

DATOS AGREGADOS Y DESAGREGADOS EN EL ANÁLISIS DE DATOS DE GARANTÍA

Jorge Mario Obando*
SERGIO YÁÑEZ**

RESUMEN

Los datos de garantía tratados estadísticamente son útiles para la predicción de reclamaciones futuras, para establecer comparaciones entre reclamaciones de productos y mejoras de ellos y estimación de la confiabilidad. Este análisis exige la disponibilidad de información diaria, que muchas veces sólo está disponible en forma agregada semanal o mensual debido a la logística de la empresa. En el presente trabajo se estudia el efecto que presenta la agregación de datos en los resultados del análisis.

PALABRAS CLAVE: confiabilidad; análisis de datos de garantía; procesos de conteo; tasa de intensidad.

ABSTRACT

Warranty data, statistically well handled, is useful to predict future claims, to make comparisons between products claims, to estimate products field reliability, and to identify reliability improvements that can be made to the product. This analysis requires the availability of daily records that sometimes are only available in weekly or monthly total (grouped data) due to the nature of the operations of the organization. In this work the effect of the grouped data in the final results of the analysis is studied.

KEY WORDS: reliability; warranty data analysis; counting processes; intensity rate.

* Ingeniero Administrador, EIA. Profesor Escuela de Ingeniería de Antioquia. pjorgeo@eia.edu.co

** Ingeniero Industrial. Magíster en Ciencias y Magíster en Estadística. Director Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín. syanez@unalmed.edu.co

1. INTRODUCCIÓN

Generalmente las empresas venden sus productos cubiertos con un período de tiempo (Ω) de garantía. Esto lo hacen por dos razones: la primera, la garantía hace parte del servicio posventa de la compañía; la segunda, la garantía hace parte de la calidad del producto. Así, un período de garantía ofrece un servicio posventa y un producto confiable.

Las empresas esperan que el número de unidades que fallen dentro del período de garantía sea mínimo, pues cada reclamación por garantía se convierte en un costo para la empresa, además de indicar que el producto falló antes de lo esperado. Si, por el contrario, el número de reclamaciones por garantía es alto, entonces existen problemas en el diseño o en alguna parte de la línea de producción, en otras palabras, la confiabilidad del producto es baja.

Es por esto por lo que el análisis estadístico de los datos de garantía se puede utilizar como un método de control estadístico de la calidad, como un medio para el estudio de la confiabilidad del producto y como herramienta para pronóstico de costos.

En la sección 2 se tratan los temas más importantes acerca del análisis de datos de garantía: el estimador de la tasa de reclamaciones por garantía y su varianza; además, se estudia el caso especial de agregación de los períodos de reclamación. La agregación de los registros de ventas lleva un supuesto acerca de las ventas diarias, esto se analiza en la sección 3. Finalmente, en la sección 4 se desarrolla el estudio de una simulación que compara los resultados de un análisis de datos de garantía a partir de datos de reclamaciones y ventas desagregados y agregados en un modelo de generación de reclamaciones específico.

2. ANTECEDENTES

El análisis de garantías consiste en el empleo del análisis estadístico para estimar la tasa de recla-

maciones por unidad dentro del período de garantía (Kalbfleisch et al., 1991; Lawless y Kalbfleisch, 1992; Lawless y Nadeau, 1995; Lawless y Kalbfleisch, 1996), calcular intervalos de confianza para datos agregados y no agregados (Kalbfleisch et al., 1991; Lawless y Kalbfleisch, 1996; Lawless, 1998), determinar la influencia de variables explicatorias (Lawless y Nadeau, 1995), calcular el efecto sobre los costos y pronosticar el impacto futuro de las reclamaciones por garantía (Lawless, 1998; Blischke y Murthy, 2000), estimar la función de confiabilidad de los productos (Kalbfleisch et al., 1991; Lawless, 1998) y, más recientemente, realizar control de calidad sobre el proceso de producción (Wu y Meeker, 2002).

La estimación de la tasa de reclamaciones se basa en la edad de la unidad al realizar la reclamación, es decir, el tiempo transcurrido desde su venta hasta la reclamación; sin embargo, el cálculo de la varianza de dicho estimador depende de los supuestos con respecto al proceso de generación de reclamaciones. Lawless y Kalbfleisch (1996) establecen tres escenarios:

1. Las reclamaciones por unidad en una edad específica son variables aleatorias independientes con distribución de Poisson (Kalbfleisch et al., 1991)
2. Las reclamaciones por unidad en una edad específica son variables aleatorias independientes, pero se presenta un factor de sobredispersión debido a las condiciones de operación y de uso de cada unidad, y
3. La presencia de correlación entre reclamaciones de distintas edades, debida a que las condiciones de operación y de uso generan un deterioro más rápido de las piezas dentro del período de garantía, lo que conduce a mayor número de reclamaciones.

Por otra parte, es común que los sistemas de registro se diseñen de acuerdo con las necesidades y limitaciones de la empresa, por lo tanto, se pueden presentar bases de datos, entre otros casos, con agregación de los registros (semanal, mensual o en otros intervalos de tiempo), tanto de ventas como de

garantías; con retraso en el registro de las reclamaciones; con registros vacíos en algunos días del año por cierre. De ellos, los casos más comunes son la agregación de los datos y el retraso en los registros.

La agregación de los datos lleva a estimar las ventas diarias y a realizar supuestos acerca de la edad de las unidades, lo cual tiene impacto en el estimador de la tasa de reclamaciones y su correspondiente varianza, pues se introduce una variación extra debida al desconocimiento de las ventas diarias (Lawless, 1998). Asimismo, el retraso en los registros lleva a la subestimación de la tasa de reclamaciones, por tanto, se debe definir una función de probabilidad de retrasos para corregir este problema.

