# Modelos estadísticos para los costos en carteras colectivas de Seguros de salud*

Arelly Ornelas Vargas**
Montserrat Guillén Estany***
Manuela Alcañiz Zanón****

**SUMARIO**



Fecha de recepción: 3 de Agosto de 2011
Fecha de aceptación: 28 de Noviembre de 2011

* El artículo resulta de gran interés en el ámbito científico de la aplicación de técnicas estadísticas avanzadas a la evaluación de modelos actuariales de riesgo. En concreto aplica modelos de regresión lineal múltiple generalizados para determinar la influencia que tienen sobre los seguros de asistencia sanitaria, dos factores de riesgo fundamentales: la edad y el sexo del asegurado.

** Asistente de investigación, Universitat de Barcelona, MA Estadística e Investigación Operativa, BA Ciencias Actuariales.
Correo electrónico: arelly.ornelas@ub.edu

*** Catedrática de Universidad, Universitat de Barcelona, PhD Ciencias Económicas y Empresariales, MSc Matemáticas.
Correo electrónico: mguillen@ub.edu

**** Profesora Titular de Universidad, Universitat de Barcelona, PhD Ciencias Económicas y Empresariales, MSc Matemáticas.
Correo electrónico: malcaniz@ub.edu

## Resumen

El objeto de este análisis es la utilización de factores de riesgo en un modelo que analice los costes de los siniestros. Desde una perspectiva teórica tradicional, el cálculo de la prima pura se suele basar en la predicción de la frecuencia esperada de siniestralidad, multiplicada por el costo medio de un siniestro. Es habitual que la frecuencia esperada se modelice a partir de factores de riesgo de los asegurados utilizando modelos de regresión o bien mediante la segmentación de la cartera de asegurados en grupos de riesgo. En este trabajo se analiza el comportamiento de los costos de siniestros en una cartera de seguros de salud utilizando un modelo causal que permita valorar la influencia de factores tales como la edad y el sexo no solo en la frecuencia esperada de siniestralidad sino también en el monto del siniestro cuando este se produce. La metodología empleada está basada en modelos lineales generalizados con efectos aditivos o multiplicativos. Se concluye que tanto la edad como el sexo del asegurado son factores de riesgo con influencia significativa en la severidad de las reclamaciones en seguros de salud y enfermedad, que permiten una mejor predicción de la severidad del siniestro que la utilización de la media del coste. Se apunta a la necesidad de investigar cómo integrar los modelos de frecuencia y severidad en un mismo marco.



**Palabras clave:** Seguros de salud, severidad de los siniestros, tarificación.

**Palabras clave descriptor:** Modelos estadíticos, Seguros de salud.



## Abstract

The aim of this analysis is to use risk factors in a model to analyze claim severity data. Pure premium calculation is usually based on the prediction of the expected claim frequency multiplied by the average loss. Typically, risk factors have been used by insurance companies for modeling expected claim frequency by using regression models or by segmenting the portfolio of insured risk groups. This paper analyzes the behavior of loss cost in health insurance portfolios by using a causal model designed to assess the influence of factors such as age and sex not only in the expected number of accidents but also in the claim cost. The methodology is based on generalized linear models with additive or multiplicative interaction between the rating factors. We conclude that both age and sex are risk factors with significant influence in Health Care and Medical Insurance allowing better prediction of the severity of the incident instead of averaging the cost for all policyholders. The next goal is to investigate how to integrate models of frequency and severity simultaneously.



**Key words:** health insurance, claims severity, pricing.

**Key words plus:** Statistical models, Health insurance.

## 1. INTRODUCCIÓN

La selección de los factores de riesgo es una de las principales etapas de la modelización actuarial. La tarificación tiene por objeto separar a los asegurados de tal manera que las características de grupos homogéneos de asegurados se correlacionen con su siniestralidad; además, esos grupos deberían explicar un alto porcentaje de la varianza de los montos reclamados.

En este artículo se pretende estimar las primas de un seguro de salud, de modo que las primas obtenidas representen la mejor predicción del monto de los siniestros que ocurrirán en el portafolio; es decir, las primas deben ser acordes a los tipos de riesgo que se estén suscribiendo. Así se conseguirá incrementar la rentabilidad de la cartera cobrando primas suficientes, esto es, tarifas cuyas primas netas permitan hacer frente a pagos de siniestros, así como a las obligaciones derivadas de la operación del seguro, tales como las comisiones de agentes y gastos de administración.

En primer lugar, se presenta de manera general el método del sesgo mínimo descrito en Ismail y Jemain[1]. En segundo lugar, se detalla el análisis realizado aplicando el método descrito a datos reales de una aseguradora y además se muestra la aplicación para datos públicos concernientes a las aseguradoras que ofrecen productos en el ramo de Salud en México. A partir de los valores estimados de la severidad se procede al cálculo de la prima pura para cada clase identificada. En la última parte se efectúa una discusión y se apuntan algunas líneas de trabajo futuras.

1 Ismail, N. y Jemain, A.A., *Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation*. Casualty Actuarial Society Forum, Primavera, 2005, págs. 367-394.

## 2. NOTACIÓN Y METODOLOGÍA

La prima que cobrará la aseguradora a sus clientes se calcula considerando las características del riesgo que se cubre; para esto se deben identificar los factores que afectan tanto a la frecuencia como a la severidad de las reclamaciones. Así, se asigna la misma prima a los individuos que tengan un gasto esperado similar. Por tanto, para llevar a cabo una buena tarificación, generalmente se involucran tanto métodos estadísticos como el juicio profesional[2].

