

ACADEMIA DE CIENCIAS DE MORELOS, A.C.

¿Comentarios y sugerencias?, ¿Preguntas sobre temas científicos? CONTACTANOS: editorial@acmor.org.mx



Encuestas y elecciones: estadística, privacidad y verificabilidad

Candidato 1	<input type="radio"/>	Candidato 1	<input type="radio"/>	Candidato 1	<input type="radio"/>
Candidato 2	<input type="radio"/>	Candidato 2	<input type="radio"/>	Candidato 2	<input type="radio"/>
Candidato 3	<input type="radio"/>	Candidato 3	<input type="radio"/>	Candidato 3	<input type="radio"/>
Candidato 4	<input type="radio"/>	Candidato 4	<input type="radio"/>	Candidato 4	<input type="radio"/>

1. Sistema de tres boletas. Se muestran tres boletas vacías, con la lista completa de candidatos en cada una.

W. Luis Mochán
 Instituto de Ciencias Físicas,
 UNAM
 Academia de Ciencias de Morelos
 mochan@fis.unam.mx

“¿Es Ud. un violador? ¿Pederasta? ¿Drogadicto? ¿Trabaja de sicario? ¿Cobra derechos de piso? ¿Engaña a su pareja? ¿Es cliente de prostíbulos? ¿Guarda una fortuna en un paraíso fiscal? ¿Evade el pago de impuestos? ¿Vende a sus alumnos las calificaciones? ¿Saca acordeones en los exámenes? ¿Plagia sus artículos de investigación? ¿Vende sus opiniones en redes sociales? ¿Vende sus votos en las elecciones federales y estatales? Su respuesta a las preguntas anteriores será mantenida en una completa confidencialidad, no haremos pública su identidad y la información será empleada únicamente con fines estadísticos.”

En el muy hipotético caso de que alguna de las respuestas a las preguntas previas fuese afirmativa, ¿contestaría Ud. honestamente una encuesta planteada en los términos anteriores? Es muy probable que no, que la encuesta no le inspire demasiada confianza, que dude de la advertencia de confidencialidad y que tema que una pequeña indiscreción vuelva públicas sus faltas, destruya su reputación, arruine su vida y quizás hasta lo conduzca a la cárcel. Seguramente no contestaría la encuesta, o la contestaría pero no con toda veracidad.

¿Qué puede hacer entonces un investigador, probablemente un sociólogo, para obtener respuestas veraces que le permitan diagnosticar, cuantificar y hacer estadísticas sobre algunos de los aspectos negativos de la sociedad en que está inmerso? Una encuesta directa no funcionaría, pues no recibiría respuestas honestas. Tendrá entonces que recurrir a otras herramientas y técnicas alternativas a la encuesta, o quizás use otras preguntas para intentar obtener la información y su grado de veracidad sin que el encuestado lo sospeche. Plantear buenas encuestas es en ocasiones todo un arte.

Hay sin embargo una técnica poco conocida [ver por ejemplo la referencia 1] que permite plantear

directamente preguntas aparentemente comprometedoras y que sin embargo permite al encuestado contestar con toda honestidad sin riesgo a su porvenir. La idea es pedir al encuestado que tire una moneda al aire sin que el encuestador lo vea, que conteste afirmativamente si cae *águila*, y que sólo conteste la verdad si cae *sol*. Si bajo estas reglas le pregunto si usted es drogadicto y usted responde que sí, yo no podría saber si en verdad lo es o si simplemente el volado resultó en *águila*. De esta manera, su reputación quedaría totalmente a salvo, incluso si yo publico en todos los diarios y afirmo en los noticieros radiofónicos y televisivos que usted contestó que sí es drogadicto. Ante cualquier acusación, usted puede defenderse diciendo, “quizás no lo soy, pudo ser el volado, no les voy a decir y nunca lo sabrán.”

