

# Métodos econométricos para la valoración cualitativa y cuantitativa del daño corporal en el seguro del automóvil

Miguel Ángel Santolino Prieto

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tesisenxarxa.net](http://www.tesisenxarxa.net)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tesisenred.net](http://www.tesisenred.net)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tesisenxarxa.net](http://www.tesisenxarxa.net)) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE BARCELONA

U

B

DEPARTAMENT D'ECONOMETRIA, ESTADÍSTICA I ECONOMIA ESPANYOLA

Risc en Finances i Assegurances

Institut de Recerca en Economia Aplicada

# Métodos econométricos para la valoración cualitativa y cuantitativa del daño corporal en el seguro del automóvil

Miguel Angel Santolino Prieto

Marzo 2007

Tesis doctoral para optar al Título de Doctor  
Directora: Dra. Mercedes Ayuso Gutiérrez  
Doctorado en Estudios Empresariales  
Especialidad en Ciencias Actuariales y Financieras  
Bienio 2002-2004

**B.U.B Secció d'Economia**  
Diagonal, 690, 08034 Barcelona  
T. 93 402 19 66

# Capítulo 6

## Predicción de la indemnización (I): Modelo lineal mixto para datos no balanceados

### 6.1 Introducción

Sin lugar a dudas, el modelo clásico de regresión lineal múltiple es el método más utilizado en econometría para el análisis de fenómenos económicos. Este modelo estadístico, bajo ciertos supuestos, tiene importantes ventajas que justifican su extensa aplicación en el campo de la economía, así como en otras áreas de conocimiento. En primer lugar, las estimaciones de los parámetros obtenidas, tanto por mínimos cuadrados ordinarios (MCO) como por máxima verosimilitud (MV), son óptimas entre una amplia gama de posibles estimadores. Por otro lado, la teoría subyacente en estos procesos de estimación, especialmente para el caso de MCO, es de sencilla comprensión y produce estimaciones fácilmente interpretables.

Desafortunadamente, en numerosas ocasiones el conjunto de supuestos sobre la forma del modelo y las relaciones entre sus partes en los que se basa la regresión lineal clásica son demasiado restrictivos para capturar la relación existente entre los datos. Los modelos lineales mixtos son una extensión del modelo lineal estándar en los que se eliminan algunas de estas limitaciones. Concretamente, el modelo lineal mixto permite correlación y heterocedasticidad entre los datos dotándose así de una mayor versatilidad, aunque todavía asume normalidad en la distribución. Robinson (1991) muestra, mediante ejemplos, como el empleo de modelos lineales mixtos es

de gran utilidad en campos aparentemente tan diversos como son la geología, la biología o la economía, entre otros. Una interesante aplicación de los modelos lineales mixtos en el ámbito actuarial es la presentada por Frees *et al.* (1999, 2001). Los autores demuestran, con datos longitudinales, la relación existente entre los modelos de credibilidad y los modelos lineales mixtos.

El supuesto de normalidad de los datos puede también relajarse, extendiéndolo a las distribuciones de la familia exponencial en los Modelos Lineales Generalizados Mixtos (MLGM). Antonio *et al.* (2005, 2006) proporcionan un excelente desarrollo de estos modelos en el campo actuarial.

El objetivo en el presente capítulo es estimar el coste de indemnización otorgado en sentencia judicial a una víctima de un accidente de tráfico, por el daño corporal sufrido. Para estimarlo, proponemos ajustar un modelo log-lineal mixto. Sugerimos que la compañía aseguradora puede considerar el coste de indemnización estimado, como el límite máximo de negociación con la parte contraria, por encima del mismo, debería ir a juicio. La estructura del capítulo se detalla a continuación. En el siguiente apartado se detallan las diferentes técnicas estadísticas que se aplican para datos balanceados y datos no balanceados. En la sección 6.3 se describe brevemente el modelo lineal mixto. Por último, se muestra la aplicación del modelo mixto para el cálculo del límite máximo de negociación en la indemnización económica.

## **6.2 Datos balanceados vs. datos no balanceados**

En el análisis de datos longitudinales, frecuentemente nos encontraremos que la estructura de los datos no es balanceada, es decir, que no disponemos del mismo número de observaciones para cada individuo, o éstas no fueron tomadas en el mismo momento del tiempo.

Una de las principales críticas a los modelos lineales multivariantes es su dificultad de aplicación cuando los datos son no balanceados. En este sentido, el modelo lineal (de efectos) mixto resuelve este inconveniente de una forma natural, y

puede ser utilizado para datos balanceados y no balanceados indistintamente. El modelo lineal mixto puede ser considerado como la combinación de dos regresiones en un único modelo estadístico. En una primera fase, se aplica una regresión lineal para estimar el perfil del individuo. Posteriormente, en una segunda fase, también mediante técnicas de regresión multivariante, se diferencia del perfil obtenido, entre los efectos fijos (aquellos comunes para todos los individuos) y los efectos propios del sujeto. Por este motivo, el modelo lineal mixto también es conocido como modelo de dos etapas (*two-stages model*).

Aunque la estructura del modelo mixto puede aplicarse a datos balanceados y no balanceados, la distinción entre ambos es fundamental, ya que las técnicas de inferencia a considerar no serán las mismas. Véase Searle *et al.* (1992) para el desarrollo de modelos mixtos con ambos tipos de datos.

### 6.3 Modelo lineal mixto

El modelo lineal mixto es una extensión del modelo de regresión lineal, en el que se incorporan efectos aleatorios en la estructura de la media. Supongamos que disponemos de una base de datos con  $N$  sujetos, donde  $n_i$  representa el número de observaciones para cada sujeto  $i$ ,  $1 \leq i \leq N$ . El modelo lineal mixto quedaría definido mediante la siguiente forma funcional,

$$y_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad (6.1)$$

donde  $y_i$  es el vector con  $n_i \times 1$  observaciones para el sujeto  $i$  y  $\boldsymbol{\beta} (p \times 1)$  es el vector de  $p$  parámetros desconocidos, común para todos los individuos (efectos fijos). Por su parte,  $\mathbf{b}_i (q \times 1)$  es el vector de  $q$  parámetros con efectos aleatorios para el sujeto  $i$ .  $\mathbf{X}_i (n_i \times p)$  y  $\mathbf{Z}_i (n_i \times q)$  son las matrices de los valores observados para los  $p$  efectos fijos y los  $q$  efectos aleatorios. Finalmente,  $\boldsymbol{\varepsilon}_i (n_i \times 1)$  es el vector de  $n_i$  perturbaciones aleatorias para el sujeto  $i$ . Cabe remarcar que en (6.1) no estamos imponiendo la condición de igualdad en el número de observaciones para cada sujeto, por lo que esta modelización puede aplicarse tanto para datos balanceados como para datos que no lo sean.

El modelo presentado en (6.1) asume que las observaciones de cada sujeto pueden ser ajustadas mediante un modelo de regresión lineal, en el cual algunos parámetros de la regresión son iguales para toda la población y otros, en cambio, son específicos de cada individuo. En general, siguiendo la nomenclatura de Verbeke y Molenberghs (2000), el modelo lineal mixto se define como aquel que cumple las siguientes condiciones,

$$\begin{aligned}
 y_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \varepsilon_i, \\
 \mathbf{b}_i &\sim N(0, \mathbf{D}), \\
 \varepsilon_i &\sim N(0, \boldsymbol{\Sigma}_i), \\
 \mathbf{b}_1, \dots, \mathbf{b}_N, \varepsilon_1, \dots, \varepsilon_N &\text{ son independientes.}
 \end{aligned} \tag{6.2}$$

En (6.2) asumimos los tradicionales supuestos de independencia entre sujetos,  $E[\varepsilon_i \varepsilon_j | \mathbf{X}] = 0$  si  $i \neq j$ , así como independencia entre los coeficientes de regresión aleatorios y los términos de perturbación,  $E[\mathbf{b}_i \varepsilon_j | \mathbf{X}] = 0 \quad \forall i, j$ . Del mismo modo, establecemos que tanto los coeficientes de regresión aleatorios como los términos de perturbación se distribuyen según una normal con media 0 y matriz de varianzas y covarianzas  $\mathbf{D} (q \times q)$  y  $\boldsymbol{\Sigma}_i (n_i \times n_i)$ , respectivamente, siendo ambas matrices definidas positivas.

De la anterior definición establecemos que  $y_i$  se distribuye según una normal multivariante (MVN) con valor esperado  $E[y_i] = \mathbf{X}_i \boldsymbol{\beta}$  y matriz de varianzas y covarianzas  $\text{Var}(y_i) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \boldsymbol{\Sigma}_i$ . Cabe señalar que la matriz  $\boldsymbol{\Sigma}_i$  depende del sujeto  $i$  únicamente para establecer su dimensionalidad ( $n_i \times n_i$ ), pero los parámetros que la forman son independientes del individuo considerado. Al igual que en el modelo de regresión general, los efectos fijos determinan la media y, con la introducción de los efectos específicos del sujeto (efectos aleatorios), modelizamos la estructura de la varianza.

### 6.3.1 Estimación de los parámetros

Consideremos la siguiente combinación en un solo modelo de los  $N$  modelos de regresión definidos en (6.2),

$$y = X\beta + Zb + \varepsilon, \quad (6.3)$$

donde  $y = (y_1', \dots, y_i', \dots, y_N')$ ,  $X = (X_1', \dots, X_i', \dots, X_N')$  y  $\varepsilon = (\varepsilon_1', \dots, \varepsilon_i', \dots, \varepsilon_N')$ . De forma similar,  $Z$  es la matriz con bloques  $Z_i$  en la diagonal principal y ceros en el resto. Supongamos que la matriz de varianzas y covarianzas del modelo es conocida,  $V = ZDZ' + \Sigma$ . Henderson (1950; 1959; 1963) demostró que cualquier  $\hat{\beta}$  y  $\hat{b}$  que sean solución de las siguientes ecuaciones lineales (conocidas como las ecuaciones del modelo mixto),

$$\begin{pmatrix} X'\Sigma^{-1}X & X'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & Z'\Sigma^{-1}Z + D^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}' \\ \hat{b}' \end{pmatrix} = \begin{pmatrix} X'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{pmatrix}, \quad (6.4)$$

son estimadores óptimos, entendiéndose como tales aquellos que satisfacen una serie de propiedades estadísticas deseables. En particular,  $\hat{\beta}$  es el vector de estimadores insesgados con mínima varianza (conocido por sus siglas en inglés, estimadores BLUE, *Best Linear Unbiased Estimators*) y coincide con los estimadores Mínimos Cuadrados Generalizados (MCG),  $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ . Por su parte,  $\hat{b}$  es el predictor<sup>65</sup> insesgado de mínima varianza (también conocido por su acrónimo en inglés BLUP, *Best Linear Unbiased Predictor*). De este modo, el predictor BLUP del efecto aleatorio para el individuo  $i$  tiene la siguiente estructura<sup>66</sup>,

---

<sup>65</sup> El término “predictor” se utilizó para diferenciarlo del término “estimador” en los efectos fijos. Goldberger (1962) fue la primera persona en acuñar la expresión *Best Linear Unbiased Predictor* para la estimación de los  $b$  que cumplieran las propiedades estadísticas de insesgabilidad y mínima varianza. Posteriormente, Henderson introduciría su acrónimo, BLUP.

<sup>66</sup> El modelo lineal mixto puede también estimarse mediante técnicas bayesianas. En este sentido, los efectos aleatorios  $b$ , se considerarían como variables aleatorias distribuidas a priori según una normal con valor esperado nulo y matriz de covarianzas  $D$ . Una vez observados los datos,  $Y_i = y_i$ , la distribución a posteriori de  $b$ , tendría la siguiente estructura,

$$f(b_i | y_i) = \frac{f(y_i | b_i) f(b_i)}{\int_{\Omega} f(y_i | b_i) f(b_i) db_i},$$

$$\hat{b}_i = \mathbf{DZ}'_i \mathbf{V}_i^{-1} (y_i - \mathbf{X}_i \hat{\beta}). \quad (6.5)$$

Nótese que el predictor definido en (6.5) se puede interpretar como el porcentaje de error  $e_i = y_i - \mathbf{X}_i \hat{\beta}$  atribuible al individuo  $i$ , determinándose dicho porcentaje como una ratio entre la covarianza de  $y_i$  y  $b_i$ ,  $\text{cov}(y_i, b_i) = \mathbf{DZ}'_i$ , y la varianza total de  $y_i$ ,  $\mathbf{V}_i$ . Puesto que  $\hat{\beta}$  y  $\hat{b}_i$  son funciones lineales de  $y_i$ , podemos definir sus varianzas como,

$$\begin{aligned} \text{var}(\hat{\beta}) &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \\ \text{var}(\hat{b}_i) &= \mathbf{DZ}'_i \left[ \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{X}_i (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'_i \mathbf{V}_i^{-1} \right] \mathbf{Z}_i \mathbf{D}, \end{aligned} \quad (6.6)$$

respectivamente. Nótese que  $\text{var}(\hat{b}_i)$  no tiene en cuenta la variación de  $b_i$ . Por este motivo, para calcular el error de estimación  $\hat{b}_i - b_i$  (o, más correctamente, error de predicción), normalmente se utiliza la varianza del error,  $\text{Var}(\hat{b}_i - b_i) = \mathbf{D} - \text{Var}(\hat{b}_i)$  (Laird y Ware, 1982; Verbeke y Molenberghs, 2000).

### 6.3.2 Estimación de la matriz de varianzas y covarianzas

Lamentablemente, en la realidad, casi nunca conocemos el valor de los componentes de la varianza y, por tanto, deben ser también estimados. Existe voluminosa literatura respecto a los métodos de estimación de los parámetros de la matriz de varianzas y

donde  $f(y_i | b_i)$  es la función de densidad de  $y_i$  condicionada a  $b_i$ ,  $f(b_i)$  es la función de densidad a priori de  $b_i$  y  $\Omega$  su espacio parametral. Si consideramos como estimador puntual de la distribución a posteriori de  $b_i$ , su valor esperado, obtenemos,

$$\hat{b}_i = \mathbb{E}[b_i | Y_i = y_i] = \mathbf{DZ}'_i \mathbf{V}_i^{-1} (y_i - \mathbf{X}_i \beta),$$

donde el estimador  $\hat{b}_i$  depende de los parámetros desconocidos  $\beta$ ,  $\mathbf{D}$  y  $\Sigma_i$ . Al sustituir estos parámetros por sus estimadores, obtenemos los predictores EBLUP de los efectos aleatorios (*Empirical Best Linear Unbiased Predictors*), también conocidos como estimadores EB (siglas en inglés de *Empirical Bayes*). Véase Laird y Ware (1982), Harville y Carriquiry (1992) o Kubokawa y Srivastava (2002), para una revisión extensa.



covarianzas (Searle *et al.*, 1992; Rao, 1997). Rao (1971) propuso un método no iterativo MIVQUE para la estimación de los componentes de la varianza, pero los más extensamente aplicados son los que se basan en la optimización de la función de verosimilitud. Para una comparación entre métodos, véase Swallow y Monahan (1984), Robinson (1987), Baltagi *et al.* (2002), entre otros.

Supongamos que  $\alpha$  es el vector de parámetros desconocidos que componen la matriz de varianzas y covarianzas de  $y_i$ ,  $V_i(\alpha)$ . De este modo, la función de verosimilitud a maximizar tiene la siguiente estructura,

$$L_{MV}(\beta, \alpha) = \prod_{i=1}^N \left\{ (2\pi)^{-n_i/2} \cdot \det(V_i(\alpha))^{-1/2} \cdot \exp\left(-\frac{1}{2}(y_i - X_i\beta)' V_i^{-1}(\alpha)(y_i - X_i\beta)\right) \right\}. \quad (6.7)$$

El método de máxima verosimilitud (MV) maximiza la función (6.7) respecto a los vectores de parámetros desconocidos  $\alpha$  y  $\beta$  de forma simultánea. A menudo nos encontraremos que el modelo lineal mixto que pretendemos estimar tiene un elevado número de efectos fijos. Patterson y Thompson (1971) propusieron un método de estimación que tiene en cuenta la pérdida de grados de libertad que se produce al estimar los efectos fijos. Ellos denominaron al procedimiento como “Máxima Verosimilitud Modificada” (o Máxima Verosimilitud Restringida, MVR). El método de MVR consiste en maximizar la función de verosimilitud conjunta, pero no para todos los datos, puesto que los efectos fijos no afectan a la estructura de error, sino para el conjunto de contrastes de error que tienen valor esperado nulo (véase Harville, 1977). La función de verosimilitud restringida puede escribirse como,

$$L_{MVR}(\beta, \alpha) = \det\left(\sum_{i=1}^N X_i' V_i^{-1}(\alpha) X_i\right)^{-1/2} \cdot L_{MV}(\beta, \alpha). \quad (6.8)$$

Con relación a que método de estimación utilizar, no existe una clara preferencia de uno de los dos en la literatura. Cuando los datos son balanceados, una ventaja de los estimadores MVR frecuentemente mencionada es que estos estimadores coinciden con los estimadores ANOVA. Ahora bien, Graybill y Hultquist (1961) ya demostraron que, si los datos son balanceados y normalmente distribuidos, los estimadores MV de los componentes de la varianza, aunque difieran

de los estimadores de análisis de la varianza (ANOVA), son los estimadores insesgados de mínima varianza (véase Robinson, 1987). En cambio, para muestras no balanceadas, existe cierto consenso de que el método de MVR tiene un mejor comportamiento. En general, debemos esperar que se acentúen las diferencias entre ambos métodos a medida que aumente el número de parámetros fijos a estimar. En cualquier caso, sea cual sea el método elegido, tanto la función (6.7) como la (6.8) debe ser maximizada mediante algún proceso iterativo de maximización como, por ejemplo, el algoritmo de Newton-Rapshon o BHHH (véase Kenneth, 2002).

Para ambos métodos de estimación, los parámetros del vector  $\alpha$  pueden estar limitados a un determinado espacio parametral  $\psi$ . Este hecho ocurre, por ejemplo, en el caso de las varianzas, es decir, para los parámetros de la diagonal principal de la matriz de covarianzas, ya que no pueden tomar valores negativos, y provoca que sus estimadores sean, en el mejor de los casos, aproximadamente insesgados (Harville y Jeske, 1992; Sun *et al.*, 2003).

### 6.3.3 Contraste de hipótesis

Una vez estimado el modelo lineal mixto, el siguiente paso debe consistir en realizar inferencia estadística sobre sus parámetros.

#### *Contraste de hipótesis sobre los parámetros de los efectos fijos*

En general, estaremos interesados en contrastar la siguiente hipótesis lineal:

$$H_0 : \mathbf{L}\boldsymbol{\beta} = 0, \quad (6.9)$$

para una matriz conocida  $\mathbf{L}$ . Existen  $F$ - y  $t$ -estadísticos de (6.9) únicamente para algunos modelos balanceados y para casos muy especiales de modelos no balanceados. En la mayoría de ocasiones, por tanto, los estadísticos de contraste serán aproximadamente  $F$ - o  $t$ -distribuidos,

$$F = \frac{\left[ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right]' \mathbf{L}' (\mathbf{L} \hat{\text{var}}(\hat{\boldsymbol{\beta}}) \mathbf{L}')^{-1} \mathbf{L} \left[ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right]}{\text{rango}(\mathbf{L})} \sim F_{\text{rango}(\mathbf{L}), \hat{\nu}}^* \quad (6.10)$$

donde  $\hat{\text{var}}(\hat{\beta})$  es la matriz de covarianzas estimada de  $\hat{\beta}$ , es decir, sustituyendo  $V$  por su estimador  $\hat{V}$  en la primera ecuación de (6.6). Los grados de libertad del denominador  $\hat{v}$  deben ser estimados de los datos, existiendo en la literatura métodos alternativos de estimación (por ejemplo, en la opción DDFM del procedimiento Proc Mixed de SAS V.9.1.3., se ofrecen 5 métodos diferentes implementados para calcularlos<sup>67</sup>). Ahora bien, sólo en el caso de muestras muy pequeñas, cabe esperar diferencias apreciables en el nivel de significación de los parámetros en función del método elegido para el cálculo de los grados de libertad (Verbeke y Molenberghs, 1997; 2000).

### ***Contraste de hipótesis sobre los componentes de la varianza***

En el caso de realizar contrastes sobre los parámetros de la varianza en modelos balanceados, bajo los supuestos de normalidad e independencia de los efectos aleatorios y los componentes de error, los estadísticos óptimos casi siempre existen y, además, coinciden con los  $F$ -estadísticos (Mathew y Sinha, 1988). Para modelos no balanceados, en cambio, muy raramente se podrá derivar un estadístico de contraste óptimo, aún en el caso de que existan  $F$ -estadísticos exactos (Mathew y Sinha, 1992; Öfversten, 1993; Khuri, Mathew y Sinha, 1998). El tradicional análisis de estos modelos se basa en  $F$ -estadísticos aproximados. Especialmente, mediante el contraste de Wald, en el que dividimos el parámetro estimado por su error estándar. Este estadístico es válido sólo para muestras grandes. El error estándar asintótico de los estimadores de los parámetros de la varianza se obtiene de la inversa de la matriz (con signo negativo) de las derivadas parciales de segundo orden de la función de verosimilitud. Lamentablemente, el contraste de Wald, al basarse en la aproximación normal, resulta inadecuado para parámetros cuyo valor real esté cercano a los límites del espacio parametral. Este sería el caso, por ejemplo, de los componentes de la varianza cercanos a cero.

---

<sup>67</sup> Contain, Betwithin, Residual, Satterth y Kenwardroger (SAS Institute Inc., 2003).

Un segundo tipo de contraste y probablemente el más utilizado, es el basado en el estadístico de la ratio de verosimilitud. En efecto, el problema en el contraste de hipótesis puede ser considerado como de elección entre dos modelos lineales diferentes, uno bajo la hipótesis nula y el otro, bajo la alternativa. De este modo, el primer modelo sería un caso especial del otro. El estadístico de contraste es, por tanto,  $-2$  veces el valor del logaritmo de la función de verosimilitud maximizada bajo la hipótesis nula y la hipótesis alternativa. Este estadístico sigue una distribución  $\chi^2$  con grados de libertad igual a la diferencia entre los grados de libertad de los dos modelos.

Por último, un procedimiento alternativo para realizar el contraste de hipótesis, sería mediante la simulación de la distribución empírica de los parámetros. Tsui y Weerahandi (1989) introdujeron el concepto de inferencia generalizada y definieron la variable de contraste generalizada para el contraste de hipótesis (Véase Apéndice E). Autores como Zhou y Mathew (1994) y Khuri *et al.* (1998) han comparado esta metodología con estadísticos aproximadamente  $F$ -distribuidos para diferentes situaciones. En el capítulo 7 nos basaremos en la noción de inferencia generalizada para el cálculo de intervalos de confianza de los parámetros de la varianza.

#### **6.4 Aplicación a la valoración de daños corporales**

En la primera parte del capítulo hemos definido el modelo lineal mixto con datos no balanceados. En esta sección implementamos el modelo descrito para la estimación del coste de indemnización de los siniestros de daños personales en el seguro del automóvil. Concretamente, nos situamos en el momento que la víctima está totalmente recuperada y la compañía debe decidir negociar con la parte contraria o ir a juicio. Dado que nuestra muestra está compuesta por siniestros de daños corporales cuya indemnización fue concedida mediante sentencia judicial, planteamos que la compañía puede utilizar la metodología propuesta para estimar el límite económico máximo de negociación de la indemnización con el demandante. Al final del capítulo 7 mostramos la aplicación del método presentado en el cálculo de las reservas por siniestros pendientes de liquidación.

### 6.4.1 El modelo de regresión

Como señalábamos en el capítulo 4 la muestra ha sido elaborada a partir de una selección aleatoria de expedientes de siniestros con víctimas en accidentes de circulación de una de las principales compañías aseguradoras del ramo del automóvil del mercado español. La característica en común de los siniestros de la muestra es que la indemnización económica ha sido fijada mediante sentencia judicial, al no haberse llegado a un acuerdo amistoso previo de indemnización entre las partes.

La información proporcionada por la muestra hace referencia a 159 expedientes de siniestros que representan un total de 204 víctimas. Para cada lesionado, disponemos de información sobre su lesión en diferentes momentos durante el tiempo que el expediente del siniestro ha estado abierto dentro de la compañía aseguradora (apertura, informes periciales, forense). Además, conocemos la indemnización económica otorgada finalmente a la víctima (sentencia en firme y, en su caso, apelación). En el capítulo 5 hemos demostrado que, mediante la utilización de modelos lineales generalizados, la compañía aseguradora puede mejorar la estimación del nivel de gravedad de la víctima con respecto a la gravedad considerada directamente por los médicos periciales de la compañía y por el médico forense. Al final de dicho capítulo mostrábamos como esta mejor clasificación de la gravedad permite mejorar el cálculo de provisiones por siniestros pendientes de la compañía.

En el presente apartado se persigue estimar, mediante una regresión lineal multivariante, el coste que le supondrá a la compañía aseguradora indemnizar a la víctima. Para obtener dicha estimación, nos situamos en una fase crucial de la vida del siniestro: cuando la víctima está totalmente recuperada, pero aún no se ha celebrado el juicio. En ese momento, la compañía dispone de los informes periciales realizados por sus médicos y, en determinados casos, del informe forense. En concreto, utilizaremos la información recogida en el primer y en el último informe pericial (con el fin de reflejar una primera valoración médica de la lesión, y la valoración cuando la víctima está totalmente recuperada) e incluiremos, además, una variable de control que indique la actuación del médico forense.

Si la compañía aseguradora dispone de una estimación fiable del coste total de la indemnización, puede utilizar dicha estimación como herramienta de negociación con la parte contraria, así como base para provisionar el coste del siniestro. No olvidemos que, en la mayoría de las ocasiones, tanto la compañía como el lesionado, estarán interesados en llegar a un acuerdo amistoso sobre la cuantía de compensación por el daño corporal y, de este modo, evitar un proceso judicial que puede llegar a ser muy costoso y dilatado en el tiempo.

