

Modelos Actuariales No Vida: Enfoque Riesgo Individual

Implementación Modelo Jerárquico para cálculo de primas

Gibrán Peniche González-Carpio

Otoño 2019

Contenido

1	Introducción	1
1.1	Planteamiento	1
2	Análisis Exploratorio de Datos	2
3	Tratamiento de Datos Faltantes	4
4	Ajuste de Modelos	4
4.1	Ajuste de densidades	5
5	Selección de Modelos	5
5.1	Licitación	5
6	Cálculo de Primas	7
6.1	Prima de riesgo individual	7
6.2	Prima de riesgo agregada	10
7	Capital Mínimo requerido	10
7.1	Capital Operativo	10
7.2	Capital Mínimo Requerido	10

1 Introducción

La administración del riesgo es importante para cualquier empresa. En el sector asegurador no es la excepción a esta situación, y de hecho, el riesgo de insolvencia se vuelve particularmente relevante para las aseguradoras dada la función social de los productos de seguros.

Para la administración del riesgo, es importante poder caracterizarlo y definirlo de tal suerte de poder cuantificar en términos monetarios la exposición que se tiene al mismo.

En nuestro caso estamos en el contexto del seguro de automoviles para lo cual contamos con una base de datos que contiene la siguiente información.

DatosRDA refiere a 1,340 reclamos que detalla las siguientes características:

- **CASENUM**: Numero de caso para identificar el reclamo (numerico con valores de 1 a 1340)
- **ATTORNEY**: Indicadora si el reclamante tuvo representacion legal (factor con valores =1 si tuvo representacion y =2 si no)
- **CLMSEX**: Género del reclamante (factor con valores =1 para masculino y =2 para femenino)
- **MARITAL**: Estado marital (factor con valores =1 para casado, =2 para soltero, =3 para viudo, y =4 para divorciado/seprado)
- **CLMINSUR**: Indicadora si el conductor del vehiculo tenia seguro (factor con valores =1 si estaba asegurado, =2 si no estaba asegurado, y =3 si no aplica)
- **SEATBELT**: Indicadora si el reclamante llevaba o no el cinturon de seguridad o sistema de retencion infantil (factor con valores =1 si estaba asegurado, =2 si no estaba asegurado, y =3 si no aplica)
- **CLMAGE**: Edad del reclamante (numerico)
- **LOSS**: Perdida economica total del reclamante (en miles de USD)

Nos interesa la **distribucion de probabilidad agregada del monto de reclamo total, la prima de riesgo agregada, la prima de riesgo individual** para la cobertura para este tipo de reclamos y el **el capital mínimo requerido** para garantizar que la compania de seguros no incurra en insolvencia en un año de operación dado.

1.1 Planteamiento

Nos encontramos en el enfoque de **riesgo individual** y nos interesa la distribución del monto agregado de reclamcion, al que llamaremos Y , con el fin de cuantificar la exposición al riesgo de la aseguradora en este portafolio de pólizas en particular.

En primer lugar, realizaremos un proceso de *licitación* de modelos con el fin de determinar que distribución sigue el monto de reclamación.

Posteriormente realizaremos un ejercicio de *diversificación de riesgos* con el fin de precisar la exposición de la aseguradora dadas las características del contrante.

Finalmente cuantificaremos la exposición de la compañía medida a través de la prima de riesgo con el fin de determinar los requerimientos de capital para evitar una escenario de insolvencia en este portafolio de pólizas.

2 Análisis Exploratorio de Datos

Veamos las características principales de las variables.

