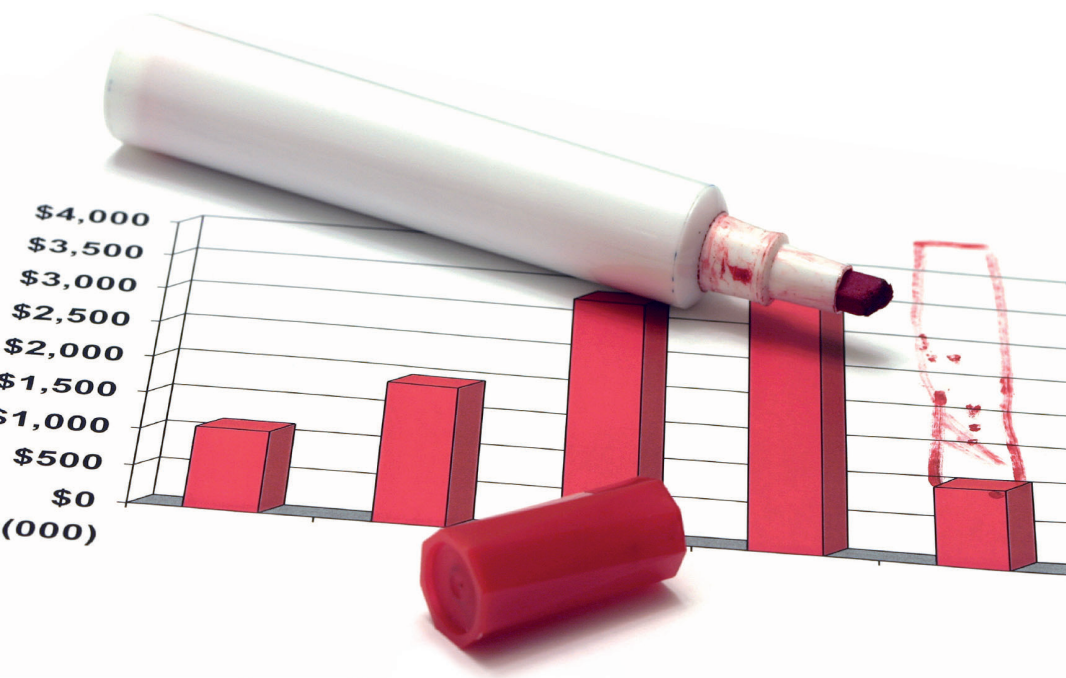


ARES y MARES

Cómo mentir con estadísticas



Darrell Huff

CRÍTICA

Darrell Huff

Cómo mentir con estadísticas

Traducción castellana
de Octavio Freixas

ARES Y MARES

ARES y MARES
es una marca editorial dirigida por
Carmen Esteban

Primera edición: junio de 2011

Primera edición en esta presentación: mayo de 2015

No se permite la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea éste electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito del editor. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (Art. 270 y siguientes del Código Penal)

Dirijase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra. Puede contactar con CEDRO a través de la web www.conlicencia.com o por teléfono en el 91 702 19 70 / 93 272 04 47

Título original: *How to Lie With Statistics*

Diseño de la cubierta: Jaime Fernández

Ilustración de cubierta: © iStock

Composición: gama, sl

© 1954, 1982 Darrell Huff

© 1954, 1982 de las ilustraciones: Irving Geis

© 1965 de la traducción: Octavio Freixas Ortega

© 2011 de la revisión: Editorial Crítica

© Editorial Planeta, S. A., 2015

Av. Diagonal 662-664, 08034 Barcelona (España)

Crítica es un sello editorial de Editorial Planeta, S. A.

editorial@ed-critica.es

www.ed-critica.es

ISBN: 978-84-9892-848-8

Depósito legal: B. 8287-2015

2015. Impreso y encuadernado en España por Book Print



Capítulo 1

LA MUESTRA QUE PRESENTA UN FACTOR DE INFLUENCIA EN SÍ MISMA

«El promedio de los componentes de la promoción 1924 de la Universidad de Yale —informaba la revista *Time*, refiriéndose a un artículo aparecido en el *Sun* de Nueva York— gana 25.111 dólares al año.»

