Un modelo estadístico para la elaboración de una tarifa de daños a automóviles de turismo

Por Francisco REINA PROCOPIO y Agustín SANS Y DE LLANOS

I

Nota previa.

La presente Comunicación tiene un simple carácter informativo. Y ello porque se limita a dar a conocer una de las soluciones teóricas, con su correlativa realización práctica, ofrecida recientemente al Seguro español, en su vertiente de mayor importancia: la del Seguro Voluntario de Vehículos a motor (categoría 1.ª: Turismos).

No podía escapar España, ni su Institución Aseguradora, a la problemática socioeconómica que plantea el rápido crecimiento del Parque automovilístico. Por ello, el tratamiento actuarial que se venía dando a la determinación de las primas del Seguro de Daños era preciso abandonarlo, porque para pasar a una Tarifa con tendencia "personalizadora" del riesgo, lo que implica una valoración de sus principales factores integrantes, era necesario afrontar el estudio de un modelo capaz de medir aquéllos.

Como se verá, la solución hallada no es más que la aplicación de la regresión múltiple. Pero al lector no pasará por alto que la experiencia es interesante y que su materialización exige serios esfuerzos.

II

El modelo estadístico.

El modelo estadístico-actuarial seguido para el estudio está basado, de un lado, en el análisis de la frecuencia media de siniestros por vehículo y año, y, por otro lado, en el análisis del valor medio del coste del siniestro.

Para estos análisis se tomó el esquema de la regresión y correlación de la Estadística Matemática.

Esto ha permitido buscar grupos homogéneos de riesgos, es decir, agrupar características de vehículos, personas, y otras varias, en las que se presentan iguales riesgos. Ello permite que, cuando se calcula la prima, ésta corresponda al verdadero riesgo corrido por el ente asegurador.

Dentro de cada uno de los grupos homogéneos de riesgos, actúa indudablemente el azar. Esta influencia del azar es recogida tomando el valor medio correspondiente al conjunto de valores que aparecen en cada uno de esos grupos citados. En resumen, se han separado del fenómeno de la siniestralidad las causas del azar, de las causas condicionantes del riesgo.

Como causas condicionantes del riesgo se tomaron en principio las siguientes con su grado de variabilidad:

— *Variables que hacen referencia a las características del vehículo*, como son:

Potencia, antigüedad, peso, vía o ancho, y valor.

— *Variables que hacen referencia a las características del conductor*, tales como:

Edad, antigüedad del permiso de conducir, y profesión.

— *Variables que hacen referencia a otras características*, tales como:

Uso del vehículo y Zona por la que circula.

Con estos tres tipos de variables en juego y en base a una muestra representativa de las Carteras de nueve importantes Compañías de Seguros, se procedió a efectuar los cálculos según el modelo matemático que se expone a continuación.

Al estudiar el riesgo de Daños propios, dentro del marco de las variables consideradas, se mide cuál es la importancia de cada una de ellas, utilizándolas al mismo tiempo para elaborar la Tarifa de primas.

Llamamos

x_1 al número de siniestros por póliza y año,

x_2 al valor de la indemnización por siniestro,

Se admite que x_1 y x_2 son funciones del siguiente conjunto de variables:

x_3 = edad del conductor
x_4 = antigüedad carnet
x_5 = profesión (cuantificada)
x_6 = potencia vehículo
x_7 = peso vehículo
x_8 = vía o ancho vehículo
x_9 = antigüedad vehículo
x_{10} = valor vehículo
x_{11} = Zona (cuantificada)
x_{12} = Uso (cuantificada)

Decimos que,

$$x_1 = f_1(x_3, x_4 \ldots x_{12})$$
$$x_2 = f_2(x_1, x_3, x_4 \ldots x_{12})$$

Y admitimos para nuestros efectos prácticos que estas funciones son de tipo lineal, es decir:

$$x_1 = b_{1,3}\, x_3 + b_{1,4}\, x_4 + \ldots + b_{1,12}\, x_{12} + k_1 \qquad [1]$$
$$x_2 = b_{2,1}\, x_1 + b_{2,3}\, x_3 + \ldots + b_{2,12}\, x_{12} + k_2 \qquad [2]$$

