

Análisis del coste de los siniestros en una compañía de seguros utilizando las distribuciones asimétricas skew-normal y skew-t

Analysis of the cost of claims at an insurance company using skew-normal and skew-t asymmetric distributions

¹*Carlos López de Castilla Vásquez y ²Diego Alonso Mendoza Quevedo

Resumen

La presente investigación tiene por objetivo principal determinar si las distribuciones asimétricas *skew-normal* y *skew-t* son buenos modelos para describir datos relativos al coste de los siniestros. Para lo cual se analizaron dos conjuntos de datos de una compañía de seguros, ajustando las distribuciones en estudio y las tradicionales a estos datos. Luego se compararon las distribuciones empleando el Criterio de Información de Akaike (AIC) y del Logaritmo de la función de Verosimilitud, complementados con la prueba de bondad de ajuste Kolmogorov-Smirnov, obteniendo resultados positivos. Además, se calcularon medidas de riesgo como el Valor en Riesgo (VaR) y el Valor en Riesgo Condicional (TVaR), que brindaron mayor solidez a los resultados previos: los costes de los siniestros en los seguros se ajustan bien a las distribuciones asimétricas *skew-normal* y *skew-t*.

Palabras clave en español: Distribucion asimetrica, skew-normal, skew-t, VaR, TVaR

Abstract

The main purpose of this research is to determine if the skew-normal and skew-t asymmetric distributions are good models to describe data related to the claim cost. Therefore, two data sets of an insurance company were analyzed, fitting the studied and the traditional distributions to said data. Then distributions were compared using the Akaike Information Criterion (AIC) and the Likelihood function Logarithm, supplemented by the Kolmogorov-Smirnov goodness-of-fit test, obtaining positive results. In addition, risk measurements were calculated, such as the Value-at-Risk (VaR) and the Tail Value-at-Risk (TVaR) which provided greater strength to the previous results: the claim costs in the insurances fit well to the skew-normal and skew-t asymmetric distributions.

Key words: Asimetric distribution, skew-normal, skew-t, VaR, TVaR.

1. Introducción

En muchos campos de estudio los datos recolectados se ajustan a una distribución *normal*; por ejemplo, en economía y finanzas es el modelo más popular; Eling, (2011). Sin embargo, esto no siempre ocurre, como es el caso de los datos relacionados al coste de los siniestros en el campo de los seguros; Vernic, (2006) que se ajustan mejor a distribuciones asimétricas.

El modelo asimétrico que se empleó en la presente investigación, que además incluye a la distribución *normal* como un caso particular, es la distribución *skew-normal* introducida por Azzalini (1985). En los últimos años esta distribución se ha estudiado con mayor profundidad, generando nuevos modelos como por ejemplo la distribución *skew-t* desarrollada por Azzalini y Capitanò (2003), que también se analizó en la presente.

Por lo tanto, se espera determinar si las distribuciones *skew-normal* y *skew-t* son buenos modelos para describir datos relativos al coste de los siniestros en una compañía

de seguro, con el fin de lograr que la empresa pueda utilizar estos modelos como herramientas para prever modificaciones en sus coberturas y pueda valorar los beneficios que le reporta el seguro (en el presente estudio se analiza el Seguro Complementario de Trabajo en Riesgo (SCTR)).

2. Materiales y métodos

Tipo de investigación

Con el fin de lograr los objetivos de la investigación se aplicó una investigación exploratoria y descriptiva, puesto que se realizó una revisión profunda de la bibliografía y se describió en detalle la variable coste del siniestro.

Diseño de investigación

Es una investigación no experimental de tipo longitudinal debido a que se consiguió datos relativos al coste de los siniestros a lo largo del tiempo, desde 1998 hasta 2010.

¹Departamento Académico de Estadística e Informática, Universidad Nacional Agraria La Molina, Lima, Perú. E-mail: clopez@lamolina.edu.pe
*Autor de correspondencia.

²Ingeniero Estadístico e Informático de la UNALM

Formulación de hipótesis

Las distribuciones asimétricas (*skew-normal* y *skew-t*) son buenos modelos para describir datos relativos al coste de los siniestros en una compañía de seguros.

Población

La población en el primer grupo lo constituyen 326 datos del coste de los siniestros por cobertura de *Invalidez*, que fueron aprobados y pagados totalmente por la empresa. La población en el segundo grupo está formada por 197 datos del coste de los siniestros por cobertura de *Sobrevivencia*, que fueron aprobados y pagados parcialmente por la empresa.

Metodología aplicada

Los pasos a seguir para poder contrastar la hipótesis de investigación son:

1. Análisis descriptivo de los datos.
2. Análisis descriptivo del logaritmo de los datos.
3. Ajuste de los datos a las distribuciones en estudio (*skew-normal* y *skew-t*) y a las distribuciones tradicionales.
4. Comparación de las distribuciones analizadas mediante diferentes criterios estadísticos.
5. Estimación de medidas de Riesgo (VaR y TVaR) que complementen los resultados.