En ocasiones, la compañía aseguradora no puede realizar los correspondientes informes médico-periciales. Las causas más frecuentes son el desconocimiento de la residencia de la víctima o que ésta no ha permitido que se le visitara, entre otras. Para el siguiente estudio, consideraremos únicamente aquellos casos en los que la compañía dispone de ambos informes periciales. En este sentido, la compañía aseguradora había realizado los dos informes periciales para 114 lesionados del conjunto de la muestra. En la Tabla 6.1 presentamos la definición y los principales descriptivos de las variables utilizadas para la estimación del modelo. Dichas variables (excepto *indemn* y *forense*), ya han sido definidas en el capítulo 5.

**TABLA 6.1 Variables explicativas incluidas en el modelo y estadísticos descriptivos**

		Media	Desv.Est
$y$	<i>indemn</i> Logaritmo del total de la indemnización concedida por el juez.	8,260	1,306
$x_1$	<i>turi</i> 1 si el vehículo de la víctima es un turismo; 0 en caso contrario.	0,596	0,493
$x_2$	<i>sexo</i> 1 si es un hombre; 0 si es mujer.	0,509	0,502
$x_3$	<i>edad</i> Edad de la víctima (en decenas).	3,868	1,686
$x_4$	<i>mismo</i> 1 si se realizó un único informe pericial (mismo informe el inicial y el final); 0 en caso contrario.	0,316	0,467
$x_5$	<i>sec_ul</i> Número de secuelas concedidas en el último informe.	1,114	1,655
$x_6$	<i>varsec</i> Número de secuelas concedidas en el último informe menos el número de secuelas del primer informe.	0,009	0,917
$x_7$	<i>dbi_ul</i> Número de días de baja incapacitado para el trabajo según el último informe.	53,132	63,027
$x_8$	<i>dbni_ul</i> Número de días de baja no incapacitado para el trabajo habitual según el último informe.	37,596	59,699
$x_9$	<i>vardbi</i> Número días de baja incapacitado para el trabajo según el último informe menos el número de días de baja incapacitado según el primer informe.	2,079	37,601
$x_{10}$	<i>forense</i> 1 si forense puntúa secuelas; 2 si forense marca las secuelas pero no las puntúa; 3 si forense no concede secuelas; 4 si no hay informe forense.	2,342	0,930

Nuestro objetivo consistirá en ajustar un modelo econométrico que explique la indemnización concedida a la víctima  $i$  (variable *indemn*) mediante el resto de regresores de la Tabla 6.1. Las variables *turi*, *sexo* y *edad* recogen información del siniestro antes de los informes periciales, es decir, en el momento de apertura del expediente del siniestro por parte de la compañía. El resto de variables, excepto la última, reflejan aspectos de la lesión considerados por el médico perito de la compañía, por tanto, después de los correspondientes informes periciales. En el modelo de regresión propuesto no se ha tomado directamente la información proporcionada por el primer informe pericial, sino las variaciones entre informes, con el fin de evitar problemas de multicolinealidad. Por último, la variable *forense* es una variable categórica con 4 categorías:

- *forense*=1. El médico forense ha visitado a la víctima, ha considerado que tiene secuelas y ha valorado la gravedad de las mismas, otorgándoles una puntuación en base a la horquilla fijada en el correspondiente baremo médico.
- *forense*=2. El médico forense ha visitado a la víctima, ha considerado que tiene secuelas y se ha limitado a definir las, pero sin valorar su gravedad<sup>68</sup>.
- *forense*=3. El médico forense ha visitado a la víctima y ha considerado que no tiene secuelas.
- *forense*=4. No se ha realizado informe forense. Éste es el caso en que la demanda sigue el cauce civil y no el penal.

En el capítulo 3 describimos la acción judicial y señalábamos en qué situaciones el médico forense debía realizar un informe médico del lesionado. Aunque la compañía aseguradora tiene el derecho de recibir el informe médico-forense que se le haya realizado a la víctima que debe indemnizar, en ocasiones, no lo obtiene hasta poco antes de la celebración del juicio. Por este motivo, en relación al conjunto de información proporcionado por el informe forense, en el modelo de

---

<sup>68</sup> El forense no está obligado a puntuar las secuelas, sino únicamente a que las secuelas que otorgue coincidan con las descritas en el baremo legal médico.

regresión propuesto hemos incluido únicamente la variable *forense*. Esta variable, al considerar sólo información general del informe, es previsible que la compañía aseguradora la obtenga en un período de tiempo más corto.

Supondremos en nuestra modelización que la relación entre la variable dependiente  $y_i$  (logaritmo de la indemnización otorgada) y las variables explicativas es lineal. Ahora bien, nuestra base de datos presenta ciertas características que nos inducen a pensar que un ajuste clásico de regresión lineal puede producir perturbaciones heterocedásticas y/o correlacionadas.

Por un lado, como hemos mencionado en capítulos anteriores, el informe forense es la única información médica no-parcial de la que dispone el juez para fijar la indemnización. En este sentido, parece lógico pensar que, probablemente, existirá una mayor variabilidad cuando el juez no dispone del informe forense que si dispone de él, o cuando el forense no ha valorado las secuelas que cuando les ha otorgado una puntuación. En conclusión, deberemos contrastar si la varianza de las perturbaciones es diferente en función del valor que tome la variable categórica *forense*, es decir, si el modelo de regresión lineal presenta heterocedasticidad por grupos en el término de perturbación. En la Tabla 6.2 presentamos las frecuencias observadas en la muestra para las diferentes categorías de la variable *forense*.

**TABLA 6.2 Frecuencias para la variable *forense***

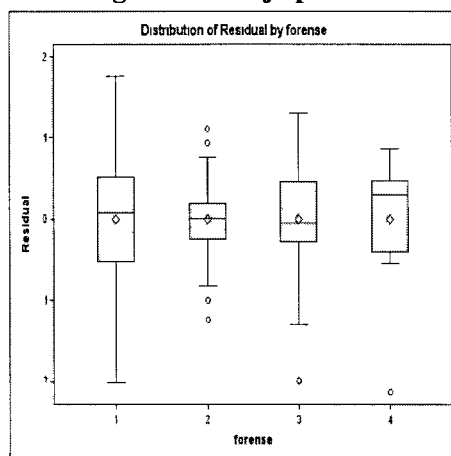
<i>Categoría (k)</i>	<i>Número de casos (N<sub>k</sub>)</i>	<i>Frecuencia (f<sub>k</sub>)</i>
<i>forense=1</i>	24	0,211
<i>forense=2</i>	39	0,342
<i>forense=3</i>	39	0,342
<i>forense=4</i>	12	0,105
<i>Total</i>	114	1,000

Al analizar el gráfico de los residuos para la variable *forense* (Gráfico 6.1), obtenidos a partir del ajuste de un modelo de regresión lineal clásico para explicar la variable dependiente, tomando como regresores el resto de variables de la Tabla 6.1,



observamos que la varianza del término de perturbación no parece independiente de la actuación del médico forense, mostrándose una evidencia gráfica de la idea previa de potencial heterocedasticidad por grupos<sup>69</sup>.

**GRÁFICO 6.1 Diagrama de caja para la variable *forense***



Además, hemos de tener en cuenta que, aunque el juez sigue las pautas marcadas en el baremo legal vigente para la valoración económica del daño corporal de la víctima, el sistema de valoración es flexible para poder ajustarse adecuadamente a las peculiaridades de cada caso (elección del porcentaje de corrección por perjuicio económico, deuda dineraria o deuda de valor, tipo de interés a aplicar, etc.). Esta flexibilidad puede introducir elementos de subjetividad por parte del juzgador, por lo que podría suceder que la indemnización final variase en función de la persona que la valore. En otras palabras, que existiría una propensión constante e individual de cada sujeto de desviarse respecto al valor medio, siempre en la misma dirección. De lo anterior se deriva, por tanto, que si un mismo individuo fija la indemnización para más de un lesionado, estos casos puedan estar correlacionados.

Por otro lado, es evidente que no es posible recoger en un modelo econométrico todos los factores que pueden influir en la determinación de la indemnización económica. La indemnización final dependerá de elementos tan

<sup>69</sup> El contraste de verosimilitud es significativo,  $-2(\ln L_0 - \ln L_1) = 9,26$  (0,05).

diversos como la actuación de los abogados durante el juicio, si son convincentes los testigos o el tipo de pruebas presentadas, entre muchos otros. Parece lógico pensar, por tanto, que aquellos casos cuyas indemnizaciones hayan sido determinadas en una misma sentencia judicial, puedan violar el supuesto clásico de independencia entre observaciones.

Nótese que los supuestos considerados en los dos párrafos anteriores, aunque guardan relación, son diferentes. En el primer párrafo planteamos que es posible que el juez, por su propia condición humana, influya en la indemnización y, por tanto, se debería calibrar el modelo teniendo en cuenta la desviación respecto al valor esperado medio de indemnización que es atribuible al individuo que establece la indemnización. La estimación del efecto individual esperado debería basarse en la jurisprudencia disponible para el sujeto considerado en situaciones análogas.

En cambio, en el siguiente párrafo, suponemos que las víctimas indemnizadas en una misma sentencia tienen multitud de efectos comunes que influyen en la cuantía final de las compensaciones económicas otorgadas. Uno de estos efectos, pero no el único, sería que hayan sido dictaminadas por el mismo individuo. La mayoría de los factores que influyen en las indemnizaciones fijadas en una misma sentencia no serán directamente observables por el investigador, por lo que no podrán recogerse en variables explicativas, pero se deberán tener en cuenta, considerando la potencial correlación entre estas observaciones.

El tratamiento econométrico para ambos casos es muy similar. El primer supuesto (efecto individual del sujeto) podría considerarse como una extensión del segundo (correlación entre observaciones de una misma sentencia). Ambos modelos compartirán la misma estructura de la varianza del término de error, considerando correlación entre las perturbaciones. La diferencia estriba en el cálculo de los efectos individuales. En el primer caso, la jurisprudencia existente puede ayudar a calibrar la desviación individual esperada del individuo en situaciones análogas a la estudiada. Por el contrario, si consideramos el caso de correlación entre observaciones de una misma sentencia, muchos de los factores no son observables, por lo que deberemos

suponer que todos los factores que influyen en una sentencia no volverán a repetirse en las mismas circunstancias (no existen dos sentencias iguales). Por tanto, si nuestro objetivo es predecir futuras indemnizaciones, no tendrá sentido ajustar el modelo con los perfiles individuales de las sentencias, y deberemos basar la predicción únicamente en el valor esperado medio de indemnización.

Cuando no se disponga de valoraciones anteriores realizadas por un mismo sujeto en situaciones análogas a la estudiada, únicamente se podrá calcular el valor esperado medio de la compensación, como sucede en el caso de correlación de observaciones de una misma sentencia. En este caso, estaremos asumiendo que existe autocorrelación en el término de perturbación derivada de que el mismo sujeto establece las indemnizaciones, pero consideraremos que el efecto individual esperado es nulo, debido a la falta de información sobre compensaciones previas. En nuestra muestra existe una total equivalencia entre individuos y sentencias, es decir, cada sentencia ha sido dictada por una persona distinta, por lo que no nos es posible diferenciar entre los dos modelos planteados (Tabla 6.3). En el presente estudio proponemos estimar un modelo lineal (de efectos) mixto donde el sujeto pasa a ser la sentencia analizada, en la cual se dictamina la indemnización a pagar por la compañía. De este modo, las indemnizaciones económicas concedidas a los diferentes lesionados en una misma sentencia, consideramos que son observaciones de un mismo sujeto.

**TABLA 6.3 Frecuencias para la variable *sentencia***

	<i>Número de casos (n<sub>g</sub>)</i>	<i>Frecuencia (f<sub>g</sub>)</i>
<i>Sentencia afecta a un lesionado</i>	74	0,804
<i>Sentencia afecta a dos lesionados</i>	14	0,152
<i>Sentencia afecta a tres lesionados</i>	4	0,044
<i>Total</i>	92	1,000

Como señalábamos al inicio de este capítulo, el modelo lineal mixto se interpreta como un modelo en dos etapas. En una primera fase, estimaremos los efectos fijos comunes para todos los individuos, teniendo en cuenta la potencial

existencia de heterocedasticidad y autocorrelación en las perturbaciones. Si consideramos que la correlación viene motivada porque las compensaciones económicas fueron otorgadas en la misma sentencia (y, por consiguiente, han influido multitud de factores no observables), nos detendremos aquí y no pasaremos a la segunda fase. Del mismo modo, si suponemos que la correlación se debe principalmente a que la valoración ha sido realizada por el mismo individuo, cuando queramos predecir la indemnización económica otorgada por una persona de quién no disponemos de información previa, no podremos estimar la desviación esperada individual y, por tanto, nos basaremos, de nuevo, únicamente en los efectos fijos (sin pasar a la segunda fase). Por el contrario, si disponemos de información previa sobre valoraciones anteriores, podremos, en la segunda fase, calibrar el modelo lineal ajustado, estimando la desviación individual esperada del individuo respecto a la indemnización económica media (efecto aleatorio). En resumen, mediante esta modelización, la regresión clásica sería un caso específico del modelo lineal mixto en el que los efectos aleatorios  $b$  son iguales a cero y, por tanto, innecesaria la segunda fase de estimación. El modelo propuesto sigue la siguiente estructura,

$$\begin{aligned}
 y_i &= \mathbf{X}_i \boldsymbol{\beta} + \text{interacciones} + \mathbf{Z}_i b_i + \varepsilon_i, & i=1, \dots, 92, \\
 b_i &\sim N(0, \sigma_\alpha^2), \\
 \varepsilon_i &\sim N(0, \bigoplus_{k=1}^4 \sigma_{\varepsilon_k}^2 \mathbf{I}_{n_k}), & \sum_{k=1}^4 n_k = n_i, \\
 b_1, \dots, b_{92}, \varepsilon_1, \dots, \varepsilon_{92} &\text{ son independientes,}
 \end{aligned} \tag{6.11}$$

donde  $y_i$  es el vector con  $n_i \times 1$  observaciones para la sentencia  $i$ , siendo  $1 \leq n_i \leq 3$ ;  $k$  la  $k$ -ésima categoría de la variable *forense*, y el símbolo  $\bigoplus$  representa la suma directa de matrices.  $\boldsymbol{\beta}(11 \times 1)$  es el vector de parámetros desconocidos de los efectos fijos para las variables explicativas presentadas en la Tabla 6.1, incluyendo un término constante,  $\beta_0$ . Por su parte,  $b_i$  es el efecto específico del sujeto  $i$ , con valor esperado nulo y varianza  $\sigma_\alpha^2$ . Finalmente,  $\mathbf{X}_i(n_i \times 11)$  y  $\mathbf{Z}_i(n_i \times 1)$  son las matrices de valores observados y  $\varepsilon_i(n_i \times 1)$  es el vector de perturbaciones aleatorias con varianza del error dependiente de la categoría de la variable *forense* (heterocedasticidad por grupos). Fíjense que  $n_i$  no es igual para todas las sentencias (Tabla 6.3), sino que toma valor

uno para la mayoría de los casos (74 sentencias), dos para 14 sentencias, y sólo para 4 casos es igual a tres, por lo que los datos son fuertemente no-balanceados. En la literatura existen múltiples trabajos en los que se han estimado modelos lineales mixtos extremadamente no-balanceados (véase, por ejemplo, Harville y Fenech, 1985; Zhou y Mathew, 1994; El-Bassiouni, 1994; Park y Burdick, 2003; 2004; Arendacká, 2005).

Con la modelización presentada en (6.11) conseguimos dos objetivos. Por un lado, al igual que con los modelos lineales generales, al establecer la estructura de la matriz de varianzas y covarianzas, evitamos posibles problemas de autocorrelación y/o heterocedasticidad entre los datos. Por otro lado, y a diferencia de los modelos lineales generales, podemos estimar resultados teniendo en cuenta perfiles individuales de los sujetos, y no solo los valores esperados medios.

### 6.4.2 Resultados

El método de estimación de los parámetros ha sido el de maximización de la función de verosimilitud restringida<sup>70</sup> (como el implementado en SAS). En primer lugar, respecto a los supuestos sobre la estructura de la varianza del modelo, en la Tabla 6.4 se muestra el contraste de la ratio de verosimilitud. Mediante este contraste comparamos si es necesario estimar el modelo lineal mixto o, podríamos haber considerado un modelo clásico de regresión lineal, es decir, planteamos como hipótesis nula  $H_0: \sigma_\alpha^2 = 0$  y  $\sigma_{\epsilon,1}^2 = \dots = \sigma_{\epsilon,4}^2 = \sigma_\epsilon^2$ . Como puede observarse, el estadístico  $\chi$ -cuadrado, igual a 12,435, es significativo al 5%, con lo que no podemos aceptar el cumplimiento de las hipótesis del modelo de regresión clásico.

**TABLA 6.4 Contraste de ratio de verosimilitud**

$\chi$ -cuadrado	<i>g.l.</i>	<i>p</i> -valor
12,435	5	0,029

-2Residual lnL<sub>0</sub>=271,360; -2Residual lnL<sub>1</sub>=283,795.

<sup>70</sup> Simultáneamente se ha aplicado también el método de Máxima Verosimilitud, obteniendo resultados muy similares.

En relación a la estimación de los parámetros, de acuerdo con los resultados obtenidos (Tabla 6.5), observamos que a excepción de la variables *turi* y *edad*, el resto de efectos fijos tienen coeficientes significativos. Respecto la variable *sexo*, el signo negativo del parámetro indica que las mujeres reciben una mayor indemnización económica que los hombres. Por otro lado, el coeficiente de la variable *mismo* es significativo y con signo negativo, lo que nos estaría indicando que si el perito de la compañía aseguradora examinó una única vez a la víctima, la indemnización económica esperada será menor que si necesitó más de un chequeo médico. Este resultado parece lógico, puesto que los lesionados que hayan sido examinados una única vez serán probablemente los que tuvieron una recuperación más rápida y, por tanto, los menos graves.

**TABLA 6.5 Estimación de los parámetros (intervalos de confianza al 90%)**

		Coeficiente	Std.Error	p-valor	Límite inferior	Límite superior
$\beta_0$	Constante	8,402	0,290	0,000*	7,920	8,884
$x_1$	<i>turi</i>	-0,218	0,126	0,107	-0,442	0,005
$x_2$	<i>sexo</i>	-0,764	0,213	0,004*	-1,144	-0,384
$x_3$	<i>edad</i>	0,052	0,039	0,209	-0,018	0,121
$x_4$	<i>mismo</i>	-0,872	0,191	0,001*	-1,213	-0,531
$x_5$	<i>sec_ul</i>	0,218	0,052	0,001*	0,126	0,311
$x_6$	<i>varsec</i>	-0,259	0,089	0,013**	-0,419	-0,100
$x_7$	<i>dbi_ul</i>	0,009	0,001	0,000*	0,006	0,011
$x_8$	<i>dbni_ul</i>	0,005	0,001	0,001*	0,003	0,007
$x_9$	<i>vardbi</i>	-0,005	0,001	0,015**	-0,008	-0,002
$x_{10c}$	1 si <i>forense</i> =3; 0 en caso contrario	-0,714	0,14	0,000*	-0,970	-0,458
<i>inter</i>	1 si <i>sexo</i> =0 y <i>mismo</i> =0; 0 en caso contrario	-0,657	0,249	0,021**	-1,100	-0,213
$\sigma_\alpha^2$	Varianza del efecto aleatorio	0,015	0,117	0,450	0,010	1,22E80
$\sigma_{\varepsilon,1}^2$	Varianza del error si <i>forense</i> =1	0,761	0,253	0,001*	0,476	1,454
$\sigma_{\varepsilon,2}^2$	Varianza del error si <i>forense</i> =2	0,185	0,136	0,086***	0,076	1,144
$\sigma_{\varepsilon,3}^2$	Varianza del error si <i>forense</i> =3	0,481	0,148	0,001*	0,310	0,868
$\sigma_{\varepsilon,4}^2$	Varianza del error si <i>forense</i> =4	0,661	0,323	0,020**	0,345	1,876

Número de observaciones:114; \* indica nivel significación del 1%; \*\* indica nivel significación del 5% \*\*\* indica nivel de significación del 10%.

Respecto a las variables que recogen la gravedad considerada por el médico de la compañía aseguradora en el último informe pericial (*sec\_ul*, *dbi\_ul* y *dbni\_ul*), como cabía esperar, están positivamente relacionadas con la indemnización final de la víctima.

Los coeficientes de las variables *varsec* y *vardbi* son significativos. Este resultado es muy interesante puesto que nos está indicando que la información relativa al primer informe pericial tiene capacidad explicativa sobre la indemnización económica otorgada a la víctima, aunque dispongamos de la valoración médica de la víctima cuando ya está recuperada. Una tarea no obvia es la interpretación del signo de sus parámetros (ambos negativos). El que estos parámetros tengan signo negativo nos señalarían que si el perito médico ha otorgado un mayor número de secuelas o días de baja en el informe inicial que en el final, este diferencial hace incrementar el valor esperado de la indemnización. Estos resultados estarían corroborando los obtenidos en el capítulo 5, cuando valorábamos la gravedad de la víctima mediante modelos lineales generalizados. Una posible explicación que ofrecimos era que, en este primer informe, la compañía realiza una valoración “generosa” de la gravedad de la víctima porque busca la suficiencia en la provisión económica del siniestro pendiente. Por el contrario, en el último informe, la compañía realiza una valoración más “ajustada” porque es consciente de que, aquello que considere, será decisivo en el pago frente a la parte contraria. Esta interpretación no es aplicable cuando las variables *varsec* y *vardbi* toman un valor positivo. Ahora bien, el análisis detallado de los casos indica que en la mayoría de los siniestros en que existen diferencias de valoración de la gravedad entre informes periciales, en el inicial era más elevada que en el final.

Es importante remarcar que el coeficiente de la variable *edad* del modelo estimado no es significativo, mientras que en el capítulo 5 señalábamos que la edad y la gravedad de la lesión estaban positivamente relacionadas. Por otra parte, cuando describimos en el capítulo 3 el funcionamiento del baremo, recordábamos que la asignación económica legal por cada punto médico concedido era menor a medida que aumentaba la edad de la víctima. Por tanto, el efecto de la relación inversa entre

valor monetario del punto y edad, quedaría compensado con el de la relación positiva entre gravedad y edad, y, como consecuencia, la edad de la víctima no tendría capacidad explicativa sobre la indemnización económica final.

Por último, señalar que la variable *forense* también tiene incidencia en la estructura de la media. En particular, la tercera categoría dicotomizada tiene coeficiente significativo y valor negativo, por lo que cuando el forense haya visitado a la víctima y haya considerado que el accidente no le ha dejado secuelas médicas, la indemnización económica esperada será menor. Para el resto de categorías de la variable *forense*, se han obtenido parámetros no significativos y decidimos excluirlas del modelo.

Finalmente, observamos que la interacción entre la variable *sexo* y *mismo* tiene coeficiente significativo. Concretamente, si la víctima es mujer y los informes periciales son diferentes, la indemnización esperada disminuirá. Esta variable actúa como compensadora de ambos efectos por separado, es decir, si la víctima es mujer (*sexo=0*) la indemnización esperada es mayor, al igual que si los dos informes son diferentes (*mismo=0*), pero si se dan ambos efectos al mismo tiempo, la interacción (*inter=1*) reduce el valor esperado.

En relación a la estimación de los componentes de la varianza (Tabla 6.5), en primer lugar, observamos un reducido valor para el estimador del parámetro que mide la correlación de las indemnizaciones otorgadas en la misma sentencia  $\hat{\sigma}_c^2$ . Este resultado es lógico, puesto que, como mostrábamos en la Tabla 6.3, la mayoría de sentencias afectan a una única víctima. Además, en los casos que afecta a más de una, disponemos como máximo de tres observaciones. Respecto al resto de parámetros, resaltar que la mayor variación en el término de perturbación se produce en los casos que el forense ha puntuado las secuelas (*forense=1*) y en los que no hay informe forense (*forense=4*). Respecto a este último resultado, si el juez no dispone de información imparcial de la gravedad de la lesión, como sería el caso que no hay informe forense, parece razonable pensar que la dispersión respecto al valor esperado medio de compensación será mayor.



En cambio, un resultado contrario al que presuponíamos es que haya una mayor variación en el término de error cuando el forense puntúa las secuelas que cuando no las valora. El modelo mostrado en la Tabla 6.5 no recoge información sobre las secuelas consideradas por el forense, ni la puntuación que, en su caso, les haya otorgado, sino únicamente captura en la estructura de la varianza, si el forense consideró que la víctima tenía secuelas y si las puntuó (variable *forense*). Partiendo de la hipótesis de que la gravedad considerada por el médico forense ejerce una gran influencia sobre la otorgada en sentencia, una posible explicación es que, cuando el forense fija las secuelas de la víctima pero no las puntúa, es más probable que coincida su evaluación con la considerada por el médico perito (es decir, probablemente existirán menos discrepancias entre cuáles son las secuelas que padece la víctima que en la valoración de la gravedad de las mismas). En definitiva, en los casos en que el forense no ha puntuado las secuelas, el número de secuelas considerado por el perito médico de la compañía (variable *sec\_ul*) se comportaría como variable *proxy* del número de secuelas considerado por el médico forense y, por tanto, estaría reflejando la gravedad otorgada en sentencia. En cambio, cuando el forense valora las secuelas, es probable que la valoración del médico perito no represente con tanta precisión el nivel de gravedad considerado por el forense y, en consecuencia, tampoco el otorgado en sentencia.

