



## EL VALOR DE LAS MUESTRAS REPRESENTATIVAS








## Lo dicen nuestros clientes y así es.

Somos fieles a nuestros principios y somos fieles a nuestros clientes.

Por eso, nos esforzamos por responder siempre con el máximo de calidad, rigurosidad y con total transparencia. Solo así se consigue crecer junto a nuestros clientes y mantener su confianza.

Con ODEC, en buenas manos.

-  **Captura de datos**
-  **Tratamiento de información**
-  **Presentación de resultados**
-  **Desarrollo de software**
-  **Outsourcing de servicios**

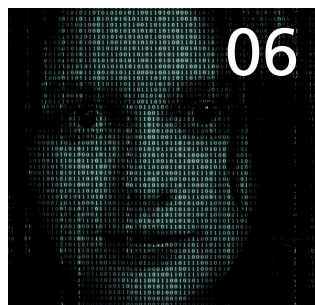


*your data partner*

902 519 090

odec@odec.es

www.odec.es



## Márgenes de error

- 06 Los errores muestrales 'a teima' de nuestro querido Carlos Lamas  
Isabel S. Peleteiro y Jesús Pascual

## Muestras y tecnología

- 12 Tickets de compra, mobile y la importancia de utilizar muestras para eliminar sesgos  
David Díez Mocha

## Observación

- 16 Observando el cosmos (humano)  
Joan S. Alós

## Audiencias digitales

- 18 Los paneles en la era digital  
Jaime Agulló

## Metodologías complementarias

- 26 En busca de la mejor foto  
Mercedes Fernández y Daniel Núñez

## Datos y población

- 30 Muestras reresentativas en la era digital  
Macarena Estévez

## Neuromarketing

- 34 La muestra y la revolución tecnológica: mención al neuromarketing  
Álvaro Saavedra Montejo

## Dimensión

- 38 El tamaño siempre importa  
José Andrés Gabardo

## La opinión del Consejo Editorial

- 42 El tamaño mató a la representatividad  
Eduardo Madinaveitia





38° SEMINARIO  
**TELEVISIÓN**  
SITGES 2023



**¡GRACIAS DE CORAZÓN!**



I&M n° 155 | abril 2023

#### Consejo Editorial

David Alameda, *profesor de Marketing en la Universidad Pontificia de Salamanca* / Miguel Cervantes, *profesor de marketing e ITM de la Universidad de León* / Carles Manteca, *consultor de estrategia, medios, marketing, contenidos e innovación* / Eduardo Madinaveitia / Xavier Moraño, *director de Empírica* / Margarita Ollero, *chief data & insight officer en Havas Media Group* / Felipe Romero, *socio director de The Cocktail Analysis* / Juana Rubio-Romero, *departamento de publicidad y marketing de la Universidad Antonio de Nebrija de Madrid*

#### Vocal de Publicaciones

David Alameda, *profesor de Marketing en la Universidad Pontificia de Salamanca*

#### Publicidad

secretaria@ia-espana.org

#### Secretaría Insights Analytics España y publicidad

Tel. 637 209 380 / secretaria@ia-espana.org

#### Coordinación

Maite Sáez

#### Dirección de arte y maquetación

Maribel Guerrero / Ana Pinos

#### Impresión

Gráficas Iglesias Impresores / San Romualdo, 26  
28037 Madrid / Tel. 91 754 44 75 / iglesias@agd.es

#### Depósito Legal

M-35944-1979 ISSN: 1131-6144

Incluida en el índice de revistas de prestigio científico reconocido, OM de 2-12-94, BOE 289 y la Resolución Ministerial de 6-11-96, BOE 280, elaborado por CENAI para el área de Ciencias Económicas y Empresariales. EJEMPLAR GRATUITO.

*La revista Investigación & Marketing no se hace responsable de las opiniones vertidas por sus colaboradores.*

C/ Ferraz 35 - 3º dcha. / 28008 Madrid

Tel: 637 209 380

www.ia-espana.es

*Esta revista se imprime en papel ecológico.*



/ PRESENTACIÓN

## Representatividad

Las muestras representativas, un ejercicio de ingeniería estadística que busca conseguir una fotografía lo más fidedigna posible de una población a través de una parte de ella, seleccionada teniendo en cuenta su representación sobre los distintos grupos por edad, sexo, rol, etcétera, podría parecer que tienen sus días contados en un entorno en el que lo digital multiplica la posibilidad de acceso a los datos de los individuos.

No obstante, tras la explosión de esta última realidad, con el crecimiento exponencial de herramientas tecnológicas que cada día facilitan millones y millones de datos, la idea de que lo importante sigue siendo el tratamiento de estos y no su posesión, nos lleva a reflexionar en este número de *Investigación & Marketing* sobre el valor de las muestras representativas. Máxime cuando, en paralelo, la necesidad de la integración de fuentes, modelos y técnicas es un imperativo creciente en el ejercicio de la investigación sobre cualquier aspecto.

El tamaño de las muestras, su comprensión, el exquisito respeto en la formulación de cada uno de los aspectos que configuran un modelo representativo son algunas de las consideraciones que se vierten en los artículos que se publican en este número. Así como también su idoneidad para calibrar y enriquecer los datos que aporta la captura digital. Más en un momento en el que las metodologías de inteligencia artificial parecen trasladarnos a otro entorno (también en el ámbito de la investigación).

Este número se planeó como un pequeño homenaje a uno de los profesionales que más defendió a lo largo de toda su trayectoria la rigurosidad y el valor de la representatividad, Carlos Lamas, fallecido el pasado mes de diciembre y cuyo recuerdo recorre algunos de los textos que incluye.

# LOS ERRORES MUESTRALES 'A TEIMA'<sup>1</sup> DE NUESTRO QUERIDO CARLOS LAMAS

Este número de la revista 'Investigación & Marketing' se planteó como un pequeño homenaje a Carlos Lamas, (fallecido el pasado mes de diciembre), gran defensor del rigor, la exactitud en el manejo de los términos y el valor de los muestreos probabilísticos. Y en este artículo, sus autores reivindican su predicamento.



.....  
**Isabel S. Peleteiro Ramos**  
Socia-directora general de  
IMOP Insights.



.....  
**Jesús Ramón Pascual Adrián**  
Socio fundador de IMOP  
Insights.

**"El error de muestreo es solamente la punta del iceberg dentro de los errores de los estudios por encuesta. El error total de los estudios es mucho mayor que el que denota el error de muestreo. Sin embargo, en la práctica se pone el énfasis principal en el error de muestreo porque hay elegantes fórmulas matemáticas que estiman su magnitud, mientras que la magnitud de los otros errores se ve demasiado difícil de estimar".** Herbert Weisberg (*Total Survey Error Approach*).

Con este artículo pretendemos reflexionar sobre uno de los temas que dominaba nuestro querido y ausente Carlos Lamas, y cuyo predicamento ha caído con frecuencia en el desierto.

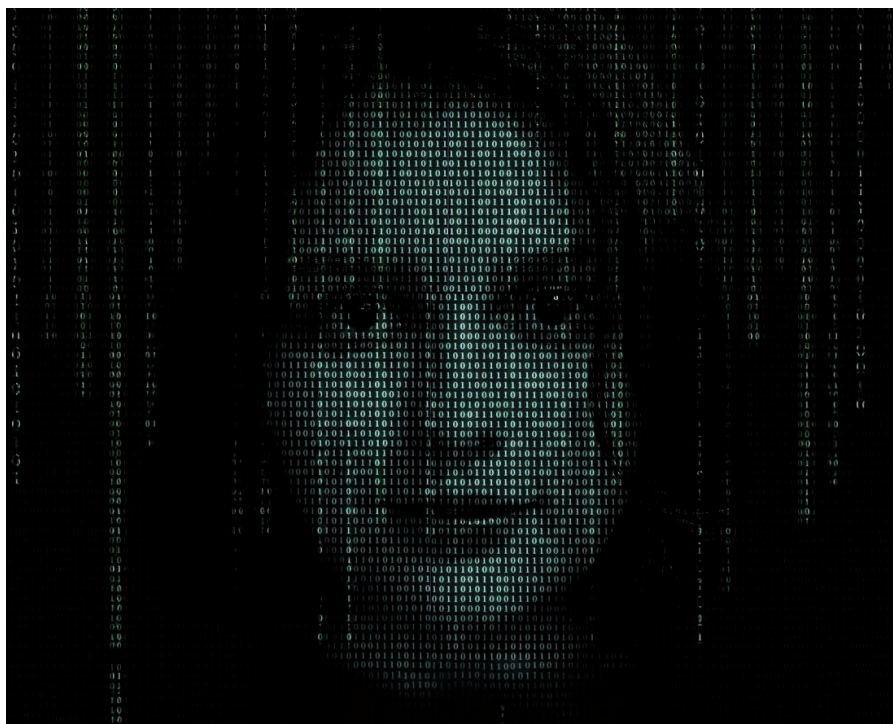
En sus últimos años, ya retirado de la vida profesional, Carlos mantuvo su interés en profundizar en todo lo referente a los márgenes de error muestral. Sinceramente, poca aportación personal hay en este artículo –hasta la cita es referencia de Carlos–, simplemente recogemos con nuestras propias palabras algunas de

sus reflexiones y enseñanzas y algunas certezas que se pueden encontrar en cualquier manual de estadística básica pero que, curiosamente, parecemos olvidar en la práctica cotidiana.

Cuando queremos llevar a cabo una estimación por muestreo y buscamos resultados fidedignos se debe tener en cuenta:

1. el tamaño de la muestra,
2. el método de selección de los elementos de la muestra,
3. el procedimiento de cálculo de las estimaciones (estimadores).

Carlos era un claro defensor de los muestreos probabilísticos. No era ningún ingenuo y sabía que en las encuestas es muy difícil conseguir un muestreo probabilístico, entre otras razones por la ausencia de marcos muestrales, las crecientes negativas a contestar y la dificultad diferencial para contactar con algunos colectivos. Por esto, lo que preconizaba era aproximarse lo más posible a un muestreo probabilístico cuidando el proceso de selección del hogar a través de la selección aleatoria de los individuos dentro del mismo, con aplazamiento de entrevistas y con la repetición de llamadas/visitas al hogar en diferentes días y horarios.



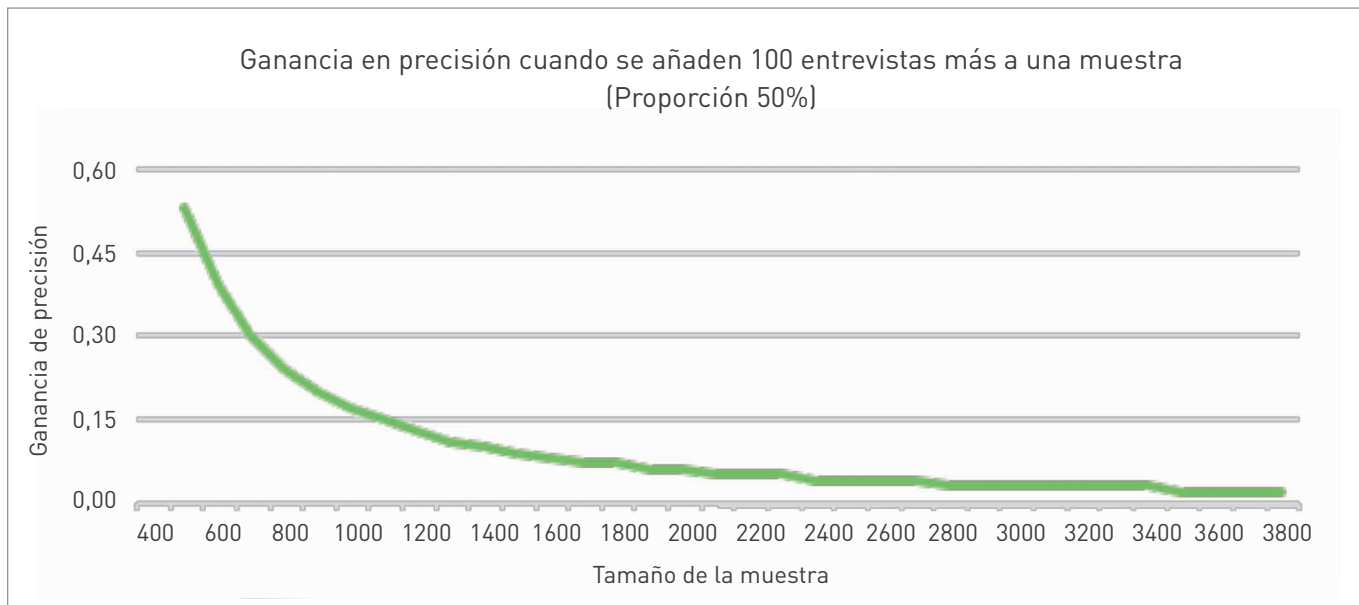
## EL MUESTREO PROBABILÍSTICO MITIGA FACTORES ESPURIOS QUE PUEDEN AFECTAR A LA SELECCIÓN DEL ENTREVISTADO O A SU PROPENSIÓN A PARTICIPAR EN UN ESTUDIO

Carlos comentaba que en el sector con frecuencia confundimos aleatorio con probabilístico, cuando lo probabilístico exige de lo primero, pero no necesariamente a la inversa. Como es conocido, para que un muestreo sea probabilístico todas las personas del universo al que nos dirigimos deben tener una probabilidad de ser entrevistadas distinta de cero y conocida. El muestreo probabilístico mitiga factores espurios que pueden afectar a la selección del entrevistado o a su propensión a participar en un estudio. En este sentido, en la medida en que las muestras se alejan del muestreo probabilístico empiezan a presentar ciertos sesgos.

