

Fastbook 09

Medición de Eficacia Publicitaria

Experimentos



09. Experimentos

En este fastbook nos centraremos en los experimentos como alternativa de modelización del ROI. Para ello, veremos cómo funcionan los experimentos en plataformas como Facebook y Google y hablaremos sobre los geo-experimentos. Finalmente, veremos cómo podemos utilizar los resultados de este tipo de medición para calibrar el resto de métodos.

Autor: Carlos Real Ugena

- ¿Qué es un experimento?
- Experimentos en Facebook
- Geo-experimentos
- Calibración de resultados con experimentos

¿Qué es un experimento?

X Edix Educación

Empezamos con la **definición genérica de diseño experimental**.

Un diseño experimental es una prueba o serie de pruebas en las cuales existen cambios deliberados en las variables de entrada de un proceso o sistema, de tal manera que sea posible observar e identificar las causas de los cambios que se producen en la respuesta de salida.

El diseño experimental existe desde hace mucho tiempo y sus orígenes provienen de la agricultura. Los agricultores aprendían a través de innumerables pruebas, cada una con solo una variable (por ejemplo, tipo de suelo, fertilizante), que cambiaban para mejorar su cosecha con el tiempo. Mantenían todas las cosas iguales, pero modificaban una variable cada vez para determinar qué factor producía el mayor beneficio.

En nuestro caso, la realización de un experimento es una forma comprobada de evaluar la eficacia de los anuncios, sin exagerar ni subestimar su verdadero impacto.

Anteriormente, ya vimos la siguiente simplificación del impacto de la publicidad.



Los dos tipos de modelos vistos hasta ahora (MMM y atribución) son técnicas que se aplican para medir el ROI, pero en ciertas ocasiones pueden sobreestimar o infravalorar el efecto de la publicidad. Esto se debe a que en algunas ocasiones lo que están midiendo es correlación y no causalidad. Motivado por esta dificultad, surgen los experimentos para medir el retorno de la publicidad, sin caer en el problema de la correlación.

Hay dos tipos principales de experimentos que se pueden emplear:

- Prueba controlada aleatorizada o ensayo controlado aleatorizado (en inglés, *randomized controlled trial*, RCT).
- Test A/B.

Veamos, a continuación, la diferencia entre los dos.

1

Prueba controlada aleatorizada (RCTs)

Es un tipo de experimento científico utilizado principalmente para conocer el efecto de un tratamiento sobre una población. La idea básica de una prueba controlada aleatorizada es ofrecer, de manera aleatoria, el tratamiento a un grupo de individuos (**grupo test o de tratamiento**) y a otro no (**grupo de control o de comparación**).

Pensemos en un caso real de experimento en otro ámbito para entenderlo mejor. Varios investigadores utilizaron este tipo de experimento para probar la reducción del tabaquismo a partir de inhaladores de nicotina orales.



Accede desde [este enlace](#) al experimento.

A continuación, un breve resumen de la metodología:

Objetivo

Determinar si el uso de un inhalador de nicotina oral reduce a largo plazo el tabaquismo y si el uso de este reemplazo de nicotina es seguro.

Diseño

Prueba controlada aleatorizada, controlado con placebo. Cuatro meses de ensayo con un seguimiento de dos años.

Lugar

Dos clínicas pulmonares de hospitales universitarios en Suiza.

Participantes

400 voluntarios sanos, reclutados a través de anuncios en los periódicos, dispuestos a reducir su tabaquismo pero que no pueden dejar de fumar inmediatamente.

Intervención

Inhalador activo o placebo según sea necesario durante un máximo de 18 meses, animando a los participantes a limitar su tabaquismo tanto como sea posible.

Principales medidas de resultado

Número de cigarrillos fumados por día desde la semana seis hasta el punto final.
Disminución verificada mediante una medición de monóxido de carbono exhalado en cada momento en comparación con la medición en la línea de base.

Resultados

A los cuatro meses se logró una reducción sostenida del tabaquismo en 52 (26%) participantes en el grupo activo y 18 (9%) en el grupo placebo ($P < 0,001$; prueba de Fisher). Las cifras correspondientes después de dos años fueron 19 (9,5%) y 6 (3,0%) ($P = 0,012$).

Conclusión

Los inhaladores de nicotina lograron de manera efectiva y segura una reducción sostenida del tabaquismo durante 24 meses. La reducción con o sin sustitución de la nicotina puede ser un primer paso factible hacia el abandono del hábito de fumar en personas que no pueden o no desean dejar de fumar abruptamente.

Como podemos ver, utilizando dos grupos de pacientes, test y control, se ha podido medir la efectividad de los inhaladores de nicotina. **En el caso de la publicidad, estaremos interesados en realizar el mismo tipo de análisis en el que queremos medir la eficacia publicitaria a través de dos muestras de usuarios.** Por ejemplo, un grupo será impactado con publicidad y otro no, para posteriormente analizar el comportamiento de ambos grupos.