Table 1: Table continues below

Abogado	Genero	EstadoCivil	Seguro
Con Abogado:685	Femenino :742	Casado :624	Asegurado : 120
Sin Abogado:655	Masculino:586	Divorciado: 35	No Aplica : 41
NA	NA's : 12	Soltero :650	No Asegurado:1179
NA	NA	Viudo : 15	NA
NA	NA	NA's : 16	NA
NA	NA	NA	NA

Cinturon	Edad	Monto
Con Cinturon:1270	Generación X:743	Min. : 0.005
Sin Cinturon: 22	Generación Z:408	1st Qu.: 0.640
NA's : 48	NA's :189	Median : 2.331
NA	NA	Mean : 5.954
NA	NA	3rd Qu.: 3.995
NA	NA	Max. :1067.697

A continuación exploramos la distribución del total de las observaciones.

Monto de reclamación

Histograma Y Diagrama de Caja Y brazos

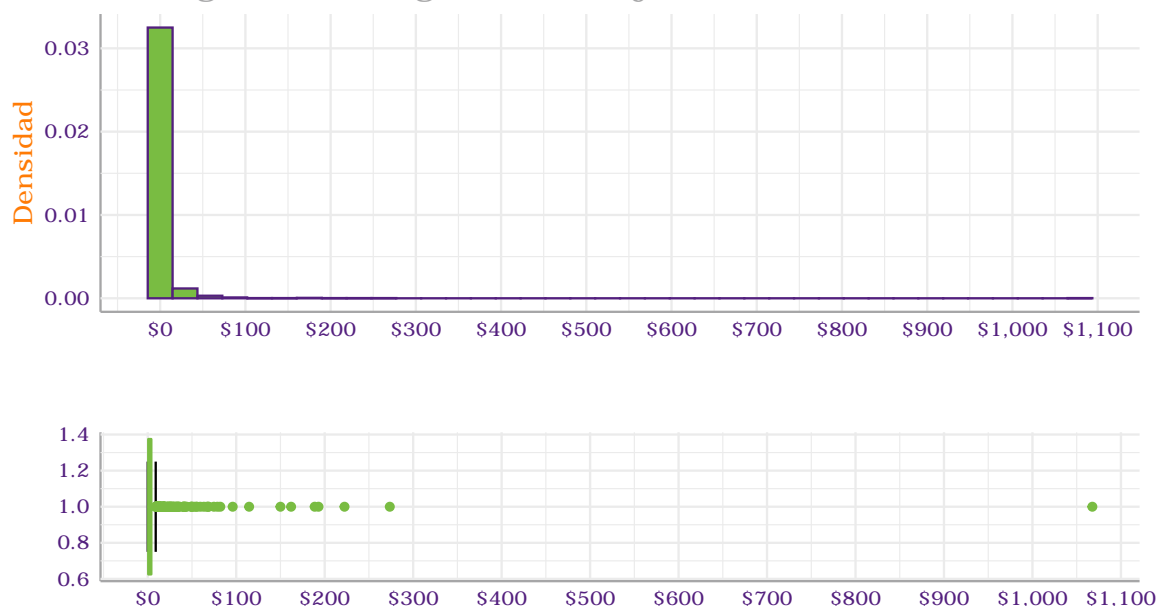
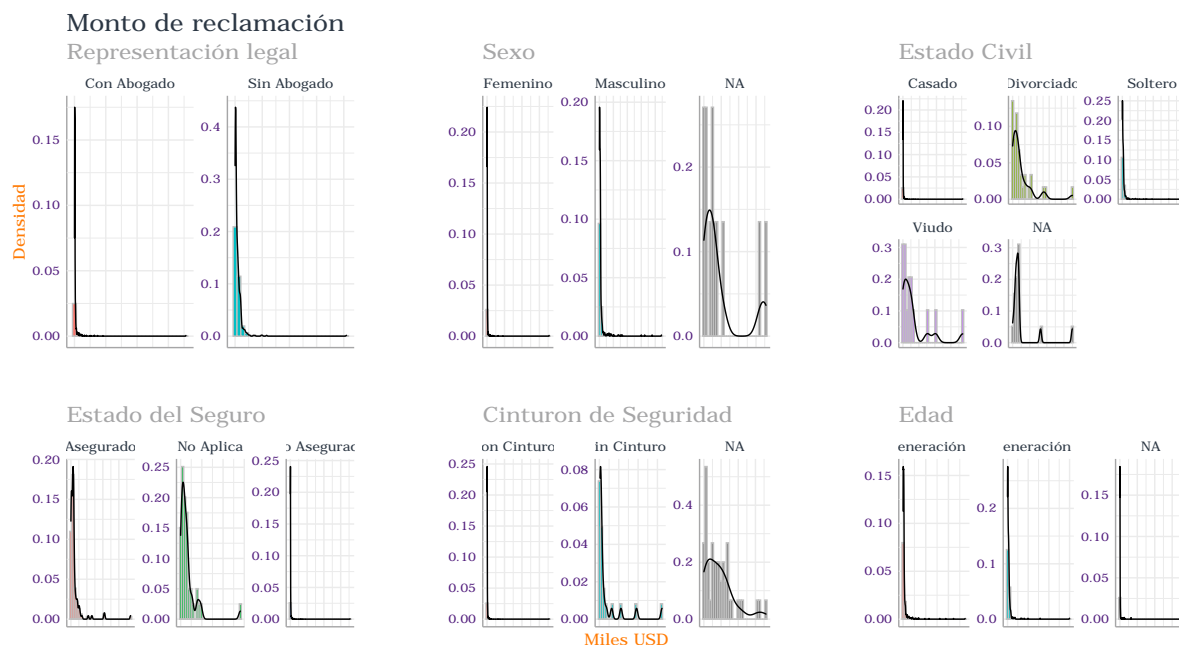


