

## Taller 1: Estimación Chain-Ladder

Brayan David Rincón Piñeros<sup>a,c</sup>,

Francisco Albeiro Gomez Jaramillo<sup>b,c</sup>

<sup>a</sup>*Estudiante de Maestría en Actuaría y Finanzas*

<sup>b</sup>*Profesor, Departamento de Matemáticas*

<sup>c</sup>*Universidad Nacional de Colombia, Bogotá, Colombia*

# 1 ENTEDIMIENTO DEL NEGOCIO

## 1.1 Contexto

Históricamente, los seguros han acompañado el desarrollo de la humanidad. Los primeros surgieron en la antigüedad, cuando pequeños grupos de personas se unían para protegerse colectivamente de eventos fortuitos e inesperados. Con el tiempo, los seguros han evolucionado y se han convertido en herramientas que ofrecen protección frente a una amplia variedad de riesgos en distintos contextos.

Este avance ha sido impulsado por el desarrollo de la estadística y las matemáticas, lo que permite que los seguros sean objeto de estudios rigurosos. Esta rigurosidad confiere a los seguros un carácter formal y sólido, aspectos especialmente importantes dada su relevancia global. Este nivel de análisis riguroso posibilita que las compañías aseguradoras mantengan la solidez financiera y dispongan de los recursos necesarios para cubrir pérdidas o siniestros, aspectos que tienen un impacto directo en sus asegurados.

Con este contexto en mente, el presente proyecto busca estudiar y proponer una metodología alternativa a la técnica Chain-Ladder. El objetivo de este enfoque es estimar los flujos futuros de dinero necesarios para cubrir pérdidas que han ocurrido, pero aún no se han reportado a la compañía aseguradora.

### 1.1.1 Sector asegurador

Es importante entender que el sector asegurador está compuesto por diferentes entidades, entre las cuales se encuentran las compañías aseguradoras, los reguladores y otros organismos que participan de forma directa o indirecta en el mercado asegurador, como por ejemplo los reaseguradores. Para cada una de estas entidades, es crucial contar con solidez financiera para cubrir futuras pérdidas. Además, es fundamental conocer la solidez financiera de los otros actores involucrados, ya que la falta de solidez en alguno de ellos podría tener un efecto negativo en toda la red financiera del sector asegurador.

Por otro lado, es importante considerar la estructura interna de las compañías aseguradoras, la cual se organiza en lo que se conoce como 'ramos'. Un ramo es un conjunto de seguros que agrupa riesgos de características similares. Algunos ejemplos de ramos son: ramo de vida, ramo de automóviles, ramo de terremoto o ramo de responsabilidad civil.

En particular este proyecto se concentra en estudiar y analizar las pérdidas del ramo de responsabilidad civil (Liability).

### **1.1.2 Responsabilidad civil**

Los seguros de responsabilidad civil permiten a una persona, conocida como tomador, trasladar a otra entidad, en este caso la aseguradora, el riesgo de ser considerada responsable civilmente por causar daños a un tercero. Por ejemplo, el dueño de una vivienda sería responsable de los daños causados por objetos que pudieran caer o ser arrojados desde la vivienda al exterior, en caso de que esto ocurra el riesgo es cubierto por la compañía aseguradora.

Desde el enfoque del proyecto, un aspecto importante a considerar en el ramo de responsabilidad civil es expuesto por Nieto y Tamayo (2018), donde se indica: 'En particular, en negocios como vida individual, gastos médicos, responsabilidad civil, etc., la evolución del reporte de los siniestros es estacional'. Teniendo en cuenta esta característica de estacionalidad, es importante considerar aplicar técnicas y modelos que permitan capturar de forma precisa estos patrones estacionales, con el fin de realizar estimaciones más acertadas de las futuras pérdidas.

### **1.1.3 Metodología Chain-Ladder**

Chain-Ladder es una técnica utilizada con frecuencia en la industria aseguradora. Esta metodología se emplea para prever la cantidad de dinero que una compañía de seguros necesitará reservar para cubrir futuras reclamaciones, basándose en el comportamiento histórico de los datos disponibles relacionados con reclamaciones pasadas. Como descripción general de la técnica, Chain-Ladder toma como referencia los datos asociados a reclamaciones de períodos anteriores. Con base en estos datos, se realiza la estimación de las reclamaciones futuras que podrían surgir. Es importante tener en cuenta que estas reclamaciones futuras pueden provenir de dos fuentes:

- a. Siniestros incurridos, pero no reportados (Incurred But Not Reported, IBNR): En este caso, se intenta estimar las pérdidas asociadas a siniestros que ya han ocurrido pero que aún no han sido reportados a la compañía aseguradora.