Sabido es que el plano de regresión mínimo-cuadrático correspondiente a (desarrollamos para [1]):

$$x_1/x_3\ x_4\ \ldots\ x_{12}$$

viene dado por

$$\psi(b_{1,3}, b_{1,4}, \ldots, b_{1,12}) =$$

$$= \sum_{i=1}^{n_1} [x_{1i} - (b_{1,3} x_3 + b_{1,4} x_4 + \ldots + b_{1,12} x_{12} + k_1)]^2 \frac{1}{n_1} =$$

$$= \text{mínimo} \qquad [3]$$

Para calcular los b_{ij} resolveremos el sistema de ecuaciones normales que aparece al igualar a cero las derivadas parciales respecto a las b_{ij} y k_1

$$\frac{\partial \psi}{\partial k_1} = -2 \sum_{i=1}^{n_1} [x_{1i} - b_{1,3} x_3 - b_{1,4} x_4 - \ldots -$$

$$- b_{1,12} x_{12} - k_1] \frac{1}{n_1} = 0$$

de donde

$$k_1 = \bar{x}_1 - b_{1,3} \bar{x}_3 - \ldots - b_{1,12} \bar{x}_{12}$$

que sustituyendo en [3] brinda:

$$\psi = \sum_{i=1}^{n_1} [(x_{1i} - \bar{x}_1) - b_{1,3}(x_{3i} - \bar{x}_3) - \ldots -$$

$$\ldots - b_{1,12}(x_{12,i} - \bar{x}_{12})]^2 = \text{mínimo}$$

Estableciendo las siguientes derivadas:

$$\frac{\partial \psi}{\partial b_{1,3}} = -2 \sum [(x_{1i} - \bar{x}_1) - b_{1,3}(x_{3i} - \bar{x}_3 - \ldots -$$

$$- b_{1,12}(x_{12,i} - \bar{x}_{12})] (x_{3i} - \bar{x}_3) \frac{1}{n_1} = 0$$

....

$$\frac{\partial \psi}{\partial b_{1,12}} = -2 \sum [(x_{1i} - \bar{x}_1) - b_{1,3}(x_{3i} - \bar{x}_3) - \ldots -$$

$$- b_{1,12}(x_{12,i} - \bar{x}_{12})] (x_{12,i} - \bar{x}_{12}) \frac{1}{n_1} = 0$$

Reduciendo y llamando a

$$\sum_{i=1}^{n1} (x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k)\frac{1}{n_1} = l_{jk}$$

tenemos

$$\left.\begin{array}{l} b_{1,3}\, l_{3,3} + \dots + b_{1,12}\, l_{12,3} = l_{1,3} \\ b_{1,3}\, l_{3,4} + \dots + b_{1,12}\, l_{12,4} = l_{1,4} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ b_{1,3}\, l_{3,12} + \dots + b_{1,12}\, l_{12,12} = l_{1,12} \end{array}\right\}$$

Resulta, por la regla de Kramer,

$$b_{1,h} = \frac{\begin{vmatrix} l_{1,h} & \dots\dots & l_{12,h} \\ l_{1,4} & \dots\dots & l_{12,4} \\ \vdots & \dots\dots & \vdots \\ \vdots & & \vdots \\ \vdots & & \vdots \\ l_{1,12} & \dots\dots & l_{12,12} \end{vmatrix}}{\begin{vmatrix} l_{3,h} & \dots\dots & l_{12,h} \\ l_{3,4} & \dots\dots & l_{12,4} \\ \vdots & \dots\dots & \vdots \\ \vdots & & \vdots \\ \vdots & & \vdots \\ l_{3,12} & \dots\dots & l_{12,12} \end{vmatrix}} = \frac{-L_{1h}}{L_{11}}$$

siendo L_{jk} los adjuntos del determinante de momentos:

$$L = \begin{vmatrix} l_{1,1} & l_{1,3} & \dots & l_{1,12} \\ l_{3,1} & l_{3,3} & \dots & l_{3,12} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ l_{12,1} & l_{12,3} & \dots & l_{12,12} \end{vmatrix}$$