La aleatorización ayuda a garantizar que los resultados se deben a la campaña publicitaria. Es una forma eficaz de hacer comparaciones justas para que podamos concluir que los resultados son causales.

Si la aleatorización se realiza correctamente, estos dos grupos serán equivalentes en todas las dimensiones (por ejemplo, edad y sexo) excepto en el tratamiento (por ejemplo, exposición al anuncio).

2

Test A/B

El test A/B consiste en comparar dos versiones de una misma página web, aplicación o anuncio para comprobar cuál de las dos versiones es más eficiente.

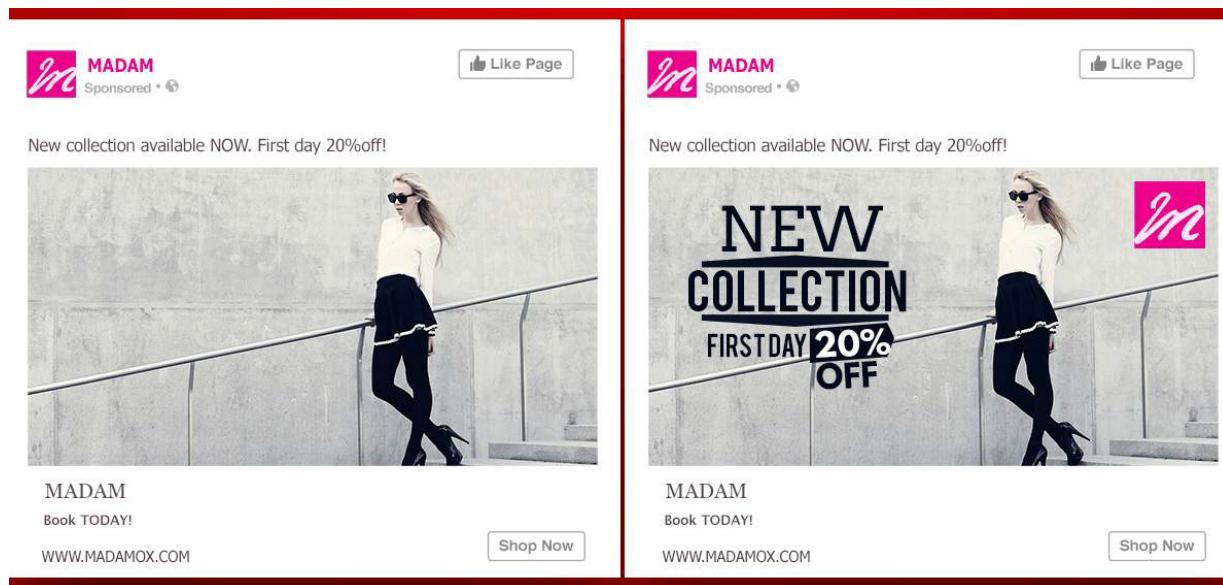
Estas variaciones, llamadas A y B, se muestran de forma aleatoria a los distintos usuarios de la página web. Una parte de ellos verá la versión A y la parte restante verá la versión B.

¿Cuál es la principal diferencia con las pruebas controladas aleatorizadas?

Que en este caso no hay grupo de control.

Las dos poblaciones o grupos de usuarios serán impactados, por ejemplo, por distintas variaciones del anuncio y el test A/B servirá para determinar cuál funciona mejor.

Este tipo de experimento nos servirá para decidir qué anuncio funciona mejor. Por ejemplo, podemos medir cuál de las siguientes dos creatividades es mejor para generar ventas:



Fuente: [Blog de Jeff Bullas](#)

Lección 2 de 4

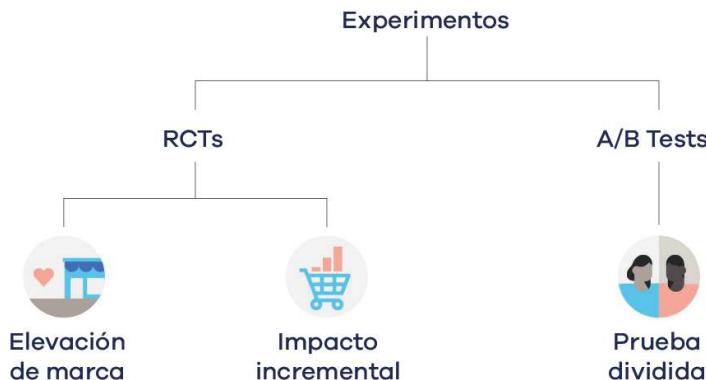
Experimentos en Facebook

X Edix Educación

Tanto Facebook como Google permiten utilizar sus plataformas para lanzar experimentos que determinen la incrementalidad de la publicidad. Dado que ambas plataformas funcionan y miden KPIs similares, nos centramos en los experimentos de Facebook en esta sección.

