

CHAID en Investigación de Mercados, Entendiendo a los Segmentos

Por Manuel González, Millward Brown

En la Investigación de Mercados es de uso común una gran cantidad de técnicas para la exploración de las relaciones entre las variables medidas en los sujetos de estudio. Técnicas básicas como las que emanan de la Estadística Descriptiva, de la Estadística Inductiva, y otras más sofisticadas, como el Análisis de Factores, sirven muy bien para ayudar al investigador a describir o explorar cómo **piensan** o cómo **sienten** aquellos que han respondido una encuesta. Estas técnicas, son de poca utilidad para describir cómo **son** estas personas, no son de mucha ayuda para **segmentar** a la población estudiada en grupos. El Análisis Multivariado sí provee al investigador de mecánicas para este tipo de segmentaciones, por ejemplo, el Análisis de Conglomerados (*Cluster Analysis*). Si bien la técnica es poderosa, presenta al menos tres problemas. Primero: la violación de los supuestos subyacentes (normalidad multivariada de los datos) lleva a serios conflictos de estabilidad en las soluciones y puede invalidar los esfuerzos. Segundo: los grupos formados, si bien estadísticamente diferentes, en no pocas ocasiones, son difíciles de interpretar. Tercero: la técnica no permite formar grupos que maximicen una variable de respuesta, opción deseable en muchos casos para la Investigación de Mercados.

EVOLUCIÓN AL CHAID

En la segunda mitad del siglo pasado, se desarrolló una técnica que logra burlar estos problemas. Está basada en pruebas de significancia chi-cuadrada sobre tablas de contingencia. De manera automática, el algoritmo detecta interacciones entre las variables. De aquí surge el nombre: CHAID es un acrónimo de *Chi-Squared Automatic Interaction Detection*. Como está basado en tablas de contingencia y pruebas chi-cuadrada asociadas, no existen supuestos de normalidad subyacentes. Dado que las variables explicativas siempre provienen de cruces de la información y no de combinaciones de éstas, los grupos siempre son fáciles de interpretar. La técnica necesita que se seleccione una variable a explicar, entonces, los grupos siempre maximizarán la respuesta en ésta o minimizarán un riesgo asociado.

CHAID SOBRE RUEDAS

Veamos el siguiente ejemplo más o menos hipotético. Un fabricante de automóviles deseaba conocer la intención de compra de un coche deportivo. La agencia de investigación aplicó un cuestionario con las siguientes preguntas:

1. Género del entrevistado (Género)
2. Edad en años cumplidos al momento de la entrevista (Edad)
3. Disposición a correr riesgos (Riesgos)
4. Interés por atraer al sexo opuesto (Atraer)
5. Intención de compra del modelo deportivo (Intención)

Cada una de estas variables aceptaba dos respuestas: Hombre o Mujer para el género; de 21 a 30 o de 31 a 40 para la edad; me gusta o no me gusta correr riesgos para la disposición a correr riesgos; me interesa o no me interesa atraer al sexo opuesto y pretendo o no pretendo comprar el auto deportivo. La Tabla 1 muestra los resultados de tabular la variable de interés (Intención) *versus* las otras cuatro medidas recabadas. A simple vista, los resultados lucen difíciles de interpretar dado que la intención de compra está dividida en partes iguales dentro de la muestra.

Un CHAID ayudó a la agencia a encontrar los segmentos de mayor intención de compra. La técnica calculó la estadística chi-cuadrada en cada una de estas tablas de contingencia (ver Tabla 2) encontrando que el mejor predictor para **Pretendo es Disposición a Correr Riesgos** por tener el valor más alto para la estadística.

La salida del CHAID es un árbol de decisión como el mostrado en la Figura 1.

RESULTADOS DEL EJERCICIO

Es claro que la intención de compra para el auto deportivo es mucho más alta (87%) en el segmento formado por las personas que les **gusta correr riesgos** y les **interesa atraer al sexo opuesto** que entre el total de la muestra (50%) o el segmento de personas a los que **no les gusta correr riesgos** (25%). Pudo haber sorprendido que las variables demográficas no desempeñen un papel.

Desde un punto de vista de mercadotecnia, información como ésta es mucho más útil que simples segmentaciones basadas en criterios demográficos. Es mucho más fácil crear una campaña de comunicación llamativa para interesar a personas a quienes les gusta correr

riesgos y desean atraer al sexo opuesto que una para hombres de 25 a 35 años de edad de NSE ABC+.

VENTAJAS DE LA HERRAMIENTA

Como herramienta de segmentación el CHAID presenta bondades importantes. Primero, la técnica no está basada en distribución probabilística alguna. Se fundamenta, únicamente, en pruebas de bondad de ajuste chi-cuadrada sobre tablas de contingencia. Éstas, dada una muestra de tamaño aceptable, casi siempre funcionan bien. Segundo, permite determinar una variable a maximizar; en el caso del ejemplo, la intención de compra del auto deportivo. Esto es muy deseable y no necesariamente está disponible en otras técnicas de segmentación. Tercero, la clasificación en segmentos es siempre fácil de interpretar, lo que no sucede, por ejemplo, con el Análisis de Conglomerados. Cuarto, la técnica garantiza, por la forma en que se aplica, que los segmentos siempre tendrán un sentido estadístico: todos serán diferentes y serán los mejores posibles dados los datos.

EL CHAID, MÁS ALLA DE LA SEGMENTACIÓN

Pero, el CHAID no es sólo una herramienta de segmentación. A partir de los segmentos, puede utilizarse para hacer predicciones. Por ejemplo, sabiendo qué tipo de hogares respondieron bien a promociones en el pasado, con la aplicación de la técnica es posible predecir qué hogares responderán a promociones en el futuro. La técnica permite evaluar riesgos asignando valores de retorno o gasto a los diferentes niveles de las variables de respuesta: ¿cuánto cuesta entregar la promoción y qué retorno recibo si participan?, por ejemplo. También puede utilizarse para entender qué quiso decir del consumidor con las respuestas que da a nuestras encuestas. No es raro que en los cuadros de imagen incluyamos atributos del estilo es **una marca para alguien como yo**. Asociando las respuestas a este atributo con los demás presentes en el mismo cuadro podemos entender qué hace que un yogurt sea **una marca para alguien como yo**.

CONCLUSIÓN

En resumen, si bien no es la única herramienta de segmentación disponible para el investigador de mercado, el CHAID sí es una opción para el entendimiento de segmentos poblacionales, la asignación de riesgos y hasta la comprensión de respuestas. Es una técnica

que permite hacer inferencias y predicciones sin apelar a fuertes supuestos probabilísticas, que se comporta bien en tanto la muestra tenga un tamaño aceptable.