

Opinión

Tres artículos

Este artículo es el primero de una serie de tres, que se publicarán en número sucesivos de ANUNCIOS, en los que Jan Hofmeyr (en la imagen), director mundial de innovación de Synovate, filial de investigación el grupo Aegis, trata sobre diferentes problemas relativos a la predicción de los comportamientos de los consumidores, los métodos que se usan para hacerla y cómo obtener los mejores resultados. Este primero trata sobre la diferencia de validez entre las respuestas agregadas y las individuales, y los restantes se ocuparán respectivamente de la satisfacción del consumidor y la intención de compra, y de la verdadera naturaleza de la predicción.



Diferencia entre validez agregada y validez a nivel individual, y

En nuestro sector existe mucha confusión en lo referente a la predicción. Estos artículos sobre investigación se han redactado para ayudar a aclarar dicha confusión. En este primer artículo hablaré de la diferencia que hay entre validez agregada y validez a nivel individual. El artículo concluye con un breve resumen sobre el rendimiento de las distintas herramientas para medir la fidelidad.



ENCUESTAS Y PANELES

No es habitual realizar una investigación que combine los datos de una encuesta con los que se registren en los paneles. Este artículo se basa en tres diarios y dos paneles de identificación de los productos comprados. Proceden de Market Facts/Synovate y TNS. Su valor radica en que por lo general se cree que dichos paneles son mejores que las encuestas a la hora de medir la conducta.

Lo que aprenderemos será lo siguiente:

- Los resultados agregados o globales de una encuesta pueden guardar gran correlación con la cuota de mercado y sin embargo ser erróneos en cuanto a quién compra una marca y con qué frecuencia la compra.
- Esto ocurre debido a un *error de compensación mutua*, es decir, para mucha gente que dice que comprará una marca y no lo hace, hay otros que dicen que no comprarán una marca y después sí lo hacen.
- Un *error de compensación mutua* provoca el fenómeno de buena validez agregada pero escasa validez a nivel individual, es decir, está en lo cierto globalmente pero equivocado a nivel individual.
- Cuando uno se encuentra

Los resultados agregados o globales de una encuesta pueden guardar gran correlación con la cuota de mercado y sin embargo ser erróneos en cuanto a quién compra una marca y con qué frecuencia la compra.

con una validez escasa a nivel individual, no se puede *responder con seguridad* a las preguntas sobre quiénes son sus clientes fieles ni por qué, es decir, no se puede estar seguro del análisis de los perfiles ni del de los *drivers*.

- La razón es que si uno se equivoca sobre quién le es fiel en la realidad, quizá se acabe obteniendo una modelización y unos perfiles de un grupo de personas erróneo¹.

• La solución es utilizar indicadores con la mejor validez posible a escala individual. El indicador del Attitudinal Equity (AE) de Brand Value Creator (BVC) parecen ser de los mejores indicadores.

ALGUNOS EJEMPLOS DEL 'ERROR DE COMPENSACIÓN MUTUA'

1. Investigación sobre las diferentes metodologías de investigación 23: la validez de una pregunta acerca de la intención de compra

Antes de que se convirtiera en Synovate, Market Facts Inc. publicó una importante serie de artículos sobre la *Investigación de las distintas metodologías de investigación*. El informe 23 se titula *Midiendo la intención de compra: en qué medida es válida la estimación*. Utilizando una escala sencilla de "sí" / "no", plasma los resultados de un estudio en el que se preguntó a los participantes si tenían pensado comprarse un microondas "en los próximos 18 meses". Dieciocho meses después, se volvió a contactar con los encuestados para averiguar si se habían comprado o no un microondas.

En la encuesta inicial, un 9% dijo que se compraría un microondas. En el seguimiento, un 7% afirmó haberlo hecho. Este es un resultado relativamente bueno a nivel global, es decir, a nivel agregado. Pero oculta una verdad complicada: sólo un 26,5% de quienes dijeron que iban a comprarse un microondas lo acabaron haciendo. La mayoría no lo hizo. La razón por la cual el resultado global seguía correspondiéndose razonablemente bien con lo que ocurrió es que mucha gente que había dicho que no iba a comprarse un microondas al final sí lo hizo.

En eso consiste el *error de compensación mutua*: por cada persona que da una respuesta errónea, hay otra que también lo hace, pero en dirección contraria. Como resultado, la validez agregada es buena pero la validez a nivel individual es baja; acertamos en términos globales, pero fallamos a nivel individual.