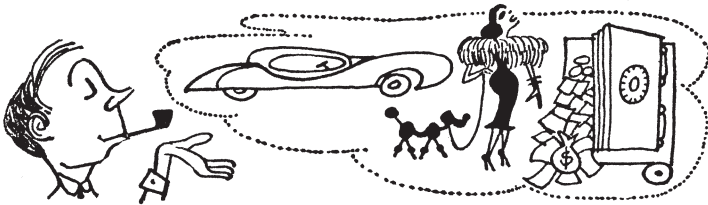
¡Magnífico!

Pero, espere un momento. ¿Qué significa esta cifra impresionante? ¿Según parece, es prueba de que si envía a su hijo a Yale no tendrá usted que trabajar en su vejez ni él tampoco?

Dos cosas referentes a esta cifra saltan a la vista a la primera ojeada recelosa. Es de una sorprendente precisión y tiene pocas probabilidades de ser exacta.

Es extremadamente difícil llegar a conocer hasta el último dólar de los ingresos medios de cualquier grupo que se formó hace tiempo. Tampoco es muy probable que usted recuerde con tal precisión sus propios ingresos del año pasado, a menos que éstos provengan enteramente de un único salario; pero pocas veces, los ingresos del orden de 25.000 dólares provienen totalmente de un salario; la gente que se encuentra en este nivel disfruta, probablemente, de inversiones bien distribuidas.

Además, este atrayente promedio está calculado, sin duda, partiendo de las cantidades que los graduados de Yale *dijeron* que ganaban. Incluso si en el año 1924 figuraban todos en el cuadro de honor por su integridad, no podemos estar seguros de que todos ellos continúen siendo honestos al cabo de un cuarto de siglo y de que estos informes sean por tanto fidedignos. Algunas personas, cuando se les pregunta acerca de sus ingresos, exageran por vanidad o por optimismo. Otros los minimizan, especialmente cuando se trata de las declaraciones a Hacienda, y después de hacerlo, temen contradecirse en cualquier otro documento. ¿Quién sabe hasta dónde pueden llegar los inspectores? Es posible que estas dos tendencias, la de sobreestimar y la de subestimar, se neutralicen, pero no es probable. Seguramente una tendencia será más fuerte que la otra, pero ¿cuál de las dos?



Empezamos a explicarnos una cifra de la cual nos dice el sentido común que difícilmente puede representar la verdad. Ahora pongamos el dedo en el probable origen del error más patente, origen que puede hacer figurar la cifra de 25.111 dólares como «ingresos medios» de algunos hombres cuya media real podría muy bien hallarse hacia la mitad de esta cantidad.

Se trata de la técnica del muestreo, que es la espina dorsal de la mayor parte de las estadísticas que usted haya de encontrar en toda clase de temas. La base de esta técnica es simple, aunque en la práctica sus sutilezas han conducido a toda clase de caminos equívocos, algunos de los cuales distan de ser respetables. Si se tiene un saco de alubias, rojas y blancas, existe sólo una manera de averiguar exactamente cuántas hay de cada color: ¡contarlas! Sin embargo, puede averiguarse con aproximación la cantidad de alubias rojas de una forma más fácil; se coge un puñado de alubias, se cuentan las que lo componen y se da por sentado que la proporción será la misma en todo el saco. Si la muestra es lo bastante extensa y bien seleccionada, representará al conjunto con bastante aproximación en la mayoría de los casos. Si no es así, puede ser menos precisa que una hipótesis sensata, y no tiene nada que la apoye, aparte de un aire de aparente precisión científica. Es una triste verdad el hecho de que detrás de lo que leemos o creemos saber están las conclusiones derivadas de muestras subjetivas o demasiado pequeñas, o ambas cosas a la vez.

El informe sobre los universitarios de Yale deriva de una muestra. Podemos estar seguros de ello, porque la razón nos dice que nadie puede ponerse en contacto con todos los miembros vivientes de la promoción de 1924. Al cabo de veinticinco años, debe de haber muchos cuyas direcciones se desconozcan. Y, entre los de direcciones conocidas, muchos no contestarán al cuestionario, sobre todo por tratarse de preguntas de carácter personal. Con alguno de los cuestionarios que se envían por correo, un porcentaje de respuestas de un 5 a un 10 por 100 suele considerarse suficientemente elevado. El que nos ocupa puede haber sufrido mejor suerte, pero nunca hasta alcanzar un 100 por 100 de respuestas.



