

**UNIVERSIDAD NACIONAL AGRARIA  
LA MOLINA**

**FACULTAD DE ECONOMÍA Y PLANIFICACIÓN  
ESPECIALIDAD DE ESTADÍSTICA E INFORMÁTICA**



**“ANÁLISIS DEL COSTE DE LOS SINIESTROS EN UNA COMPAÑÍA  
DE SEGUROS UTILIZANDO LAS DISTRIBUCIONES  
ASIMÉTRICAS SKEW-NORMAL Y SKEW-T”**

**Presentado por:**

**DIEGO ALONSO MENDOZA QUEVEDO**

**TESIS PARA OPTAR EL TÍTULO DE  
INGENIERO ESTADÍSTICO E INFORMÁTICO**

**Lima – Perú**

**2013**

## **Dedicatoria**

A mi tía Vicky, por todas sus buenas enseñanzas.

Que descanses en paz.

## **Agradecimiento**

En primer lugar a mi madre, por todo el amor y apoyo que me ha brindado para la ejecución de este proyecto. En segundo lugar a mi hermano, que sin quererlo, me ha motivado indirectamente a terminar mi tesis. Y en tercer lugar a todas aquellas personas que de alguna u otra manera pusieron su granito de arena para poder cumplir este objetivo. Gracias!

## Índice general

|     |                                                                                    |    |
|-----|------------------------------------------------------------------------------------|----|
| I.  | Introducción                                                                       | 1  |
| II. | Revisión de Literatura                                                             | 2  |
|     | 2.1 Sistema de Seguros                                                             | 3  |
|     | 2.1.1 Estadística Actuarial                                                        | 3  |
|     | 2.1.2 Clases de Seguros                                                            | 4  |
|     | a. Seguro patrimoniales o seguros de daños                                         | 4  |
|     | b. Seguros de vida                                                                 | 4  |
|     | 2.1.3 Seguro Complementario de Trabajo en Riesgo (SCTR)                            | 5  |
|     | 2.1.4 Funcionamiento del Sistema de Seguros                                        | 6  |
|     | a. Número de los siniestros                                                        | 6  |
|     | b. Coste de los siniestros                                                         | 7  |
|     | 2.1.5 Valor en Riesgo ( <i>VaR</i> ) y Valor en Riesgo Condicional ( <i>TVaR</i> ) | 7  |
|     | 2.2 Asimetría                                                                      | 8  |
|     | 2.2.1 Simetría de una distribución                                                 | 8  |
|     | 2.2.2 Asimetría de una distribución                                                | 10 |
|     | 2.2.3 Verificación de la asimetría                                                 | 12 |
|     | 2.3 Distribución <i>Skew-Normal</i>                                                | 18 |
|     | 2.3.1 Introducción                                                                 | 18 |
|     | 2.3.2 Definición y Propiedades                                                     | 19 |
|     | 2.3.2.1 Definición                                                                 | 19 |
|     | 2.3.2.2 Propiedades                                                                | 20 |
|     | 2.3.3 Distribución <i>skew-normal</i> con parámetros de locación<br>escala         | 21 |
|     | 2.3.3.1 Definición                                                                 | 21 |
|     | 2.3.3.2 Propiedades                                                                | 22 |
|     | 2.3.4 Generación de una distribución <i>skew-normal</i>                            | 23 |
|     | 2.3.5 Inferencia sobre los parámetros de la distribución <i>skew-</i>              | 24 |

|         |                                                                             |    |
|---------|-----------------------------------------------------------------------------|----|
|         | <i>normal</i>                                                               |    |
|         | 2.3.5.1 Método de los momentos                                              | 24 |
|         | 2.3.5.2 Método de Máxima Verosimilitud                                      | 26 |
| 2.4     | Distribución <i>Skew-t</i>                                                  | 30 |
| 2.4.1   | Introducción                                                                | 30 |
| 2.4.2   | Definición y Propiedades                                                    | 31 |
| 2.4.2.1 | Definición                                                                  | 31 |
| 2.4.2.2 | Propiedades                                                                 | 33 |
| 2.4.3   | Generación de una distribución <i>skew-t</i>                                | 34 |
| 2.4.4   | Inferencia clásica                                                          | 36 |
| 2.4.4.1 | Método de los momentos                                                      | 36 |
| 2.4.4.2 | Método de Máxima Verosimilitud                                              | 37 |
| 2.5     | Pruebas de Bondad de Ajuste                                                 | 39 |
| 2.5.1   | Prueba Kolmogorov-Smirnov                                                   | 39 |
| 2.6     | Criterios de Selección del Mejor Modelo                                     | 41 |
| 2.6.1   | Criterio de Información de Akaike (AIC)                                     | 41 |
| III.    | Materiales y Métodos                                                        | 45 |
| 3.1     | Materiales                                                                  | 45 |
| 3.2     | Metodología de la Investigación                                             | 46 |
| 3.2.1   | Tipo de Investigación                                                       | 46 |
| 3.2.2   | Diseño de investigación                                                     | 46 |
| 3.2.3   | Formulación de hipótesis                                                    | 46 |
| 3.2.4   | Identificación de las variables                                             | 47 |
| 3.2.5   | Población y muestra                                                         | 47 |
| 3.2.6   | Metodología aplicada                                                        | 47 |
| IV.     | Resultados y Discusión                                                      | 49 |
| 4.1     | Análisis Estadístico                                                        | 49 |
| 4.1.1   | Descripción de los conjuntos de datos                                       | 49 |
| 4.1.2   | Análisis de los datos de Invalidez y Supervivencia                          | 49 |
| 4.1.3   | Análisis del logaritmo natural de los datos de Invalidez<br>y Supervivencia | 52 |

|       |                                                                                                      |    |
|-------|------------------------------------------------------------------------------------------------------|----|
| 4.2   | Estimación de los parámetros de las distribuciones <i>skew-normal</i> y <i>skew-t</i>                | 54 |
| 4.3   | Comparación de los modelos usando el criterio del logaritmo de la función de verosimilitud y el AIC. | 55 |
| 4.4   | Pruebas de Bondad de Ajuste Kolmogorov – Smirnov                                                     | 57 |
| 4.5   | Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR)                                           | 59 |
| V.    | Conclusiones                                                                                         | 63 |
| VI.   | Recomendaciones                                                                                      | 64 |
| VII.  | Referencias Bibliográficas                                                                           | 65 |
| VIII. | Anexos                                                                                               | 67 |

## Índice de cuadros

|                                                                                                                                                 |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Cuadro 1: Interpretación de Resultados de los Coeficientes de Asimetría.                                                                        | 17 |
| Cuadro 2: Estadísticas descriptivas del coste de los siniestros para los grupos de Invalidez y Supervivencia                                    | 50 |
| Cuadro 3: Estadísticas descriptivas del logaritmo del coste de los siniestros para los grupos de Invalidez y Supervivencia                      | 52 |
| Cuadro 4: Parámetros estimados de las distribuciones <i>skew-normal</i> y <i>skew-t</i> .                                                       | 54 |
| Cuadro 5: Criterio del Logaritmo de la función de Verosimilitud y Criterio de Información de Akaike (AIC).                                      | 56 |
| Cuadro 6: Prueba de bondad de ajuste Kolmogorov – Smirnov.                                                                                      | 58 |
| Cuadro 7: Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) usando el percentil 99%                                                    | 60 |
| Cuadro 8: Parámetros estimados de las distribuciones <i>exponencial</i> , <i>normal</i> , <i>logística</i> , <i>log-normal</i> y <i>weibull</i> | 80 |

## Índice de figuras

|                                                                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1: Ejemplo de una distribución simétrica.                                                                                                    | 9  |
| Figura 2: Tipos de asimetrías presentados por una distribución.                                                                                     | 11 |
| Figura 3: Gráfico de la función de distribución <i>skew-normal</i> estándar con valores diferentes de $\lambda$ .                                   | 20 |
| Figura 4: Gráficos de las funciones de densidad de la distribución <i>skew-t</i> : (a) Parámetro de asimetría fijo. (b) Parámetro de curtosis fijo. | 32 |
| Figura 5: Gráficos descriptivos del coste de los siniestros para los grupos de Invalidez (izquierda) y Supervivencia (derecha).                     | 51 |
| Figura 6: Gráficos descriptivos del logaritmo del coste de los siniestros para los grupos de Invalidez (izquierda) y Supervivencia (derecha).       | 53 |
| Figura 7: Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) con percentiles del 90% al 99.9%                                               | 62 |



## 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:** *Distribución asimétrica, skew-normal, skew-t, coste, siniestro, seguros.*

## **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:** *Asymmetric Distribution, skew-normal, skew-t, cost, claim, insurance.*

## I. INTRODUCCION

En muchos campos de estudio, la distribución normal es el modelo simétrico que mejor se ajusta a los datos recolectados; por ejemplo, en economía y finanzas es el modelo más popular. Sin embargo, la distribución normal no siempre es el modelo más apropiado para describir conjuntos de datos, como en el campo de los seguros donde las distribuciones asimétricas se ajustan mejor a los datos relacionados al coste de los siniestros<sup>12</sup> (Eling, 2011).

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

Debido a que existen muy pocas investigaciones a la fecha que usen las distribuciones *skew-normal* y *skew-t* para ajustar datos de reclamos<sup>3</sup> en seguros, se sugiere la necesidad de realizar más estudios, utilizando nuevos conjuntos de datos de empresas aseguradoras, que justifiquen el uso de las distribuciones mencionadas para

---

<sup>1</sup> El coste de los siniestros es la remuneración que paga la empresa aseguradora a sus asegurados expresado generalmente en moneda (en nuestro estudio es en nuevos soles) cuando estos últimos tienen un siniestro.

<sup>2</sup> Según Sarabia Gómez y Vázquez (2007) un *siniestro* es un suceso que produce un daño previsto en el contrato de seguro y que da lugar al cumplimiento de las obligaciones contraídas por la aseguradora mediante la reposición del bien o la indemnización al asegurado.

<sup>3</sup> Se hablará indistintamente de siniestro o reclamación, aunque en la práctica pudiera ocurrir un siniestro sin que se dé parte del mismo a la compañía aseguradora (Artís et. al., citado por Sarabia et al. 2007).

modelar datos de siniestros en las compañías de seguros (Bolance et. al., citado por Eling 2011).

Por lo tanto, se espera determinar silas distribuciones *skew-normal* y *skew-tson* modelos que mejor describen el coste de los siniestros en una compañía de seguros<sup>4</sup>, con el fin de lograr que la empresa pueda utilizar estos modelos como herramientas para prever modificaciones en sus coberturas (en el presente se analizan dos conjuntos de datos de dostipos de coberturas: Invalidez y Sobrevivencia) y pueda valorar los beneficios que le reporta el seguro (en el presente estudio se analiza el Seguro Complementario de Trabajo en Riesgo (SCTR)).

Para identificar los mejores modelos se compararon las distribuciones asimétricas mencionadas con otras distribuciones tradicionales, empleando el Criterio de Información de Akaike (AIC) y el Criterio del Logaritmo de la función de Verosimilitud, complementado ambos resultados con la prueba de bondad de ajuste Kolmogorov – Smirnov. Además, se simulan datos de los mejores modelos y se comparan con los datos empíricos o reales mediante medidas de riesgo estimadas como el Valor en Riesgo (VaR) y el Valor en Riesgo Condicional (TVaR), con la idea de reafirmar los resultados anteriores.

Por último, el estudio es relevante porque servirá de referencia para futuras investigaciones estadísticas e informáticas aplicadas en finanzas o en actuarial<sup>5</sup>, que la Universidad Nacional Agraria La Molina, específicamente la especialidad de Estadística e Informática, abordará en su nueva malla curricular<sup>6</sup>, manteniendo así una educación de calidad, acorde a las demandas del mercado.

---

<sup>4</sup> El nombre de la compañía de seguros no se revelará por razones de confidencialidad.

<sup>5</sup> La *Estadística Actuarial* es la ciencia que proporciona las herramientas necesarias para el estudio de ciertas actividades económicas que se llevan a cabo en las compañías de seguros (Saravia et. al., 2008)

<sup>6</sup> La especialidad de Estadística e Informática de la Facultad de Economía y Planificación de la UNALM ha incorporado en su malla de estudios que regirá a partir del año 2012-I, el curso de Estadística Actuarial. Ver ANEXO 1.