2. ¿Quiénes son mis 'heavy drinkers'?

Hace unos años, mis compañeros y yo tuvimos la oportunidad de testar la validez de los hábitos de consumo declarados por los participantes de un estudio internacional sobre refrescos (Ver Louw et al, 2005). Los participantes (n: 899) realizaron una encuesta típica sobre uso y actitudes en la que se les preguntaba con qué frecuencia bebían distintas marcas de refrescos. El cliente utilizó una pregunta sobre *frecuencia* para determinar los perfiles de los consumidores *heavy*,

Validez agregada y validez a por qué es importante / 1

medium y *light* de su bebida, y para entender las motivaciones de sus consumidores *heavy* o fidelizados. Los *heavy drinkers* eran quienes afirmaban beber la marca del cliente al menos una vez al día.

Luego, los participantes rellenaban un cuaderno todos los días durante una semana para apuntar lo que bebían en realidad. ¿Cómo eran dichos resultados en comparación con los primeros?

Con respecto a todas las marcas, un 42% de quienes afirmaron beber una marca diariamente no anotaron que la bebieron durante la semana siguiente. Esta disonancia entre la encuesta y los datos del diario se repitió en todos los niveles de frecuencia. Sin embargo, las cuotas de mercado agregadas para ambos conjuntos de datos coincidían con lo que el cliente se esperaba. Esto se debió a que aunque muchas personas afirmaron ser *heavy drinkers* en un método (por ejemplo, en la encuesta), hubo *light drinkers* que pasaron a ser *heavy drinkers* en el otro (por ejemplo, en el diario). Y por ello, las cuotas de mercado fueron relativamente precisas a nivel agregado, pero los datos que dieron muchas personas en la encuesta o en el diario eran erróneos a nivel individual.

Nuestro sector ha tendido a anteponer la veracidad de los datos del diario a la de los datos de la encuesta. Sin embargo, en sentido estricto, no podemos saber cuáles eran erróneos. Quizá quienes afirmaron ser *heavy drinkers* en la encuesta normalmente eran *heavy drinkers* (aunque, precisamente esa semana, no). Pero se puede ver el problema del cliente: ¿cómo responde la pregunta sobre quiénes son sus verdaderos *heavy drinkers*? ¿Debería tener en cuenta a quienes afirmaron ser *heavy drinkers* en la encuesta, o a quienes dijeron en el diario que la habían consumido mucho?

3. ¿Quién utiliza mis marcas?

Louw et al (2005) puso un segundo ejemplo de disonancia entre el uso que se declaraba en un panel diario y en una encuesta. Estos datos proceden del Reino Unido. Se obtuvieron de un importante fabricante de utensilios domésticos y productos para el cuidado personal y abarcó veintiuna marcas procedentes de tres categorías de productos (por razones de confidencialidad con respecto al cliente no puedo concretar más). El panel diario estaba formado por personas que realizaban las compras del hogar. Ha estado en funcionamiento durante dos años en los que hemos tenido una oportunidad única de encuestar a los panelistas.

El cliente es un gran inversor en investigación convencional mediante encuestas. Suele utilizar una pregunta sobre *uso reciente* para determinar cuántas personas utilizan sus marcas, quiénes son y por qué lo hacen. En este caso, el cliente identificó a los usuarios con la siguiente pregunta: “¿Qué marcas ha utilizado en los últimos tres meses?”. Por eso, preguntamos a los panelistas por las marcas que habían utilizado en los últimos tres meses. Esto nos permitió comparar las marcas que los panelistas habían afirmado comprar para su hogar con las marcas que después afirmaron haber utilizado.

De entre todas las categorías y marcas, un 62% de las marcas que los panelistas dijeron haber utilizado “en los últimos tres meses” no se anotaron en los diarios. Sin embargo, en una cantidad parecida de ocasiones, los panelistas anotaron en los diarios haber comprado marcas que no mencionaron en la encuesta. En ambos conjuntos de datos, las penetraciones de marca encajaban con lo que el cliente sabía de las penetraciones a partir de los datos del panel de identificación de los productos comprados.