Así pues, hallamos que la cifra de ingresos se basa en una muestra formada por componentes de la promoción cuyas direcciones se conocen y que contestaron al cuestionario. ¿Es representativa esta muestra? Es decir, ¿puede aceptarse que este grupo sea equivalente, en cuanto a la cifra de ingresos, al grupo no representado, formado por aquellos cuyo paradero se desconocía o que no contestaron?



¿Quiénes son los corderos extraviados cuyas «direcciones se desconocen»? ¿Son los grandes capitalistas —los hombres de Wall Street—, los directores de empresa, los directores comerciales y de fábricas? No; no es difícil dar con las direcciones de los acomodados. Muchos de ellos figuran en el *Who's Who in America*

y otros volúmenes de referencia, incluso en caso de que olviden mantener contacto con la asociación de ex alumnos.

Es probable suponer que los nombres extraviados son los de aquellos quienes, después de conseguir su licenciatura en Artes y Oficios, no alcanzaron el éxito que se prometían. Son los empleados, los mecánicos, los vagabundos, los alcohólicos sin empleo, los artistas y escritores que apenas logran subsistir... hombres de los que se necesitaría media docena o más para alcanzar la cifra de 25.111 dólares sumando sus ingresos. Estos hombres no asisten a las reuniones de su promoción, aunque sólo sea por el hecho de que no pueden afrontar los gastos de desplazamiento.



De entre los que fueron localizados, ¿quiénes tiraron el cuestionario a la papelera más cercana? No podemos estar muy seguros, pero una hipótesis con bastantes probabilidades de certeza es que se trata de los que no ganaban suficiente dinero como para presumir de sus ingresos. Son, en cierta manera, como aquel individuo que encontró una nota cogida con un clip al sobre de su primera paga, sugiriendo-

le que considerase el importe de su salario como asunto confidencial y no como motivo de confidencia con sus compañeros.

—No se preocupe —le dijo a su jefe—. Estoy tan avergonzado de mi sueldo como lo está usted.

Es evidente que la muestra ha omitido dos grupos que con toda probabilidad reducirían la media. La cifra de 25.111 dólares empieza a explicarse. Suponiendo que esa cifra sea representativa, lo que representa es sin duda el grupo especial de miembros de la promoción de 1924 cuyas direcciones se conocen y que están dispuestos a colaborar y publicar los ingresos de que disfrutaban. Incluso en tal caso, hay que partir del supuesto de que estos hombres dicen la verdad.

Formular un supuesto no es tarea fácil. La experiencia obtenida en los estudios de muestras que llamamos estudios de mercado nos demuestra que es casi imposible.

En cierta ocasión se llevó a cabo una encuesta casa por casa, con la finalidad de conocer el número de lectores de revistas; la pregunta clave de la encuesta era: ¿Qué revistas leen los miembros de su familia? Cuando se tabularon y analizaron las respuestas, resultó que a mucha gente le gustaba *Harper's* y muy poca gente leía *True Story*. Ahora bien, existían cifras de los editores demostrando claramente que mientras *True Story* editaba millones de ejemplares, *Harper's* sólo editaba unos cientos de miles. ¿Quizá nos dirigimos a un público inadecuado?, se preguntaron los planificadores de la encuesta. Pero no era así; el cuestionario se había pasado en todos los barrios de la ciudad. La única conclusión razonable fue que muchos de los entrevistados no habían dicho la verdad. La encuesta sirvió solamente para revelar el grado de esnobismo de los informantes y descubrió que si se desea saber lo que lee cierta clase de público, no es informativo preguntárselo a los interesados.

Podría obtenerse mejor información visitándoles y diciéndoles que se deseaba comprar todas las revistas viejas que tuvieran. Después, no quedaría sino comparar el número de revistas de tono universitario y los folletines. E incluso este sistema no informa de lo que lee la gente, sino sólo de aquello con que se halla en contacto.

Por igual motivo, cuando lea otra vez que el americano medio (se habla mucho de él, y la mayoría de las veces lo que se dice está poco fundamentado) se cepilla los dientes 1,02 veces al día —cifra que se me acaba de ocurrir, pero que puede ser tan buena como la facilitada por cualquier otra persona— pregúntese: ¿Cómo se ha podido averiguar tal cosa? ¿Acaso una mujer, después de leer en innumerables anuncios que quienes no se lavan los dientes ofenden a la sociedad, confesará a un desconocido que no suele hacerlo con regularidad? Estos resultados estadísticos pueden tener significado para quien quiera saber solamente la opinión de la gente acerca de la higiene dental, pero nos dicen bien poco sobre la frecuencia con que las cerdas se aplican a los incisivos.