- ⓘ Si quieras aprender cómo funciona Google, puedes hacerlo leyendo el contenido de [este link](#). Si quieres saber cómo es la plataforma de Facebook accede a [Facebook Ads](#).

Facebook resume los experimentos que se pueden lanzar a través de Facebook Ads en un diagrama como el siguiente:



En resumen, Facebook permite ejecutar dos tipos de RCTs:

- Brand lift (elevación de marca).
- Conversion lift (impacto incremental).
- Un tipo de test A/B que se denomina split testing (prueba dividida).

Veamos, a continuación, qué soluciones proporciona Facebook para cada tipo de experimento.

1

Brand lift

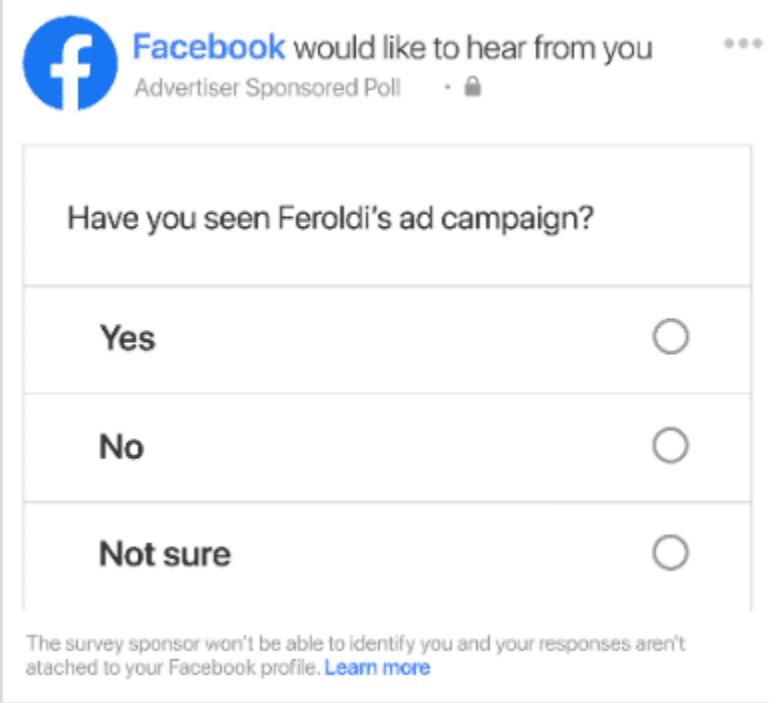
Es una prueba controlada aleatorizada que sirve para medir cuál es la influencia de una campaña en intangibles sobre la marca. Va a ser el test recomendado cuando el objetivo de una campaña sea incrementar el conocimiento o la percepción de una marca. Por ejemplo, se empleará durante lanzamientos de nuevas marcas o productos.

La manera de llevar a cabo este test es a través de encuestas. Facebook ofrece opciones flexibles para adaptar las preguntas de su prueba en función de las necesidades y objetivos comerciales.

Se pueden incluir hasta tres preguntas en el test. Cada persona verá solo una de esas preguntas, elegida al azar. Para cada pregunta de la encuesta, al menos una respuesta de ella, representa un resultado afirmativo con respecto al objetivo comercial que representa.

A continuación, varios ejemplos de preguntas que se pueden hacer de la parte de arriba del funnel de compra (proceso de compra).

Notoriedad publicitaria



The image shows a Facebook poll interface. At the top, it says "Facebook would like to hear from you" and "Advertiser Sponsored Poll". Below the question, there are three options: "Yes", "No", and "Not sure", each with a radio button next to it. At the bottom, a note states: "The survey sponsor won't be able to identify you and your responses aren't attached to your Facebook profile. [Learn more](#)".

Have you seen Feroldi's ad campaign?

Yes

No

Not sure

The survey sponsor won't be able to identify you and your responses aren't attached to your Facebook profile. [Learn more](#)

Notoriedad de marca

 Facebook would like to hear from you • • •
Advertiser Sponsored Poll • 

Have you seen Feroldi's new product line?

Yes

No

Not sure

The survey sponsor won't be able to identify you and your responses aren't attached to your Facebook profile. [Learn more](#)

Notoriedad de marca (top of mind)

 Facebook would like to hear from you ***
Advertiser Sponsored Poll • 

Which of the following brands comes to mind when you first think of men's fashion?

Joystick Studios

Feroldi's

Spruce

Club:Soda

None of the above

The survey sponsor won't be able to identify you and your responses aren't attached to your Facebook profile. [Learn more](#)

También se pueden realizar preguntas de la parte intermedia del funnel. Familiaridad con la marca.

 Facebook would like to hear from you

Advertiser Sponsored Poll • ⓘ

How familiar are you with Feroldi's name?