Una vez definidos los grupos de riesgo que mejor reflejen el costo de los siniestros, se puede calcular la severidad y el número de reclamaciones observadas para cada uno de ellos, que denotaremos como $(c_i, y_i)$, respectivamente, con $i = 1,2,\ldots, n$, donde i representa la categoría de pólizas o grupo de riesgo. Estas variables serán las empleadas en el ajuste del modelo, teniendo en cuenta que el costo agregado esperado resulta de multiplicar $y_i \times c_i$.

Escribiendo el modelo de regresión a ajustar en notación vectorial se tiene **c** el vector con los montos medios de las reclamaciones como la variable de respuesta, el vector **y** con el recuento de las reclamaciones como los pesos y la matriz de variables explicativas **X** donde $x_{ij}$ toma valor ó indicando la presencia o ausencia del j-ésimo factor de riesgo en la categoría i-ésima. La matriz tiene una primera columna de unos. El vector de parámetros $\boldsymbol{\beta}$ a estimar por el modelo tiene dimensión p.

El costo medio esperado de los siniestros ajustado por el modelo puede ser resultado de una relación aditiva o multiplicativa (entre otras) de los grupos de riesgo; este último implica que la frecuencia relativa para una celda es el producto de las frecuencias relativas de su fila y columna[3]. En particular, usaremos el modelo multiplicativo (log-lineal). Entonces el valor ajustado para la -ésima categoría de pólizas será:

$$f_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \tag{1}$$

donde $f$ es el vector del monto medio de las reclamaciones ajustadas. Se espera que cada $f_i$ sea tan cercano al valor real observado $c_i$ como sea posible, por lo que se debe encontrar una manera de minimizar las diferencias entre estos independientemente del modelo que se haya empleado para estimar la severidad de las reclamaciones.

2 McClenahan, C.L., *Ratemaking*. Foundations of Casualty Actuarial Science, 4a. edición, 2001, págs. 75-148.

3 Feldblum, S. y Brosius, J.E., *The Minimum Bias Procedure: A Practitioner's Guide*. Actuarial Society, 2003, vol. 90, n° 172, págs. 196-273.

## 2.1 El método de mínimo sesgo

Estadísticamente existen diferentes técnicas para realizar tarificación, entre ellas el método de mínimo sesgo y los modelos lineales generalizados. Mildenhall, S. J.[4] muestra que ambas técnicas se relacionan cuando se utiliza la máxima verosimilitud. En este artículo en particular se presentará la metodología de mínimo sesgo.

Quienes introdujeron el método del sesgo mínimo fueron Bailey y Simon[5], que además propusieron una serie de criterios para que una tarifa fuera aceptable. Indicaron que la tarifa debe reproducir la experiencia del colectivo asegurado de manera general y en cada caso particular, es decir, debe ser ajustada a lo observado y equilibrada para los grupos de riesgo. Además la tarifa usará criterios de credibilidad, lo que significa que pueden utilizarse las primas puras de un grupo (o incluso del colectivo) para aproximar las de otro, en el caso, por ejemplo, de que este último sea muy poco numeroso o no haya suficientes datos para efectuar los cálculos. Como tercer criterio señalan que las tarifas deben desviarse lo mínimo posible de los costos reales para el máximo número de personas. El cálculo del precio de la prima de seguros debe producir una tarifa para cada grupo lo más cercana posible a la experiencia real, de tal forma que las diferencias se puedan considerar aleatorias.

Para comprobar el primer criterio propusieron calcular el siguiente cociente:

$$\frac{\sum_i y_i f_i}{\sum_i y_i c_i} \tag{2}$$

Para que se cumpla el equilibrio entre lo observado y lo ajustado se desea que dicho cociente sea lo más próximo a uno.

Una manera de corroborar el tercer criterio es medir la magnitud de los errores, es decir, cuánto se alejan en promedio los datos ajustados de los datos observados; esto se consigue aplicando el test de la diferencia absoluta.

En cuanto al cuarto criterio sugieren un estadístico para verificar el ajuste de las primas basado en la distribución $\chi^2$,

$$\hat{\chi}^2 = \sum_i \frac{y_i}{f_i}(c_i - f_i)^2 \tag{3}$$

4 Mildenhall, S.J., *A Systematic Relationship between Minimum Bias and Generalized Linear Models*. Proceedings of the Casualty Actuarial Society, 1999, vol. 86, n° 164, págs. 393-487.

5 Bailey, R.A. y Simon, L.J., *Two Studies in Automobile Insurance Ratemaking*. ASTIN Bulletin. 1960, vol. 1, n° 4, págs. 192-217.

De la ecuación (2) se deriva el modelo de mínimo sesgo, de la idea de que la diferencia entre $c_i$ y $f_i$ debe ser mínima. Ismail y Jemain[6] reescribieron esta ecuación considerando además que estas diferencias tienen pesos $w_i$, y que de esta forma se puede llegar a una manera más eficiente de estimar los parámetros y, por lo tanto, la severidad esperada:

$$\sum_i w_i (c_i - f_i)^2 = 0, \qquad j = 1, \ldots, p, \qquad w_i = y_i x_{ij}. \tag{4}$$

Minimizar las diferencias equivale a calcular la derivada respecto a $\boldsymbol{\beta}$ e igualar a cero la ecuación (4):

$$\sum_i w_i (c_i - f_i) \frac{\partial f_i}{\partial \beta_j} = 0 \qquad j = 1, 2, \ldots, p.$$

Las ecuaciones resultantes se pueden resolver de forma iterativa para obtener cada $\boldsymbol{\beta}_j$. La idea general es dar un valor inicial al vector $\boldsymbol{\beta}$, minimizar las diferencias con este valor, obtener un nuevo vector $\boldsymbol{\beta}$ y realizar este procedimiento hasta que soluciones consecutivas sean idénticas o muy cercanas[7].