El hiperplano de regresión mínimo-cuadrática buscado para [1] de x_1 sobre $x_3, x_4, \ldots, x_{12}$ será:

$$x_1 = \bar{x}_1 - \frac{L_{1,3}}{L_{1,1}} (x_3 - \bar{x}_3) - \frac{L_{1,4}}{L_{1,1}} (x_4 - \bar{x}_4) - \ldots - \frac{L_{1,12}}{L_{1,1}} (x_{12} - \bar{x}_{12})$$

De igual forma, para [2] sería:

$$x_2 = \bar{x}_2 - \frac{L_{2,1}}{L_{2,2}} (x_2 - \bar{x}_2) - \frac{L_{2,3}}{L_{2,2}} (x_3 - \bar{x}_3) - \ldots - \frac{L_{2,12}}{L_{2,2}} (x_{12} - \bar{x}_{12})$$

Se llama la atención sobre el hecho de que n_1 y n_2 correspondientes a [1] y a [2] no son iguales, ya que n_1 representa el número de pólizas y n_2 el número de siniestros, respectivamente, de la muestra.

Conocidos los dos hiperplanos, podemos ir prediciendo los valores de x_1 y x_2 fijando unas variables y haciendo que las restantes tomen valores sucesivos.

Cálculo de la prima con una franquicia a deducible de todo siniestro

La segunda distribución, es decir, la del valor de la indemnización se ajustó por la distribución logarítmico-normal, es decir,

$$L(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_0 \frac{1}{u} e^{-\frac{1}{2\sigma^2}(\log u - \mu)^2} du$$

Haciendo el cambio $y = \log x$ se tiene:

$$L(y) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2\sigma^2}(v - \mu)^2} dv$$

y se ha transformado en Normal de parámetros μ y σ.

Para calcular los momentos de la distribución logarítmico-normal citada tenemos:

$$E[x^s] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sy}\, e^{-\frac{1}{2\sigma^2}[y-\mu]^2}\, du$$

Multiplicando y dividiendo por $e^{s\mu}\, e^{\frac{1}{2}s^2\sigma^2}$ obtenemos la expresión:

$$E[x^s] = \frac{e^{s\mu+\frac{1}{2}s^2\sigma^2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{sy}\cdot e^{-\frac{1}{2\sigma^2}[y-\mu]^2}}{e^{s\mu+\frac{1}{2}s^2\sigma^2}}\, dy =$$

$$= \frac{e^{s\mu+\frac{1}{2}s^2\sigma^2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{s(y-\mu)-\frac{1}{2}s^2\sigma^2}\, e^{-\frac{1}{2\sigma^2}[y-\mu]^2}\, dy =$$

$$= \frac{e^{s\mu+\frac{1}{2}s^2\sigma^2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(y\mu)^2-2\sigma^2 s\ (y-\mu)+s^2\sigma^4]}\, dy =$$

$$= \frac{e^{s\mu+\frac{1}{2}s\sigma^2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(y-\mu-s\sigma^2)}\, dy =$$

$$= e^{s\mu+\frac{1}{2}s^2\sigma^2}$$

que para S = 1 —momento ordinario de primer orden— se convierte en

$$E[x] = e^{\mu+\frac{1}{2}\sigma^2}$$

y por tanto la varianza será:

$$D^2[x] = E[x^2] - [E(x)]^2 = e^{2\mu+\frac{1}{2}4\sigma^2} - e^{2\mu+\sigma^2} =$$
$$= e^{2\mu+\sigma^2}(e^{\sigma^2}-1) = [E(x)]^2(e^{\sigma^2}-1)$$

para ajustar por el método de los momentos esta distribución, se igualará la media y la varianza obtenida de los datos de la muestra a

$$E[x] = e^{\mu+\frac{1}{2}\sigma^2}$$
$$D^2[x] = [E(x)]^2(e^{\sigma^2} - 1)$$

respectivamente.