Table 3: Momentos Muestrales

Media	Mediana	Varianza	Sesgo	Kurtosis
5.953	2.331	1098	25.72	794.6

Del visual y resumen de momentos anterior destaca una distribución con **sesgo** y un **rango intercuartílico** reducido, así como **observaciones de cola**.

Por otro lado vale la pena estudiar la distribución del monto de reclamación condicional a las variables **cualitativas** incluidas en la base de datos.



La variables de **Representación legal** así como la de **Cinturón** son variables que se observan *ex-post*, sin embargo, del análisis anterior encontramos evidencia de que es una variable relevante para el monto de la reclamación, tanto cómo para el monto en el sentido de la magnitud de la indemnización así como en el número de reclamaciones.

Explorando el monto de reclamación por **Género** del reclamante parece ser que los hombres tienen más reclamos, existe cierta propensión a que el monto sea mayor en las colas de las mujeres. Llama la atención la distribución de los datos faltantes, que podría deberse a errores de captura o identificación no binaria de género.

Avancemos ahora a estudiar el monto de reclamo por **Estado Civil**. Existen diferencias importantes en las distribuciones en varios de los momentos centrales incluso en términos de multimodalidad.

Continuamos con la variable **Estado del Seguro**. Donde encontramos importantes diferencias en término del monto del reclamo.

Ahora explorando la variable **Cinturón de Seguridad**, que al igual que **Representación legal** y **Estado del Seguro**, es una variable que se observa *ex-post*, sin embargo vale la pena realizar un ajuste a la prima de riesgo por las propenciones de estas variables.

Finalmente realicemos el análisis de **Edad**, recordemos que esta variable ha sido codificada en **términos generacionales**.¹

¹Brechas de 7 años

Llama la atención que, al menos en nuestra base de datos, ocurre que los montos se concentran en dos grupos generacionales. Recomendamos profundizar con más información más adelante en este fenómeno.

Vale la pena profundizar en la condicionalidad del monto con respecto a las diferentes variables, por cuestiones de tiempo y extensión del proyecto, consideramos el análisis exploratorio visual como evidencia suficiente para realizar un ejercicio de diversificación de riesgos. Recomendamos en particular explorar momentos muestrales y proporciones de los datos de cada tipo con respecto al total para robustecer, en particular, la inclusión de las variables cualitativas.