También pensaba que del término representatividad se hace un uso frecuente y poco riguroso. La representatividad no es una característica de las muestras, es el objetivo que el diseño muestral busca, es nuestra aspiración, ni siquiera una buena selección aleatoria la garantiza.

La dificultad real de llevar a cabo este tipo de muestreo justifica que en las fichas técnicas, cuando se habla de los márgenes de error muestral, se diga “en el supuesto de un muestreo aleatorio simple tendremos un margen de error muestral de ...” ¿Qué jugarretas nos juegan a veces las palabras! Llamar muestreo aleatorio simple al más difícil de conseguir, porque, ¿cuándo disponemos en investigación social de un marco muestral tan completo que nos permita conocer y garantizar a priori la equiprobabilidad de cada miembro del universo?

El margen de error de una estimación depende del estimador de la proporción y del tamaño de la muestra. A medida que se incrementa el tamaño de la muestra disminuyen los errores absolutos para un mismo estimador de la proporción. De forma similar, si mantenemos un mismo tamaño muestral, a medida que baja la proporción desde el 50% se reduce el



## CONTRA LA LÓGICA APARENTE DEL LENGUAJE, UN MENOR MARGEN DE ERROR MUESTRAL ABSOLUTO NO SIGNIFICA SIEMPRE UNA MAYOR PRECISIÓN

Observando la tabla y el gráfico vemos que cuando aumentamos 100 entrevistas a la muestra cuando esta es pequeña, ganamos bastante precisión.

A medida que la muestra va creciendo cada vez que añadimos 100 entrevistas es menor la ganancia en precisión.

Esto nos indica que a partir de un tamaño muestral no merece la pena incrementarla porque la ganancia en precisión va a ser muy pequeña.

margen de error muestral absoluto (y se incrementa el relativo).

Observando la tabla y el gráfico vemos que cuando aumentamos 100 entrevistas a la muestra cuando esta es pequeña, ganamos bastante precisión.

A medida que la muestra va creciendo cada vez que añadimos 100 entrevistas es menor la ganancia en precisión. Esto nos indica que a partir de un tamaño muestral no merece la pena incrementarla porque la ganancia en precisión va a ser muy pequeña.

Pero, volvamos a las fichas técnicas. En ellas se suelen remarcar los márgenes de los errores muestrales diciendo: en el caso de  $p=q=50\%$  el margen de error muestral es de  $\pm 3\%$  para un nivel de confianza del 95%; otro de

los errores recurrente en el sector, el símbolo del porcentaje, no se trata del 3% de la estimación, sino de  $\pm 3$  puntos porcentuales.

Pero, en cualquier caso, esto dice poco de la precisión de los datos. Primero, porque no hay un error muestral para una encuesta, sino que cada estimación (cada pregunta) tiene su propio margen de error muestral. El 50% solamente nos da información de aquellas estimaciones con ese valor. Y, segundo, este margen de error muestral es el absoluto que nos marca la horquilla en que se va a mover la estimación, pero deberíamos contemplar también el error relativo —el gran olvidado—.

De nuevo, contra la lógica aparente del lenguaje, un menor margen de error



muestral absoluto no significa siempre una mayor precisión. La precisión de la estimación de una proporción viene dada por el error relativo. Desde una perspectiva estadística, el error relativo es el resultado de dividir el margen de error muestral absoluto por la estimación de la proporción, ahora sí, con el símbolo % (porcentaje que el error representa sobre el valor de la estimación).

Es necesario trabajar con el error relativo a la hora de conocer la precisión de la estimación, porque el margen de error muestral absoluto nos sirve solamente para definir la horquilla en la que se mueve la estimación.

En el cuadro y en el gráfico observamos que, para un mismo tamaño de la muestra, si la proporción se reduce por debajo del 50%, también se va reduciendo el margen de error muestral absoluto y, contrariamente, se va incrementando el error relativo. De aquí la necesidad de buscar la acuracidad o cualidad que combina precisión –asociada al tamaño de la muestra y la variabilidad de

## LA REPRESENTATIVIDAD NO ES UNA CARACTERÍSTICA DE LAS MUESTRAS, ES EL OBJETIVO QUE EL DISEÑO MUESTRAL BUSCA, ES NUESTRA ASPIRACIÓN

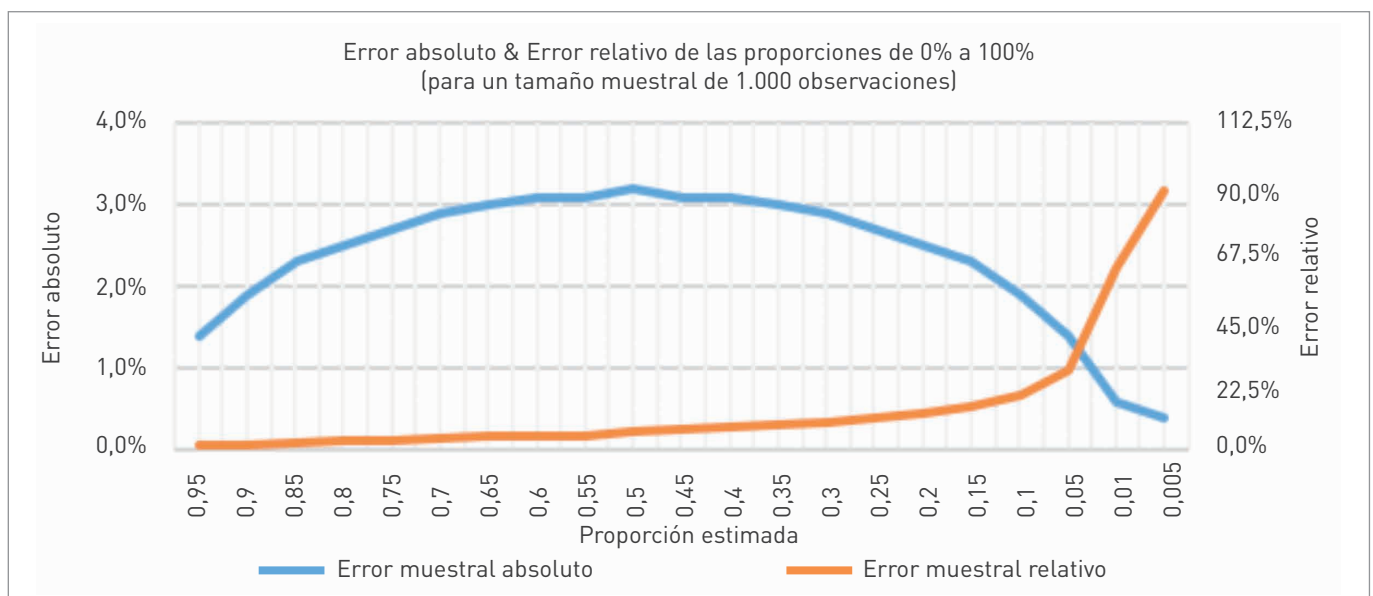
los participantes– y exactitud o proximidad entre el resultado obtenido y el valor real, lo que depende de la representatividad y de la ausencia de sesgo.



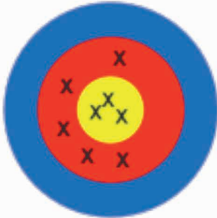
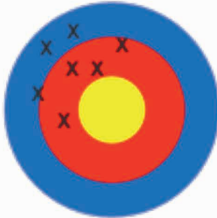
Reflexionando sobre la diferencia que hay entre preciso y exacto, observamos que con un tamaño suficiente de la muestra podemos conseguir precisión, es decir, que si repitiésemos la misma muestra en las mismas condiciones, obtendríamos dentro del margen de error las mismas estimaciones, pero hay sin duda otros factores que pueden estar desviando la estimación de la proporción real, múltiples sesgos

que tienen todas las encuestas; por ejemplo, sesgos de deseabilidad social, sesgos del instituto, del entrevistador, etcétera. En casi todos los casos estos sesgos –que Carlos llamaba operativos– imposibles de medir, incrementan el margen de error. No sabemos con exactitud en qué medida en cada caso, pero lo incrementan, de ahí la falacia de poner decimales al margen de error y no ruborizarnos.

Los diseños muestrales teóricos subyacentes en las encuestas no son simples: varias fases en el proceso de selección de la muestra (estudios polietápicos), estratificación previa

En el cuadro y en el gráfico observamos que, para un mismo tamaño de la muestra, si la proporción se reduce por debajo del 50%, también se va reduciendo el margen de error muestral absoluto y, contrariamente, se va incrementando el error relativo.



	Exactitud	Inexactitud (error sistemático)
Precisión		
Imprecisión (baja reproducibilidad)		

los oyentes de una cadena de radio, etcétera.

Para ir terminando, creemos importante insistir en que al ver un dato obtenido por encuesta debemos reflexionar sobre cómo ha sido obtenido, su grado de precisión y capacidad para reflejar la realidad en un momento dado. Quizás deberíamos reflexionar entre todos sobre mejores prácticas para las fichas técnicas. Ante la presunta avalancha de encuestas que se nos avecinan hasta el 28 M, esta advertencia no es baladí. Antes de descalificar a todo el sector, seamos modestos y no le pidamos a las encuestas más de lo que nos pueden dar, pero exijamos que no nos den menos.

Con este artículo lo único que hemos pretendido es homenajear a nuestro amigo y maestro, Carlos Lamas, repitiendo sus enseñanzas. Si a alguien este artículo le ha motivado para aplicar sus recomendaciones, bienvenido sea a la senda del rigor que tanto predicó. ■

de la población con diferentes sistemas de afijación muestral por estrato, intervención de algún efecto conglomerado y la aplicación de procesos de ajuste/equilibrio a la muestra obtenida.

Factores todos ellos que se alejan del esquema de muestreo aleatorio simple y que complican substancialmente el cálculo teórico correcto de los márgenes del error de muestreo. El cálculo del error estándar en los diseños complejos exige la utilización de procedimientos especiales: series de Taylor o diferentes técnicas de replicación o *bootstrapping*.

### EQUILIBRAJE

En nuestro país estamos acostumbrados a utilizar el equilibrio para reproducir la representatividad de la muestra en aquellas variables de la población que son conocidas: sexo, edad, ubicación geográfica, estatus, tamaño del hogar, etcétera. Pero al equilibrar para conseguir esta 'representatividad', el peso de los individuos resulta desigual, en

unos casos estos pesos estarán más dispersos y en otros casos estarán más concentrados.

Es lógico pensar que a mayor dispersión de los pesos, más se incrementará el margen de error de muestreo absoluto. "Al dar diferentes pesos a los entrevistados para conseguir la representatividad de la muestra, estamos reduciendo la precisión de las estimaciones". 'Nuestros mayores'<sup>2</sup> nos han dado métodos para corregir el efecto del diseño y estimar el error en el caso de muestras ponderadas (obviamente en el supuesto de un muestreo probabilístico), pero en el día a día nos tememos que es una práctica poco frecuente.

Tampoco hay que olvidarse de los márgenes de error muestrales en las variables cuantitativas para estimar el total y el promedio. O del margen de error muestral de las diferencias. Por ejemplo, la diferencia que pueda haber entre dos olas del EGM al estimar los lectores de un diario, de una revista,

### Notas

1. 'Teima', del gallego: idea persistente.
2. La referencia a Leslie Kish es ineludible.



Adéntrate en la mente de los consumidores para guiar a las empresas en su estrategia y sus planes de innovación.

 A tiempo completo

 10 meses

 Madrid

# Master in Market Research and Consumer Behavior



# TICKETS DE COMPRA, 'MOBILE' Y LA IMPORTANCIA DE UTILIZAR MUESTRAS PARA ELIMINAR SESGOS

La tecnología, es indudable, brinda a la investigación grandes y nuevas oportunidades para resolver cuestiones y aportar valor a los proyectos. Pero ello no supone denostar la utilización de muestras representativas que, en algunos casos, siguen aportando interesantes matices que se escapan a la automatización. Este artículo aborda la necesaria convivencia de metodologías y las bondades específicas de cada una de ellas.