### ***Contraste de hipótesis sobre los parámetros de la varianza***

Como mencionamos en la sección anterior (apartado 6.3.3), para muestras no balanceadas, el estadístico aproximado de Wald no es un estadístico de contraste óptimo. En particular, señalábamos que este estadístico podía ser especialmente inadecuado para parámetros de la varianza cuyo valor real estuviera cercano a cero. Para nuestro conjunto de datos, en la Tabla 6.5 encontramos algún indicio sobre la no fiabilidad del estadístico. Concretamente, en el caso del parámetro de la varianza de los efectos aleatorios  $\sigma_\alpha^2$ . En primer lugar, el parámetro estimado tiene un valor cercano a su límite inferior. Además, el límite superior del intervalo toma un valor exagerado (1,22E80). Este intervalo se ha calculado mediante la aproximación de Satterthwaite. En el siguiente capítulo analizaremos con profundidad esta

aproximación y mostraremos alternativas de cálculo de los intervalos de confianza, cuando, como en el presente estudio, la aproximación de Satterthwaite no ofrece resultados satisfactorios.

Por el contrario, el estadístico de Wald funciona adecuadamente para el resto de los componentes de la varianza. Esto se debe a que no se dan de forma tan pronunciada ninguna de las causas anteriormente mencionadas. En particular, si analizamos el número de casos para cada una de las categorías de la variable *forense* (Tabla 6.2), se observa una muestra más balanceada y con mayor número de observaciones para cada categoría de la variable. Además, los estimadores de los parámetros no se encuentran próximos a su límite inferior (Tabla 6.5). Por último, los límites de los intervalos de confianza estimados para estos parámetros no muestran valores desproporcionados (Tabla 6.5). Park y Burdick (2003) obtuvieron resultados similares para el caso de un modelo mixto no balanceado homocedástico con dos componentes de la varianza, mostrando que el contraste era inadecuado para el componente de la varianza del efecto aleatorio  $\sigma_\alpha^2$ , pero no para el del término de perturbación  $\sigma_\varepsilon^2$ .

### ***Estimación de los efectos aleatorios***

Como mencionamos en el apartado 6.4.1, si consideramos que existe una desviación individual en cada sentencia respecto a la indemnización esperada media, podremos especificar el modelo con la inclusión de los efectos aleatorios. Tomando de nuevo la fórmula (6.5) para la estimación de los efectos aleatorios, observamos que el modelo lineal incorpora un porcentaje del error de estimación asociado a cada sentencia individual,  $y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}$ . En concreto, el efecto aleatorio tiene la siguiente forma,

$$\hat{\mathbf{b}}_i = \hat{\sigma}_\alpha^2 \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1} (y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}),$$

donde  $\hat{\mathbf{V}}_i = \hat{\sigma}_\alpha^2 \mathbf{Z}_i \mathbf{Z}_i' + \bigoplus_{k=1}^4 \hat{\sigma}_{\varepsilon_k}^2 \mathbf{I}_{n_{ik}}$  con  $\sum_{k=1}^4 n_{ik} = n_i$ . Nótese que el porcentaje del error que se asignará a cada sujeto queda determinadazo por  $\hat{\sigma}_\alpha^2 \mathbf{Z}_i' \hat{\mathbf{V}}_i^{-1}$  y, por tanto, variará

únicamente en función de la/s categoría/s de la variable *forense* observada/s,  $\hat{\sigma}_{\varepsilon_i}^2$  para  $k=1, \dots, 4$ . A modo de ejemplo, permítanme mostrar el siguiente escenario.

*Escenario (I)*. Supongamos que estamos interesados en estimar la indemnización de un lesionado de 20 años de edad, quién tuvo un accidente con su moto, para quien la compañía dispone de los dos informe médicos periciales, en los cuales se consideró que había necesitado de 35 días de recuperación (impeditivos) y que el accidente no le había dejado secuelas. Además, la compañía sabe que el procedimiento judicial sigue el cauce civil (no interviene médico forense) y que en el moto iba un ocupante que también resultó lesionado y se espera que sea valorado en la misma sentencia. Con esta información, a la compañía se le plantean dos posibles situaciones a considerar en la estimación del valor esperado del logaritmo de la indemnización para esta víctima.

- A) Caso A. Suponer que en la valoración de la indemnización para el individuo analizado habrá correlación con la indemnización otorgada al pasajero por múltiples factores que no son observables directamente. Esta correlación quedará, por tanto, recogida en los términos de perturbación y es tenida en cuenta al realizar la estimación por Mínimos Cuadrados Generalizados. En la predicción de la indemnización de la víctima, por consiguiente, estimamos el valor esperado medio (únicamente los efectos fijos),

$$\hat{y} = 8,402 - 0,764 + 0,052 \times 3 + 0,009 \times 35 = 8,109.$$

- B) Caso B. Analizar jurisprudencia que permita estimar la desviación individual respecto al valor medio en sentencias previas (incorporar, por tanto, efectos aleatorios). Consideremos, por ejemplo, que disponemos de información sobre la indemnización concedida por el juez a dos víctimas, en sentencias anteriores. En este sentido, el error de estimación de la primera víctima fue de 0,6 y para la segunda víctima de 0,246. Respecto a la varianza del término de perturbación, en el primer caso, la categoría de la variable *forense* tomaba valor 1 (el forense ha puntuado las secuelas otorgadas) y, por tanto, la

varianza estimada es  $\hat{\sigma}_{\varepsilon,1}^2 = 0,761$ . Para el segundo caso, la variable *forense* tenía valor 2 (el forense ha definido las secuelas pero no las ha valorado) y varianza  $\hat{\sigma}_{\varepsilon,2} = 0,185$ . De este modo, la estimación de la desviación esperada individual respecto al valor medio en dicha sentencia ( $j$ ) sigue la siguiente estructura,

$$\hat{b}_j = (0,015 \quad 0,015) \begin{pmatrix} 0,761+0,015 & 0 \\ 0 & 0,185+0,015 \end{pmatrix}^{-1} (0,6 \quad 0,246)' = 0,031.$$

Cuando incorporamos el efecto aleatorio EBLUP en el modelo, el valor esperado del logaritmo de la indemnización para la nueva víctima, es,

$$\hat{y}_j = 8,402 - 0,764 + 0,052 \times 3 + 0,009 \times 35 + 0,031 = 8,140.$$

Nótese que en el caso de que no hubiéramos dispuesto de jurisprudencia previa que nos permitiese estimar la desviación individual respecto al valor medio, la predicción de la indemnización habría consistido en el valor esperado medio, es decir, considerando únicamente los efectos fijos (ídem que en el Caso A).

Cabe recordar, sin embargo, que en la estimación del modelo realizada (Tabla 6.5), el parámetro de varianza asociado al efecto aleatorio no era significativo y con un intervalo de confianza de amplitud desmesurada, provocado, como ya hemos comentado, por trabajar con una muestra extremadamente no balanceada. En el siguiente capítulo buscamos obtener un intervalo de variación para el parámetro de varianza asociado al efecto aleatorio, más que una predicción puntual del mismo, que nos ayude a solucionar el problema de significación provocado por la muestra de trabajo.

# Capítulo 7

## Predicción de la indemnización (II): Intervalo de confianza para componentes de la varianza

### 7.1 Introducción

En numerosas ocasiones, más que contrastar una hipótesis sobre un valor puntual de la varianza del efecto aleatorio, estaremos interesados en obtener un rango de valores (intervalo) en el que cabría esperar que, con un nivel de confianza determinado, se encuentre la verdadera  $\sigma_\alpha^2$ . En la literatura existen diferentes procedimientos para construir intervalos de confianza de los componentes de la varianza en los modelos lineales mixtos no-balanceados. Entre los métodos más frecuentes destacan aquellos basados en el estadístico de Wald, especialmente los que utilizan la aproximación de Satterthwaite. Mediante esta aproximación se tiene en cuenta que el límite inferior del intervalo del parámetro no puede tomar un valor negativo. El intervalo calculado presenta la siguiente forma,

$$\frac{2Z^2 \hat{\sigma}_\alpha^2}{\chi_{2Z^2, 1-\frac{w}{2}}^2} \leq \sigma_\alpha^2 \leq \frac{2Z^2 \hat{\sigma}_\alpha^2}{\chi_{2Z^2, \frac{w}{2}}^2} \quad (7.1)$$

donde  $\hat{\sigma}_\alpha^2$  es la estimación por Máxima Verosimilitud (Restringida) de  $\sigma_\alpha^2$ ,  $1-w$  es el nivel de confianza y  $Z$  el estadístico de Wald de dicha estimación,  $\hat{\sigma}_\alpha^2 / \widehat{\text{se}}(\hat{\sigma}_\alpha^2)$ . Esta técnica de cálculo de los intervalos de confianza está implementada en la mayoría de

paquetes estadísticos. Sin embargo, como ya señalábamos en el capítulo anterior, esta aproximación ofrece buenos resultados únicamente cuando disponemos de un número de observaciones  $n_i$  elevado para cada individuo, o cuando los datos son balanceados (Burdick y Graybill, 1992).

Otro método muy popular de cálculo de intervalos de confianza es el propuesto directamente por Wald (1940, 1947) para modelos con dos componentes de la varianza. Wald construye un intervalo de confianza para la ratio de los componentes  $\sigma_a^2/\sigma_\varepsilon^2$  mediante la función de distribución de una determinada cuantía central. Posteriormente, Seely y El-Bassiouni (1983) generalizan este método para poder aplicarlo a modelos con más de dos componentes de la varianza, señalando las condiciones (necesarias y suficientes) que debe satisfacer el modelo mixto correspondiente para que el método de Wald pueda ser aplicado. Posteriormente, Harville y Fenech (1985) tienen en cuenta las limitaciones computacionales de esta técnica y simplifican su cálculo sustituyendo en la cuantía central los diferentes valores propios por su valor medio. Basándose en las aproximaciones de Harville y Fenech (1985), El-Bassiouni (1994) considera límites modificados de las colas de las distribuciones para obtener intervalos de confianza de menor longitud. Otros autores han derivado los intervalos de confianza de los componentes de la varianza desde la teoría bayesiana (véase, por ejemplo, Michalski, 2003). Finalmente, Weerahandi (1993, 1995) utiliza la metodología de inferencia generalizada, introducida por Tsui y Weerahandi (1989) en el ámbito de los contrastes de hipótesis, para construir intervalos de confianza de los parámetros.

En el capítulo anterior remarcábamos que en muestras fuertemente no balanceadas, el estadístico de Wald no era una medida adecuada para realizar el contraste de hipótesis de los parámetros de la varianza cuyo valor estuviera cercano a cero. Del mismo modo, al principio de este apartado, hemos señalado que la aproximación de Satterthwaite, basada en el estadístico de Wald, no funciona apropiadamente si no disponemos de un gran número de casos para cada individuo. Por otro lado, en los resultados mostrados en el capítulo anterior (apartado 6.4.2),

vimos como el valor del estimador de  $\sigma_\alpha^2$  para nuestra muestra estaba cercano a su límite inferior, indicando que era un resultado lógico debido al bajo número de observaciones disponibles para cada uno de los individuos (máximo tres víctimas en una misma sentencia). En este punto, planteamos la hipótesis que el bajo valor obtenido del estimador  $\sigma_\alpha^2$  y la ineficacia del contraste y del intervalo de confianza estimados se deben en gran medida a la estructura de los datos utilizados. Por este motivo seleccionamos una metodología alternativa para calcular el rango de posibles valores, con un nivel de confianza determinado, en el que se hallará  $\sigma_\alpha^2$ . Nuestro objetivo en este capítulo es utilizar la metodología de inferencia generalizada para construir el intervalo de confianza de  $\sigma_\alpha^2$ . En la última parte del capítulo utilizaremos los resultados obtenidos para realizar una predicción por intervalos de las cuantías indemnizadoras, y propondremos una posible aplicación de los mismos por parte de una compañía aseguradora.

## 7.2 Intervalo de confianza generalizado

Tomemos la definición clásica de intervalo de confianza,

$$\Pr[I(\mathbf{X}) \leq \delta \leq S(\mathbf{X})] = 1 - \omega, \quad (7.2)$$

donde  $\mathbf{X}=(X_1, \dots, X_n)$  es una muestra aleatoria de una determinada distribución con parámetro desconocido  $\delta$ . Por su parte,  $I(\mathbf{X})$  y  $S(\mathbf{X})$  son estadísticos que satisfacen la ecuación (7.2) al nivel de confianza  $1 - \omega$ ,  $0 \leq \omega \leq 1$ . Normalmente, tendremos una familia de intervalos de confianza posibles. Por tanto, deberemos buscar un intervalo de confianza que cumpla un conjunto de propiedades deseables. Existen diferentes técnicas estadísticas para la obtención de dicho intervalo. Un procedimiento extensamente aplicado es utilizar una determinada función real  $R=(\mathbf{X}, \delta)$  con ciertas propiedades, denominada cuantía central (*pivotal quantity*).

Supongamos ahora que la función de distribución de  $\mathbf{X}$  depende de un vector de parámetros desconocidos  $\mathbf{v}=(\delta, \zeta)$ , dónde  $\delta$  es el parámetro de interés, y  $\zeta$  es un vector de parámetros molestos. Lamentablemente, debido a la presencia de

parámetros molestos, en la mayoría de ocasiones no existirán estadísticos que cumplan la ecuación (7.2) para todos los posibles valores de  $\zeta$ . Para solucionar esta limitación de los intervalos de confianza, Weerahandi (1993, 1995) extendió su definición, permitiendo a  $I(\cdot)$  y  $S(\cdot)$  depender de la variable aleatoria observable  $\mathbf{X}$ , pero también de los datos realmente observados  $\mathbf{x}=(x_1, \dots, x_n)$ . Como sucede en el caso convencional, el intervalo de confianza de  $\delta$  puede obtenerse de nuevo basándose en la noción de *cuantía central*.

**Definición 7.1 (Weerahandi, 2004).** *La función  $R = r(\mathbf{X}, \mathbf{x}, \mathbf{v})$  es una cuantía central generalizada (CCG), si y sólo si:*

- *su distribución de probabilidad no tiene parámetros desconocidos (Propiedad A).*
- *$R_{obs} = r(\mathbf{x}, \mathbf{x}, \mathbf{v})$  no depende del vector de parámetros molestos  $\zeta$  (Propiedad B).*

La primera propiedad nos asegura que podemos encontrar un subconjunto  $\rho_\omega$  del espacio muestral de  $R$  para cualquier nivel de confianza  $(1-\omega)$  considerado. Por su parte, la propiedad B garantiza que el intervalo de confianza estimado de  $\delta$  se basará exclusivamente en los datos observados  $\mathbf{x}$  (véase Apéndice D). Cabe remarcar que, igual que sucede con los intervalos de confianza convencionales, la utilización de la cuantía central no es el único mecanismo para construir un intervalo de confianza de  $\delta$ . Por ejemplo, mediante la función potencia generalizada podríamos llegar a resultados equivalentes a los obtenidos con  $R = r(\mathbf{X}, \mathbf{x}, \mathbf{v})$  (véase Apéndice E).

### 7.2.1 Propiedades frecuentistas de los intervalos de confianza generalizados

Consideremos la definición clásica de muestreo repetido, la cual establece que si repetimos un experimento en las mismas condiciones en un gran número de ocasiones, los intervalos de confianza estimados, con nivel de confianza  $(1-\omega)$ , incluirán el verdadero parámetro el  $(1-\omega) \times 100\%$  de las veces. Esta propiedad, aunque muy importante, es fundamentalmente teórica, puesto que en la realidad casi nunca dispondremos de un elevado número de muestras repetidas de un mismo experimento. Del mismo modo, de la definición de intervalo de confianza



generalizado se derivan propiedades frecuentistas atractivas, pero de difícil aplicación. En particular, los intervalos de confianza generalizados satisfacen la propiedad que establece que, tomando remuestreos de nuestra variable aleatoria y calculando el intervalo de confianza del parámetro de interés en función de cada nueva muestra simulada y de los datos realmente observados, si el número de repeticiones del experimento es suficientemente grande, los intervalos estimados contendrán el parámetro de interés en el  $(1-\omega)\times 100\%$  de las veces. Ambas propiedades son muy similares. En concreto, la propiedad de muestreo repetido implica la segunda, aunque en sentido inverso sólo se cumple de forma condicional, esto es, fijado un conjunto de datos determinado  $x$ . Es decir, los intervalos de confianza generalizados satisfacen la propiedad de muestreo repetido condicionalmente, dado  $x$ .

No obstante, en la literatura existen numerosos trabajos, basados en simulaciones, que han estudiado el comportamiento no condicional de los intervalos de confianza generalizados, demostrando que el grado de cobertura real de estos intervalos es al menos tan grande como el nivel de confianza elegido (Griffiths y Judge, 1992; Zhou y Mathew, 1994). Por tanto, los intervalos de confianza generalizados cumplirían aproximadamente el principio de muestreo repetido también de forma no condicional, teniendo incluso un mejor comportamiento que otras aproximaciones disponibles. Weerahandi (1995) demostró formalmente que el intervalo de confianza generalizado con nivel de confianza de  $(1-\omega)\times 100\%$  es un intervalo de confianza (clásico) que satisface la propiedad de muestreo repetido con un nivel de confianza de al menos  $(1-2\omega)\times 100\%$ .

### 7.2.2 La elección de la cuantía central generalizada

Desafortunadamente, encontrar una adecuada  $R$  puede resultar una tarea tediosa. El primer paso para la simplificación del problema estadístico debería consistir en encontrar el conjunto de estadísticos mínimo suficientes en los que se basará  $R$ . En aquellos casos que el conjunto esté formado por más de un estadístico, podemos obtener posteriores reducciones de la dimensionalidad del problema imponiendo los

criterios de *invariabilidad* y *similitud*. Ahora bien, aún aplicando las técnicas citadas anteriormente, en numerosas situaciones nos encontraremos que no hay una única  $R$ , y serán necesarios argumentos complementarios que justifiquen la elección de la cuantía central generalizada (CCG) seleccionada.

La definición de *invariabilidad* y *similitud* utilizada por Weerahandi (1993, pp.901-902) para el ámbito de la estimación por intervalos es la siguiente.

**Definición 7.2.** Consideremos la variable aleatoria  $X$  con función de distribución  $F(X, \delta)$ , siendo  $\Omega$  el espacio parametral de  $\delta$  y  $G$  un grupo de transformaciones del espacio muestral  $\chi$ . Si para cualquier  $g \in G$  y  $\delta \in \Omega$  dados, existe un único  $\delta^*$  tal que la función de distribución de  $g(X)$  es  $F(g(X), \delta^*)$ , entonces la familia de distribuciones  $F(X, \delta)$  es invariable bajo el grupo de transformaciones  $G$ . Supongamos que el parámetro  $\delta$  no se ve afectado por el grupo de transformaciones  $G$ ,  $\delta^* = \delta$ . La región de confianza generalizada de  $\delta$  es invariable si para cualquier observación de  $X$ ,  $\mathbf{x} \in \chi$ , se cumple que  $\Omega(\mathbf{x}) = \Omega(g(\mathbf{x}))$  para todo  $g \in G$ .

Tsui y Weerahandi (1989) demuestran que si la distribución de  $R$  y  $R_{obs}$  depende de  $\mathbf{x}$  únicamente a través del conjunto de máximo invariantes (estadísticos con mínima dimensión; véase Lehmann (1986) para una definición formal), entonces se puede derivar cualquier región de confianza generalizada invariable de  $\delta$  a partir de  $R$ .

**Definición 7.3.** Un subconjunto  $\chi(\mathbf{x}, \mathbf{v})$  del espacio muestral  $\chi$  y el intervalo estimado  $\Omega_{\chi}(\mathbf{x}, \mathbf{v}) = \{\delta \in \Omega \mid \mathbf{x} \in \chi(\mathbf{x}, \mathbf{v})\}$  son similares en  $\zeta$ , si la  $\Pr[\mathbf{X} \in \chi(\mathbf{x}, \mathbf{v})]$  no depende del vector de parámetros molestos  $\zeta$ .

Es decir, una cuantía que dependa de los datos observados  $\mathbf{x}$  sólo puede ser utilizada como cuantía central para obtener intervalos sobre el parámetro de interés, si la podemos calcular sin tener información del(os) parámetro(s) molesto(s). Nótese que esta definición está muy relacionada con la primera propiedad de la cuantía central generalizada. Por tanto, aunque el concepto de *similitud* haya sido definido

como mecanismo de reducción en la dimensionalidad, en realidad es necesario para obtener un intervalo de confianza factible dado un conjunto de datos  $\mathbf{x}$ .

### 7.3 Aplicación en un modelo con dos componentes de varianza

Consideremos de nuevo el modelo presentado en (6.3) para el caso de dos componentes de la varianza,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \quad (7.3)$$

donde  $\mathbf{y}$  es el vector de  $n \times 1$  observaciones,  $\mathbf{X}(n \times p)$  y  $\mathbf{Z}(n \times q)$  son matrices conocidas y  $\boldsymbol{\beta}(p \times 1)$  es el vector de  $p$  parámetros desconocidos. Por su parte,  $\mathbf{b}$  y  $\boldsymbol{\varepsilon}$  son vectores aleatorios no-observables normalmente distribuidos y mutuamente independientes,  $\mathbf{b} \sim N_q(0, \sigma_b^2 \mathbf{I}_q)$  y  $\boldsymbol{\varepsilon} \sim N_n(0, \sigma_\varepsilon^2 \mathbf{I}_n)$  tal que  $\text{cov}(\boldsymbol{\varepsilon}, \mathbf{b}) = 0$ . Adicionalmente, asumimos que se cumplen las siguientes desigualdades  $n > \text{rango}(\mathbf{X}, \mathbf{Z}) > \text{rango}(\mathbf{X}) \equiv p > 0$  (Harville y Fenech, 1985; El-Bassiouni, 1994; Park y Burdick, 2003).

Nuestro propósito consiste en construir un intervalo de confianza generalizado para  $\sigma_\alpha^2$ . Dado que la construcción de un intervalo para los componentes de la varianza es invariable bajo el grupo de transformaciones  $G = \{y \rightarrow y + \mathbf{X}\boldsymbol{\kappa}, \boldsymbol{\kappa} \in R^p\}$ , siendo  $\boldsymbol{\kappa}$  un vector  $(p \times 1)$  arbitrario, podemos reducir el problema mediante el estadístico máximo invariable en  $G$ ,  $\mathbf{u} = \mathbf{W}'\mathbf{y}$  donde  $\mathbf{W}$  es la matriz  $n \times (n-p)$  tal que  $\mathbf{W}'\mathbf{W} = \mathbf{I}_{n-p}$  y  $\mathbf{W}'\mathbf{X} = 0$ . La distribución del estadístico propuesto tendrá la siguiente estructura,

$$\mathbf{u} \sim N_{n-p}(0, \sigma_\alpha^2 \mathbf{V} + \sigma_\varepsilon^2 \mathbf{I}) \quad (7.4)$$

donde  $\mathbf{V} = \mathbf{W}'\mathbf{Z}\mathbf{Z}'\mathbf{W}$ . Tomemos la descomposición espectral de  $\mathbf{V} = \sum_{l=0}^m \lambda_l \mathbf{E}_l \mathbf{E}_l'$ , donde  $\lambda_l$  representa los  $m$  distintos valores propios de  $\mathbf{V}$ ,  $\lambda_m > \dots > \lambda_1 > \lambda_0 = 0$ <sup>71</sup>, y  $\mathbf{E}_l$  es

---

<sup>71</sup>  $\lambda_0 = 0$  si la matriz  $\mathbf{V}$  es singular, lo cual se cumple en la gran mayoría de los casos (Arendacká, 2005).

la matriz  $(n-p) \times r_l$  cuyas columnas son los vectores propios ortonormales de  $\lambda_l$ , con multiplicidad  $r_l$ , siendo  $r_l = \text{rango}(\mathbf{E}_l \mathbf{E}_l') > 0$ . Olsen *et al.* (1976) demostraron que  $Q_l = \mathbf{u}' \mathbf{E}_l \mathbf{E}_l' \mathbf{u}$ , con  $l=0,1,\dots,m$ , determinan los estadísticos mínimo suficientes para la familia de distribuciones (7.4) y se distribuyen independientemente con la siguiente estructura,

$$Q_l \sim (\sigma_\varepsilon^2 + \lambda_l \sigma_\alpha^2) \chi_{r_l}^2 \quad \text{con } l=0,1,\dots,m. \quad (7.5)$$

Para el caso particular de  $\lambda_0$ , se cumple que su multiplicidad es igual a  $r_0 = n - \text{rango}(\mathbf{X}, \mathbf{Z})$ . Si tenemos en cuenta la siguiente igualdad  $Q_0 = \mathbf{y}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y} - \hat{\mathbf{b}}'\mathbf{U}$ , donde  $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  es el proyector ortogonal del espacio columna de  $\mathbf{X}$ ,  $\mathbf{U} = \mathbf{Z}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}$  y  $\hat{\mathbf{b}}$  es cualquier solución lineal de  $\mathbf{Z}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{Z}\hat{\mathbf{b}} = \mathbf{U}$ , podemos calcular el error cuadrado medio del modelo (7.3) mediante la fórmula  $S_\varepsilon^2 = Q_0 / r_0$ . El lector interesado, encontrará la demostración en El-Bassiouni (1994).