Quizás ahora se pregunte ¿para qué sirve una encuesta si la mitad de las respuestas están determinadas por el azar? El encuestador sabe de antemano que aproximadamente la mitad de las respuestas serán sí sólo por el resultado de los volados, sin tener relación con las respuestas reales. Por lo tanto, debe tirar a la basura esas encuestas y hacer su análisis estadístico sobre todas las demás. Pero ¿cuáles sí/s son los que deberá tirar y cuáles deberá conservar? ¡No importa! Una vez contados, todos los sí/s son iguales. No importa cuál se tire y cuál se conserve. Los resultados reales de la encuesta son los que arroje la mitad de los resultados que no sean desechados.

Quizás un ejemplo valga más que mil explicaciones. Imagine una encuesta en que preguntamos a diez mil jóvenes preparatorianos ($P=10,000$) si copiaron en su último examen y les explicamos el procedimiento descrito arriba. Suponga que el resultado es que seis mil contestan que sí ($Q=6,000$). De esas seis mil respuestas, aproximadamente 5,000 se deben a que el resultado de los correspondientes volados fue *águila*. Tiramos entonces a la basura $P/2=5,000$ de esas respuestas afirmativas. Nos quedan entonces $S=Q-P/2=1,000$ sí/s y $N=P-Q=4,000$ no/s auténticos. En este caso, podemos concluir que el 20% de los muchachos

(1,000/5,000) copiaron. Por lo tanto, de los 10,000 encuestados, alrededor de 2,000 copiaron. De ellos, unos 1,000 contestaron que sí pues la moneda cayó en *águila*, junto con otros 4,000 que no copiaron pero que también contestaron afirmativamente por el mismo motivo. Habría otros mil que contestaron que sí por haber copiado y responder honestamente la encuesta, pero nunca podremos saber quiénes de los seis mil que contestaron que sí fueron los que realmente copia-

ron. Entonces los alumnos no tendrían por qué temer el contestar honestamente. La probabilidad de que un joven que haya contestado que sí sea culpable de copiar es apenas de 1,000/6,000 16.7%, por lo que responder honestamente no lo incrimina. Se pierde muy poca privacidad por participar honestamente en una encuesta de este tipo, mientras no se aplique repetidamente. Un tema de investigación en estadística es el estudio de la pérdida de privacidad al publicarse información sobre encuestas [ref. 2].

El problema de este método de encuestar es que el encuestador debe ganarse la confianza del encuestado, este debe entender un poquito de estadística y entender por qué no corre riesgo al contestar con toda sinceridad. Quizás por ello no se han popularizado.

Candidato 1	<input checked="" type="radio"/>	Candidato 1	<input type="radio"/>	Candidato 1	<input type="radio"/>
Candidato 2	<input type="radio"/>	Candidato 2	<input checked="" type="radio"/>	Candidato 2	<input type="radio"/>
Candidato 3	<input checked="" type="radio"/>	Candidato 3	<input checked="" type="radio"/>	Candidato 3	<input type="radio"/>
Candidato 4	<input type="radio"/>	Candidato 4	<input type="radio"/>	Candidato 4	<input checked="" type="radio"/>
2866091845		8271916108		4540483265	

2. Se muestran las tres boletas como las podría haber llenado un simpatizante del candidato 3. Noten que viendo una sola de las boletas no podríamos saber qué candidato fue el favorito. Los números aleatorios añadidos al final de cada boleta permiten al votante identificarla.

ron. Entonces los alumnos no tendrían por qué temer el contestar honestamente. La probabilidad de que un joven que haya contestado que sí sea culpable de copiar es apenas de 1,000/6,000 16.7%, por lo que responder honestamente no lo incrimina. Se pierde muy poca privacidad por participar honestamente en una encuesta de este tipo, mientras no se aplique repetidamente. Un tema de investigación en estadística es el estudio de la pérdida de privacidad al publicarse información sobre encuestas [ref. 2].

El problema de este método de encuestar es que el encuestador debe ganarse la confianza del encuestado, este debe entender un poquito de estadística y entender por qué no corre riesgo al contestar con toda sinceridad. Quizás por ello no se han popularizado.