Una vez obtenido el vector inicial de parámetros $\boldsymbol{\beta}_{(0)}$, las derivadas $\frac{\partial f_i(\beta)}{\partial \beta_j}$ matriz $Z_{(0)}$ de dimensiones $n \times p$ y la matriz de pesos $W_{(0)}$ de tamaño $n \times n$, estos dos últimos evaluados en el valor inicial $\boldsymbol{\beta}_{(0)}$, Ismail y Jemain[8] proponen el uso de series de Taylor para obtener el vector:

$$\boldsymbol{\beta}_{(1)} = (\mathbf{Z}^t_{(0)} \mathbf{W}_{(0)} \mathbf{Z}_{(0)})^{-1} \mathbf{Z}^t_{(0)} \mathbf{W}_{(0)} (\mathbf{c} - \mathbf{s}_{(0)}),$$

contando que $\mathbf{s}_{(0)}$ es el vector de tamaño $n \times 1$ cuya componente -ésima viene dada por:

$$s_i = f_i\,(\beta_{(0)}) - \sum_{j=1}^{p} \beta_{j(0)} z_{ij(0)}.$$

Los detalles para el modelo multiplicativo y para el modelo aditivo se pueden ver en los trabajos de Ismail y Jemain[9].

6 Ismail, N. y Jemain, A.A. *Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation*. Casualty Actuarial Society Forum, Primavera, 2005, págs. 367-394.

7 Bailey, R.A. y Simon, L.J., *Two Studies in Automobile Insurance Ratemaking*. ASTIN Bulletin, 1960, vol. 1, n° 4, págs. 192-217.

8 Feldblum, S. y Brosius, J.E., *The Minimum Bias Procedure: A Practitioner's Guide*. Actuarial Society, 2003, vol. 90, n° 172, págs. 196-273; Ismail, N. y Jemain, A.A., *Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation*. Casualty Actuarial Society Forum, Primavera, 2005, págs. 367-394.

9 Feldblum, S. y Brosius, J.E., *The Minimum Bias Procedure: A Practitioner's Guide*. Actuarial Society, 2003, vol. 90, n° 172, págs. 196-273.; Ismail, N. y Jemain, A.A., *Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation*. Casualty Actuarial Society Forum, Primavera, 2005, págs. 367-394.

Ismail y Jemain[10] también muestran que se puede realizar el ajuste suponiendo una distribución y estimando por máxima verosimilitud. La Tabla 1 muestra los pesos asociados a los diferentes ajustes.

**Tabla 1: Pesos para el Modelo Multiplicativo. Fuente: Ismail y Jemain, [6].**

| Modelos | $w_i$ donde $\sum_i w_i(c_i - f_i) = 0$ | | |
|---|---|---|---|
| Sesgo cero | $y_i x_{ij}$ | Densidad Poisson | $\frac{y_i}{f_i}\frac{\partial f_i}{\partial \beta_j}$ |
| Mínimo $\chi^2$ | $\frac{y_i(c_i + f_i)}{f_i^2}\frac{\partial f_i}{\partial \beta_j}$ | Densidad Exponencial | $\frac{1}{f_i^2}\frac{\partial f_i}{\partial \beta_j}$ |
| Densidad Normal | $y_i^2 \frac{\partial f_i}{\partial \beta_j}$ | Densidad Gamma | $\frac{y_i}{f_i^2}\frac{\partial f_i}{\partial \beta_j}$ |
| Densidad Log - Normal | $y_i \frac{\partial f_i}{\partial \beta_j}$ | | |

Feldblum y Brosius[11] presentan ejemplos intuitivos de cómo se realiza la estimación por el método de mínimo sesgo, tanto si se consideran efectos multiplicativos o aditivos, bajo el supuesto de sesgo cero, tratando de corroborar la lista de criterios propuestos por Bailey.

## 3. APLICACIÓN AL SEGURO DE GASTOS MÉDICOS MAYORES EN MÉXICO

El Seguro de Gastos Médicos Mayores (GMM) en México está diseñado para cubrir los gastos médicos originados por un accidente y/o enfermedad cubierto en la póliza, una vez que el monto rebasa la franquicia. Se enfoca a restaurar y rehabilitar la salud del asegurado, por lo que se puede decir que es curativo, pues la cobertura comienza una vez que se ha visitado al médico.

Los seguros de Salud y Gastos Médicos Mayores se pueden contratar de manera Individual (cuando brindan protección al individuo y/o familia contratante) y en Grupo o Colectivo (cuando brindan protección a los empleados de los sectores empresariales más diversos, así como a los miembros de asociaciones y agrupaciones legalmente establecidas).

10 Ismail, N. y Jemain, A.A., *Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation*. Casualty Actuarial Society Forum, Primavera, 2005, págs. 367-394.

11 Feldblum, S. y Brosius, J.E., *The Minimum Bias Procedure: A Practitioner's Guide*. Actuarial Society, 2003, vol. 90, n° 172, págs. 196-273.

Al hacer la suscripción del riesgo se debe ser cuidadoso en seleccionar subgrupos de asegurados con las mismas características y así poder cuantificar de una mejor forma el riesgo que la compañía aseguradora asume. Para cumplir con lo anterior es primordial identificar y seleccionar los factores de riesgo más relevantes, es decir, las características de los asegurados que definan aumentos o disminuciones en la siniestralidad y que conjuntamente expliquen un gran porcentaje de la varianza de los costos reclamados.