Partiendo de la independendencia entre los distintos  $Q_l$  y de la siguiente igualdad  $\sum_{l=1}^m r_l = \text{rango}(\mathbf{V}) = \text{rango}(\mathbf{X}, \mathbf{Z}) - \text{rango}(\mathbf{X}) = s$  (El-Bassiouni, 1994), se deriva,

$$\begin{aligned} W &= \frac{Q_0}{\sigma_\varepsilon^2} = \frac{r_0 S_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{r_0}^2 \\ T &= \sum_{l=1}^m \frac{Q_l}{\sigma_\varepsilon^2 + \lambda_l \sigma_\alpha^2} \sim \chi_s^2, \end{aligned} \quad (7.6)$$

donde  $W$  y  $T$  pueden utilizarse en la construcción de la cuantía central generalizada  $R$ . Park y Burdick (2003, 2004) definieron  $R$  como la solución de  $\sigma_\alpha^2$  en la siguiente ecuación no lineal,

$$\sum_{l=1}^m \frac{q_l}{\frac{r_0 S_\varepsilon^2}{W} + \lambda_l R} = T, \quad (7.7)$$

siendo  $q_l$  y  $s_\varepsilon^2$  los valores observados de  $Q_l$  y  $S_\varepsilon^2$ , respectivamente. Nótese que la distribución de  $R$  queda definida por medio de las distribuciones conjuntas de  $W$  y  $T$  y, por tanto, no depende de ningún parámetro desconocido, cumpliéndose así la primera propiedad de la cuantía central generalizada. Por otro lado, su valor observado  $R_{obs}$  es el parámetro de interés  $\sigma_\alpha^2$  y, por tanto, no depende del parámetro molesto  $\sigma_\varepsilon^2$  (Propiedad B),

$$\sum_{l=1}^m \frac{q_l}{\frac{r_0 s_\varepsilon^2}{\sigma_\varepsilon^2} + \lambda_l R_{obs}} = \sum_{l=1}^m \frac{q_l}{\sigma_\varepsilon^2 + \lambda_l \sigma_\alpha^2} = t_{obs} \Rightarrow R_{obs} = \sigma_\alpha^2.$$

Desafortunadamente, la ecuación (7.7) no tiene una sencilla solución analítica, por tanto, deberemos utilizar alguno de los procesos iterativos disponibles (p.ej. Newton, Bisección, etc.) para su cálculo en las  $J$  simulaciones, siendo  $t_j$  y  $w_j$  los  $j$ -ésimos pares de valores simulados,  $j=1, \dots, J$ . De este modo, obtendremos el intervalo de confianza generalizado de  $\sigma_\alpha^2$  con nivel de confianza del  $(1-\omega) \times 100\%$ , a partir de los percentiles de la distribución simulada  $[R_{\omega/2}^*, R_{1-\omega/2}^*]$ .

### 7.3.1 La unicidad del intervalo de confianza generalizado

Olsen *et al.* (1976) demostraron en su artículo que el estadístico suficiente (7.5) era completo únicamente para aquellos casos que  $V$  tuviera como máximo dos valores propios diferentes,  $m \leq 1$ . Posteriormente, Zhou y Mathew (1994) mostraron formalmente, en el área del contraste de hipótesis, que cuando el estadístico suficiente no es completo, la variable de contraste generalizada no es única y, consecuentemente, tampoco lo son los intervalos de confianza generalizados que se deriven (véase Apéndice E para definición de variable de contraste generalizada).

Autores como Arendacká (2005) se han centrado en el problema de elección de la variable de contraste más adecuada para el modelo (7.3) cuando ésta no es única. Mediante simulaciones, la autora compara entre diferentes variables de contraste en

función de la amplitud y la probabilidad de cobertura de los intervalos de confianza que producen. Entre sus conclusiones señala que la variable de contraste equivalente a la cuantía central definida en (7.7) es de las que demuestra un mejor comportamiento para los distintos escenarios planteados. Por tanto, la utilización de la cuantía central generalizada de Park y Burdick quedaría justificada para el cálculo del intervalo de confianza del parámetro de interés  $\sigma_\alpha^2$  del modelo (7.3), aún para aquellos casos que exista más de una variable de contraste generalizada. Cabe mencionar que el problema de elección entre variables de contraste es una línea de investigación abierta en la cual la literatura existente aún no es muy extensa.

### 7.3.2 El caso heterocedástico

Como señalábamos en el capítulo 6, en el modelo propuesto, la varianza del término de perturbación no es constante, sino que depende de la actuación del médico forense. El modelo lineal mixto estimado mostraba la siguiente estructura,

$$y \sim N_n(\mathbf{X}\beta, \bigoplus_{k=1}^4 \sigma_{\varepsilon_k}^2 \mathbf{I}_{n_k} + \sigma_\alpha^2 \mathbf{Z}\mathbf{Z}') \quad (7.8)$$

donde  $k$  es la  $k$ -ésima categoría de la variable *forense*,  $n_k$  el número total de observaciones para cada una de las  $k$  categorías,  $\sum_{k=1}^4 n_k = n$ , y el símbolo  $\bigoplus$  representa la suma directa de matrices. Debido a la presencia de heterocedasticidad por grupos, debemos transformar el conjunto de datos observados para realizar la estimación del intervalo de confianza generalizado de  $\sigma_\alpha^2$ .

En primer lugar, reescribimos la matriz  $\bigoplus_{k=1}^4 \sigma_{\varepsilon_k}^2 \mathbf{I}_{n_k}$  como  $\sigma_\varepsilon^2 \mathbf{\Omega}$ , donde,

$$\mathbf{\Omega} = \begin{pmatrix} \gamma_1 \mathbf{I}_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_4 \mathbf{I}_{n_4} \end{pmatrix}_{n \times n},$$

siendo  $\gamma_k$  una escala arbitraria que satisface  $\gamma_k \geq 0$  y al menos un  $\gamma_k > 0$  para  $k=1, \dots, 4$ . Nótese que  $\mathbf{\Omega}(n \times n)$  es una matriz diagonal definida positiva y, por lo tanto, puede ser

descompuesta de la forma  $\Omega = \mathbf{C}\Lambda\mathbf{C}'$ , siendo su inversa  $\Omega^{-1} = \mathbf{P}'\mathbf{P}$ , con  $\mathbf{P}' = \mathbf{C}\Lambda^{-1/2}$ . Si multiplicamos todos los elementos de la regresión lineal por  $\mathbf{P}$ ,

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\mathbf{Z}\mathbf{b} + \mathbf{P}\boldsymbol{\varepsilon} \Rightarrow \mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{Z}^*\mathbf{b} + \boldsymbol{\varepsilon}^*, \quad (7.9)$$

obtenemos que la regresión transformada es equivalente a la definida en (6.3),

$$\mathbf{y}^* \sim N_n(\mathbf{X}^*\boldsymbol{\beta}, \sigma_\varepsilon^2\mathbf{I}_n + \sigma_\alpha^2\mathbf{Z}^*\mathbf{Z}^{*\prime}), \quad (7.10)$$

con perturbaciones  $\boldsymbol{\varepsilon}^*$  homocedásticas. Desafortunadamente, la matriz  $\Omega$  es desconocida por lo que deberemos sustituirla por su estimador  $\hat{\Omega}$ .

## 7.4 Resultados

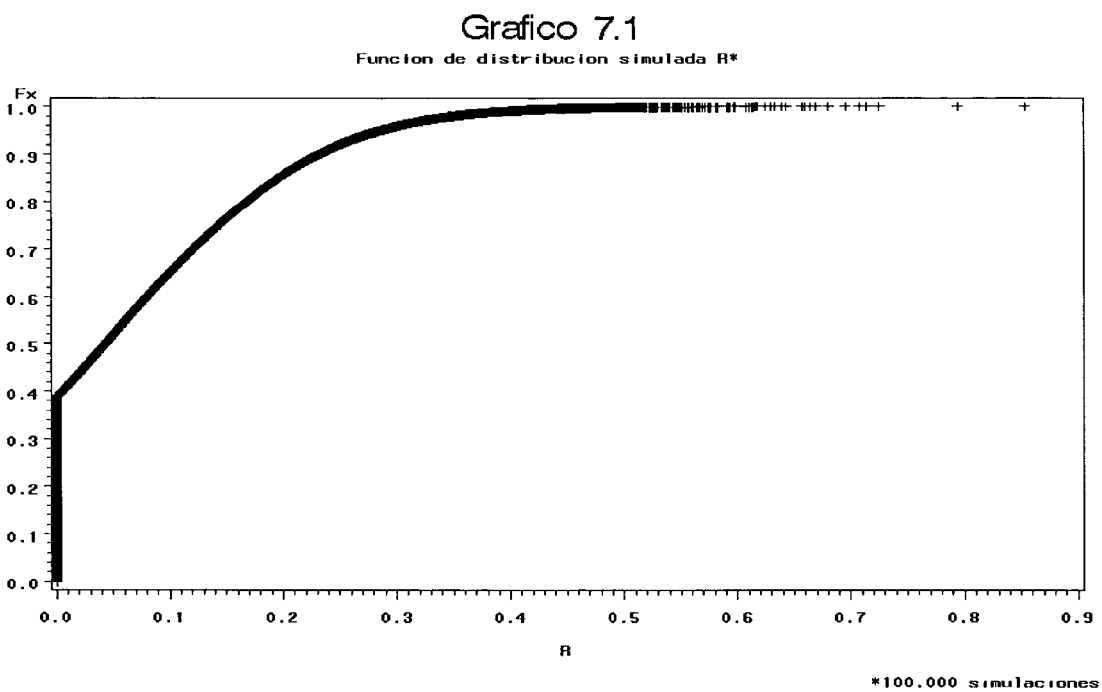
En primer lugar, necesitamos estimar la matriz  $\Omega$ . Para ello, podemos utilizar los estimadores máximo-verosímiles restringidos de los parámetros de la varianza del término de perturbación presentados en el capítulo anterior. Si tomamos los estimadores MVR de  $\sigma_{\varepsilon,k}^2$  de la Tabla 6.5 y los  $n_k$  de la Tabla 6.2, podemos definir el estimador MVR de  $\hat{\Omega}$  como,

$$\bigoplus_{k=1}^4 \hat{\sigma}_{\varepsilon_k}^2 \mathbf{I}_{n_k} = \begin{pmatrix} 0,761\mathbf{I}_{24} & \cdots & 0 & 0 \\ \vdots & 0,185\mathbf{I}_{39} & & \vdots \\ 0 & & 0,481\mathbf{I}_{39} & 0 \\ 0 & \cdots & 0 & 0,661\mathbf{I}_{12} \end{pmatrix} = 0,185 \begin{pmatrix} 4,113\mathbf{I}_{24} & \cdots & 0 & 0 \\ \vdots & \mathbf{I}_{39} & & \vdots \\ 0 & & 2,6\mathbf{I}_{39} & 0 \\ 0 & \cdots & 0 & 3,573\mathbf{I}_{12} \end{pmatrix} = \hat{\sigma}_\varepsilon^2 \hat{\Omega}$$

Todos los cálculos han sido realizados en entorno SAS (v.9.1). Para la solución de la ecuación no-lineal (7.7) se ha utilizado el método de bisección, basándonos en el algoritmo ofrecido por Park y Burdick (2003). La función de distribución simulada se ha obtenido en base a 100.000 simulaciones.

Si analizamos la función de distribución simulada (Gráfico 7.1), observamos que está truncada en cero. Este truncamiento se debe a que  $\sigma_\alpha^2$  es la varianza del efecto aleatorio y, por tanto, no puede tomar valores negativos. Por este motivo, en

cada iteración, cuando obtenemos un valor negativo de la cuantía central generalizada simulada  $R^* < 0$ , se iguala a cero. Entre otros resultados, destacar, por ejemplo, que con una probabilidad mayor del 60% el coeficiente de  $\sigma_\alpha^2$  será positivo ( $\Pr[R^* > 0] > 0,6$ ), o que tomará un valor inferior a 0,3 con una probabilidad superior al 90% ( $\Pr[R^* < 0,3] > 0,9$ ).



En la Tabla 7.1 se muestra el intervalo de confianza generalizado estimado para el parámetro de la varianza asociado al efecto aleatorio, para niveles de confianza del 90% y del 95%. En este sentido, para un nivel de confianza del 95%, el límite superior del intervalo de confianza generalizado del componente de la varianza tomará valor 0,335 y, para el 90%, un valor de 0,286.

**TABLA 7.1 Intervalo de Confianza Generalizado para  $\sigma_\alpha^2$**

<i>Nivel de Confianza (%)</i>	<i>Límite Inferior</i>	<i>Límite superior</i>
0,90 ( $\omega=0,10$ )	0	0,286
0,95 ( $\omega=0,05$ )	0	0,335



Finalmente, partiendo de la ecuación (6.5), que recoge la estimación del efecto aleatorio (EBLUP), podemos definir la máxima desviación individual respecto al valor esperado medio imputable al individuo  $i$ , de la siguiente forma,

$$\hat{b}_{i(\max)} = \hat{\sigma}_\alpha^{*2} \mathbf{Z}_i' \hat{\mathbf{V}}_i^{*-1} (y_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) = \hat{\sigma}_\alpha^{*2} \mathbf{Z}_i' \hat{\mathbf{V}}_i^{*-1} \mathbf{e}_i,$$

donde  $\hat{\sigma}_\alpha^{*2}$  es el límite superior del intervalo de confianza, tal que  $\Pr[\hat{\sigma}_\alpha^2 \leq \hat{\sigma}_\alpha^{*2} | \mathbf{x}] = 1 - \alpha/2$ ,

y  $\hat{\mathbf{V}}_i^* = \hat{\sigma}_\alpha^{*2} \mathbf{Z}_i \mathbf{Z}_i' + \bigoplus_{\varepsilon_k} \hat{\sigma}_{\varepsilon_k}^2 \mathbf{I}_{n_{i,k}}$  con  $\sum_{k=1}^4 n_{i,k} = n_i$ .

Nótese que para la estimación de  $\hat{b}_{i(\max)}$  no hemos tenido en cuenta el efecto de  $\hat{\sigma}_\alpha^{*2}$  en la estimación del vector de parámetros  $\hat{\boldsymbol{\beta}}$ . Ahora bien, el interés del estimador  $\hat{b}_{i(\max)}$  radica en, fijado un determinado error de estimación  $\mathbf{e}_i$ , determinar cual es el porcentaje máximo de este error que se le puede atribuir al efecto aleatorio. Dicho porcentaje se obtiene como una ratio entre la covarianza máxima del efecto aleatorio y la variable dependiente,  $\hat{\sigma}_\alpha^{*2} \mathbf{Z}_i'$ , respecto a la varianza total transformada  $\hat{\mathbf{V}}_i^*$ . Esta estimación es interesante para la compañía aseguradora, puesto que le indica cual sería la variación máxima respecto a la cuantía media estimada que podría asociarse a cada individuo. Para mostrar su incidencia, consideremos de nuevo el escenario presentado en el capítulo 6.

*Escenario (II)*. Nuestro objetivo es estimar la indemnización económica de un lesionado, en un accidente de moto, de 20 años de edad. El médico perito ha señalado la existencia de 35 días de baja y la no existencia de secuelas, en ambos informes periciales. Se trata de una demanda civil y se dispone de sentencias previas dictadas por el mismo sujeto ( $j$ ) que ha de juzgar el caso. Con esta información, la predicción de la cuantía indemnizatoria puede obtenerse de diferentes formas:

- A) Estimación puntual del logaritmo de la indemnización, considerando los efectos fijos más el efecto aleatorio (capítulo 6),

$$\hat{y}_j = 8,402 - 0,764 + 0,052 \times 3 + 0,009 \times 35 + 0,031 = 8,140.$$

B) Estimación puntual del logaritmo de la indemnización utilizando los estimadores de los efectos fijos y el efecto aleatorio derivado del estimador máximo del parámetro de la varianza, a un nivel de confianza del 95% (Tabla 7.1). El primer paso consistirá en calcular la desviación esperada (con el estimador máximo del parámetro de la varianza) en base a la jurisprudencia existente,

$$\hat{b}_{j(\max)} = (0,335 \quad 0,335) \begin{pmatrix} 0,761 + 0,335 & 0 \\ 0 & 0,185 + 0,335 \end{pmatrix}^{-1} (0,6 \quad 0,246)' = 0,341.$$

Posteriormente, teniendo en cuenta el efecto aleatorio calculado  $\hat{b}_{j(\max)}$ , la estimación del logaritmo de indemnización será la siguiente,

$$\hat{y}_j^* = 8,402 - 0,764 + 0,052 \times 3 + 0,009 \times 35 + 0,341 = 8,450.$$

En este ejemplo, el estimador del efecto aleatorio EBLUP, así como el obtenido con el límite superior del intervalo de confianza generalizado para  $\sigma_\alpha^2$ , tienen signo positivo, incrementando, por tanto, el valor esperado de la indemnización. Ahora bien, como hemos comentado anteriormente, estos estimadores son desviaciones individuales esperadas respecto al valor medio de la indemnización, por lo que perfectamente podrían haber tomado valores negativos.

## 7.5 Límite máximo de negociación en la cuantía indemnizatoria

En la práctica, en la mayoría de ocasiones se logra un acuerdo privado entre las partes en la cuantía económica y, por tanto, no es necesario llegar a juicio para la fijación de la indemnización. Como mencionamos en el capítulo 6, la compañía aseguradora está muy interesada en predecir el valor de la indemnización económica en sentencia, puesto que, esta predicción puntual, le señala el coste esperado del siniestro si el juicio se celebra, y, por tanto, puede considerarla como el nivel máximo de negociación con la parte contraria. Es decir, como un instrumento que le

indique a la entidad el límite máximo para llegar a un acuerdo con la parte contraria, por encima del mismo, por ende, es preferible continuar con el proceso judicial. Por otro lado, también será de gran interés para la compañía, con un nivel de confianza determinado, predecir la cuantía máxima esperada de indemnización otorgada en sentencia judicial, puesto que, este extremo superior del intervalo de predicción, le indicará a la compañía el margen de flexibilidad del que dispone respecto al límite máximo de negociación predicho.

En este apartado, nos centraremos en el cálculo de la predicción puntual y del intervalo de predicción en los modelos lineales mixtos. En particular, mostraremos como calcular el extremo superior del intervalo de predicción y su aplicación en la negociación. Finalmente, analizaremos su efecto en el cálculo de las provisiones por siniestros pendientes.

Supongamos ahora que deseamos predecir la indemnización económica de un nuevo lesionado. Este problema se puede considerar un caso específico de la predicción de una variable aleatoria  $\tau$  basándonos en el valor del vector aleatorio observable  $y$ ,

$$\tau = \lambda' \beta + \delta' b, \quad (7.11)$$

con  $E[\tau] = \lambda' \beta$ ,  $\text{Var}(\tau) = \sigma_a^2 \delta' \delta$  y  $\text{Cov}(y, \tau) = \sigma_a^2 Z \delta$ , siendo  $\lambda$  y  $\delta$  dos vectores conocidos. Tomemos  $d(y)$  como un predictor puntual de  $\tau$ . Nuestro objetivo ha de ser buscar un  $d(y)$  que sea insesgado o aproximadamente insesgado,  $E[d(y) - \tau] \approx 0$ , y con el mínimo error cuadrado medio (ECM) posible,  $m = E[(d(y) - \tau)^2]$ . Cabe recordar que si  $d(y)$  es insesgado,  $m$  coincide con la varianza del error de predicción. Como señalamos en el capítulo 6, Henderson (1950, 1975) demostró que si los componentes de la varianza son conocidos, el predictor insesgado de  $\tau$  que tiene mínima varianza es,

$$d(y) = \lambda' \hat{\beta} + \sigma_a^2 Z \delta V^{-1} (y - X \hat{\beta}) \quad (7.12)$$

donde  $\hat{\beta}$  es el vector de estimadores MCG de los parámetros y  $V$  es la matriz de varianzas y covarianzas de  $y$ . El intervalo de predicción, con nivel de confianza  $(1-\omega)\times 100\%$ , tiene entonces la siguiente estructura,

$$d(y) + t_{\omega(n-p)} \sqrt{m} \quad (7.13)$$

siendo  $t_{\omega(n-p)}$  el valor superior  $\omega$  de la distribución  $t$ -student, con  $(n-p)$  grados de libertad. Lamentablemente, los parámetros  $\sigma_{\epsilon,k}^2$  y  $\sigma_{\alpha}^2$  para  $k=1,\dots,K$ , en la mayoría de ocasiones serán desconocidos y, por tanto, deberemos sustituir  $m = f(\sigma_{\epsilon,k}^2, \sigma_{\alpha}^2)$  por su estimador  $\hat{m} = f(\hat{\sigma}_{\epsilon,k}^2, \hat{\sigma}_{\alpha}^2)$ . La relación entre ambas expresiones es la siguiente,

$$m = \hat{m} + \text{Var}[e^*]$$

donde  $e^*$  refleja la diferencia entre el predictor EBLUP y el predictor BLUP. Autores como Prasad y Rao (1990), Harville (1989) o Harville y Jeske (1992) ofrecen diferentes expresiones, exactas y aproximadas, para  $\text{Var}[e^*]$  en los modelos lineales mixtos. En la práctica, debido a que las expresiones propuestas, o no son factibles, o son computacionalmente muy intensas, normalmente se calcula el intervalo de predicción aproximado por medio de la varianza del error de predicción<sup>72</sup>,

$$d(y) + t_{\omega(n-p)} \sqrt{\text{var}[d(y)-\tau]},$$

donde  $\text{var}[d(y)-\tau]$  tiene la siguiente estructura (Harville y Jeske, 1992),

$$\text{var}[d(y)-\tau] = \hat{\sigma}_{\alpha}^2 \delta' \delta - \hat{\sigma}_{\alpha}^4 \delta' Z' \hat{V}^{-1} Z \delta + [\lambda' - \hat{\sigma}_{\alpha}^2 \delta' Z' \hat{V}^{-1} X] (X' \hat{V}^{-1} X)^{-1} [\lambda - \hat{\sigma}_{\alpha}^2 X' \hat{V}^{-1} Z \delta]. \quad (7.14)$$

Por otro lado, de forma equivalente a cuando estimábamos el modelo en el capítulo 6, si la compañía aseguradora no dispone de jurisprudencia previa o,

---

<sup>72</sup> Véase Tuchscherer *et al.* (1998) para una comparación, mediante simulaciones, del error cuadrático medio  $m$  y la varianza del error de predicción.

simplemente, asume que la correlación en las indemnizaciones de la misma sentencia se debe al efecto de múltiples factores no observables, deberá realizar la predicción considerando únicamente los efectos fijos (indemnización esperada media). Nótese que sustituyendo  $\delta=0$  en la ecuación (7.14), obtenemos que la varianza del error de predicción es una forma cuadrática de la matriz de covarianzas aproximada de  $\hat{\beta}$ ,

$$\text{var}[d^*(\mathbf{y}) - \tau^*] = \lambda'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\lambda,$$

siendo  $\tau^* = \lambda'\beta$  y su estimador  $d^*(\mathbf{y}) = \lambda'\hat{\beta}$ .

Por último, señalar que, en la estimación del modelo lineal mixto, la variable dependiente había sido modificada a escala logarítmica, por lo que será preciso transformar las predicciones a su escala original. En este sentido, a modo de recordatorio, presentamos las funciones utilizadas para la transformación de los logaritmos de las indemnizaciones estimadas a su escala original (Greene, 1998),

$$\hat{\eta}_i = e^{d(y_i) + 0,5 \times \text{var}[d(y_i) - \tau_i]},$$

$$\text{var}(\hat{\eta}_i - \eta_i) = e^{2 \times d(y_i) + \text{var}[d(y_i) - \tau_i]} \left( e^{\text{var}[d(y_i) - \tau_i]} - 1 \right),$$

siendo  $\hat{\eta}_i$  el vector de predicciones para el individuo  $i$  transformado a escala original. Lógicamente, estas funciones se mantienen si sustituimos  $\delta(y_i)$  y  $\tau_i$  por  $\delta^*(y_i)$  y  $\tau_i^*$ . A continuación se utiliza, de nuevo, el escenario presentado en páginas anteriores para calcular el límite máximo de negociación.

*Escenario (III).* Consideremos nuevamente los datos del escenario presentado en los capítulos 6 y 7 para calcular la predicción puntual y la predicción por intervalos de la indemnización económica por el daño corporal del lesionado (Tabla 7.2a y Tabla 7.2b). En el primer caso (las tres primeras columnas de las tablas), suponemos que la compañía no dispone de jurisprudencia previa y que, por tanto, la desviación individual esperada es igual a cero (sólo efectos fijos). En cambio, para la predicción con efectos aleatorios (las tres últimas columnas de las tablas), asumimos que la entidad si que dispone de sentencias previas (efectos fijos más aleatorios).

**TABLA 7.2a Predicción puntual y extremo superior (95%) del intervalo de predicción para el individuo  $j$  (en logaritmos)**

<i>Indemnización media</i> ( $d^*(y)$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}[d^*(y)-\tau^*]}$ )	<i>Extremo superior</i>	<i>Indemnización</i> ( $d(y)$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}[d(y)-\tau]}$ )	<i>Extremo superior</i>
8,102	0,149	8,347	8,133	0,189	8,444

**TABLA 7.2b Predicción puntual y extremo superior (95%) del intervalo de predicción para el individuo  $j$  (escala original)**

<i>Indemnización media</i> ( $\hat{\eta}^*$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}(\hat{\eta}^*-\eta^*)}$ )	<i>Extremo superior</i>	<i>Indemnización</i> ( $\hat{\eta}$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}(\hat{\eta}-\eta)}$ )	<i>Extremo superior</i>
3.337,911	500,122	4.160,612	3.466,360	661,036	4.553,766

De acuerdo con los resultados de la Tabla 7.2b, en el caso de considerar únicamente los efectos fijos, el asegurador debería preferir ir a juicio si la víctima no acepta una indemnización menor a 3.337,91€, sabiendo que, si finalmente se celebra el juicio, el máximo de indemnización económica que podría otorgarse a la víctima en sentencia es de 4.160,61€. Si, por el contrario, también tenemos en cuenta la desviación individual esperada, el límite máximo de negociación de la indemnización para este lesionado será de 3.466,36€, siendo consciente de que, en sentencia, la indemnización otorgada podría alcanzar a lo sumo los 4.553,77€.

