Sesgos y errores en las pruebas de producto

por Jorge A. Valdés
Evamerc

Como lo hemos presentado en números anteriores, las pruebas monádicas y las comparativas pueden ser enormemente eficientes; aunque, sin duda, también presentan sus propios riesgos. Todo depende de las aplicaciones que de ellas se hagan. A lo largo de este artículo, el autor propone analizar aquellos aspectos que ofrecen la posibilidad de sesgos y errores, mismos que, deben tomarse en cuenta para tomar decisiones en el terreno de las pruebas de producto.

La reflexiones metodológicas de este texto surgen como resultado de otro publicado por Laura Ruvalcaba y que aborda el eterno "dilema de la prueba monádica vs. las comparativa"[1]. En él, la autora concluye que "la vida es percibida de manera monádica, aunque en las decisiones se deben tomar en cuenta todas las cartas de la baraja". Semejante proposición me hizo reflexionar sobre tan importante tema, así que expondré otros elementos de juicio que se refieren a este tópico.

Efectivamente, se trata de un tema que siempre causará polémica y cuya comprensión dependerá de la cultura de cada empresa. Aquéllas orientadas al perfeccionamiento del producto, se inclinan más por las pruebas de tipo comparativo; mientras que las más enfocadas al desarrollo de conceptos, tienden a emplear evaluaciones monádicas. El uso de cualquiera de las dos posibilidades tiene como propósito que el mayor número de consumidores esté satisfecho. No obstante, puede decirse que las primeras son más obsesivas en el desempeño del producto y que, por oposición, las segundas son un poco más tolerantes en este elemento constitutivo del *mix de marketing*.

Por otro lado, existe la tendencia a hacer evaluaciones comparativas cuando no se quieren correr riesgos en el desempeño del producto. Al ser éstas pruebas más ácidas, se pierde la oportunidad de tener un producto competitivo en el mercado, pues se rechazan opciones que pudieran ser bien acogidas por los consumidores.

En cambio, combinando el resultado de ambos tipos de estudios, se pueden obtener resultados de productos exitosos. Veamos cómo.

La experiencia recomienda precaución

Se puede coincidir en que una prueba monádica es más realista y una comparativa es una prueba forzada, ya que no es habitual que el consumidor tenga a la mano varios productos al mismo tiempo para luego decidir cuál es su preferido. ¿Qué hacer, entonces? Organizamos primero, qué nos ofrecen las pruebas monádicas.

Éstas, técnicamente, sólo tienen un tipo de sesgo en su implementación, y es el denominado "sesgo de cortesía", en el cual la gente, como agradecimiento por el producto que les acaba de ser entregado, tiende a calificar alto su desempeño. El riesgo, en estos casos, es de carácter interpretativo, ya que si no se tiene una referencia o norma con la cual comparar, entonces no se podrá saber si el resultado es bueno o malo —o, como coloquialmente se dice, no se puede conocer "¿qué tanto es tantito?" o "¿qué tan poco es poquito?". El problema se complica aún más cuando el mismo estudio se hace en diferentes ciudades y en distintos países, ya que en cada lugar el sesgo puede ser diferente.

En el caso de tener alguna norma, conviene tener cuidado con algunos de sus elementos comunes; recodemos que "lo obvio provoca ceguera". Así, tratemos de ver un poco más allá de los datos, pues sólo así podremos evitar conclusiones erróneas.

Les relato el siguiente caso: en alguna ocasión, a punto de terminar el trabajo de campo, una persona nos pidió incluir en los cuadros del reporte analítico la norma de intención de compra que ellos tenían para comparar los resultados. Al preguntar sobre el origen de tal referencia, la respuesta fue que ése era el único dato del que se disponía.

En la aplicación de las pruebas monádicas debe tenerse en cuidado con las llamadas "pruebas de cortesía" pues a menudo las personas tienden a agradecer el producto que se les ha entregado como cortesía evaluándolo muy alto.