A continuación se introduce la notación y se expone el modelo de análisis estadístico de datos de garantías basado en la edad.

2.1 Notación

El análisis de datos de garantía se basa en la edad a del producto, medida como el número de días de servicio de la unidad i desde el día de su venta, d_i , hasta el día t en el que se presenta la reclamación. Para ello, suponga que las unidades $i=1, 2, \dots, N$ se venden en los días $d=0, 1, 2, \dots, \tau$; así, $N(d)$ es el número de unidades vendidas el día d . Además, suponga que los registros incluyen todas las reclamaciones reportadas al productor hasta el día T , esto permite definir:

$n^T(d, a) \rightarrow$ Número de unidades vendidas el día d de edad a

$\lambda(a) \rightarrow$ Tasa de intensidad de reclamaciones por unidad de edad a ($a=1, 2, \dots$)

$\Lambda(a) \rightarrow$ Número esperado de reclamaciones por unidad hasta la edad a

Por otra parte, es posible que las reclamaciones tarden algún tiempo en ser reportadas al productor, debido, entre otros, a la logística propia de las empresas; por esto, se define:

$f(r) \rightarrow$ Probabilidades de que una reclamación se retrase r días en ser reportado ($r=0, 1, 2, \dots$)

$R^T(a) \rightarrow$ Unidades de edad a en riesgo reportadas hasta T

2.2 Análisis de reclamaciones por garantías

2.2.1 Estimador de la tasa de reclamaciones por garantía

Si $n^T(d, a, r)$ es el número de reclamaciones de unidades vendidas el día d , de edad a y que tuvieron un retraso de r días en ser reportadas al productor, de acuerdo con Kalbfleisch, Lawless y Robinson (1991):

$$n^T(d, a, r) \sim \text{Poisson}(\mu^T(d, a, r)) \quad (1)$$

$$\mu^T(d, a, r) = N(d)\lambda(a)f(r)$$

Entonces, al contar con los registros de los $n^T(d, a, r)$ para $d+a+r \leq T$, y suponer que $f(r)$ ($r=0, 1, 2, \dots$) y $N(d)$ ($d=0, 1, 2, \dots$) son conocidos, se tiene la ecuación de verosimilitud:

$$L = \prod_{d=0}^{\tau} \prod_{a=0}^{\tau-d} \prod_{r=0}^{\tau-d-a} \frac{e^{-\mu^T(d, a, r)} [\mu^T(d, a, r)]^{n^T(d, a, r)}}{n^T(d, a, r)!} \quad (2)$$

$$\Rightarrow \begin{cases} \lambda(a) = \frac{n^T(a)}{R^T(a)} \\ \Lambda(a) = \sum_{u=0}^a \lambda(u) \end{cases} \quad (3)$$

Donde:

$$n^T(a) = \sum_{d=0}^{\tau-a} n^T(d, a)$$

$$R^T(a) = \sum_{d=0}^{\tau-a} N(d)F(T-(d+a)) \quad (4)$$

$$F(r) = \sum_{u=0}^r f(u)$$

No obstante, Kalbfleisch, Lawless y Robinson (1991) señalan que (1) realmente es una simplificación de modelos más generales, los cuales tienen en cuenta que:

- La distribución de las reclamaciones varía de unidad a unidad. Esto se debe a que las tasas y condiciones de uso de cada unidad son diferentes.
- Otra fuente de variación de los $\lambda(a)$ es el efecto calendario: las reclamaciones por garantía son difíciles (o imposibles) de realizar en fines de semana (más aun los domingos y festivos en el caso colombiano). Así, para las unidades vendidas un lunes, por ejemplo, $\lambda(7k - 1) = 0$ ($k = 1, 2, \dots$).

En estos casos, Kalbfleisch, Lawless y Robinson (1991) precisan que el modelo (1) es plausible si se tiene en cuenta que los conteos $n^T(d, a, r)$ obtenidos por la superposición de los procesos de muchas unidades son aproximadamente Poisson cuando la intensidad de las reclamaciones es pequeña; además, el parámetro $\lambda(a)$ debe ser interpretado como el número marginal de reclamaciones por unidad de edad a .

Finalmente, el modelo (1) es válido porque los estimadores máximos verosímiles son también cuasiverosímiles (o ecuaciones de mínimos cuadrados generalizados) y llevan a estimadores concretos válidos con supuestos más generales.

2.2.2 Varianza del estimador de la tasa de reclamaciones por garantía

Es precisamente en este punto donde hay que tener en cuenta las variantes del modelo (1) mencionadas, pues la variabilidad de unidad a unidad en las distribuciones del número de reclamaciones y el efecto calendario no son tenidos en cuenta en el primer modelo, aunque, como ya se mencionó, esto no afecta el estimador de $\lambda(a)$ propuesto en (3); por lo tanto, se tienen varios estimadores de la varianza de $\Lambda(a)$, dependiendo de los supuestos del modelo que se esté trabajando:

1. Si realmente (1) se cumple, es decir, los $n^T(d, a, r)$ son variables aleatorias Poisson independientes, entonces los $n^T(d, a)$ también son variables aleatorias independientes Poisson con media $\mu^T(d, a) = N(d)\lambda(a)F(r)$, y, en consecuencia (Lawless y Kalbfleisch, 1996), $n^T(a) = R^T(a)\lambda(a)$ y la varianza de $\Lambda(a)$ se estima con:

$$V_p(\hat{\Lambda}(a)) = \sum_{a=0}^{\infty} \frac{\lambda(a)}{R^T(a)} \quad (5)$$

2. Ahora, si se consideran supuestos más generales, como lo es la variabilidad en las distribuciones de reclamaciones de unidad a unidad, entonces se presentará variación adicional a la Poisson (Lawless y Kalbfleisch, 1996; Lindsey, 1997). Otra posibilidad, aun más general, es considerar que puede presentarse un grado de correlación entre los $n^T(d, a)$ (Lawless y Kalbfleisch, 1996; Lawless, 1998). Estos supuestos se discutirán en un trabajo futuro, ya que, por el momento, están por fuera de los objetivos de este artículo.