Sistema de seguros

Según la Asociación Peruana de Afiliados de Seguros (2012), el Sistema de Seguros es un sistema de protección del hombre y de su patrimonio frente a diversos hechos que amenazan su integridad, su vida, su interés y su propiedad. El Seguro garantiza la indemnización de un capital para reparar la pérdida que aparezca en cualquier momento, recibiendo como contraprestación un precio por adelantado por el servicio de protección que ofrece.

Estadística Actuarial

Saravia, Gómez y Vásquez (2007) definen la Estadística Actuarial (también llamada Matemática Actuarial) como una ciencia que brinda las herramientas necesarias para estudiar ciertas actividades económicas que llevan a cabo las compañías de seguros. Estas actividades aparecen generalmente ligadas al término *riesgo*, habitual en los escenarios actuariales. Por riesgo entenderemos cualquier suceso que pueda ocurrir y que suponga una pérdida, generalmente cuantificable en términos económicos.

Seguro Complementario de Trabajo en Riesgo (SCTR)

Es el tipo de seguro que se analizó en la presente investigación y según la SBS (2012) es un seguro obligatorio que otorga coberturas por accidente de trabajo y por enfermedad profesional a los trabajadores, empleados y obreros, que tienen la calidad de afiliados

regulares del Seguro Social de Salud y que laboran en un centro de trabajo en el que la entidad empleadora realiza las actividades de riesgo.

Valor en Riesgo (VaR - Value at Risk)

Se define como la pérdida máxima esperada durante un determinado intervalo de tiempo, bajo condiciones normales de mercado y dentro de un nivel de confianza establecido.

Valor en Riesgo de la Cola (TVaR - Tail Value at Risk):

Es la media de los valores que superan el VaR. Actuarialmente, se considera mejor medida del riesgo que el VaR. La ventaja de usar la medida del TVaR como una medida del riesgo es que proporciona la suma total de pérdidas por encima de un cierto nivel de confianza.

Distribución Skew-Normal

La distribución *skew-normal* (normal asimétrica traducida al español) fue formalmente introducida por Azzalini (1985). A partir de entonces, las propiedades probabilísticas de esta distribución fueron ampliamente estudiadas.

Una variable aleatoria Z tiene distribución normal asimétrica estándar con parámetro de asimetría λ , denotado por $SN(\lambda)$, si su densidad es dada por:

$$f(z)=2\phi(z)\Phi(\lambda z), \quad z \in R \quad (1)$$

donde λ es un valor definido en R , y $\phi(\cdot)$ y $\Phi(\cdot)$ denotan la función de densidad de probabilidad y la función de distribución acumulada de la distribución normal estándar.

La ecuación (1) muestra el comportamiento de $f(z)$ para diferentes valores de λ , y es fácil notar que la distribución normal estándar está incluida como caso particular de la normal asimétrica estándar cuando $\lambda=0$.

Una variable aleatoria Y tiene distribución skew-normal ($Y \sim SN(\mu, \sigma, \lambda)$) con parámetro de asimetría λ , parámetro de locación μ y el parámetro de escala σ , ($\sigma > 0$) si su densidad es dada por:

$$f(y)=\frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\left(\frac{y-\mu}{\sigma}\right)\right), y \in R \quad (2)$$

Notar que si $Z \sim SN(\lambda)$ y $Y = \mu + \sigma Z$, entonces $Y \sim SN(\mu, \sigma, \lambda)$. O sea, cualquier combinación lineal de una variable aleatoria normal asimétrica estándar también será una distribución normal estándar.

Estimación de parámetros

Considere $Z=(Z_1, Z_2, \dots, Z_n)$ una muestra n aleatoria de tamaño $Z \sim SN(\lambda)$ de una variable aleatoria. La función de verosimilitud es dada por

$$L(\lambda; z) = \prod_{i=1}^n 2\phi(z_i)\Phi(\lambda z_i)$$

Para obtener el estimador de máxima verosimilitud para λ es preciso maximizar el logaritmo de la función de verosimilitud, que es dado por:

$$\log L(\lambda; z) = \sum_{i=1}^n [\log(2\phi(z_i)) + \log\Phi(\lambda z_i)].$$

La primera derivada es dada por:

$$\frac{\partial \log L(\lambda; z)}{\partial \lambda} = \sum_{i=1}^n \frac{z_i \phi(\lambda z_i)}{\Phi(\lambda z_i)}$$

donde las raíces de esa derivada pueden ser obtenidas numéricamente. Sea $X = \mu + \sigma Z$, esto es, $X \sim SN(\mu, \sigma, \lambda)$. Considerando una muestra aleatoria de tamaño n su función de verosimilitud es:

$$L(\mu, \sigma, \lambda; \mathbf{x}) = \left(\frac{2}{\sigma}\right)^n \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2} \left[\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2}\right]\right\} \prod_{i=1}^n \Phi\left(\lambda \frac{(x_i - \mu)}{\sigma}\right)$$