Del análisis anterior llama atención varios datos no disponibles correspondientes a variables cualitativas a los que daremos tratamiento a continuación. Además existe diferencias sustanciales de la distribución del monto agregado asociado a las variables cualitativas incorporadas en la base de datos, lo cual sustenta la hipótesis de no **homogeneidad** invitando a realizar un ejercicio de *diversificación* de riesgos posterior a la licitación de modelos.

3 Tratamiento de Datos Faltantes

Primero realizamos un diagnóstico general de los datos faltantes. Generaremos un data frame del tipo *tabular* para indicar que valores van a ser imputados.

Utilizamos la función *imputeMCA* de la librería **missMDA** que utiliza Análisis de Correspondencia Múltiple iterando sobre proporciones de datos que si están disponibles existentes.

Table 4: Datos Completos (continued below)

Abogado	Genero	EstadoCivil	Seguro
Con Abogado:685	Femenino :752	Casado :633	Asegurado : 120
Sin Abogado:655	Masculino:588	Divorciado: 35	No Aplica : 41
NA	NA	Soltero :657	No Asegurado:1179
NA	NA	Viudo : 15	NA

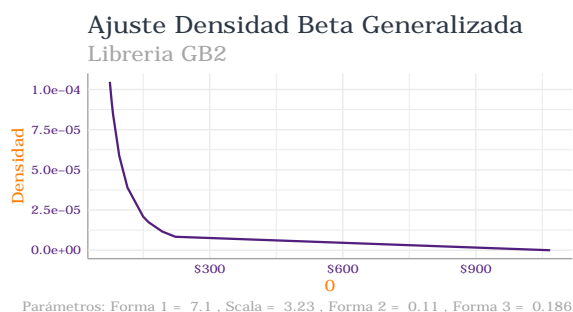
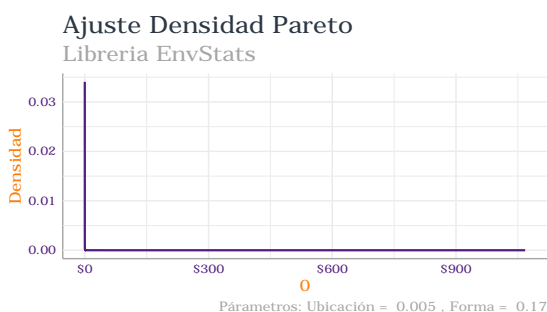
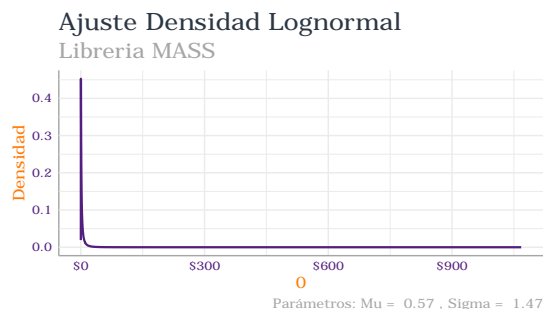
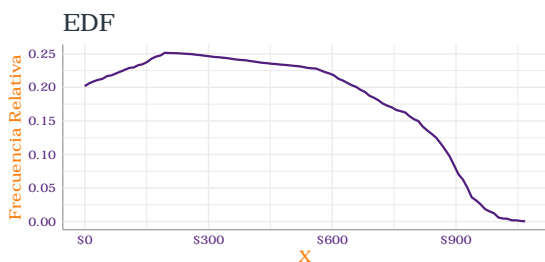
Cinturon	Edad
Con Cinturon:1318	Generación X:926
Sin Cinturon: 22	Generación Z:414
NA	NA
NA	NA