Supongamos un nuevo escenario con los mismos datos para el lesionado, pero, en cambio, asumimos que la jurisprudencia previa revela una desviación negativa ( $\hat{b}_j = -0,028$ ). Lógicamente, como se puede observar en la Tablas 7.3a y 7.3b, cuando pronosticamos el valor esperado medio de la indemnización y su extremo superior, estos valores no se ven afectados con respecto a los obtenidos en las Tablas 7.2a y 7.2b, influyendo únicamente en la predicción de la indemnización (y al extremo superior del intervalo) cuando también tenemos en cuenta los efectos aleatorios.

**TABLA 7.3a Predicción puntual y extremo superior (95%) del intervalo de predicción para el individuo  $j'$  (en logaritmos)**

<i>Indemnización media</i> ( $d^*(y)$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}[d^*(y)-\tau^*]}$ )	<i>Extremo superior</i>	<i>Indemnización</i> ( $d(y)$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}[d(y)-\tau]}$ )	<i>Extremo superior</i>
8,102	0,149	8,347	8,074	0,191	8,388

**TABLA 7.3b Predicción puntual y extremo superior (95%) del intervalo de predicción para el individuo  $j'$  (escala original)**

<i>Indemnización media</i> ( $\hat{\eta}^*$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}(\hat{\eta}^*-\eta^*)}$ )	<i>Extremo superior</i>	<i>Indemnización</i> ( $\hat{\eta}$ )	<i>Std.Err.Predic</i> ( $\sqrt{\text{var}(\hat{\eta}-\eta)}$ )	<i>Extremo superior</i>
3.337,911	500,122	4.160,612	3.269,004	630,118	4.305,55

***Provisión para siniestros pendientes***

Como mencionábamos al inicio del capítulo, la compañía aseguradora está muy interesada en obtener una predicción lo más acertada posible del coste que le supondrá indemnizar al lesionado. Por un lado, como mostrábamos en el apartado anterior, esta predicción podrá utilizarla como límite de negociación para lograr acuerdos amistosos con la parte contraria. Por otra parte, la entidad aseguradora, como cualquier otra entidad, al cierre del ejercicio económico debe demostrar que dispone de reservas suficientes que le garanticen hacer frente en el futuro a los riesgos que tenga contraídos. En este sentido, como señalábamos en el capítulo 5, los procesos judiciales normalmente se dilatan durante muchos meses antes de que se dicte sentencia. Por lo tanto, la mayoría de expedientes de la entidad aseguradora referentes a siniestros de daños corporales, cuya indemnización haya sido solicitada por el cauce judicial, no serán liquidados durante el mismo ejercicio económico de apertura del expediente y la compañía deberá provisionarlos adecuadamente antes del cierre del ejercicio.

Una derivación directa del trabajo realizado consistirá, por tanto, en dotar a provisiones por siniestros pendientes, el valor de la predicción estimada de la indemnización de cada lesionado. En las Tablas 7.4a y 7.4b se muestran los

resultados agregados de esta aplicación (para el total de siniestros analizados), diferenciando entre la predicción individual obtenida, si consideramos únicamente los efectos fijos, o si también tenemos en cuenta los efectos aleatorios. Adicionalmente, a modo de comparación, mostramos la provisión total obtenida si dotáramos a reservas el valor estimado del extremo superior del intervalo de predicción, con nivel de confianza del 95%.

**TABLA 7.4a Provisión para siniestros pendientes (efectos fijos)**

	Coste total (€)	Coste total /Provisión (%)
Coste total de los siniestros de acuerdo con lo otorgado en sentencia	925.011,95	—
Provisión total de los siniestros de acuerdo con las predicciones del modelo (efectos fijos)	907.100,07	98,06%
Provisión total de los siniestros de acuerdo con los extremos superiores* pronosticados por el modelo (efectos fijos)	1.165.292,49	125,97%

\* Nivel de confianza del 95%.

**TABLA 7.4b Provisión para siniestros pendientes (efectos fijos + aleatorios)**

	Coste total (€)	Coste total /Provisión (%)
Coste total de los siniestros de acuerdo con lo otorgado en sentencia	925.011,95	—
Provisión total de los siniestros de acuerdo con las predicciones del modelo (efectos fijos + aleatorios)	907.870,04	98,15%
Provisión total de los siniestros de acuerdo con los extremos superiores* pronosticados por el modelo (efectos fijos+aleatorios)	1.187.026,04	128,32%

\* Nivel de confianza del 95%.

Como se puede apreciar, a nivel agregado, no existen diferencias significativas entre aplicar el criterio de provisionar cada siniestro abierto por la predicción del coste de indemnización, considerando únicamente los efectos fijos, o teniendo en



cuenta, además, las desviaciones esperadas individuales. Este resultado es lógico, puesto que, para el caso de los efectos aleatorios, cuando agregamos las predicciones, se compensan entre las desviaciones individuales con signo positivo y aquellas de signo negativo. Por tanto, la estimación del coste de indemnización del lesionado teniendo en cuenta los efectos aleatorios (desviaciones individuales), encontraría su máxima justificación para el cálculo del límite máximo de negociación con la parte contraria, puesto que, en este caso, la compañía debe negociar la cuantía indemnizatoria individualmente. Ahora bien, cuando agregamos el coste de las compensaciones individuales, como en el cálculo de las reservas por siniestros pendientes, las dos metodologías presentadas se comportarían satisfactoriamente.



# Capítulo 8

## Conclusiones

### 8.1 Conclusiones

La elevada accidentalidad en las carreteras europeas y el convencimiento, por parte de la sociedad, de que puede ser combatida con acciones concretas, ha motivado que, al amparo del Programa de Acción Europeo de Seguridad Vial aprobado por la Comisión Europea en el año 2003, en la mayoría de países europeos se estén impulsando programas de seguridad vial que persiguen reducir las enormes consecuencias humanas, sociales y económicas que se derivan de los accidentes de tráfico. En España, un claro ejemplo de medida que pretende incidir sobre el elevado número de accidentes ha sido la entrada en vigor del carnet por puntos, desde el 1 de julio de 2006.

Desde el punto de vista del asegurador, la reducción del número de accidentes con víctimas experimentada en las carreteras españolas en los últimos años, ha originado, obviamente, una disminución en el número de siniestros por daños personales de la cobertura de responsabilidad civil en el seguro del automóvil. No obstante, el continuo incremento del coste unitario de estos siniestros no siempre ha permitido que las reducciones en la accidentalidad se traduzcan en una menor siniestralidad para el sector, calculada como el producto del número de siniestros por el coste de los mismos. Este incremento del coste está motivado por múltiples factores, tanto económicos, como sociales o legales, y en un sector tan competitivo

como el del seguro del automóvil (donde trasladar los incrementos de los costes a las primas cada vez resulta más complicado), supone que las compañías aseguradoras se enfrenten a importantes desafíos si desean que continúe siendo un ramo con resultados técnicos positivos en el futuro.

En primer lugar, al redactar la póliza, el asegurador debe definir con precisión los límites del riesgo asegurable. Una vez ha ocurrido el riesgo, la mejor herramienta de la entidad aseguradora para lograr la contención en el coste consiste en realizar una adecuada tramitación del siniestro hasta su liquidación. Analizando la vida de estos siniestros nos damos cuenta de que normalmente son comunicados rápidamente a la entidad aseguradora, pero que, en cambio, pueden permanecer abiertos durante varios ejercicios económicos hasta su liquidación. Por lo tanto, la compañía aseguradora debe prestar especial atención al cálculo de las reservas por siniestros pendientes de liquidación (RBNS). Por otro lado, en la mayoría de los siniestros por daños personales, la cuantía de la compensación económica es acordada entre las partes, por lo que la entidad aseguradora debe disponer de una estrategia bien definida para afrontar la negociación con la parte contraria.

### ***Reservas por siniestros pendientes de liquidación***

El cálculo de las reservas es una de las principales actividades de una compañía aseguradora. La entidad frecuentemente estima el coste de compensación de los siniestros por daños personales (y, por tanto, provisiona) basándose en los informes médicos periciales que ha realizado a la víctima durante el período de recuperación. Esta práctica puede no estar estimando correctamente el coste de los siniestros, debido a las discrepancias habitualmente existentes entre la gravedad considerada en los informes periciales y la finalmente otorgada en sentencia judicial. Por otro lado, las técnicas estadísticas desarrolladas en la literatura para el cálculo de reservas se han basado, tradicionalmente, en datos agregados en los triángulos de desarrollo. Sin embargo, como señalan England y Verrall (2002), para el cálculo de las reservas, “los modelos basados en siniestros individuales, más que en datos agregados en triángulos, son probablemente un beneficio”. Siguiendo esta reflexión y puesto que la

evaluación de la gravedad de una víctima es a menudo de carácter cualitativo, en la tesis aplicamos un modelo logit ordenado heterocedástico para explicarla. Esta modelización estaría incluida en la metodología de la estimación estadística individual de Taylor y Campbell (2002). Además, asumimos que la información del siniestro no fluye continuamente dentro de la compañía aseguradora, sino que se va ampliando por etapas, por lo que estimamos el modelo de forma secuencial, en cuatro de las etapas más importantes por las que transcurre el expediente antes de su liquidación. En concreto, consideramos la información disponible por la compañía en el momento de apertura del siniestro, con la realización del informe pericial inicial, del informe pericial final y, finalmente, del informe forense. Al final del ejercicio económico, provisionamos los siniestros pendientes de liquidación según la gravedad de los mismos, asignando a cada caso el coste medio de la categoría estimada.

En relación a los atributos de la víctima, de acuerdo con nuestros resultados, la edad y el sexo tienen capacidad explicativa en la gravedad esperada de la lesión, durante las sucesivas etapas de la estimación. El resto de información recogida exactamente en el momento de apertura del siniestro, en cambio, deja de mostrar parámetros significativos a medida que la compañía dispone de los informes médicos periciales. Respecto a dichos informes, la información relativa a la primera evaluación de la víctima (informe pericial inicial) parece guardar una estrecha relación con la gravedad final de la misma, incluso cuando la compañía ya dispone del último informe pericial. Este resultado concluimos que puede deberse a que el informe pericial inicial y el final persiguen objetivos diferentes. Por un lado, el primer informe estaría más enfocado al cálculo de la provisión del siniestro. Por el contrario, el último informe puede actuar como una herramienta de negociación para llegar a un acuerdo sobre la indemnización de la víctima o, en última instancia, como prueba en el juicio.

Otro resultado interesante es que el número de secuelas y días de baja otorgados por el perito médico, estarían reflejando mejor la gravedad final de la víctima que la directa puntuación de las secuelas, también valorada por el perito médico. Ello se debe, probablemente, a que los primeros son datos más objetivos que

la medición de la intensidad de la gravedad. En cuanto a la clasificación de la gravedad, resaltar que el porcentaje de siniestros correctamente estimados por el modelo mejora sustancialmente cuando incorporamos la información del informe forense, indicándonos la enorme influencia que esta información tiene en la valoración de la gravedad final considerada por el juez en sentencia. Por último, como resultado de la modelización, mostramos como, al asignar el coste medio de la categoría de gravedad estimada del siniestro, la provisión por siniestros pendientes que obtenemos es adecuada, representando alrededor del 95% del coste final total de los siniestros, en las tres últimas etapas consideradas. Estos resultados mejoran notablemente el cálculo de provisiones realizado por la entidad aseguradora en base a los informes periciales médicos.

### ***Límite máximo de negociación***

En la segunda parte de la tesis, planteamos un modelo que ayude a la compañía aseguradora a decidir entre negociar la compensación económica con la parte contraria, o ir a juicio. Nuestra muestra está compuesta por siniestros con daños corporales cuya indemnización ha sido concedida mediante sentencia judicial, por lo que planteamos que el coste estimado puede ser considerado como el límite máximo de negociación con la parte contraria, por encima del cual la compañía preferirá ir a juicio. Puesto que la negociación se realiza cuando la víctima está totalmente recuperada, en la estimación del coste asumimos que la compañía dispone de todos los informes médicos y ajustamos un modelo log-lineal mixto que tenga en cuenta la existencia de heterocedasticidad y correlación en los datos. La primera viene provocada por la influencia que la valoración del médico forense (si existe) tiene en la indemnización final concedida. La segunda, en cambio, por el hecho de que diferentes víctimas pueden ser valoradas en una misma sentencia.

Respecto al análisis individual de los parámetros estimados, las variables explicativas muestran un comportamiento muy parecido (en cuanto a signo de los coeficientes y niveles de significación) al obtenido en el modelo logit ordenado heterocedástico secuencial, confirmando, de este modo, la estrecha relación existente

entre la gravedad de la lesión y el coste final de indemnización. Una excepción a destacar es que el parámetro de la variable edad de la víctima ahora no es significativo, debido a que, por un lado, a mayor edad, mayor es la probabilidad de que la lesión sea grave. Sin embargo, por el otro, a mayor edad, menor valor económico otorga el baremo a la puntuación de las secuelas. El hecho de que la edad de la víctima no muestre capacidad explicativa en la estimación de la indemnización económica final, puede venir provocada por una compensación entre ambos efectos.

En cuanto a la correlación en las observaciones, planteamos la hipótesis de que pueda venir motivada por el hecho de que, en algunas ocasiones, la indemnización a las diferentes víctimas de un mismo accidente viene detallada en una misma sentencia. En la fijación de la compensación existen diversos criterios económicos que influyen como, por ejemplo, la aplicación del baremo vigente en el año del accidente o el vigente en el año de la sentencia, el tipo de interés por mora a considerar, el porcentaje de corrección por perjuicios económicos, entre otros. La elección de estos criterios puede ser subjetiva y depender de la persona que determina la indemnización final. La especificación de modelos mixtos permite estimar las desviaciones individuales respecto al valor esperado medio de indemnización (efectos aleatorios), teniendo en cuenta el análisis de la jurisprudencia anterior. La estructura de la muestra utilizada nos lleva a aplicar inferencia generalizada (Tsui y Weerahandi, 1989) en el cálculo del intervalo de confianza del componente de la varianza asociado al efecto aleatorio.

Por último, se demuestra que si dotamos a reservas, por cada siniestro abierto, el máximo de negociación económica estimado, las reservas agregadas cubren prácticamente el cien por cien del coste total de los siniestros, independientemente de si hemos considerado en la estimación las desviaciones individuales, o únicamente el valor esperado medio de la indemnización.

## **8.2 Aportaciones de la investigación**

Este trabajo se estructura en dos partes claramente diferenciadas. La primera parte se centra en el cálculo de la provisión económica para un siniestro durante el tiempo

que éste permanece abierto dentro de la compañía. En cambio, en la segunda parte, se ofrecen mecanismos que ayuden a la compañía aseguradora a la hora de decidir entre negociar la indemnización con la parte contraria o ir a juicio. Para la consecución de estos objetivos, se proponen diferentes metodologías. En el análisis de las aportaciones de este trabajo, hemos creído conveniente mantener también esta estructura.

### ***Reservas por siniestros pendientes de liquidación***

Tradicionalmente, la utilización de los modelos logit ordenados para la modelización de la gravedad de las víctimas de accidentes de circulación, se ha realizado desde el punto de vista de la seguridad vial. Esta metodología no había sido aplicada en el cálculo de las reservas por siniestros pendientes de liquidación (RBNS), es decir, desde el punto de vista de una entidad aseguradora.

La modelización presentada en este trabajo, puesto que se basa en siniestros individuales, puede incluirse dentro de la metodología de estimación estadística individual (EEI). Ahora bien, a diferencia de la EEI clásica, no modeliza el comportamiento de una variable cuantitativa (coste pendiente del siniestro abierto), sino el de una variable cualitativa (gravedad final del siniestro), por lo que representa una contribución innovadora en la misma. Otro aspecto novedoso es su carácter dinámico. La entidad aseguradora no dispone de toda la información del siniestro en el momento de la apertura, sino que la va incorporando a lo largo de su vida. La modelización secuencial tiene en cuenta que el proceso de entrada de información dentro de la entidad aseguradora es dinámico (incorporación de nuevos datos con la recepción de los informes médicos) y con memoria (la entrada de nueva información no supone olvidar la que previamente existía, sino que la actualiza).

La ventaja de la modelización utilizada en la tesis respecto a la EEI clásica (dirigida a estimar costes) es su mayor estabilidad en el tiempo, ya que en esta última influyen aspectos económicos como la inflación, el coste de los servicios médicos, el nivel de los salarios, etc. Esta característica permite, por ejemplo, estimar la probabilidad de que la víctima padezca una determinada gravedad en sus lesiones,



para las diferentes etapas de la vida del siniestro, sin necesidad de considerar supuestos adicionales sobre la variación en los precios. Del mismo modo, cabe esperar que no se produzcan variaciones importantes en el comportamiento de las variables explicativas, si aplicamos esta metodología a siniestros cerrados en un período diferente al de nuestra muestra.

Las aportaciones mencionadas en este apartado dan lugar al artículo de Ayuso y Santolino “Predicting automobile claims bodily injury severity with sequential Ordered Logit Models” aceptado en *Insurance: Mathematics and Economics*. Asimismo, dan lugar al artículo “La valoración del daño corporal en Europa” publicado en *Actualidad Aseguradora*.

### ***Límite máximo de negociación de la indemnización***

En la segunda parte del trabajo podemos destacar dos tipos de aportaciones. La primera es en el campo de investigación de los daños corporales y, por otra parte, la contribución relacionada con la metodología considerada. En cuanto al área de investigación, a pesar de que la mayoría de las compensaciones por daños personales se pactan entre las partes, existe escasa literatura sobre su tratamiento. Los pocos trabajos empíricos existentes se han limitado a valorar la efectividad de la negociación sobre la indemnización. El dotar a la compañía aseguradora de una herramienta que le permita establecer hasta qué cuantía indemnizatoria es preferible llegar a un acuerdo con la parte contraria o cuando, superado ese nivel, es preferible ir a juicio, es un enfoque que no ha sido tratado hasta la fecha.

Respecto a la metodología propuesta, en esta parte de la tesis hemos estimado el coste de indemnización del siniestro en sentencia judicial, basándonos en datos individuales, por lo que la modelización aplicada debe también englobarse en la metodología de la estimación estadística individual (EEI). En particular, sugerimos un modelo log-lineal mixto que tenga en cuenta la heterocedasticidad por grupos y correlación de los datos. En general, la presencia de heterocedasticidad por grupos en el término de perturbación ha sido una cuestión poco tratada en la teoría de los modelos mixtos, y resulta novedosa en la metodología EEI. Por otro lado, mediante

la interpretación ofrecida de los efectos aleatorios del modelo log-lineal mixto, como posibles desviaciones individuales en la cuantificación de la indemnización económica respecto al valor esperado medio de compensación, se muestra de forma sencilla como los modelos lineales generalizados (MLG) pueden ser considerados como un caso particular de los modelos lineales generalizados mixtos (MLGM).

Por último, y como consecuencia de que nuestra muestra es extremadamente no balanceada, observamos problemas de fiabilidad en el intervalo de confianza estimado para el parámetro de la varianza asociado al efecto aleatorio. Por este motivo aplicamos una metodología alternativa para su cálculo, basada en la teoría de inferencia generalizada. Esta aproximación es reciente en la literatura estadística y se ha desarrollado principalmente en el ámbito teórico. En este trabajo se muestra una aplicación empírica. En concreto, consideramos el intervalo de confianza generalizado desarrollado por Park y Burdick (2003) para dos componentes de la varianza (caso homocedástico), y lo extendemos al caso de presencia de heterocedasticidad por grupos.

Las aportaciones mencionadas en este apartado dan lugar al documento de trabajo de Santolino y Ayuso “Forecasting the maximal limit of negotiation for bodily injury claims in automobile insurance”, serie de documentos de trabajo del *Institut de Recerca en Economia Aplicada*, y a un capítulo del libro Guillén, M. y Sarabia, J.M. (Eds.) *Investigación en Seguros y Gestión de Riesgos*.

### **8.3 Futuras líneas de investigación**

#### ***Reservas por siniestros pendientes de liquidación***

En la exposición de líneas futuras, en primer lugar hemos de diferenciar, en la metodología desarrollada para el cálculo de reservas por siniestros pendientes de liquidación, entre la modelización secuencial de la gravedad y el criterio adoptado para la asignación del coste a la categoría estimada.

En cuanto a la modelización de la gravedad, cabe señalar que la regresión logit es un modelo econométrico de clasificación. Una primera línea futura, por tanto,

consistirá en comparar la capacidad predictiva del modelo logit propuesto con la de otras técnicas de clasificación disponibles, como son, por ejemplo, las técnicas discriminantes, los árboles de decisión, las regresiones por *splines* o las redes neuronales artificiales.

En relación al criterio de asignación del coste, en la cuantificación final del daño corporal se tienen en cuenta más conceptos que la puntuación de las secuelas, lo que puede originar importantes diferencias en la cuantía indemnizatoria final otorgada a víctimas con una misma categoría de gravedad de la lesión (por ejemplo, individuos con menos de 15 puntos y una elevada indemnización). En consecuencia, considerar únicamente el coste medio muestral de la categoría puede ser un criterio excesivamente simple y, por tanto, cabría complementarlo con otros momentos de la distribución como, por ejemplo, la varianza muestral.

#### ***Límite máximo de negociación de la indemnización***

En la estimación de los efectos aleatorios pueden utilizarse técnicas bayesianas. Algunas de las ventajas de esta aproximación son que permite considerar estructuras más complicadas para el predictor lineal o utilizar diferentes distribuciones para los efectos aleatorios (mediante técnicas de maximización de la verosimilitud, en numerosos paquetes estadísticos únicamente está implementada la distribución normal). En este sentido, una posible línea futura de investigación consistirá en comparar las estimaciones obtenidas mediante el criterio de maximización de la verosimilitud, y las que obtendríamos aplicando principios bayesianos.

Finalmente, en el capítulo 7 de esta tesis señalábamos que la estimación del intervalo de confianza generalizado para el componente de la varianza asociado al efecto aleatorio, se deriva de una variable de contraste generalizada y que ésta raramente es única. Por lo tanto, una extensión natural de la presente investigación se basará en comparar entre diferentes variables de contraste generalizadas, en función de la amplitud y la probabilidad de cobertura de los intervalos de confianza que generen.

### ***Fraude***

Por último señalar que otra línea de investigación futura consistirá en complementar la presente investigación con el análisis del fraude en los siniestros por daños corporales. Principalmente, investigar como incide en el cálculo de las reservas RBNS y el coste de indemnización, la sospecha de que en un siniestro se haya producido un comportamiento fraudulento con el fin de conseguir una indemnización de mayor cuantía a la que verdaderamente correspondería.

# Chapter 8

## Conclusions

### 8.1 Conclusions

The high traffic accident rate in Europe and the conviction of the society that it can be combated with specific actions have motivated most of European countries to establish road safety programs. These programs have been launched under the protection of the European road safety action program, approved by the European Commission in 2003, and they pursue to reduce the socioeconomic consequences of automobile accidents. In Spain, for example, a measure that tries to affect the high accident rate is the new legislation concerning the driving licence with penalty points, which came into force from July 1<sup>st</sup> 2006.

From the insurer viewpoint, the reduction in the number of injured victims involved in road accidents in Spain during last years has led to a decrease in the frequency of Bodily Injury (BI) claims in the motor liability insurance. Nevertheless, the decline in the frequency of claims has been compensated for the continuous increment in the unitary cost of BI claims and, as result, the overall BI cost in the automobile insurance has not decreased. The increase in the cost of BI claims is caused by economic, social and legal reasons. On the other hand, in recent years automobile insurance has become more competitive business for insurers. Therefore, they can hardly translate BI claim cost increments into higher premiums. The combination of both effects (costs rising and constant premiums) leads insurers to

face new challenges if they wish that automobile insurance continues being a business with positive technical results in the future.

Firstly, insurers should pay close attention to the policy wording. At this stage underwriters must define clearly and in great detail which risks are covered and in which limits of insurability. When the underwritten risk occurs, processing the claim efficiently until its closure is the best available mechanism for controlling the final claim cost. Monitoring the life of BI claims, we notice that these claims are normally reported to the insurer in a short period. Nevertheless, they may remain opened inside the company during several financial years before being compensated. Therefore, insurance company should take special care to calculate reserves for Reported But Not Settled claims (RBNS reserves). On the other hand, in most auto bodily injury liability claims the final compensation is negotiated (and agreed) between the claimant and the insurer. Consequently, the insurer should have a well-defined negotiation strategy to face the negotiation process.

### ***Reserves for Reported But Not Settled claims***

Calculating reserves for bodily injury (BI) claims becomes a main task for automobile insurers. The insurer frequently estimates (and therefore reserves) bodily injury compensation according to his own medical reports. These reports are made by medical experts of the company who visit to the victims during the recovery period. This practice may misestimate the final cost, because in many cases there exist huge differences between the injury severity awarded by the judge and the severity considered by the medical expert. On the other hand, the statistic techniques developed in the actuarial literature for calculating reserves are based primarily on aggregated data in run-off triangles. England and Verrall (2002) point out that “models based on individual claims, rather than data aggregated into triangles, are likely to be of benefit”. Following this idea and since the evaluation of the victim’s severity is often only qualitative, we apply in this thesis a heteroscedastic ordered logit model to explain it. The methodology covered by this section falls within the general class of statistical case estimation (SCE) defined by Taylor and Campbell

(2002). In addition, we assume that the BI claim information available to the insurer does not flow continuously, but it is expanded at different stages during the life of the claim file. For this reason, we estimate the model sequentially, considering different information levels of the claim, and incorporating this information gradually. The model is applied in four of the most important phases of the claim processing. Namely, we estimate the heteroscedastic ordered logit model with the available claim information after the claim is reported to the company, after the first internal medical report, after the final internal medical report (full recovery of the victim), and after the forensic report. At the end of the financial year, we reserve outstanding claims allocating the mean cost of the corresponding category to each prediction.