2.3 Datos agregados

Observe que hasta el momento, los estimadores propuestos suponen el conocimiento de las $n^T(d, a)$ y $N(d)$, es decir, se cuenta con registros de reclamaciones, bien sea unidad por unidad o día a día; pero la organización logística de muchas empresas impide que se tenga esta clase de información tan detallada. En otras palabras, muchas empresas sólo tienen acceso a registros de reclamaciones y ventas totales semanales, quincenales, mensuales o en cualquier otro tipo de agregación temporal.

De acuerdo con Lawless y Kalbfleisch (1996), si los períodos de agregación de las reclamaciones y los reportes son del mismo tamaño, basta hacer las sustituciones necesarias en los estimadores propuestos y no habrá mayor diferencia, tal como lo muestran los autores en su artículo al comparar resultados con base en registros diarios y agregados mensualmente. Sin embargo, si los períodos de agregación de reclamaciones y reportes son de diferente tamaño (Lawless y Kalbfleisch, 1996), o, más aun, cuando los registros de ventas también están agregados en algún intervalo de tiempo (Lawless, 1998), es necesario hacer una extensión del análisis para realizar los ajustes necesarios.

Para ello se suponen los conjuntos de edad $A_j = [a_{j-1}, a_j)$ y los intervalos de tiempo en los que ocurren las fallas $P_k = [t_{k-1}, t_k)$.

Si se define $n^{*T}(t, a) = n^T(t - a, a)$, se tiene su análogo agregado:

$$n^{*T}(P_k, A_j) = \sum_{t=t_{k-1}}^{t_k-1} \sum_{a=a_{j-1}}^{a_j-1} n^{*T}(t, a)$$

En el caso de datos agregados, se desea estimar:

$$\Lambda(A_j) = \sum_{a=a_{j-1}}^{a_j-1} \lambda(a) \quad (6)$$

Para ello, se define primero el número de reclamaciones reportadas de unidades con edad j :

$$n^T(A_j) = \sum_{a=a_{j-1}}^{a_j-1} \sum_{t=0}^T n^{*T}(t, a)$$

Sin embargo, recuerde que no se tienen registros diarios; además, como todos los datos están agregados, T es el límite superior del último de los P_k . Así, si $T \in P_m$, entonces $T = t_m - 1$, por lo tanto:

$$n^T(A_j) = \sum_{k=1}^m n^{*T}(P_k, A_j) = \sum_{a=a_{j-1}}^{a_j-1} n^T(a) \\ \Rightarrow E(n^T(A_j)) = \sum_{a=a_{j-1}}^{a_j-1} \lambda(a) R^T(a)$$

Ahora bien, Lawless (1998) afirma que la suposición de $\lambda(a)$ constante en A_j es válida; no obstante, esto en realidad depende de las características particulares del producto y de la amplitud de los intervalos de agregación y, por ello, otros supuestos serán necesarios en casos más especiales.

Tomando el supuesto de Lawless, se tiene que:

$$E(n^T(A_j)) = \lambda(a) \sum_{a=a_{j-1}}^{a_j-1} R^T(a) \times \frac{a_j - a_{j-1}}{a_j - a_{j-1}} \\ = \Lambda(A_j) R^T(A_j)$$

Donde:

$$R^T(A_j) = \frac{1}{a_j - a_{j-1}} \sum_{a=a_{j-1}}^{a_j-1} R^T(a) \quad (7)$$

De todo lo anterior se concluye que:

$$\hat{\Lambda}(A_j) = \frac{n^T(A_j)}{R^T(A_j)} \quad (8)$$

Ahora bien, (7) supone el manejo de registros diarios de ventas. Más adelante se analizará la expresi-

ón con el supuesto de agregación de los registros de ventas. Por el momento, se supondrá que los $R^T(A_j)$ son conocidos para la estimación de la varianza de (8). De acuerdo con Lawless y Kalbfleisch (1996) y Lawless (1998), se pueden calcular estimadores análogos a (5), pero teniendo en cuenta la agregación:

$$\hat{V}_p(\hat{\Lambda}(A_j)) = \frac{\hat{\Lambda}(A_j)}{R^T(A_j)} \quad (9)$$

3. EXTENSIÓN DEL ANÁLISIS POR LA AGREGACIÓN DE LAS VENTAS

Note que las expresiones (7) a (9) suponen el conocimiento de las ventas diarias $N(d)$, lo cual no es posible si los registros de ventas también están agregados en algún intervalo de tiempo. En este caso, Kalbfleisch, Lawless y Robinson (1991) proponen un estimador para $R^T(A_j)$, el cual está basado en aceptar que las $N(d)$ son constantes en un intervalo $[a, b]$: Sean $D_k = P_k - A_j$ y $N(D_k) = \sum_{d \in D_k} N(d)$ las ventas en el período D_k , y sea $T \in P_m$ (así $T = t_m - 1$); entonces :

$$\bar{R}^T(a) = \sum_{k=1}^m \left[\frac{N(D_k)}{d_k - d_{k-1}} \times \sum_{a=d_{k-1}}^{d_k-1} F(T - (d+a)) \right] \\ \bar{R}^T(A_j) = \frac{1}{a_j - a_{j-1}} \sum_{a=a_{j-1}}^{a_j-1} \bar{R}^T(a) \quad (10)$$

$$\Rightarrow \bar{\Lambda}(A_j) = \frac{n^T(A_j)}{\bar{R}^T(A_j)} \quad (11)$$

$$\bar{\Lambda}_j = \sum_{a=1}^j \bar{\Lambda}(A_j)$$

Ello resolvería el problema del desconocimiento de $N(d)$ en (7) y (8). Además, (9) subestima la varianza del estimador, dado que también se tiene incertidumbre acerca de los $N(d)$, es decir, el hecho de estimar $N(d)$ introduce una nueva fuente de variación que debe tenerse en cuenta en el modelo.