Se dice que un río no puede sobrepasar el caudal de la fuente que lo origina. Sin embargo, podría conseguirse si existiera un equipo de bombas escondido en algún lugar. Del mismo modo también es verdad que el resultado del estudio de una muestra no puede ser mejor que la muestra estudiada. Cuando los datos han sido filtrados a través de distintas fases de manipulación estadística y reducidos a una media expresada en decimales, el resultado empieza a presentar una aureola de convicción que sólo se podría ver empañada por una revisión cuidadosa de la muestra.

¿El diagnóstico precoz del cáncer salva algunas vidas? Es probable que sí, pero si examináramos las cifras utilizadas para probarlo, podríamos deducir que no. Los informes del Connecticut Tumor Registry se remontan al año 1935 y parecen presentar un sustancial aumento de cinco años de supervivencia entre este año y 1941. La realidad es que estos datos se tomaron a partir de 1941, y todo lo anterior se obtuvo reconstruyendo datos antiguos. Muchos pacientes se habían marchado de Connecticut, y era imposible saber si sobrevivieron o murieron. Según el doctor Leonard Engel, que informó con posterioridad acerca de este estudio, la influencia de este factor es «suficiente para explicar toda la mejora que se pretende haber alcanzado en porcentaje de supervivencia».

Para que un informe basado en una muestra tenga valor, debe utilizar una muestra representativa, donde se hayan eliminado todos los posibles factores de influencia. En este punto las cifras de Yale muestran su escaso valor. En este punto también, muchas de las cosas que usted lee en los periódicos y en las revistas revelan su inherente falta de significado.

Un psiquiatra informó una vez de que prácticamente todo el mundo está neurótico. Aparte del hecho de que tal uso de la pala-

bra «neurótico» destruye todo su significado, vamos a examinar la muestra utilizada por el doctor. Es decir ¿a quién observó el psiquiatra? Resulta que llegó a esta edificante conclusión partiendo del estudio de sus pacientes, que distan mucho de representar una muestra de la población. Si una persona no tuviera algún problema no hubiera acudido a un psiquiatra.



Examine dos veces lo que lea, y evitará creer una cantidad enorme de cosas que no son verdad.

Vale la pena tener en cuenta también que la representatividad de una muestra puede ser destruida con la mayor facilidad, tanto por influencia de factores visibles como invisibles. Es decir, incluso en el caso de que no pueda demostrarse que existe un factor de influencia apreciable, conserve cierto grado de escepticismo sobre los resultados, siempre que haya una posibilidad de influencia en alguna parte. Y siempre la hay.

Las elecciones presidenciales de 1948 y de 1952 serían suficientes para probar lo que acaba de afirmarse, por si quedase alguna duda.

Para más pruebas, retrocedamos hasta el año 1936, cuando ocurrió el célebre desastre del *Literary Digest*. Los diez millones de abonados, entre poseedores de teléfono y suscriptores del *Digest*, que aseguraron a los editores de la desahuciada revista que Landon alcanzaría 370 votos y Roosevelt 161, procedían de una relación que predijo con acierto la elección de 1932. ¿Cómo podía existir un factor de influencia en una relación que ya había sido probada?

Todo el mundo apoya a Alf



Existía, tal como se demostró a posteriori. Las personas que disponían de teléfono y estaban suscritas a revistas en el año 1936 no eran una sección representativa de los votantes. Económicamente formaban un grupo especial, y constituían una muestra influenciada porque predominaban en ella los que resultaron ser votantes republicanos. La muestra eligió a Landon, pero los votantes pensaron de modo distinto.