In relation to the information obtained when the claim is opened, the victim's age and gender have explanatory capacity of the injury severity during the successive stages of the model estimation. The rest of variables lose explanatory capacity at the moment when the insurance company has a professional assessment of the damages (medical reports). Regarding the medical reports, our results show that the initial medical report still provides with relevant information in order to explain the injury severity, even when the company already has the final medical report. We suggest that this result is due to the fact that medical reports may pursue different aims which are not necessarily compatible. Whereas the first medical report would be closer to the objective of calculating the claim provision, the final medical report would be also used by the company as a negotiation tool or, as a last resort, as evidence at the trial.

Among the information considered by the insurer's medical expert, another interesting result is that the number of sequelae, or the number of recovery days would reflect the underlying severity better than direct assessment of the sequelae. This result is likely due to the fact that the former variables collect more objective information than the latter one. Note that when the medical expert assesses the sequelae, he is measuring the intensity of the injury. Concerning the last stage, let us emphasize that the percentage of cases accurately estimated by the model increases

notably when the forensic report information is included. This result indicates the strong influence of the forensic evaluation on the severity awarded by the judge. Finally, if we allocate the mean cost of the corresponding category to the predicted severity, we show that the resulting estimated provision for outstanding claims is adequate. Namely, the reserve calculated by means of this technique represents around the 95% of the total cost in the last three stages of the model.

### *Negotiation limit*

In the second part of this thesis, we specify a model which may help the insurer in the negotiation process. Particularly, the suggested methodology assists him to decide whether to negotiate with the plaintiff for the settlement of the BI claim, or to go to trial. Notice that the fact that the compensation amount was determined by judicial verdict is the common characteristic of the BI claim records in our database. In consequence, we suggest the insurer may consider the estimated claim cost as the negotiation limit, i.e. the maximum amount that the underwriter would be willing to offer the claimant in the negotiation process. Since the two parties negotiate the compensation when the victim is fully recovered, we assume that the insurance company already has all medical reports at the negotiation moment. We fit a mixed log-linear model to estimate the individual monetary compensation awarded in judicial sentence. Heteroscedasticity and autocorrelation are considered. The former is due to the influence of the forensic valuation (if it exists) on the final compensation and the latter occurs when more than one claimant (BI victim) is involved in the judicial sentence.

Regarding the statistical analysis of the parameters, the regressors behave very similar (in coefficient sing and significance level) to those showed in the estimated sequential heteroscedastic ordered logit model. This result would confirm the close relationship between the injury severity and the final compensation amount. An exception is the lack of significance of the victim age coefficient in the estimated mixed model. In the first part of the thesis, we demonstrated that the victim's age is positively correlated with the injury severity. On the other hand, the monetary value



stipulated in the legislative scale for the injury assessment is inversely related to the victim's age. We propose that the lack of explanatory capacity of the victim's age in the compensation limit may be due to the fact that the aforementioned effects counteract each other.

Concerning the correlation, BI victims involved in the same automobile accident are often indemnified in the same judicial sentence. In this sense, we assume that the correlation among observations occurs when more than one victim is compensated in the same verdict. This correlation is caused from the fact that there are several economic criteria which influence the final settlement, such as applying the legislative scale which was in force in the accident year or in the sentence year, the applicable percentage of correction for economic damages or the delay interest rate, etc. The choice of these criteria may be subjective and it would depend on the person who sets the final compensation. Mixed models permit us to estimate individual deviations from the expected mean value of compensation (random effects). These estimates of the individual variations are based on previous jurisprudence. Due to the framework of the sample, we consider generalized inference (Tsui and Weerahandi, 1989) in the estimate of the confidence interval on the variance component related to the random effect.

Lastly an application to claim reserving is presented. In particular, we allocate to reserves the predicted maximum negotiation value of each outstanding claim. To compute the maximum negotiation value we differentiate between the expected mean cost (fixed effects) and the situation when the individual deviation is also included in the prediction (mixed effects). We show that the aggregated provision obtained covers almost the 100% of the total claims cost regardless of the estimate considered (only fixed effects or also with random effects).

## **8.2 Contributions of this research**

This thesis is structured into two different parts. The first part is focused on the individual estimate of the BI claim provision during the time that the claim is opened inside the company. On the contrary, in the second part are showed techniques which

may help insurance companies in the claim settlement process. Namely, we develop a tool which indicates the insurer the maximum amount to offer in the negotiation before going to trial. In each part of the thesis, the implemented methodologies to achieve the aforementioned objectives are notably different. For this reason, we believe it is a good idea to keep this division in the current section.

### ***Reserves for Reported But Not Settled claims***

Traditionally, estimating injury severity of victims involved in automobile accidents (by means of logit models) has been done from the road safety viewpoint. This methodology has not been applied to estimate reserves for Reported But Not Settled Claims (RBNS) to date, i.e. from the insurer viewpoint.

Since the suggested model estimates individual BI claims, it could be included into the general class of statistical case estimation (SCE) methodology. Unlike the classic SCE methodology, our model fits a qualitative variable (injury severity) instead of a quantitative variable (cost of outstanding claims). This approach would represent a new contribution into the SCE methodology. Another contribution of the proposed methodology is its dynamic framework. The insurer normally doesn't know all of the claim information at the moment when it is reported. In fact, the company expands gradually the claim information during the claim's life. The sequential model takes into account that the available claim information is a dynamic process (new claim information in the incoming medical reports) and with memory (the insurer does not forget the previous information, but he updates it).

Comparing our approach with the classic SCE methodology, the main advantage of our approach is its independence of economic factors. Indeed, the classic SCE methodology estimates the claim cost and therefore depends on a lot of economic factors, such as the inflation rate, medical expenses or the wages level, etc. By means of the ordered logit model, for instance, we may calculate the probability of a certain injury severity at different stages of the claim's life. The independence characteristic of our model allows us then to compare the probability transition among the claim stages without making additional assumptions regarding price

variations. Furthermore, if we apply our methodology to estimate the severity of claims which were settled in a period different from the period under review, we could expect not important variations in the behaviour of the explanatory variables.

The aforementioned contributions of this section lead to the Ayuso and Santolino's paper "Predicting automobile claims bodily injury severity with sequential Ordered Logit Models" accepted in *Insurance: Mathematics and Economics*. These contributions also lead to the paper "La valoración del daño corporal en Europa" published in *Actualidad Aseguradora*.

### *Negotiation limit*

In the second part of this thesis we may distinguish between two kinds of contributions. The first one is the innovation in the field of automobile bodily injury research and, the second one, the contribution related to the applied methodology. Regarding the branch of knowledge, despite of the fact that most motor bodily injury (BI) settlements are negotiated between the parties involved, relatively little research has been done in this area. This previous research was primarily focused on evaluating the effectiveness of the negotiation process, i.e. its effect on the final settlement. In this thesis we provide the insurer with a tool which informs him about the maximum amount to offer in the negotiation process. It is a new approach in the negotiation area.

In relation to the methodology used in this second part, we estimate the BI claim compensation awarded in judicial sentence. This methodology is based on individual claims and therefore it falls within the general class of statistical case estimation (SCE) as well. In particular, we develop a log-linear mixed model which considers autocorrelation and groupwise heteroscedasticity. To deal with groupwise heteroscedasticity could be considered as rather unusual in the mixed model theory, and is completely new in the SCE methodology. In addition, we show in a straightforward form that the Generalized Linear Models (GLM) are a particular case of Generalized Linear Mixed Models (GLMM). Namely we demonstrate that the estimated random effects of the mixed log-linear model may be considered as

expected deviations in the compensation settlement, i.e. individual variations from the mean value of the BI compensation.

Finally, we observed that the estimated confidence interval on the variance component related to the random effect may be unreliable, because of the extremely unbalanced sample sizes of this study. For this reason, in the last chapter we deal with an alternative methodology to compute confidence intervals on variance components. In particular, we use generalized inference to estimate the confidence interval on the variance component related to the random effect. The generalized inference theory is a rather recent approach in the statistical literature and it has been mainly developed theoretically. This paper presents one of its possible empirical applications. In particular, we consider the generalized confidence interval suggested by Park and Burdick. They deal with the construction of a confidence interval on the variance component related to the random effect in mixed linear models with two variance components (homoscedastic case). We extend this confidence interval to the case of groupwise heteroscedasticity.

The aforementioned contributions of this section lead to the paper “Forecasting the maximal limit of negotiation for bodily injury claims in automobile insurance” in the series of working papers of the Applied Economics Research Institute (IREA), and to a chapter in the book Guillén, M. y Sarabia, J.M. (Eds.) *Investigación en Seguros y Gestión de Riesgos*.

### **8.3 Further research**

#### ***Reserves for Reported But Not Settled claims***

Regarding the suggested methodology to estimate RBNS reserves, firstly we should distinguish between the sequential modelling of the BI severity and the subsequent criterion of allocating to each prediction the mean cost of the corresponding category.

Concerning the BI severity modelling, the ordered logit regression is a statistical classification model. Therefore, a first future line of research could be to evaluate the predictive capacity of the ordered logit model comparing with other

available classification techniques. Some of these techniques are, for example, discriminant techniques, decision trees, regressions splines models or artificial neuronal networks, among others.

In relation to the cost allocation criterion, let us point out that the final compensation depends on different factors. One of them is the final score for sequelae, but there are many others. In consequence, two victims classified in the same category of injury severity, they could receive very different compensation amounts (like, for example, a person with a score less than 15 points for the sequelae, but with a high compensation awarded). Therefore, allocating to reserves the mean cost of each category could be a far too simple criterion and it could be supplemented with other moments of the cost distribution like, for example, the sample variance.

### ***Negotiation limit***

Instead of frequentist inference, Bayesian techniques may be used to estimate the random effects. This approach has several advantages comparing to frequentist methods of inference. For example, Bayesian approach permits us to assume other distributions for the random effects different from the normal (in several statistical software packages the normal distribution is the only available). Additionally, more complex structures for the linear predictor may be considered using Bayesian techniques. Further research will consist, for instance, in comparing the estimates when maximum likelihood and Bayesian techniques are applied.

Finally, in the chapter 7 of this thesis we remarked that the estimated generalized confidence interval on the variance component related to the random effect is derived from a generalized test variable. We pointed out that the generalized test variable is rarely unique (and therefore neither the confidence interval derived). As a consequence, a logical extension of the present research would be considering different generalized test variables in the construction of the generalized confidence intervals and comparing them, i.e. in function of the coverage probability and length of the confidence intervals yielded.

***Fraud***

In conclusion, we would like to mention that the fraud analysis in automobile bodily injury claims is another future line of research. Principally, we are interesting in studying the effect of fraud suspicious on the RBNS calculation and the claim compensation payment.

# Apéndices

## Apéndice A. Legislación básica sobre el seguro del automóvil

### *Legislación estatal básica sobre el aseguramiento de vehículos a motor*

- *Ley de uso y circulación de vehículos de motor 122/1962, de 24 de diciembre.*
- *Decreto 632/1968, de 21 de marzo.* Se publicó el texto refundido de la Ley 122/1962, en el que se incluyeron las modificaciones necesarias para la creación del título ejecutivo en el proceso penal.
- *Real Decreto legislativo 1304/1986, de 28 de junio.* Se procedió a adaptar el texto refundido de la ley de uso y circulación de vehículos a motor, aprobado por decreto 632/1968, al ordenamiento jurídico comunitario. En particular, en lo relativo al aseguramiento de la responsabilidad civil derivada de la circulación de automóviles, lo que supuso la modificación de la normativa del seguro obligatorio del automóvil y la ampliación de las funciones del Consorcio de Compensación de Seguros.
- *Real Decreto 2641/1986, de 30 de diciembre (nuevo Reglamento).* Se adapta la ley de uso y circulación de vehículos de motor 122/1962, al ordenamiento jurídico comunitario. En el seguro de responsabilidad civil se establece la cobertura integral de los gastos de

asistencia médica y hospitalaria a las víctimas, siempre que se preste en centros sanitarios reconocidos por el Consorcio.

- *Resolución de 25 de enero de 1989, publicada 8/2/89*, del Consorcio de Compensación de Seguros, por la que se publica el convenio de asistencia sanitaria privada para accidentes de tráfico para el año 1989 (primer convenio de asistencia sanitaria con instituciones privadas que posteriormente se ha ido actualizando).
- *Resolución de 4 de mayo de 1989, publicada 22/5/89*, del Consorcio de Compensación de Seguros, por la que se publica el convenio marco de asistencia sanitaria para accidentes de tráfico para el año 1989 con instituciones sanitarias públicas (primer convenio de asistencia sanitaria con instituciones públicas que posteriormente se ha ido actualizando).
- *Orden 6601/1991, de 5 de marzo*, por la que se daba publicidad a un sistema de valoración (de aplicación voluntaria) de los daños personales en el seguro de Responsabilidad Civil ocasionados por medio de vehículos de motor, y se consideraba al mismo como procedimiento apto para calcular las provisiones técnicas o prestaciones pendientes correspondientes a los siniestros de dicho seguro. Este sistema incorporaba un mecanismo de actualización anual automática de las indemnizaciones por muerte o incapacidad, que consistía en multiplicar el salario mínimo interprofesional vigente en cada momento por un número de mensualidades determinado.
- *Real Decreto 1559/1992, de 18 de diciembre*. Se elevó el importe mínimo de cobertura de responsabilidad civil en el seguro obligatorio de vehículos a motor, como consecuencia de la voluntad de los legisladores en una aproximación gradual a la legislación de la Comunidad Económica Europea.
- *Resolución 4781/1993, de 1 de febrero*, de la Dirección General de Seguros, por la que se daba publicidad a las cuantías de las indemnizaciones por incapacidad y muerte que resultasen de aplicar el sistema de valoración de daños personales.



Motivo: En el RD 44/1993 se elevó el salario mínimo interprofesional y, por tanto, era necesario actualizar las indemnizaciones.

- *Ley 30/1995, de 8 de noviembre, de Ordenación y Supervisión de los Seguros Privados (LOSSP)*. En su Disposición adicional tercera, se modificó la anterior Ley de Uso y Circulación de Vehículos de Motor, 122/1962, pasando a denominarse *Ley sobre Responsabilidad Civil y Seguro en la Circulación de Vehículos a Motor*. Esta Ley supuso un sustancial cambio en la regulación de la normativa del seguro de suscripción obligatoria de automóviles y la adaptación a la Directiva europea 90/232/CEE, relativa a la aproximación de las legislaciones de los Estados Miembros.

Cabe destacar dos aspectos novedosos de la ley 30/1995. Por un lado, se aprobó un sistema obligatorio para la valoración de los daños y perjuicios causados a las personas, que se aplicaría a la valoración de todos los daños ocasionados a las personas en accidentes de circulación (*baremo*). Por otro lado, se establece la necesidad de crear un fichero informativo de vehículos asegurados que permita, a las personas implicadas en un accidente de circulación, averiguar, a la mayor brevedad posible, las circunstancias relativas al contrato de seguro y a la entidad aseguradora correspondiente (el fichero FIVA).

- *Real Decreto 2486/1998, de 20 de noviembre, por el que se aprueba el Reglamento de Ordenación y Supervisión de los Seguros Privados*. Regula la actividad de las entidades aseguradoras, tanto nacionales como extranjeras.

- *Resolución de 22 de febrero de 1999, de la Dirección General de Seguros*. Mediante esta Resolución se introduce por primera vez el apartado de día de baja no impeditivo en la Tabla V del *baremo*.

- *Real Decreto 7/2001, de 12 de enero, por el que se aprueba el Reglamento sobre la Responsabilidad Civil y Seguro en la Circulación de Vehículos a motor* (nuevo Reglamento que sustituye al RD 2641/1986). Se establecen los requisitos, la forma y la

periodicidad en que las aseguradoras deben remitir la información correspondiente al fichero FIVA, para conseguir que las personas implicadas en un accidente de circulación conozcan las circunstancias relativas al contrato de seguro y a la entidad aseguradora. También se reguló la obligación de las entidades aseguradoras de suministrar la información relativa a los vehículos asegurados al Ministerio de Economía y Hacienda, mediante su remisión al Consorcio. Finalmente, incorpora, con carácter definitivo, la obligación recogida en la Directiva 84/5/CEE, de aproximación de las legislaciones de los Estados Miembros en el seguro de Responsabilidad Civil.

- *Ley 34/2003, de 4 de noviembre, de modificación y adaptación a la normativa comunitaria de la legislación de seguros privados.* Se modifica la Tabla VI del *baremo* con el objetivo de aproximarla a un futuro baremo europeo.
- *Real Decreto legislativo 8/2004, de 29 de octubre.* Se aprueba el texto refundido con las modificaciones introducidas en la Ley 34/2003, para adaptarse a la normativa europea. En particular, se introducen las garantías necesarias para que una víctima residente en España sea indemnizada cuando el vehículo que causa el accidente, no estaba asegurado, independientemente del Estado Miembro de estacionamiento habitual del vehículo. También se incluye la nueva Tabla VI del *baremo*.
- *Proyecto de reforma de la Ley de responsabilidad civil y seguro de circulación,* aprobado por el Consejo de Ministros, el 7 de febrero de 2007, con el fin de adaptarse a la quinta Directiva. Entre otras propuestas, las aseguradoras dispondrán de tres meses desde el siniestro, para realizar una oferta de indemnización. Además, deberán indemnizar los daños ocasionados, en caso de robo del vehículo asegurado. Por último, se elimina la posibilidad que tenían las compañías aseguradoras de repetir el importe de la indemnización pagada contra el asegurado, por aquellas causas que estuvieran recogidas en el contrato del seguro.

### ***Legislación comunitaria básica sobre el seguro del automóvil***

Hasta la fecha se han elaborado cinco Directivas sobre el seguro del automóvil en el marco de la Unión Europea, además de numerosas recomendaciones y decisiones de la Comisión Europea.

- *La primera Directiva (72/166/CEE), de 24 de abril de 1972.* Mediante esta directiva, se estableció la obligatoriedad de aseguramiento para los Estados Miembros.
- *La segunda Directiva (84/5/CEE), de 30 de diciembre de 1983.* En esta Directiva se fijan los límites de aseguramiento obligatorio para los daños materiales y los daños personales. Por otra parte, se establece la imposibilidad, por parte de las entidades aseguradoras, de oponerse a indemnizar a las víctimas en base al incumplimiento, por parte del asegurado, de las condiciones de la póliza.
- *La tercera Directiva (90/232/CEE), de 14 de mayo de 1990.* Por esta Directiva, se garantiza la cobertura de los ocupantes del vehículo, excepto al conductor, en todo el territorio de la UE.
- *La cuarta Directiva (2000/26/CE), de 16 de mayo de 2000,* se centra en la aproximación de las legislaciones de los Estados Miembros en el seguro de responsabilidad civil. El principal objetivo de esta Directiva es mejorar la defensa de los derechos de los perjudicados en siniestros ocurridos en un Estado Miembro distinto al de residencia.
- *La quinta Directiva (2005/14/CE), de 11 de mayo de 2005.* Se extiende la defensa de los derechos de los perjudicados introducida en la cuarta Directiva, permitiendo a cualquier víctima reclamar desde su país de residencia contra un asegurador que tenga su sede en un Estado Miembro. Por otro lado, se incrementa la cuantía mínima de aseguramiento obligatorio por responsabilidad civil.

## Apéndice B. Cuestionario realizado para cada lesionado de la muestra

Número de siniestro

Número de víctima

### I. Información del siniestro

Fecha siniestro \_\_\_ / \_\_\_ / \_\_\_

Fecha apertura \_\_\_ / \_\_\_ / \_\_\_

Fecha cierre expte. \_\_\_ / \_\_\_ / \_\_\_

Gravedad de las lesiones  M  G  LE  DB

Culpa  Asegurado  Dudosa  Contrario

Tipo vehículo (del lesionado)  Moto  Turismo  Camión/Furgoneta  Peatón/ciclista

Valor Total de Pagos (por lesionado) \_\_\_\_\_ €

Coste días de baja Hospitalarios	€
Coste días de baja Impeditivos	€
Coste días de baja No Impeditivos	€
Costes Secuelas/Invalidez	€
Coste Gastos de Curación	€
Coste Garantía Ocupantes	€
Coste Gastos de Gestión	€
Otros costes	€

### II. Datos del lesionado

Edad \_\_\_\_\_ años

Sexo Hombre  Mujer

Estado civil  Casado  Soltero  Divorciado  Pareja Hecho  Viudo

Actividad  Activo C/A  Activo C/P  Estudiante  Jubilado  Desocupado

Tipo de víctima  Ocupante Propio  Conductor Contrario  Ocupante Contrario  Peatón/Ciclista

### III. Informe perito médico inicial

Fecha visita \_\_\_ / \_\_\_ / \_\_\_

Motivo no realización  No se deja visitar  Paradero desconocido  Cía. no realizó  
 Desconocimiento existencia lesionado

Profesión \_\_\_\_\_

Secuelas

Código	Estado	Ptos.	Código	Estado	Ptos.
Puntuación total secuelas					

Días de baja

	No.
Días de baja Hospitalarios	
Días de baja Impeditivos	
Días de baja No impeditivos	

Incapacidad permanente  Ninguna  Parcial  Total  Absoluta

Gran Invalidez Sí  No

**IV. Informe perito médico final**

Fecha visita \_\_\_/\_\_\_/\_\_\_

Motivo no realización  No se deja visitar  Paradero desconocido  Cía. no realizó  
 Desconocimiento existencia lesionado

Secuelas

Código	Ptos.	Código	Ptos.
Puntuación total secuelas			

Días de baja

	No.
Días de baja Hospitalarios	
Días de baja Impeditivos	
Días de baja No impeditivos	

Incapacidad permanente  Ninguna  Parcial  Total  Absoluta

Gran Invalidez Sí  No

**V. Informe médico forense**

Fecha informe médico \_\_\_ / \_\_\_ / \_\_\_

Puntúa secuelas  Sí  No  Marca Horquilla

Médico forense emisor

Clinica Médico Forense	
Forense del juzgado instructor	
Forense juzgado de paz o exhortado	
Médico cabecera (exhorto)	

Secuelas

Código	Ptos.	Código	Ptos.
Puntuación total secuelas			

Días de baja

	No.
Días de baja Hospitalarios	
Días de baja Impeditivos	
Días de baja No impeditivos	

Incapacidad permanente  Ninguna  Parcial  Total  Absoluta

Gran Invalidez Sí  No

Contiene otras consideraciones cuantificables económicamente \_\_\_\_\_

**VI. Sentencia**

Código del juzgador \_\_\_\_\_

Juzgado Tipo  1ªInst.  1ªInst.e Inst  Penal No. \_\_\_\_\_ De \_\_\_\_\_

Fecha sentencia \_\_\_ / \_\_\_ / \_\_\_

Tipo de sentencia Civil  Penal  Ejecutivo

Costas Sí  No  A cargo de: Cía.Aseg  Contrario  Compartidas

Factor reducción por concurrencia de culpa en accidente \_\_\_\_\_ (en % C.O/Contrario)

Aplicación del baremo: Año del siniestro  Año de la sentencia  Otro

Puntuación de las secuelas \_\_\_\_\_

Valoración de las secuelas: \_\_\_\_\_ €

Días de baja (Puntuación)

Días de baja Hospitalarios	
Días de baja Impeditivos	
Días de baja No impeditivos	

Días de baja (Valoración)

Días de baja Hospitalarios	€
Días de baja Impeditivos	€
Días de baja No impeditivos	€

Factor corrección por perjuicios económicos \_\_\_\_\_%      Afecta DB   Sí    No

Daños morales complementarios (secuela >75 puntos o concurrentes > 90puntos) \_\_\_\_\_€

Incapacidad permanente  Ninguna  Parcial  Total  Absoluta   Valoración: \_\_\_\_\_€

Gran invalidez   Sí    No    Valoración básica: \_\_\_\_\_€

Elementos correctores a la indemnización por Gran Invalidez (concepto y valor) \_\_\_\_\_

Muerte   Sí    No    Valoración básica: \_\_\_\_\_€

Elementos correctores a la indemnización por muerte (concepto y valor) \_\_\_\_\_

Costes económicos justificados (excepto daños materiales) \_\_\_\_\_€

Costes por daños materiales (al vehículo) \_\_\_\_\_€

Total sentencia \_\_\_\_\_€

Intereses  Sin intereses  Legal + 50%  20% a partir del 2º año  20% desde fecha siniestro

Apelación   No

                  Sí   Confirme  Revoca parcial a la baja  Revoca parcial al alta  Revoca total

Observaciones \_\_\_\_\_

**VII. Apelación**

Fecha sentencia \_\_\_/\_\_\_/\_\_\_

Motivo de la apelación \_\_\_\_\_  
 \_\_\_\_\_

Parte apelante: Cía.Aseg  Contrario  Ambas

Costas Sí  No  A cargo de: Cía.Aseg  Contrario  Compartidas

Factor reducción por concurrencia de culpa en accidente \_\_\_\_\_ (en % Cía/Contrario)

Aplicación del baremo: Año del siniestro  Año de la sentencia  Otro

Puntuación de las secuelas  Valoración de las secuelas: \_\_\_\_\_ €

Días de baja (Puntuación)

Días de baja Hospitalarios	<input type="text"/>
Días de baja Impeditivos	<input type="text"/>
Días de baja No impeditivos	<input type="text"/>

Días de baja (Valoración)

Días de baja Hospitalarios	<input type="text"/>	€
Días de baja Impeditivos	<input type="text"/>	€
Días de baja No impeditivos	<input type="text"/>	€

Factor corrección por perjuicios económicos \_\_\_\_\_ % Afecta DB Sí  No

Incapacidad permanente  Ninguna  Parcial  Total  Absoluta Valoración: \_\_\_\_\_ €