4 Ajuste de Modelos

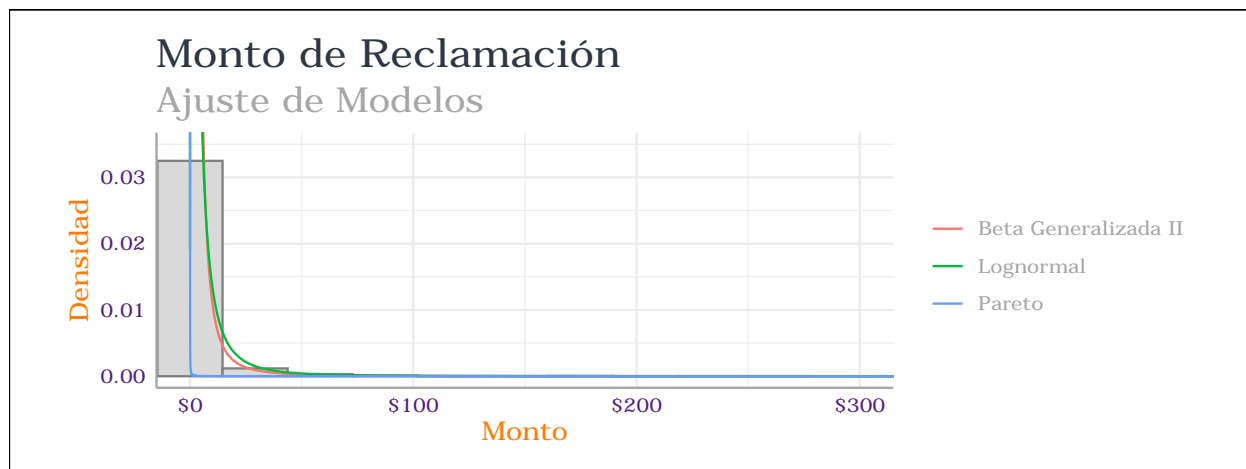
A continuación comenzaremos el proceso de licitación de modelos de probabilidad. Entre los candidatos encontramos al modelos *Lognormal*, *Pareto*, y *Beta Generalizada del segundo tipo*. En las siguientes secciones procederemos a realizar el ajuste. A pesar de que en clase estudiamos también la distribución *Pareto Generalizado* y la distribución *Gamma Generalizada*. La primera es un modelo que se utiliza generalmente para el estudio de valores extremos y el segundo presenta problemas de estimación que por cuestiones de tiempo y extensión del presente trabajo no resolveremos. Por la razones mencionadas anteriormente los dos modelos anteriores **no** participaran en el proceso de licitación.

La información del modelo no tiene etiquetas temporales, por esta razón tomamos una muestra aleatoria correspondies al 80% de los datos para entrenar el modelo y el restante 20% como conjunto de prueba para el modelo.

4.1 Ajuste de densidades



Dado el ajuste anterior visualmente tenemos lo siguiente.



5 Selección de Modelos

5.1 Licitación

Realizamos la licitación de modelos por los criterios de **Verosimilitud Predictiva**, **Kolmogorov-Smirnoff** y **Anderson Darling**. Buscamos escoger la distribución que mejor ajuste los datos, para utilizarla más adelante de tal suerte de determinar la prima de riesgo del portafolio.

5.1.1 Verosimilitud predictiva

Table 6: Verosimilitud Predictiva Por este criterio la distribución que mejor ajusta los datos es la Distribución Beta Generalizada del segundo tipo.

lognormal	pareto	beta
-600.4	-845.7	-593.2

5.1.2 Kolmogorov Smirnov

Table 7: Prueba Kolmogorov-Smirnov

T_05	lognormal	pareto	betagen
0.08429	0.113	0.4041	0.05775

Bajo la prueba **K-S**, que por cierto es más sensible a valores más cercanos a la mediana, volvemos a concluir que la distribución Beta Generalizada del segundo tipo es la que mejor ajusta los datos.

5.1.3 Anderson Darling

Table 8: Prueba Anderson-Darling

T_05	lognormal	pareto	betagen
2.492	2.957	7.995	1.094

Hay que observar que la densidad Beta Generalizada del segundo tipo no es acotada. Por esta razón el test *A-D* no es una buena métrica de contraste en este caso. Sin embargo, por este criterio la distribución lognormal se impondría a la distribución pareto

5.1.4 Selección y Comentarios

Después del proceso de licitación concluimos que de la densidad *Beta Generalizada del tipo 2* es la densidad que mejor ajusta los datos.