I know a lot about this product	<input type="radio"/>
I know a fair amount this product	<input type="radio"/>
I know a little about this product	<input type="radio"/>
I've only heard the name	<input type="radio"/>
I've never heard of the product	<input type="radio"/>

The survey sponsor won't be able to identify you and your responses aren't attached to your Facebook profile. [Learn more](#)

Y también de la parte de abajo del funnel. Intención de compra.

Facebook would like to hear from you
Advertiser Sponsored Poll • ⓘ

Will you buy Feroldi's the next time you shop for men's clothing?

Yes

No

Not sure

I am not familiar with this brand

The survey sponsor won't be able to identify you and your responses aren't attached to your Facebook profile. [Learn more](#)

En cuanto a la distribución de las encuestas, un usuario no puede ver más de una en siete días, independientemente de la plataforma. Facebook muestra las preguntas de encuestas en Facebook, Instagram, su red de anuncios y Messenger.

Es posible que una pregunta de encuesta no se publique en la misma ubicación que el anuncio. Por ejemplo, si la impresión se publicó en Instagram, la persona podría recibir la encuesta en Facebook.



Facebook proporciona una [guía](#) con puntos clave a la hora de ejecutar un experimento brand lift.

2

Conversion lift

Al igual que brand lift, conversion lift es una prueba controlada aleatorizada que utiliza grupos de test y control para medir el número de conversiones incrementales causadas por los anuncios de Facebook.

En el caso anterior de brand lift, nos encontrábamos ante información declarada que puede ser más o menos precisa, pero, en este caso, analizaremos la información observada.

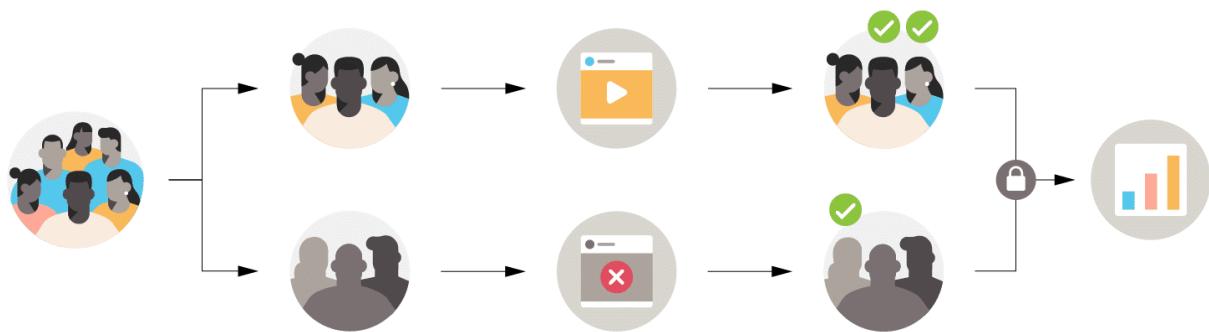
Un anunciante puede ejecutar una prueba de conversion lift en una campaña publicitaria de específica en Facebook o en toda su publicidad de Facebook.

Este test está diseñado para ayudar a determinar el impacto de la publicidad en Facebook, independientemente de otros esfuerzos de marketing. Es decir, sirve para responder la siguiente pregunta.

¿Cuántas conversiones adicionales se produjeron como resultado de los anuncios de Facebook?

Esto se conoce como conversion lift (o impacto incremental en español).

El siguiente gráfico resume la manera de medir las conversiones incrementales:



¿Cómo interpretamos el diagrama anterior?

La audiencia se separa en dos grupos: grupo de test (parte superior) y grupo de control (parte inferior).

Al grupo test se le muestra la campaña y al grupo de control no.

De esta manera, se pueden medir las conversiones del grupo 1 y del grupo 2.