A diferencia del ejemplo de las bebidas, no se trata de una falta de correspondencia entre lo que dijo la gente que solía hacer (por ejemplo, beber una marca “una vez a la semana”) y lo que acababa haciendo (no beberla). Se trata de una disonancia entre lo que la gente escribe en un diario acerca de lo que ha comprado y lo que dice haber utilizado en una encuesta. Aunque es cierto que pueden haber comprado algunas marcas para que las utilicen otros, es poco probable que otra persona haya comprado las marcas que utilizaban. Las principales causas de esta

disonancia son: a) la gente no es muy meticulosa cuando rellena un diario; b) la gente no recuerda lo que ha hecho; o c) un poco las dos cosas.

Puede parecer poco probable que la gente haya acertado un porcentaje tan pequeño de marcas que habían comprado en los últimos tres meses (38%), sin embargo, hay que recordar que, en el ejemplo anterior de la *Investigación sobre las distintas metodologías de investigación*, un porcentaje de personas aún menor acertaron lo que iban a comprar (26,5%).

4. Cuota de gasto del comercio al por menor

Un último ejemplo: en 2007 y en 2008 tuvimos la oportunidad de encuestar a personas en un panel de identificación de los productos comprados sobre sus actitudes hacia los dieciséis comercios minoristas más importantes de productos alimenticios de Italia (Ver Hofmeyr et al, 2008). El panel pertenecía a una de las compañías líderes mundiales en investigación de mercados especializada en recabar información actitudinal mediante paneles. Una de las preguntas de su encuesta trataba de la satisfacción del cliente.

En la Tabla 1 aparece la correlación entre la satisfacción agregada de los clientes y las cuotas de mercado de los comercios al por menor utilizando puntuaciones medias y porcentajes de clientes satisfechos.

La media es $R = 0,635$. Sin embargo, a nivel individual, la correlación entre lo que los participantes se gastaban en cada comercio y su satisfacción con cada comercio era $R = 0,27$ (2007) y $R = 0,22$ (2008). El motivo de la baja correlación a nivel individual es que hay muchos participantes que dan una buena puntuación a un comercio al por menor aunque no lo utilicen mucho. Sin embargo, hay otros panelistas que dan al comercio una puntuación baja pero compran mucho allí.

Si está familiarizado con BVC, podrá entender por qué puede haber disonancia entre el índice de satisfacción a nivel individual y la cuota de gasto. Los dos motivos principales son:

- *Market barriers* (barreras del mercado): las personas no siempre pueden utilizar las marcas que desean. En tal caso, utilizan otras marcas que no valoran tan bien;
- *Relative ratings* (puntuaciones relativas): la puntuación absoluta que consigue una marca no es lo importante. Lo que importa es cómo una persona puntúa una marca con respecto a las demás que podría utilizar.

Tabla 1. Correlación agregada entre la satisfacción del cliente y la cuota de mercado

| | 2007 | 2008 |
|------------------------|------|------|
| Media | 0,61 | 0,63 |
| % clientes satisfechos | 0,62 | 0,68 |
| Correlación media | 0,62 | 0,65 |
| TOTAL | 3712 | 2414 |

Continúa en página 18

5. Resumen

Me he fijado en cinco conjuntos de datos procedentes de cuatro países y seis categorías de producto que van desde productos duraderos a bienes de rotación rápida pasando por servicios. En cada caso, los resultados que se hallaron en el estudio (es decir, la intención de compra, frecuencia declarada, uso pasado declarado y satisfacción del cliente) encajaron con lo que se sabía de las penetraciones reales y cuotas del mercado. Sin embargo, el análisis a nivel individual mostró que:

- La mayoría de las personas que dijeron que comprarían un microondas “en los próximos 18 meses” no lo hizo. No obstante, el resultado global de la penetración en el mercado de los microondas fue bastante preciso porque mucha gente que dijo que no compraría un microondas, sí lo hizo.

- Una gran proporción de personas que admitieron beber una marca frecuentemente (al menos una vez al día) anotaron que no habían bebido nada en la semana siguiente. Sin embargo, las cuotas de mercado encajaron con las expectativas debido a la gran cantidad de los llamados *light drinkers* que afirmaron haberla bebido.

- Una gran proporción de personas que admitieron haber comprado ciertas marcas en el diario no mencionaron que las utilizaban en la encuesta. Sin embargo, las penetraciones de la marca encajaron con las expectativas porque una cantidad parecida de personas admitió haber utilizado marcas no habiendo anotado que las compraban.