Por otro lado dada la evidencia a favor de un ejercicio de diversificación (*i.e* no se cumple el supuesto de **Homogeneidad**) en la secciones subsecuentes realizaremos un ajuste de modelos vía jerarquización.

A pesar de que el proceso de licitación arrojó como un modelo más adecuado a la ya mencionada distribución, ocurre que por el momento las herramientas disponibles para realizar el ajuste de jerarquías para esta distribución no está disponible. Si bien construir el código para la implementación de este modelo no es difícil está más allá de los objetivos del curso y lamentablemente sujetos a las restricciones de tiempo y extensión del proyecto no es será posible implementarlo. Sin embargo valdría la pena explorar más adelante el ajuste vía modelo jerárquico para dicha distribución.

Por la razones enlistadas anteriormente en los subsecuente se utilizará el modelo *Log-Normal* para realizar el ajuste del modelo jerárquico.

Table 9: Codificación

Status	Identificador
Masculino Generación X Casado	S1
Femenino Generación X Soltero	S2
Masculino Generación Z Soltero	S3
Masculino Generación X Divorciado	S4
Femenino Generación X Casado	S5
Masculino Generación X Soltero	S6
Femenino Generación Z Soltero	S7
Femenino Generación X Divorciado	S8
Femenino Generación Z Casado	S9
Masculino Generación Z Casado	S10
Femenino Generación X Viudo	S11
Masculino Generación X Viudo	S12

6 Cálculo de Primas

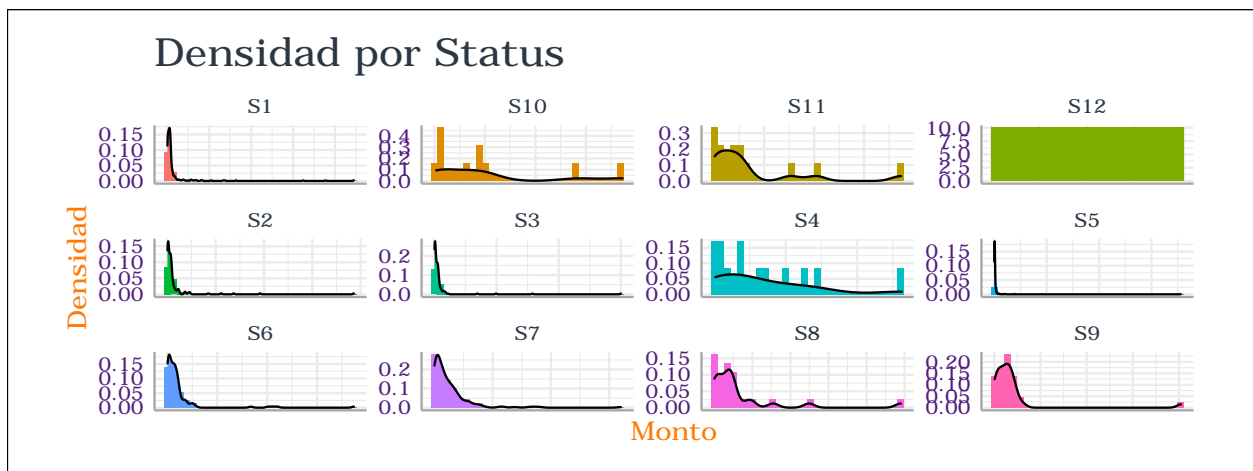
6.1 Prima de riesgo individual

Comenzamos el cálculo de la prima individual por diversificación de riesgos vía modelos jerárquicos.

Construimos primero una etiqueta que incorpora la información de las variables **Edad**, **Estado Civil** y **Género** de tal suerte de crear un *status* único para cada una de las combinaciones de las 3 variables que nos permitan caracterizar la distribución del monto dependiendo de las características del asegurado.

