

7

Discusión y conclusiones

7.1. Limitaciones de este estudio

Esta tesis tiene objetivos fuertemente metodológicos, y para poder llevarlos a cabo ha partido de una base de datos típica en Marketing. Eso no quiere decir que el objeto de estudio, la fidelidad, o más concretamente la conducta de redención de puntos en una tarjeta de fidelización sea secundario. Sin embargo puede parecer a la luz de los resultados que hemos profundizado poco en el estudio conductual de la fidelidad.

El Marketing actual, en su enfoque más reciente, CRM, sigue dissociando los registros automatizados de los datos más puramente psicológicos, actitudinales o de preferencias. En la revisión teórica sobre fidelidad llevada a cabo en el capítulo 2 hemos visto cómo en la práctica hay dos campos separados de investigación: uno conductual, en el que se realizan registros automáticos de acciones observables, y otro enfoque actitudinal basado en encuestas o entrevistas. ¿No es posible hacer que ambos converjan? El CRM como filosofía, pero también como conjunto de herramientas tecnológicas que permitan cumplir unos objetivos empresariales, entre los cuales destaca la fidelidad, debiera ser capaz de llevar a cabo esta integración.

Por tanto en nuestro estudio hemos realizado una investigación con datos secundarios a partir de una base de datos en la que podíamos predecir, pero no podíamos explicar. Esta es una de las principales limitaciones de este estudio.

Sin embargo, a la vez, una de las principales novedades es la de proponer una metodología que puede integrar eficazmente todo tipo de datos heterogéneos, puesto que las curvas ROC son independientes tanto de las escalas de medida, como de los orígenes de los datos. De hecho, en Medicina se utilizan para evaluar instrumentos que integran tanto datos objetivos (mediciones biológicas) como subjetivos (valoraciones de los médicos). Una de las aplicaciones más curiosas de las revisadas en el capítulo 3 es un estudio que propone 3 únicas preguntas para detectar alcoholismo (Gordon et al., 2001). Independientemente de su éxito o no en el empeño, lo destacable es que con una metodología firmemente basada en la evaluación de la evidencia empírica, es posible economizar y precisar mucho en nuestros objetivos. El enorme desarrollo de las herramientas de "*screening*" en Medicina prueba que este enfoque es, en la práctica, muy fructífero.

Pero, ¿qué hemos podido comprobar en relación con "la comparación de modelos de curvas ROC", que es el título de esta tesis?

7.2. Modelos de curvas ROC: conclusiones

Las curvas ROC empíricas no cabe duda que son una herramienta potentísima, y que con el esquema de contraste de hipótesis estadísticas sin supuestos de partida en cuanto a la forma de las distribuciones, una metodología a la vez sencilla y potente que ha visto reflejadas estas características en la enorme cantidad de estudios que la utilizan. Bien es verdad que en unos campos más que en otros, pero creemos que esta tesis puede cumplir su papel de avanzar en que un sector en el que pueden jugar un papel importante, como es el Marketing. Sector en el que por cierto es difícilísimo encontrar aplicaciones de las mismas, aun cuando en algunos casos se utilizan los mismos conceptos.

Dicho esto, tenemos también que proporcionar conclusiones sobre el modelo de curvas binormales. Como se ha comentado en la introducción, no fue hasta la aparición del programa NCSS 2004 (Hintze, 2003) que tanto los contrastes estadísticos no paramétricos como la estimación de curvas binormales estuvo disponible en un paquete generalista de estadística. Hasta esta fecha, la estimación se tenía que hacer "a mano", o

mediante el software de Metz, ROCKIT (Gilhuijs, 2000; en <http://xray.bsd.uchicago.edu/krl/index.htm>), que por supuesto tiene bastante inconvenientes para integrarse dentro de los procesos tan tecnificados de análisis en Marketing.

Las curvas ROC binormales aportan mayor potencia a la estimación tanto del área bajo la curva como del análisis coste-beneficio y de valor predictivo. Ello a costa de establecer un fuerte supuesto: ambos grupos deben mostrar una distribución gaussiana. Dependiendo del contexto, este supuesto puede ser fácil o difícil de cumplir, pero en cualquier caso se echa en falta un enfoque de evaluación de la tolerancia de este modelo a desviaciones de la curva normal, tal y como sucede con otras muchas técnicas estadísticas que también suponen la curva normal. El mayor inconveniente para su aplicación será por tanto conseguir demostrar que la distribución sigue una curva normal, aunque existen métodos de hacerlo de forma óptima, nos encontraremos con el escollo de que si se aplican de una manera en el grupo de positivos, puede ser que produzcan efecto contrario al deseado en el grupo de negativos.