.....  
David Diez Mocha  
'Sales development director'  
en Gelt  
\_\_\_\_\_

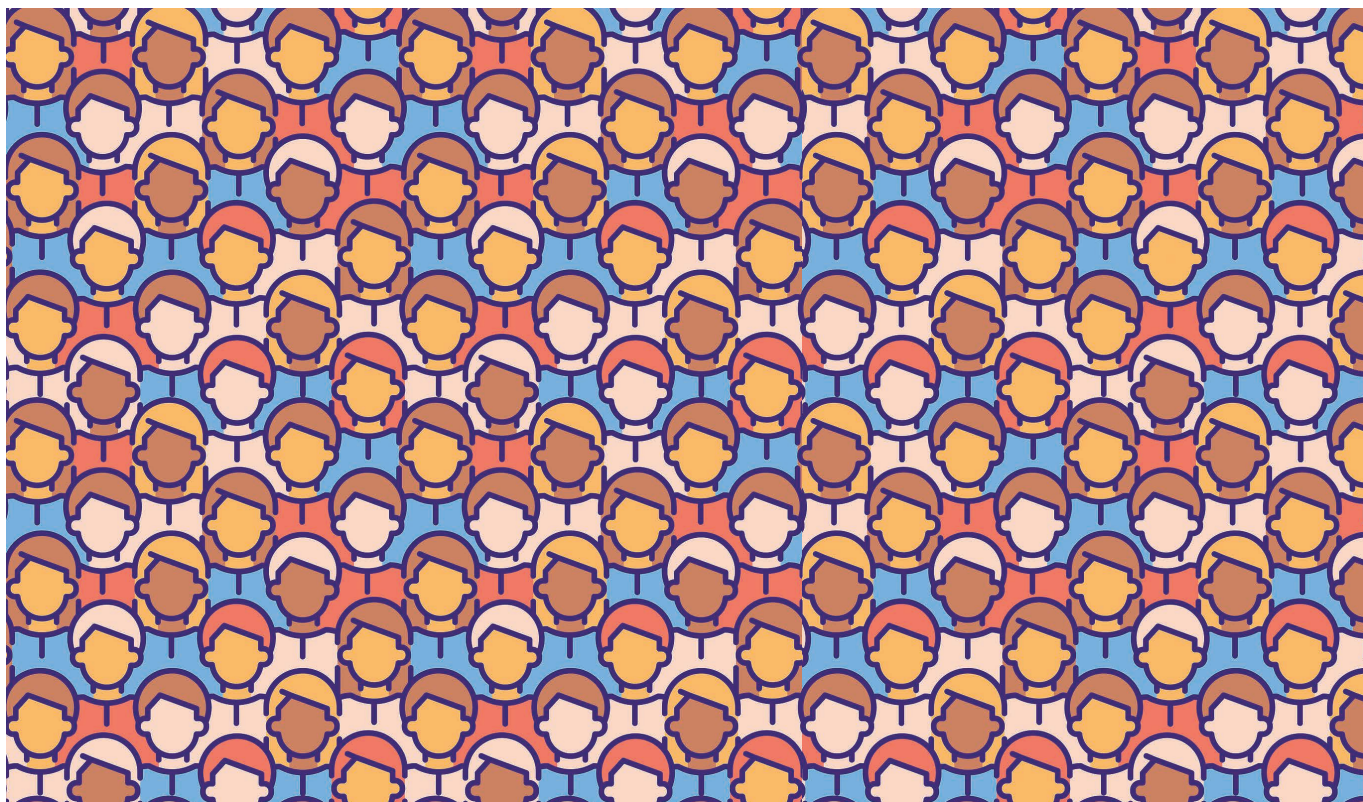
“Mas caro el collar que el perro”. Veníamos, yo al menos por mis años, de usar muestras representativas como única manera de llevar a cabo estudios recurrentes o puntuales, que de otra manera eran imposibles de abordar, ya fuera por la inexistencia de información completa o por el coste desmesurado de acceder a los datos necesarios. Esta realidad, hacía que la inversión necesaria fuera más elevada que los potenciales beneficios de los resultados. Pero hoy en día, la tecnología ha hecho que este decorado cambie por completo. Tanto, como que aquella carestía de datos completos y fiables ha mutado a una situación donde la mayoría de las veces nos enfrentamos a situaciones en las que contamos con demasiada información y la definición de muestras representativas nos ayuda a separar el polvo de la paja y, entre otras cosas, también a asegurar la solvencia y fortaleza de los resultados, algo vital en las decisiones estratégicas de alto impacto.

Por tanto, creo que el dinamismo y evolución de la tecnología en las últimas décadas nos está otorgando

a los profesionales de la investigación de mercados grandes oportunidades de resolver cuestiones y aportar valor a los proyectos de maneras que hace poco nos hubiera costado imaginar. Mas rápido, más eficaz, más eficiente, más profundo, más sólido. Creo que, en este aspecto, habrá poco debate al respecto.

No pienso, sin embargo, que la utilización de muestras representativas esté en tela de juicio, ni siquiera en estos momentos en los que la tecnología (y las posibilidades que nos abre) se ha vuelto indispensable.

En nuestro caso particular, en el que el uso de tecnologías como OCR, *deep learning*, o inteligencia artificial son protagonistas desde el minuto uno hasta el último *insight* que extraemos, la definición y utilización de muestras representativas es uno de los valores fundamentales y necesarios en la práctica totalidad de los proyectos que abordamos con nuestros clientes. Fundamentalmente seleccionamos y definimos muestras con métodos no probabilísticos, dada la gran cantidad de datos, usuarios e histórico de los



que disponemos. En definitiva, sin la definición de muestras sería imposible eliminar sesgos que a la postre lastrarían las conclusiones, los *insights* y, por extensión, la rentabilidad de las decisiones a tomar.

Nuestros clientes, en su mayoría marcas de gran consumo y *retailers* en áreas de marketing, *trademarketing* y *research*, suelen centrar sus objetivos y proyectos en:

- Mejorar el conocimiento del *shopper* y sus hábitos de compra e histórico de consumo para adelantarse a sus posibles requerimientos y decisiones de compra.
- Detectar o confirmar tendencias, a veces incipientes, en nuestros hábitos de compra y consumo.

- Entender en profundidad los comportamientos y necesidades de los compradores.
- Identificar desventajas competitivas.
- Comprender nuestros patrones de compra en las visitas a cada supermercado, mediante modelos de clasificación y predicción de compra.

En definitiva, *insights* que acaban formando parte de decisiones estratégicas inteligentes en diferentes ámbitos, como pueden ser la definición de producto, comunicación, packaing, precio, acciones promocionales y/o surtido, retención y fidelización de clientes o potencial por categoría y código postal, entre otros.

El uso de muestras, por tanto, lejos de ser una opción en nuestro contexto,

## ¿SERÁ EL USO Y APLICACIÓN DE DETERMINADAS IAS EL SIGUIENTE PASO EN NUESTRA INDUSTRIA?

# EL USO DE MUESTRAS, LEJOS DE SER UNA OPCIÓN EN NUESTRO CONTEXTO, ES UNA HERRAMIENTA ABSOLUTAMENTE IMPRESCINDIBLE PARA ELIMINAR SESGOS Y SOBRE INDEXACIONES, QUE, DE OTRA MANERA, NOS LLEVARÍAN A CONCLUSIONES PARCIALES Y A PROMOVER DECISIONES ERRÓNEAS EN LAS ESTRATEGIAS DE NUESTROS CLIENTES

es una herramienta absolutamente imprescindible para eliminar sesgos y sobre indexaciones que, de otra manera, nos llevarían a conclusiones parciales y a promover decisiones erróneas en las estrategias de nuestros clientes.

En cuanto a la fusión de fuentes de información, afortunadamente estamos viendo cómo se ha convertido en una realidad, donde la tecnología vuelve a ser nuestra aliada en aspectos claves, como los procesos de automatización e integración de diferentes fuentes de datos, estructurados y no estructurados. ¿Será el uso y aplicación de determinadas IAs el siguiente paso en nuestra industria? Una cuestión apasionante, pero sin respuesta al menos por mi parte y de momento. Quizá sea una posible temática para una próxima edición de *Investigación & Marketing*.

Como ejemplo, basta observar el tiempo que se tardó en llegar a 100 millones de usuarios (Fuente: @engineers\_feed) en:

- Teléfono: 75 años
- Teléfono móvil: 16 años
- World Wide Web: 7 años
- iTunes: 6,5 años
- Twitter: 5 años
- Facebook: 4,5 años
- WhatsApp: 3,5 años
- Instagram: 2,5 años
- App Store Apple: 2 años
- ChatGPT: 2 meses

Donde probablemente no haya vuelta atrás, sea en el mundo 100% digital, entorno dominado desde

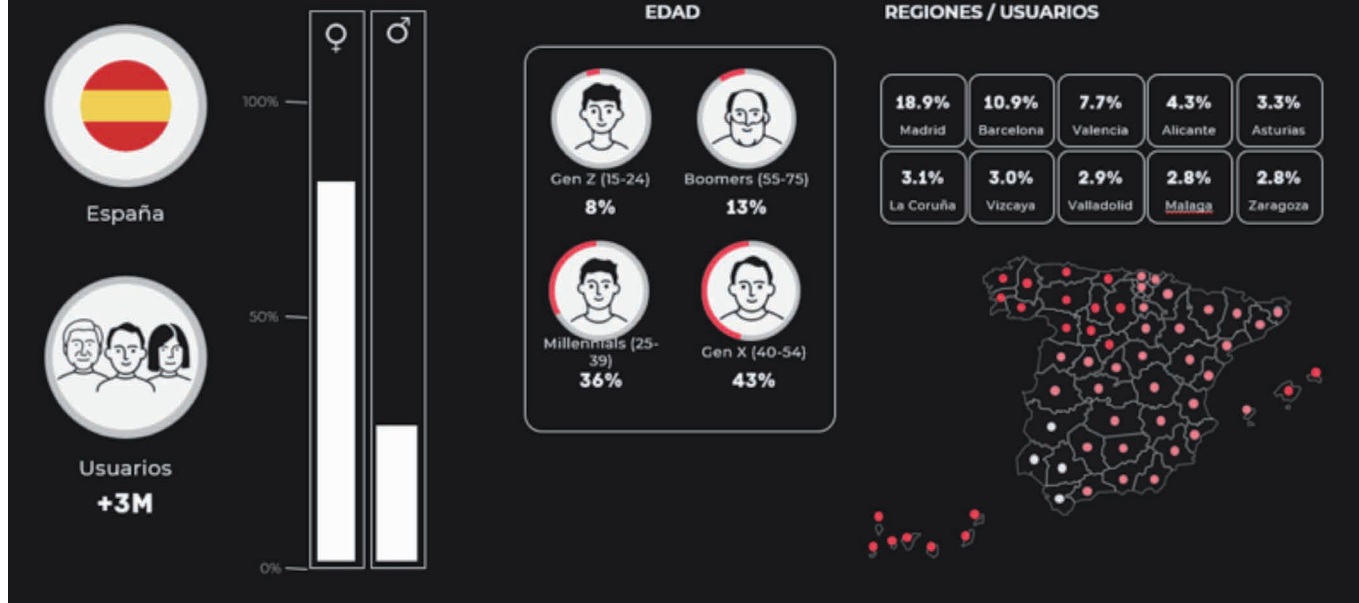
hace tiempo por los datos en tiempo real, donde una vez más la tecnología ha jugado un papel absolutamente catalizador. Herramientas en este caso relacionadas con soluciones en el campo de *dashboarding*, seguimiento, trazabilidad, atribución y automatización de eventos, hacen que el uso y definición de muestras representativas tenga una importancia menor.

Ambos métodos, estudios de investigación de mercado con muestras y el uso de la tecnología como la inteligencia artificial, tienen en nuestra experiencia sus ventajas y desventajas, dependiendo del objetivo y del contexto específico de la investigación.

Sin embargo, en algunos casos los estudios de investigación de mercado clásicos con muestras pueden ser mejores que el uso de la inteligencia artificial debido a las siguientes razones:

1. **Flexibilidad en el diseño de la muestra:** La investigación de mercado con muestras permite a los investigadores seleccionar la muestra de acuerdo con criterios específicos que se adapten al objetivo de la investigación. Por ejemplo, se pueden seleccionar grupos específicos de clientes o consumidores en función de su ubicación geográfica, edad, ingresos, hábitos de compra, etcétera. En cambio, la inteligencia artificial puede no ser capaz de seleccionar muestras tan específicas y personalizadas.

# Composición del panel



2. **Validación de resultados:** Los estudios de investigación de mercado con muestras permiten la validación de resultados a través de técnicas estadísticas y de análisis. Esto significa que los resultados obtenidos de la muestra pueden ser generalizados y aplicados a una población más amplia con cierto nivel de certeza. En cambio, la inteligencia artificial puede no proporcionar esta validación, debido a la falta de una muestra representativa.

3. **Comprensión del contexto:** La investigación de mercado con muestras también permite a los investigadores entender el contexto en el que se realiza la investigación, lo que puede afectar a los resultados. Por ejemplo, el contexto social, económico, cultural, etcétera, pueden influir en la forma en que los consumidores toman decisiones

de compra. La inteligencia artificial puede no ser capaz de tener en cuenta este contexto y puede producir resultados inexactos.