Para ello, sea $\tilde{R}_r = \{\tilde{R}^r(a); a = 0, 1, 2, \dots\}$; además, suponga que los $\tilde{R}^r(a)$ son insesgados e independientes del conteo de reclamaciones $n^r(d, a)$ y:

$$\hat{\lambda}(a) = \frac{n^r(a)}{\tilde{R}^r(a)} = \frac{n^r(a)}{\tilde{R}^r(a)} \times \frac{\tilde{R}^r(a)}{\tilde{R}^r(a)} = \tilde{\lambda}(a) \times \frac{\tilde{R}^r(a)}{\tilde{R}^r(a)}$$

$$\hat{\Lambda}(A_j) = \frac{n^r(A_j)}{\tilde{R}^r(A_j)} = \frac{n^r(A_j)}{\tilde{R}^r(A_j)} \times \frac{\tilde{R}^r(A_j)}{\tilde{R}^r(A_j)} = \tilde{\Lambda}(A_j) \times \frac{\tilde{R}^r(A_j)}{\tilde{R}^r(A_j)}$$

Entonces, recordando que:

$$V(X) = E(V(X|Y)) + V(E(X|Y))$$

y haciendo $X = \hat{\Lambda}(a)$ y $Y = \tilde{R}_r$, se tiene:

$$V(\hat{\Lambda}(a)) = E(V(\hat{\Lambda}(a)|\tilde{R}_r(a))) + V(E(\hat{\Lambda}(a)|\tilde{R}_r(a)))$$

$$= E(V(\hat{\Lambda}(a)|\tilde{R}_r(a))) + V\left(\sum_{u=0}^{\infty} \lambda(u) \frac{\tilde{R}^r(u)}{\tilde{R}^r(u)}\right) \quad (12)$$

Ahora bien, (12) tiene dos términos. Según Lawless (1998), el primer término se calcula reemplazando $\tilde{\lambda}(a)$ y $\tilde{R}^r(a)$ en (5); mientras que para el segundo término se hace necesario establecer una distribución para $\tilde{R}^r(u)/R^r(u)$ para, mediante simulación, poder reemplazar $\lambda(a)$ por $\hat{\lambda}(a)$.

De la misma forma, se puede hacer la extensión para el caso agregado:

$$V^{(A)}(\hat{\Lambda}(A_j)) = E(V(\hat{\Lambda}(A_j)|\tilde{R}_r(A_j))) + V(E(\hat{\Lambda}(A_j)|\tilde{R}_r(A_j)))$$

$$= E(V(\hat{\Lambda}(A_j)|\tilde{R}_r(A_j))) + V\left(\Lambda(A_j) \frac{\tilde{R}^r(A_j)}{\tilde{R}^r(A_j)}\right) \quad (13)$$

Ya en (13) se siguen los lineamientos propuestos por Lawless (1998), pero teniendo en cuenta la agregación.

4. SIMULACIÓN

Para generar las ventas diarias, se truncó y discretizó una distribución normal con parámetros $\mu=8$ y $\sigma=1$ en el intervalo [5,12] (por simplicidad, esta distribución se denotará en adelante como $N_T^D(\mu, \sigma, a, b)$, donde [a,b] es el intervalo de truncamiento); se generaron 240.000 números aleatorios de esta distribución y se agruparon en 8.000 conjuntos no traslapados de 30 números. Cada observación mensual correspondió a la suma de los 30 números aleatorios de cada conjunto; así se generó una serie de 8.000 observaciones de ventas mensuales, a la cual se le realizó una prueba de bondad de ajuste con una distribución discreta con un nivel de confianza del 95% y arrojó el resultado de la tabla 1.

Tabla 1. Prueba de Kolmogorov-Smirnov para el ajuste de las 8.000 observaciones mensuales generadas

Distribución	Binomial(264; 0,852956)
Estadístico KS	9,03x10 ⁻³
Valor p	0,527

donde KS es la abreviatura de Kolmogorov-Smirnov.

Luego se parametrizó la distribución binomial en términos del vector de parámetros $\theta' = [\mu, \sigma^2]$ y se supuso que, al desconocer las ventas diarias, se tendría una estimación de su distribución de probabilidad dada por Binomial($\frac{1}{30}\theta$). En este caso, el resultado es Binomial(9; 0,852956).

Conociendo las distribuciones de ventas mensuales y diarias, se generaron 7.620 observaciones de $\tilde{R}^r(A_j)$ y $R^r(A_j)$, $j=1, 2, \dots, 12$, a las cuales se les realizó una prueba de bondad de ajuste, con los resultados que muestra la tabla 2.

Tabla 2. Distribuciones para $\tilde{R}^r(A_j)$ y $R^r(A_j)$.