Para los Seguros de Salud los factores más evidentes a tener en cuenta son la edad y el sexo; se espera que durante los primeros años de vida las reclamaciones de siniestros sean altas, para después tener un descenso en la niñez y adolescencia, y un aumento en la edad adulta, seguido de un segundo descenso en la vejez.

## 3.1 Descripción de los datos

La base de datos para realizar este trabajo contiene información real de reclamaciones realizadas entre los años 2006, 2007 y 2008 para seguros de Grupo únicamente. Aunque la base está conformada con todos los siniestros sucedidos, se han eliminado aquellos cuya causa sea parto y cesárea, así como aquellos que tengan que ver con padecimientos que requieran pagar una sobre prima para cubrirlos, pues la idea es crear una tarifa base y, a partir de esta, ir incrementado el costo con las coberturas adicionales contratadas.

Para el análisis solo se toman en cuenta el número y monto de siniestros pagados por la aseguradora, es decir, no se contempla ni la cantidad de dinero pagada en concepto de franquicia, ni el coaseguro, que son a cargo del asegurado. Todos los montos que se presentan están expresados en pesos mexicanos.

La manipulación de la base de datos se hizo con el software Visual FoxPro 6.0 y con Excel 2003. Para los análisis estadísticos se usó R versión 2.10.1. En el apéndice se recogen los programas en R utilizados para ajustar los modelos, que se basan en los programas publicados en Ismail y Jemain[12].

La edad media de los asegurados es de 29 años, la composición por sexo es de hombres (49,95%) y mujeres (50,05%). La severidad es de 25,301 pesos, de un total de 20,336 reclamaciones, y un monto total pagado de 514,529,148 pesos.

Los siniestros que cuestan menos dinero son los que tienen más reclamaciones; tan solo con los montos menores a 12,700 pesos ya se alcanza el 52,3% de frecuencia acumulada, si bien el monto reclamado en total no es alto comparado con los mon-

12 Smail, N. y Jemain, A.A., *Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation*. Casualty Actuarial Society Forum, Primavera, 2005, págs. 367-394.; Ismail, N. y Jemain, A.A., *Comparison of Minimum Bias and Maximum Likelihood Methods for Claim Severity*. Casualty Actuarial Society E-Forum, 2009, págs. 243-275.

tos elevados que tienen un pequeño número de las reclamaciones que tienen lugar. Así, una póliza de seguros se puede ver afectada tanto por la frecuencia como por la severidad de las reclamaciones.

En el mercado asegurador, normalmente, se presentan las tarifas de los seguros de gastos médicos por quinquenios. En la Tabla 2 se presentan el número de siniestros y los montos reclamados por edad.

**Tabla 2: Descripción de siniestros por edad.**

| Edad | Total | % | Suma | % | Monto medio |
|---|---|---|---|---|---|
| **[0,5)** | 2,609 | 12,8 | 56,108,421 | 10,9 | 21,505.72 |
| **[5,10)** | 1,236 | 6,1 | 18,389,158 | 3,6 | 14,877.96 |
| **[10,15)** | 931 | 4,6 | 16,125,581 | 3,1 | 17,320.71 |
| **[15,20)** | 906 | 4,5 | 19,590,565 | 3,8 | 21,623.14 |
| **[20,25)** | 1,338 | 6,6 | 29,723,664 | 5,8 | 22,215.00 |
| **[25,30)** | 2,515 | 12,4 | 52,222,050 | 10,1 | 20,764.23 |
| **[30,35)** | 3,170 | 15,6 | 71,572,411 | 13,9 | 22,578.05 |
| **[35,40)** | 2,608 | 12,8 | 67,458,180 | 13,1 | 25,865.87 |
| **[40,45)** | 1,807 | 8,9 | 53,882,450 | 10,5 | 29,818.73 |
| **[45,50)** | 1,316 | 6,5 | 47,354,082 | 9,2 | 35,983.34 |
| **[50,55)** | 827 | 4,1 | 34,207,985 | 6,6 | 41,363.95 |
| **[55,60)** | 526 | 2,6 | 19,645,445 | 3,8 | 37,348.75 |
| **[60,65)** | 300 | 1,5 | 14,290,074 | 2,8 | 47,633.58 |
| **[65,70)** | 127 | 0,6 | 7,885,484 | 1,5 | 62,090.43 |
| **[70,90]** | 120 | 0,6 | 6,073,598 | 1,2 | 50,613.32 |
| **Total** | 20,336 | 100 | 514,529,148 | 100 | |

**Fuente:** Elaboración propia.

## 3.2 Estimación de los montos ajustados en el modelo multiplicativo

Los factores que interesan en mayor medida son el sexo y la edad categorizada en quinquenios. Con ellos se deben construir variables indicadoras para cada una de las categorías, siempre teniendo en cuenta que el último nivel de la categoría no debe ser incluido, pues se puede derivar de los demás niveles. En total se obtienen 30 clases como resultado de multiplicar 2 (número de categorías para Sexo) por 15 (número de quinquenios considerados para la Edad). No existen categorías con cero individuos, por lo que el ajuste se realiza sobre las 30 clases posibles.

Para el Sexo se ha creado una variable que vale uno si el asegurado es mujer y cero si es hombre. Para la variable Edad, que tiene 15 niveles, se han creado 14 variables indicadoras, siendo la categoría de mayores de 69 años la que actúa como categoría base. Así, resulta que la matriz $\chi$ tiene dimensión 30 (clases) $\times$ 16 (número de $\boldsymbol{\beta}$'s a estimar, más la intersección).