## II. REVISIÓN DE LITERATURA

### 2.1 Sistema de Seguros

Según la Asociación Peruana de Seguros (APESEG), 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.

El Seguro constituye una solución a la necesidad que sentimos de vernos protegidos ante la ocurrencia de hechos imprevistos, cuyas consecuencias superen nuestra capacidad individual para repararlas (APESEG).

#### 2.1.1 Estadística Actuarial<sup>7</sup>

Saravia, Gómez y Vázquez (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.

---

<sup>7</sup> Ver Terminología Actuarial en ANEXO 2

### **2.1.2 Clases de Seguros**

Según la SBS se tienen las siguientes clases:

#### **a. Seguros patrimoniales o seguros de daños**

Bajo esta clasificación de seguros, se recogen todos los seguros cuyo fin es el de reparar una pérdida sufrida en el patrimonio del tomador del seguro, quien usualmente es además el asegurado. Dentro de esta clasificación, se encuentran los seguros destinados a resarcir al asegurado de una pérdida sufrida directamente en sus bienes y los seguros de responsabilidad, que garantizan al asegurado contra la responsabilidad civil en que pueda incurrir ante terceros, por actos de los que sea responsable.

#### **b. Seguros de vida**

En esta clase de seguros, el objeto del seguro es la vida del asegurado, y el pago de la indemnización pactada en el contrato de seguros, depende de la muerte o de la sobrevivencia del asegurado a una fecha determinada. Los seguros de vida pueden ser contratados por toda la vida del asegurado o por un periodo de tiempo determinado. En el primer caso se trata de un seguro de vida universal o un seguro de vida entera y la aseguradora pagará la indemnización a la muerte del asegurado, siempre que a dicha fecha el contrato se encuentre vigente. En el segundo caso se trata de un seguro de vida temporal, en el que la aseguradora pagará la indemnización acordada si el asegurado fallece dentro de la vigencia del contrato, si sobrevive a dicho plazo no habrá lugar al pago de indemnización.

### **2.1.3 Seguro Complementario de Trabajo en Riesgo (SCTR)<sup>8</sup>**

Es el tipo de seguro que se analizó en la presente investigación y según la SBS 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 descritas en el Anexo 5 del Decreto Supremo N° 009-97-SA, Reglamento de la Ley de Modernización de la Seguridad Social en Salud<sup>9</sup>.

Son asegurados obligatorios del SCTR, la totalidad de los trabajadores del centro de trabajo en el cual se desarrollan las actividades de riesgo previstas en el Anexo 5 de dicho Decreto Supremo, sean empleados u obreros, sean eventuales, temporales o permanentes.

La Cobertura de Invalidez, Muerte y Sepelio, que protege obligatoriamente al asegurado o sus beneficiarios contra los riesgos de invalidez o muerte producida como consecuencia de accidente de trabajo o enfermedad profesional; otorga las siguientes prestaciones mínimas:

- Pensión de Sobrevivencia
- Pensión de Invalidez
- Cobertura de los Gastos de Sepelio

Esta cobertura la contrata el empleador con una Compañía de Seguros o la Oficina de Normalización Previsional (ONP). Para nuestro estudio, se analizarán los datos de una aseguradora con respecto a la remuneración base que reciben los asegurados o beneficiarios por concepto de las coberturas de Invalidez y Sobrevivencia del SCTR.

---

<sup>8</sup> Información de la página web de la SBS.

<sup>9</sup> Ver ANEXO 3



#### **2.1.4 Funcionamiento del Sistema de Seguros**

Según Bolance et. al (2007) los factores que influyen en el acontecimiento de un suceso contra el cual un bien puede asegurarse pueden ser múltiples y diversos. Por ejemplo, en el seguro de automóviles: un siniestro puede ocurrir debido a causas como la edad del conductor, antigüedad del vehículo, cilindrada del mismo, etc., en cualquiera de los casos, el asegurado paga una cantidad de dinero, la prima de seguro al asegurador, la compañía de seguros, y desde ese instante comienza la cobertura sobre el bien asegurado. Desde este punto de vista, el negocio del seguro no se diferencia, en principio, de cualquier otro tipo de negocio. La empresa aseguradora vende un producto que es la seguridad, a cambio de un precio que es la prima. La seguridad, que es el producto que se comercializa, se garantiza a todos los clientes, aunque la entrega sólo se efectúa a una parte de ellos. El tiempo además es un factor que juega a favor del asegurador, puesto que las indemnizaciones se reparten, posponiéndose y dando lugar, entretanto, a un monto de ahorro que forman las provisiones técnicas o reservas. Luego, desde un punto de vista financiero, el tomador de seguros es el prestamista que proporciona el crédito al asegurador para que fabrique el producto (la seguridad), convirtiéndose el asegurador en un colocador de los fondos que no se consumen periódicamente de entre todos aquellos que le han sido prestados.

Sin embargo, en el negocio del seguro aparecen incertidumbres para el asegurador que no suelen aparecer en otro tipo de negocio. Según Ferreira y Garín (2010) estas incertidumbres también llamadas variables aleatorias, generalmente son de dos tipos: **el número de siniestros y el coste de los mismos.**

##### **a. Número de los siniestros**

El comportamiento del número de siniestros nos dará idea de la frecuencia con la que se producen dichos siniestros.

**b. Coste de los siniestros**

El coste de cada siniestro, nos ayuda a valorar en términos económicos la ocurrencia de los mismos.

Esta última, es la variable aleatoria que se analizará en la presente investigación.

**2.1.5 Valor en Riesgo (*VaR*) y Valor en Riesgo Condicional (*TVaR*)**

**a. Valor en riesgo (*VaR* – *Value at Risk*):**

Cuantil de la función de distribución que deja un determinado porcentaje (por ejemplo al 95% o 99%) a su izquierda y el resto a la derecha.

**b. Valor en Riesgo Condicional (*TVaR* – *Tail Value at Risk*):**

Media aritmética de los valores que superan al valor en *VaR* (por ejemplo al 95% o 99%).

Actuarialmente, ésta última se considera mejor medida del riesgo que el *VaR*. La ventaja de usar la medida del *TVaR* como una medida del riesgos que proporciona la suma total de pérdidas por encima de un determinado percentil.

## 2.2 Asimetría

Según Macerau (2012), la palabra asimetría significa ausencia de simetría, es decir, ausencia de armonía, de estética, de proporcionalidad o de equilibrio. En estadística, la simetría o asimetría es observada en la distribución de los datos.

Un factor importante en el análisis de los datos es identificar el tipo de asimetría presentado en la distribución de las observaciones, puesto que cuando se ha identificado, se procura utilizar distribuciones de probabilidad que consigan modelar el comportamiento asimétrico de las observaciones.

En los siguientes ítems se definen simetría y asimetría de una distribución de probabilidad continua y se describen algunas formas para verificar la asimetría de una distribución.

### 2.2.1 Simetría de una distribución

La simetría de una distribución se evalúa en relación a su centralidad, que puede ser su media o su mediana.

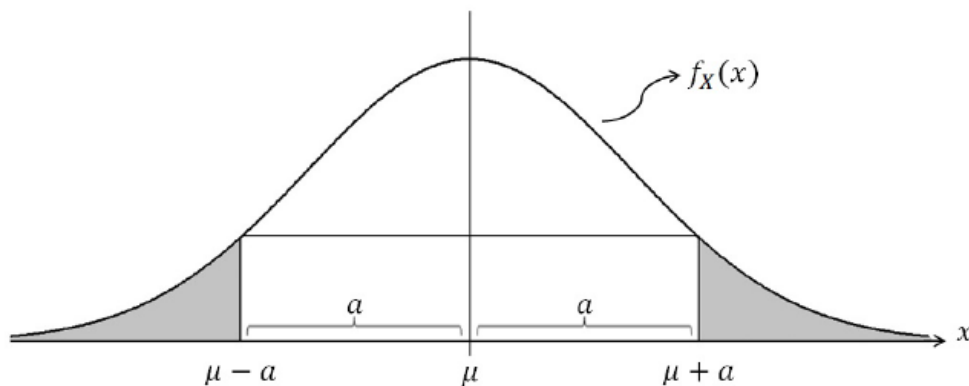
En una distribución simétrica, el área debajo de la curva de la función de densidad de probabilidad,  $f_X(x)$ , contiene la misma área en ambos lados de la media o de la mediana de la distribución, es decir,

$$\int_{-\infty}^{\mu} f_X(x) dx = \int_{\mu}^{\infty} f_X(x) dx \quad \text{o} \quad \int_{-\infty}^{m_d} f_X(x) dx = \int_{m_d}^{\infty} f_X(x) dx, \quad (1)$$

donde  $\mu$  y  $m_d$  son la media y la mediana de la distribución, respectivamente; y  $x$  son los valores que la variable aleatoria continua  $X$  puede tomar, con  $x \in \mathbb{R}$ .

La figura 1 representa un ejemplo de una distribución simétrica, con forma de seno, donde un lado de la distribución es la imagen esperada del otro;  $\mu$  representa la media de la distribución,  $a$  una distancia cualquiera,  $\mu - a$  y  $\mu + a$  son puntos en el eje de los valores de  $x$ , que representan los valores que la variable aleatoria continua  $X$  puede asumir, y  $f_X(x)$  representa la función de densidad de probabilidad, con  $x \in \mathbb{R}$ .

**Figura 1: Ejemplo de una distribución simétrica**



Fuente: Walkiria Macerau (2012)

Se observa a través de la figura 1, que los puntos  $\mu - a$  y  $\mu + a$  son equidistantes en relación a  $\mu$ . En términos de la probabilidad, se puede decir que la probabilidad acumulada por debajo del punto  $\mu - a$  es igual a la probabilidad acumulada por encima del punto  $\mu + a$ , es decir,

$$\Pr(X < (\mu - a)) = \Pr(X > (\mu + a)), \quad (2)$$

donde  $\mu - a \in \mathbb{R}$  y  $\mu + a \in \mathbb{R}$ .

Escribiendo de otra forma tenemos,

$$F_X(\mu - a) = 1 - F_X(\mu + a), \quad (3)$$

donde  $F_X(\cdot)$  es la función de distribución acumulada.

### 2.2.2 Asimetría de una distribución

Consideramos como asimétrica cualquier distribución que no tiene la condición de asimetría.

En una distribución asimétrica, el área debajo de la curva de la función de densidad de probabilidad,  $f_X(x)$ , no contiene la misma área en ambos lados de la media o de la mediana de la distribución, y si tomamos los puntos  $\mu - a$  y  $\mu + a$ , que están alejados a una distancia  $a$  de  $\mu$ , tenemos que la probabilidad acumulada por debajo del punto  $\mu - a$  no será igual a la probabilidad acumulada por encima del punto  $\mu + a$ , es decir,

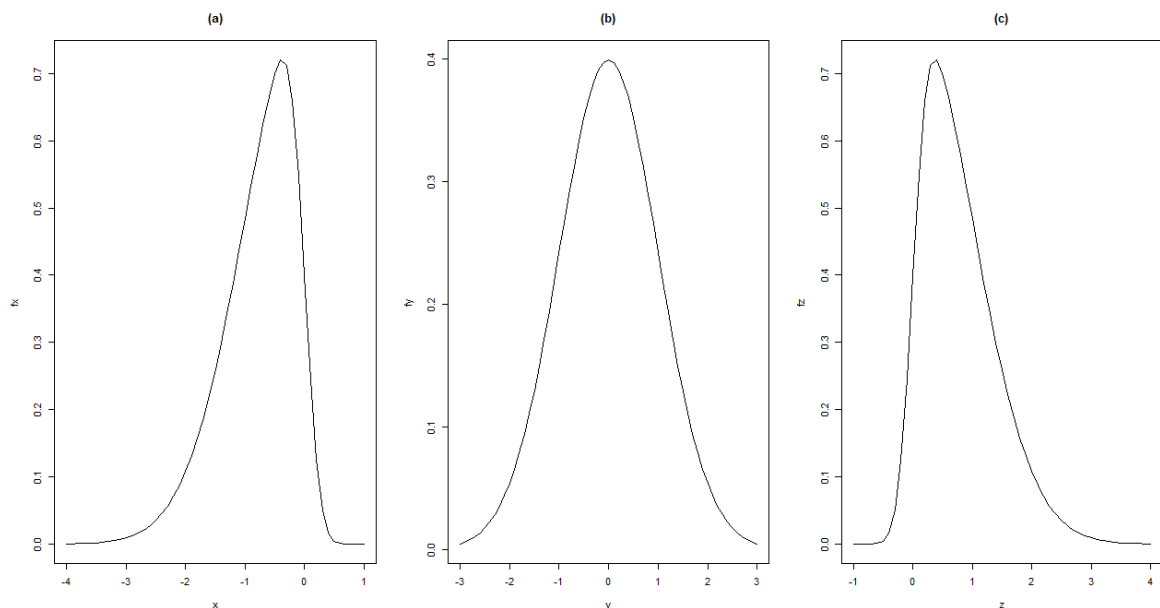
$$\Pr(X < (\mu - a)) \neq \Pr(X > (\mu + a)), \quad (4)$$

La asimetría de una distribución puede ser a la izquierda o a la derecha. La asimetría a la izquierda, indica que la probabilidad en un intervalo del lado izquierdo de la moda de la distribución es menor que la probabilidad en otro intervalo de la misma longitud que el primero situado en el lado derecho de la moda de la distribución. La asimetría a la

derecha, indica que esa probabilidad del lado derecho de la moda de la distribución es menor que la probabilidad del lado izquierdo de la moda de la distribución, suponiendo que la distribución sea unimodal.