La muestra básica es la llamada «aleatoria», la cual se selecciona por azar, tal como indica la palabra, partiendo de una «población», vocablo con que el experto en estadísticas describe el con-

junto del cual la muestra es una parte. Se toma uno de cada diez nombres buscados en un fichero o relación. De un sombrero se extraen cincuenta papeletas dobladas. Se entrevista a una de cada veinte personas que pasan por Market Street (pero recuerde que ésta no sería una muestra de la población del mundo, ni de los Estados Unidos, ni de San Francisco, sino solamente de la gente que pasaba por Market Street en aquel momento). Con motivo de una encuesta de opinión, una entrevistadora dijo que ella buscaba a la gente en una estación de ferrocarril porque «en una estación puede encontrarse a toda clase de personas». Debieron señalarle que las madres con niños pequeños, por ejemplo, debían ofrecer una representación muy exigua en su estación de ferrocarril.

Para comprobar una muestra tomada al azar se procede según el criterio siguiente: ¿Tiene cada nombre o cosa del grupo total la misma posibilidad de formar parte de la muestra?

La muestra escogida al azar es la única que puede examinarse con completa confianza por medio de la teoría estadística, pero existe un factor desfavorable a la misma. Es tan difícil y cara de obtener, en muchos casos, que queda descartada por su coste. Un sustituto más económico, usado en todo el mundo para trabajos tales como las encuestas de opinión y estudios de mercado, es la llamada muestra al azar por estratos.

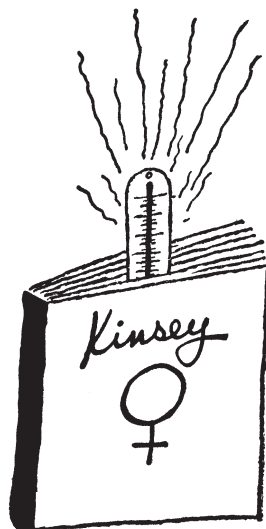
Para obtener esta muestra estratificada se divide la población en partes y se toman grupos proporcionales a las mismas. Justamente entonces pueden empezar sus problemas: la información acerca de la proporcionalidad tal vez no sea correcta. Usted instruye a los entrevistadores para que procuren hablar con un número determinado de personas de color, con tal o cual porcentaje de personas que se encuentren en distintos niveles de salarios,

con un número especificado de granjeros, etc. Al mismo tiempo, todo el grupo debe dividirse por igual en personas de más de cuarenta años y de menos de cuarenta. Esto suena bien pero, ¿qué ocurre? Tratándose de blancos o de negros, el entrevistador acertará la mayoría de las veces; en cuanto a los ingresos, cometerá mayor número de errores; y por lo que se refiere a los granjeros, ¿cómo clasificaría usted a un hombre que trabaja en una granja parte de su tiempo y dedica la otra parte a trabajar en la ciudad? Hasta la cuestión de la edad puede presentar algunos problemas, que se solucionan fácilmente entrevistando a las personas que queden claramente por debajo y por encima de los cuarenta. En tal caso, la muestra se verá afectada por la ausencia virtual de quienes se aproximen a los cuarenta y quienes los sobrepasen ligeramente. No hay por dónde escapar.

Más aún: ¿cómo se obtiene una muestra al azar dentro de la estratificación? El sistema utilizable a primera vista es empezar con una relación completa y dirigirse a los nombres que se escojan al azar; pero esto resulta demasiado caro. Por lo tanto, sale usted a la calle y su muestra se ve afectada por los que están en casa; va usted de puerta en puerta y deja de ver a la mayor parte de las personas que tienen un empleo; se dedica a las entrevistas nocturnas y faltarán los que van al cine o a un club nocturno.

La labor de la encuesta de opinión se convierte en una batalla contra las distintas influencias, y libran esta batalla constantemente todas las organizaciones de encuestas con buena reputación. El lector de informes debe recordar que esta batalla nunca se gana. No se debe leer ninguna conclusión en el sentido de que «el 65 por 100 de la población americana» es contrario a una u otra cosa, sin formularse la pregunta: ¿el 65 por 100 de qué población americana?

Así ocurre con *The Female Volume*, el famoso estudio sobre la sexualidad americana del doctor Alfred C. Kinsey. El problema, al igual que con cualquier otro estudio basado en una muestra, es cómo leerlo (o leer un resumen divulgativo del mismo) sin dar por ciertas cosas que pueden no serlo.