Una vez descrita esta experiencia, no nos queda más que advertir del riesgo que dicha práctica entraña. Si lo que se desea es comparar contra normas, debemos asegurarnos de responder a las siguientes preguntas con fines metodológicos:

1. ¿La escala de la pregunta es del mismo tamaño?
2. ¿En qué posición del cuestionario se encuentra ubicada la pregunta?
3. ¿El marco muestral es el mismo?
4. ¿El tipo de prueba es el mismo?
5. ¿El tamaño de muestra con el que fueron realizados los estudios —o el estudio— para determinar la norma es estadísticamente estable?

Pareciera innecesario, ahondar en el primer punto; sin embargo, llega a suceder que se compara el resultado del *top box* o del *top two boxes* sin tener claro si la escala es de 7 o de 5 puntos o inclusive de 9, por lo que podríamos estar "comparando peras con manzanas" y errar en la conclusión.

El punto número dos, se basa en una afirmación que pudiera parecer extremista: "después de la primera pregunta, todas las demás están sesgadas". No obstante, debemos de considerarla como punto de partida, ya que, conforme se va avanzando en el cuestionario, el sujeto investigado razona más su respuesta. Se da incluso el caso en que se le hacen a éste cuestionamientos que tal vez nunca antes hubieran pasado

No puede afirmarse categóricamente que las pruebas monádicas resultan superiores frente a las comparativas o viceversa. Su elección dependerá de los fines que cada empresa tenga.

por su mente. Si la pregunta crítica de evaluación se hace después de una larga batería de interrogantes, se acumulan más pensamientos de los que regularmente se hacen, por lo cual el resultado global puede presentar un sesgo.

Ante este panorama, el procedimiento aconsejable consiste en colocar las preguntas decisivas al inicio del cuestionario. Así se evitarán posibles condicionamientos. Las interrogantes restantes son de diagnóstico y bien pueden ubicarse después.

El tercer punto es fundamental y, aunque pareciera obvio, debe observarse siempre, pues en ocasiones se omite la aclaración del tipo de público del que proviene la información. Es muy diferente el tipo de respuesta de niños vs. adultos, hombres vs. mujeres, usuarios de una categoría vs. no usuarios, practicantes de alguna actividad vs. no practicantes, por mencionar algunos casos comunes. Es decir, cada grupo tiene un comportamiento diferente para evaluar productos.

El cuarto punto trata sobre los distintos tipos de pruebas, sean éstas monádicas, protomonádicas, monádicas-secuenciales o sólo comparativas.

Las normas metodológicamente más puras son las monádicas, ya que no tienen referencia de otro producto evaluado previamente. El sujeto responde de acuerdo con la experiencia acumulada a lo largo de su vida.

Los sesgos y la manera en que afectan una prueba

Con respecto a las normas monádicas secunciales, conviene señalar que, al probar más de un producto, los resultados estarán sesgados. Con el fin de que esta desviación sea igual en todos los elementos del experimento, se efectúa una rotación, es decir, los productos se alternan con el fin de que cada uno tenga la misma oportunidad de ser evaluado en todas las posiciones posibles.

Hasta este momento, todo parecería marchar según lo programado, porque tenemos control sobre la prueba, pero es pertinente preguntarse de qué manera afecta el sesgo. Lo primero que observamos es que los resultados de una prueba monádico-secuencial obtendrán valores menores que los de una prueba monádica simple.

Para demostrar cómo se da esto, tomaré como ejemplo un estudio de productos de consumo realizado el año pasado. Se trató de una prueba ciega monádico-secuencial de cuatro elementos. Se realizaron tres celdas pareadas, donde los productos B, C y D fueron comparados con el A. Tal y como fue hecho el diseño, 50% del producto de cada celda se probó en primera posición y el restante en segunda. Al final se aplicó una prueba comparativa. Este tipo de metodología permite analizar los resultados monádicos, monadico-secuenciales y los comparativos.