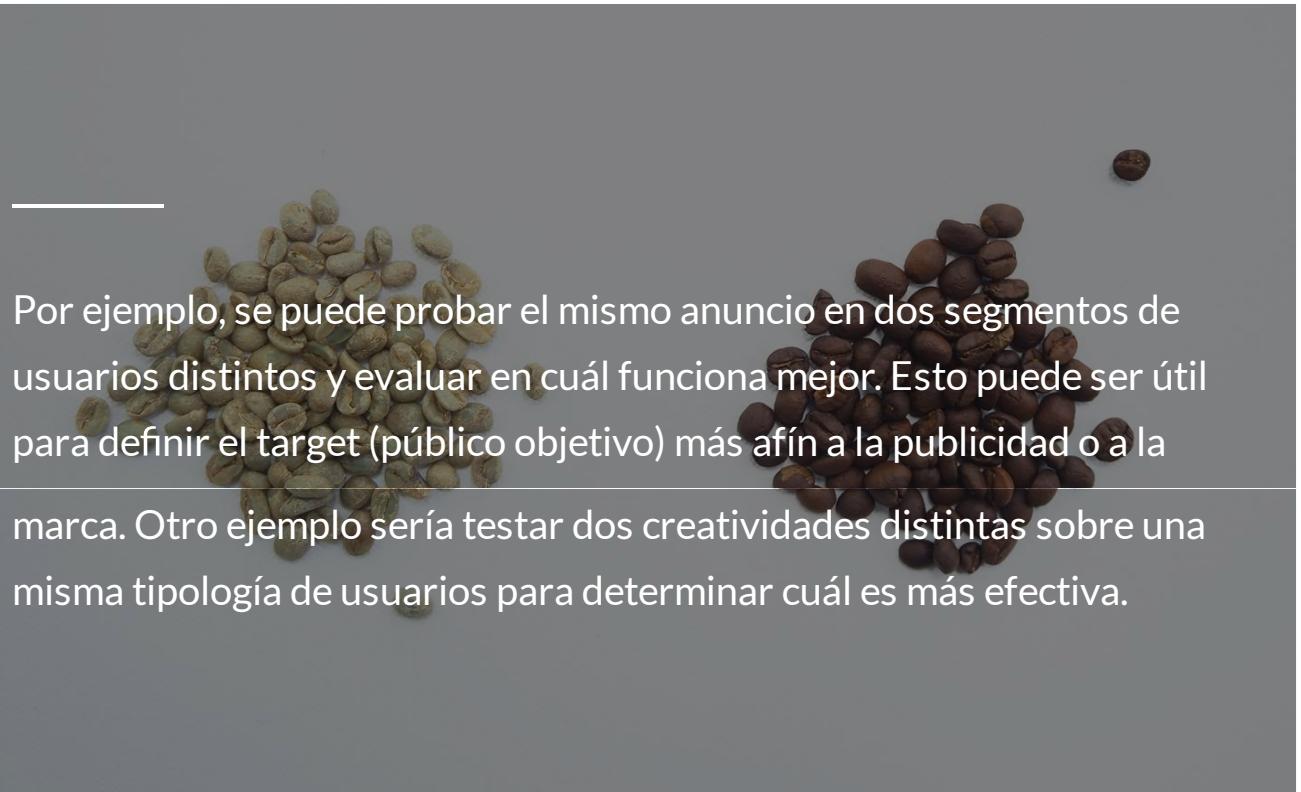
La diferencia entre conversiones será el número de conversiones incrementales causadas por la campaña.

3

Split testing

En este caso, nos encontramos ante un **test A/B**. El objetivo principal de este tipo de medición va a ser optimizar la ejecución de campañas. Como vimos previamente, no emplearemos este método si lo que queremos es estimar el incremento generado por la publicidad, sino evaluar 2 o más acciones al mismo tiempo.

Split testing (prueba dividida en español) **es una forma de que los anunciantes puedan determinar qué tácticas publicitarias producen el mejor resultado** para su campaña en función de sus indicadores clave.



Por ejemplo, se puede probar el mismo anuncio en dos segmentos de usuarios distintos y evaluar en cuál funciona mejor. Esto puede ser útil para definir el target (público objetivo) más afín a la publicidad o a la marca. Otro ejemplo sería testar dos creatividades distintas sobre una misma tipología de usuarios para determinar cuál es más efectiva.

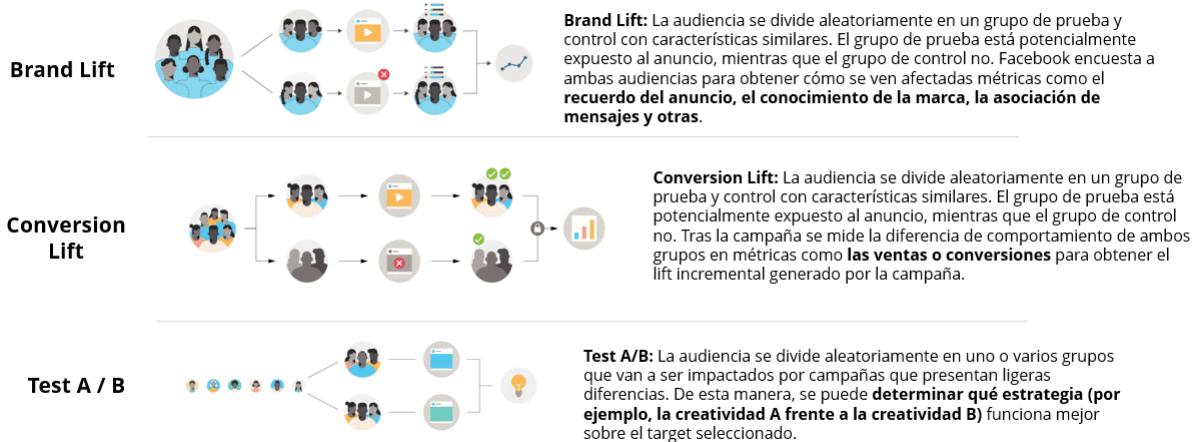
¿Es mejor enseñar el precio o un descuento? Cada persona, basada en su intuición, podría responder a la pregunta anterior. Sin embargo, **el test A/B nos dará un número real del impacto de cada creatividad en la métrica seleccionada.**

En un test A/B, la audiencia se divide en grupos aleatorios que no se superponen. Esta aleatorización sirve para garantizar que los resultados no estén sesgados, es decir, que haya otros factores que desvíen los resultados.