y su matriz de información de Fisher $I(\mu, \sigma, \lambda)$ es dada por

$$\begin{bmatrix} \frac{1 + \lambda^2 a_0}{\sigma^2} & \frac{E(Z) \frac{1 + 2\lambda^2}{1 + \lambda^2} + \lambda^2 a_1}{\sigma^2} & \frac{\sqrt{2/\pi}(1 + \lambda^2)^{3/2} - \lambda a_1}{\sigma} \\ \frac{E(Z) \frac{1 + 2\lambda^2}{1 + \lambda^2} + \lambda^2 a_1}{\sigma^2} & \frac{2 + \lambda^2 a_2}{\sigma^2} & \frac{\lambda a_2}{\sigma} \\ \frac{\sqrt{2/\pi}(1 + \lambda^2)^{3/2} - \lambda a_1}{\sigma} & \frac{\lambda a_2}{\sigma} & a_2 \end{bmatrix}$$

donde

$$a_k = E_Z \left[\frac{z^k \phi(\lambda z)}{\Phi(\lambda z)} \right], k = 0, 1, 2.$$

Uno de los problemas encontrados en esa matriz es el hecho de que es singular cuando $\lambda=0$ lo que puede impedir que sea evaluada la existencia de asimetría utilizando pruebas basadas en la información de Fisher, como por ejemplo, el test de Wald. La mayor dificultad encontrada con el método de máxima verosimilitud es que estudios de simulación han demostrado que el estimador de máxima verosimilitud de λ puede ser infinito, aunque el verdadero valor de ese parámetro sea finito. Una alternativa es usar el algoritmo EM que requiere la obtención de una variable latente. Otra alternativa es utilizar la parametrización centrada de Azzalini (1985), de la forma como fue propuesto en Pewsey (2000).

Distribución Skew-t

Se plantea como alternativa más robusta al modelo *skew-normal* la versión asimétrica de la distribución *t*-student. Esta distribución llamada *skew-t* (distribución *t* asimétrica traducido al español) es una distribución asimétrica que no incluye sólo a la distribución *t*-student como un caso particular, además también a las distribuciones normal y *skew-normal*, en casos límites.

Una variable aleatoria Z es denominada *skew-t* estándar, con parámetros de asimetría ($\lambda \in R$) y curtosis ($v \in R_+^*$), si su función de densidad de probabilidad es dada por:

$$f_Z(z) = 2t_v(z)T_{v+1}\left(\lambda z \sqrt{\frac{1+v}{v+z^2}}\right) \quad (3)$$

donde t_v denota la función de densidad de probabilidad de una distribución *t*-student estándar con v grados de libertad, $T_{(v+1)}$ la función de distribución acumulada de una distribución *t*-student estándar con $v+1$ grados de libertad y $-\infty < z < \infty$. Como notación representamos $Z \sim ST(v, \lambda)$.

El parámetro λ regula la forma de la distribución. Valores negativos de λ indican asimetría negativa y valores positivos de λ indican asimetría positiva. La densidad definida en (3) será simétrica cuando $\lambda=0$, y coincidirá con la densidad de la distribución *t*-student estándar.

Con la introducción de los parámetros de posición ($\mu \in R$) y escala ($\sigma > 0$), conseguimos obtener una generalización de la distribución *skew-t* estándar. De esa manera, una variable aleatoria tendrá una “distribución *skew-t* con parámetros de posición y escala”, o simplemente *skew-t*, si $Y = \mu + \sigma Z$, donde $Z \sim ST(v, \lambda)$. La notación utilizada será $Z \sim ST(\mu, \sigma^2, v, \lambda)$.

Estimación de parámetros

Considere $Z = (Z_1, Z_2, \dots, Z_n)$ una muestra aleatoria simple de tamaño n de la variable aleatoria Z con distribución $ST(v, \lambda)$. En este caso, la función de verosimilitud es dada por:

$$L(v, \lambda; \mathbf{z}) = \prod_{i=1}^n 2t_v(z_i)T_{v+1}\left(\lambda z_i \sqrt{\frac{1+v}{v+z_i^2}}\right)$$

Si consideramos que el número de grados de libertad es conocido ($v > 0$), la función de verosimilitud dependerá apenas del parámetro λ , luego:

$$L(v, \lambda; \mathbf{z}) \propto \prod_{i=1}^n T_{v+1}\left(\lambda z_i \sqrt{\frac{1+v}{v+z_i^2}}\right)$$

Para esta situación, si $\forall z_i > 0$, entonces $L(v, \lambda; \mathbf{z})$ es una función monótona creciente en λ y, como consecuencia, el estimador que maximiza la función de verosimilitud será

infinito. Utilizando el mismo raciocinio, verificamos que si $\forall z_i < 0$, entonces el estimador de máxima verosimilitud será menos infinito.

Los estimadores de máxima verosimilitud, tanto para la distribución *skew-t* estándar, como para la distribución *skew-t* con parámetros de posición y escala no tienen forma cerrada. Por eso, las estimaciones de máxima verosimilitud deben ser obtenidas mediante métodos numéricos.