A_j	$\tilde{R}^r(A_j)$ y $R^r(A_j)$	KS	VALOR p	AD	VALOR p
1	LogN(-0,5733,0,44948,1,73962E-02)	6,70E-03	0,863	2,98	0,94
2	LogN(-0,519226,0,410141,1,89091E-02)	7,22E-03	0,796	0,298	0,94
3	LogN(-0,246491,0,211119,2,41345E-02)	1,12E-02	0,268	0,666	0,588
4	LogN(-0,244596,0,209835,2,56013E-02)	1,07E-02	0,318	0,798	0,483
5	LogN(-0,302889,0,257905,2,61828E-02)	1,06E-02	0,328	0,851	0,446
6	Beta(0,788804,1,26737,17,4135,23,1054)	7,72E-03	0,723	0,34	0,905
7	LogN(-0,160265,0,144657,3,4887E-02)	6,62E-03	0,872	0,413	0,836
8	LogN(-9,24807E-02,8,75519E-02,4,12505E-02)	5,70E-03	0,956	0,244	0,973
9	LogN(-1,46594E-02,1,79873E-02,5,08777E-02)	7,02E-03	0,822	0,301	0,937
10	LogN(-4,94681E-02,5,29852E-02,5,92457E-02)	7,34E-03	0,779	0,288	0,947
11	LogN(-1,31473E-02,-4,35388E-03,7,97683E-02)	5,80E-02	0,95	0,25	0,97
12	LogN(0,133338,-0,180109,0,126276)	7,19E-03	0,8	0,371	0,877

donde KS y AD son los estadísticos de las pruebas de Kolmogorov-Smirnov y Anderson-Darling, respectivamente. Además, la notación $\text{LogN}(\mu, \sigma, x_0)$ hace alusión a una distribución lognormal con parámetros μ y σ y punto de localización (inicio de la distribución) en x_0 .

Se realizaron entonces 100 simulaciones (indexadas $s=1, 2, \dots, 100$) de ventas y reclamaciones para $d=0, 1, 2, \dots, 364$, con los conteos $n_i(a)$ siguiendo una distribución Poisson ($\lambda=0,002$) y $f(r)$ dada por:

$$f(r) = \begin{cases} N_T^D(7, 2, 0, 15) & \text{si } 0 \leq r \leq 15 \\ N_T^D(22, 4, 16, 30) & \text{si } 16 \leq r \leq 30 \\ N_T^D(37, 2, 31, 45) & \text{si } 31 \leq r \leq 45 \end{cases}$$

Y para cada una de ellas se calcularon $\lambda_i(a)$ y $V_p(\hat{\lambda}(a))$ de acuerdo con (3) y (5), respectivamente; para luego calcular:

$$\hat{\Lambda}(A_j) = \sum_{u=0}^{\infty} \hat{\lambda}(u)$$

$$\tilde{\Lambda}_j = \sum_{u=1}^j \hat{\Lambda}(A_j)$$

$$V_p^{(A)}(\hat{\Lambda}(A_j)) = \sum_{u=0}^{\infty} V_p(\hat{\lambda}(u)) \quad (14)$$

También se calcularon $\tilde{\Lambda}_j(A_j)$, $\tilde{R}^r(A_j)$, para después determinar el primer término de (13) reemplazándolos en (9).

Finalmente, acudiendo a los resultados mostrados en la tabla 2 y a los $\tilde{\Lambda}_j(A_j)$ obtenidos, se calculó el segundo término de (13) para $j=1, 2, \dots, 12$, siguiendo los pasos descritos a continuación:

1. Generar 10.000 valores de $k_{s,m} = \tilde{R}_s^r(A_j) / R_s^r(A_j)$, $m=1, 2, \dots, 10.000$.
2. Calcular: $\hat{\Lambda}_{s,m}(A_j) = \tilde{\Lambda}_j(A_j) \times k_{s,m}$

1 La notación binomial (n, p) denota una distribución binomial con n ensayos y probabilidad de éxito p.

3. Calcular la varianza muestral para la s-ésima simulación:

$$\hat{V}\left(\hat{\lambda}_s(A_j) \frac{R^r(A_j)}{\hat{R}^r(A_j)}\right) = \frac{1}{9999} \sum_{i=1}^{10000} \left[\hat{\lambda}_{s,\mu}(A_j) - \frac{1}{10000} \sum_{i=1}^{10000} \hat{\lambda}_{s,\mu}(A_j) \right]^2$$

5. ANÁLISIS DE RESULTADOS

La idea central del trabajo es realizar una comparación entre análisis de datos de garantía desagregados y agregados para determinar el efecto que tiene la agregación en los resultados finales. Por motivos de espacio y practicidad, aquí sólo se muestran las comparaciones y resultados para $j=3, 7$; para los otros valores de j , los resultados son análogos.

En la figura 1 se muestran las comparaciones de $\hat{\lambda}_s(A_j)$ con $\bar{\lambda}_s(A_j)$ para $j = 3, 7$. Note que prácticamente no hay diferencia entre ambos estimadores, esto se verifica en la tabla 3 donde se prueba la hipótesis nula $E(\hat{\lambda}_s(A_j)) = E(\bar{\lambda}_s(A_j))$ a partir de los estimadores muestrales

$$\bar{\lambda}_s(A_j) = \frac{1}{10000} \sum_{i=1}^{10000} \hat{\lambda}_{s,i}(A_j) \quad \bar{\lambda}_r(A_j) = \frac{1}{10000} \sum_{i=1}^{10000} \bar{\lambda}_{s,i}(A_j)$$

$$S_{1,j}^2 = \frac{1}{9999} \sum_{i=1}^{10000} (\hat{\lambda}_{s,i}(A_j) - \bar{\lambda}_s(A_j))^2 \quad S_{2,j}^2 = \frac{1}{9999} \sum_{i=1}^{10000} (\bar{\lambda}_{s,i}(A_j) - \bar{\lambda}_r(A_j))^2$$

La hipótesis nula se acepta para $j= 1, 2, \dots, 10$.