Gran invalidez Sí  No  Valoración: \_\_\_\_\_ €

Muerte Sí  No  Valoración: \_\_\_\_\_ €

Costes económicos justificados (excepto daños materiales) \_\_\_\_\_ €

Costes por daños materiales (al vehículo) \_\_\_\_\_ €

Total de la nueva sentencia \_\_\_\_\_ €

Intereses  Sin intereses  Legal + 50%  20% a partir del 2º año  20% desde fecha siniestro

Observaciones \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_



## Apéndice C. Descripción de las variables de la muestra

Nombre	Descripción	Perdidos	Número de casos	No proceden
<b><i>I. Información del siniestro</i></b>				
<i>id1</i>	Número de siniestro.	0	204	0
<i>id2</i>	Número de víctima.	0	204	0
<i>g1</i>	Fecha siniestro.	0	204	0
<i>g2</i>	Fecha apertura.	0	204	0
<i>g3</i>	Culpa del siniestro (1 si asegurado; 2 contrario; 3 dudosa).	0	204	0
<i>g4</i>	Tipo de vehículo del lesionado (1 si moto; 2 turismo; 3 furgoneta; 4 peatón/ciclista).	0	204	0
<i>g5</i>	Valor total de pagos (por lesionado).	0	204	0
<i>g5a</i>	Coste por días de baja hospitalarios.	0	204	0
<i>g5b</i>	Coste por días de baja improductivos.	0	204	0
<i>g5c</i>	Coste por días de baja no improductivos.	0	204	0
<i>g5d</i>	Costes por secuelas/invalidez.	0	204	0
<i>g5e</i>	Coste por gastos de curación.	0	204	0
<i>g5f</i>	Coste por garantía de ocupantes.	0	204	0
<i>g5g</i>	Coste por gastos de gestión.	0	204	0
<i>g5h</i>	Otros costes.	0	204	0
<b><i>II. Datos del lesionado</i></b>				
<i>l1</i>	Edad del lesionado.	0	204	0
<i>l2</i>	Sexo (1 si hombre; 0 en caso contrario).	0	204	0
<i>l3</i>	Estado civil (1 si casado; 2 soltero; 3 divorciado; 4 pareja de hecho; 5 viudo).	18	186	0
<i>l4</i>	Actividad laboral (1 si activo por cuenta ajena; 2 activo por cuenta propia; 3 estudiante; 4 jubilado; 5 desocupado).	25	179	0
<i>l5</i>	Tipo de víctima (1 si ocupante propio; 2 conductor contrario; 3 ocupante contrario; 4 peatón/ciclista).	0	204	0
<b><i>III. Informe pericial inicial</i></b>				
<i>p1</i>	Fecha de la visita médica.	78	119	7
<i>p2</i>	Motivo de no realización del informe (1 si no se deja visitar; 2 paradero desconocido; 3 compañía no lo realizó; 4 desconocimiento existencia lesionado).	0	197	7
<i>p3</i>	Profesión.	90	107	7
<i>p4.1a</i>	Código de la primera secuela.	78	119	7
<i>p4.1b</i>	Estado 1ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.1c</i>	Puntuación de la primera secuela.	78	119	7

<i>p4.2a</i>	Código de la segunda secuela.	78	119	7
<i>p4.2b</i>	Estado 2ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.2c</i>	Puntuación de la segunda secuela.	78	119	7
<i>p4.3a</i>	Código de la tercera secuela.	78	119	7
<i>p4.3b</i>	Estado 3ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.3c</i>	Puntuación de la tercera secuela.	78	119	7
<i>p4.4a</i>	Código de la cuarta secuela.	78	119	7
<i>p4.4b</i>	Estado 4ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.4c</i>	Puntuación de la cuarta secuela.	78	119	7
<i>p4.5a</i>	Código de la quinta secuela.	78	119	7
<i>p4.5b</i>	Estado 5ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.5c</i>	Puntuación de la quinta secuela.	78	119	7
<i>p4.6a</i>	Código de la sexta secuela.	78	119	7
<i>p4.6b</i>	Estado 6ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.6c</i>	Puntuación de la sexta secuela.	78	119	7
<i>p4.7a</i>	Código de la séptima secuela.	78	119	7
<i>p4.7b</i>	Estado 7ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.7c</i>	Puntuación de la séptima secuela.	78	119	7
<i>p4.8a</i>	Código de la octava secuela.	78	119	7
<i>p4.8b</i>	Estado 8ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.8c</i>	Puntuación de la octava secuela.	78	119	7
<i>p4.9a</i>	Código de la novena secuela.	78	119	7
<i>p4.9b</i>	Estado 9ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.9c</i>	Puntuación de la novena secuela.	78	119	7
<i>p4.10a</i>	Código de la décima secuela.	78	119	7
<i>p4.10b</i>	Estado 10ª secuela (1 si permanente; 0 en caso contrario).	78	119	7
<i>p4.10c</i>	Puntuación de la décima secuela.	78	119	7
<i>p4</i>	Puntuación total de las secuelas.	78	119	7
<i>p5.1</i>	Número días de baja hospitalarios.	78	119	7
<i>p5.2</i>	Número días de baja improductivos.	78	119	7
<i>p5.3</i>	Número días de baja no improductivos.	78	119	7
<i>p6</i>	Incapacidad permanente (1 si ninguna; 2 parcial; 3 total; 4 absoluta).	78	119	7
 <b><i>IV. Informe pericial final</i></b>				
<i>u1</i>	Fecha de la visita médica.	77	120	7
<i>u2</i>	Motivo de no realización del informe (1 si no se deja visitar; 2 paradero desconocido; 3 compañía no realizó; 4 desconocimiento existencia lesionado).	0	197	7
<i>u3.1a</i>	Código de la primera secuela.	77	120	7
<i>u3.1b</i>	Puntuación de la primera secuela.	77	120	7
<i>u3.2a</i>	Código de la segunda secuela.	77	120	7

<i>u3.2b</i>	Puntuación de la segunda secuela.	77	120	7
<i>u3.3a</i>	Código de la tercera secuela.	77	120	7
<i>u3.3b</i>	Puntuación de la tercera secuela.	77	120	7
<i>u3.4a</i>	Código de la cuarta secuela.	77	120	7
<i>u3.4b</i>	Puntuación de la cuarta secuela.	77	120	7
<i>u3.5a</i>	Código de la quinta secuela.	77	120	7
<i>u3.5b</i>	Puntuación de la quinta secuela.	77	120	7
<i>u3.6a</i>	Código de la sexta secuela.	77	120	7
<i>u3.6b</i>	Puntuación de la sexta secuela.	77	120	7
<i>u3.7a</i>	Código de la séptima secuela.	77	120	7
<i>u3.7b</i>	Puntuación de la séptima secuela.	77	120	7
<i>u3.8a</i>	Código de la octava secuela.	77	120	7
<i>u3.8b</i>	Puntuación de la octava secuela.	77	120	7
<i>u3.9a</i>	Código de la novena secuela.	77	120	7
<i>u3.9b</i>	Puntuación de la novena secuela.	77	120	7
<i>u3.10a</i>	Código de la décima secuela.	77	120	7
<i>u3.10b</i>	Puntuación de la décima secuela.	77	120	7
<i>u3</i>	Puntuación total de las secuelas.	77	120	7
<i>u4.1</i>	Número días de baja hospitalarios.	77	120	7
<i>u4.2</i>	Número de días de baja improductivos.	77	120	7
<i>u4.3</i>	Número de días de baja no improductivos.	77	120	7
<i>u5</i>	Incapacidad permanente (1 si ninguna; 2 parcial; 3 total; 4 absoluta).	77	120	7
<b><i>V. Informe forense</i></b>				
<i>f0</i>	1 si emisión del informe forense; 0 en caso contrario.	0	204	0
<i>f1</i>	Fecha de la realización del informe forense.	2	176	26
<i>f2</i>	Puntuación de las secuelas el médico forense (1 si puntúa; 2 marca horquilla; 3 no concede secuelas; 4 no hay informe).	0	178	26
<i>f3</i>	Médico forense emisor (1 si clínica médico forense, 2 forense del juzgado instructor; 3 forense exhortado; 4 médico de cabecera).	0	178	26
<i>f4.1a</i>	Código de la primera secuela.	0	178	26
<i>f4.1b</i>	Puntuación de la primera secuela.	0	112	92
<i>f4.2a</i>	Código de la segunda secuela.	0	178	26
<i>f4.2b</i>	Puntuación de la segunda secuela.	0	112	92
<i>f4.3a</i>	Código de la tercera secuela.	0	178	26
<i>f4.3b</i>	Puntuación de la tercera secuela.	0	112	92
<i>f4.4a</i>	Código de la cuarta secuela.	0	178	26
<i>f4.4b</i>	Puntuación de la cuarta secuela.	0	112	92
<i>f4.5a</i>	Código de la quinta secuela.	0	178	26
<i>f4.5b</i>	Puntuación de la quinta secuela.	0	112	92
<i>f4.6a</i>	Código de la sexta secuela.	0	178	26

<i>f4.6b</i>	Puntuación de la sexta secuela.	0	112	92
<i>f4.7a</i>	Código de la séptima secuela.	0	178	26
<i>f4.7b</i>	Puntuación de la séptima secuela.	0	112	92
<i>f4.8a</i>	Código de la octava secuela.	0	178	26
<i>f4.8b</i>	Puntuación de la octava secuela.	0	112	92
<i>f4.9a</i>	Código de la novena secuela.	0	178	26
<i>f4.9b</i>	Puntuación de la novena secuela.	0	112	92
<i>f4.10a</i>	Código de la décima secuela.	0	178	26
<i>f4.10b</i>	Puntuación de la décima secuela.	0	112	92
<i>f4.11</i>	Puntuación total de las secuelas.	0	112	92
<i>f5.1</i>	Número de días de baja hospitalarios.	0	178	26
<i>f5.2</i>	Número de días de baja improductivos.	0	178	26
<i>f5.3</i>	Número de días de baja no improductivos.	0	178	26
<i>f6</i>	Incapacidad permanente (1 si ninguna; 2 si parcial; 3 si total; 4 si absoluta).	0	178	26
<b>VI. Sentencia</b>				
<i>s0</i>	Código juzgador.			
<i>s1a</i>	Tipo de Juzgado (1 si 1ª Instancia; 2 si 1ª Instancia e Instrucción; 3 si Penal).	0	204	0
<i>s1b</i>	Número del Juzgado.	0	204	0
<i>s1c</i>	Población del Juzgado.	0	204	0
<i>s2</i>	Fecha de la sentencia.	0	204	0
<i>s3</i>	Tipo de sentencia (1 si civil; 2 penal; 3 ejecutivo).	0	204	0
<i>s4</i>	1 si hay costas procesales; 0 en caso contrario.	0	204	0
<i>s4a</i>	Carga de las costas (1 si al asegurador; 2 al contrario; 3 compartidas).	0	46	158
<i>s7</i>	Factor de reducción por concurrencia de culpa en accidente (%)	0	204	0
<i>s8</i>	Aplicación del baremo (1 si año del siniestro; 2 año de sentencia; 3 mismo año siniestro y sentencia; 4 otro criterio).	0	204	0
<i>s9</i>	Puntuación de las secuelas.	1	203	0
<i>s10</i>	Valoración de las secuelas.	1	203	0
<i>s11.1a</i>	Número de días de baja hospitalarios.	0	204	0
<i>s11.1b</i>	Valoración de los días de baja hospitalarios.	0	204	0
<i>s11.2a</i>	Número de días de baja improductivos.	0	204	0
<i>s11.2b</i>	Valoración de los días de baja improductivos.	1	203	0
<i>s11.3a</i>	Número de días de baja no improductivos.	0	204	0
<i>s11.3b</i>	Valoración de los días de baja no improductivos.	0	204	0
<i>s12</i>	Factor de corrección por perjuicios económicos.	0	204	0
<i>s12a</i>	1 si afecta el factor de corrección a los días de baja; 0 en caso contrario.	0	109	95
<i>s13</i>	Daños morales complementarios (descripción).	0	204	0
<i>s14.a</i>	Incapacidad permanente (1 si ninguna; 2 parcial; 3 total; 4	0	204	0

	absoluta).			
<i>s14.b</i>	Valoración económica de la incapacidad permanente.	0	3	201
<i>s15a</i>	1 si la víctima fallece; 0 en caso contrario.	0	204	0
<i>s15b</i>	Valoración económica de la indemnización por muerte.	0	7	197
<i>s15c</i>	Factor corrector aplicado a la indemnización por muerte.	0	7	197
<i>s16</i>	Gastos médicos u otros gastos justificados.	0	204	0
<i>s16a</i>	Gastos materiales del vehículo.	0	204	0
<i>s17</i>	Total de la indemnización económica en sentencia.	0	204	0
<i>s18</i>	Interés a aplicar (1 sin interés; si legal + 50%; 3 si 20% a partir del 2º año; 4 si 20% desde fecha siniestro).	0	204	0
<b>VII. Apelación</b>				
<i>a0</i>	Fecha de la nueva sentencia.	0	59	145
<i>a1</i>	Parte apelante (1 si asegurador; 2 contrario; 3 ambos).	0	59	145
<i>a2</i>	1 si hay costas procesales; 0 en caso contrario.			
<i>a2.1</i>	Carga de las costas (1 si al asegurador; 2 al contrario; 3 compartida).			
<i>a3</i>	Factor de reducción por concurrencia de culpa.	0	59	145
<i>a4</i>	Aplicación del baremo (1 si año del siniestro; 2 año de sentencia; 3 mismo año siniestro y sentencia; 4 otro criterio).	0	59	145
<i>a5</i>	Puntuación de las secuelas.	0	59	145
<i>a5.1</i>	Valoración de las secuelas.	0	59	145
<i>a6</i>	Número de días de baja hospitalarios.	0	59	145
<i>a6.1</i>	Valoración de días de baja hospitalarios.	0	59	145
<i>a7</i>	Número de días de baja improductivos.	0	59	145
<i>a7.1</i>	Valoración de días de baja improductivos.	0	59	145
<i>a8</i>	Número de días de baja no improductivos.	0	59	145
<i>a8.1</i>	Valoración de días de baja no improductivos.	0	59	145
<i>a9</i>	Factor de corrección por perjuicios económicos.	0	59	145
<i>a9.1</i>	1 si afecta el factor de corrección a los días de baja; 0 en caso contrario.	0	59	145
<i>a10</i>	Incapacidad permanente (1 si ninguna; 2 parcial; 3 total; 4 absoluta).	0	59	145
<i>a10.1</i>	Valoración de la incapacidad permanente.	0	59	145
<i>a11</i>	Costes materiales.	0	59	145
<i>a12</i>	Interés a aplicar (1 sin interés; si legal + 50%; 3 si 20% a partir del 2º año; 4 si 20% desde fecha siniestro).	0	59	145
<i>a13</i>	Total indemnización de la nueva sentencia.	0	59	145
<i>y0</i>	Fecha de cierre del expediente.	0	204	0
<i>y</i>	Coste total de la indemnización (Sentencia o Apelación).	0	204	0

## Apéndice D. Cuantía central generalizada

Consideremos un vector aleatorio observable  $\mathbf{X}$  con función de distribución  $F(\mathbf{x}, \mathbf{v})$ , donde  $\mathbf{v}=(\delta, \zeta)$  es el vector de parámetros desconocidos, siendo  $\delta$  el parámetro de interés y  $\zeta$  un vector de parámetros molestos.

Dada una cuantía central generalizada  $R$ , si fijamos un nivel de confianza determinado  $(1-\omega^*)$ , aunque  $\mathbf{v}$  es desconocido, podemos calcular el subespacio del espacio muestral de  $R$ ,  $\rho_{\omega^*} \subset \rho$ , como aquél que satisface la siguiente ecuación (Propiedad A),

$$\Pr[R \in \rho_{\omega^*}] = 1 - \omega^*, \quad \forall \omega^* \in (0, 1). \quad (\text{D.1})$$

Finalmente, puesto que  $R_{obs}$  no depende de  $\zeta$  (Propiedad B), obtenemos el intervalo de confianza generalizado de  $\delta$  con nivel de confianza  $(1-\omega^*)$  como el subconjunto del espacio paramétrico  $\Omega_{\omega^*}(R_{obs}) \subset \Omega$  definido con la siguiente expresión,

$$\Omega_{\omega^*}(R_{obs}) = \{ \delta \in \Omega \mid r(\mathbf{x}, \mathbf{x}, \mathbf{v}) \in \rho_{\omega^*} \}. \quad (\text{D.2})$$

El lector interesado encontrará la demostración en Weerahandi (1995, 2004).

## Apéndice E. Intervalo de confianza generalizado basado en el p-valor

Como ocurre en la inferencia clásica, los intervalos de confianza también pueden ser contruidos a partir del contraste de hipótesis. Tomemos de nuevo el vector aleatorio observable  $\mathbf{X}$  con función de distribución  $F(\mathbf{x}, \mathbf{v})$ , siendo  $\mathbf{v}=(\delta, \zeta)$  el vector de parámetros desconocidos, y planteemos el siguiente contraste de hipótesis,

$$H_0 = \delta \leq \delta_0, \quad H_1 = \delta > \delta_0 \quad (\text{E.3})$$

donde  $\delta$  es el parámetro de interés y  $\zeta$  un vector de parámetros molestos. En la teoría convencional, la idea del contraste de significación se basa en definir un subconjunto insesgado del espacio muestral  $\chi$  con los datos observados  $\mathbf{x}$  en el límite (región extrema) y computar la probabilidad de este subconjunto bajo la hipótesis nula ( $p$ -valor). Para la obtención de la región extrema, primero se debe ordenar  $\chi$  mediante el estadístico de contraste.

Desafortunadamente, debido a los parámetros molestos  $\zeta$ , normalmente no se podrá obtener una región extrema con probabilidad independiente de  $\zeta$  mediante la utilización de un estadístico de contraste convencional. Para lograr dicha independencia, Tsui y Weerahandi (1989) amplían la definición de estadístico de contraste, convirtiéndolo en función dependiente de  $\mathbf{X}$ , pero también de  $\mathbf{x}$  y  $\mathbf{v}$ ,  $T=T(\mathbf{X},\mathbf{x},\mathbf{v})$ . Los autores denominan a  $T(\mathbf{X},\mathbf{x},\mathbf{v})$  variable de contraste generalizada y, evidentemente, debe cumplir ciertas propiedades, concretamente:

1.  $t_{obs}=T(\mathbf{x},\mathbf{x},\mathbf{v})$  no depende de parámetros desconocidos.
2. Si fijamos  $\delta$ , la distribución de  $T$  no depende de  $\zeta$ .
3. Si fijamos  $\mathbf{x}$  y  $\zeta$ ,  $\Pr[T \geq t|\delta]$  es función no-decreciente de  $\delta$ , por tanto,  $T$  es estocásticamente creciente en  $\delta$ .

Nótese que la definición ofrecida para la Cuantía Central Generalizada es muy similar a la de  $T$ , incluyendo para esta última una propiedad adicional (concretamente, la tercera), la cual garantiza la ordenación de  $\chi$ . De este modo, Weerahandi (1995) define el  $p$ -valor generalizado de la siguiente forma

$$p = \Pr[T \geq t_{obs} | \delta = \delta_0], \quad (\text{E.4})$$

y a la función potencia generalizada como,

$$\pi(\mathbf{x}, \delta) = \Pr[T \geq t_{obs} | \delta]. \quad (\text{E.5})$$

Si  $T$  es una variable aleatoria continua, se puede construir el intervalo de confianza generalizado a partir de la función potencia generalizada. Fijando  $\mathbf{x}$ , se cumple que:

$$\Pr[\gamma_1 \leq \pi(\mathbf{X}, \mathbf{x}, \delta) \leq \gamma_2] = \gamma_2 - \gamma_1, \quad (\text{E.6})$$

para  $\gamma_1, \gamma_2 \in [0, 1]$ . Por tanto,  $\pi(\mathbf{X}, \mathbf{x}, \delta)$  es una Cuantía Central Generalizada con distribución uniforme. Nótese que la ecuación (E.6) se cumple condicionalmente, fijado  $\mathbf{x}$ , por lo tanto, la propiedad de muestreo repetido de los intervalos de confianza también se cumplirá únicamente de forma condicional.



## Referencias bibliográficas

Abdel-Aty, M. (2003). "Analysis of driver injury severity levels at multiple locations using ordered probit models." *Journal of Safety Research*, **34**(5), 597-603.

Abdel-Aty, M., Abdelwahab, H. (2004). "Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison." *Journal of Transportation Engineering*, **130**(2), 204-210.

Abdel-Aty, M., Chen, C., Schott, J. (1998). "An assessment of the effect of driver age on traffic accident involvement using log-linear models." *Accident Analysis & Prevention*, **30**(6), 851-861.

Abdel-Aty, M., Keller, J. (2005). "Exploring the overall and specific crash severity levels at signalized intersections." *Accident analysis & Prevention*, **37**(3), 417-425.

Actualidad Aseguradora (2006). *En su justo valor*. Actualidad Aseguradora, 10 de Abril, 32.

Adams, M., Burton, M., Hardwick, P. (2003). "The Determinants of Credit Ratings in the United Kingdom Insurance Industry." *Journal of Business Finance & Accounting*, **30**(3-4), 539-572.

Al-Ghamdi, A. (2002). "Using logistic regression to estimate the influence of accident factors on accident severity." *Accident Analysis & Prevention*, **34**(6), 729-741.

- Al-Ghamdi, A. (2003). "Analysis of traffic accidents at urban intersections in Riyadh." *Accident Analysis & Prevention*, **35**(5), 717-724.
- Antonio, K., Beirlant, J. (2005). "Applications of Generalized Linear Mixed Models in Actuarial Statistics." *Insurance: Mathematics and Economics*, en prensa.
- Antonio, K., Beirlant, J., Hoedemakers, T., Verlaak, R. (2006). "Lognormal mixed models for reported claim reserves." *North American Actuarial Journal*, **10**(1), 30-48.
- Aragón, M. (1997). "Dos sorprendentes hallazgos: la ilegalidad de la ley y la soberanía del juez: Comentario a la sentencia 280/1997 de 26 de marzo, de la Sala de lo Civil del Tribunal Supremo." *La Ley*, **3**.
- Arendacká, B. (2005). "Generalized confidence intervals on the variance component in mixed linear models with two variance components." *Statistics*, **39**(4), 275-286.
- Artís, M., Ayuso, M., Guillén, M. (1999). "Modelling different types of automobile insurance fraud behaviour in the Spanish market." *Insurance: Mathematics and Economics*, **24**(1-2), 67-81.
- Artís, M., Ayuso, M., Guillén, M. (2002). "Detection of automobile insurance fraud with discrete choice models and misclassified claims." *Journal of Risk and Insurance*, **69**(3), 325-340.
- Austin, R., Faigin, B. (2003). "Effect of vehicle and crash factors on older occupants." *Journal of Safety Research*, **34**(4), 441-452.
- Ayuso, M. (1998). *Modelos econométricos para la detección del fraude en el seguro del automóvil*. Tesis doctoral, Universitat de Barcelona.
- Ayuso, M., Santolino, M. (2006a). "Predicting automobile claims bodily injury severity with sequential Ordered Logit Models." *Insurance: Mathematics and Economics*, doi:10.1016/j.insmatheco.2006.09.008.

- Ayuso, M., Santolino, M. (2006b). "La valoración económica del daño corporal en Europa." *Actualidad Aseguradora (Suplemento)*, **11**, 67-70.
- Baltagi, B.H., Song, S.H., Jung, B.C. (2002). "A comparative study of alternative estimators for the unbalanced two-way error component regression model." *The Econometrics Journal*, **5**(2), 480-493.
- Barrese, J., Doeringhaus, H., Nelson, J. (1995). "Do Independent Agent Insurers Provide Superior Service? The Insurance Marketing Puzzle." *The Journal of Risk and Insurance*, **62**(2), 297-308.
- Belhadji, E.B., Dionne, G., Tarkhani, F. (2000). "A Model for the Detection of Insurance Fraud." *The Geneva Papers on Risk and Insurance-Issues and Practice*, **25**(4), 517-538.
- Bell, P. (2006). "The Social Construction of Bodily Injury." *Geneva papers on Risk and Insurance-Issues and Practice*, **31**, 340-356.
- Boison, L.R., Lehman, S.G. (2004). "New Jersey Automobile No-Fault Study: Analysis of the Cost Effects of AB 3531 and SB 2533." *Journal of Insurance Regulation*, **23**(1), 43-62.
- Bond, E.W., Crocker, K.J. (1997). "Hardball and the soft touch: The economics of optimal insurance contracts with costly state verification and endogenous monitoring costs." *Journal of Public Economics*, **63**(2), 239-264.
- Borooah, V.K. (2002). *Logit and Probit: Ordered and Multinomial Models*. Thousand Oaks, California: Sage.
- Brockett, P., Derrig, R.A., Golden, L., Levine, A., Alpert, M. (2002). "Fraud Classification Using Principal Component Analysis of RIDITs." *Journal of Risk and Insurance*, **69**(3), 341-371.