Una vez obtenida la distribución ajustada, en la que aparecerán μ y σ con sus valores numéricos, se procederá a truncar la distribución para obtener la prima correspondiente a la franquicia a.

Es sabido que la función de densidad de la distribución truncada será:

$$f_t(x) = \begin{cases} 0 & \text{para } x \leqq a \\ \dfrac{f(x)}{F(b) - F(a)} & \text{para } a < x \leqq b \\ 1 & \text{para } x > b \end{cases}$$

Como nos interesa truncarla solamente para el valor a de la franquicia, b será ∞ y por tanto $F(\infty) = 1$.

$$f_t(x) = \begin{cases} 0 & \text{para } x \leqq a \\ f(x) \dfrac{1}{1 - F(a)} & \text{para } x > a \end{cases}$$

Habrá que calcular ahora el valor de la función de distribución en el punto a, es decir,

$$L(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^a \frac{1}{u} \cdot e^{-\frac{1}{2\sigma^2}(\log u - \mu)^2} du$$

Para encontrar este valor y poder buscar en las tablas de la distribución Normal efectuaremos, como hicimos antes, el cambio $y = \log x$ convirtiendo la distribución en Normal de parámetros μ y σ

$$L(\log a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log a} e^{-\frac{1}{2\sigma^2}(v-\mu)^2} dv$$

y tipificando la variable la convertimos en Normal (0,1);

$$t = \frac{v-\mu}{\sigma}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\log a - \mu} e^{-\frac{1}{2}t} \cdot dt$$

valor que encontramos fácilmente en las tablas de la Distribución Normal, y, por tanto, el de

$$f_t(x) = \frac{1}{1 - F(a)} f(x) \quad \text{para} \quad x > a$$

Una vez obtenidos estos valores tendremos que calcular la media de la nueva distribución truncada que ha resultado.

Antes teníamos para obtener los momentos ordinarios

$$E[x^s] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sy} e^{-\frac{1}{2\sigma^2}[y-\mu]^2} dy$$

y ahora tendremos, poniendo la función de densidad de la distribución truncada obtenida:

$$E_t[x^s] = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{1 - F(a)} \int_{-\infty}^{\infty} e^{sy} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy$$

que por lo mismo de antes nos lleva a

$$E_t[x] = \frac{1}{1 - F(a)} \cdot e^{\mu + \frac{1}{2}\sigma^2} = \frac{1}{1 - F(a)} E[x]$$

por lo que *la media de la distribución con franquicia es igual a la de la distribución completa multiplicada por*

$$\frac{1}{1 - F(a)}$$

valor este último obtenido anteriormente.

Como la franquicia la tenemos que restar del importe de todos los siniestros que nos han quedado en la distribución truncada, el valor que tendremos para calcular la prima con franquicia será:

$$E_t[x] - a$$

puesto que

$$E[x - a] = E[x] - a$$

En estas condiciones, llamando *A* a la media de la distribución sin truncar de las indemnizaciones, la media de la distribución truncada será:

$$\frac{A}{1 - F(a)} - a = \bar{x}_2^a$$

Rehaciendo la distribución de frecuencias, se calculará su nueva media, x_1^a.

Por tanto, la prima con franquicia de nivel *a* será:

$$P = \bar{x}_1^a \; \bar{x}_2^a$$

III

Realización práctica.

El primer paso era el de obtener información estadística suficiente. Elegido como período de observación el año 1965, hubo de hacerse una consulta probabilística a la realidad mediante muestreo aleatorio estratificado con afijación proporcional del parque asegurado, a fin de poder inferir a través de los datos suministrados por la muestra los parámetros del modelo estadístico.

Los datos obtenidos para cada póliza fueron los siguientes:

— Entidad aseguradora.