Con respecto a los estimadores de la varianza, la figura 2 muestra que existen diferencias entre $V_p^{(D)}(\hat{\lambda}(A_j))$ y $V_p^{(A)}(\hat{\lambda}(A_j))$. Para medir mejor el efecto de la agregación en la varianza del estimador, se calculó la diferencia porcentual $\Delta_{j,s}$, $j=3, 7$ (figura 3), entre las estimaciones dadas por (14) y (13)

$$\Delta_{j,s} = \frac{V_p^{(A)}(\hat{\lambda}(A_j)) - V_p^{(D)}(\hat{\lambda}(A_j))}{V_p^{(D)}(\hat{\lambda}(A_j))} \times 100\%$$

Esta diferencia porcentual fluctúa entre el 3,8% y el 8% para $k=3$, y entre -0,5% y 9,2% para $k=7$. No obstante, en la tabla 3 se observa la similitud entre los estimadores de la varianza de las muestras Monte Carlo, $S_{1,j}^2$ y $S_{2,j}^2$.

Por otra parte, en la tabla 4 se muestra que la amplitud de los intervalos de confianza para $\hat{\lambda}_j$ y $\bar{\lambda}_j$, l y l^* , respectivamente es muy similar y cercana a cero su diferencia porcentual

$$\delta_j = \frac{|l - l^*|}{l} \times 100\%$$

Con todo esto se comprueba que las distribuciones de $\hat{\lambda}_s(A_j)$ y $\bar{\lambda}_s(A_j)$ son aproximadamente iguales y que la agregación de datos no tiene un efecto significativo en los resultados del análisis de datos de garantía.

Como nota final, se aclara que no se incluyen resultados para $j = 12$, porque son atípicos, lo mismo ocurre para $j = 11$. Esto ocurre porque el tiempo de garantía (Ω) es igual al tiempo de observación de la base de datos, lo cual genera, para el caso de datos desagregados, que en algunas réplicas de la simulación se obtengan valores de $R^r(a)$ cercanos a cero para $a \geq 350$ y, por lo tanto, valores de $\lambda(a)$ altos, así se presente un solo reclamo.

Tabla 3. Pruebas de hipótesis para la igualdad de medias de los estimadores de la tasa de reclamaciones con base en datos desagregados y agregados

J	$\hat{\lambda}_r(A_j)$	$\bar{\lambda}_r(A_j)$	$S_{1,j}^2$	$S_{2,j}^2$	z	Valor p
1	0,0624	0,0624	2,51E-05	2,51E-05	-0,015	0,506
2	0,0563	0,0563	2,54E-05	2,54E-05	-0,0145	0,5058
3	0,0626	0,0626	3,23E-05	3,23E-05	-0,0254	0,5101
4	0,0605	0,0605	3,34E-05	3,33E-05	0,0111	0,4956
5	0,0627	0,0627	4,08E-05	4,08E-05	0,0251	0,49
6	0,0606	0,0606	4,60E-05	4,59E-05	0,0259	0,4897
7	0,0629	0,0629	5,87E-05	5,85E-05	0,0466	0,4814
8	0,0631	0,0631	7,15E-05	7,12E-05	0,0671	0,4733
9	0,0614	0,0614	9,72E-05	9,65E-05	-0,0644	0,5257
10	0,0645	0,0645	1,64E-04	1,59E-04	0,3123	0,3774
11	0,0649	0,0644	4,09E-04	3,53E-04	1,7974	0,0361