La figura 2.2 presenta tres ilustraciones de tipos de asimetría presentados por una distribución, donde (a) representa una distribución asimétrica a la izquierda, (b) una distribución simétrica y (c) representa una distribución asimétrica a la derecha.

**Figura 2: Tipos de asimetrías presentados por una distribución**



Fuente: Elaboración propia<sup>10</sup>

La distribución en la figura 2 (a) posee una asimetría a la izquierda, esto es, la menor masa de probabilidad de la distribución está concentrada en el lado izquierdo del gráfico; la distribución presentada en la figura 2 (c) posee una asimetría a la derecha, esto es, la menor concentración de la distribución está en el lado derecho, y la figura 2 (b) presenta una distribución simétrica.

---

<sup>10</sup> Ver códigos en ANEXO 4

Groeneveld & Meeden (1984), citado por Macerau (2012) han presentado una definición formal de la asimetría. Los autores usan el tercer momento estandarizado para evaluar la asimetría de la distribución de una variable aleatoria  $X$ . Se representa por  $\gamma_1$  y está definida por:

$$\gamma_1 = E \left[ \left( \frac{X - \mu}{\sigma} \right)^3 \right] = \frac{\mu_3}{\sigma^3}, \quad (5)$$

donde  $\mu_3$  es el tercer momento con respecto a la media  $\mu$  y  $\sigma$  es la desviación estándar de la variable aleatoria  $X$ .

Esta medida sólo es válida cuando el tercer momento de  $X$  existe. En caso el tercer momento no exista, recurrimos a otras formas para verificar la asimetría de una distribución.

### 2.2.3 Verificación de la asimetría

Las medidas de tendencia central como la media, la moda y la mediana, se pueden utilizar para verificar la asimetría de una distribución.

Suponga que la distribución tiene la forma de la figura 2 (a), esto es, la distribución asimétrica es a la izquierda, entonces la moda y la mediana se encuentran a la derecha de la media, la cual tiene una tendencia para moverse en la dirección de los puntos extremos. Ahora, si la distribución tiene la forma de la figura 2 (c), esto es, la distribución asimétrica es a la derecha, entonces la moda y la mediana estarán a la izquierda de la media. Y para una distribución que tiene la forma de la figura 2 (b), esto es, la distribución es simétrica, entonces las medidas de tendencia central tienden a ser iguales.

Se puede estudiar la asimetría de una distribución unimodal, es decir, de distribuciones que presentan una única moda, utilizando el siguiente intervalo,

$$\mu - M_o, \quad (6)$$

donde  $\mu$  y  $M_o$  son la media y la moda de la distribución, respectivamente.

Si  $\mu - M_o < 0$ , entonces la asimetría es a la izquierda (o negativa), si  $\mu - M_o = 0$  la asimetría es nula, o la distribución es simétrica, y si  $\mu - M_o > 0$  la asimetría es a la derecha (o positiva).

La relación entre la media y la moda presentada en (2.6) es también denominada la Medida de Asimetría de Pearson. Esta medida posee dos limitaciones en la aplicación. La primera es debido a que es una medida absoluta, entonces el resultado está expresado en términos de la unidad original de la medida de la variable, es decir, el resultado cambia cuando la unidad de medida cambia. La segunda es porque la misma magnitud absoluta de la asimetría tiene diferentes significados para diferentes series de datos con diferentes grados de variabilidad.

Una medida que supera las limitaciones enunciadas por la Medida de Asimetría de Pearson es el Coeficiente de Asimetría de Pearson dado por

$$A_S^{(1P)} = \frac{\mu - M_o}{\sigma} \quad (7)$$

donde  $\sigma$  es la desviación estándar de la distribución.



La expresión (2.7) también se conoce como el Primer Coeficiente de Asimetría de Pearson. La aplicación de dicha expresión envuelve otra dificultad puesto que puede suceder que una distribución presente más de una moda, mientras que una distribución presenta solamente una mediana.

No obstante, existe una relación empírica también atribuida a Pearson entre la media, la mediana y la moda, donde para una distribución moderadamente asimétrica, como las distribuciones con  $0.15 < |A_S^{(1P)}| < 1$ , se tiene:

$$\mu - M_o \approx 3(\mu - M_d), \quad (8)$$

donde  $M_d$  es la mediana de la distribución. Sustituyendo el lado derecho de (8) en (7), se puede reescribir el Coeficiente de Asimetría de Pearson como

$$A_S^{(2P)} \cong \frac{3(\mu - M_d)}{\sigma}, \quad (9)$$

que es conocido como el Segundo Coeficiente de Asimetría de Pearson, según Kenney & Keeping (1962), citado por Macerau (2012).

Si la media o la desviación estándar no están disponibles, se puede calcular la asimetría utilizando los cuartiles, por medio del Coeficiente Cuantil de Asimetría dado por:

$$A_S^{(B)} = \frac{Q_3 - 2M_d + Q_1}{Q_3 - Q_1}, \quad (10)$$

donde  $Q_1$  y  $Q_3$  son el primer y tercer cuartil de la distribución, respectivamente, según Yule (1912) y Bowley (1920), citado por Macerau (2012). El Coeficiente Cuantil es conocido como Coeficiente de Asimetría de Bowley.

Se puede utilizar también el Coeficiente Percentil, conocido como el Coeficiente de Asimetría de Kelley dado por:

$$A_S^{(K)} = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}, \quad (11)$$

donde  $P_{10}$ ,  $P_{50}$  y  $P_{90}$  son el décimo, quincuagésimo y nonésimo percentil, de la distribución de Kelly, respectivamente (Kelly 1924, citado por Macerau 2012).

Otra medida de verificación de la asimetría es el Coeficiente del Momento de Asimetría o Coeficiente de Asimetría de Fisher. Esta medida utiliza el tercer momento central con respecto a la media, y se define por:

$$A_S^{(F)} = \frac{m_3(\mu)}{\sigma^3}, \quad (12)$$

donde  $m_3(\mu)$  es el tercer momento central y  $\sigma$  es la desviación estándar de la distribución, Abramowitz (1972) citado por Macerau (2012). Dicho coeficiente es el mismo que el definido por Groeneveld & Meeden, que se presentó en la ecuación (5).

Una medida de asimetría en términos de la moda fue introducida recientemente por Arnold & Groeneveld (1995), citado por Macerau (2012). Los autores asumen en esa medida que la moda de la distribución existe y es única, y la definen como uno menos dos veces la función de distribución acumulada a la izquierda de la moda, es decir,

$$\gamma = 1 - 2F_X(M_o), \quad (13)$$

la medida satisface  $-1 < \gamma < 1$ , siendo que  $\gamma = 0$  significa que la distribución es simétrica.

La elección de cual coeficiente utilizar depende de la dificultad para calcular las medidas estadísticas y del comportamiento del conjunto de observaciones que se está analizando. El cuadro 2.1 presenta una interpretación simple del resultado para los coeficientes de Pearson (primero y segundo), Bowley, Kelley y Fisher, y también para la medida de asimetría de Arnold & Groeneveld (1995).

**Cuadro 1: Interpretación de Resultados de los Coeficientes de Asimetría**

| <b>Resultado</b>  | <b>Distribución</b>       |
|-------------------|---------------------------|
| $ A_S^{(*)}  < 0$ | Asimétrica a la izquierda |
| $ A_S^{(*)}  = 0$ | Simétrica                 |
| $ A_S^{(*)}  > 0$ | Asimétrica a la derecha   |

(\*) indica los coeficientes de Pearson (1° y 2°), Bowley, Kelley y Fisher.

Fuente: Walkiria Macerau (2012)

El resultado del Primer Coeficiente de Asimetría de Pearson también permite calcular el grado de asimetría de una distribución. Si el valor absoluto del resultado es  $|A_S^{(1P)}| < 0.15$  la asimetría de la distribución es considerada débil, si  $0.15 \leq |A_S^{(1P)}| < 1$  la asimetría es considerada moderada, y si  $|A_S^{(1P)}| \geq 1$  la asimetría es considerada fuerte.

## 2.3 Distribución *Skew-Normal*

### 2.3.1 Introducción

El análisis estadístico para el estudio de los datos continuos ha sido desarrollado, en gran parte, con base en el modelo normal. Sin embargo, el supuesto de simetría de los datos puede hacer que sean hechas inferencias poco apropiadas sobre los parámetros de interés. Una alternativa para esta situación sería flexibilizar el supuesto de normalidad, considerando una clase de distribuciones asimétricas que poseen como un caso particular a la distribución normal (Gomes, 2009).

A pesar que la distribución normal sea el modelo de probabilidad más utilizado en estadística, en la práctica existen muchos fenómenos que no pueden ser descritos por la distribución normal, ni tampoco por otras distribuciones simétricas. Estos fenómenos se encuentran en diversas áreas como ciencias actuariales, finanzas, economía, demografía, ingeniería, ciencia ambiental y ciencias médicas, entre otros; donde se tiene que realizar aplicaciones de las distribuciones asimétricas (Johnson *et al.* 1994 y 1995, Seshadri 1999, citado por Gomes 2009).

La distribución *skew-normal* (normal asimétrica traducida al español) fue formalmente introducida por Azzalini en 1985. A partir de entonces, las propiedades probabilísticas de esta distribución fueron ampliamente estudiadas. Por ejemplo en un estudio que realizó Arrellano-Valle & Azzalini el 2006 (citado por Gomes 2009) se presenta la parametrización centrada de esa distribución, corrigiendo el problema de singularidad de la matriz de información de Fisher cuando el parámetro de asimetría es nulo. Además, el artículo muestra la correlación (unificación) de las diversas propuestas originadas durante los años, para el caso multivariado de la distribución *skew-normal*.

En esta investigación se presenta una revisión de la distribución *skew-normal* univariada, así como sus propiedades y los métodos de inferencia para los parámetros de la distribución. También se considera otra representación hecha por Henze

(1986), el cual obtiene la densidad propuesta por Azzalini a través de la transformación de variables.