- Las puntuaciones en cuanto a satisfacción del cliente que dieron los panelistas a dieciséis comercios de productos alimenticios no encajaron mucho con la cuota de gasto que obtuvo cada comercio al por menor de cada panelista. Sin embargo, las puntuaciones en cuanto a satisfacción global encajaron con la cuota de mercados a nivel agregado.

Todos los proyectos ilustran el problema del error que aparece a nivel individual.

POR QUÉ ES IMPORTANTE DIFERENCIAR ENTRE VALIDEZ AGREGADA Y VALIDEZ A NIVEL INDIVIDUAL

Imagine que está observando los resultados de cualquiera de los estudios anteriores: ¿cómo respondería las preguntas sobre quiénes son sus usuarios potenciales (*intención de compra*) o quiénes son sus actuales usuarios fidelizados (*uso reciente, frecuencia, satisfacción*)?

Veamos, por ejemplo, el estudio de los microondas. En este caso, más del 70% de las personas que dijeron que se comprarían uno, no lo hicieron. Puede que llegue a la conclusión de que sus mejores clientes potenciales sean mujeres con una renta elevada y con hijos viviendo en casa. Si es así, quizá esté equivocado. Puede que hubiera acertado si las personas que *sí acabaron comprando uno* fueran otras mujeres con una renta elevada y con hijos viviendo en casa que dijeron en la encuesta que no iban a hacerlo (pero correría un riesgo si se fiara de esto).

Lo mismo ocurre con el resto de estudios: quizá haya llegado a una conclusión acerca de quiénes son los usuarios de su marca (productos para el cuidado personal y utensilios domésticos), quiénes son sus *heavy drinkers* frente a los *light drinkers* (refrescos), o quiénes son sus clientes fidelizados (comercio al por menor). Puede que no esté en lo cierto.

El problema también se aplica al intentar averiguar por qué las personas compran ciertas marcas, es decir, lo que tradicionalmente se ha denominado *análisis de los drivers*. En todos los ejemplos anteriores quizá habría obtenido una modelización de un grupo de personas erróneo. Averiguar qué se asocia con un “aumento en el consumo de las bebidas” (refrescos) o con la satisfacción del cliente (comercios al por menor) podría inducir a error si las personas involucradas no fueran *heavy drinkers* o no utilizaran mucho dicho comercio al por menor.

En resumen: no se puede estar seguro de los intentos que pretenden responder preguntas sobre “quién” o “por qué” si todo de lo que se dispone es de una validez agregada. Cuanto mejor sea la validez a nivel individual de sus medidas, más seguro de las conclusiones se podrá estar.

UNA PEQUEÑA COMPARACIÓN DE LA VALIDEZ AGREGADA Y A NIVEL INDIVIDUAL DE LAS DISTINTAS MEDIDAS

En mi segundo artículo de esta serie, me centro en las correlaciones a nivel agregado e individual entre los distintos indicadores clásicos de fidelidad y conducta. El tercer artículo analiza en detalle el rendimiento de BVC. No obstante, pensé que sería buena idea incluir en este artículo los resultados de la comparación que se verán en el segundo artículo sobre los distintos indicadores de fidelidad².

Tabla 2. Comparación de la asociación entre los distintos indicadores de fidelidad y la conducta

| | Res. agregados | ...con... | A nivel individual | ...con... |
|---------------------------------|----------------|---|--------------------|---|
| Satisfacción del cliente | 0,75 | ...índices anteriores de crecimiento, cuota de mercados | 0,19 | ...cambio de marca, siguiente marca comprada, cuota de gasto del cliente... |
| Intención de compra | 0,52 | ...% de compra, ventas... | 0,27 | ...siguiente marca comprada, cuota de gasto del cliente... |
| Net Promoter Score | 0,83 | ...índice anterior de crecimiento... | Peor que otras | ...probabilidad de comprar más... |
| Attitudinal Equity | 0,85 | ...cuota de mercados... | 0,55 | ...cuota de gasto del cliente... |
| Total Equity (AE + NBE) | 0,90 | ...cuota de mercados... | No disponible | n/a |

Aparecen en la Tabla 2.