4. **Intuición y experiencia:** Los investigadores que realizan estudios de investigación de mercado con muestras tienen intuición y experiencia en la interpretación de los datos recopilados. Pueden proporcionar información adicional sobre los hallazgos y explicar los resultados en función de su conocimiento del mercado y la industria. En cambio, la inteligencia artificial puede producir resultados precisos, pero no puede proporcionar interpretaciones basadas en la intuición y la experiencia.

En resumen, los estudios de investigación de mercado con muestras pueden ser más apropiados

en situaciones donde se requiere una comprensión detallada y contextualizada de los clientes y consumidores, así como una validación estadística de los resultados. La inteligencia artificial puede ser útil en situaciones donde se requiere el análisis de grandes cantidades de datos, la identificación de patrones y tendencias, y la toma de decisiones automatizada. ■

# OBSERVANDO EL COSMOS (HUMANO)

Una exquisita reflexión sobre el sentido, el valor y la responsabilidad de la investigación, en esencia, así como respecto a la evolución metodológica y la necesaria convivencia de utilizar distintos métodos de observación para conseguir una imagen lo más fidedigna posible.



Joan S. Alós

Divulgador científico 'former president' Aedemo

**Otra mirada.** Estamos aprendiendo cada vez más, y más rápidamente del cosmos, básicamente por dos factores:

- a) Nos situamos fuera del sistema.
- b) Exploramos simultáneamente en una gran diversidad de longitudes de onda.

Situarse fuera del sistema significa dejar atrás la atmósfera que nos rodea. Significa alejarse de nuestro planeta, buscar lugares de observación estables, como puntos de Lagrange gravitacionalmente compensados, e incluso dejar atrás el mismo sistema solar.

Explorar significa utilizar todos nuestros 'sentidos', explorar en el espectro de la luz visible, pero también en infrarrojos, como que el James Webb, o en rayos X o gamma, los más penetrantes y letales, sin olvidarnos de las microondas y el espectro radiofónico y acústico. Traducido a escala humana, significa

'mirar' sin prejuicios y aceptar que un solo sistema o método de observación es totalmente insuficiente para comprender el universo que nos rodea.

Para completar este cuadro de exploración científica, añadiría:

- c) Los humanos nos hacemos preguntas para poner a punto métodos, sensores, inteligencia artificial que puedan observar y proporcionarnos datos.
- d) Interpretamos y contrastamos estos datos dentro de un marco teórico que, llegado el caso, podemos llegar a cambiar todo o en parte.

Bien, sin ánimo de aguar la fiesta, me pregunto qué hacemos cuando lo que se trata es de observar el 'cosmos humano', de lo que, entre otras disciplinas, se ocupa la investigación social y de mercados. Os invito a una reflexión abierta sobre ello: observar.





## NO SOMOS HISTORIADORES, NI NOTARIOS. LOS DATOS QUE OBTENEMOS PUEDEN SER ASÉPTICOS, PERO LOS RECOGEMOS PARA INFLUIR EN SU EVOLUCIÓN

En ciencia vale hacerse preguntas, pero no vale preguntar, pedirle por ejemplo a un experto, supongamos filósofo o teólogo, qué piensa acerca del universo. Se trata de observar, de obtener datos que otros puedan contrastar de forma independiente.

Saber para decidir, influir, cambiar. No somos historiadores, ni notarios. Los datos que obtenemos pueden ser asépticos, pero los recogemos para influir en su evolución.

No hay un único método de obtención de información que sea suficientemente representativo de un conjunto humano. Es preciso explorar en todas las longitudes onda. Si, por circunstancias, nos vemos constreñidos al uso de un único instrumento, para interpretarlo debemos basarnos en la evolución de

las mediciones y correlacionarla con aquellos factores que hayan podido contribuir a dicha evolución.

La inteligencia artificial es nuestra aliada para la obtención de indicadores que nos permitan de la forma más sencilla posible, evaluar el efecto e importancia de las complejas variables que afectan al cosmos humano.

Hay un dicho: “Únicamente si te haces preguntas inteligentes, las respuestas también lo serán”. Por lo tanto, pregúntate, observa y mira de entender, hasta donde sea posible, por qué pasa lo que pasa. ¿Seguimos hablando? jsalos62@gmail.com

Que la investigación te sea propicia. ■



# LOS PANELES EN LA ERA DIGITAL

Frente a la promesa inicial de que medir la audiencia digital –o cualquier otro aspecto relacionado con digital– sería más fácil que hacerlo en entornos analógicos, la realidad ha demostrado que aquella primera idea no se iba a cumplir. Este artículo repasa los diferentes pasos dados en las poco más de dos décadas de historia de la medición de las audiencias digitales y nos sitúa en el momento actual.

A comienzos de este siglo, cuando la primera burbuja digital terminaba de estallar y sus efectos aún no habían calado fuera de los mercados de bolsa más anglosajones, lo que entonces llamábamos internet vivía sus primeros pasos en el terreno de la investigación de mercados y, en particular, de la investigación de audiencias.

Para quienes no lo vivieron –o para quienes lo hayan olvidado–, en España nos encontramos, hacia finales de 2001, con hasta tres operaciones de medición de audiencia basadas en paneles de individuos: **Nielsen Netratings**, rama digital del gran árbol de la estadounidense Nielsen;

**Mediametrix**, una novedosa compañía, también estadounidense, nacida al calor del mundo digital, y **NetValue**, aportación europea –francesa en concreto– al panorama del momento.

Nadie quería perderse una oportunidad de crecimiento importante en unos años en los que las rondas de inversores permitían alegrías económicas desconocidas hasta entonces. Y el sector de la medición de audiencia no fue menos, siempre a caballo de la metodología estrella, los paneles de individuos con audímetros instalados, heredada de la medición de audiencia de televisión. Como bien



.....  
**Jaime Agulló**  
Managing VP de  
Comscore para Iberia  
\_\_\_\_\_



dijo en aquellos años el mayor experto español en paneles, Carlos Lamas, internet era un fenómeno que pedía a gritos paneles para medirlo. Y, al menos durante una década, tuvo razón.

Por tanto, la primera aproximación a la medición de audiencia digital, es decir, a la medición del comportamiento de consumo digital de los individuos y no a la medición del uso de la tecnología digital, que ya existía mediante las herramientas tecnológicas disponibles, se basó en la reconversión de las metodologías y tecnologías utilizadas históricamente para medir la audiencia de televisión: paneles de individuos con audímetros instalados.

En España, Nielsen Netratings lanzó sus primeras mediciones en 2001, en paralelo a la aparición de Mediametrix y poco antes de que NetValue se uniera al trio. La configuración metodológica y técnica era similar a la de la televisión, si bien con algunas diferencias clave.

Se trataba de recoger información de un panel de varios miles de individuos que formaban una muestra en principio representativa de la población internauta, que en aquel momento era una pequeña minoría todavía, aunque en rápida expansión. Esta muestra había aceptado colaborar instalando en su ordenador un software de audimetría capaz de recoger el comportamiento del usuario del ordenador en cuestión capturando las direcciones web de los sucesivos contenidos consumidos en ese dispositivo.

Y aquí se encuentra una primera diferencia con la televisión: frente al uso gregario de la televisión de entonces, centro de la vida del hogar, el ordenador era un elemento de uso principalmente individual, aunque pudiera tener varios usuarios. Quedaba mucho para la televisión actual, que migra a toda velocidad de los televisores a los dispositivos digitales personales, llámense ordenadores, tabletas o móviles. Además, el ordenador podía ser usado en casa o desde el trabajo, siendo en este segundo caso un ordenador propiedad de la empresa para la que trabajase el individuo panelista.

Por tanto, a diferencia de la televisión, los paneles digitales apuntaron a medir únicamente individuos, no individuos en hogares, con la complejidad de que

esos individuos podían ser sucesivos usuarios –distintos– de un mismo ordenador o tener acceso a este únicamente en horario laborable de lunes a viernes y fuera de su hogar. Eso sí: lo digital se ahorró discutir durante años sobre eso de los ‘invitados’.

Otra diferencia fue qué se medía en cada caso. Mientras en la televisión la clave para saber qué se medía era asignar correctamente la frecuencia de entrada de cada canal con la cadena que correspondiera y hacer un buen minutado de los contenidos emitidos, las cosas en internet no eran tan fáciles. No había tal ‘frecuencia’ asignable a una oferta de contenido, sino una sucesión de direcciones web –urls– a las que el ordenador había accedido y que podían –o no– pertenecer a sitios web reconocibles. En muchos casos, las urls contaban una historia cada vez más intrincada de contenidos descargados en el equipo –tal vez no visualizados– que se superponían y se complementaban, usando en ocasiones diversos proveedores para ser servidos al usuario a medida que internet se sofisticaba. Como si hubiéramos querido medir una televisión en la que cada parte de los contenidos en pantalla tuviera un origen y configuración técnica distinta, en frecuencias diferentes y en ocasiones opacas para el audímetro, es decir, como si ahora quisiéramos medir

EN LOS MEDIOS DIGITALES ACTUALES, CAPACES DE LLEGAR A CIFRAS DE AUDIENCIA IMPENSABLES PARA LOS ANTIGUOS SOPORTES EN PAPEL U ONDA HERCIANA, NO EXISTE UNA AUDIENCIA MASIVA QUE COINCIDA EN ESPACIO Y TIEMPO EN UN CONTENIDO EN CONCRETO, SALVO EN MUY CONTADAS OCASIONES



con audímetros tradicionales todo lo que aparece en la pantalla de nuestra televisión conectada.

Una tercera diferencia, crucial para entender todo lo que ha venido después, fue la atomización del consumo de internet. La diversidad de contenidos y de sitios web era grande ya a comienzos de los 2000 y, año tras año, no hizo sino crecer exponencialmente. Y no se trataba de crecer por la vía de sitios web y contenidos divulgativos o informativos. No, casi cualquiera tenía siempre en mente la posibilidad de sacarle unos ingresos publicitarios a sus contenidos. Si ya de salida los medios o sitios web que medir, por su interés comercial, multiplicaban por cien los que había sido necesario medir en la televisión, en apenas un par de años comenzó a ser necesario multiplicarlos por diez mil.

Pero, si esta fragmentación ya era de por sí letal para los paradigmas tradicionales de las metodologías de

medición, la ruptura del concepto de audiencia masiva fue definitivo para forzar un cambio. Porque una cosa es un medio de masas y otra una audiencia masiva. En los medios digitales actuales, medios de masas capaces de llegar a cifras de audiencia impensables para los antiguos soportes en papel u onda herciana, no existe una audiencia masiva que coincida en espacio y tiempo en un contenido en concreto, salvo en muy contadas ocasiones, como puede ser un gran acontecimiento deportivo o social. Consumimos lo que queremos, cuando lo queremos consumir y desde donde lo queremos consumir. Que la plataforma que nos lo ofrece sea única ayuda a agregar audiencia alrededor de esa plataforma, pero no necesariamente lo consumiremos en el mismo momento ni desde el mismo dispositivo que el resto de esa audiencia masiva. Por tanto, los paneles, diseñados para medir eventos masivos desde la lógica de la estadística aplicada a

## A MEDIDA QUE LAS SOLUCIONES MESTIZAS DE PANELES Y ETIQUETAS FUERON DESARROLLÁNDOSE, EL PAPEL DE LAS MEDICIONES PANEL FUE SIENDO RECONDUCIDO HACIA LA CUALIFICACIÓN DE LO QUE LA MEDICIÓN TECNOLÓGICA DEVOLVÍA

comportamientos masivos como los de la televisión tradicional, encontraron pronto una seria limitación en ese nuevo consumo que, hoy día, es el consumo normal de la mayoría de la población... y sigue creciendo.

El siguiente elemento claramente diferencial con respecto a la televisión fue que, mientras en ella las cosas siempre se han movido a poca velocidad y con un entorno de medición aislado, en los dispositivos con acceso a internet las cosas eran bien distintas. Los audímetros de televisión se conectaban a dispositivos que no estaban abiertos a ninguna red de comunicaciones, por lo que simplemente recogían un uso unidireccional, mientras que los softwares de audimetría no tenían barreras de protección frente al uso cotidiano de una máquina que sufría todo tipo de cambios casi diariamente: nuevos softwares, nuevos hardwares, actualizaciones de software, problemas técnicos, antivirus, resets...

Si la vida de un audímetro de televisión siempre tuvo una componente azarosa, con los problemas de los panelistas poco disciplinados, demasiado conscientes de estar siendo medidos o simplemente escurridizos a la hora de dejarse medir, los audímetros digitales les superaron por mucho. La privacidad de los individuos ganó peso como un

factor relevante a la hora de gestionar esos paneles y, en el caso de los ordenadores propiedad de empresas, la audimetría colisionó directamente con las políticas de seguridad recomendadas.