— Número de Póliza.

— Número de matrícula.

— Marca y modelo del vehículo.

— Uso: se distinguió entre "Particular" (individual), "Particular" (empresas), "Servicio público" (con taxímetro), "Servicio público" (sin taxímetro) y "Auto-Escuela".

— Zona: se optó por el criterio del lugar habitual del garaje.

— Fecha de nacimiento.

— Fecha del permiso de conducir.

— Profesión: se establecieron cuatro grandes grupos genéricos en función del grado de utilización del vehículo.

— Frecuencia de siniestros e importe pagado (o estimado), de cada uno de ellos.

Todos estos datos fueron reflejados en una ficha cumplimentada manualmente y que constituyó el documento base para su perforación en tarjetas (dos para cada póliza).

Con independencia de ello se perforaron fichas maestras con los valores de las variables correspondientes a las características de cada uno de los tipos de turismos que formaron parte de la muestra. Con estos dos grupos de fichas perforadas, y en base al modelo matemático, los técnicos de la Casa I. B. M. pasaron a la elaboración de los siguientes programas de cálculo:

1.° Cálculo de los valores modales de las variables, tomando los vehículos característicos del Parque Nacional, y tres Zonas, que fueron: Barcelona, Madríd y Resto de España.

2.° Cuantificación de las variables cualitativas: Zona, Uso y Profesión, según criterio de la frecuencia media.

3.° Lectura de la muestra completa y cálculo de las medias de las variables: frecuencia y siniestralidad.

4.° Cálculo de los elementos de los determinantes de los momentos y sus adjuntos. A partir de ello se obtuvieron los coeficientes (b_{ij}) de las ecuaciones de los Hiperplanos de Regresión.

5.° Perforación de dichas muestras (cada una de las cuales hacía referencia a las características de un solo tipo vehículo) y cálculo de la Prima que corresponde a cada vehículo. Estas primas aparecieron en un gran tabulado.

6.° Cálculo de la franquicia según el modelo matemático expuesto.

7.° Tabulado de los descuentos correspondientes a cada nivel de franquicia prefijado.

Señalamos algunas peculiaridades del programa número 1 para dar idea de su contextura.

Para la variable "Potencia del vehículo" se establecieron 5 contadores con el siguiente escalado:

1.ª 00 a 07 HP.
2.ª 08 a 11 HP.
3.ª 12 a 15 HP.
4.ª 16 a 20 HP.
5.ª 21 en adelante HP.

Para la variable "Peso", se tomó el escalado con 6 contadores:

1.º 0 a 600 Kg.
2.º 601 a 800 Kg.
3.º 801 a 1.000 Kg.
4.º 1.001 a 1.500 Kg.
5.º 1.501 a 2.000 Kg.
6.º 2.001 en adelante Kg.

Para la variable "Vía o ancho" se tomaron 3 contadores con el escalado

100-125 centímetros
126-150 centímetros
De 151 en adelante

El "Valor del vehículo" exigió 9 contadores, correspondientes a la escala de valores:

60 — 80 — 100 — 120 — 180 — 250 — 350 — 500 — Más de 500
(Referido a 1.000 pesetas)

Finalmente, para la variable "Uso" bastaron 5 contadores, correspondientes a los cinco usos elegidos.

El cálculo electrónico, realizado por los computadores IBM 360/40 y 1401, y sin cuyo apoyo fundamental era inútil el modelo, obtuvo los determinantes de momentos a partir de los cuales pudieron tenerse todos los adjuntos y hallar los coeficientes b_{ij} buscados.