## 2.3.2 Definición y Propiedades

### 2.3.2.1 Definición

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

$$f(z) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R} \quad (14)$$

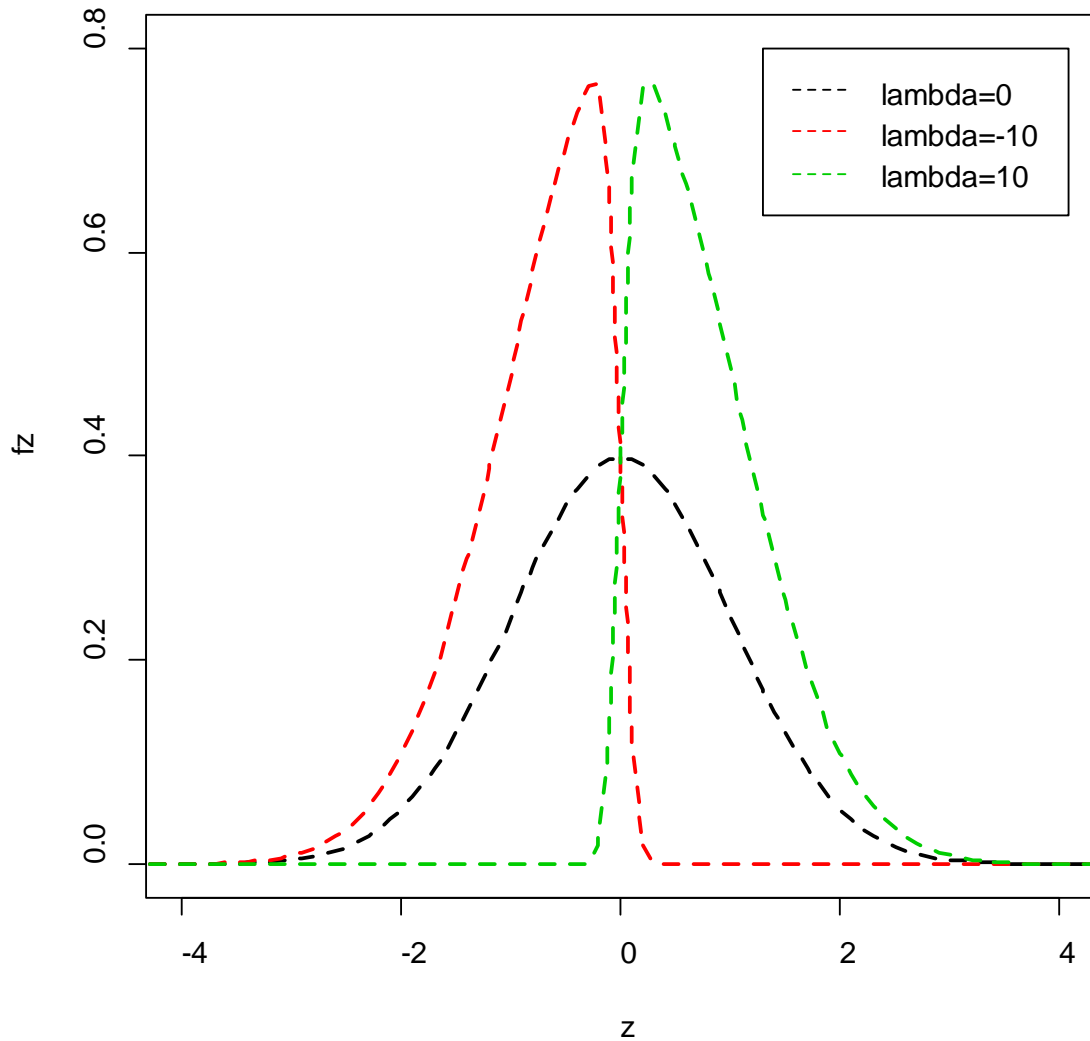
donde  $\lambda$  es un valor definido en  $\mathbb{R}$ ,  $\phi(\cdot)$  y  $\Phi(\cdot)$  denotan la función de densidad de probabilidad y la función de distribución acumulada de la distribución normal estándar, respectivamente.

La ecuación (14) muestra el comportamiento de  $f(z)$  para diferentes valores de  $\lambda$ , y es fácil notar que la distribución normal estándar está incluida como caso particular de la *skew-normal* estándar cuando  $\lambda = 0$ .

La función de distribución acumulada (f. d. a.) de la *skew-normal* estándar, denotada por  $F(z; \lambda)$ , es dada por

$$F(z; \lambda) = 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t)\Phi(u) du dt. \quad (15)$$

**Figura 3: Gráfico de la función de distribución *skew-normal* estándar con valores diferentes de  $\lambda$**



Fuente: Elaboración propia<sup>11</sup>

### 2.3.2.2 Propiedades

- a.  $SN(0)$  es igual a  $N(0,1)$ .
- b. Si  $Z \sim SN(\lambda)$  entonces  $Y = |Z| \sim HN(0,1)$ , donde  $HN(0,1)$  representa la distribución *half-normal* con densidad dada por  $f(y) = 2 \phi(y)I_{[y>0]}(y)$ .

<sup>11</sup> Ver códigos en ANEXO 5

- c. Si  $Z \sim SN(\lambda)$  entonces cuando  $\lambda \rightarrow \infty$ ,  $Z \xrightarrow{D} HN(0,1)$ . Donde  $\xrightarrow{D}$  indica convergencia en distribución.
- d. Si  $Z \sim SN(\lambda)$  entonces  $-Z \sim SN(-\lambda)$ .
- e. Si  $Z \sim SN(\lambda)$  entonces la f. d. p. de  $Z$  es unimodal y  $\log f(z)$  es una función cóncava.
- f.  $1 - F(-z, \lambda) = F(z, -\lambda)$ .
- g.  $F(z, 1) = \{\Phi(z)\}^2$ .
- h. Si  $Z \sim SN(\lambda)$  entonces  $Z^2 \sim \chi_1^2$ .

### 2.3.3 Distribución *skew-normal* con parámetros de locación escala

#### 2.3.3.1 Definición

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

$$f(y) = 2 \frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\lambda \left(\frac{y - \mu}{\sigma}\right)\right), y \in \mathbb{R} \quad (16)$$

donde  $\phi(\cdot)$  y  $\Phi(\cdot)$  denotan la función de densidad de probabilidad y la función de distribución acumulada de la distribución normal estándar, respectivamente.



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

### 2.3.3.2 Propiedades

i. Si  $Y \sim SN(\mu, \sigma, \lambda)$  entonces  $X = a + bY \sim SN(a + b\mu, b\sigma, \lambda)$ ,  $a, b \in \mathbb{R}$ .

j. La función generadora de momentos de  $Y$  es dada por

$$M_y(t) = 2 \exp\left(\frac{(t - \mu)^2}{2\sigma^2}\right) \Phi\left(\delta \left(\frac{y - \mu}{\sigma}\right)\right), \quad (17)$$

donde  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ .

k. La media de  $Y$  es dada por

$$E(Y) = \mu + \sigma\delta \sqrt{\frac{2}{\pi}} \quad (18)$$

l. La varianza de  $Y$  es dada por

$$Var(Y) = \sigma^2 \left[1 - \frac{2}{\pi} \delta^2\right] \quad (19)$$

- m. El tercer momento y el coeficiente de asimetría de  $Y$  son dados, respectivamente por

$$E(Y^3) = \mu^3 + 3\mu^3\sigma\delta\sqrt{\frac{2}{\pi}} + 3\mu\sigma^2 + 3\sigma^3\delta\sqrt{\frac{2}{\pi}} - \sigma^3\delta^3\sqrt{\frac{2}{\pi}},$$

$$\gamma_1 = \sqrt{\frac{2}{\pi}}\delta^3\left[\frac{4}{\pi} - 1\right]\left[1 - \frac{2}{\pi}\delta^2\right]^{-3/2}$$

donde  $\gamma_1$  asume valores en el intervalo  $(-0.99527, 0.99527)$ .

#### 2.3.4 Generación de una distribución *skew-normal*

Según Gomes (2009) existen diversas formas de obtener la clase de las distribuciones *skew-normal* (Freitas, 2006): construcción por condicionamiento, representación de Henze (1986), estadísticas de orden, la versión de Cartinour (1990) y la versión de Arellano-Valle & Azzalini (2006). La construcción es hecha a partir de una distribución normal bivariada en los tres primeros casos, mientras que Cartinour (1990) parte de una distribución normal multivariada, en la cual todas las componentes del vector aleatorio son truncadas en un intervalo cerrado. A continuación se describe la presentación de Henze (citado por Gomes 2009).

La representación de Henze consiste en construir la distribución *skew-normal* como una combinación lineal de variables aleatorias independientes, y este método es muy eficiente en simulaciones, pues a partir de esa representación se puede implementar fácilmente algoritmos computacionales para la generación de muestras.

Sean  $X_0 \sim N(0,1)$  y  $X_1 \sim N(0,1)$  variables aleatorias independientes,  $\delta \in (-1,1)$  y  $Z = \delta|X_0| + \sqrt{1 - \delta^2}X_1$ . Entonces  $Z \sim SN(\lambda)$ , donde  $\lambda = \frac{\delta}{\sqrt{1-\delta^2}}$ .

### 2.3.5 Inferencia sobre los parámetros de la distribución *skew-normal*

Se presenta dos métodos clásicos de inferencia para los parámetros del modelo *skew-normal*: el método de los momentos y el método de máxima verosimilitud, según Gomes (2009).

#### 2.3.5.1 Método de los Momentos

El método de los momentos consiste en estimar los parámetros poblacionales a través de sistemas de ecuaciones que involucran a los momentos muestrales. Como se vio en la ecuación (17), la función generadora de momentos para una variable  $Y \sim SN(\mu, \sigma, \lambda)$  es dada por

$$M_y(t) = 2 \exp\left(\frac{(t - \mu)^2}{2\sigma^2}\right) \Phi\left(\delta \left(\frac{y - \mu}{\sigma}\right)\right)$$

Considere  $X_1, X_2, \dots, X_n$  una muestra estandarizada de tamaño  $n$  de la variable aleatoria  $Y$

$$X = \frac{Y - \bar{y}}{s}$$

donde  $\bar{y} = \sum_{i=1}^n y_i/n$  y  $s^2 = \sum_{i=1}^n (y_i - \bar{y})^2/n - 1$  son, respectivamente, la media y la varianza muestral de  $Y$ .

Entonces  $X \sim SN(\mu_X, \sigma_X, \lambda)$ , donde

$$\mu_X = \frac{\mu - \bar{y}}{s},$$

$$\sigma_X = \frac{\sigma}{s}.$$

Se tiene que  $\bar{x} = 0$ ,  $s_X = 1$  y  $m_3 = \frac{1}{n} \sum_{i=1}^n x_i^3$ . Igualando los momentos muestrales a los momentos poblacionales se obtiene el siguiente sistema de ecuaciones

$$\mu_X + \sigma_X \delta \sqrt{\frac{2}{\pi}} = 0$$

$$\sigma_X^2 \left[ 1 - \frac{2}{\pi} \delta^2 \right] = 1$$

$$\mu_X^3 + 3\mu_X^3 \sigma_X \delta \sqrt{\frac{2}{\pi}} + 3\mu_X \sigma_X^2 + 3\sigma_X^3 \delta \sqrt{\frac{2}{\pi}} - \sigma_X^3 \delta^3 \sqrt{\frac{2}{\pi}} = m_3.$$

De este sistema,

$$\hat{\mu}_X = -\frac{1}{s} \left( \frac{2}{4 - \pi} m_3 \right)^{\frac{1}{3}},$$

$$\hat{\sigma}_X = \sqrt{1 + \hat{\mu}_X^2}$$

$$y \quad \hat{\delta} = -\frac{\hat{\mu}_X}{\hat{\sigma}_X} \sqrt{\frac{2}{\pi}}$$

Por lo tanto los estimadores de momentos de  $\mu$ ,  $\sigma$  y  $\lambda$  son dados por

$$\hat{\mu} = \bar{y} + s\hat{\mu}_X,$$

$$\hat{\sigma} = \sqrt{s^2(1 + \hat{\mu}_X^2)}$$

$$y \quad \hat{\lambda} = \frac{\hat{\delta}}{\sqrt{1 - \hat{\delta}^2}}$$

siendo este último sobre la condición  $|\hat{\delta}| < 1$ , caso contrario  $\hat{\lambda}$  no está definido.

Cuando  $\delta = 0$ , se verifica que los estimadores encontrados no se comportan bien en el caso particular (distribución normal), pues el estimador de momentos  $\hat{\sigma}$  sobreestima el parámetro  $\sigma$  y  $\hat{\mu}$  sub o sobreestima  $\mu$  dependiendo del signo de  $\hat{\mu}_X$ . Para más detalles, ver Pewsey (2000), donde es propuesta la utilización de la parametrización centrada de Azzalini (1985) para disminuir los efectos de sobre o subestimación de los parámetros.

### 2.3.5.2 Método de Máxima Verosimilitud

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

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

Si  $\lambda$  es positivo y suficientemente grande, la probabilidad de  $z_i > 0$  es razonablemente grande, para todo  $i = 1, 2, \dots, n$ . En este caso,  $L(\lambda; \mathbf{z})$  es una función monótona creciente en  $\lambda$  y, por lo tanto, el estimador que maximiza la función de verosimilitud será infinito, el que llevaría a inferir que  $Z \sim HN(0, 1)$ . Ese mismo raciocinio sigue cuando  $\lambda$  es suficientemente distante, a la izquierda de 0.