Hay que tener en cuenta que todavía no tenemos la validación a nivel individual del Total Equity (TE). Por lo tanto, lo que podemos afirmar sobre el TE es que supera a otros indicadores a nivel agregado, y que es probable que tenga un buen desempeño debido a que incluye el AE. A nivel individual, el indicador *Attitudinal Equity* de Synovate supera fácilmente a los indicadores clásicos de fidelidad.

En este sentido, cabe destacar que un fabricante líder a nivel mundial organizó recientemente un test controlado del indicador *Attitudinal Equity* (AE) que emplea Synovate frente a ocho indicadores propuestos por algunos de nuestros rivales líderes. El AE empleaba menos tiempo de estudio y superó a todos los demás.

Una gran proporción de personas que admitieron beber una marca frecuentemente (al menos una vez al día) anotaron que no habían bebido nada en la semana siguiente. Sin embargo, las cuotas de mercado encajaron con las expectativas.

Jan Hofmeyr
Director de innovación de Synovate

BIBLIOGRAFÍA

- Anónimo, fecha desconocida: “Measuring buying intention: How valid is the estimate”, Research on Research, 12, Market Facts Inc.
- Hofmeyr, Jan (2010): “The validity of customer satisfaction, purchase intention, and Net Promoter score,” *Prediction in Marketing Research*, 2, Synovate Laboratories
- Hofmeyr, Jan (2010): “What is prediction?,” *Prediction in Marketing Research*, 3, Synovate Laboratories
- Hofmeyr, Jan, Victoria Goodall, Martin Bongers, and Paul Holtzman (2008): “A new measure of brand attitudinal equity based on the Zipf distribution”, *International Journal of Market Research*, 50:2
- Louw, Alice, Vanessa Withington, and Alida Jansen (2005): “Self-reported Behavior: Fact or Fiction”, artículo sin publicar presentado en el congreso de la asociación South African Marketing Research Association

Notas

1. En este artículo utilizo expresiones como “quizá haya obtenido una modelización de un grupo de personas erróneo”, “no puede estar seguro de sus modelos”, etcétera. Es importante reforzar lo que quiero decir. Nuestros datos muestran claramente que muchas personas no están haciendo lo que afirman en las encuestas. Sin embargo, aún puede haber acertado en parte acerca de *quién* o *por qué* siempre que otras personas idénticas también se hayan equivocado acerca de lo que hacen. Por eso, utilizo expresiones como “quizá sea erróneo”.

Dicho esto, también es importante destacar que los análisis de perfiles y de *drivers* a menudo están asociados a sesgos estadísticos de los datos. En otras palabras, usted identifica grupos en los que tiene el mayor *índice de impacto* o que tienen ciertas motivaciones en particular. Por eso, serán, por definición, más difíciles de hallar en el resto del estudio. La única solución es buscar la mejor validez a nivel individual que se pueda obtener.

En este sentido, las herramientas que ha desarrollado Synovate Laboratories están sumamente centradas en el problema de la validez a nivel individual. Por ello, Synovate está bien posicionado para aprovechar lo que preveo será un énfasis creciente en la *validez a nivel individual* dentro de nuestro sector.

2. El problema de la incertidumbre en lo que la gente nos cuenta es uno de los motivos por los que los profesionales del marketing invierten tanto dinero en medirlo directamente. Es, por ejemplo, por lo que la auditoría y los paneles de identificación de los productos comprados han sido por lo general más numerosos que las encuestas actitudinales.

A este respecto, es importante saber el papel que las nuevas tecnologías están desempeñando en la medición directa. Extraer información de bases de datos y del panel de identificación de los productos comprados es una forma anticuada (pero muy importante) de medir lo que las personas hacen cuando realizan operaciones bancarias o salen de una tienda. Pero cada vez somos más capaces de medir la actividad cerebral, o hacer un seguimiento de lo que la gente mira siguiendo sus ojos. Y, por supuesto, internet es una fuente masiva de conducta directa.

De nuevo, Synovate está bien posicionada para aprovechar las formas en que estas nuevas tecnologías transformarán la medición de las encuestas. En concreto, el énfasis de Synovate Laboratories en la *eficacia de la medición y la validez mejorada a nivel individual*, combinado con la idea de un *adecuado seguimiento que se adapte*, nos equipa para participar en el futuro con seguridad. Nuestros grandes estudios de *tracking* llevados a cabo por los equipos de Customer Experience de Estados Unidos están entre los mejores ejemplos del mundo de *tracking* adecuado y adaptado.