámetro principal y que la respuesta sólo puede ser SI o NO. Pero no es el caso de las encuestas electorales.

Generalmente, los comerciantes de encuestas no informan cómo se reclutó a los encuestadores, qué experiencia tenían, qué entrenamiento recibieron, cómo fueron supervisados; si se verificó la veracidad de las respuestas en un número razonable de cuestionarios o no; y si se verificó la consistencia entre el dato introducido al computador y el dato emitido por el informante. Abra los ojos, señor candidato.

Los comerciantes de encuestas esconden muy bien que mientras más preguntas se hace, más se confunde al informante y menor es la confiabilidad de las respuestas. Las respuestas que están al final de un largo cuestionario son mucho menos confiables que las que están al principio. También, prefieren no aplicar tests de consistencia a los datos recogidos. Algunos de ellos podrían mostrar que los resultados a algunas preguntas se distribuyen de manera diferente a las distribuciones estadísticas más conocidas. En ese caso la información tendría que ser revisada porque es anómala.

Por último, los comerciantes de encuestas prefieren no realizar la expansión de resultados al universo, cumpliendo las exigencias de la décima de hipótesis. Es decir, no hacen este ejercicio: Si 27% de los informantes en la muestra dicen que votarán por Juan ¿qué porcentaje de los votantes del país hará lo mismo y con qué confianza podemos afirmarlo? La respuesta no es 27% automáticamente. La respuesta caerá en un intervalo que dependerá de la hipótesis y la hipótesis nula, la confiabilidad buscada, el valor del estadístico elegido y de la distribución estadística que se adopte para realizar el test. No se obtendrá los mismos resultados aplicando la distribución normal que la distribución hipergeométrica o la de Poisson. La elección de la distribución no se puede hacer “al gusto del cliente”.

Obviamente, excluyo por completo esos remedos de “encuestas” que algunos realizan por radios, canales de televisión, Redes sociales y otros medios. No tienen seriedad alguna, no son confiables e inducen al engaño. Aun más: suelen estar diseñadas para engañar.

En fin, me parece que los políticos están siendo presas de supuestos concedores del espíritu profundo de la opinión pública. Tal cosa no existe. Sólo existen aproximaciones sucesivas, que serán menos inexactas cuanto más cerca nos encontremos de la fecha de las elecciones generales y cuanto más deseosos estén los informantes de hacer conocer su verdadera opción (si lo están, naturalmente).