Cada grupo ve conjuntos de anuncios que son idénticos excepto por la variable que está probando: la audiencia, las ubicaciones de los anuncios o la creatividad.

Facebook considera fiables todos aquellos test que tengan una confianza de, al menos, el 75%.

Para terminar, podemos resumir los tres tipos de experimentos de la siguiente forma:

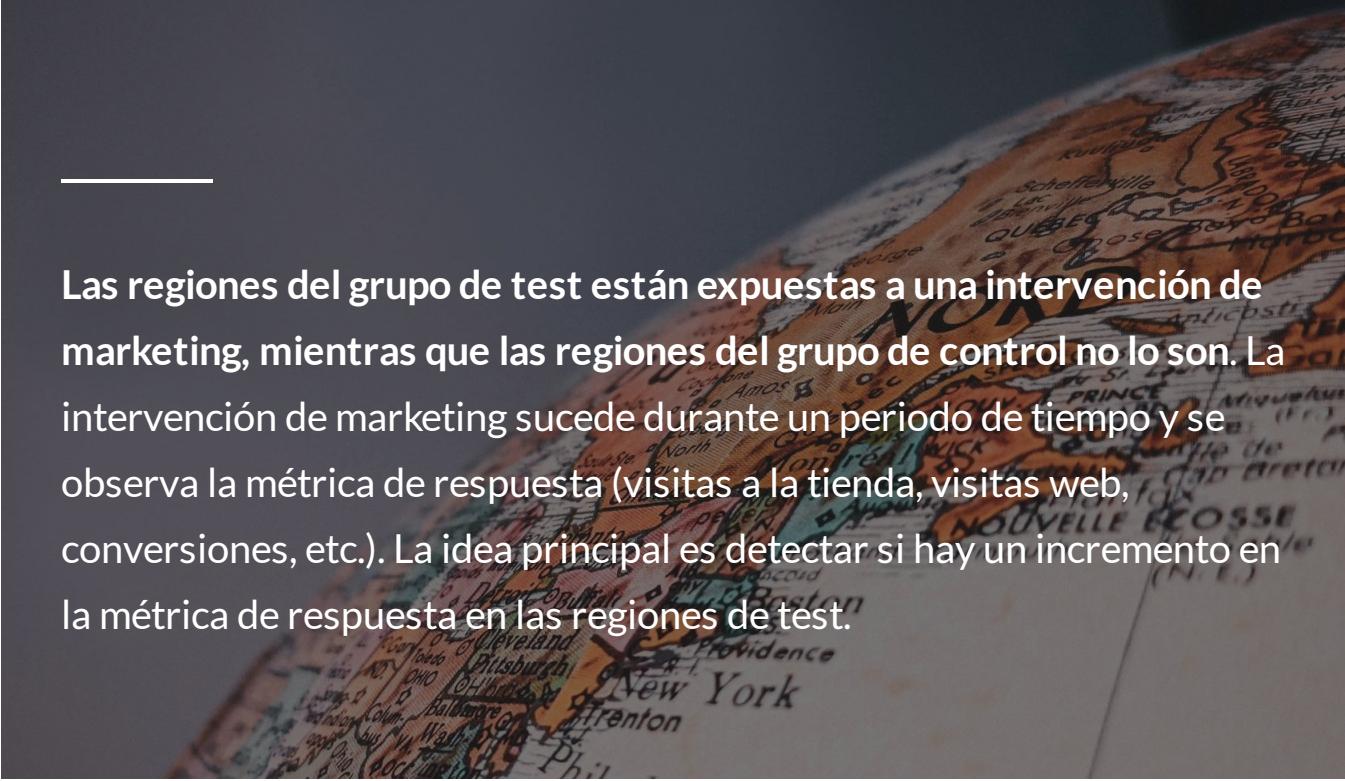


Geo-experimentos

X Edix Educación

Los **geo-experimentos** son similares a los experimentos vistos previamente, pero **utilizan geografías / regiones para definir su grupo de control y test**, en lugar de usuarios o cookies web.

En estos experimentos, **las regiones geográficas se dividen en dos grupos: test y control**. Dichas regiones tienen que ser lo más comparables entre sí, lo cual no es siempre fácil. Imaginémonos que nos encontramos en un país en el que todas las ciudades tienen dinámicas distintas. En este caso, será muy complicado definir estos grupos para que sean comparables.



Las regiones del grupo de test están expuestas a una intervención de marketing, mientras que las regiones del grupo de control no lo son. La intervención de marketing sucede durante un periodo de tiempo y se observa la métrica de respuesta (visitas a la tienda, visitas web, conversiones, etc.). La idea principal es detectar si hay un incremento en la métrica de respuesta en las regiones de test.