Las ecuaciones obtenidas por el ordenador fueron:

Primer Hiperplano.

x_1 = frecuencia de siniestralidad

$$x_1 = -\ 0{,}00025\ x_3 - 0{,}02064\ x_4 + 0{,}1840\ x_5 + 0{,}04312\ x_6 - 0{,}00133\ x_7 + 0{,}02497\ x_8 - 0{,}02932\ x_9 - 0{,}00105\ x_{10} + 1{,}0792\ x_{11} + 0{,}3883\ x_{12} - 2{,}66128$$

Segundo Hiperplano.

x_2 = valor medio de la indemnización

$$7x_2 = -\ 101{,}845\ x_1 + 7{,}612\ x_2 + 23{,}156\ x_4 + 268{,}47\ x_5 - 99{,}627\ x_6 + 0{,}830\ x_7 + 3{,}110\ x_8 - 31{,}676\ x_9 + 13{,}649\ x_{10} - 1.205{,}710\ x_{11} - 257{,}284\ x_{12} + 2.511{,}903$$

El paso al cálculo de las primas puras como producto de x_1 y x_2, para cada uno de los vehiculos, se realizó a través de nuevas fichas maestras, cada una de las cuales incluía los siguientes datos: Marca y Modelo; Peso del Vehículo; Vía o ancho; Potencia, y Valor, correspondientes a los turismos que se deseaba tarificar.

Dando a estas variables, en los hiperplanos, los valores referentes a un vehículo por medio de la ficha maestra, quedaban aún por determinar las variables distintas a las de las características del vehículo. Se combinaron, entonces, todos los valores posibles de tales variables, dando lugar en cada caso a una Prima que aparecía impresa en un Tabulado o Parrilla, de la que se reproduce en Anexo su estructura.

Huelga decir que, dado el gran número de turismos, las Parrillas resultaron de gran extensión. Ello, unido al escaso influjo de algunas de las variables elegidas "a priori", condujo, en el deseo lógico de simplificación, a suprimir en la presentación comercial las variables de poca influencia.

Es evidente que este estudio no es completo por cuanto no se incluye en él la estimación de los parámetros de los hiperplanos, ni la bondad de la hipótesis de linealidad efectuada por razones de tipo práctico. Pero se subraya el carácter informativo de esta Comunicación, que aspira a dar a conocer la solución dada para el Seguro español en un plazo de realización extremadamente corto, y además, dada la necesidad de presentarla dentro del plazo marcado por la Comisión Organizadora del Congreso, que coincidió con el de ejecución del trabajo.

Del estudio de los dos aspectos señalados, o sea estimación de los caminos intuídos, de superposición de dos leyes de siniestralidad y quizá otros que esclarezcan más el fenómeno estudiado.

Anexo.

Estructura de la tarifa (antes de su simplificación)

Marca y modelo	Antigüedad	Turismos — Prima de datos						ZONA 3					USO 1				
		Primera profesión				Segunda profesión				Tercera profesión				Cuarta profesión			
		Carnet 1		Carnet 2													
		Edad 1	Edad 2	Edad 1	Edad 2	Edad 1	Edad 2	Edad 1	Edad 2	Edad 1	Edad 2	Edad 1	Edad 2	Edad 1	Edad 2	Edad 1	Edad 2
Volkswagen 1200 Exp	2	9111	9440	7828	8075	8518	8831	7240	7471	8691	9008	7411	7647	8178	8482	6901	7124
	7	7867	8159	6566	6776	7310	7587	6012	6207	7472	7753	6173	6373	6990	7258	5695	5881
Volkswagen 1500	2	9731	10027	8069	8282	9091	9371	7432	7631	9277	9562	7618	7821	8722	8993	7066	7256
	7	8366	8625	6685	6862	7761	8005	6084	6247	7937	8186	6259	6426	7413	7649	5739	5893
Volkswagen 1500 S	2	9259	9533	7516	7709	8623	8883	6885	7062	8808	9072	7069	7251	8258	8508	6522	6691
	7	7898	8137	6137	6294	7299	7522	5542	5683	7473	7701	5715	5861	6954	7169	5200	5332

Carnet 1 = Menos de 4 años.
Carnet 2 = Más de 4 años.
Antigüedad 2 = Menos de 4,5 años.
Antigüedad 7 = Más de 4,5 años.

Edad 1 = Menos dc 30 años.
Edad 2 = Más de 30 años.