Considere ahora solamente muestras donde el estimador de máxima verosimilitud es finito. Para obtener el estimador de máxima verosimilitud para  $\lambda$  es preciso maximizar el logaritmo de la función de verosimilitud, que es dado por

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

La primera derivada es dada por

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

donde las raíces de esa derivada pueden ser obtenidas numéricamente.

Sabiendo que la información de Fisher es dada por

$$I(\lambda) = nE_Z \left[ \frac{Z^2 \phi^2(\lambda Z)}{\Phi(\lambda Z)} \right],$$

se obtiene

$$I(\lambda) = n \int 2z^2 \phi(z) \frac{\phi^2(\lambda z)}{\Phi(\lambda z)} dz.$$

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

$$L(\mu, \sigma, \lambda; \mathbf{x}) = \left(\frac{2}{\sigma}\right)^n \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2} \left[ \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} \right]\right\} \prod_{i=1}^n \Phi\left(\lambda \left(\frac{x_i - \mu}{\sigma}\right)\right), \quad (20)$$

y su matriz de información de Fisher es dada por

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

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

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 (ver Freitas, 2006) es que estudios de simulación han demostrado que el estimador de máxima verosimilitud de  $\lambda$  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 (Azzalini, 1985), de la forma como fue propuesto en Pewsey (2000).



## 2.4 Distribución *Skew-T*

### 2.4.1 Introducción

Como se mencionó anteriormente, el análisis estadístico desarrollado para el estudio de variables aleatorias continuas está basado en el modelo normal aunque para eso es necesario asegurar condiciones como simetría y además, un determinado valor de curtosis. Entonces, para evitar la omisión de cualquiera de esas condiciones, se proponen modelos alternativos que han sido desarrollados en la literatura. Por ejemplo el modelo t-student, es considerado un modelo más robusto que el normal, pues permite reducir la influencia de valores extremos en las inferencias.

A pesar de lo descrito, tanto la distribución normal como la distribución t-student asumen la simetría de los datos. Aunque tengamos distribuciones asimétricas bien conocidas, como por ejemplo la distribución gamma, que es no invariante por transformación lineal, y cuyo espacio muestral está definido apenas para los valores positivos, sería bastante interesante obtener distribuciones asimétricas menos restrictivas y que consideren a las distribuciones simétricas conocidas como un caso particular. En ese sentido, como se vio en la anterior sección, Azzalini propuso la distribución *skew-normal* como una generalización del modelo normal. Sin embargo, existen casos en que la muestra no solo tiene un comportamiento asimétrico, sino que más que eso, hay la presencia de valores extremos, debido a eso la utilización de la distribución *skew-normal* para el modelamiento del conjunto de datos puede ser ineficaz.

En situaciones de este tipo, se plantean como alternativas más robustas 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. Todo esto se verá a continuación para el caso univariado (Godoy, 2007).

## 2.4.2 Definición y Propiedades

### 2.4.2.1 Definición

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

$$f_Z(z) = 2t_\nu(z)T_{\nu+1}\left(\lambda z \sqrt{\frac{1+\nu}{\nu+z^2}}\right), \quad (21)$$

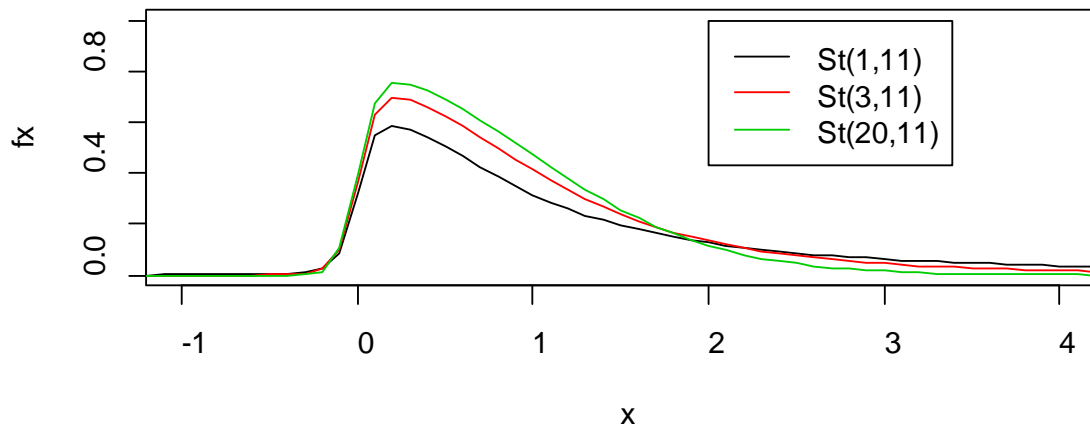
donde  $t_\nu$  denota la función de densidad de probabilidad de una distribución t-student estándar con  $\nu$  grados de libertad,  $T_{\nu+1}$  la función de distribución acumulada de una distribución t-student estándar con  $\nu + 1$  grados de libertad y  $-\infty < z < \infty$ . Como notación representamos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

En la figura 4 presentamos algunos gráficos que ilustran el comportamiento de la función de densidad. En la figura 4 (a) se muestra el gráfico para  $\lambda$  igual a 0, 2, 10 y 50, fijando el valor de  $\nu$  en 5. Y en la figura 4 (b) se muestra para valores de  $\nu$  iguales a 2, 5 y 15, fijando el valor de  $\lambda$  en 10.

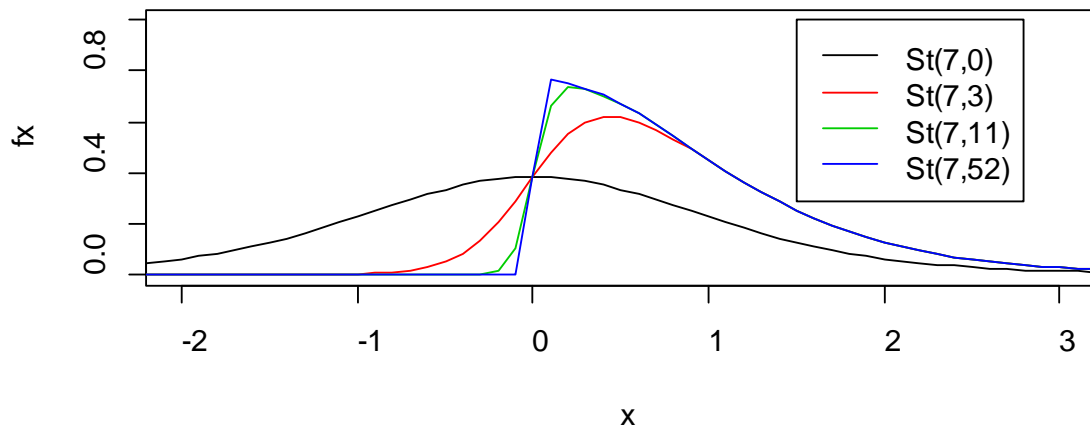
**Figura 4: Gráficos de las funciones de densidad de la distribución *skew-t*:**

**(a) Parámetro de asimetría fijo. (b) Parámetro de curtosis fijo.**

**(a) Parámetro de Asimetría fijo**



**(b) Parámetro de Curtosis fijo**



Fuentes: Elaboración propia<sup>12</sup>

El parámetro  $\lambda$  regula la forma de la distribución. Valores negativos de  $\lambda$  indican asimetría negativa y valores positivos de  $\lambda$  indican asimetría positiva. La densidad definida en (21) será simétrica cuando  $\lambda = 0$ , y coincidirá con la densidad de la distribución *t*-student estándar.

<sup>12</sup> Ver códigos en ANEXO 6

Con la introducción de los parámetros de posición ( $\mu \in \mathbb{R}$ ) y escala ( $\sigma > 0$ ), conseguimos obtener una generalización de la distribución *skew-t* estándar. De esa manera, una variable aleatoria  $Y$  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  $Z \sim ST(\mu, \sigma^2, v, \lambda)$ .

#### 2.4.2.2 Propiedades

A continuación se presenta algunas propiedades de la distribución t-asimétrica. De esa forma, considere  $Z \sim ST(v, \lambda)$  y  $Y \sim ST(\mu, \sigma^2, v, \lambda)$ . Así:

- a.  $|Z| \sim HT(v)$ . La notación  $HT(v)$  indica que  $|Z|$  tiene distribución t-positiva o *half-t* con  $v$  grados de libertad.
- b. Si  $v = 1$  entonces  $Z \sim SC(\lambda)$ . La notación  $SC(\lambda)$  se refiere a la distribución Cauchy-asimétrica con parámetro de asimetría  $\lambda$ .
- c. Cuando  $\lambda \rightarrow \infty$ ,  $Z$  converge a una  $HT(v)$ .
- d. Cuando  $v \rightarrow \infty$ ,  $Z$  converge a una  $SN(\lambda)$ .
- e.  $-Z \sim ST(v, -\lambda)$ .
- f.  $Z^2 \sim F\text{-Snedecor}(1, v)$ .
- g.  $F_Z(z; v, -\lambda) = 1 - F_Z(-z; v, \lambda)$ .
- h. Si  $Y \sim St(\mu, \sigma^2, v, \lambda)$  y  $Y_1 = a + bY$  entonces  $Y_1 \sim ST(a + b\mu, b^2\sigma^2, \text{signo}(b)\lambda, v)$

A través de la propiedad  $h$ , se puede afirmar que la distribución *skew-t* es cerrada para transformaciones lineales, es decir, cualquier combinación lineal de una variable aleatoria con distribución *skew-t* será también una distribución *skew-t*.

La **función generadora de momentos** de la distribución *skew-t* estándar es dada por:

$$M_Z(t) = \int_0^\infty h_W(w) M_X\left(\frac{t}{\sqrt{w}}\right) dw, \quad (22)$$

donde  $h_W(\cdot)$  corresponde a la función de densidad de probabilidad de una variable aleatoria  $W$  con distribución  $Gamma\left(\frac{v}{2}, \frac{v}{2}\right)$  y  $M_X(t)$  es la función generadora de momentos de la distribución *skew-normal* estándar, dada por:

$$M_Z(t) = 2e^{\frac{t^2}{2}} \Phi(\delta t), \text{ con } \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}. \quad (23)$$

A través de la función generadora de momentos conseguimos obtener los momentos de orden par e impar de la distribución *skew-t* estándar y, por consiguiente, la esperanza, la varianza, los coeficientes de asimetría y curtosis.

### 2.4.3 Generación de una distribución *skew-t*

Las notaciones  $Gamma(a, b)$ ,  $t_2(\mu, \Sigma, v)$ ,  $SN(\lambda)$  y  $SN(\mu, \sigma^2, \lambda)$  se refieren a las distribuciones gamma, t-student bivariada, *skew-normal* estándar y *skew-normal* con parámetros de posición y escala, respectivamente. Presentamos a continuación las siguientes cuatro maneras de obtener  $Z \sim ST(v, \lambda)$ :

- a. Si  $X \sim SN(\lambda)$ ,  $W \sim Gamma\left(\frac{v}{2}, \frac{1}{2}\right)$  y  $X$  y  $W$  son independientes, entonces  $Z =^d \frac{X}{\sqrt{\frac{W}{v}}}$ .
- b.  $Z|W \sim SN\left(0, \frac{1}{W}, \lambda\right)$  donde  $W \sim Gamma\left(\frac{v}{2}, \frac{v}{2}\right)$ .
- c. Si  $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim t_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix}, v \right]$ , entonces  $Z =^d X_2 | X_1 > 0$ , con  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ .
- d. Si  $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim t_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, v \right]$ , entonces  $Z =^d \delta |X_1| + \sqrt{1-\delta^2} X_2$ , con  $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$ .

La primera forma de obtener la distribución *skew-t* (ítem a) es similar a la construcción de la distribución t-student. La segunda (ítem b) representa la distribución *skew-t* como una mezcla del parámetro de escala de la distribución *skew-normal* (ésta representación es fundamental en la obtención de la función generadora de momentos de la distribución *skew-t*). La tercera forma (ítem c) es conocida como modelo de truncamiento oculto. Y la última (ítem d) se muestra útil en la simulación de la variable aleatoria  $Z$  y es conocida en la literatura como representación estocástica.

## 2.4.4 Inferencia Clásica

De manera similar a lo planteado en la distribución *skew-normal*, se presentan dos métodos clásicos de inferencia para los parámetros del modelo *skew-t*: el método de los momentos y el método de máxima verosimilitud, según Godoi (2007).