Para cada clase se ha calculado el costo medio de siniestro pagado $c_i$, como resultado de dividir la suma total pagada en cada clase entre el número de reclamaciones $y_i$.

En la Tabla 3 se presentan los exponenciales de las $\boldsymbol{\beta}$'s estimadas y los resultados del test $\chi^2$ y el test de las diferencias absolutas. El ajuste que minimiza el test de $\chi^2$ es el Mínimo $\chi^2$, seguido por el ajuste de sesgo cero. De los métodos por Máxima Verosimilitud el mínimo del test de $\chi^2$ se tiene para la distribución Gamma; el valor mínimo de la diferencia absoluta es para el ajuste Normal seguido por la Gamma. Además, cabe notar que los parámetros de regresión estimados no son tan diferentes entre distribuciones; a excepción del ajuste log-normal, la mayor diferencia se da en el término independiente estimado. El peor ajuste lo produce la distribución exponencial, que toma el máximo valor para los dos test realizados.

Para calcular los valores ajustados de monto promedio pagado basta realizar el producto de los parámetros estimados para las características deseadas, sin olvidar el término constante, y multiplicar el resultado por mil. Por ejemplo, si se quiere estimar el costo de siniestro estimado para una mujer de 50 años con el método de Máxima verosimilitud bajo la Normal, se ha de multiplicar:

$$50.8185 \times 0.8294 \times 0.9448 \times 1000 = 39{,}822$$

Y si fuera un hombre de 50 años:

$$50.8185 \times 0.8294 \times 1000 = 42{,}149$$

Entonces el monto de siniestro medio esperado para una mujer de 50 años es de 39,827 pesos y para un hombre de la misma edad un siniestro medio esperado de 42,152 pesos.

En prácticamente todos los casos, en los ajustes para costos medios más altos es donde más diferencia existe entre la severidad ajustada y la observada. Finalmente, teniendo los valores esperados de severidad se puede calcular la prima pura.

**Tabla 3: Exponencial de betas estimadas**

| | Mínimo sesgo | | Máxima verosimilitud | | | | Otros | |
|---|---|---|---|---|---|---|---|---|
| | Sesgo cero | Mín $\chi^2$ | Normal | Exponencial | Gamma | Log-Normal | Mín Cuadrados | $\chi^2$ Modificada |
| **Constante** | 53.0692 | 56.2887 | 50.8185 | 55.1924 | 51.7971 | 0.7900 | 55.0011 | 46.5711 |
| **Edad 0** | 0.4198 | 0.3960 | 0.4311 | 0.4198 | 0.4265 | 0.9271 | 0.4086 | 0.4773 |
| **Edad 5** | 0.2905 | 0.2746 | 0.2963 | 0.2932 | 0.2955 | 0.8928 | 0.2822 | 0.3289 |
| **Edad 10** | 0.3384 | 0.3190 | 0.3486 | 0.3369 | 0.3436 | 0.9070 | 0.3298 | 0.3852 |
| **Edad 15** | 0.4243 | 0.3999 | 0.4378 | 0.4198 | 0.4299 | 0.9278 | 0.4149 | 0.4830 |
| **Edad 20** | 0.4395 | 0.4144 | 0.4490 | 0.4361 | 0.4439 | 0.9308 | 0.4315 | 0.4997 |
| **Edad 25** | 0.4118 | 0.3882 | 0.4233 | 0.4065 | 0.4165 | 0.9249 | 0.4032 | 0.4684 |
| **Edad 30** | 0.4470 | 0.4216 | 0.4600 | 0.4415 | 0.4525 | 0.9326 | 0.4372 | 0.5081 |
| **Edad 35** | 0.5110 | 0.4823 | 0.5263 | 0.5054 | 0.5178 | 0.9451 | 0.4989 | 0.5801 |
| **Edad 40** | 0.5886 | 0.5547 | 0.6038 | 0.5830 | 0.5952 | 0.9581 | 0.5768 | 0.6698 |
| **Edad 45** | 0.7100 | 0.6695 | 0.7302 | 0.7033 | 0.7192 | 0.9757 | 0.6937 | 0.8073 |
| **Edad 50** | 0.8168 | 0.7819 | 0.8295 | 0.8112 | 0.8216 | 0.9865 | 0.8083 | 0.9006 |
| **Edad 55** | 0.7322 | 0.6968 | 0.7661 | 0.7101 | 0.7389 | 0.9771 | 0.7211 | 0.8166 |
| **Edad 60** | 0.9345 | 0.8820 | 0.9680 | 0.9191 | 0.9455 | 1.0009 | 0.9159 | 1.0608 |
| **Edad 65** | 1.2193 | 1.1505 | 1.2521 | 1.2145 | 1.2373 | 1.0260 | 1.1882 | 1.3847 |
| **Género** | 0.9159 | 0.9166 | 0.9448 | 0.8639 | 0.9376 | 0.9934 | 0.8885 | 0.9178 |
| $\chi^2$ | **2,921.46** | **2,888.15** | 3,097.01 | **3,387.44** | **3,014.45** | 1.03 | **3,027.86** | 3,220.13 |
| **Dif. Absoluta** | 0.0530 | 0.0529 | **0.0478** | **0.0673** | **0.0494** | 0.0204 | 0.0580 | 0.0525 |