Pruebas de Bondad de Ajuste

Las pruebas de bondad de ajuste indican si es razonable o no asumir que una muestra aleatoria proviene de una distribución específica. La prueba de Kolmogorov – Smirnov es usada para decidir si una muestra proviene de una población con una distribución específica. Puede ser aplicada tanto para datos discretos y continuos agrupados (incluso si algunos autores no están de acuerdo con este punto) como para variables continuas. Esto se basa sobre una comparación entre la función de distribución acumulada empírica (ECDF por sus siglas en inglés) y la función de distribución teórica definida como $F(x) = \int_{-\infty}^x f(y; \theta) dy$, donde $f(y; \theta)$ es la función de densidad de probabilidad.

Criterio de Información de Akaike (AIC)

El Criterio de Información de Akaike (AIC) admite que el modelo *verdadero*, es decir, un modelo con la verdadera distribución de los datos es desconocido, y que dentro de los modelos que están siendo validados, ninguno es considerado el que realmente describe la variable en estudio, Akaike (1974). El modelo con menor valor de AIC es el modelo que mejor se a los datos.

3. Resultados y discusión

Las estadísticas descriptivas de la Tabla 1 muestran que las remuneraciones de Sobrevivencia son mayores con respecto a la asimetría y la kurtosis que las de Invalidez. Pero ambas tienen asimetría a la derecha. Esta característica también se refleja en los altos valores del Valor en Riesgo y el Valor en Riesgo Condicional. Este primer indicio hace pensar que una distribución asimétrica sea la más adecuada para ajustar estos tipos de datos.

En la Figura 1, se presentan los histogramas y sus correspondientes gráficos de normalidad Q-Q plots para ambos conjuntos de datos. En los histogramas se observa claramente un comportamiento asimétrico para las variables Invalidez y sobrevivencia.

En un segundo paso se analizan el logaritmo de los datos que es una práctica ampliamente utilizada en estadística y en la ciencia actuarial con el fin de disminuir los valores extremos que influyen en el propósito de modelamiento (Bolance *et. al.* 2008, citado por Eling 2011).

Tabla 1. Estadísticas descriptivas del coste de los siniestros para los grupos de Invalidez y Sobrevivencia

Estadísticas Descriptivas	Invalidez	Sobrevivencia
Nº de observaciones	326.00	197.0
Media (X)	2636.70	1659.8
Desv. Estánd. (X)	1659.74	1259.8
Asimetría (X)	0.64	2.2
Kurtosis (X)	-0.57	5.5
Mínimo	282.33	322.1
Máximo	7066.17	7299.2
99% Cuantil (Valor en Riesgo)	6764.62	6713.4
Media (X/X > Valor en Riesgo)	6886.50	7015.9

La Tabla 2 presenta las estadísticas descriptivas del logaritmo de los datos: número de observaciones, la media, la desviación estándar, asimetría, kurtosis, los valores mínimo y máximo, el percentil 99% y la pérdida media, si la pérdida está por encima del percentil 99%.

Tabla 2. Estadísticas descriptivas del logaritmo del coste de los siniestros para los grupos de Invalidez y Sobrevivencia

Estadísticas Descriptivas	Invalidez	Sobrevivencia
Nº de observaciones	326.00	197.00
Media (X)	7.65	7.20
Desv. Estánd. (X)	0.73	0.63
Asimetría (X)	-0.44	0.36
Kurtosis (X)	-0.61	-0.10
Mínimo	5.64	5.77
Máximo	8.86	8.90
99% Cuantil (Valor en Riesgo)	8.82	8.81
Media (X/X > Valor en Riesgo)	8.84	8.86

En la Figura 2, se presentan también los histogramas y sus correspondientes gráficos de normalidad Q-Q plots para el logaritmo natural de las variables Invalidez y Sobrevivencia. Después de tomar el logaritmo natural, los datos de

Invalidez aún mantienen una gran asimetría, mientras que los de Sobrevivencia se asemejan más a una distribución normal; sin embargo, en ambos casos, se puede apreciar desviaciones de normalidad, pero con colas no muy extremas.

Todos los parámetros estimados en la presente investigación están calculados bajo el método de la estimación de máxima verosimilitud. La Tabla 3 presenta los parámetros estimados de las distribuciones *skew-normal* y *skew-t* para el coste de los siniestros y el logaritmo natural de los mismos, tanto para el grupo de Invalidez como para el grupo de Sobrevivencia.