En este tipo de experimentos, hay tres fases clave.

1

Diseño del experimento

En esta fase vamos a determinar el **mejor método para dividir las áreas geográficas en dos grupos**. Es la fase que llevará más tiempo y que nos asegurará que podamos comparar los resultados posteriormente.

Para llevar a cabo la selección de áreas, se puede optar por una de estas metodologías:

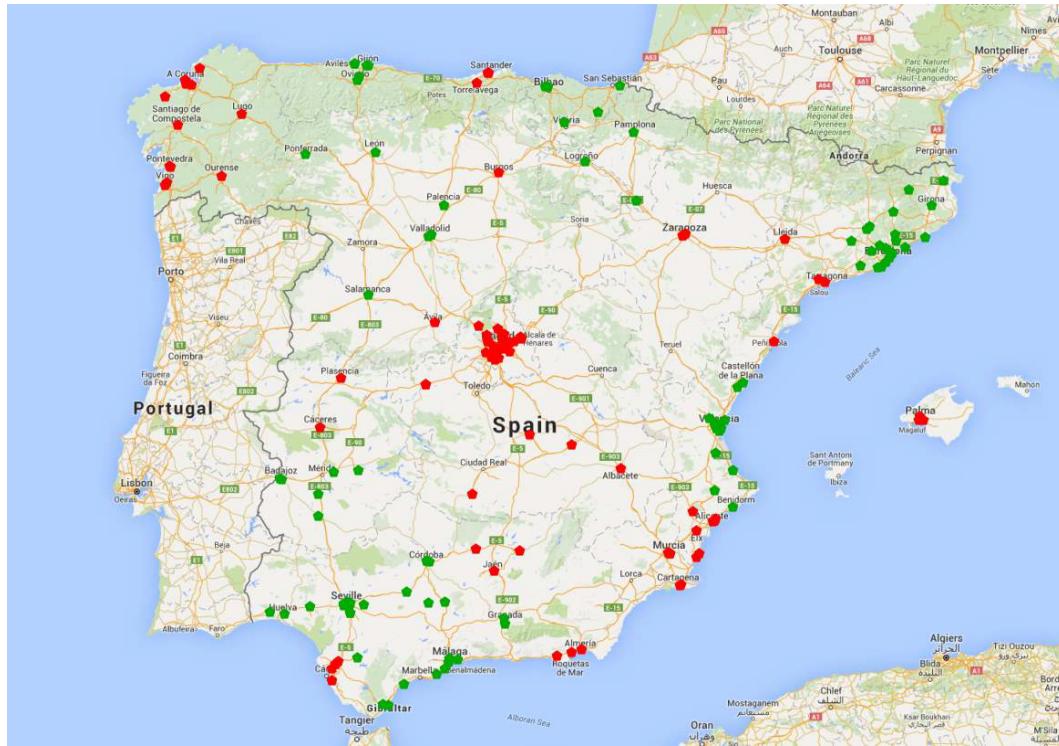
Selección aleatoria

En este caso, tomaremos las regiones de manera aleatoria y formaremos nuestro grupo test y control. Aquí podemos también emparejar regiones, por ejemplo, ciudades para luego comparar los resultados. Este método funcionará cuando las regiones sean lo suficientemente similares entre ellas.

Selección no aleatoria

En este caso, haremos un análisis basado en clustering (segmentación) o análisis de similitud de series temporales para definir las regiones que estarán en el grupo de control y en el de test. En este proceso, también ‘descubriremos’ el parecido entre regiones para llevar a cabo este tipo de estudios.

El resultado es tener un mapa con las distintas regiones como el siguiente:



Mapa con la localización de los puntos de venta de los grupos test (verde) y control (rojo).

Adicionalmente, podremos tener emparejadas las ciudades para poder realizar comparativas adicionales.

2

Ejecución de la campaña

Una vez definido el periodo para evaluar los resultados, se lanzará la campaña en las regiones del área test, garantizando que las personas que están en las regiones de control no son impactadas con ella.

Para poder llevar a cabo esta fase es importante que se cumplan los siguientes requisitos:

Control de la plataforma

No todas las plataformas permiten lanzar campañas a nivel de ciudad/región. Esto es importante a la hora de definir la campaña que queremos evaluar.

Control de variables

Es necesario asegurar que el resto de variables son similares. Si, por ejemplo, hubiera otras acciones de marketing afectando a algunas de las regiones (por ejemplo, radio local), los resultados podrían no ser comparables.

Presupuesto

El presupuesto debe ser suficiente para que tenga un impacto significativo en las variables a analizar. Podríamos acabar asumiendo que la publicidad no está generando un incremento en negocio cuando el problema ha sido el volumen de inversión, y no la efectividad del medio.