- b. Siniestros que aún no se han resuelto: Corresponde a siniestros que ya están bajo el dominio de la compañía aseguradora, pero que, debido a su complejidad, aún se encuentran en estudio y eventualmente se deberá incurrir en pagos futuros.

La esencia de la metodología Chain-Ladder está fundamentada en la idea de que las reclamaciones de un año en particular suelen desarrollarse de forma similar a las de años anteriores.

### Pasos básicos para implementar Chain-Ladder:

1. **Crear un triángulo de pérdidas:** Los datos históricos se organizan en un triángulo en el que cada fila representa un año de origen de las reclamaciones, y cada columna representa un período de tiempo acumulativo desde ese año de origen.

Pagos acumulados de reclamaciones		Año de desarrollo			
		0	1	2	3
Año de accidente	2011	600	680	720	740
	2012	620	695	730	
	2013	680	760		
	2014	720			

2. **Calcular factores de desarrollo:** Para cada par de años consecutivos dentro de una fila, se calcula un factor de desarrollo, que es la razón entre las pérdidas acumuladas en el año más reciente y las pérdidas acumuladas en el año anterior.

$$\text{Desarrollo año 2 al año 3: } \frac{740}{720} = 1.0278$$

$$\text{Desarrollo año 1 al año 2: } \frac{720+730}{680+695} = 1.0545$$

$$\text{Desarrollo año 0 al año 1: } \frac{680+695+760}{600+620+680} = 1.1237$$

3. **Promediar factores de desarrollo:** Se toma un promedio de los factores de desarrollo para cada columna con el fin de estimar cómo se desarrollarán las pérdidas en el futuro.

4. **Proyectar pérdidas futuras:** Utilizando los factores de desarrollo promedio, se proyectan las pérdidas futuras para cada año de origen y se suman para obtener el total de la reserva necesaria.

La metodología Chain-Ladder es popular y está ampliamente adoptada por compañías aseguradoras; sin embargo, tiene ciertas limitaciones. Por ejemplo, asume que los patrones de desarrollo de las pérdidas en el pasado son un buen predictor de los patrones futuros, lo cual no es necesariamente correcto. Además, es sensible a variaciones en los datos, lo que puede afectar las estimaciones.

## **1.2 Objetivos y criterios de éxito del negocio**

### **Objetivo general**

Optimizar la asignación de recursos financieros en la compañía aseguradora mediante la mejora de la precisión en la estimación de factores de reserva para siniestros incurridos, pero no avisados en el ramo de responsabilidad civil.

### **Objetivos específicos**

- i. Identificar las limitaciones y oportunidades de mejora en el proceso actual de estimación de reservas para siniestros incurridos, pero no avisados, de tal forma que permita reducir la incertidumbre y riesgos financieros asociados.
- ii. Incrementar la eficiencia operativa en la gestión de siniestros a través de un sistema de estimación de reservas más preciso.

## **1.3 Objetivos y criterios de éxito de la minería de datos**

### **Objetivo general**

Desarrollar y validar un modelo de machine learning que mejore el error asociado al método Chain-Ladder en la proyección de pagos futuros en el ramo de responsabilidad civil.

### **Objetivos específicos**

- i. Examinar los datos históricos de siniestros para identificar patrones, correlaciones y posibles outliers que puedan afectar la proyección.
- ii. Utilizar técnicas de modelado estadístico y machine learning para crear modelos que proyecten con mayor precisión los pagos futuros de siniestros.

- iii. Validar los modelos construidos mediante validación cruzada y comparar su rendimiento versus Chain Ladder.

## 2 ENTENDIMIENTO DE LOS DATOS

### 2.1 Recolección de los datos

Los [datos](#) asociados a este proyecto se han obtenido desde la web de la CAS (Casualty Actuarial Society), para esto la CAS obtuvo datos a través de la base de datos de la NAIC (National Association of Insurance Commissioners).

Para este proyecto se hace uso del conjunto de datos asociado a Responsabilidad civil de producto (Product Liability Data Set), este conjunto de datos se encuentra disponible en el repositorio del proyecto en github.

Estos datos se encuentran almacenados un archivo .csv y no poseen relación con otra fuente de datos.