Cabe señalar que las variables **Representación Legal**, **Seguro** y **Cinturón** se observan *ex-post*, por lo que no van a ser incluidas en el modelo. Sin embargo serán consideradas más adelante.

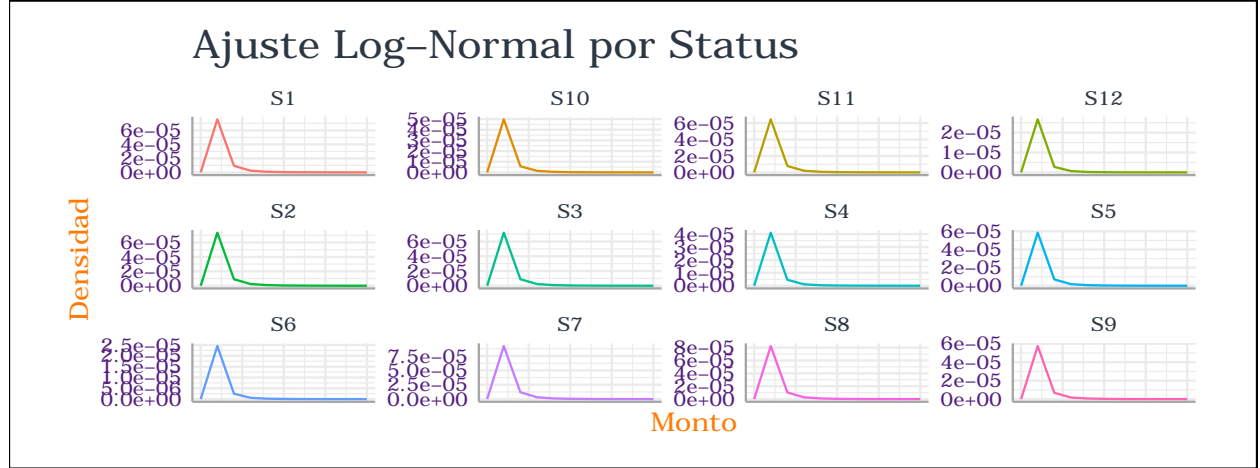
La distribución dados los estatus se comporta como sigue:



Llama la atención la multimodalidad de algunos status, por lo que se recomienda un acercamiento de modelación vía mezclas que no exploraremos en este trabajo.

Comenzamos por realizar el ajuste del modelo. La distribución del monto Y condicional al status será de la forma:

$$Y|\iota_i \sim \text{LogNormal}(y|\mu_i, \sigma^2)$$



6.1.1 Prima Base

Encontramos la prima de riesgo básica $\pi_{i,base}$ por estatus que corresponde a $E[X|\iota_i]$ con $i = \{Status_i\}$ tal que $Status_i = \{ \text{Masculino Generación X Casado, Femenino Generación X Soltero, Masculino Generación Z Soltero, Masculino Generación X Divorciado, Femenino Generación X Casado, Masculino Generación X Soltero, Femenino Generación Z Soltero, Femenino Generación X Divorciado, Femenino Generación Z Casado, Masculino Generación Z Casado, Femenino Generación X Viudo, Masculino Generación X Viudo} \}$, cómo sigue.

Table 10: Prima Básica por Status

Status	Prima Base
S1	0.7145194
S2	0.6979925
S3	0.6865280
S4	0.3957176
S5	0.5767529
S6	0.1447292
S7	0.8224561
S8	0.7598198
S9	0.5674323
S10	0.4924394
S11	0.6287465
S12	0.1850209

6.1.2 Factor de Recarga

Sin embargo aun debemos incorporar ajustes al monto por la probabilidad de que asegurado contrate representación legal y lleve el cinturón de seguridad para generar un simil al *factor de recarga* de la forma $\Theta_{1,2} = (1 + \theta_1)(1 + \theta_2)p_1p_2$, de tal manera que la prima $\pi_{recarga}$ quede definida como $\Theta\pi_{base}$. Aquí p_1 y p_2 representan la propensión de los asegurados a solicitar asistencia legal en un areclamación y a utilizar el cinturón de seguridad mientras manejan.