Alineación de equipos

Es importante que los distintos equipos (marketing, producto, comercial, etc.) estén alineados para que el experimento funcione. Si, por ejemplo, se lanza una campaña en digital ofreciendo un descuento en tienda, la sincronización debe ser total. Imagina lo que ocurriría si al ir a la tienda te dicen que no existe ese descuento.

3

Evaluación de los resultados

Para evaluar los resultados, compararemos la métrica o métricas bajo estudio en los distintos períodos temporales:

Periodo pretest

Antes de que la campaña haya empezado.

Periodo test

Cuando la campaña se está ejecutando.

Periodo post-test

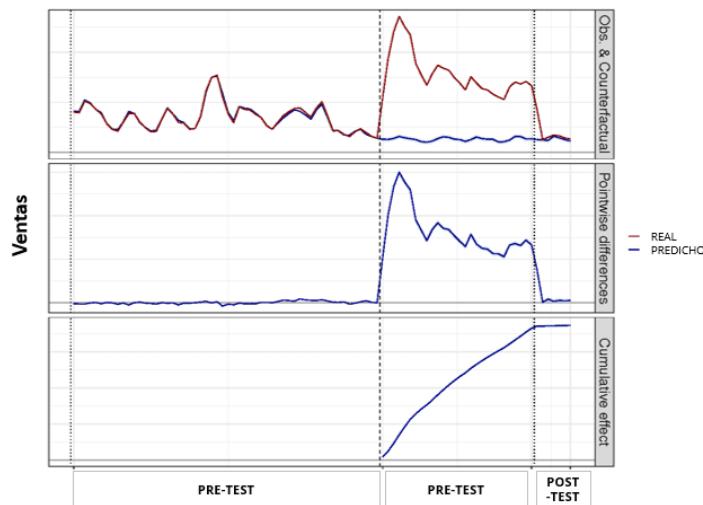
Periodo posterior que podría estar todavía afectado por la campaña. Recordemos que la publicidad tiene un efecto inmediato y posterior que puede ser recogido con el ad-stock.

En cuanto al método, lo habitual será utilizar un counterfactual analysis (análisis contrafactual) que permite comparar lo que sucede dado un cambio, versus lo que debería haber sucedido si no hubiera ocurrido algún cambio en primer lugar. Existe una gran variedad de métodos estadísticos para realizar este test.



Uno muy empleado en geo-experimentos es el modelo [Causal Impact](#) desarrollado por Google.

A continuación, una representación de lo que se obtendría aplicando este tipo de modelo:



Como podemos observar en el gráfico, el impacto de la campaña ha sido alto. La diferencia entre el valor real y el predicho que se muestra en el segundo gráfico es el valor incremental sobre ventas que podremos otorgar a la campaña. Habitualmente, se trabajará con intervalos de confianza para evaluar la confianza de los resultados.

Calibración de resultados con experimentos

 Edix Educación

Ahora que ya hemos visto distintos modelos de medición (marketing mix modeling, atribución y experimentos), ¿cómo podríamos utilizar los resultados de los experimentos?

Una tendencia seguida por los anunciantes es calibrar los resultados de MMM o atribución a partir de los experimentos.

Recordemos que los experimentos son los más indicados para medir la causalidad, por lo que los resultados estarán más cercanos a la realidad que el resto de métodos.

Cabe destacar que realizar experimentos suele ser más costoso (en dinero y tiempo) y no permite medir la eficacia de todos los medios, por lo que combinar resultados será indispensable.

¿Cómo podemos calibrar MMM o atribución con resultados CLE? Una opción es definir un enfoque metodológico donde se utiliza uno o varios experimentos para restringir MMM o atribución, ayudándolo a aprender y reflejando el tamaño del efecto medido en CLE.

En el siguiente artículo elaborado por Deloitte, se detalla un caso real de aplicación de esta metodología:

Deloitte.

Exploring Marketing Mix Modeling (MMM) and Conversion Lift Experiment (CLE) blending
The Alshaya Group/H&M Case developed by Deloitte

03. Calibration / validation method algorithm

04. Results for the Alshaya / H&M discussion

Fuente: [Deloitte.](#)

Podemos decir que hay **tres beneficios principales** de ejecutar experimentos para medir la incrementalidad de la publicidad:

Respuesta a preguntas específicas

El diseño experimental nos permite responder interrogantes que tengamos sobre la publicidad y su impacto en el negocio.

Facilidad de implementación —

En Google, Facebook, TikTok y la mayor parte de plataformas están incluyendo módulos para poder evaluar las campañas a través de esta metodología de una manera sencilla.

Gold-Standard de la medición —

Un experimento válido y controlado nos dirá si el marketing en una plataforma específica crea un valor incremental. Otros métodos de medición (MMM y atribución) son útiles, pero pueden confundir causalidad con correlación.

Fácil de entender —

Los experimentos controlados son transparentes y fáciles de entender. Los especialistas en marketing de cualquier nivel pueden comprender los resultados.

¡Enhорabuena! Fastbook superado

edix

Creamos Digital Workers