En este informe entran en juego por lo menos tres factores de muestreo. Las muestras de la población escogidas por el doctor Kinsey (primer factor) distan mucho de ser tomadas al azar y tal vez no sean del todo representativas, pero se trata de muestras amplias en comparación con cualquier otro trabajo llevado a cabo anteriormente, y sus cifras deben aceptarse como informativas e importantes, si bien no es necesario tomarlas al pie de la letra. Quizá sea más importante recordar que cualquier cuestionario no es sino

una muestra (segundo factor) de las posibles preguntas, y que la contestación de la señora no es sino una muestra (tercer factor) de su actitud y su experiencia sobre cada una de las cuestiones.

Los componentes del equipo de investigadores pueden influir en los resultados dándoles un matiz distinto. Durante la segunda guerra mundial, el National Opinion Research Center envió dos equipos de entrevistadores a hacer preguntas a quinientos negros de una ciudad del Sur. Uno de los equipos estaba formado por entrevistadores blancos y el otro por entrevistadores de color.

Una de las preguntas era: «¿Recibirían los negros mejor o peor trato si los japoneses conquistasen Estados Unidos?». Los entrevistadores negros informaron que el 9 por 100 de los entrevistados había contestado que recibirían «mejor» trato, mientras que los entrevistadores blancos obtuvieron solamente un 2 por 100 de tales respuestas. Y mientras que los entrevistadores negros informaron de que sólo un 25 por 100 creía que los negros recibirían peor trato, los entrevistadores blancos alcanzaban el porcentaje del 45 por 100.

Cuando se substituyó en la pregunta a «japoneses» por «nazis», los resultados fueron similares.

La tercera pregunta exploró actitudes que podían basarse en sentimientos contenidos en las dos primeras. «¿Qué cree más importante: concentrarse para atacar al Eje o hacer que la democracia funcione mejor en nuestro país?». La respuesta del 39 por 100, según los entrevistadores negros, fue «atacar al Eje». Esta misma respuesta fue la del 62 por 100 según los entrevistadores blancos.

Nos encontramos ante una tendencia producida por factores desconocidos. El más palpable es una influencia que siempre debe tenerse en cuenta cuando se lean los resultados de las encuestas: se trata del deseo de quedar bien. ¿Sería extraño que al contestar a una pregunta afectada de una connotación de desleal-

CÓMO MENTIR CON ESTADÍSTICAS

tad en tiempos de guerra, un negro del Sur dijera a un blanco lo que había de sonarle bien, en lugar de su verdadera opinión? También es posible que los distintos grupos de entrevistadores se dirigieran a distintas clases de personas. En cualquier caso, los resultados presentan un factor tendencioso que les hace perder todo valor. Usted puede juzgar por sí mismo cuántas conclusiones basadas en encuestas contienen semejantes influencias, lo cual las hace perder todo su valor, y sin que, por otra parte, haya ningún medio de comprobación para desenmascararlas.



Usted ya dispone de pruebas suficientes para actuar, si sospecha que en general las encuestas están influenciadas en uno u otro sentido. En el caso del error del *Literary Digest*, este factor tendencioso resulta de dirigirse a las personas que poseen más dinero, mejor educación, mejor información y mayor actividad, mejor aspecto, mejor comportamiento convencional, y unas costumbres más fijas que el promedio de la población que teóricamente representan.

Fácilmente puede verse el efecto de esta circunstancia. Supongamos que usted es un entrevistador y que le han asignado una esquina de una calle, con una entrevista a realizar. Usted localiza a dos hombres que parecen encajar en la categoría fijada: más de cuarenta años y habitantes de la ciudad. Uno ofrece un aspecto limpio y cuidado. El otro va sucio y parece poco amigable. Con intención de terminar pronto su trabajo, usted se acerca al individuo de aspecto más agradable, y todos sus colegas en todo el país toman la misma decisión.

Los sentimientos más contrarios a las encuestas de opinión podrían hallarse entre los círculos liberales o de izquierdas, donde se cree que las encuestas están generalmente falseadas. Tras este punto de vista se encuentra el hecho de que los resultados de las encuestas raramente coinciden con las opiniones y deseos de quienes no tienen ideas conservadoras. En Estados Unidos, por ejemplo, las encuestas parecen elegir siempre a los republicanos incluso cuando los votantes, poco después, hacen lo contrario.

En realidad, como acabamos de ver, no hace falta que la encuesta esté falseada, es decir, que se tergiversen los resultados deliberadamente a fin de crear una falsa impresión. La tendencia de la propia muestra a presentar un factor de influencia puede falsearla automáticamente.