**Fuente:** Elaboración propia.

## 3.3 Cálculo de la prima pura

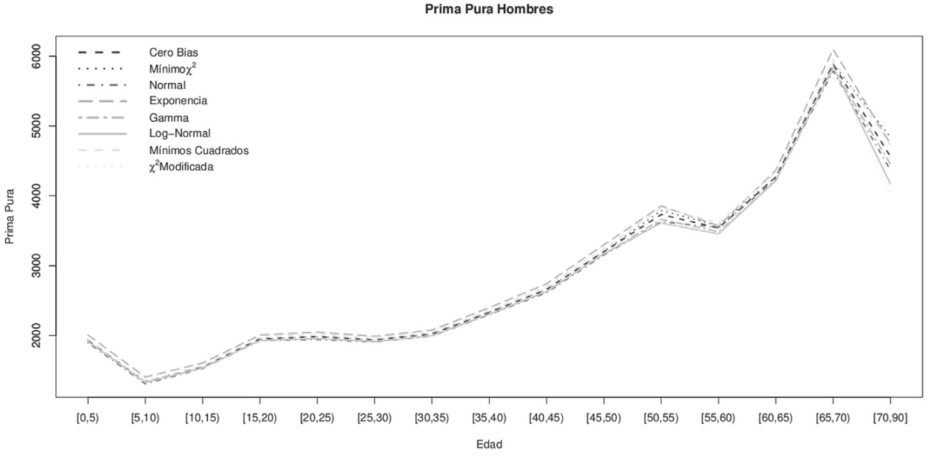
La prima pura se puede calcular como el producto de la frecuencia esperada de las reclamaciones por la severidad media de una reclamación. Para calcular la frecuencia de las reclamaciones hace falta tener el número de asegurados en vigor, separados por cada clase de edad y sexo, y además conocer el tiempo en el que estuvieron expuestos (en 2008 se tenían 218,734 asegurados). Así, dividiendo el número de reclamaciones entre la distribución de asegurados en vigor se obtiene la frecuencia de las reclamaciones.

Para todas las primas calculadas se cumple que estas son mayores para las mujeres, aunque la diferencia con los varones es mínima, por lo que las tarifas son bastante cercanas. Las menores diferencias entre prima pura para hombres y mujeres se dan con el ajuste exponencial y el de mínimos cuadrados, donde en promedio el incremento en la prima para mujeres es de menos del 3; en cambio, para los demás ajustes este va del 6 al 9.

Para ambos sexos se tiene que para la edad de 0 a 5 años el costo es mayor que en el periodo de la adolescencia y juventud, creciendo para la edad adulta y bajando finalmente para los asegurados de más de 70 años, lo que se puede explicar por el hecho de que en teoría ya no deben existir asegurados de esta edad por no cubrirlos este tipo de seguro que está pensado para un colectivo de trabajadores y sus familiares. Aunque se hagan excepciones de inclusión, estas son pocas y, por lo tanto, hay un escaso número de reclamaciones de siniestros de personas de más de 69 años.

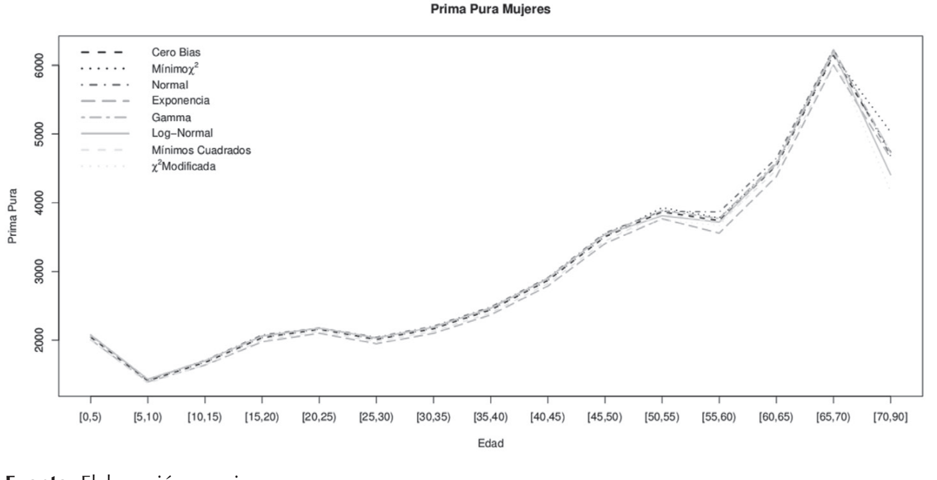
En las Figuras 1 y 2 se muestran las primas ajustadas por todos los métodos, separados por hombres y mujeres. La prima pura resultado del ajuste con la exponencial para los hombres es superior a las demás; en cambio para las mujeres ocurre lo contrario. Para los hombres, las diferencias más grandes ocurren para las edades de (50,55), (65,70) y (70,90). Para las mujeres, las mayores diferencias tienen lugar a partir de la edad (55,60).