Construiremos el factor de recarga a partir del exceso en valor esperado de la distribución del monto de reclamo exceso condicional a las variables **Representación Legal** y **Cinturón de Seguridad** del valor esperado de la distribución no condicional del monto de reclamo. Resultado en $\Theta = 1.22$

Así, la prima modificada por la incidencia de Representación legal y Cinturón de seguridad es de:

Table 11: Prima Modificada A

Status	Prima Recarga
S1	0.8734553
S2	0.8532523
S3	0.8392376
S4	0.4837401
S5	0.7050444
S6	0.1769224
S7	1.0054013
S8	0.9288323
S9	0.6936505
S10	0.6019765
S11	0.7686033
S12	0.2261765

6.1.3 Prima de Riesgo por el principio de Varianza

Definimos el *Coefficiente de Variación* $CV = \frac{\sigma}{\mu}$. Para los valores estimados en la sección **Ajuste de Modelos** para la densidad log-normal. Tenemos un valor de $\mu = 0.571$ y $\sigma = 1.47$ resultando en un $CV = 2.58$.

De este análisis concluimos que una unidad adicional en incertidumbre contribuye para nuestra parametrización propuesta con 2.58 unidades de riesgo al portafolio agregado.

Si bien hay que atender a cuestiones de mercado, reservas y de solvencia, en el presente trabajo unicamente con base en el criterio anterior proponemos $\alpha = \frac{CV}{2}$ y dada la varianza estimada del modelo anterior de 2.11 unidades definimos la **Prima de Riesgo** π_{riesgo} como $\pi_{base} + \alpha V[Y]$ por lo que el recargo por principio de varianza será \$ 2.72.

6.1.4 Tarificación

Dependiendo deL **STATUS** definido anteriormente la **Prima individual** que deberá cobrar la aseguradora estará tarificada como sigue:

Table 12: Tarificación en Dólares

Status	Prima de Riesgo
S1	3.593
S2	3.572
S3	3.558
S4	3.203
S5	3.424
S6	2.896
S7	3.725
S8	3.648
S9	3.413
S10	3.321
S11	3.488
S12	2.945

6.2 Prima de riesgo agregada

Por el principio de varianza y para el mismo factor de α dado el modelo **Log Normal** la prima de riesgo será el promedio ponderado por los pesos de los status, esto ya que la prima de riesgo por principio de varianza cumple el principio de *aditividad*.

De esta manera la **Prima Colectiva** queda definida como $\Pi = \sum_{i=1}^L \omega_i \pi_i$ donde π_i representa la i -ésima prima individual de las L jerarquías del portafolio y ω_i el número de pólizas relativo al total de la i -ésima jerarquía del portafolio.

Así la prima colectiva está dada por \$3.48 miles de USD,

7 Capital Mínimo requerido

7.1 Capital Operativo

De la sección **Selección de Modelos**, *Selección y Comentarios* se planteó trabajar con el modelo lognormal. Ahora bien, definimos el capital operativo K_{op} como sigue $Pr(Y > K_{op}) = 0.995 \iff K_{op} = q_y(0.995)$.., $Y \sim \logNormal(y|\mu, \sigma^2)$. Por lo que en nuestro caso y para este portafolio de polizas en particular el **Capital Operativo** será de \$ 78.45 miles de USD.

7.2 Capital Mínimo Requerido

Finalmente el Capital Mínimo Requerido K_{min} queda definido como $K_{min} = K_{op} - \Pi$. Si en nuestro caso la Prima Colectiva es de \$3.48 el Capital Operativo es de \$78.45 el **Capital Mínimo Requerido** es de \$74.96 miles de USD.