### 2.2 Descripción y metadatos

El conjunto de datos esta formado por un total de 7000 observaciones y 13 variables. Estas variables son:

VARIABLE	DESCRIPCION
GRCODE	Código de la compañía según la NAIC (incluye grupos aseguradores e individuales).
GRNAME	Nombre de la compañía según la NAIC (incluye grupos aseguradores e individuales).
AccidentYear	Año del accidente (de 1988 a 1997).
DevelopmentYear	Año de desarrollo (de 1988 a 1997).
DevelopmentLag	Desfase en el año de desarrollo (AY-1987 + DY-1987 - 1).
IncurLoss_R1	Perdidas incurridas y gastos asignados reportados al final del año.
CumPaidLoss_R1	Perdidas pagadas acumuladas y gastos asignados al final del año.
BulkLoss_R1	Reservas por pérdidas en bloque e IBNR (Reservas para siniestros ocurridos, pero no reportados) y gastos de defensa y contención de costos reportados al final del año.
EarnedPremDIR_R1	Primas ganadas en el año de incurrencia: directas y asumidas.
EarnedPremCeded_R1	Primas ganadas en el año de incurrencia: cedidas.
EarnedPremNet_R1	Primas ganadas en el año de incurrencia: netas.
Single	1 indica una Única entidad, 0 indica un asegurador de grupo.
PostedReserve97_R1	Reservas publicadas en el año 1997 tomadas del Underwriting and Investment Exhibit -Part 2A que incluye pérdidas netas no pagadas y gastos no pagados por ajuste de pérdidas.

## 2.3 Calidad de los datos

Se realiza la verificación de valores únicos y la existencia de datos faltantes (valores nulos) por variable, donde se obtienen los siguientes resultados:

VARIABLE	VALORES UNICOS	VALORES NULOS	TIPO
GRCODE	70	0	Cualitativa
GRNAME	70	0	Cualitativa
AccidentYear	10	0	Cualitativa
DevelopmentYear	19	0	Cualitativa
DevelopmentLag	10	0	Cualitativa
IncurLoss_R1	1417	0	Cuantitativa
CumPaidLoss_R1	1122	0	Cuantitativa
BulkLoss_R1	821	0	Cuantitativa
EarnedPremDIR_R1	425	0	Cuantitativa
EarnedPremCeded_R1	240	0	Cuantitativa
EarnedPremNet_R1	392	0	Cuantitativa
Single	2	0	Cualitativa
PostedReserve97_R1	50	0	Cuantitativa

De este análisis se identifican variables cualitativas y cuantitativas, también se observa no existen datos faltantes. Para las variables cuantitativas se obtiene el análisis de medidas de tendencia central, dispersión y posición:

	IncurLoss_R1	CumPaidLoss_R1	BulkLoss_R1	EarnedPremDIR_R1	EarnedPremCeded_R1	EarnedPremNet_R1
<b>Conteo</b>	7000	7000	7000	7000	7000	7000
<b>Media</b>	2133,64	1262,71	463,16	3908,3	618,72	3289,57
<b>Desviación estándar</b>	10837,22	7437,6	3433,73	18108,13	3307,02	15112,64
<b>Mínimo</b>	-504	-862	-423	-12	-90	-22
<b>P25%</b>	0	0	0	5	0	3
<b>P50%</b>	5	1	0	120	7	84,5
<b>P75%</b>	249	121,25	8	1031	92	650,25
<b>Máximo</b>	142234	129300	87677	185016	39845	150180

Es importante mencionar que la presencia de valores negativos en los registros de reserva es inherente a la naturaleza de la gestión de reservas en una compañía aseguradora. En casos en los que aparecen valores negativos, se ha efectuado una liberación de la reserva. Por ejemplo, supongamos que, en un determinado año, la aseguradora reserva \$100 para un siniestro específico. Sin embargo, tras un análisis más detenido, se determina que el costo final del siniestro será de \$70. En este caso, los \$30 restantes previamente reservados deben ser liberados de la

cuenta de reserva. Este monto se registra como un valor negativo, en este caso, -\$30, para indicar su retirada de la reserva. De forma similar, puede ocurrir con las primas dados los movimientos contables que existen al realizar devoluciones, cambios en la vigencia de las pólizas o los valores asegurados.

#### **2.4 Exploración inicial**

### **3 PREPARACIÓN DE LOS DATOS**

#### **3.1 Selección**

#### **3.2 Construcción de los datos a trabajar**

### **4 BIBLIOGRAFÍA**