Incluso, con el andar del tiempo y la evolución tecnológica, se descubrió que los audímetros iban dejando de medir cada vez más cosas, a medida que el material que recogían –las urls de los contenidos consumidos– dejaba de ser propio de un determinado site y se sustituía por urls aparentemente pertenecientes a las grandes plataformas tecnológicas –Google, Apple, Facebook–. Todo ello, claro está, a cambio de acelerar o expandir el alcance del contenido en cuestión. Como resultado, buena parte del consumo de contenidos aparecía como contenidos pertenecientes a esos grandes operadores, cuando en la práctica no eran más que soportes técnicos para facilitar el consumo de contenidos ajenos.

Y todo ello, como no podía ser de otra manera, aderezado con una componente económica que pronto se demostró preocupante. Las rondas de inversores para iniciativas digitales no duraron eternamente con la alegría de su comienzo: los individuos no se unieron masivamente a los incipientes medios digitales –ni mucho menos

pagaron las suscripciones que por entonces se pusieron en marcha–, y las campañas publicitarias en internet –atadas a los adservers, su herramienta tecnológica de gestión, siempre propiedad de alguien– tampoco fueron la panacea, sobre todo en un país como España, con la televisión más barata de Europa gracias al recientemente ¿fenecido? GRP.

En resumen, internet se volvió mayoritariamente gratuito y eso planteó un serio problema de viabilidad para la medición vía paneles: ni eran suficientemente grandes para medir todo lo que se les pedía, ni podían serlo dada la falta de interés en pagarlos y las propias limitaciones de la estadística en un terreno de audiencias atomizadas, ni podían medir todo lo que estaba siendo consumido, sobre todo lo consumido a través de las grandes plataformas tecnológicas estadounidenses.

Las soluciones que se plantearon a ese estado de cosas tomaron un camino de mestizaje que, con distintos nombres –hibridación, unificación– dependiendo de la empresa que las puso en marcha, apuntaban a lo mismo: utilizar los datos procedentes de las herramientas tecnológicas para apuntalar las carencias de los paneles de individuos con audímetros. En esa carrera se vieron envueltas dos de las empresas de medición iniciales, Nielsen Netratings, que adquirió a Netvalue, y Comscore, que nació sobre la base de los paneles de la desaparecida Mediametrix.

### HIBRIDACIÓN

En paralelo al nacimiento y evolución de los paneles de medición de audiencias, la tecnología de medición de eventos técnicos dentro

de los contenidos digitales había evolucionado también. Las palabras estrella eran etiqueta y, a poca distancia, *cookie*. Las etiquetas –tags en inglés– son líneas de código que se añaden a la programación de los contenidos digitales y permiten que los servidores del dueño de esos códigos reciban una alerta –una llamada– cuando esos contenidos son consumidos, obteniendo así una medición técnica y no humana del número de veces que un contenido es consumido por algún dispositivo. Más aún, mediante el añadido de una instrucción de descarga de cookie a dicha etiqueta, el dispositivo que ha consumido ese contenido será marcado e individualizado con una cookie frente a otros dispositivos, permitiendo hacer un seguimiento de sus sucesivos consumos en ese u otros proveedores de contenidos.

Esa información técnica, que recogía todo el consumo de contenidos desde una óptica tecnológica y censal –y no muestral– permitió elaborar metodologías combinadas que tomaron datos de paneles y de etiquetas para construir una visión del consumo de contenidos más completa, evitando tanto las limitaciones estadísticas como las tecnológicas de los audímetros: donde el alcance del audímetro o el tamaño muestral no llegaban, podía obtenerse un dato, censal y no humano, sobre el que se podía construir una hipótesis razonable acerca de cuántos humanos y de qué tipo había detrás.

A finales de la década del 2000, las principales compañías de medición de audiencia digital operaban ya metodologías de medición combinadas –hibridadas o unificadas– con una componente de panel

DE ALGÚN MODO,  
LO QUE HABÍA  
COMENZADO  
COMO UN USO DE  
LA TECNOLOGÍA  
PARA APUNTALAR  
LOS PANELES, HA  
ACABADO SIENDO  
UN USO DE LOS  
PANELES PARA  
CUALIFICAR LA  
TECNOLOGÍA



procedente de los paneles de usuarios ya clásicos, y una componente tecnológica, con datos censales procedentes de las etiquetas. Justo a continuación, estamos hablando ya de los comienzos de la década de 2010, llegó la revolución móvil digital: todos pasamos de llevar en el bolsillo un teléfono móvil digital con una pequeña pantalla a tener una pantalla táctil a la que llamábamos teléfono móvil –para desconcierto de los más jóvenes de la casa, que poco saben de teléfonos fijos–.

Durante los primeros veinte años de los medios digitales, de mediados de los 90 del siglo XX a mediados de la década de 2010, el desarrollo y potencial del soporte digital estuvo limitado enormemente por la necesidad de recurrir a dispositivos poco amigables para el común de los mortales: los ordenadores. Demasiada tecnología para que un ciudadano de a pie se moviera con soltura por la oferta de contenidos y servicios digitales, así que para conseguir la definitiva explosión de la digitalización

de nuestras vidas y consumos, el punto clave fue la posibilidad de transferir a nuestro teléfono móvil inteligente buena parte de lo que hacíamos con una pantalla, un teclado, una CPU y un módem.

El efecto sobre el volumen y diversidad del consumo de contenidos y servicios móviles fue inmediato, con un crecimiento imparable que, por ejemplo, migró aceleradamente buena parte de nuestra vida desde nuestra cartera de piel a nuestro teléfono móvil. Para la medición de audiencias se abrió un reto adicional a los ya comentados: ¿era posible medir audiencia en el entorno y, sobre todo, medir el consumo de un mismo individuo desde distintos dispositivos?

Como casi siempre en las cosas digitales, las primeras experiencias al respecto se desarrollaron en Estados Unidos y las dos compañías dedicadas a medir audiencia allí en aquel momento, Nielsen Netratings y Comscore, comprobaron inmediatamente la dificultad del reto.

Si bien se desarrollaron rápidamente aplicaciones que, instaladas en los dispositivos móviles, permitieron medir audiencia del mismo modo que en los ordenadores, aunque con sus mismas limitaciones –visibilidad limitada de los consumos realizados, dificultad de mantenimiento en el tiempo del software instalado y atomización de audiencias–, la medición de dispositivos móviles añadió sus propias complejidades a las ya conocidas: existencia de al menos dos sistemas operativos móviles diferentes y relevantes –iOS de Apple y Android de Google– que requerían soluciones técnicas distintas y una mucha mayor conciencia de privacidad entre sus usuarios. Ya no se trataba de que el audímetro hubiera entrado en el salón comedor del individuo o incluso en su mesa de trabajo... es que se le pedía que ahora entrase en su bolsillo y en sus momentos de mayor individualidad e intimidad.

Las experiencias de ambas compañías estadounidenses llevaron a las mismas conclusiones: el reclutamiento y

LA WFA LO DESCRIBIÓ CON CLARIDAD EN SUS PRINCIPIOS PARA UNA NUEVA MEDICIÓN: LOS ANUNCIANTES MUNDIALES SIGUEN NECESITANDO, INCLUSO EN ESTOS TIEMPOS PROGRAMÁTICOS, QUE SE MIDAN AUDIENCIAS Y QUE SE MIDAN CON TRANSPARENCIA, NEUTRALIDAD Y RIGOR TÉCNICO, VENGAN DE DONDE VENGAN LOS COMPONENTES DE DICHAS MEDICIONES

mantenimiento de usuarios de dispositivos móviles era tan o más complejo y caro como el reclutamiento de hogares para los paneles de televisión, soportaba las mismas o mayores limitaciones en la recogida de datos que los audímetros de los ordenadores y, además, complicaba y encarecía la construcción del panel definitivo al añadir una capa de exigencia muestral adicional, consistente en mantener muestras con medición conjunta en ordenador y móvil para todas las cuotas establecidas en el diseño muestral en un entorno de alta volatilidad de los panelistas como el que producía la combinación de cambio tecnológico y concienciación de privacidad.

Se comprobó que los paneles multidispositivo implicaban unos enormes costes de reclutamiento y mantenimiento, cuando se percibía ya que su capacidad de capturar el consumo de los individuos iba a ser cada vez menor a medida que ganaba peso el uso de plataformas tecnológicas externas o de contenidos servidos desde entornos protegidos, situaciones en las que los audímetros dejaban de ser útiles. Con todo ello, la reacción de ambas compañías fue buscar soluciones alternativas que no requirieran la construcción de esos paneles, en ambos casos por la vía de buscar la conexión del consumo en distintos dispositivos a través de la información tecnológica recogida con etiquetados.

Sin embargo, este pragmatismo estadounidense no ha sido completamente adoptado en Europa, donde, pese a los antecedentes, al menos dos países –Reino Unido y España– apostaron en los últimos años por proyectos de medición basados en esos paneles multidispositivo descartados por las compañías pioneras. Con diferentes tiempos y suertes, en ambos casos se han repetido los aprendizajes ya sabidos: la tremenda dificultad y coste de mantener una muestra de ese tipo, si es que se desea hacerlo bien, y su imposibilidad para recoger buena parte del consumo actual de contenidos digitales vehiculados a través de plataformas tecnológicas que ciegan la visibilidad de los audímetros y derivan en datos de audiencia incompletos y menores de lo esperado... a un coste que suele ser poco sostenible en el tiempo.

La necesidad de integrar los datos procedentes de una metodología panel y de una tecnología de etiquetado tuvo un efecto positivo adicional, en ocasiones buscado y en otras simplemente encontrado. Al alinear los datos internos de las empresas medidas con algunos de los datos medidos externamente, la información censal de gestión técnica de un medio digital y su perspectiva censal medida con etiquetas del medidor de audiencia acabaron siendo prácticamente iguales, lo cual ahorró muchas discusiones... a costa de un segundo efecto no tan balsámico: por primera vez en



medición de audiencia, la colaboración o participación de los medios medidos era relevante y prácticamente imprescindible para que se realizase la medición, impactando esta nueva realidad sobre el paradigma tradicional de la medición de audiencia como elemento externo, neutral y ajeno a la actividad de los medios medidos.

Este nuevo estado de cosas ha abierto la puerta a infinidad de discusiones acerca de la bondad de los datos y no era más que un anticipo de la discusión general que ahora nos debe preocupar más: en este momento de consumos de contenidos en entornos cerrados, si queremos medición es necesaria la integración de datos procedentes de los medios y plataformas medidos, luego el papel del medidor debe evolucionar desde ser un proveedor neutral de soluciones de medición completas y cerradas hacia ser el muñidor neutral de entramados de datos que integren a todos los operadores posibles, con el fin de completar la visión global del consumo de los individuos.

Y, en ese nuevo rol de los medidores como muñidores de entramados de datos, ¿qué papel les espera a los paneles en este nuevo entorno?

A medida que las soluciones mestizas de paneles y etiquetas fueron desarrollándose, el papel de las mediciones panel fue siendo reconducido hacia la cualificación de lo que la medición tecnológica devolvía. Es decir: cuantificación de audiencias



basada en datos tecnológicos –censales– pero con un proceso de cualificación –de humanización y de perfilado demográfico– basado en los paneles, porque, en definitiva, lo que se busca al hablar de medir audiencias es medir individuos. De algún modo, lo que había comenzado como un uso de la tecnología para apuntalar los paneles, ha acabado siendo un uso de los paneles para cualificar la tecnología.

LA WFA (World Federation of Advertisers) lo describió con claridad en sus principios para una nueva medición: los anunciantes mundiales siguen necesitando, incluso en estos tiempos programáticos, que se midan audiencias y que se midan con transparencia, neutralidad y rigor técnico, vengan de donde vengan los componentes de dichas mediciones.

En ese entorno de necesidad de transparencia, neutralidad y rigor se mueven los paneles hoy día.

A medida que estamos construyendo sistemas de medición más completos y sofisticados, basados en la integración de datos procedentes de plataformas y operadores de contenidos, el rol de los paneles está dejando de ser el de proveer el dato final de audiencia para pasar a ser el elemento de cualificación de las audiencias de todos ellos y, cada vez con más relevancia, ser la fuente de verdad en esas integraciones, la mayoría de ellas basadas en algoritmos que necesitan tener una cierta guía extraída de la realidad.

Ese es el nuevo rol de los paneles en la medición de audiencia: ser los lazarillos de tanto algoritmo como ahora estamos implementando, todos ellos inteligentes pero ciegos a la realidad que les rodea. Una realidad que, aunque sea de un modo incompleto y con trazos poco definidos, los paneles siguen siendo capaces de reflejar. ■

Una encuesta, señalan los autores de este artículo, no deja de ser una fotografía en 4K de un momento concreto que contribuye a aumentar la calidad de la secuencia pixelada que ofrece el *big data* y, por tanto, contar mejores historias que ayuden a tomar decisiones más acertadas. Es decir, ambas metodologías están condenadas a entenderse y complementarse.