El camino es por supuesto la integración de estas técnicas (incluyendo las no paramétricas) en los programas estadísticos. Sin ello será imposible la difusión más allá de conjuntos más o menos amplios de investigadores aplicados.

En relación también con este tema de los modelos, destacamos las palabras de Hintze (2003) en el sentido que hasta que no apareció el libro de Zhou *et al.* (2002) no hubo una recopilación unificada del tema. Y esto es en parte porque el mundo de la estadística parece no haberse enterado de la existencia de estos métodos. Tomemos por ejemplo un manual para médicos fuertemente metodológico, el de Shoukri y Pause (1999) y no encontraremos ni una referencia a las curvas ROC! Este manual habla de la sensibilidad y especificidad, propone un capítulo entero para los estudios de prevalencia, pero ni siquiera menciona alguno de los artículos de Hanley y Mc.Neil, o de Metz, o de Swets. Lo mismo hemos comprobado que sucede en varios campos en los que hemos llevado a cabo una revisión bibliográfica.

Sirva por tanto este trabajo para difundir estas técnicas, después de haber mostrado (que no demostrado) su potencia y aplicación a un campo tan complicado como el de la investigación de mercados.

7.2 La eficacia predictiva en el mundo real

Leyendo a la mayoría de los autores con relevancia en este particular campo de las curvas ROC, una de las principales conclusiones es que no existe, por el momento, y quizá nunca haya una regla simple que permita optimizar la decisión en la tabla de doble entrada que sirve de punto de partida al análisis ROC. Esto es así porque la estimación de costes y beneficios plantea enormes dificultades prácticas, quizá también porque no se trata de un esquema común en varios campos. En particular, hemos comprobado que en Marketing, el objetivo se marca sobre los verdaderos positivos, o los verdaderos negativos, no se incide en las falsas alarmas o en las omisiones. Es común encontrarse resultados que indican que "se ha alcanzado al 80% de la población", aunque quizá no se dice que "a costa de tener un 50% de falsas alarmas", con todos los costes que ello implica.

Conceptos normativos como la tasa de prevalencia pueden no tener tanto sentido en contextos tan abiertos y dinámicos como el Marketing. Hemos comprobado que necesitamos conocer esta tasa, pero también podemos variar entre diferentes tasas "candidatas" (que bien pueden ser objetivos de negocio). Por ejemplo, las reglas de negocio expresadas en la tabla 2.3, que reproducimos a continuación, pueden tener su reflejo en el esquema de análisis de curvas ROC y permitir de esta manera un enfoque estándar y potente para la evaluación de los resultados frente a los objetivos. Como ejemplo de autor que lleva esto a cabo está el excelente artículo de Baesens *et al.* (2004), que debido a su recentísima publicación no ha podido ser incorporado mejor a este trabajo, pero que creemos que marca un hito.

| Tipo de cliente | % | Automatizada | Personalizada | Objetivos |
|------------------------|----------|---------------------|----------------------|---|
| "Top Tier" | 20 | 5% | 95% | Mantener la fidelidad y la conducta de compra incrementando la relación personal. |
| "Golden" | 62 | 40% | 60% | Crear diálogo <i>online</i> y <i>offline</i> con cada individuo. Incrementar la frecuencia de compra y el porcentaje de presupuesto dedicado. |
| "Lower Tier" | 18 | 95% | 5% | Beneficiarse de una interacción automatizada que sea efectiva en costes para incrementar la actitud de fidelidad con los clientes. |

Tabla 7.1 (también 2.3). Clasificación de los clientes en función de su interés (según Johnson y Leger (1999))

Unas palabras más sobre el carácter diferente de los sectores de aplicación, que hemos mantenido entre la Medicina (o la Psicología Clínica) y el Marketing, es la confianza en la tecnología. Nos referimos con esto a la creencia que habrá tecnologías que permitirán solucionar los problemas de una manera lo más automática posible. Este tema continúa la exposición anterior sobre la tasa de prevalencia, etc. Si en Medicina tenemos "screening", pero intentando mantener al máximo la solidez de las conclusiones mediante una validación cruzada de los índices usando una metodología estándar, en Marketing la palabra de moda es "minería de datos". Esto es, algoritmos que naveguen en enormes bases de datos para localizar "la regla perdida".