### 2.4.4.1 Métodos de los momentos

Sea  $Z = (Z_1, Z_2, \dots, Z_n)$  una muestra aleatoria simple de tamaño  $n$  de la variable aleatoria  $Z$  con distribución  $ST(\nu, \lambda)$ . Considere  $\bar{Z} = \frac{\sum_{i=1}^n Z_i}{n}$  y  $m_2 = \frac{\sum_{i=1}^n Z_i^2}{n}$ . Tenemos que los estimadores para los parámetros de asimetría y curtosis son dados respectivamente, por:

$$\tilde{\lambda} = \frac{\bar{Z} \Gamma\left(\frac{\tilde{\nu}}{2}\right)}{\sqrt{\frac{\tilde{\nu}}{n} \Gamma^2\left(\frac{\tilde{\nu}-1}{2}\right) - \bar{Z}^2 \Gamma^2\left(\frac{\tilde{\nu}}{2}\right)}} \quad y$$

$$\tilde{\nu} = \frac{2m_2}{m_2 - 1}$$

Notar que la existencia del estimador  $\tilde{\lambda}$  está condicionada a que el radicando sea mayor que cero, así como el estimador  $\tilde{\nu}$  solo existe cuando el denominador de la fracción es positivo y  $m_2 > 0$ . De esa forma, estas condiciones de existencia se traducen en:

$$\bar{Z} \in \left[ -\sqrt{\frac{\tilde{\nu}}{n} \frac{\Gamma\left(\frac{\tilde{\nu}-1}{2}\right)}{\Gamma\left(\frac{\tilde{\nu}}{2}\right)}}, \sqrt{\frac{\tilde{\nu}}{n} \frac{\Gamma\left(\frac{\tilde{\nu}-1}{2}\right)}{\Gamma\left(\frac{\tilde{\nu}}{2}\right)}} \right] \quad y$$

$$m_2 > 1.$$

También, se nota que a causa de la condición de existencia muy restrictiva de  $\lambda$  y  $\nu$ , este método de los momentos se aplica a pocas muestras. Además, en el caso en que la distribución considerada es la distribución *skew-t* con parámetros de posición y escala, se tiene que las estimaciones de momentos deben ser obtenidas a través de métodos numéricos.

#### 2.4.4.2 Método de Máxima Verosimilitud

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

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

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

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

Para esta situación, si para todo  $i, z_i > 0$ , entonces  $L(\nu, \lambda; \mathbf{z})$  es una función monótona creciente en  $\lambda$  y, como consecuencia, el estimador que maximiza la función de verosimilitud será infinito. Utilizando el mismo raciocinio, verificamos que si para todo  $i z_i < 0$ , entonces el estimador de máxima verosimilitud será menos infinito.



Según Sartori (2006), citado por Godoy (2007), menciona que cuando consideramos la muestra como descrita anteriormente, pero con  $\nu > 0$  conocido, la probabilidad de obtener una muestra con el estimador de máxima verosimilitud para  $\lambda$  infinito es dada por:

$$P(Z < 0)^n + P(Z > 0)^n = \left(\frac{1}{2} - \frac{1}{\pi} \arctan \lambda\right)^n + \left(\frac{1}{2} + \frac{1}{\pi} \arctan \lambda\right)^n \quad (24).$$

Comenta también que cuando consideramos ambos parámetros de la distribución *skew-t* estándar desconocidos, la probabilidad de obtener el estimador de máxima verosimilitud del parámetro de asimetría infinito es menor que lo mencionado en (24). Tal situación se debe a la estimación del parámetro de grados de libertad.

Además de eso, la estimación para el parámetro de grados de libertad también puede ser infinita. Un valor de  $\hat{\nu}$  infinito indica que la muestra fue generada a partir del modelo *skew-normal*. De la misma forma que un valor  $\hat{\lambda}$  infinito, indica que la muestra es proveniente de un modelo t-positivo. Sin embargo, existe una probabilidad no despreciable de que las distribuciones con parámetro de asimetría finito generen muestras con estimaciones de máxima verosimilitud infinitas.

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.

## 2.5 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 (Ricci, 2005). Estas tienen una forma de prueba de hipótesis donde las hipótesis nula y alternativa son:

$H_0$ : los datos muestrales provienen de una distribución establecida.

$H_1$ : los datos muestrales no provienen de una distribución establecida.

Estas pruebas son llamadas algunas veces como *pruebas ómnibus* y ellas son *distribuciones libres*, lo que significa que no dependen de la función de densidad de probabilidad. La prueba chi-cuadrado ( $\chi^2$ ) es la prueba de bondad de ajuste más antigua planteada por Karl Pearson (1900). La prueba puede ser pensada como una comparación formal de un histograma con la densidad ajustada. La prueba  $\chi^2$  se puede aplicar a las distribuciones discretas o continuas, mientras que las pruebas de Kolmogorov - Smirnov y Anderson – Darling se limitan a distribuciones continuas.

### 2.5.1 Prueba de Kolmogorov – Smirnov

Según Ricci (2005), la prueba de Kolmogorov – Smirnov, que se utiliza en la presente investigación, 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.

Dado  $n$  puntos de datos ordenados  $X_1, X_2, \dots, X_n$ , la ECDF es definida como:

$$F_n(X_i) = N(i)/n$$

donde  $N(i)$  es el número de puntos menores a  $X_i$  ( $X_i$  son ordenados del valor más pequeño al más grande). Esto es una función *paso*, que se incrementa con  $1/n$  en el valor de cada punto de los datos ordenados.

La prueba estadística utilizada es:

$$D_n = \sup_{1 \leq i \leq n} |F(x_i) - F_n(x_i)|$$

que es el extremo superior entre la diferencia en valor absoluto entre la función de distribución acumulada empírica y la función de distribución acumulada teórica.

La hipótesis con respecto a la forma de la distribución es rechazada si la prueba estadística,  $D_n$ , es más grande que el valor crítico obtenido de una tabla, o, lo que es lo mismo, si el p-valor es más bajo que el nivel de significancia (normalmente 0.05).

La prueba Kolmogorov – Smirnov es más poderosa que la prueba  $\chi^2$  cuando el tamaño muestral no es tan grande. Para grandes tamaños ambas pruebas tienen el mismo poder. La mayor limitación de la prueba Kolmogorov - Smirnov es que la distribución debería ser especificada, esto es, locación, escala, y parámetros de forma no pueden ser estimados desde los datos muestrales. Debido a esto, muchos analistas prefieren usar la prueba de Bondad de Ajuste Anderson-Darling. Sin embargo, la prueba Anderson-Darling está disponible para pocas distribuciones específicas.

## 2.6 Criterios de Selección del Mejor Modelo

### 2.6.1 Criterio de Información de Akaike (AIC)

Según Macerau (2012), 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 AIC intenta escoger dentro de los modelos que están disponibles, aquel que minimice la divergencia de Kullback – Leibler (KL).

La divergencia de Kullback – Leibler (KL) es la distancia entre dos modelos probabilísticos, o entre dos distribuciones de probabilidades. Es el valor esperado del logaritmo de la razón de las densidades de los modelos que están siendo probados, es decir, un integral de las diferencias entre las densidades, en escala logarítmica en todo espacio donde la variable aleatoria  $X$  está definida, ponderada por una distribución  $f_X(x)$ .

La divergencia de KL, conforme Kullback & Leibler (1951) y Kullback (1968), es dada por:

$$D_{KL}(f_X(x)||g_X(x; \theta)) = I(f; g) = \int_{-\infty}^{\infty} f_X(x) \log\left(\frac{f_X(x)}{g_X(x; \theta)}\right) dx \quad (25),$$

donde  $f_X(x)$  es la función de densidad del modelo “verdadero”,  $g_X(x; \theta)$  es la función de densidad del modelo “propuesto”, y  $\theta$  es el vector de parámetros.

Reescribiendo la ecuación (25), tenemos:

$$I(f, g) = \int_{-\infty}^{\infty} f_X(x) \log(f_X(x)) dx - \int_{-\infty}^{\infty} f_X(x) \log(g_X(x; \theta)) dx \quad (26).$$

Si pensamos en términos de esperanza, la ecuación (4.12) puede ser escrita como

$$I(f, g) = E_f[\log[f_X(x)]] - E_g[\log[g_X(x; \theta)]] \quad (27),$$

con la esperanza siendo tomada en relación al modelo “verdadero”  $f_X(x)$ .

El primer término de la ecuación (27) será siempre una constante, es decir,  $E_f[\log[f_X(x)]] = C$ , pues depende del modelo “verdadero”. Sustituyendo esa constante en (27) tenemos:

$$I(f, g) - C = -E_f[\log[g_X(x; \theta)]] \quad (28),$$

Como no conocemos el modelo “verdadero” y los valores del vector de parámetros  $\theta$  del modelo propuesto  $g_X(x; \theta)$ , no tenemos como calcular el valor de la expresión (28).

Una estrategia es estimar el comportamiento del modelo “verdadero” utilizando una muestra grande, es decir,  $E_f[\log[g_X(x; \theta)]] \approx E_x[\log[g_X(x; \theta)]]$ , que es el logaritmo de la función de verosimilitud del modelo en la muestra, entonces  $E_x[\log[g_X(x; \theta)]] = \log\{L[g_X(\theta/x)]\}$ .

Si la muestra ( $x$ ) es suficientemente grande, y las estimaciones ( $\hat{\theta}$ ) del vector de parámetros fueron consistentes, entonces  $E_x[\log[g_X(x; \theta)]] \approx E_x[\log[g_X(x; \hat{\theta})]]$ .

Si las estimaciones ( $\hat{\theta}$ ) de los parámetros fueron las de máxima verosimilitud (EMV), esa esperanza es el máximo del logaritmo de la función de verosimilitud del modelo en la muestra, es decir,  $E_x[\log[g_X(x; \theta)]] = \log\{L[g_X(\hat{\theta}/x)]\}$ , donde  $\hat{\theta}$  es el vector de EMV.

Por lo tanto, la ecuación (28) puede ser escrita como

$$I(f, g) - C \approx -\log\{L[g_X(\hat{\theta}/x)]\} \quad (29),$$

esto quiere decir, que el máximo del logaritmo de la función de verosimilitud es la estimación de la menor distancia relativa posible para el modelo propuesto. Ello engloba dos divergencias, la divergencia de aproximación entre el modelo “verdadero” y el modelo propuesto y la divergencia entre los valores de los parámetros y las estimaciones de máxima verosimilitud.

Akaike (1974), percibió que cuando la muestra del modelo “verdadero” y la muestra utilizada para estimar los parámetros eran la misma muestra, el máximo valor del logaritmo de la función de verosimilitud era un distribución sesgada, y el sesgo era positivo y proporcional al número  $k$  de parámetros del modelo. Entonces una estimación corregida es

$$\log\{L[g_X(\hat{\theta}/x)]\} - k \quad (30),$$

luego la estimación del criterio de información de Akaike, más conocido como AIC es dado por

$$\widehat{AIC} = -2 \log \left\{ L \left[ g_x \left( \frac{\hat{\theta}}{x} \right) \right] \right\} + 2k \quad (31),$$

donde  $\log \left\{ L \left[ g_x \left( \frac{\hat{\theta}}{x} \right) \right] \right\}$  es el logaritmo de la función de verosimilitud del modelo propuesto con vector de parámetros  $\theta$ , y  $k$  es el número de parámetros.

El modelo con menor valor de AIC es el modelo que mejor se a los datos.

### III. MATERIALES Y MÉTODOS

#### 3.1 Materiales

Los materiales y equipos que se utilizaron en la presente investigación son los siguientes:

- a) Una computadora laptop marca DELL, con un procesador Intel (R) Core(TM)2 CPU T5200 @ 1.60 GHz 1.60 GHz, con una memoria RAM de 2.96 GB y un sistema operativo Windows 7 Ultimate de 32 bits.
- b) El *software* estadístico R versión 2.15.0.
- c) El *software* de edición Microsoft Word 2010.
- d) Una memoria USB de 2 Gb para el almacenamiento de la información
- e) Un impresora Epson Multiuso 3800
- f) Un millar de papel bond tamaño A4.



## **3.2 Metodología de la Investigación**

### **3.2.1 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 costedel siniestro.

### **3.2.2 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).

Se analizó dos conjuntos de datos, según el tipo de cobertura del SCTR<sup>13</sup>, a los cuales se les aplica una secuencia de pasos ordenados descritos en el ítem “Metodología Aplicada”, que se detalla más adelante.

