

¿ Están los Datos "Ponderados"? Una Nota Metodológica



Javier Alagón
Estadística Aplicada

¿Cuántas veces hemos estado, al inicio de una presentación, de un estudio, enfrentándonos a la pregunta sobre si los datos se encuentran "ponderados"? ¿En cuántas ocasiones se pone en la balanza la credibilidad del estudio dependiendo de la respuesta que se dé?

¿Qué significa que el estudio se encuentre "ponderado"? Aunque aún no he visto la respuesta en libros de Investigación de Mercados o de Estadística, de manera convencional, un estudio se encuentra "ponderado" si a la información se le han aplicado "factores de expansión" que reproduzcan la estructura de las variables sociodemográficas típicas (sexo, edad y nivel socioeconómico) y/o de variables adicionales tales como "tamaño de región" o "niveles de consumo" de una categoría.

La expansión anterior se realiza en la última etapa del procesamiento de la información, de la misma manera en que se haría una expansión tradicional para un muestreo estratificado: los tamaños de los estratos en la población bajo estudio, se utilizan para "expandir" la información. De esta manera, si en la muestra se entrevistaron hombres y mujeres en proporción 40-60, por ejemplo, la expansión en una población con una proporción 48-52, se realizaría simplemente mediante la utilización del factor 48/40 para "ponderar" las respuestas de los hombres y de 52/60 para las mujeres, ajustando al final el tamaño de muestra resultante de la utilización de dichos factores.

Hasta aquí no habría ningún problema en la interpretación de datos ponderados. Quizás el problema surge cuando se ha realizado un muestreo "parcial", por conveniencia, en ciertos grupos poblacionales, que no necesariamente pueden "expandirse" al total de la población.

La ilustración típica de la situación anterior, es la pretensión de utilización de datos recabados en México, Guadalajara y Monterrey como "representativos" a nivel nacional o a nivel nacional urbano, o para las ciudades más grandes del país. Por supuesto que el recabar información en estas tres ciudades dan más perspectiva que de haberla recabado en una sola, típicamente México. Pero de ahí, a que podamos leer la información como "representativa" a nivel nacio-

nal, o a nivel nacional urbano, hay mucha distancia, aunque se lleve a cabo cualquier esquema de ponderación de la información. No tan sólo eso, un problema recurrente al que nos enfrentamos los investigadores para resolver esta encrucijada es la determinación de los "factores" que representen el peso de las tres ciudades para alcanzar la pretendida "representatividad".

Las salidas convencionales al crucigrama anterior son tres, **ninguna de las cuales puede considerarse como óptima:**

- 1) A cada una de las ciudades se les asigna un factor proporcional al tamaño de la población de interés para cada una de las ciudades. De esta manera, la información de la ciudad de México "pesa" aproximadamente seis veces más que la correspondiente para Guadalajara y Monterrey. En este caso, la población objetivo (universo de referencia) se confina a la población en estas tres ciudades y no se pretende hacer inferencias sobre otros universos.
- 2) A cada una de las ciudades se les asignan factores proporcionales al tamaño de ciertas "regiones" de referencia. Es decir, se hace el supuesto, de que la medición hecha en una ciudad, es "representativa" de toda una región. Por ejemplo, el que los datos recabados en Guadalajara sean "representativos" de una "Zona Occidente" o "Zona Bajío". Este supuesto, puede resultar incorrecto e implicará problemas en la utilización de la información de manera adecuada para estrategias de mercadotecnia.
- 3) Ante la imposibilidad de representar adecuadamente el peso de la información, se asignan factores de acuerdo a un esquema independiente de ponderación. Por ejemplo, el que a la ciudad de México se le asigne un ponderador de 0.5 y a las otras dos ciudades de 0.25. Las lecturas para el total de la información, se hacen a la luz de este supuesto.

Tal y como señalamos anteriormente, ninguna de las tres soluciones puede considerarse como óptima. Quizás la primera incorpora un poco más de realis-

mo; sin embargo, la humildad ejercida en la ponderación muchas veces se desvirtúa al tratar de extrapolar a poblaciones más amplias de las que originalmente se deseaba representar.

En ocasiones, el esquema anterior de ponderación no resulta suficiente para la lectura de todas las variables bajo estudio. Por ejemplo, cualquier pregunta que esté relacionada con el consumo declarado de marcas, aún si se encuentra ponderada bajo los esquemas anteriores de proporcionalidad al tamaño de la población, no reproducirá el ordenamiento típico de marcas de fuentes externas de información (basadas en ventas y auditorías de canales o tiendas), a menos que se incorpore información sobre estadísticas de consumo.

Por ejemplo, para el mercado de tequilas, quizás resulte insuficiente (e inclusive, incorrecto) la ponderación a nivel poblacional de las principales variables sociodemográficas: el consumo de tequila puede ser menor (en términos de volumen, no de incidencia o penetración) en ciertos grupos, por ejemplo, entre mujeres o en la región Sureste de la República Mexicana. Si los datos no se ponderan con factores que representen el nivel de consumo de dichos grupos, cierta información del estudio puede verse seriamente distorsionada (estadísticas de consumo declarado de marcas, por ejemplo).

En estas ocasiones, puede resultar de gran utilidad realizar un esquema doble de ponderación: primero, a nivel poblacional, para las variables sociodemográficas básicas y segundo, a nivel consumo, para aquellas variables en donde se tenga información de fuentes externas sólidas.

A pesar de lo anterior, si en un estudio no se realizó un muestreo que incluya TODOS los aspectos de la población bajo interés, NO se podrá hacer inferencias estadísticas válidas para TODA la población. Lo más importante es reconocer, que si por razones presupuestales o logísticas, no se tiene información de todo el grupo objetivo, las inferencias que se realicen del estudio son aplicables a un SUBGRUPO de la población de interés.

Los factores de ponderación deberán ser utilizados únicamente para mejorar el nivel de representatividad de la muestra en este SUBGRUPO, y no pretender jamás, que la ponderación salvará lo que no se hizo con un esquema de muestreo más amplio. La humildad y la honestidad intelectual en este caso, es mucho más importante que el deseo de hacer extrapolaciones para toda la población, provenientes de un esquema incompleto de muestreo.

La solución al problema de representatividad referido en este artículo se encuentra en los siguientes dos puntos:

- 1) El reconocimiento *a priori* de que los niveles de representatividad de la muestra, dependen del tipo de muestreo que se realice.
- 2) Realizar una selección probabilística, no por conveniencia, de los individuos en el estudio. Esto incluye la fase de selección de ciudades o regiones, dependiendo del nivel de representatividad.

Con el cumplimiento de los puntos anteriores, no evitaremos que la primera pregunta que se nos haga en una presentación de un estudio se refiera a la ponderación del mismo, pero sí estaremos mucho más preparados para responderla coherentemente y con honestidad.





EUROPEAN SOCIETY FOR OPINION AND MARKETING RESEARCH

CALENDARIO DE EVENTOS

2002

Congress and Trade Exhibition 2002
Consolidation or Renewal? From Research to Customer Insight
22-25 September; Barcelona Spain

Telecom Conference and Exhibition
20-22 October; Oslo Norway

Worldwide Qualitative Research Conference and Exhibition
10-12 November; Boston USA

Asia Pacific Conference and Exhibition
1-3 December; Singapore

Para recibir mayor información sobre cualquiera de estos eventos contactarse con:

ESOMAR: Tel: +31-20-664.21.41 Fax: +31-20-664.29.22 www.esomar.nl	En México: Ana Cristina Covarrubias Tel: 56.51.48.23 Fax: 55.93.09.29 E-mail: accovas@avantel.net
---	--