7.3 Directrices futuras

En cierto modo esta tesis también es una muestra del debate entre el paradigma estadístico tradicional "artesanal" o de análisis "a mano" (si hiciera falta), frente al automatizado y, en cierto modo, "de caja negra". ¿Tiene mayor eficacia predictiva? Ése es el problema, que muchos estudios no se han parado a establecerla. O si lo han hecho, no utilizan una metodología que proporcione resultados comparables, con lo que volvemos a estar un poco como al principio.

En este punto es donde este trabajo es menos representativo del estado de la cuestión. Nos explicamos. No se ha hecho minería de datos como tal, puesto que un árbol de decisión es una técnica relativamente antigua, y ya sabíamos que subóptima frente a las técnicas estadísticas optimizadas, mucho más cuando ya habíamos estudiado con tanto detalle nuestros datos. Pero nuestra principal crítica es ese enfoque de "caja negra", en el que los pasos intermedios dejan de tener explicación, y todo reside en el resultado. Que por otro lado no nos será fácil comprobar o contrastar.

Pongamos por caso el árbol de decisión que se ha realizado en el capítulo 6. Pequeñas variaciones en la base de datos pueden dar lugar a reglas bastante diferentes. ¿De qué nos sirve entonces un modelo que tenemos que cambiar con frecuencia? ¿Es que la ventaja (¿para el analista? ¿para el cliente del analista?) de poder plantear las reglas en lenguaje naturales es mayor que los costes de recálculo, o de pérdida de capacidad predictiva?

Efectivamente, en el mundo actual de enormes bases de datos, y gracias a una tecnología portentosa en cuanto a capacidad de cálculo, el gran problema es definir previamente qué es lo que consideramos por un patrón "interesante", "valioso" o "útil", que será necesariamente multidimensional. Para un problema estadístico tradicional, unos cuantos miles de datos serían suficiente, si no muchos. Pero la tecnología moderna de bases de datos permite mantener bases de millones de registros. Por ejemplo, AT&T tiene 100 millones de clientes, y gestiona 200 millones de llamadas diarias en su red de larga distancia. Se habla no ya de *gigabytes*, sino de *terabytes* de información.

Los ficheros ya no son entonces una única base de datos en forma de matriz, sino múltiples bases de datos interrelacionadas, con una estructura jerárquica, por ejemplo, y que no permiten una visualización estándar.

Hand (1998) señala consecuencias más sutiles del enorme tamaño de bases de datos:

"En el pasado, en muchas de las situaciones en que han trabajado los estadísticos, el problema ha sido el de falta de datos más que el de abundancia. Por eso la estrategia ha sido fijar el error de tipo I en un valor razonable, 1%, 5% ó 10%, y recoger suficientes datos como para proporcionar potencia suficiente a las hipótesis alternativas. Sin

embargo, cuando los datos existen en superabundancia, esta estrategia es cuestionable. Los resultados de estas pruebas llevarán a evidencias suficientemente fuertes de que existen hasta los más pequeños errores, efectos que pueden no tener ningún sentido práctico. En vez de "significación estadística" quizá nos debamos concentrar en "la significación sustantiva": ¿el efecto encontrado es interesante o no?"

Otro problema común en el mundo actual son los "datos contaminados". Lo hemos visto en este trabajo, en el que ha habido que realizar ajustes manuales para solucionar pequeños problemas puntuales, pero que para algunas técnicas pueden tener consecuencias nefastas. Unos datos limpios son un requisito para muchos análisis estadísticos. Hay libros enteros, de todo, para solucionar la detección de *outliers* y datos *missing*. Una solución es volver a los datos originales, con procedimientos que permitan sortear estos inconvenientes.

Esto afecta al análisis de curvas ROC, por supuesto más al modelo binormal que al modelo empírico, pero en este trabajo hemos mostrado que con unas técnicas relativamente sencillas, y creemos que fácilmente comprensibles a personas que no tienen por qué tener un entrenamiento muy especializado en Estadística, se puede desarrollar una metodología de predicción para un problema complicado como el que se ha estudiado. Insistimos ahora que estas técnicas se tienen que integrar en el trabajo cotidiano no ya de académicos o médicos, sino también de analistas en campos como la investigación de mercados.