# EN BUSCA DE LA MEJOR FOTO

Sales desde Sarria hasta Santiago, y aunque se pueda pensar que al llegar a la Catedral ya está, puedes seguir caminando hasta Finisterre, y aún así, no, no es el final del camino. Esta andadura se lleva realizando cientos de años por miles de personas y así seguirá siendo en un futuro. El trayecto se va adaptando a una nueva era, y por mucho tiempo que pase o nuevas rutas que surjan, no desaparecerá. ¿Por qué? Porque ofrece una experiencia que ninguna otra senda puede ofrecer.

de tamaño 'reducido'... En esta línea reflexionaba Eduardo Madinaveitia (alguien a quien no necesitamos presentar en este foro) en un artículo en su blog: "la fascinación por las grandes bases de datos desestructuradas y los programas que permiten tratarlas y extraer conclusiones a partir de muestras de las que se sabe muy poco, ha provocado la pérdida de prestigio de los estudios basados en muestras representativas de tamaño reducido".



**Mercedes Fernández**  
Directora de proyectos en 40dB



**Daniel Núñez**  
Analista de datos en 40dB

Algo similar ocurre con la investigación de mercado, social, política (nunca tan de actualidad como con el vigente presidente del CIS)... El rápido avance de la tecnología ha supuesto nuevas formas de recopilar datos, muchos datos, de obtenerlos de manera muy rápida y con costes muy bajos, algo que, desde hace tiempo, parece amenazar las muestras representativas

Sin embargo, estamos convencidos de que todos estos avances en el tratamiento de datos no suponen el fin de la investigación de mercados, porque, igual que el Camino, esta ofrece algo diferencial y exclusivo.

Lo importante es no caer en una pelea entre 'muchos datos' y 'tamaño reducido', sino entender que se



## LA IMPORTANCIA (Y LA NECESIDAD) DE COMPLEMENTAR LAS ENCUESTAS DE MUESTRAS REPRESENTATIVAS CON EL ANÁLISIS DE DATOS MASIVOS (*BIG DATA*) VIENE DETERMINADA PORQUE LAS FORTALEZAS DE UNA METODOLOGÍA SON LAS DEBILIDADES DE LA OTRA, Y VICEVERSA

trata de metodologías con objetivos diferentes que pueden y deben entenderse como complementarias. No debemos interpretar las muestras representativas de tamaño reducido como algo de un pasado en el que no era posible acceder y manejar grandes volúmenes de datos. Debemos poner el acento en algo de lo que también nos alerta Eduardo: en muchas ocasiones sabemos poco de esas grandes muestras.

En nuestro caso, nos hemos encontrado con proyectos concretos en los que la encuesta a una muestra representativa ha sido imprescindible para corregir importantes sesgos en las encuestas a la base de datos del cliente... Tu olfato de investigador te dice que los resultados que arroja la encuesta enviada por *mail* a la base de datos del cliente son en algunos aspectos contraintuitivos y te enfrentas al reto de validar (o no) la hipótesis de partida. Aunque no sin esfuerzo, convencimos al cliente de que nos dejara realizar la misma encuesta a una

muestra representativa de la población general donde encontraríamos a clientes de su marca. Las principales resistencias del cliente tenían que ver con dos aspectos: el sentido que parecían haber pedido las encuestas a muestras representativas (para qué, si tengo una base de datos con miles de registros...) y el mayor coste de estas encuestas frente a preguntar a tu propia base de datos. La encuesta dentro de una muestra representativa nos reveló importantes diferencias con los resultados obtenidos en la base de datos en variables sociodemográficas básicas (sexo, edad, nivel de estudios...) y en otras de mayor calado que no revelaremos porque permitirían identificar al cliente.

Este ejemplo, reciente y real, matiza lo que oímos en un I+A Talks entre Kiko Llaneras y Narciso Michavila acerca de que se podían extrapolar los resultados de encuestas hechas a la base de datos de cliente poniendo en valor las sinergias entre ambas metodologías.

# NI EL ANÁLISIS DE DATOS MASIVOS ES EL SANTO GRIAL QUE HA VENIDO A DESTERRAR LAS ENCUESTAS, NI ESTAS PUEDEN DEJAR DE EVOLUCIONAR

## FORTALEZAS Y DEBILIDADES

La importancia (y la necesidad) de complementar las encuestas de muestras representativas con el análisis de datos masivos (*big data*) viene determinada porque las fortalezas de una metodología son las debilidades de la otra, y viceversa. Entre las ventajas del análisis de *big data* encontramos la gran cantidad de datos que nos proporcionan, el análisis en tiempo real o la flexibilidad a la hora de aplicar distintas técnicas de investigación. En contra, distintos aspectos relacionados con la calidad de los datos, que pueden ser incompletos, estar sesgados o, en algunos casos, incluso ser falsos. Aspectos, todos ellos, que nos llevarían a unos resultados alejados de la realidad que podrían hacernos tomar decisiones equivocadas. En muchas ocasiones, la falta de contexto de los datos o la información incompleta sobre el tema objeto del estudio dificulta e

incluso impide su análisis. Y algo que debemos de tener siempre presente, evitar como hemos hecho siempre con las encuestas (código ESOMAR), el “mal uso” de los datos personales que puede haber dentro de ese gran conjunto de datos (recordemos el escándalo de Cambridge Analytica y Facebook).

Estas debilidades del *big data* pueden ser compensadas con información obtenida en encuestas a muestras representativas de nuestro público objetivo. Como en cualquier metodología, si cuidamos todos los pasos (diseño muestral, cuestionario, controles de calidad en el trabajo de campo, etcétera) los datos declarativos son también ‘*data*’ de una gran calidad con los que podemos controlar posibles sesgos, reduciendo errores y aumentando la precisión y el rigor del análisis. En cualquier caso, siempre proporcionan y/o amplían el contexto de análisis abordando nuevas aristas del problema. Entre las ventajas de las encuestas encontramos también los estudios longitudinales que nos permiten ver evoluciones y tendencias. En su contra, los principales problemas de las encuestas residen en el coste ‘extra’ que suponen y en los tiempos que, aunque son cada vez menores, no pueden competir con los datos ‘en tiempo real’.

Las encuestas y el *big data* están destinados a entenderse y complementarse. Ni el análisis de datos masivos es el Santo Grial que ha venido a desterrar las encuestas, ni estas pueden dejar de evolucionar. Como en otras amenazas que hemos vivido en

la profesión de investigador, debemos abandonar posiciones miedosas y/o defensivas para seguir evolucionando (solo hay que recordar que hace 20 años las encuestas se hacían en la calle con papel y un perforador) y pensar cómo esa gran cantidad de datos puede complementar nuestro trabajo. Por poner solo un ejemplo, en nuestro caso hemos llevado a cabo proyectos en los que el análisis de lo que se estaba diciendo en redes sociales sobre un tema ha sido un excelente punto de partida que nos ha permitido, además, recoger en un cuestionario aspectos que quizás no hubiésemos incluido.

Como profesionales de la investigación nos dedicamos a contar historias a partir de datos y para hacerlo tenemos que cuidar la calidad de nuestra materia prima. Una encuesta no deja de ser una fotografía en 4K de un momento concreto y el *big data* son un montón de fotografías secuenciales pixeladas (se pueden ver tendencias, pero no graduarlas). Con esa foto en 4k seremos más capaces de aumentar la calidad de esa secuencia pixelada y poder contar mejores historias que ayuden a tomar mejores decisiones. ■

# LA ARROBA DE I+A

El encuentro de referencia de los investigadores sociales y data scientists españoles.

**1 DE JUNIO DE 2023**

**09:00 – 16:30 Espacio COAM**  
Calle de Hortaleza, 63. 28004, Madrid

**Una jornada para descubrir, discutir y disfrutar**

**Toda la info en:**  
**[www.ia-espana.org](http://www.ia-espana.org)**

# MUESTRAS REPRESENTATIVAS EN LA ERA DIGITAL

En un contexto en el que la posibilidad de acceder a los datos ha aumentado vertiginosamente debido a la digitalización, la utilización de muestras representativas sigue siendo fundamental porque no siempre la información ofrecida por esos datos representa al conjunto de la población. Así se explica en este artículo.

**El concepto de muestra representativa** proviene de la estadística y la investigación de mercados. Desde hace mucho tiempo, los investigadores han utilizado muestras de una población más grande para hacer inferencias sobre dicha población a la que era imposible acceder en su totalidad.

Sin embargo, en la nueva era que estamos viviendo parece que la accesibilidad a los datos ha dejado de ser un problema. El avance exponencial de la tecnología ha hecho posible conectarse con prácticamente cualquier persona en cualquier parte del mundo. Y esto, de alguna manera, ha hecho que hablemos menos de las muestras representativas.

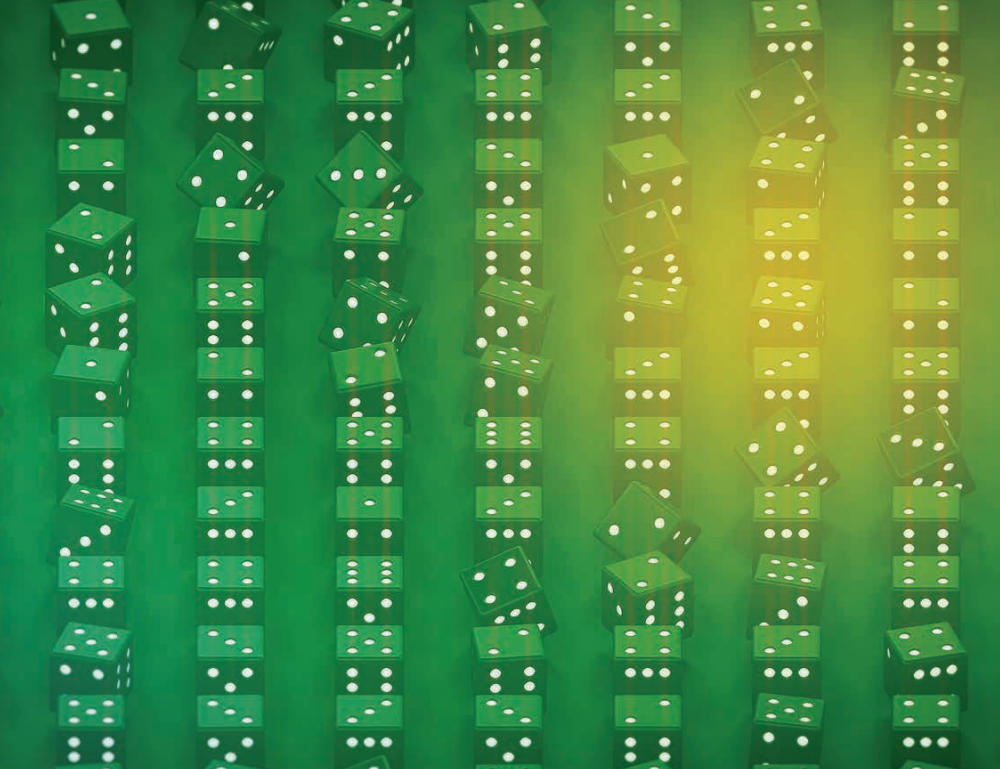
Pero hay que tener cuidado. Observar millones de personas, por ejemplo, en una red social como Twitter

en España, no es necesariamente representativo de lo que haga o diga la sociedad española en su totalidad. Por dos razones. La primera, que el comportamiento en una red social puede estar sesgado hacia la manera de expresarse en ella. Por ejemplo, en Twitter mucha gente expresa más la negatividad. Por eso, mirar lo que se dice en Twitter no es mirar lo que se dice en España. La segunda, que las personas que se expresan en la red social no se distribuyen necesariamente como lo hace la población española, en el sentido de porcentaje de hombres vs mujeres, grupos de edad o clase social.

En el caso de los datos de navegación de una página web, para obtener una muestra representativa deberemos tener claro qué universo es el que queremos representar. De la misma manera que sucede con las redes sociales, será difícil representar el universo España, salvo que sea una página en la que se entre a través de login y de la que tengamos datos personales de los navegantes. ¿Qué se hace habitualmente? Se realiza lo que se conoce como muestreo de usuarios activos, que consiste en



.....  
**Macarena Estévez**  
Fundadora del Círculo de  
Ingenio Analítico  
\_\_\_\_\_



## EL AVANCE EXPONENCIAL DE LA TECNOLOGÍA HA HECHO POSIBLE CONECTARSE CON PRÁCTICAMENTE CUALQUIER PERSONA EN CUALQUIER PARTE DEL MUNDO. Y ESTO, DE ALGUNA MANERA, HA HECHO QUE HABLEMOS MENOS DE LAS MUESTRAS REPRESENTATIVAS

seleccionar una muestra aleatoria de usuarios que visitan el sitio web en un momento dado. En este caso, el universo no será España, sino “gente que visita la página web”. Lo que estamos haciendo es representar a todos los que navegan por la página web con una muestra que debe ser lo suficientemente grande y representativa para poder obtener conclusiones válidas y significativas sobre el comportamiento general de los usuarios en la página web.

En uno y otro caso, será además importante considerar la privacidad de los usuarios y cumplir con las regulaciones de protección de datos en la recopilación y uso de los datos de navegación.