**Figura 1.** Comparativo de prima pura ajustada en hombres.



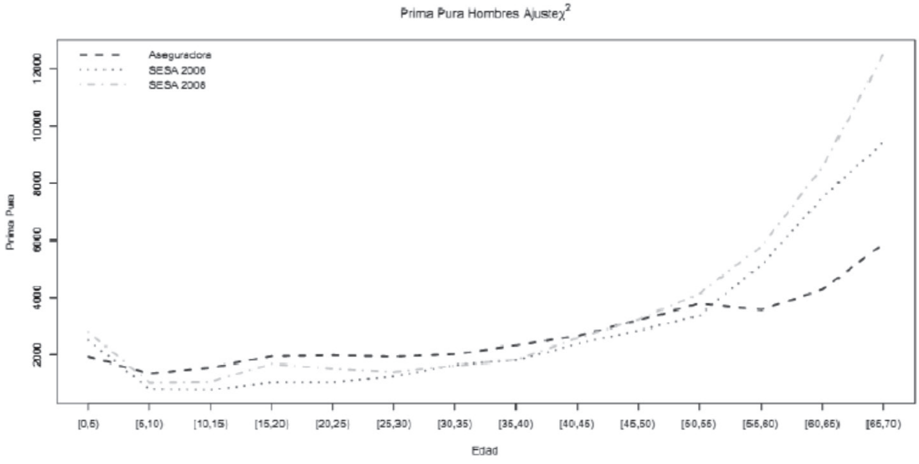

**Fuente:** Elaboración propia.

**Figura 2.** Comparativo de prima pura ajustada en mujeres.



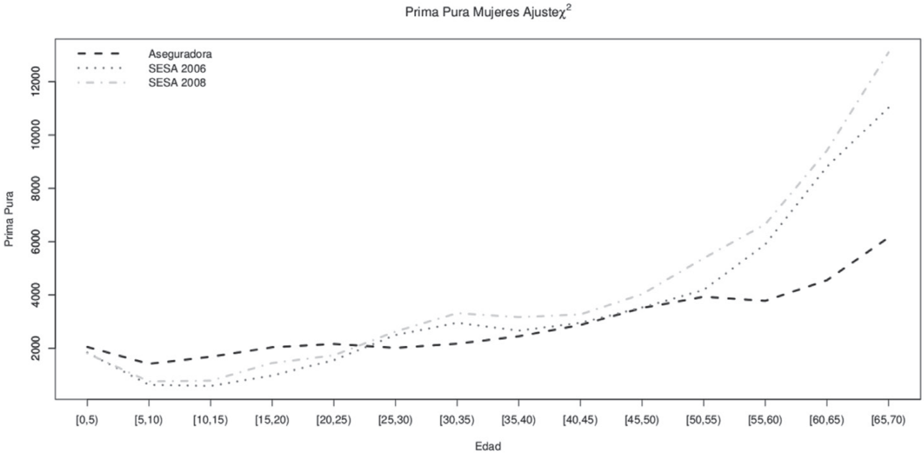

**Fuente:** Elaboración propia.

**Figura 3.** Comparativo de prima pura ajustada en tres poblaciones: Hombres.




**Fuente:** Elaboración propia.

**Figura 4.** Comparativo de prima pura ajustada en tres poblaciones: Mujeres.




**Fuente:** Elaboración propia.

## 3.4 Comparación con valores ajustados para el sector

En esta sección se hace un comparativo de las primas puras obtenidas con los datos de la aseguradora y las obtenidas por las SESA's (Sistema Estadístico del Sector Asegurador: estadísticas anuales de todas las aseguradoras). Las mayores diferencias se dan a partir de la edad de 55 años, donde las SESA's estiman primas superiores. Para los hombres, con los datos de la aseguradora solo se estiman primas puras mayores a las obtenidas por las SESA's para las edades entre 5 y 44 años. Para las mujeres, con los datos de la aseguradora solo se estiman primas puras más elevadas que las de las SESA's para las edades entre 0 y 24 años. A partir de esa edad, para todas las edades las primas estimadas por las SESA's son mayores.

## 4. CONCLUSIONES

En este trabajo se han presentado diferentes maneras de estimar el costo promedio de siniestros pagados. En general se han comparado los resultados obtenidos por el método con mínimo sesgo y con el método de Máxima Verosimilitud, suponiendo diferentes distribuciones de la severidad de los siniestros.

Si bien estos métodos se pueden aplicar tanto para modelos multiplicativos, aditivos e inversos o definidos por cualquier polinomio, en este caso solo se tomaron los modelos multiplicativos por ser el método más directo.

Los ajustes se realizaron suponiendo que no todos los asegurados tienen la misma probabilidad de reportar siniestros, por lo que se seleccionaron características clave para poder segmentar la población y asignar costos iguales para riesgos iguales. También hubo que realizar una buena manipulación de la base de datos para llegar a tener los datos en la forma correcta.

Una vez realizada la modelación, en primer lugar se concluye que el ajuste planteado no genera excesivas dificultades de implementación. Además se puede pasar de un modelo a otro con tan solo cambiar la matriz de pesos. En este punto se observó que los parámetros estimados por el modelo de sesgo cero son iguales a los estimados por el método de Máxima Verosimilitud suponiendo una distribución Poisson.

En la práctica, se aprecia que usando el test de la $\chi^2$ tanto para los datos de la aseguradora como los datos de las SESA's, el mejor ajuste fue el dado por la Mínima $\chi^2$. Para el test de la diferencia absoluta no se pudo encontrar ninguna coincidencia.

Los costos estimados por los diferentes ajustes no difieren en exceso entre ellos, por lo que, si se desconoce la distribución de la que provienen los siniestros, siempre se puede recurrir a otro ajuste que no requiera disponer de dicha información, pues solo se requiere hacer una elección de la matriz de pesos.

En cuanto a las primas puras, se concluye que es importante realizar una correcta clasificación del riesgo; como se ha visto las mujeres son las que mayor gasto originan y, por lo tanto, deben tener asignadas primas puras mayores que los varones.

Comparando las tres bases de datos analizadas, se observa que en el global de las compañías se tienen con frecuencia asegurados con edad mayor a la edad límite de aceptación, lo que provoca que para estos casos las primas calculadas sean mayores a las estimadas con los datos de la aseguradora.