Elecciones

Una elección es una especie de encuesta en que se le pregunta a la ciudadanía cuál de una lista de candidatos es su preferido. La intención de una elección es determinar cuál es la preferencia de una sociedad a partir de las preferencias de los ciudadanos. Curiosamente, este problema está mal planteado desde un punto de vista matemático y no tiene una solución totalmente aceptable [ref. 3], como ilustra el llamado teorema de Arrow [ref. 4]. Aún aceptando las paradojas que conduce todo sistema electoral, como el que ha adoptado nuestro país, hay dos dificultades esenciales que es indispensable atender. El voto debe ser *libre*, es decir, los ciudadanos deben tener

la libertad de elegir de entre los candidatos registrados que cumplan todos los requisitos legales, al que sea de su preferencia, sin que terceros puedan obligarlo a votar por uno u otro. Para ello, es necesario que el voto sea *secreto*. De no serlo, un votante podría ser premiado por un tercero de votar como se le hubiese indicado o podría ser castigado por votar de forma distinta; podría haber compra o coacción de voto. Además, el voto debe ser *verificable*, el ciudadano debe poder verificar que

su voto fue contado tal y como fue emitido, sin ser reemplazado o desechado, y debe tener una certeza razonable de que no se añadieron a la cuenta votos que no fueron emitidos por otros ciudadanos.

Nuestro sistema ha tomado algunas medidas para cumplir parcialmente con los requisitos anteriores. Cada votante tiene asignada una casilla no demasiado lejana a su domicilio para tener acceso a votar. El número de votantes por casilla está acotado. Las casillas son operadas por ciudadanos capacitados bajo la supervisión de representantes de los partidos políticos y de observadores. Los votos se marcan en papeletas dentro de cancelas, lejos de las miradas de posibles curiosos, y se depositan doblados en urnas transparentes a la vista de todos. Se toman medidas como marcar la credencial y un dedo de cada votante para impedir que vote múltiples veces en la misma jornada. Cuando el votante se presenta en la casilla se sella el renglón correspondiente en una lista nominal con nombre y datos de todos los votantes asignados a la misma. Al final de la jornada, los funcionarios cuentan los votos en presencia de testigos, escriben la información en mantas públicas y se vacían en actas que se llevan a un centro distrital donde la información es capturada, enviada a un procesador central y publicada en una base de datos preliminar antes de ser depurada e incorporada a la base de datos final. Las urnas quedan resguardadas para poder ser vueltas a contar en caso de haber motivos para sospechar

errores o manipulaciones en la primera cuenta.

A pesar de las medidas anteriores, el grado de desconfianza en los resultados electorales y en las autoridades electorales es enorme, siendo parte de la enorme desconfianza que siente la ciudadanía hacia todas las instituciones [ref. 5]. La calificación que recibió el Instituto Nacional Electoral (INE) en 2016 es 6.2/10, un poco más que la de los bancos y la televisión, pero menos que los periódicos y los empresarios. No lo culparía si Ud. pensara que *la burra no era arisca pero la hicieron*. Sería entonces muy sano contar con un sistema que garantizara los resultados electorales, en que uno pueda confiar totalmente, a pesar de los deseos reales o imaginados de los contendientes de hacer trampa.

Alternativas

Exploremos entonces alternativas que pudieran ayudar a volver al sistema más confiable. Empecemos por un sistema en que cada uno de los votos quede registrado en una base de datos. Habría una página de internet en que cualquier votante pueda verificar que su voto registrado en la base de datos coin-

Candidato 1	<input checked="" type="radio"/>	Candidato 1	<input type="radio"/>
Candidato 2	<input type="radio"/>	Candidato 2	<input type="radio"/>
Candidato 3	<input checked="" type="radio"/>	Candidato 3	<input type="radio"/>
Candidato 4	<input type="radio"/>	Candidato 4	<input type="radio"/>
2866091845		82719	

3. Otra posible papeleta válida llena por un simpatizante del candidato 3.

cida con su voto emitido. Pensándolo bien, esta idea no es tan buena. ¿Confiaríamos en los programadores que desarrollen la base de datos y la página web? ¿Quién nos aseguraría que los datos que nos muestre la página sean los mismos que los que se contabilicen? Nos podrían hacer trampa programando dos bases de datos: una con los resultados reales, cuya información mostrarían a cada ciudadano que la interroga, y otra con datos falsos que se sumen para obtener un resultado manipulado.