Otro ejemplo de muestra representativa podría ser la obtención de datos a través de encuestas online. En este caso, es importante tener en cuenta que no todos los usuarios de internet están dispuestos a participar en encuestas *online*, y esto mismo puede ser un sesgo a la hora de elegir

una muestra representativa. Aunque tengamos una muestra que parezca similar a la población española en cuanto a edad, género y clase social, podría darse el caso de que las personas elegidas, que deberán ser personas con capacidad de responder una encuesta *online*, fueran de alguna manera diferentes. Pensemos, por ejemplo, que fuesen personas que no estén a favor del avance tecnológico y que la encuesta pudiera estar afectada por esto.

Por lo tanto, cuando queramos investigar una población, saber lo que piensa, cómo decide, o por qué hace lo que hace, o bien tenemos acceso al total de la población (censo), o bien necesitamos elegir una muestra representativa, es decir, un grupo de individuos seleccionados de manera aleatoria de una población determinada, con el fin de que representen a la población de manera adecuada, evitando además posibles sesgos según lo que sea que se está investigando. Si la muestra no es

# CUANDO QUERAMOS INVESTIGAR UNA POBLACIÓN, SABER LO QUE PIENSA, CÓMO DECIDE, O POR QUÉ HACE LO QUE HACE, O BIEN TENEMOS ACCESO AL TOTAL DE LA POBLACIÓN (CENSO), O BIEN NECESITAMOS ELEGIR UNA MUESTRA REPRESENTATIVA

representativa, los resultados obtenidos pueden ser sesgados y no reflejar la realidad de la población.

## PROCESO DE SELECCIÓN

Para obtener una muestra representativa es importante seguir un proceso riguroso de selección. Una forma de hacerlo es utilizar el muestreo aleatorio simple, que consiste en seleccionar individuos al azar de la población. Esto asegura que todos los individuos tengan la misma probabilidad de ser seleccionados y, por lo tanto, que la muestra sea representativa. Por ejemplo, si queremos realizar una encuesta a la población española para conocer su opinión sobre el uso de energías renovables, podemos seleccionar al azar un número determinado de personas de todas las provincias del país, asegurando que todos tengan la misma probabilidad de ser seleccionados.

Otro modelo es el muestreo estratificado, que se utiliza cuando se sabe que la población se divide en subgrupos con características distintas. En este caso, se selecciona una muestra aleatoria de cada subgrupo, de manera que la muestra final incluya individuos de cada uno de ellos en proporción a su tamaño en la población total. En

el ejemplo anterior, podemos pensar que no va a ser lo mismo la opinión que tengan los más jóvenes que los mayores, o los hombres que las mujeres. Seleccionaremos una muestra aleatoria de cada grupo, de manera que la muestra final incluya individuos de cada subgrupo en proporción a su tamaño en la población total. De esta manera, tendremos una muestra representativa de la población española que refleje las diferentes opiniones según la edad y el género.

También se puede utilizar el muestreo por conglomerados, que consiste en dividir la población en grupos más grandes llamados conglomerados, y seleccionar al azar algunos de estos para obtener una muestra representativa. Dentro de cada conglomerado seleccionado se puede utilizar el muestreo aleatorio simple o el muestreo estratificado para seleccionar a los individuos que formarán parte de la muestra. Por ejemplo, en el caso de España podemos dividir primero la población en grupos más grandes como pueden ser las diferentes comunidades autónomas. Luego, seleccionamos al azar algunas de estas comunidades autónomas para obtener una muestra representativa. Dentro de cada comunidad autónoma seleccionada, podemos utilizar el muestreo aleatorio simple o el muestreo estratificado para seleccionar a los individuos que formarán parte de la muestra.

La elección de la técnica de muestreo adecuada dependerá del objetivo específico de la investigación y de la complejidad de la población a estudiar.





AUNQUE LA  
ACCESIBILIDAD  
A LOS DATOS  
PARECE HABER  
AUMENTADO, ES  
IMPORTANTE  
TENER EN CUENTA  
QUE OBSERVAR  
MILLONES DE  
PERSONAS EN UNA  
RED SOCIAL NO ES  
NECESARIAMENTE  
REPRESENTATIVO  
DE LO QUE PIENSA  
O HACE LA  
SOCIEDAD EN SU  
TOTALIDAD

En algunos casos, la población puede estar dividida en múltiples subgrupos con características muy diferentes, por lo que el muestreo estratificado será la técnica más adecuada para obtener una muestra representativa. En otros casos, la población puede estar dividida en conglomerados, como en el ejemplo de las comunidades autónomas en España, por lo que el muestreo por conglomerados será la técnica más adecuada. Por último, podrá haber casos sin divisiones, donde el muestreo aleatorio sea el mejor candidato.

rigurosa. En España, se pueden utilizar diferentes técnicas de muestreo como el muestreo aleatorio simple, el muestreo estratificado o el muestreo por conglomerados, dependiendo de las características de la población a investigar. ■

Resumiendo, la utilización de muestras representativas sigue siendo fundamental en la era digital en la que vivimos. Aunque la accesibilidad a los datos parece haber aumentado, es importante tener en cuenta que observar millones de personas en una red social no es necesariamente representativo de lo que piensa o hace la sociedad en su totalidad. Por lo tanto, cuando se quiere investigar una población, es necesario elegir una muestra representativa de manera

En un número dedicado a las muestras representativas, este artículo tiene por objeto desentrañar las ventajas, pero también las barreras que presenta el neuromarketing, una metodología que dista mucho de emplear grandes muestras.



# LA MUESTRA Y LA REVOLUCIÓN TECNOLÓGICA: MENCIÓN AL NEUROMARKETING

La muestra en la investigación es tan importante como lo es el agua para que las personas podamos sobrevivir. Toda investigación se realiza en torno al análisis de una muestra extraída de una población. En el contexto de la investigación, por población nos referimos a una colección finita o infinita de unidades (individuos o elementos) de las cuales se desea obtener una información, mientras que por muestra nos referimos a una parte de las unidades que forma la población.

La investigación trata de dar respuesta a un objetivo, buscando explicar y

entender un fenómeno que afecta a una determinada población a través de la muestra. Si queremos hacer una investigación sobre el voto político en España, lo ideal sería ver la cantidad de personas que pueden votar e ir preguntando una a una la información que se considere. Sin embargo, esto es muy complicado, ya que el coste económico y de tiempo es demasiado alto. Por ello, se escoge una muestra representativa de toda esa población para, posteriormente, poder inferir esas conclusiones. En definitiva, es crucial darle la importancia que merece al tratamiento de la muestra.



Álvaro Saavedra Montejo  
Profesor ayudante en la  
Universidad Pública de  
Navarra



Las tecnologías de la información y comunicación han transformado por completo la forma en que se lleva a cabo la investigación, revolucionando la recopilación, análisis y visualización de datos. Gracias a ellas, los investigadores tienen ahora acceso ilimitado a una cantidad masiva de información relevante para sus estudios, y pueden analizar grandes cantidades de datos en tiempo récord, obteniendo resultados precisos y veloces. La precisión de la recopilación de datos también ha sido mejorada de manera significativa, especialmente en estudios online o a través de dispositivos electrónicos. Además, la colaboración entre investigadores de diferentes partes del mundo se ha visto impulsada y ampliada, aumentando la eficiencia y efectividad de la investigación. Por todo esto, la visualización de los resultados de la investigación se ha optimizado, facilitando la interpretación y comunicación de los hallazgos. En resumen, las tecnologías de la información y comunicación han

revolucionado la forma en que se lleva a cabo la investigación, convirtiéndola en un proceso más eficiente, preciso y colaborativo.

Si bien los avances tecnológicos han revolucionado la recolección de datos, permitiéndonos acceder a una cantidad masiva de información nunca antes imaginada, es importante destacar cómo estos avances afectan de forma particular a la disciplina del neuromarketing. No solo es evidente que hay una mejora esencial en la recolección de datos, sino que se ha abierto un mundo de posibilidades. Sin embargo, para obtener datos en la disciplina del neuromarketing es necesario disponer de un espacio adecuado (laboratorio) donde podamos reunir a la muestra y recoger los datos de los diferentes estímulos que se han medido mediante dispositivos específicos. Gracias al avance tecnológico, los dispositivos para medir la actividad cerebral (como el encefalograma), los movimientos oculares (como el *eye-tracker*) y la respuesta emocional (como el anillo galvánico), se han vuelto más accesibles y han mejorado su capacidad para recopilar datos en pequeños intervalos temporales. Además, estos dispositivos son ahora más fáciles de usar para investigadores que no son especialistas en neurociencia.

Como es sabido, el neuromarketing es una interdisciplina reciente donde se combinan diferentes ámbitos, como el marketing, la psicología y la neurociencia para entender el comportamiento del consumidor (Ramsøy, 2015). Asimismo, el objetivo

## LA FUSIÓN DEL NEUROMARKETING CON LAS TÉCNICAS TRADICIONALES DE INVESTIGACIÓN DE MERCADOS PUEDE REVOLUCIONAR LA FORMA EN QUE LAS EMPRESAS COMPRENDAN LAS MOTIVACIONES DE SUS CLIENTES



principal del neuromarketing es entender cómo las emociones, las motivaciones y los procesos cognitivos influyen en las decisiones de los consumidores y cómo las empresas pueden aprovechar esta información para mejorar su estrategias. Para ello, es necesario disponer no solo de una instalación como se ha comentado anteriormente, sino también de una serie de dispositivos específicos para recoger los cambios que se producen en las variables corporales a través de los dispositivos mencionados anteriormente.

Aunque el coste de los dispositivos a través de los cuales se recoge la actividad psicofisiológica es 'accesible' (equipar un laboratorio ronda actualmente entre 15.000 y 25.000 euros), el precio de las herramientas neurocientíficas resulta ser un desafío mayor, con tarifas que oscilan entre 100 y 600 euros por hora de escaneo, debido a la necesidad de contratar a expertos en la materia (Casado-Aranda, L, 2021). A esto, le debemos sumar el coste que supone captar a la muestra, que varía en función

de las características específicas. En general, con cada participante se suele estar en torno a 45-60 minutos, ya que hay un proceso de calibrado de los dispositivos, de los umbrales de la actividad, de medición de los diferentes estímulos y, normalmente, el experimento se complementa con una encuesta. No obstante, la utilización de mediciones repetidas y la propia naturaleza de que los datos son objetivos y precisos, reduce la cantidad de sujetos requeridos para cada estudio. Según un análisis científico, para que la muestra y los resultados sean representativos es necesario contar con alrededor de 30 sujetos experimentales (Bazzani et al., 2020).

En resumen, podríamos decir que las principales barreras de la muestra en el neuromarketing serían: el coste de las técnicas, la dificultad/coste para reclutar participantes y el coste de los dispositivos. Además de las implicaciones éticas sobre la privacidad de los datos y la interpretación de los resultados, así como las leyes y regulaciones que rigen su uso. Sin embargo, el neuromarketing sigue siendo una herramienta valiosa para comprender mejor al consumidor y mejorar las estrategias de marketing. Por lo que, como principales ventajas de la muestra encontramos que los datos obtenidos no son declarados por el usuario, sino que se obtienen directamente de las variables fisiológicas y que, además, dependiendo del dispositivo se pueden obtener datos en milisegundos (datos continuos), lo que genera finalmente una cantidad de información significativamente

# EN LA DISCIPLINA DEL NEUROMARKETING, AUNQUE LA CAPACIDAD DE LOS DISPOSITIVOS PARA RECOGER DATOS HA MEJORADO, SU COSTE SIGUE SIENDO UN GRAN DESAFÍO, JUNTO CON EL PRECIO DE LAS HERRAMIENTAS NEUROCIENTÍFICAS

mayor que un estudio de mercado donde generalmente solo se utiliza la encuesta. Precisamente, esta gran cantidad de datos es una de las justificaciones de porqué la muestra para estos estudios es más pequeña.

Independientemente de las características propias de la muestra en el neuromarketing, los investigadores sugieren que estos estudios deben ir acompañados de otras técnicas de investigación. La fusión del neuromarketing con las técnicas tradicionales de investigación de mercados puede revolucionar la forma en que las empresas comprenden las motivaciones de sus clientes. El neuromarketing permite medir las respuestas emocionales y cognitivas de los consumidores, mientras que las técnicas tradicionales proporcionan información sobre las características demográficas y de comportamiento. Al combinar ambos enfoques se obtiene una comprensión más completa y precisa del consumidor, lo que puede llevar a estrategias de marketing más efectivas y a productos y servicios mejor adaptados a las necesidades del mercado (Gountas et al., 2019).