### **3.2.3 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.

---

<sup>13</sup> Explicado en el capítulo II: Revisión de Literatura.

### 3.2.4 Identificación de las variables

Las variables con las que se cuenta en esta investigación son:

- Código del siniestro: La variable código del siniestro es un identificador. Esta variable permitirá anotar el código numérico que corresponda al coste del siniestro
- Coste del siniestro: La variable coste del siniestro es una variable cuantitativa. Esta nos permitirá contrastar la hipótesis de la investigación: se observará si es bien ajustada por las distribuciones asimétricas (*skew-normal* y *skew-t*)

### 3.2.5 Población y muestra

La población en el primer grupo es 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 es 197 datos del coste de los siniestros por cobertura de Sobrevivencia, que fueron aprobados y pagados parcialmente por la empresa.

En ambos grupos, la unidad monetaria de los datos está representado en la moneda nacional: nuevos soles (S/.)

### 3.2.6 Metodología aplicada

Cada uno de los conjuntos de datos (llamados también grupos) estudiados en esta investigación será analizado con el siguiente procedimiento:

- ✓ Análisis descriptivo de los datos.
- ✓ Análisis descriptivo del logaritmo de los datos.

- ✓ Ajuste de los datos a las distribuciones en estudio (*skew-normal* y *skew-t*) y a las distribuciones tradicionales.
- ✓ Comparación de las distribuciones analizadas mediante diferentes criterios estadísticos<sup>14</sup>.
- ✓ Estimación de medidas de Riesgo (VaR y TVaR) que complementen los resultados.

---

<sup>14</sup> Explicado en el capítulo II: Revisión de Literatura.

## **IV. RESULTADOS Y DISCUSION**

### **4.1 Análisis estadístico**

#### **4.1.1 Descripción de los conjuntos de datos**

Los datos del presente estudio llamados coste de los siniestros, provienen de la remuneración base que indemnizó la empresa de seguros a las personas aseguradas o a sus beneficiarios, ya sean por coberturas de Invalidez o de Sobrevivencia, respectivamente. Entonces se consideran dos conjuntos de datos a analizar, cada uno con unidad de medida el nuevo sol, que tienen por objetivo comprobar la hipótesis de la investigación.

#### **4.1.2 Análisis de los datos de Invalidez y Sobrevivencia**

El cuadro 2 presenta las estadísticas descriptivas para los dos conjuntos de datos, en donde se encuentra el número de observaciones, los estimadores de los cuatro primeros momentos (media, desviación estándar, asimetría y kurtosis), los valores mínimo y máximo; además se presenta el Valor en Riesgo (VaR), considerándolo como el percentil 99% y el Valor en Riesgo Condicional (TVaR), que es también llamado la pérdida media cuando esta se encuentra por encima del VaR<sup>15</sup>.

---

<sup>15</sup> Estas medidas de riesgo, son explicadas en el capítulo II: Revisión de Literatura.

**Cuadro 2: Estadísticas descriptivas del coste de los siniestros para los grupos de Invalidez y Supervivencia**

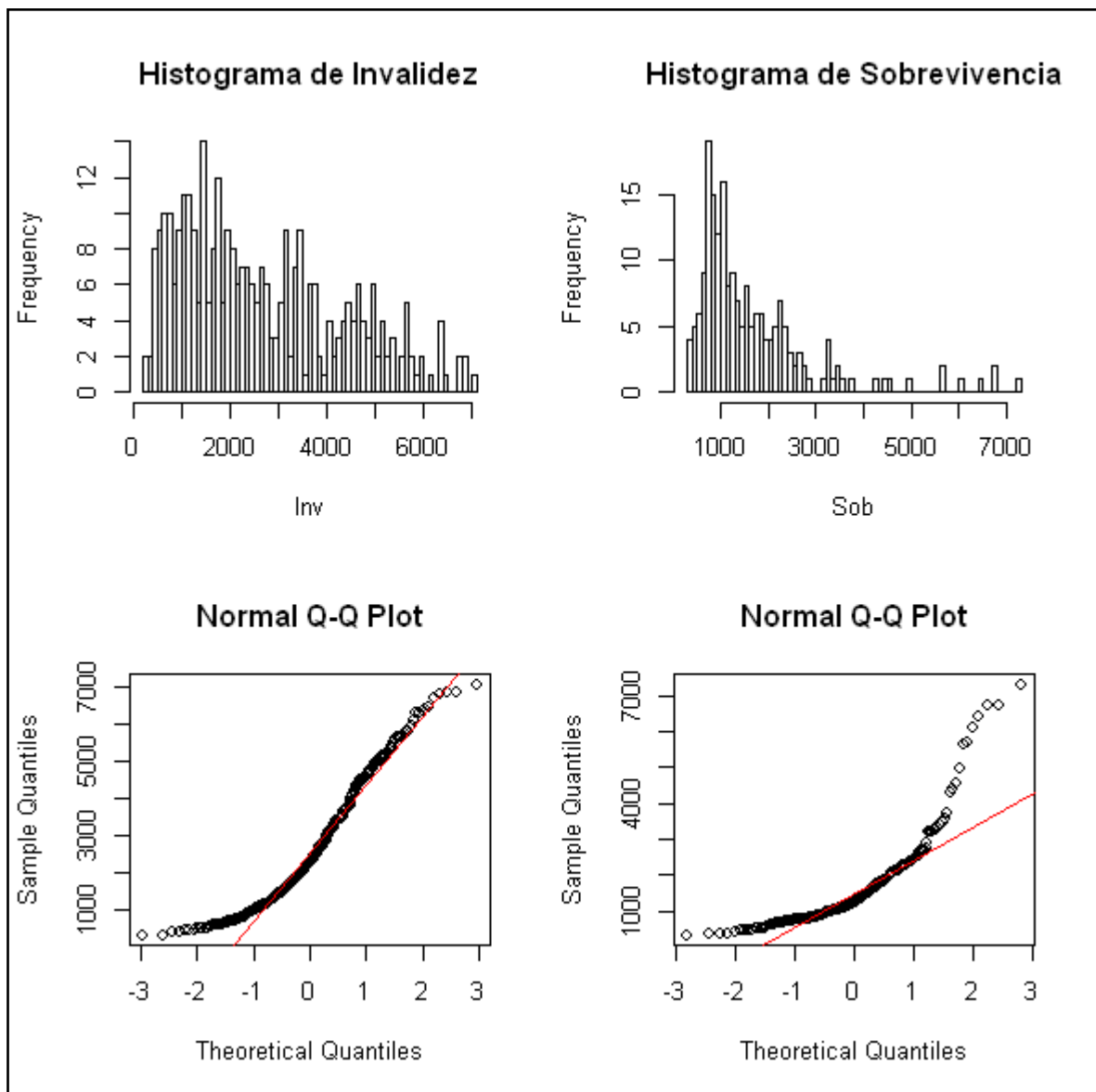
|                                                       | <b>Invalidez</b> | <b>Supervivencia</b> |
|-------------------------------------------------------|------------------|----------------------|
| Nº de observaciones                                   | 326              | 197                  |
| 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               |
| X: coste de los siniestros en una compañía de seguros |                  |                      |

Fuente: Elaboración propia

Las estadísticas descriptivas muestran que las remuneraciones de Supervivencia 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.

Sin embargo, se tiene que realizar más pruebas tanto informales como formales para comprobar que los datos de seguros siguen una distribución asimétrica. Por lo tanto en la figura 5, se presentan dos histogramas para los conjuntos de datos considerados, así como los correspondientes gráficos de normalidad Q-Q plots. Las dos figuras de la izquierda muestran los datos de remuneración por Invalidez, mientras que las de la derecha muestran los datos de remuneración por Supervivencia.

**Figura 5: Gráficos descriptivos del coste de los siniestros para los grupos de Invalidez (izquierda) y Sobrevivencia (derecha)<sup>16</sup>**



Fuente: Elaboración propia

Ambos gráficos de histogramas revelan una característica muy típica de los datos del coste de los siniestros en los seguros: una cantidad grande de costes bajos y una cantidad pequeña de costes altos.

<sup>16</sup> Ver códigos en ANEXO 7

### 4.1.3 Análisis del logaritmo natural de los datos de Invalidez y Supervivencia

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 (Bolanceet. al. 2008, citado por Eling 2011).

El cuadro 3 presenta las estadísticas descriptivas del logaritmo de los datos y se muestra el número de observaciones, la media, la desviación estándar, asimetría, kurtosis, los valores mínimo y máximo, además se presenta VaR, considerándolo como el percentil 99% y el TVaR, que es también llamado la pérdida media cuando esta se encuentra por encima del VaR.

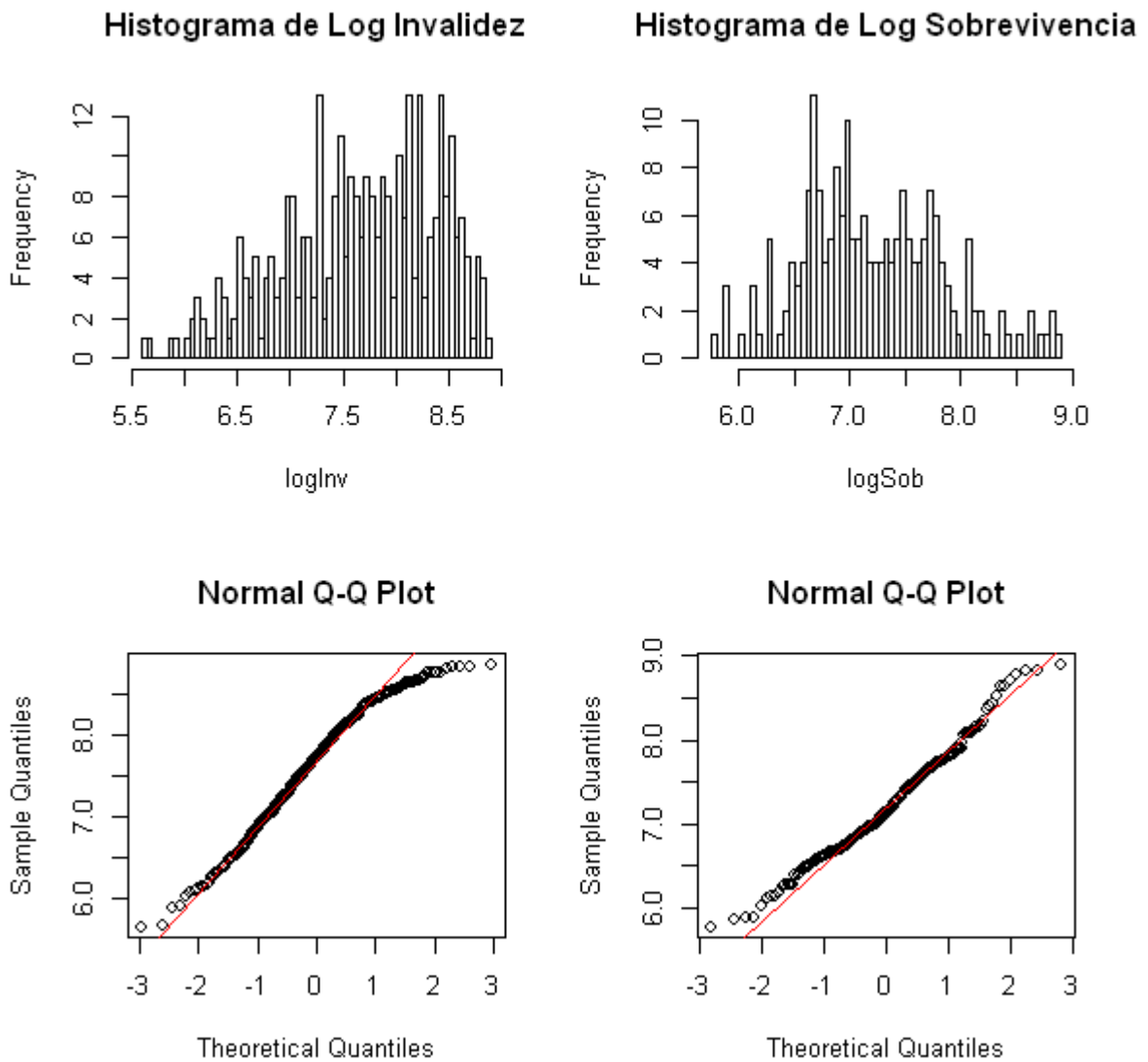
**Cuadro 3: Estadísticas descriptivas del logaritmo del coste de los siniestros para los grupos de Invalidez y Supervivencia**

|                                                                     | <b>logInvalidez</b> | <b>logSupervivencia</b> |
|---------------------------------------------------------------------|---------------------|-------------------------|
| Nº de observaciones                                                 | 326                 | 197                     |
| 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                    |
| X: logaritmo del coste de los siniestros en una compañía de seguros |                     |                         |

Fuente: Elaboración propia

La figura 6 presenta los histogramas y los gráficos normales Q-Q plots para los datos del coste de los siniestros por Invalidez (izquierda) y por Sobrevivencia (derecha), tomando el logaritmo natural de los mismos.