A destacar la facilidad con la que se puede ampliar la metodología usada para incluir más factores de riesgo y no solo dos; basta con reorganizar los datos creando tantas variables indicadoras como sea necesario, y haciendo ajustes en la programación.

El siguiente paso en esta investigación, que será abordado en un futuro próximo, será calcular la prima de tarifa contemplando los gastos tanto de adquisición como de administración y la utilidad esperada por la compañía. Además, se debe considerar la inflación en salud, pues refleja el aumento esperado en el costo de los servicios relacionados con el sector salud (medicamentos, material quirúrgico, servicio hospitalario, honorarios médicos, etc.) y tiene relevancia para un cálculo más ajustado de la prima.

## AGRADECIMIENTOS

Parte de este trabajo fue realizado durante la elaboración del trabajo de final del Master Interuniversitario de Estadística e Investigación Operativa de la Universidad Politécnica de Cataluña y la Universidad de Barcelona. La segunda autora agradece el apoyo recibido de SEJ2010-21787-C03-01.

## BIBLIOGRAFÍA

Bailey, R. A. (1963), *Insurance Rates with Minimum Bias*. Proceedings of the Casualty Actuarial Society. 50 (93): 4-11.

Bailey, R. A. y Simon, L. J. (1960), *Two Studies in Automobile Insurance Ratemaking*. ASTIN Bulletin. 1 (4): 192-217.

Beard, R. E.; Pentikäinen, T. y Pesonen. E. (1984), *Risk Theory*. London: Chapman and Hall.

Brown, R. L. (1988), *Minimum Bias with Generalized Linear Models*. Proceedings of the Casualty Actuarial Society, 75 (143): 187-217.

Feldblum, S. y Brosius, J. E. (2003), *The Minimum Bias Procedure: A Practitioner's Guide*. Actuarial Society, 90 (172): 196-273.

Ismail, N. y Jemain, A. A. (2005), *Bridging Minimum Bias and Maximum Likelihood Methods through Weighted Equation*. Casualty Actuarial Society Forum, Primavera. 367-394.

Ismail, N. y Jemain, A. A. (2009), *Comparison of Minimum Bias and Maximum Likelihood Methods for Claim Severity*. Casualty Actuarial Society E-Forum. 243-275.

Jung, J. (1968), *On Automobile Insurance Ratemaking*. ASTIN Bulletin. 5 (1): 41-48.

McClenahan, C. L. (2001), *Ratemaking*. Foundations of Casualty Actuarial Science, 4a. edición. 75-148.

Mildenhall, S. J. (1999), *A Systematic Relationship between Minimum Bias and Generalized Linear Models*. Proceedings of the Casualty Actuarial Society. 86 (164): 393-487.

## APÉNDICE

Los programas en R que se han utilizado para la estimación de los modelos son los siguientes:

```
funcion<-
function(datos,iter=1000,a=0,b=2,d=0,g=0,h=0,k=0,l=0,seed=0.01,modelo='M')
 {
      X <- as.matrix(datos)
      cost <- datos$cost
      count <- datos$count
      aseg <- datos$aseg
      new.beta <- c( rep(c(seed), dim(X)[2]))

if(modelo=='M')
  {
     for (i in 1:iter){
           beta <- new.beta
           fitted <- as.vector(exp(X%*%beta))
           Z <- diag(fitted)%*%X
           W <- diag(count^b*(count/cost)^d*fitted^g*(1-
fitted)^h*(cost+fitted)^k)*(1+a*count*fitted)^1
           r.s <- cost-fitted+as.vector(Z%*%beta)
           new.beta <- as.vector(solve(round(t(Z)%*%W%*%Z,3),tol=1e-
100)%*%t(Z)%*%W%*%r.s
           }
     fitted <- as.vector(exp(X%*%new.beta))
     chi.square <- sum((count*(cost-fitted)^2)/fitted)
     abs.difference <- sum(count*abs(cost-fitted))/sum(count*cost)
     betas<-t(t(exp(new.beta)))
     fit<-matrix(fitted*1000,15,2 ,byrow = FALSE)
     primapura<-matrix((count/aseg)*(fitted*1000),15,2 ,byrow = FALSE)

     salida<-list (exp_Betas=betas, round(chi.square,3), abs.difference=
round(abs.difference,3), CostoFit=fit,  PrimaPura=primapura)
  }


if(modelo=='A')
  {

     for (i in 1:iter) {
            beta <- new.beta
            fitted <- as.vector(X%*%beta)
            W <- diag(count^b*(count/cost)^d*fitted^g*(fitted*(1-
fitted))^h*(count+fitted)^k)
            r.s <- cost-fitted+as.vector(X%*%beta)
            new.beta <- as.vector(solve(round(t(X)%*%W%*%X,3),tol=1e-
100)%*%t(X)%*%W%*%r.s)
            }
     fitted <- as.vector(X%*%beta)
     chi.square <- sum((count*(cost-fitted)^2)/fitted)
     abs.difference <- sum(count*abs(cost-fitted))/sum(count*cost)
     betas<-t(t(beta))
     fit<-matrix(fitted*1000,15,2 ,byrow = FALSE)
     primapura<-matrix((count/aseg)*(fitted*1000),15,2 ,byrow = FALSE)

     salida<-list (betas= betas, chi.square= round(chi.square,3),   abs.difference=
round(abs.difference,3), CostoFit = fit, PrimaPura=primapura)
    print(salida)
  }
}
```