Podríamos eliminar la posibilidad de una doble contabilidad si la base de datos fuera pública, si se publicara su formato, y si se adhieran a estándares que permitieran a cualquier ciudadano y a cualquier agrupación el bajarla, interrogarla y computar el resultado de la elección. Esto es lo que se hace actualmente con las bases del Programa de Resultados Electorales Preliminares (PREP) y

ACADEMIA DE CIENCIAS DE MORELOS, A.C.

¿Comentarios y sugerencias?, ¿Preguntas sobre temas científicos? CONTÁCTANOS: editorial@acmor.org.mx



con las bases del Coteo Distrital, aunque esta base sería alrededor de mil veces más grande. Así, sería imposible mantener un doble conteo. Aún publicando la base de datos íntegra, la propuesta anterior tiene un enorme defecto: sería violatoria de la secrecía del voto. Políticos, jefes policiacos, dueños de empresas, líderes sindicales y hasta familiares y amigos podrían averiguar cómo votaron algunos ciudadanos. Eso les permitiría dar regalos o dinero a quienes voten de cierta forma o castigar laboral, económica o físicamente a quien vote de otra forma. Este sistema fomentaría la compra y coacción de votos, costumbre aparentemente muy enraizada en nuestra cultura política. ¿Cómo diseñar entonces un sistema que permita verificar los votos individuales pero no permita su compra o coacción? Una idea es asignar a cada boleta un número y en la base de datos anotar el número y el voto correspondiente, pero no anotar el nombre del votante ni su número de cédula electoral. Desafortunadamente, si los números asignados a cada boleta fueran consecutivos esto

La discusión anterior muestra las dificultades inherentes en satisfacer simultáneamente dos requisitos que deberían cumplir las votaciones: la secrecía y la verificabilidad.

Tres boletas

Recientemente, en 2006, fue inventado por Ron Rivest un ingenioso protocolo de votaciones [ref. 6] basado en ideas como las presentadas arriba en la sección sobre encuestas, que permite proteger la secrecía del voto a la vez que posibilita su verificación por parte del elector. La idea es darle a cada elector *no una, sino tres boletas*. En cada una estarían inscritos los nombres de todos los candidatos; cada uno aparecería tres veces, una vez en cada boleta (figura 1). El elector podrá votar por cualquier cantidad de candidatos en cada boleta con la restricción de que entre las tres boletas *le otorgue exactamente un voto a cada candidato excepto a uno, a quien le otorgue exactamente dos votos*. Es más fácil explicar lo anterior con un ejemplo. En la figura 1 se muestra una papeleta con 3 boletas vacías, cada una con cuatro candidatos. En la figura 2 se muestra esta papeleta correctamente llenada. En este caso el candidato 1 obtuvo una marca en la boleta de la izquierda, el 3 obtuvo dos, una en la izquierda y otra en la derecha y el candidato 4 obtuvo una marca en la boleta central. En este caso, el candidato preferido fue el 3, pues obtuvo dos marcas mientras que los demás obtuvieron solo una. Note que en las boletas izquierda y central están marcados más de un candidato. En este sistema, esto es permitido. La figura 3 es otro juego de boletas válida donde gana de nuevo el candidato 3. Este juego muestra

marca en alguna de las tres boletas, esta papeleta sería inválida pues un candidato (el 3) tiene tres marcas. Una vez que el votante ha llenado de manera válida su papeleta, esta se corta en tres boletas que se depositan en la urna. Antes de esto, se le pide al votante elegir una de las tres boletas, la que él desea, se hace una copia de la misma y se le entrega para que la conserve. Para identificar cada papeleta, justo antes de meterlas a la urna y de generar la copia para el votante se generan e imprimen números al azar (compare la fig. 1 con las otras) que no estén correlacionados entre sí ni con los datos del votante. Cuando se abra la urna y se realice la cuenta de votos, los escrutadores no sabrán qué boleta iba junto con cuáles otras en la misma papeleta, por lo que no tendrían más remedio que sumar todas las marcas. Sin embargo, cada votante pondría una marca a cada candidato y una adicional a su candidato favorito. Entonces, si en una casilla hubieran votado *N* personas, cada candidato hubiera sido marcado *N* veces más el número real de votos. De manera que en el cómputo final de marcas bastará restar el número total de votantes para obtener el número real de votos recibidos por cada candidato.