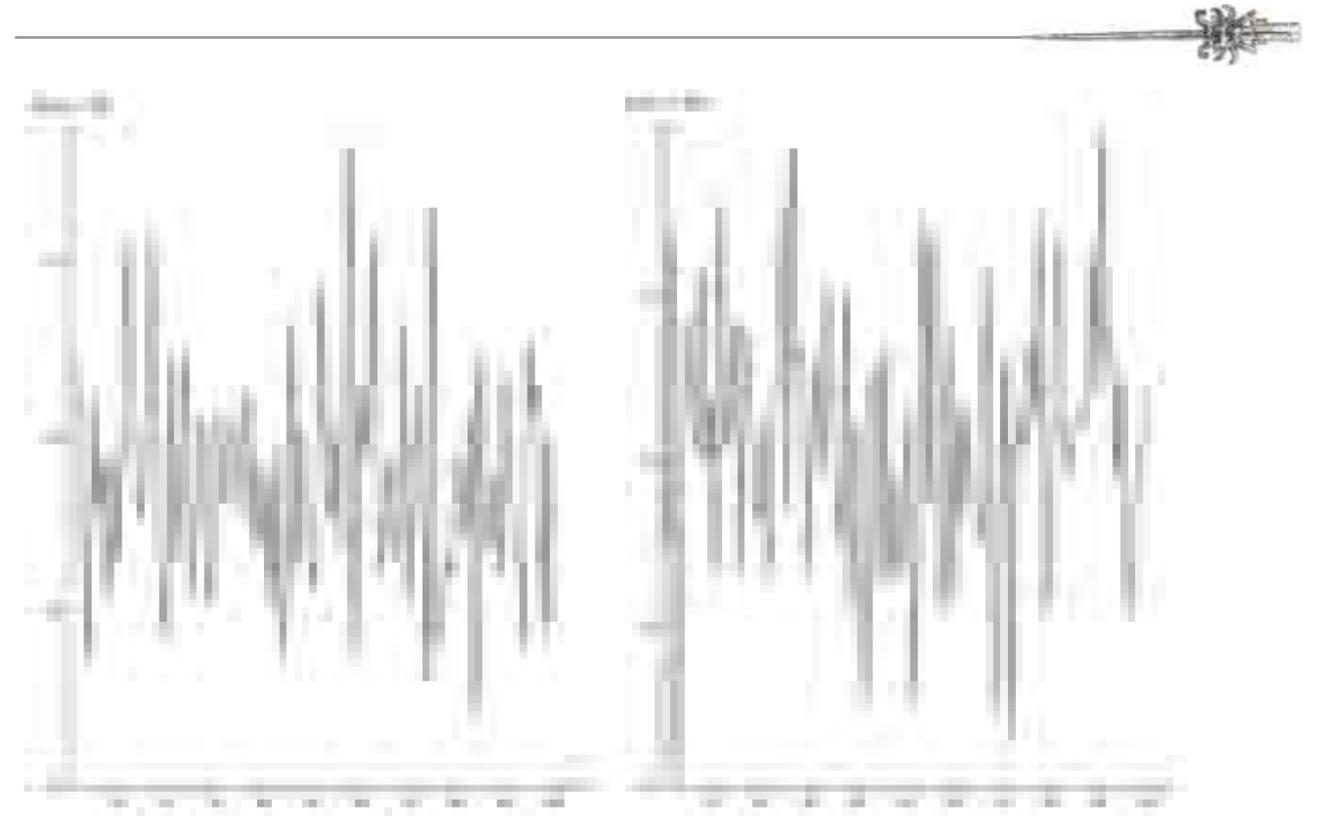


Figura 1. Gráficas comparativas de $\hat{\lambda}_s(A_j)$ con $\bar{\lambda}_s(A_j)$, $j = 3, 7$.

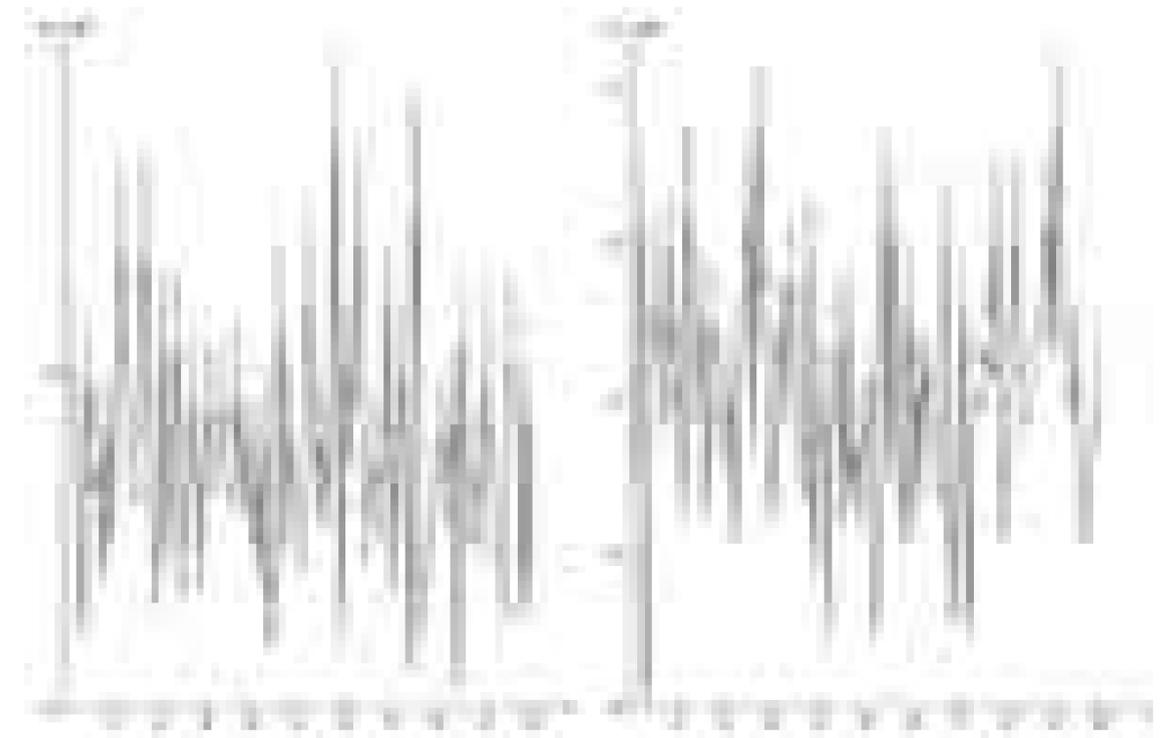


Figura 2. Gráficas comparativas de $V_p^{(D)}(\hat{\lambda}(A_j))$ con $V_p^{(A)}(\hat{\lambda}(A_j))$, $j = 3, 7$.



Figura 3. Gráficas de Δ_3 y Δ_7 .