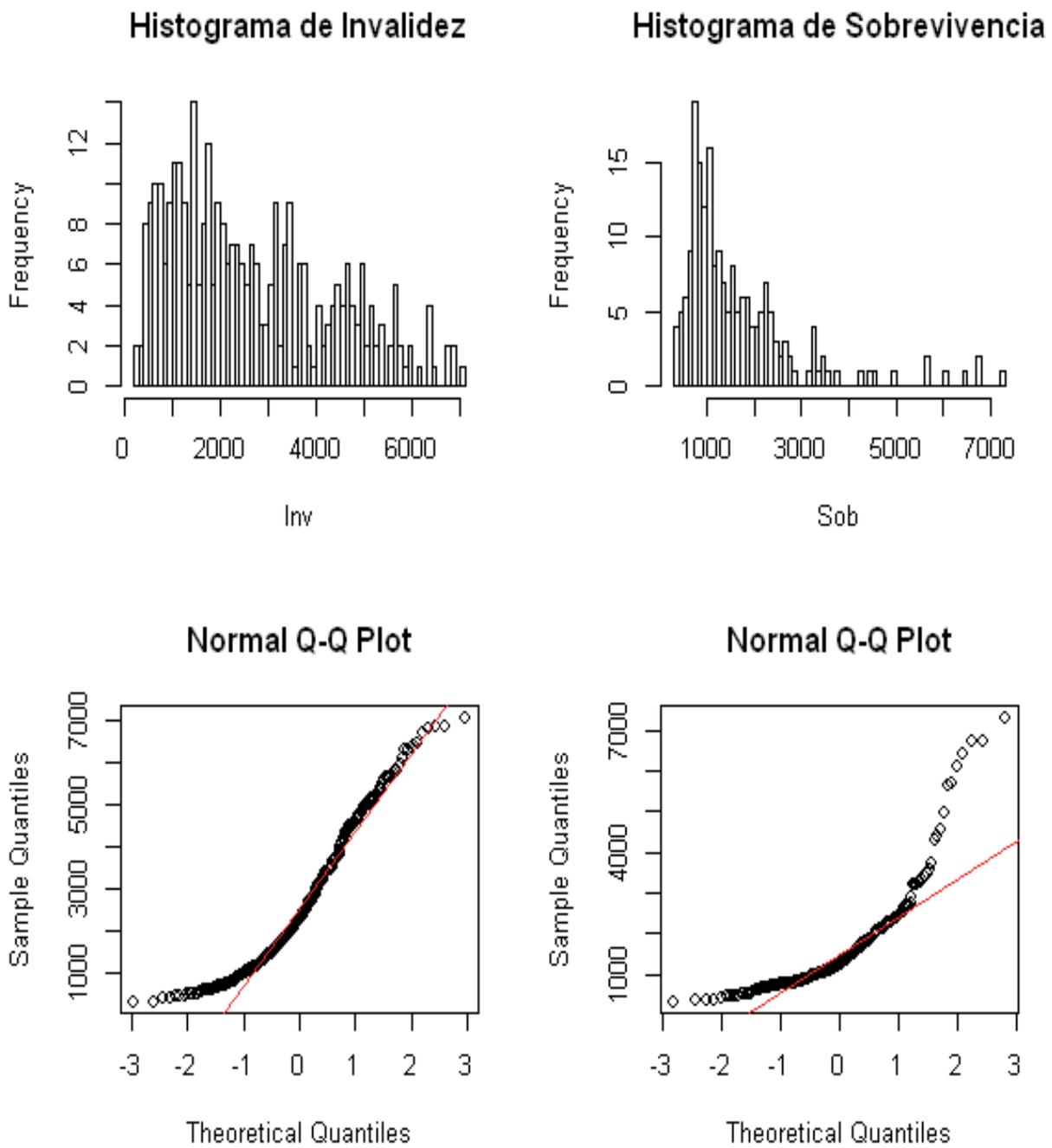


Figura 1. Gráficos descriptivos del coste de los siniestros para los grupos de Invalidez y sobrevivencia