Veamos los resultados gonádicos. Para ello, sólo seleccionaremos aquellos casos que fueron probados en primera posición de cada celda —no existe sesgo de orden—, tal como se aprecia en la siguiente tabla:

NOTA: las letras indican diferencias significativas vs. las columnas indicadas

	Prod A	Prod B	Prod C	Prod D
	(a)	(b)	(c)	(d)
	%	%	%	%
TOP BOX	20	40 acd	21	20
TOP TWO BOXES	80	89 ad	84	76
MEAN	3.9	4.25 acd	3.98	3.85
S.D.	0.88	0.76	0.78	0.91

Parece una obviedad, pero no lo es: si no se toma en cuenta el tipo de público del que provienen las respuestas, el resultado será siempre confuso. Cada universo ofrece información muy particular.

El diseño de esta prueba permite hacer una comparación de los resultados de los cuatro productos evaluados monádicamente. Se puede observar que el producto con mejor desempeño es el B, el cual registra 40% en *top box*, 89% en *top two boxes* y el valor de la media con 4.25. Asimismo, conviene destacar el nivel de la desviación estándar, el cual se puede considerar relativamente bajo, 0.76. La media es 5.6 veces mayor y asciende a 4.25/0.76.

Resultados de los productos cuando se prueban en segunda posición

	Prod A	Prod B	Prod C	Prod D
	(a)	(b)	(c)	(d)
	%	%	%	%
TOP BOX	13	34 acd	15	17
TOP TWO BOXES	49	72 ad	42	56
MEAN	3.26 c	3.92 acd	3.04	3.38 c
S.D.	1.17	1.05	1.29	1.19

NOTA: las letras indican diferencias significativas *vs.* las columnas indicadas.

Puede advertirse que todos los valores se reducen sustancialmente: el *top box* del producto B, por ejemplo, pasa de 40% a 34%; el *top two boxes* de 89% a 72% y la media, de un elevado 4.25, se reduce a 3.92. La relación de la media con respecto a la desviación estándar desciende de de 5.6 a 3.73.

Podemos decir que el "sesgo de orden", resultado de la ubicación de un producto en segunda posición, disminuye su valor. La varianza aumenta, puesto que ya se tiene una referencia.

Ahora bien, si se utiliza toda la muestra, es decir, se incluyen tanto a aquellos que probaron primero el producto como a quienes lo hicieron después, los resultados se promedian y los valores quedan en un rango determinado por estos dos grupos. Los resultados también tienen un sesgo y su resultado es inferior al de la prueba monádica.

Si nos referimos a una prueba comparativa, la conclusión es la misma: cuando un producto es evaluado en segunda posición, tenderá a preferirse menos que el primero. El sesgo se manifiesta de igual manera, es decir, los resultados del mismo producto cuando se evalúa al final llegan a ser menores que cuando se mide al inicio.

Los resultados que mostramos en este artículo son típicos, por lo cual, les invito a hacer ustedes mismos el ejercicio para que puedan comprobarlo.

Las normas —idealmente— debieran basarse en resultados de pruebas monádicas. Si esto no es posible, resulta absolutamente necesario tener conocimiento de la metodología aplicada.

Para ello, nos remitiremos al quinto punto, el cual trata del tamaño de muestra. Aunque parezca obvio existen casos en los que se toman normas de pruebas con muestras muy pequeñas que conllevan un riesgo de error muy alto. Son, por tanto, normas inestables con las que hay que tener mucho cuidado.

Conclusiones

Todas las pruebas se pueden llevar a cabo, ubicándonos en la cultura de pensamiento de cada empresa y comprendiendo la naturaleza de la información que se tiene al final de cada tipo de evaluación. De este modo no caeremos en interpretaciones que nos lleven a conclusiones incorrectas.

Jorge A. Valdés es director de Evamerc. Economista de profesión, es también maestro en ciencias de la administración. Cuenta en su haber con más de 20 años en el negocio de las encuestas. Le apasiona bucear y practica *Tai Chi* cuando la investigación se lo permite. Es padre de familia de dos hijos porque su esposa así lo decidió.

1 V. Datos, Diagnósticos y Tendencias, Boletín de la AMAI, Enero 2006.

Conviene observar un hecho esencial: que a menor tamaño de la muestra, menor estabilidad en los resultados. Un grupo sólido es aquel lo suficientemente grande como para ser considerado como representativo.