Verificabilidad

¿Qué ganamos con este sistema tan complicado? Para empezar, al menos, no hemos perdido nada, pues es posible computar el número de votos que hubiera recibido cada candidato sumando marcas y restando votantes. La idea es que la información de todas las papeletas queden registradas en una base de datos pública, de forma que cualquier votante puede revisar, o encargar a un tercero

tres papeletas, sería $2/3 \approx 66.7\%$. Así mismo, la probabilidad de poder cambiar o eliminar dos votos sería $(2/3)^2 = 4/9 \approx 44\%$, la de poder cambiar tres sería $(2/3)^3 = 8/27 \approx 29\%$ y así sucesivamente. Note que esta secuencia de números se hace pequeña rápidamente, pues cada número es más pequeño que el anterior en un factor de $2/3$. Por ejemplo, la probabilidad de poder modificar 1000 votos sin ser detectado sería $(2/3)^{1000} \approx 6 \times 10^{-351}\%$, es decir, sería 6% dividido entre un número gigante con 351 ceros. Este número es tan ridículamente pequeño que podemos afirmar que sería imposible modificar los votos sin que la ciudadanía se enterara. Incluso, si sólo un porcentaje pequeño de votantes desconfiados verificara su voto, sería imposible hacer un cambio significativo en el resultado de la elección sin que fuera detectado. Tampoco se podrían añadir votos, pues el número de boletas debería coincidir con el número de votantes (multiplicado por tres). Pero si el votante se queda con una boleta, ¿no se violaría la secrecía? No. El votante se queda con una boleta, pero el sentido de su voto no se puede conocer sin tener a mano las otras dos boletas. El votante podría tratar de convencerlos de que votó por el candidato 1, para cobrar la plata o para evitar el plomo en caso de que hubiesen intentado comprar o coaccionar su voto, mostrándonos para ello la primera boleta donde hubiera marcado a dicho candidato solamente, pero no podremos saber si en otra de las dos boletas marcó al candidato 1 o si marcó dos veces a algún otro candidato en las dos boletas restantes.

asegurar que no guardará información que pueda comprometer la secrecía o modificar los votos que no serán copiados antes de depositarlos. El votante debe poder verificar que sus marcas hayan sido registradas correctamente al momento de votar, pero debe ser imposible fotografiarlas, para impedir otro mecanismo de compra y coacción del voto. De no impedir fotografiarlas, debe permitirse que el usuario cancele su papeleta y la llene nuevamente antes de que se deposite en la urna, para que la fotografía no sea una demostración de voto. Por cuestiones prácticas, los números aleatorios podrían acompañarse de códigos de barras para poder contar los votos automáticamente en un lector óptico.

Conclusiones

Mediante el uso de algunas técnicas estadísticas se pueden diseñar encuestas que permitan a los encuestados contestar honestamente sin comprometer su reputación, aunque contengan preguntas delicadas, y se pueden diseñar protocolos electorales verificables por cada uno de los votantes sin que se violen los requisitos de secrecía.

Agradecimientos

Este trabajo fue apoyado por DGAPA-UNAM mediante el proyecto IN113016. Su origen tuvo que ver con una reunión organizada por el Dr. Alejandro Frank en el Centro de Ciencias de la Complejidad para discutir la desconfianza y escepticismo electoral. Agradezco a Juan Claudio Toledo por presentarme el protocolo de tres boletas.