- Brockett, P., Xia, X., Derrig, R.A. (1998). "Using Kohonen's Self-Organising Feature Map to Uncover Automobile Bodily Injury Claims Fraud." *Journal of Risk and Insurance*, **65**(2), 245-274.
- Brookes, R., Prevet, M. (2004). "Statistical Case Estimation Modelling: An Overview of the NSW WorkCover Model." *Institute Actuaries of Australia's Accident Compensation Seminar*, 28 November to 1 December.
- Browne, M.J., Puelz, R. (1996). "Statutory Rules, Attorney Involvement, and Automobile Liability Claims." *Journal of Risk and Insurance*, **63**(1), 77-94.
- Browne, M.J., Puelz, R. (1999). "The effect of Legal Rules on the Value of Economic and Non-Economic Damages and the Decision to File." *Journal of Risk and Uncertainty*, **18**(2), 189-213.
- Browne, M.J., Schmit, J.T. (2006). "Litigation Patterns in Automobile Bodily Injury Claims 1977-1997: Effects of Time and Tort Reforms." *Risk Management & Insurance Seminar Series: 2005/2006*, March 10, Terry College of Business.
- Browne, M.J., Wells, B.P. (1999). "Claims Adjudication in the Personal Automobile Insurance Residual Market." *Journal of Risk and Insurance*, **66**(2), 275-290.
- Burdick, R., Graybill, F. (1992). *Confidence Intervals on Variance Components*. Marcel Dekker, New York.
- Carson, J., Mannering, F. (2001). "The effect of ice warning signs on ice-accident frequencies and severities." *Accident Analysis & Prevention*, **33**(1), 99-109.
- Caudill, S., Ayuso, M. & Guillen, M. (2005). "Fraud detection using a multinomial logit model with missing information." *Journal of Risk and Insurance*, **72**(4), 539-550.
- Chang, L.Y., Wang, H.W. (2006). "Analysis of traffic injury severity: An application of non-parametric classification tree techniques." *Accident Analysis & Prevention*, **38**(5), 1019-1027.

Clogg, C.C., Shihadeh, E.S. (1994). *Statistical Models for Ordinal Variables*. Thousand Oaks, California: Sage.

Cox, D.R., Snell, E.J. (1989). *Analysis of Binary Data*. (2nd ed.), Chapman & Hall, London.

Crocker, K.J., Tennyson, S. (1999). "Costly State Falsification or Verification? Theory and Evidence from Bodily Injury Liability Claims." En: Dionne, G., Laberge-Nadeau, C. (eds.), *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, Boston: Kluwer.

Crocker, K.J., Tennyson, S. (2002). "Insurance Fraud and Optimal Claims Settlement Strategies." *Journal of Law & Economics*, 45(2), 469-507.

Cummins, J.D., Phillips, R.D., Weiss, M.A. (2001). "The Incentive Effects of No Fault Automobile Insurance." *Journal of Law & Economics*, 44(2), 427-464.

Cummins, J.D., Tennyson, S. (1996). "Moral Hazard in Insurance Claiming: Evidence from Automobile Insurance." *The Journal of Risk and Insurance*, 12, 29-50.

D'Arcy, S. (2005). "Predictive Modeling in Automobile Insurance: A Preliminary Analysis." En: *World Risk and Insurance Economics Congress*, August, Salt Lake City.

Derrig, R.A. (2002). "Insurance Fraud." *The Journal of Risk and Insurance*, 69(3), 271-287.

Derrig, R.A. (2004). "The Settlement Negotiation Process for Automobile Bodily Injury Liability Claims In the Presence of Suspicion of Fraud and Build-up." En: *Proceedings of Risk Theory Society*.

Derrig, R.A., Rempala, G.A. (2006). "A Statistical Analysis of the Effect of Anchoring in the Negotiation Process of Automobile Bodily Injury Liability Insurance Claims." En: *CAS predictive modelling seminar*, October 4, Boston.

Derrig, R.A., Weisberg, H.I. (2004). "Determinants of Total Compensation for Auto Bodily Injury Liability Under No-Fault: Investigation, Negotiation and the Suspicion of Fraud." *Insurance and Risk Management*, **71**(4), 633-662.

DFID (2003). *Guidelines for Estimating the Cost of Road Crashes in Developed Countries*. Department for International Development, United Kingdom.

DGT (2007). *Estadísticas definitivas (anuales)*. Disponible en internet: <http://www.dgt.es/estadisticas/estadisticas03.htm> (consultado el 2 de febrero de 2007).

Doerpinghaus, H. (1991). "An Analysis of Complaint Data in the Automobile Insurance Industry." *Journal of Risk and Insurance*, **58**, 120-127.

Doerpinghaus, H., Schmit, J., Yeh, J.J.H. (2003). "Personal Bias in Automobile Claims Settlement." *Journal of Risk and Insurance*, **70**(2), 185-205.

El-Bassiouni, M.Y. (1994). "Short Confidence Intervals for Variance Components." *Communications in Statistics-Theory and Methods*, **23**(7), 1915-1933.

Eluru, N., Bhat, C. (2006). "A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity." En: *TRB 85th Annual Meeting*, January 22-26, Washington DC.

England, P., Verall, R. (2002). "Stochastic claims reserving in general insurance." *British Actuarial Journal*, **8**(3), 443-544.

ETSC (1997). *Transport Accident Costs and the Value of Safety*. European Transport Safety Council, Brussels.

Fagart, M., Picard, P. (1999). "Optimal Insurance under Random Auditing." *Geneva Papers on Risk and Insurance Theory*, **24**(1), 29-54.

Fernández, J. (1999). *Valoración judicial de daños y perjuicios*. Madrid: Escuela Judicial-CGPJ (ed.).

Fernández, J. (2001). *Después de la sentencia 181/2000 del Pleno del Tribunal Constitucional*. Ponencia presentada en el Congreso Constituyente de la Asociación Española de Abogados Especializados en Responsabilidad Civil y Seguro, Granada: 15 de noviembre.

Fernández, M.J. (2006). "La Quinta Directiva *piano ma non troppo*." *Actualidad Aseguradora*, **13**, 28-29.

Frees, E.W., Young, V.R., Luo, Y. (1999). "A longitudinal data analysis interpretation of credibility models." *Insurance: Mathematics and Economics*, **24**(3), 229-247.

Frees, E.W., Young, V.R., Luo, Y. (2001). "Case studies using panel data models." *North American Actuarial Journal*, **5**(4), 24-42.

Galcerá, J.M. (2004). *La Responsabilidad Civil del Automóvil y el Daño Corporal*. Tesis del Master en Dirección de Entidades Aseguradoras y Financieras. Universidad de Barcelona.

García, C. (2002). "La anunciada inconstitucionalidad del nuevo sistema de valoración de daños personales derivados de los accidentes de circulación." *Revista internauta de práctica jurídica*, **10**.

Garrido, R., Lopez, J.M. (2005). "Accidentes de tráfico. Una pandemia del presente." *Economía de la salud*, **4**(1), 34-41.

Goldberger, A.S. (1962). "Best linear Unbiased Prediction in the Generalized Linear Regression Model." *Journal of the American Statistical Association*, **57**(298), 369-375.

Gómez, C., Muntaner, F.A. (2004). "¿Quién da más?" *Indret*, **256**.

Gossner, O., Picard, P. (2005). "On the Consequences of Behavioral Adaptations in the Cost-Benefit analysis of Road Safety Measures." *Journal of Risk and Insurance*, **72**(4), 577-599.

- Graham, L. (2005). "Ex Post Moral Hazard and Auto Bodily Injury Liability Insurance." *Rosen-Huebner-McCahan Seminar Series*, The Wharton School.
- Graybill, F.A., Hultquist, R.A. (1961). "Theorems Concerning Eisenhart's model II." *Annals of Mathematical Statistics*, **32**, 261-269.
- Greene, W.H. (1998). *Análisis Económico*, 3ª ed., Prentice Hall Iberia, Madrid.
- Griffiths, W., Judge, G. (1992). "Testing and Estimating Location Vectors When the Error Covariance Matrix is Unknown." *Journal of Econometrics*, **54**(1-3), 121-138.
- Haastrup, S., Arjas, E. (1996). "Claims reserving in continuous time: a non-parametric Bayesian approach." *ASTIN Bulletin*, **26**(2), 139-164.
- Harville, D. (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of American Statistical Association*, **72**(358), 320-338.
- Harville, D. (1989). "BLUP (Best Linear Unbiased Prediction) and Beyond." In: Gianola, D. and Hammond, K. (eds.) *Advances in Statistical Methods for Genetic Improvement of Livestock*, New York: Springer-Verlag, 239-276.
- Harville, D., Carriquiry, A. (1992). "Classical and Bayesian Prediction as Applied to an Unbalanced Mixed Linear Model." *Biometrics*, **48**(4), 987-1003.
- Harville, D., Fenech, A. (1985). "Confidence Intervals for a Variance Ratio, or for Heritability, in an Unbalanced Mixed Linear Model." *Biometrics*, **41**(1), 137-152.
- Harville, D., Jeske, D. (1992). "Mean Squared Error of Estimation or Prediction Under a General Linear Model." *Journal of the American Statistical Association*, **87**(419), 724-731.
- Henderson, C.R. (1950). "Estimation of genetic parameters." *Annals of Mathematical Statistics*, **21**(2), 309-310.



Henderson, C.R. (1963). "Selection index and expected genetic advance." En: Hanson, W.D., Robinson, H.F. (eds.), *Statistical Genetics and Plant Breeding*. Washington, DC: Publication 982, National Academy of Sciences, 141-163.

Henderson, C.R. (1975). "Best Linear Unbiased Estimation and Prediction Under a Selection Model." *Biometrics*, **31**, 423-447

Henderson, C.R., Kempthorne, O., Searle, S.R., Krosigk, C.M. (1959). "The estimation of environmental and genetic trends from records subject to culling." *Biometrics*, **15**, 192-218.

Herbers, J.A. (1994) "Choice No-Fault: Actuarial Costing Methods." *Casualty Actuarial Society*, Winter Forum.

Hillman, R., Allen, K. (2003). "Medical Malpractice Insurance: Multiple Factors Have Contributed to Premium Rate Increases." GAO-03-702.

Hosmer, D.W., Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley & Sons.

Hoyt, R., Mustard, D., Powell, L. (2006). "The Effectiveness of State Legislation in Mitigating Moral Hazard: Evidence from Automobile Insurance." *Journal of Law and Economics*, **49**(2), 427-450.

ICEA (2007a). *Siniestralidad por garantías*. Disponible en internet: [http://intranet.icea.es/Almacen/Autos/Informacion\\_aseg/2005/siniestralidad\\_autos\\_054.asp](http://intranet.icea.es/Almacen/Autos/Informacion_aseg/2005/siniestralidad_autos_054.asp) (consultado el 2 de febrero de 2007).

ICEA (2007b). *Parque de vehículos*. Disponible en internet: [http://intranet.icea.es/Almacen/Autos/Vehiculos/2005/parque\\_54.asp](http://intranet.icea.es/Almacen/Autos/Vehiculos/2005/parque_54.asp) (consultado el 2 de febrero de 2007).

Johnson, J.E., Flanigan, G.B., Winkler, D.T. (1992). "Cost Implications of No-Fault Automobile Insurance." *Journal of Risk and Insurance*, **59**(1), 116-123.

Karlaftis, M.G., Kotzampassakis, I., Kanellaidis, G. (2003). "An empirical investigation of European drivers' self-assessment." *Journal of Safety Research*, 34(2), 207-213.

Kenneth, T. (2002). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Khuri, A.I., Mathew, T., Sinha, B. (1998). *Statistical Tests for mixed linear models*. Wiley series in probability and mathematical statistics.

Kockelman, K., Kweon, Y. (2002). "Driver Injury severity: an application of ordered probit models." *Accident Analysis & Prevention*, 34(3), 313-321.

Kubokawa, T., Srivastava, M.S. (2002). "Prediction in Multivariate Mixed Linear Models." CIRJE-F-series, No.180.

Kuhnert, P., Do, K.A., McClure, R. (2000). "Combining non-parametric models with logistic regression: an application to motor vehicle injury data." *Computational Statistics & Data Analysis*, 34(3), 371-386.

Laird, N.M., Ware, J.H. (1982). "Random-effects models for longitudinal data." *Biometrics*, 38(4), 963-974.

Lascher, E.L., Powers, M.R. (2001). "Choice No-Fault Insurance: Efficiency and Equity." En: Lascher, E.L.Jr. & Powers, M.R. (eds.) *The Economics and Politics of Choice No-Fault Insurance*. Boston: Kluwer Academic Publisher.

Law Commission, The. (1999a). *Damages for personal injury: Medical, Nursing and other expenses; Collateral benefits*. Law Commission No.262. Disponible en internet: <http://www.lawcom.gov.uk/docs/lc262.pdf> (consultado el 20 de octubre de 2006)

Law Commission, The. (1999b). *Damages for personal injury: Non-pecuniary loss*. Law Commission No.257. Disponible en internet: <http://www.lawcom.gov.uk/docs/lc257.pdf> (consultado el 20 de octubre de 2006)

- Lee, C., Abdel-Aty, M. (2005). "Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida." *Accident Analysis & Prevention*, **37**(4), 75-786.
- Lehmann, E.L. (1986). *Testing Statistical Hypothesis*. New York: Wiley.
- Lewis, R. (2006). "How Important are Insurers in Compensating Claims for Personal Injury in the U.K.?" *The Geneva Papers on Risk and Insurance-Issues and Practice*, **31**(2), 323-339.
- Lewis, R., McNabb, R., Robinson, H., Wass, V. (2003). "Loss of earnings following personal injury: do the courts adequately compensate injured parties?" *Economic Journal*, **113**(491), F568-F584.
- Liao, T.F. (1994). *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. Thousand Oaks, California: Sage.
- Loughran, D.S. (2005). "Deterring Fraud: The Role of General Damage Awards in Automobile Insurance Settlement." *Journal of Risk and Insurance*, **72**(4), 551-575.
- Marter, S., Weisberg, H. (1991). "Medical Costs and Automobile Insurance: A Report on Bodily Injury Liability Claims in Massachusetts." *Journal of Insurance Regulation*, **91**, 381-422.
- Martín, M. (2001). *¿Hacia un baremo europeo para la indemnización de los daños corporales?* Ponencia presentada en el II Congreso de Responsabilidad Civil y Seguro, Córdoba, 3-4 de mayo. Disponible en internet.  
<http://civil.udg.es/cordoba/pon/martin.htm> (consultado el 20 de octubre de 2006).
- Mathew, T., Sinha, B. (1988). "Optimum Tests for Fixed Effects and Variance Components in Balanced Models." *Journal of the American Statistical Association*, **83**(401), 133-135.
- Mathew, T., Sinha, B. (1992). "Exact and Optimum Tests in Unbalanced Split-Plot Designs Under Mixed and Random Models." *Journal of the American Statistical Association*, **87**(417), 192-200.

McCullagh, P. (1980). "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society Series B*, 42(2), 109-142.

McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models*. Second Edition. London: Chapman & Hall.

McKelvey, R., Zavoyna, W. (1975). "A statistical model for the analysis of ordinal level dependent variables." *Journal of Mathematical Sociology*, 4, 103-120.

McIntosch, D., Holmes, M. (ed.) (2003). *Personal Injury Awards in EU and EFTA Countries*. The Hague: Kluwer Law International.

Michalski, A. (2003). "On some aspects of the optimal statistical inference on variance components in mixed linear models." *Tatra Mountains Mathematical Publications*, 26(1), 133-153.

Moreno, L. (2001). *Un baremo europeo de valoración del daño corporal*. Comunicación presentada en el II Congreso de Responsabilidad Civil y Seguro, Córdoba. 3-4 de mayo. Disponible en internet: <http://civil.udg.es/cordoba/com/Moreno2.htm> (consultado el 20 de octubre de 2006).

Nagelkerke, N.J.D. (1991). "A note on a general definition of the coefficient of determination." *Biometrika*, 78(3), 691-692.

Nayga, R.M., Poghosyan, A., Nichols, J.P. (2004). "Will consumers accept irradiated food products?" *International Journal of Consumer Studies*, 28(2), 178-185.

Norberg, R. (1993). "Prediction of Outstanding Liabilities in Non-Life Insurance." *ASTIN Bulletin*, 23(1), 95-115.

Ntzoufras, I., Dellaportas, P. (2002). "Bayesian prediction of outstanding claims." *North American Actuarial Journal*, 6(1), 113-136.

O'Donnell, C.J., Connor, D.H. (1996). "Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice." *Accident analysis & Prevention*, **28**(6), 739-756.

OECD-IRTAD (1994). *Under-reporting of road traffic accidents recorded by the police at the international level*. Special Report, Oslo.

Öfversten, J. (1993). "Exact Tests for Variance Components in Unbalanced Mixed Linear Models." *Biometrics*, **49**, 45-57.

Olsen, A., Seely, J., Birkes, D. (1976). "Invariant Quadratic Unbiased Estimation for Two Variance Components." *The Annals of Statistics*, **4**(5), 878-890.

OMS (2004). *Informe mundial sobre prevención de los traumatismos causados por el tránsito*. OMS, Ginebra.

Ostrom, M., Eriksson, A. (2001). "Pedestrian fatalities and alcohol." *Accident Analysis & Prevention*, **33**(2), 173-180.

Palanca, S., Taylor, D., Bayley, M., Cameron, P. (2003). "Mechanisms of motor vehicle accidents that predict major injury." *Emergency Medicine*, **15**(5-6), 423-428.

Park, D.J., Burdick, R.K. (2003). "Performance of Confidence Intervals in Regression Models with Unbalanced One-Fold Nested Error Structures." *Communications in Statistics-Simulation and Computation*, **32**(3), 717-732.

Park, D.J., Burdick, R.K. (2004). "Confidence Intervals on Total Variance in a Regression Model with an Unbalanced Onefold Nested Error Structure." *Communications in Statistics-Theory and Methods*, **33**(11), 2735-2743.

Patterson, H.D., Thompson, R. (1971). "Recovery of inter-block information when block sizes are unequal." *Biometrika*, **58**(3), 545-554.

Perez-Marín, A. (2006). *Survival methods for the analysis of customer lifetime duration in insurance*. Tesis doctoral, Universitat de Barcelona.

Pintos-Ager, J. (2000). *Baremos, Seguros y Derecho de daños*. Editorial Civitas, Madrid.

Pintos-Ager, J. (2003). "Efectos de la baremización del daño sobre la litigiosidad." *InDret*, **131**. Disponible en internet: [http://www.indret.com/rcs\\_articulos/cas/131.pdf](http://www.indret.com/rcs_articulos/cas/131.pdf) (consultado el 20 de octubre de 2006).

Prasad, N.G.N., Rao, J.N.K. (1990). "The Estimation of the Mean Squared Errors of Small-Area Estimators." *Journal of American Statistical Association*, **85**(405), 163-171.

Pratt, J.W. (1981). "Concavity of the Log Likelihood." *Journal of the American Statistical Association*, **76**(373), 103-106.

RACE (2003). *La seguridad del transporte en España*. RACE.

Rao, C.R. (1971). "Estimation of variance and covariance components - MINQUE theory." *Journal of Multivariate Analysis*, **1**, 257-275.

Rao, P. (1997). *Variance components estimation. Mixed models, methodologies and applications*. London: Chapman Hall.

Rempala, G., Derrig, R. (2004). "Modeling Hidden Exposures in Claim Severity via the EM Algorithm." En: *Astin Colloquium*, 6-9 June, Bergen.

Robinson, D.L. (1987). "Estimation and use of variance components." *The Statistician*, **36**, 3-14.

Robinson, G.K. (1991). "That BLUP Is a Good Thing: The Estimation of Random Effects." *Statistical Science*, **6**(1), 15-51.

Rogers, W.V.H. (ed.) (2001). *Damages for non-pecuniary loss in a comparative perspective*. New York: Springer.

Rothley, W. (2003). *Draft report with recommendations to the Commission on a European disability rating scale*. Committee on Legal Affairs and the Internal Market, 2003/2130 (INI).

SafetyNet (2005). *Annual Statistical Report 2005*. European Road Safety Observatory.

Sandra-Vives, T., Gómez-Alvado, F., Folch, E. (2006). *La situación del Mercado de Seguros del Automóvil*. Seminario del Col·legi d'Actuaris de Catalunya, 20 de noviembre, Barcelona.

Santolino, M., Ayuso, M., (2007). "Forecasting the maximal limit of negotiation for bodily injury claims in automobile insurance." Documento de trabajo del Institut de Recerca en Economia Aplicada de la Universitat de Barcelona (IREA).

SAS Institute Inc. (2003). *SAS 9.1.3: Help and Documentation*. Cary, NC: SAS Institute Inc.

Schmit, J.T., Browne, M.J., Lee, H. (1997). "The effect of State Tort Reforms on Claim Filings." *Risk Management and Insurance Review*, 1(1), 1-17.

Schmit, J.T., Yeh, J.H. (2003). "An Economic Analysis of Auto Compensation Systems: Choice Experiences From New Jersey and Pennsylvania." *The Journal of Risk and Insurance*, 70(4), 601-628.

SCOR Global P&C (2006). *Nivel y Evolución del Coste Medio Daño Corporal Grave por Accidentes de Circulación Ocurridos en España*. SCOR.

Scott Long, J. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, California: Sage.

Searle, S., Casella, G., McCulloch, C. (1992). *Variance Components*. Wiley series in probability and mathematical statistics.

Seely, J.F., El-Bassiouni, Y. (1983). "Applying Wald's Variance Component Test." *The Annals of Statistics*, **11**(1), 197-201.

Sohn, S.Y., Shin, H. (2001). "Technical Note Pattern recognition for road traffic accident severity in Korea." *Ergonomics*, **44**(1), 107-117.

Srinivasan, K. (2002). "Injury Severity Analysis with Variable and Correlated Thresholds: Ordered Mixed Logit Formulation." *Transportation Research Record*, **1784**, 132-142.

Stephens, D.A., Crowder, M.J., Dellaportas, P. (2004). "Quantification of automobile insurance liability: a Bayesian failure time approach." *Insurance: Mathematics and Economics*, **34**(1), 1-21.

Sun, Y., Sinha, B.K., von Rosen, D., Meng, Q. (2003). "Nonnegative estimation of Variance Components in multivariate unbalanced mixed linear models with two variance components." *Journal of Statistical Planning and Inference*, **115**(1), 215-234.

Swallow, W.H., Monahan, J.F. (1984). "Monte Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components." *Technometrics*, **26**(1), 47-57.

Swiss Re (2004). "The economics of liability losses-Insuring a moving target." *Sigma*, **6**.

Taylor, G., Capbell, M., (2002). "Statistical case estimation." Research paper number 104, Centre for Actuarial Studies, The University of Melbourne, Australia.

Disponible en internet:

<http://www.economics.unimelb.edu.au/actwww/html/no104.pdf> (consultado el 20 de octubre de 2006).



- Taylor, G., McGuire, G., Greenfield, A., (2003). "Loss reserving: past, present and future." Research paper number 109, Centre for Actuarial Studies, The University of Melbourne, Australia. Disponible en internet: <http://www.economics.unimelb.edu.au/actwww/html/no109.pdf> (consultado el 20 de octubre de 2006).
- Tennyson, S., Salsas-Forn, P. (2002). "Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objective." *The Journal of Risk and Insurance*, **69**(3), 289-308.
- Tsui, K., Weerahandi, S. (1989). "Generalized  $p$ -values in Significance Testing of Hypothesis in the Presence of Nuisance Parameters." *Journal of the American Statistical Association*, **84**(406), 602-607.
- Tuchscherer, A. Herrndörfer, G., Tuchscherer, M. (1998) "Evaluation of the Best Linear Unbiased Prediction in Mixed Linear Models with Estimated Variance Components by Means of the MSE of Prediction and the Genetic Selection Differential." *Biometrical Journal*, **40**(8), 949-962.
- Ulfarsson, G.F., Mannering, F.L. (2004). "Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents." *Accident analysis & Prevention*, **36**(2), 135-147.
- UNESPA (2006). *Memoria Social del Seguro Español 2005*. UNESPA.
- Verbeke, G., Molenberghs, G. (1997). *Linear Mixed Models in Practice: A SAS-Oriented Approach*. New York: Springer-Verlag.
- Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Viane, S., Ayuso, M., Guillen, M., VanGheel, D., Dedene, G. (2007). "Strategies to detect and prevent fraudulent claims in the automobile insurance industry." *European Journal of Operational Research*, **176**, 565-583.

Viane, S., Derrig, R.A., Baesens, B., Dedene, G. (2002). "A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection." *The Journal of Risk and Insurance*, **69**(3), 373-421.

Viscusi, W., Born, P. (2005). "Damages Caps, Insurability, and the Performance of Medical Malpractice Insurance." *Journal of Risk and Insurance*, **72**(1), 23-43.

Voas R.B., Fisher, D.A., Tippetts, S. (2002). "Children in fatal crashes: driver blood alcohol concentration and demographics of child passengers and their drivers." *Addiction*, **97**(11), 1439-1448.

Wald, A. (1940). "A note on the analysis of variance with unequal class frequencies." *Ann. Math. Stat.*, **11**(1), 96-100.

Wald, A. (1947). "A note on regression analysis." *Annals of Mathematical Statistics*, **18**, 586-589.

Wang, X., Kockelman, K. (2005). "Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury: Distinguishing the Effects of Vehicle Weight and Type." *Transportation Research Record*, **1908**, 195-204.

Weerahandi, S. (1993). "Generalized Confidence Intervals." *Journal of the American Statistical Association*, **88**, 899-905.

Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis*. New York: Springer-Verlag.

Weerahandi, S. (2004). *Generalized Inference in Repeated Measures*. New Jersey: John Wiley & Sons.

Weisberg, H., Derrig, R. (1993). "Pricing Auto No-fault and Bodily Injury Liability Coverages Using Micro-Data and Statistical Models." En: *CAS Special Edition Forum, Ratemaking seminar*.

Weisberg, H., Derrig, R. (1995). "Identification and Investigation of Suspicious Claims." AIB Cost Containment/Fraud Filing (DOI Docket R95-12), Automobile Insurers Bureau of Massachusetts, 192-245, Boston.

Weisberg, H., Derrig, R. (1998). "Quantitative Methods for detecting fraudulent automobile bodily injury claims." *Risques*, **35**, 1-21.

Zajac, S., Ivan, J. (2003). "Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut." *Accident analysis & Prevention*, **35**(3), 369-379.

Zhou, L., Mathew, T. (1994). "Some Tests for Variance Components Using Generalized  $p$  Values." *Technometrics*, **36**(4), 394-402.



B.U.B Secció d'Econòmiques  
Diagonal, 690, 08034 Barcelona  
N 102 12 68