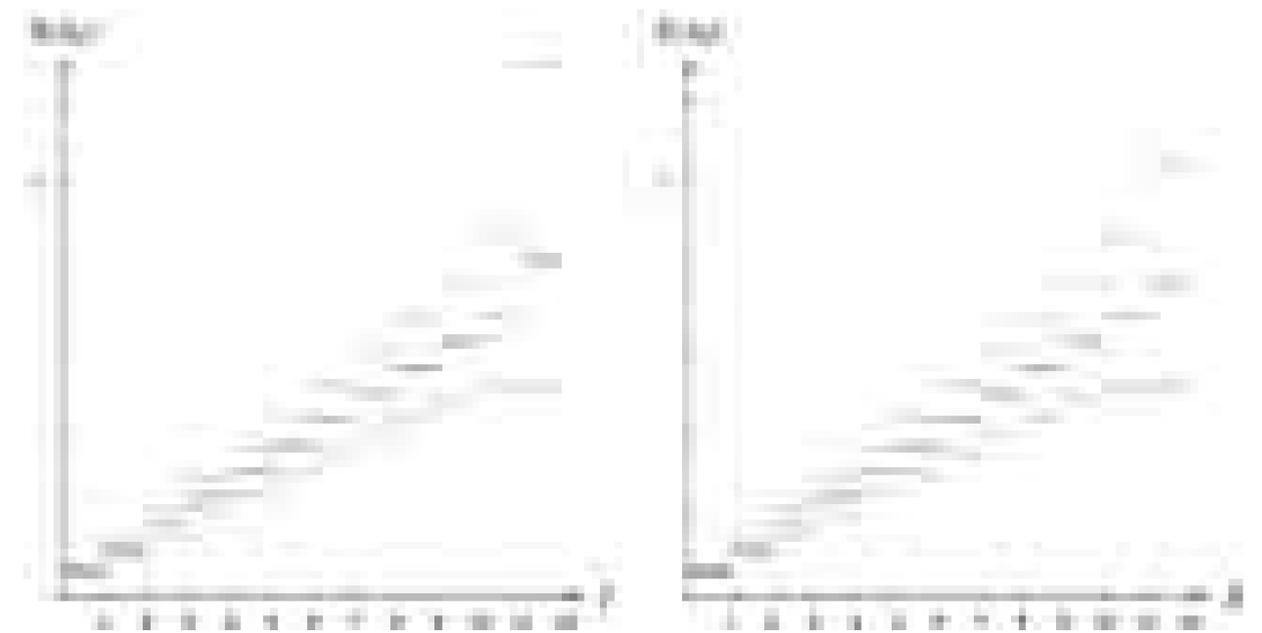


Figura 4. Intervalos de confianza Monte Carlo para los estimadores $\hat{\lambda}(A_j)$ y $\tilde{\lambda}(A_j)$.

Tabla 4. Intervalos de confianza Monte Carlo para los estimadores $\hat{\lambda}(A_j)$ y $\tilde{\lambda}(A_j)$.

j	$\hat{\lambda}_j$			$\tilde{\lambda}_j$			l	l'	δ
	Inf	Media	Sup	Inf	Media	Sup			
1	0,0527	0,0624	0,0723	0,0527	0,0624	0,0724	0,0196	0,0197	0,45%
2	0,0996	0,1187	0,1388	0,0996	0,1187	0,1389	0,0392	0,0394	0,37%
3	0,1513	0,1813	0,2129	0,1512	0,1813	0,213	0,0616	0,0618	0,40%
4	0,2008	0,2418	0,2848	0,2006	0,2418	0,2849	0,0840	0,0843	0,35%
5	0,2514	0,3045	0,3604	0,2512	0,3045	0,3604	0,1090	0,1092	0,13%
6	0,2991	0,3651	0,4346	0,2990	0,3651	0,4346	0,1356	0,1356	0,06%
7	0,3470	0,4280	0,5131	0,3472	0,4280	0,5131	0,1661	0,1659	-0,08%
8	0,3941	0,4911	0,5932	0,3942	0,4911	0,5929	0,1991	0,1988	-0,15%
9	0,4369	0,5525	0,6749	0,4371	0,5525	0,6745	0,2380	0,2373	-0,27%
10	0,4778	0,6170	0,7657	0,4782	0,6169	0,7650	0,2880	0,2868	-0,40%
11	0,5066	0,6819	0,8744	0,5100	0,6814	0,8692	0,3678	0,3591	-2,35%

4. CONCLUSIONES Y TRABAJO FUTURO

El presente artículo es sólo la parte inicial de un trabajo más amplio de comparación de análisis de datos de garantía desagregado y agregado que está en ejecución; en él se considerarán, además de otros escenarios de ventas diarias, variaciones en la naturaleza de generación de las reclamaciones; por ejemplo, un escenario para incluir consiste en suponer que las reclamaciones de distintas edades no son independientes.

Por el momento, se concluye que acudir al supuesto de Kalbfleisch et al., (1991) de suponer ventas constantes en un intervalo $[a,b]$, cuando sólo se tiene información acerca de las ventas totales de dicho intervalo, no introduce sesgo en el estimador de la tasa de reclamaciones de garantía ni incrementa su variabilidad; de hecho, las distribuciones de los estimadores con base en datos desagregados y agregados son casi iguales; es decir, la agregación de datos no tiene un efecto significativo en los resultados del análisis de datos de garantías.

BIBLIOGRAFÍA

Blischke, W. R. and D. N. P. Murthy (2000). Reliability: modeling, prediction and optimization, John Wiley & Sons.

Kalbfleisch, J. D., Lawless, J. F. and Robinson, J. A. (1991). Methods for the analysis and prediction of warranty claims. *Technometrics* 33(3), 273-285.

Lawless, J. F. (1998). Statistical analysis of product warranty data. *International Statistical Review* 66(1), 41-60.

Lawless, J. F. (2000). Statistics in reliability, *Journal of American Statistical Association* 95(451), 989-992.

Lawless, J. F. and Kalbfleisch, J. D. (1996). Statistical analysis of warranty claims data, in: W. R. Blischke and D. N. P. Murthy, eds. *Product Warranty Handbook*, Marcel Dekker, chapter 9.

Lawless, J. F. and Kalbfleisch, J. D. (1992). Some issues in the collection and analysis of field reliability data, *Survival analysis: state of the art*, pp. 141-152.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events, *Technometrics* 37(2), 158-168.

Lindsey, J. K. (1997). *Applying generalized linear models*, Springer.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical methods for reliability data*. John Wiley & Sons.

Wu, H. and Meeker, W. Q. (2002). Early detection of reliability problems using information from warranty databases. *Technometrics* 44(2), 120-133.