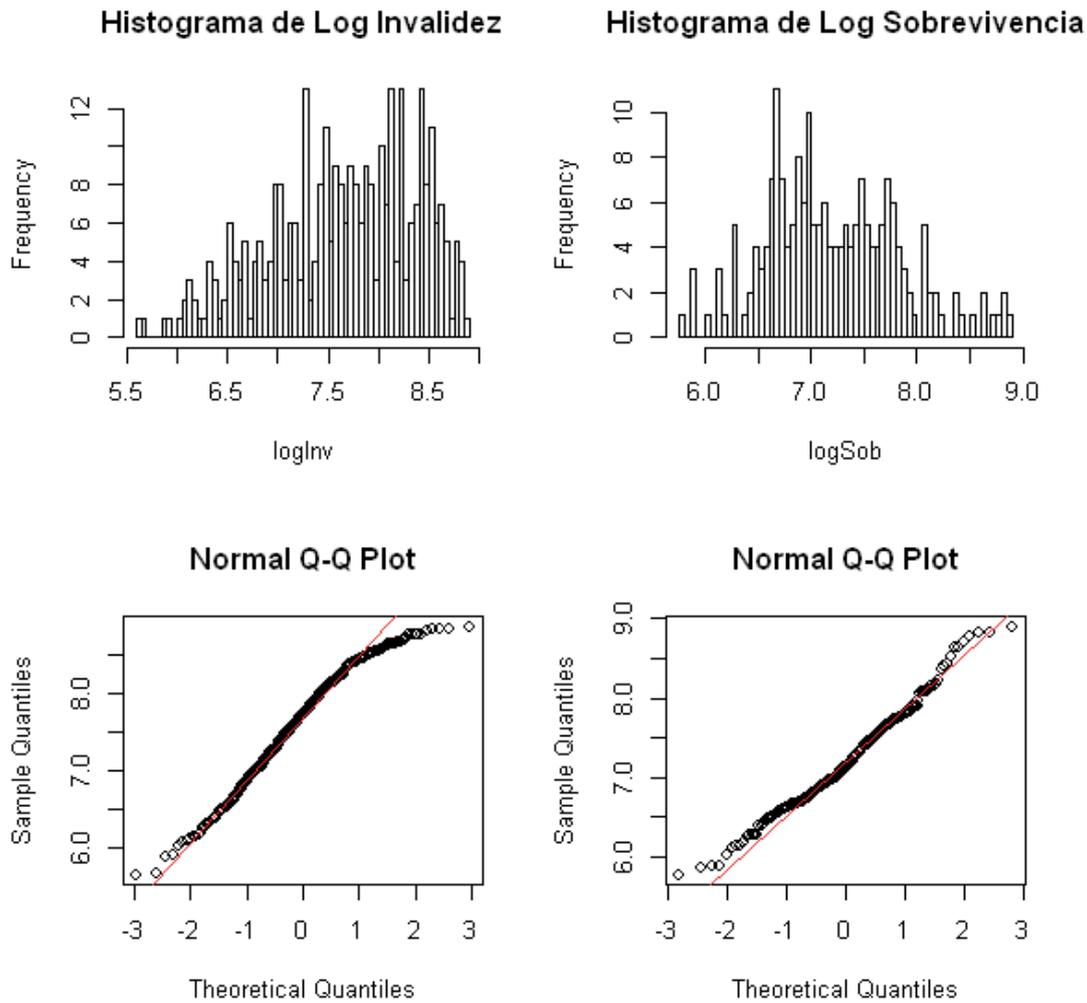


Figura 2. Gráficos descriptivos del logaritmo del coste de los siniestros para los grupos de Invalidez (izquierda) y Sobrevivencia (derecha)

Tabla 3. Parámetros estimados de las distribuciones skew-normal y skew-t

	Invalidez	Sobrevivencia	logInvalidez	logSobrevivencia
Skew-Normal				
Locación	4.5e+02	319.3	8.7	6.6
Escala	2.7e+03	1824.5	1.3	0.9
Forma	2.0e+01	1533.8	-8.7	2.0
Skew-t				
Locación	3.8e+02	497.6	8.7	6.6
Escala	2.8e+03	1041.5	1.3	0.9
Forma	3.0e+01	7.6	-8.7	2.0
Grados de Libertad (gl)	1.6e+71	2.8	17591.4	17183.8

Para el cálculo de la estimación mediante el método de máxima verosimilitud se utilizaron los paquetes *sn* (para las distribuciones *skew-normal* y *skew-t*) y *MASS* (para las demás distribuciones) del *software* estadístico R.

La Tabla 4 presenta los estadísticos para elegir el mejor modelo ubicando al principio los resultados para las distribuciones *skew-normal* y *skew-t* dado que ellas son el centro de este estudio. Luego se presentan los demás modelos, ordenados en primer lugar según el número de parámetros envueltos, y en segundo lugar según el orden alfabético.

Tabla 4. Criterio del Logaritmo de la función de Verosimilitud y Criterio de Información de Akaike (AIC)

Modelo	Simetría	Num_par	Logaritmo de la f. de Verosimilitud				AIC			
			Datos Originales		Log Datos Originales		Datos Originales		Log Datos Originales	
			Invalidez	Sobrevivencia	Invalidez	Sobrevivencia	Invalidez	Sobrevivencia	Invalidez	Sobrevivencia
<i>skew-normal</i>	F	3	-2833	-1624	-341	-186	5672	3253	688	379
<i>skew-t</i>	F	4	-2832	-1607	-341	-186	5672	3222	690	381
exponencial	F	1	-2894	-1658	-989	-586	5790	3317	1980	1174
Cauchy	V	2	-2947	-1654	-434	-226	5898	3312	872	457
Gama	F	2	-9999	-1621	-364	-187	20002	3246	732	379
normal	V	2	-2879	-1685	-359	-189	5762	3375	722	382
logística	V	2	-2887	-1661	-367	-191	5778	3327	739	386
log-normal	F	2	-2851	-1608	-367	-187	5707	3219	739	378
t (no-central)	V	3	-9999	-1645	-9999	-9999	20004	3296	20004	20004
weibull	F	2	-2842	-1632	-348	-205	5687	3269	700	413

Los resultados del cuadro 5 muestran que las distribuciones *skew-normal* y *skew-t* son modelos aptos para ajustar a los conjuntos de datos del estudio. Considerando al estadístico AIC las distribuciones *skew-normal* y *skew-t* tienen los mejores ajustes para los datos del grupo de Invalidez, mientras que para el grupo Sobrevivencia dichas distribuciones se ubican en cuarto y segundo lugar, respectivamente. Considerando el criterio del Logaritmo de la función de Verosimilitud las distribuciones *skew-normal* y *skew-t* se mantienen entre los primeros lugares de las que mejor ajustan a los grupos analizados (Invalidez y Sobrevivencia).

Para analizar y descubrir si los modelos en estudio son apropiados para modelar datos del coste de los siniestros se realiza la prueba de bondad de ajuste Kolmogorov Smirnov.