Como conclusión, la muestra es una parte fundamental en la investigación, ya que a través de ella se pueden inferir y estimar las características de una población. Aunque lo ideal sería tener acceso a toda la población, el coste económico y de tiempo resulta demasiado alto, por lo que se opta por utilizar una muestra representativa. Gracias a las tecnologías de la información y comunicación, la recopilación, análisis y visualización de

datos ha mejorado significativamente, haciendo que la investigación sea más eficiente, precisa y colaborativa. En la disciplina del neuromarketing, aunque la capacidad de los dispositivos para recoger datos ha mejorado, su coste sigue siendo un gran desafío, junto con el precio de las herramientas neurocientíficas. A pesar de ello, la naturaleza del dato extraído de cada participante reduce la cantidad de sujetos requeridos por estudio. En definitiva, la muestra sigue siendo esencial en la investigación, y aunque el avance tecnológico ha mejorado significativamente el proceso de recolección, análisis y visualización de datos, la disciplina del neuromarketing todavía presenta desafíos en cuanto al coste de las herramientas y la necesidad de contar con expertos en la materia. ■

## Referencias

- Bazzani, A., Ravaoli, S., Trieste, L., Faraguna, U., & Turchetti, G. (2020). Is EEG Suitable for Marketing Research? A Systematic Review. *Frontiers in Neuroscience*, 14. <https://www.frontiersin.org/articles/10.3389/fnins.2020.594566>
- Casado-Aranda, L. (2021). *Neurociencia del consumidor*. Piramide.
- Gountas, J., Gountas, S., Ciorciari, J., & Sharma, P. (2019). Looking beyond traditional measures of advertising impact: Using neuroscientific methods to evaluate social marketing messages. *Journal of Business Research*, 105, 121-135. <https://doi.org/10.1016/j.jbusres.2019.07.011>
- Ramsøy, T. Z. (2015). Introduction to Neuromarketing & Consumer Neuroscience. *Neurons Inc.* <https://research.cbs.dk/en/publications/introduction-to-neuromarketing-amp-consumer-neuroscience>

# EL TAMAÑO SIEMPRE IMPORTA

El presupuesto disponible es una de las variables más definitorias de cualquier proyecto que se emprenda. También, obviamente, de la utilización de una metodología u otra a la hora de llevar a cabo una investigación. Y, en el caso de las basadas en muestras representativas, también de su tamaño.

La base de cualquier estudio **demoscópico** es la muestra que se va a utilizar. De su mayor o menor calidad dependerá en gran medida que los resultados sean más o menos representativos.

Una vez que se ha decidido el sistema de muestreo o captación de entrevistados, cómo se van a procesar los datos y qué salidas e indicadores se van a producir, queda la gran incógnita de siempre: ¿cuántas entrevistas vamos a hacer, cuál es el tamaño de la muestra?

Dado que los recursos suelen ser finitos, y los monetarios mucho más, lo que se suele hacer es otra pregunta: ¿de cuánto presupuesto dispones?, para que, una vez restado el importe necesario para completar el estudio (grabación de las entrevistas, ponderación/elevación, tabulación y presentación de resultados), dividas el importe restante del presupuesto entre el coste por unidad de entrevista para obtener la muestra. El principio del que se parte es que hagas toda la muestra que te puedas permitir.

Se puede intentar hacer de una forma más científica. Cualquier estudio basado en una muestra está sujeto a

una incertidumbre que proviene de las diferentes posibilidades de seleccionar una muestra de esa población. Es lo que se conoce como error estándar o margen de error. Se puede seleccionar un tamaño de muestra adecuado para que el margen de error del estudio sea inferior a una cierta magnitud.

El error estándar se calcula como el cociente entre la desviación estándar y la raíz cuadrada de la muestra. Esto significa que si no estoy suficientemente contento con la precisión de mi estimación y pretendo mejorarla, doblar la muestra solo reducirá el error estándar dividiendo por raíz de 2, que es aproximadamente 1,414. Para reducir el margen de error a la mitad se deberá multiplicar la muestra por 4 (porque la raíz de 4 es 2).

Por otra parte, cada estimación, cada dato tendrá su propio margen de error que dependerá de la penetración que resulte sobre el universo (no es el mismo error el asociado a un dato que represente el 10% o el 20% del universo). Y por otro lado, si el dato que obtengo está referido a una parte de la muestra, por ejemplo, si en un estudio referido a toda España con 1.000 entrevistas y yo quiero calcular



José Andrés Gabardo  
Director técnico de AIMC



el dato de la comunidad de Madrid, tendré que calcular el margen de error sobre el porcentaje de Madrid pero con un tamaño de muestra referido solo a las entrevistas de Madrid.

De este modo, el margen de error deberá estar referido a la menor de las muestras que tengamos en alguno de los targets por los que vayamos a ventilar las tabulaciones.

En general, la regla del presupuesto de que disponemos es la más sensata para calcular el tamaño de la muestra.

Y tened en cuenta que una muestra de 1 en un *target* solo es representativa si el *target* se refiere a individuos únicos como a los Reyes de España en activo, porque incluso Reyes de España ahora hay dos y no te serviría la muestra. También sirve para representarte a ti mismo, que además es una muestra fácil y barata de obtener. En la mayoría de los casos, disponer de una muestra de uno es solo engañarte para tener un dato, pero ningún resultado que se obtenga con esta muestra será relevante ni significativo.

### ¿NO HAY OTRA SOLUCIÓN AL PROBLEMA DEL TAMAÑO DE LA MUESTRA?

Como norma general no existe ninguna. Pero sí que podemos utilizar la tecnología para encontrar otras vías de solución en casos concretos. La alternativa a los estudios demográficos es la realización de censos, es decir, contabilizar la totalidad del universo para evitar en gran medida los problemas del margen de error.

Es lo que se hace cuando un tema es muy importante o delicado: realizar un referéndum, no una encuesta, en el que se obtiene la opinión de todo el censo. Pero esta no es una solución que se pueda utilizar siempre. No podemos estar haciendo un referéndum cada vez que una empresa necesita conocer un dato.

Pero podemos tener otra solución. En este mundo digital en el que estamos siempre hay que preguntarse si lo que queremos medir deja alguna huella digital o es susceptible de generar una huella digital, que de forma ya

EN GENERAL,  
LA REGLA DEL  
PRESUPUESTO DE  
QUE DISPONEMOS  
ES LA MÁS SENSATA  
PARA CALCULAR  
EL TAMAÑO DE LA  
MUESTRA

EL MERCADO  
DEMOSCÓPICO  
TRADICIONAL NO  
CORRE RIESGO DE  
DESAPARECER,  
TODAVÍA HAY  
MUCHAS COSAS  
QUE NO SE PUEDEN  
TRANSFORMAR  
EN INFORMACIÓN  
DIGITAL DE  
MANERA  
AUTOMÁTICA,  
Y TENDREMOS  
QUE SEGUIR  
PREGUNTÁNDOLO  
A LA MANERA  
ACTUAL

censal sea capaz de ofrecernos un resultado.

Si yo utilizo una agenda de papel, seguramente podré organizarme el tiempo de la misma forma efectiva que si utilizo una agenda digital, pero además de poder programar la agenda digital para que me avise de mis compromisos con cierto margen, la disponibilidad de la información de mis planes en formato digital me permitiría realizar después un análisis de a qué dedico mi tiempo según esa agenda. Esto también se podría hacer con la agenda de papel, pero requeriría seguramente una cantidad de tiempo adicional que seguramente no tendría cabida en la agenda de papel.

Si lo que necesitamos medir deja huella digital o podemos hacer que lo deje, podríamos diseñar un procedimiento para obtener esa información, no en base a una muestra, sino de forma censal, es decir contabilizando todo el universo, sin estimaciones.

Si yo quiero saber cuántos pasos da un español al día, puedo preguntar a una muestra en cuánto estima sus desplazamientos diarios, estimar el tamaño de su zancada, y por cociente obtener su número de pasos. Si en lugar de esto dispongo de una aplicación que se pueda instalar en el teléfono móvil y que sea capaz de contabilizar los pasos que damos en el día, yo podré obtener una cifra mucho más ajustada a la realidad de aquellos que dispongan de mi aplicación en su móvil. Puede que nunca consiga que toda la población con móvil se instale

mi aplicación, pero si lo consiguiese sería una forma muy efectiva de medir.

El reto para los próximos años es tratar, siempre que se pueda, de digitalizar u obtener una información digitalizada de lo que queramos medir para no tener problemas de tamaño de muestra.

Pero tranquilos, que el mercado demoscópico tradicional no corre riesgo de desaparecer, todavía hay muchas cosas que no se pueden transformar en información digital de manera automática, y tendremos que seguir preguntándolo a la manera actual, espero.

Por eso seguiré diciendo que en el caso de las muestras **el tamaño sí importa.** ■





# *Mira todo lo que hemos preparado para ti*

En I+A nos movemos contigo. Hemos multiplicado nuestros eventos y actividades a lo largo del año para aprender de los mejores, compartir experiencias y crear puntos de contacto con la industria.

*¿Te vienes?*



Marzo 2023



Abril 2023



Junio 2023



Junio 2023



Septiembre 2023



Octubre 2023



Noviembre 2023



Tercer jueves de cada mes (excepto julio y agosto)

Toda la info en: [www.ia-espana.org](http://www.ia-espana.org)



# EL TAMAÑO MATÓ A LA REPRESENTATIVIDAD

Eduardo Madinaveitia es experto en investigación de medios y miembro del Comité Editorial de *Investigación & Marketing*

Durante casi cien años, los estudios basados en muestras representativas, de tamaño relativamente pequeño, han sido la base de las investigaciones sociológicas cuantitativas.

Cuando el pasado mes de diciembre murió Carlos Lamas, que siempre fue un gran defensor de este tipo de técnicas, planteé al Consejo Editorial de *I&M* que un número de este año se dedicara a analizar el papel de los estudios basados en muestras representativas: en el pasado, ahora y en el futuro, si creemos que pueden seguir siendo útiles para el conocimiento de la realidad social.

Siempre he sido bastante ingenuo y me imaginé un número con numerosas colaboraciones, tanto de personas que trabajaron con Carlos, como de otras que lo admiraran o, simplemente, que trabajan, han trabajado, o piensan seguir haciéndolo, con estas técnicas.

Pero me da la impresión de que no ha sido así, conseguir unos cuantos artículos ha sido laborioso. Más bien parece que se cumple la frase con la que titulé mi artículo de entonces: *En vías de extinción*.

La representatividad ganó una gran parte de su prestigio cuando en las elecciones presidenciales de 1936 en Estados Unidos, Gallup acertó que el demócrata Roosevelt sería el ganador con un estudio basado en solo 5.000 entrevistas. La prestigiosa revista *Literary Digest* había realizado una macroencuesta entre sus lectores en la que obtuvo dos millones de respuestas y daba como ganador al otro candidato, el republicano Landon. Parece que entre los lectores se daba un cierto sesgo conservador.

Desde entonces se han realizado cientos de miles de estudios, tanto políticos como comerciales o sociológicos que mostraron su capacidad de reflejar las opiniones del conjunto de la sociedad, redujeron la incertidumbre y ayudaron a tomar miles de decisiones.

Cuando llegó la digitalización empezamos a poder manejar grandes bases de datos desestructurados y el tamaño y el hecho de que fueran datos de personas concretas de las que se podía seguir su huella hizo que ganaran terreno para la investigación.

En paralelo, la representatividad de las muestras con las que se elaboraban estudios *online* pasó a ser algo menor, especialmente en el caso de algunos grupos sociales, que tardaron más en acceder masivamente a internet.

Además, internet dio lugar a una infinidad de fenómenos muy minoritarios que habrían precisado de muestras representativas de tamaño muy grande para alcanzar la fiabilidad necesaria.

Como consecuencia de todo eso parece que, aunque todavía muchos grandes estudios se basan en ellas, solas o en compañía de otros métodos, el futuro ya no pasa por las muestras representativas. Podrían tener aún un papel como complemento a otros estudios basados en el análisis de bases de datos, pero no como fuente única de conocimiento.

No sé si la investigación será mejor ahora, pero una época interesante, y en la que hemos analizado con éxito muchos aspectos de la sociedad, parece que toca a su fin. ■

# KANTAR

El valor de un equipo con  
una profunda comprensión  
del consumidor

- Líder destacado en España.
- Más de 1.400 especialistas en todos los sectores y disciplinas.
- Principal fuente de datos del consumidor, *shopper* y audiencia.
- Más *agile* y online que nunca: te ayudamos a tomar decisiones rápidas para un mayor retorno.



[www.kantar.com/es](http://www.kantar.com/es)



Síguenos  
en LinkedIn



Síguenos  
en Twitter



Visita  
nuestra Web



GRUPO

ANÁLISIS  
E INVESTIGACIÓN

## PRIMER GRUPO INDEPENDIENTE

de investigación de mercados  
y data science en España  
y Latinoamérica

"TRANSFORMANDO  
LA INVESTIGACIÓN  
A TRAVÉS DE LA  
TECNOLOGÍA"

MADRID |  
BARCELONA |  
BILBAO |  
VITORIA |  
ZARAGOZA |  
NUEVA YORK |  
BOGOTÁ |  
MÉXICO D.F. |  
SANTIAGO DE CHILE |  
LA PAZ |  
SÃO PAULO |

[www.analiseinvestigacion.com](http://www.analiseinvestigacion.com)

Síguenos en   @grupo\_aei