Referencias

Randomized response - Wikipedia <http://bit.ly/2zUJ87g>
 Differential privacy - Wikipedia <http://bit.ly/2B3uCu7>
 W. Luis Mochán, *Matemáticas Electorales*, Diario Unión de Morelos, 22 de junio de 2009, pgs. 34 y 35. <http://bit.ly/1iAtv8P>
 Arrow's impossibility theorem - Wikipedia, <http://bit.ly/2jDz6S5>
 México: Confianza en Instituciones 2016, <http://bit.ly/2zieohl>
 Ronald L. Rivest (2006). "The ThreeBallot Voting System", <http://bit.ly/2j7l5b7> <http://theory.csail.mit.edu/~rivest/Rivest-TheThreeBallotVotingSystem.pdf>

o 1	<input type="radio"/>	Candidato 1	<input type="radio"/>
o 2	<input checked="" type="radio"/>	Candidato 2	<input type="radio"/>
o 3	<input checked="" type="radio"/>	Candidato 3	<input type="radio"/>
o 4	<input checked="" type="radio"/>	Candidato 4	<input type="radio"/>
16108		4540483265	

no garantizaría el secreto. Un observador podría adivinar qué número le tocó a cada votante de acuerdo a su hora de llegada a la casilla, y posteriormente consultar cómo votó. Los números deberían entonces asignarse de manera azarosa, garantizando únicamente que no haya dos boletas con el mismo número. Tendría entonces que ser un número de muchos dígitos. Para poder verificar su voto el elector tendría que anotar este número. Para evitar errores, sería mejor que el votante conservara una copia de su boleta, lo cual sería además indispensable para que pudiera reclamar si su voto no fue registrado de forma correcta. La nueva dificultad estriba en que si el votante conserva una copia del número de su boleta o conserva su número, puede ser forzado a mostrarla a algún hampón quien al ver la boleta o al consultar la base de datos podría averiguar el sentido del voto, violándose el secreto electoral.

Candidato 1	<input checked="" type="radio"/>	Candidato 1	<input type="radio"/>	Candidato 1	<input checked="" type="radio"/>
Candidato 2	<input type="radio"/>	Candidato 2	<input checked="" type="radio"/>	Candidato 2	<input type="radio"/>
Candidato 3	<input checked="" type="radio"/>	Candidato 3	<input checked="" type="radio"/>	Candidato 3	<input type="radio"/>
Candidato 4	<input type="radio"/>	Candidato 4	<input checked="" type="radio"/>	Candidato 4	<input type="radio"/>
2866091845		8271916108		4540483265	

4. Papeleta inválida por tener más de un candidato con dos marcas.

que en este sistema también es válido dejar una boleta en blanco, siempre y cuando en las otras dos boletas todos los candidatos estén marcados una vez salvo uno de ellos (el 3) que quede marcado dos veces. La fig. 4 muestra un juego de boletas inválido, pues hay dos candidatos con dos marcas (el candidato 1 y el 3), la fig. 5 muestra un juego inválido pues ningún candidato tiene dos marcas y la fig. 6 muestra un juego inválido pues un candidato (el 1) no tiene ninguna marca. Aún si tuviera una

que revise, que la papeleta cuya copia conserva haya sido registrada correctamente. Para ello puede usar el número que identifica a la boleta. La base de datos se haría pública de inmediato sin que se sepa cuáles son las boletas de las que los votantes conservaron copias. Por lo tanto, si todos los votantes verificaran su voto, la probabilidad de poder cambiar o eliminar un voto sin que sea detectado por alguno de los votantes, que conserva copia de una de las

Candidato 1	<input checked="" type="radio"/>	Candidato 1	<input type="radio"/>	Candidato 1	<input type="radio"/>
Candidato 2	<input type="radio"/>	Candidato 2	<input checked="" type="radio"/>	Candidato 2	<input type="radio"/>
Candidato 3	<input checked="" type="radio"/>	Candidato 3	<input type="radio"/>	Candidato 3	<input type="radio"/>
Candidato 4	<input type="radio"/>	Candidato 4	<input checked="" type="radio"/>	Candidato 4	<input type="radio"/>
2866091845		8271916108		4540483265	

5. Boleta inválida por tener un candidato sin marca alguna. También es inválida por tener un candidato con tres marcas.