Los resultados de la Tabla 5 muestran que las distribuciones *skew-normal* y *skew-t* cumplen en la mayoría de los casos, menos en el caso donde la distribución *skew-normal* describe a los datos originales de Sobrevivencia, donde el valor de la prueba (0.124) supera ligeramente al valor crítico (0.097). Por lo cual se rechazaría la hipótesis de que los datos pertenezcan a la distribución *skew-normal*. Sin embargo, observando al logaritmo de los datos originales, se puede apreciar que,

de nuevo, la distribución *skew-normal* es un buen modelo para describir estos datos. Por lo tanto, los resultados de estas pruebas están en concordancia con los resultados del AIC y confirman la habilidad de la distribución *skew-t* y *skew-normal* para describir datos de seguros.

Se utilizaron los parámetros estimados mediante el método de máxima verosimilitud para obtener modelos de cada una de las distribuciones vistas anteriormente, para que con estos modelos se pueda simular y derivar estimadores de VaR y TVaR, y compararlas con los datos empíricos o reales. En la Tabla 6 se presentan estos estimadores que tienen como objetivo complementar los resultados anteriores. Se presentan estimadores de VaR y TVaR con un nivel de confianza del 99%.

Los resultados generados en la Tabla 6 fueron realizados usando datos del coste de siniestros simulados y consideraron un millón de números aleatorios. Los resultados son bastante estables, dado que la convergencia de la simulación de medias, desviaciones estándar, y medidas de riesgo fueron comprobadas.

Tabla 5. Prueba de bondad de ajuste Kolmogorov – Smirnov

Modelo	Datos Originales		Log Datos Originales	
	Invalidez	Sobrevivencia	Invalidez	Sobrevivencia
Valor critico	0.075	0.097	0.075	0.097
skew-normal	0.046	0.124	0.058	0.039
skew-t	0.045	0.049	0.058	0.039
Normal	0.099	0.15	0.07	0.066
exponencial	0.15	0.237	0.533	0.551
log-normal	0.07	0.066	0.078	0.052
logistica	0.092	0.136	0.061	0.06
weibull	0.053	0.108	0.051	0.084

Fuente: Elaboración propia

Tabla 6. Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) con un nivel de confianza al 99%

Modelo	Datos Originales		Log Datos Originales	
	Invalidez	Sobrevivencia	Invalidez	Sobrevivencia
Valor en Riesgo				
skew-normal	7514	5027	8.8	8.9
skew-t	7604	5834	8.8	8.9
Normal	6489	4586	9.3	8.7
exponencial	12143	7659	35.2	33.2
log-normal	11330	5847	9.6	8.8
logistica	6890	4162	9.7	8.8
weibull	7468	5091	9.0	8.6
empírico	6765	6713	8.8	8.8
Valor en Riesgo Condicional				
skew-normal	8384	5608	8.9	9.2
skew-t	8493	7818	8.9	9.2
Normal	7052	5012	9.6	8.9
exponencial	14826	9310	42.8	40.4
log-normal	14998	7410	9.9	9.1
logistica	7840	4757	10.1	9.2
weibull	8381	5792	9.1	8.7
empírico	6886	7016	8.8	8.9

En general, el VaR y el TVaR no tienen un buen performance cuando se consideran los datos originales. Sin embargo, los resultados lucen mucho mejor cuando se consideran al logaritmo natural de los datos originales del coste de los siniestros. Por ejemplo, para las distribuciones *skew-normal*, *skew-t* y normal, los estimadores teóricos tanto del VaR como del TVaR son relativamente cercanos al estimador empírico o real y por lo tanto ajustan a los datos bastante bien. En cambio, otras distribuciones, como por ejemplo la distribución log-normal proporciona altos valores para las medidas de riesgo (VaR y TVaR), lo cual enfatiza que es más extrema en las colas que las distribuciones en estudio de la presente investigación.

La Figura 3 presenta el VaR y el TVaR con percentiles del 90% al 99.9%. La línea discontinua de color plomo representa los valores observados empíricos o reales para el VaR y el TVaR, la línea continua de color plomo representa a los estimadores correspondientes a la distribución *skew-normal*, la línea discontinua de color

negro a la *skew-t*, y la línea continua de color negro a la normal.

En el caso de los VaR del logaritmo del coste de los siniestros las distribuciones *skew-normal* y *skew-t* se ajustan bien a los datos de ambos grupos (Invalidez y Sobrevivencia); situación que no sucede en el caso de la distribución normal que sobrestima a los riesgos empíricos o reales del grupo de Invalidez (Figura 7a) y subestima a estas mismas medidas de riesgo del grupo de Sobrevivencia (Figura 7b).

En cambio, si consideramos a los TVaR del logaritmo del coste de los siniestros, tanto la distribución normal como las distribuciones asimétricas en estudio se ajustan bien a los datos del grupo de Invalidez (Figura 7c), mientras que en el grupo de Sobrevivencia las distribuciones *skew-normal* y *skew-t* sobreestiman a estos riesgos empíricos o reales en la cola derecha y la distribución normal subestima a estas medidas de riesgo en todos sus niveles de confianza.

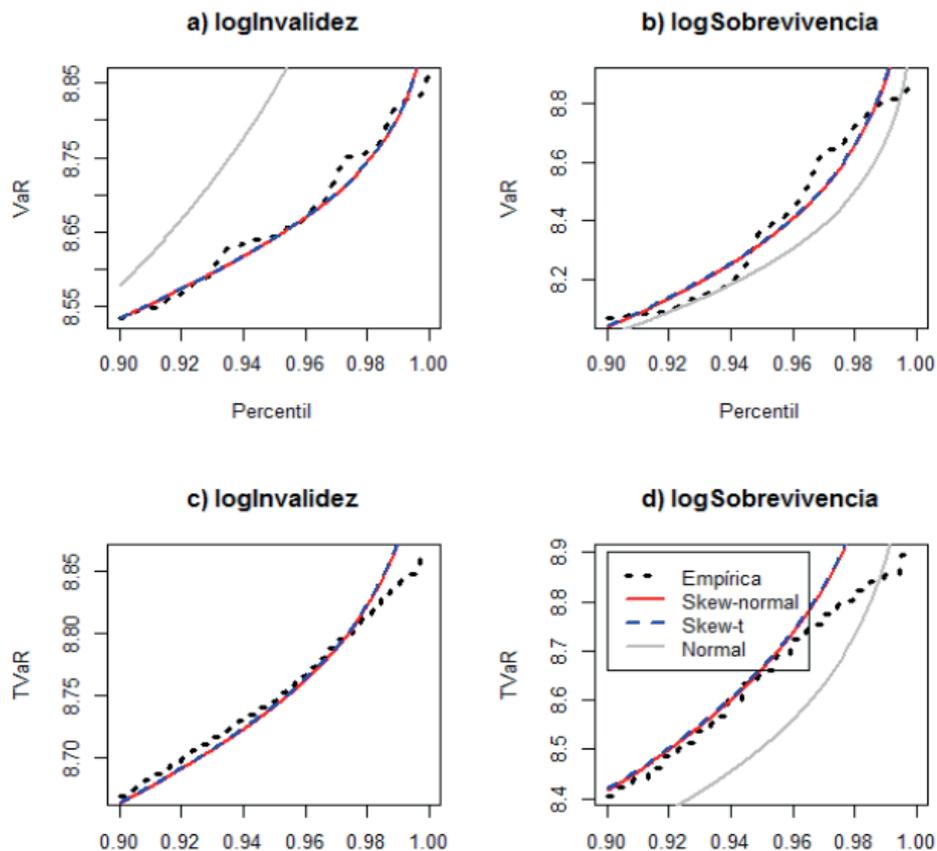


Figura 3. Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) con niveles de confianza entre 90% y 99.9%

4. Conclusiones

Se comprobó que las distribuciones asimétricas skew-normal y skew- t reflejan mejor el comportamiento de los datos de estudio al considerar sus valores extremos. En conjunto, los indicadores usados determinaron que las distribuciones skew-normal y skew- t son las que mejor describen al logaritmo del coste de los siniestros, tanto en las coberturas de seguros de Invalidez como de Sobrevivencia. Para los dos grupos de datos estudiados, la distribución skew-normal se ajusta mejor en comparación con la distribución skew- t . Los resultados confirman las ideas planteadas por Bolance et al. (2008) y Eling (2011) en sus publicaciones, que plantean la utilidad de las distribuciones estudiadas en el área de seguros.

5. Literatura citada

- Akaike, H. (1974).** A new look at the statistical model identification, *IEEE Transactions on Automatic Control* AC-19, 716-723.
- Asociación Peruana de Empresas de Seguros. Orientación. (2012).** Recuperado de <http://www.apeseg.org.pe/orientacion.html>.
- Azzalini, A. (1985).** A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12: 171-178.
- Azzalini, A. y Capitanó, A. (2003).** Distributions

generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of Statistical Society. Series B*, 65: 367-389.

Bolance, C., Guillen, M., Pelican, E., and Vernic, R., (2008). Skewed bivariate models and nonparametric estimation for the CTE risk measure, *Insurance: Mathematics and Economics*, 43(3), 386-393.

Eling, M. (2011). Fitting Insurance Claims to Skewed Distribution: Are the Skew-Normal and Skew-Student Good Models?. Working Papers on Risk Management and Insurance N° 98. Institute of Insurance Economics. University of St. Gallen.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27 (7), 859-870.

Sarabia, J.; Gómez, E. y Vázquez, F. (2007). Estadística Actuarial. Teoría y Aplicaciones. España: Editorial Pearson Educación, S.A.

Superintendencia de Banca, Seguro y AFP. Preguntas Frecuentes. (2012). Recuperado de http://www.sbs.gob.pe/0/modulos/JER/JER_Interna.aspx?ARE=0&PFL=0&JER=2170

Vernic, R. (2006). Multivariate skew-normal distribution with applications in insurance. *Insurance: Mathematics and Economics*, 38: 413-426.