**Figura 6: Gráficos descriptivos del logaritmo del coste de los siniestros para los grupos de Invalidez (izquierda) y Sobrevivencia (derecha)<sup>17</sup>**



Fuente: Elaboración propia

<sup>17</sup> Ver códigos en ANEXO 8



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

#### 4.2 Estimación de los parámetros de las distribuciones *skew-normal* y *skew-t*

Todos los parámetros estimados en la presente investigación están calculados bajo el método de la estimación de máxima verosimilitud<sup>18</sup>. El cuadro 4 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 Supervivencia.

**Cuadro 4: Parámetros estimados de las distribuciones *skew-normal* y *skew-t***

|                           | <b>Invalidez</b> | <b>Supervivencia</b> | <b>logInvalidez</b> | <b>logSupervivencia</b> |
|---------------------------|------------------|----------------------|---------------------|-------------------------|
| <b><i>Skew-Normal</i></b> |                  |                      |                     |                         |
| 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                     |
| <b><i>Skew-t</i></b>      |                  |                      |                     |                         |
| 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                 |

Fuente: Elaboración propia

<sup>18</sup> Expuesto en el capítulo II: Revisión de Literatura.

Los parámetros estimados para las otras distribuciones consideradas en la presente no se muestran en esta sección al no ser el foco de estudio, pero pueden ser consultadas en la sección de Anexos<sup>19</sup>. 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.

### **4.3 Comparación de los modelos usando el criterio del logaritmo de la función de verosimilitud y el AIC**

El cuadro 5 presenta los estadísticos para elegir el mejor modelo ubicando al principio de la tabla 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.

Según el criterio del logaritmo de la función de verosimilitud se elige el modelo (i.e., la distribución) con el menor valor absoluto y según el Criterio de Información de Akaike (AIC<sup>20</sup>) el modelo preferido es el que tiene el menor AIC.

---

<sup>19</sup> Ver ANEXO 9. Además en el ANEXO 10 se muestran los códigos en R para la elaboración de los cuadros 4, 5, 6, 7, 8 y de la figura 7.

<sup>20</sup> Expuesto en el capítulo II: Revisión de Literatura.

**Cuadro 5: 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           |

Fuente: Elaboración propia

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 obtenido tanto para el coste de los siniestros como para su logaritmo natural, 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 obtenido tanto para el coste de los siniestros como para su

logaritmo natural, las distribuciones *skew-normal* y *skew-t* se mantienen entre los primeros lugares de las que mejor ajustan a los grupos analizados (Invalidez y Supervivencia).

#### **4.4 Pruebas de Bondad de Ajuste Kolmogorov – Smirnov**

Después de lo observado anteriormente, las distribuciones asimétricas en estudio al parecer se ajustan mejor en comparación con las otras distribuciones clásicas también presentadas en el cuadro 5. Entonces se podría concluir el considerar ambas distribuciones, la *skew-normal* y la *skew-t*, cuando se modelan datos de seguros.

Sin embargo, como se elige al mejor modelo de manera relativa y gracias a los estadísticos del Logaritmo de Función Verosimilitud o del AIC (que indican la distribución que mejor se ajusta según sea el valor más grande o más pequeño, respectivamente), puede ocurrir el caso en que cuando se comparen varias distribuciones, todas ellas sean muy malas a la hora de describir a los datos. Por esa razón, para analizar y descubrir si los modelos en estudio son buenos para modelar datos del coste de los siniestros se realiza también la prueba de bondad de ajuste Kolmogorov Smirnov<sup>21</sup>.

---

<sup>21</sup> Expuesto en el capítulo II: Revisión de Literatura.

**Cuadro 6: Prueba de bondad de ajuste Kolmogorov – Smirnov**

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

Fuente: Elaboración propia

El valor crítico a comparar con un nivel de confianza del 95% es calculado por  $1.36/\sqrt{n}$  (según Tabla<sup>22</sup>), con  $n$  como el número de observaciones. Si el valor estadístico de la prueba se mantiene por debajo de este nivel, entonces no se rechaza la hipótesis planteada de que la distribución empírica de los datos pertenece a este tipo de distribución. Las distribuciones que cumplen este criterio se muestran en negrita en el cuadro 6. En dicho cuadro se analizan las siete distribuciones que tuvieron un performance razonable en el análisis del cuadro 5.

Los resultados del cuadro 6 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

<sup>22</sup> Ver Anexo 11

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.

#### **4.5 Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR)<sup>23</sup>**

Se utilizaron los parámetros estimados 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 reales. En el Cuadro 7 se presentan estos estimadores que tienen como objetivo complementar los resultados anteriores. Se presentan estimadores de VaR y TVaR usando el percentil 99%.

Los resultados generados en el cuadro 7 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.

---

<sup>23</sup> Expuesto en el capítulo II: Revisión de Literatura.

**Cuadro 7: Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) usando el percentil 99%**

| Modelo                             | Datos Originales |               | Log Datos Originales |               |
|------------------------------------|------------------|---------------|----------------------|---------------|
|                                    | Invalidez        | Sobrevivencia | Invalidez            | Sobrevivencia |
| <b>Valor en Riesgo</b>             |                  |               |                      |               |
| 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           |
| logística                          | 6890             | 4162          | 9.7                  | 8.8           |
| weibull                            | 7468             | 5091          | 9.0                  | 8.6           |
| Empírico                           | 6765             | 6713          | 8.8                  | 8.8           |
| <b>Valor en Riesgo Condicional</b> |                  |               |                      |               |
| 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           |
| Logística                          | 7840             | 4757          | 10.1                 | 9.2           |
| Weibull                            | 8381             | 5792          | 9.1                  | 8.7           |
| Empírico                           | 6886             | 7016          | 8.8                  | 8.9           |

Fuente: Elaboración propia

En general, el VaR y el TVaR no tienen un buen performance cuando se consideran los datos originales. Las distribuciones muestran una gran variación en los resultados, incluyendo en los casos de las distribuciones *skew-normal* y *skew-t*. Sin embargo, los resultados para el VaR y el TVaR lucen mucho mejor cuando se consideran al logaritmo natural de los datos originales del coste de los siniestros; los estimadores de riesgo derivados usando las distribuciones teóricas son bastante próximos a los valores empíricos o reales. 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) en el grupo de Invalidez, lo cual enfatiza que es más extrema en las colas que las distribuciones en estudio de la presente investigación.

La figura 7 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.

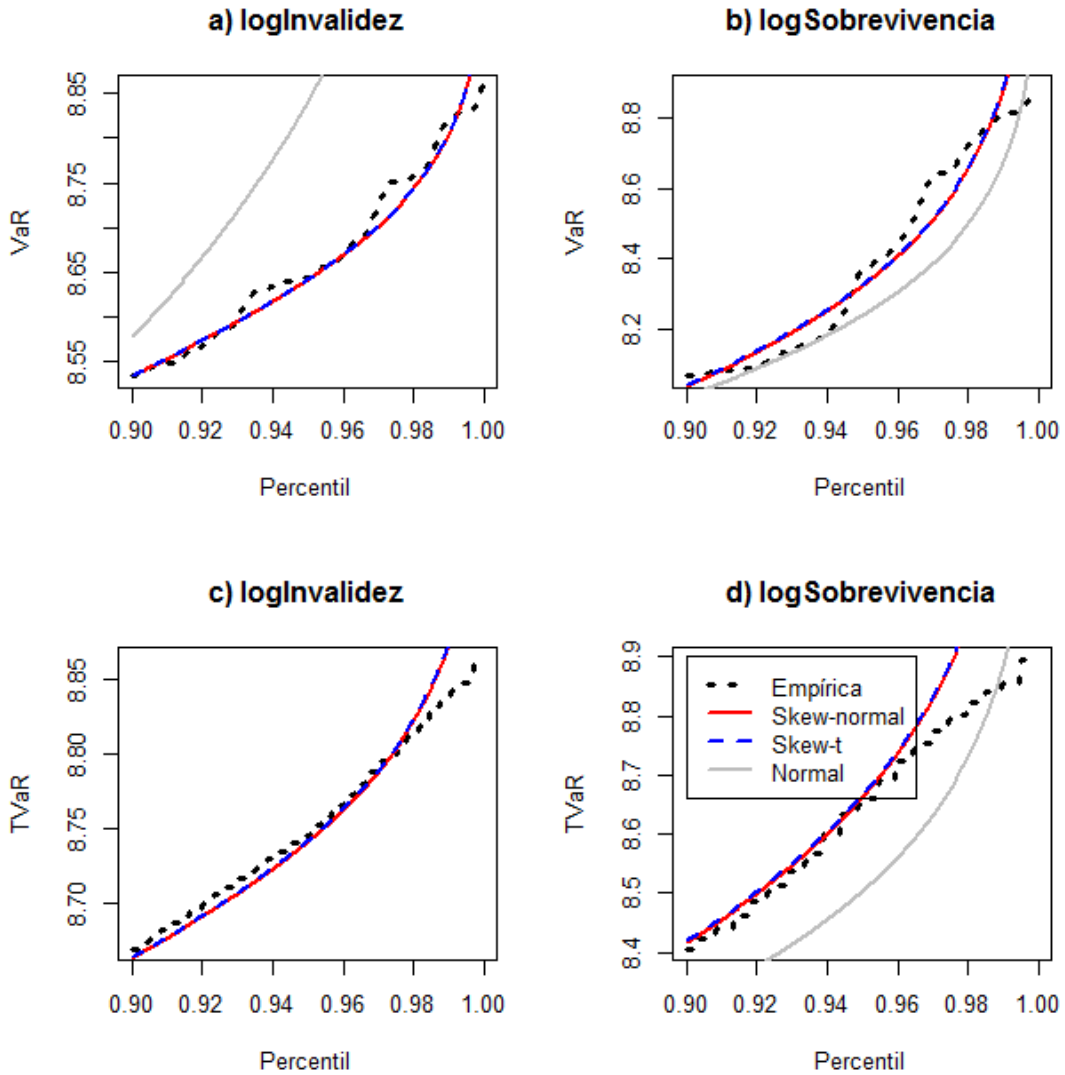
En el caso de los VaR del logaritmo del coste de los siniestros, como ya se mencionó por el AIC y las pruebas de bondad de ajuste, las distribuciones *skew-normal* y *skew-t* se ajustan bien a los datos de ambos grupos (Invalidez y Supervivencia); situación que no sucede en el caso de la distribución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

En cambio, si consideramos 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 Supervivencia las distribuciones *skew-normal* y *skew-t* sobrestiman 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 7: Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) con percentiles del 90% al 99.9%<sup>24</sup>**



Fuente: Elaboración propia

<sup>24</sup> Ver Anexo 12

## V. CONCLUSIONES

1. 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.
2. En conjunto, los indicadores del presente estudio (AIC, LogVero, K-S, VaR y TVaR) 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 Supervivencia
3. Para los dos grupos de datos estudiados, la distribución skew-normal se ajusta mejor en comparación con la distribución skew-t.
4. 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.

## **VI. RECOMENDACIONES**

1. Con los resultados obtenidos, la empresa de seguros que nos brindó los datos debe modelar su variable coste de los siniestros mediante la distribución asimétrica skew-normal.
2. Son necesarias más aplicaciones con diferentes conjuntos de datos para confirmar el uso de las distribuciones asimétricas en estudio (skew-normal y skew-t) para modelar datos de seguros (coste de los siniestros).

## VII. REFERENCIAS BIBLIOGRAFICAS

Asociación Peruana de Empresas de Seguros. Orientación <<http://www.apeseg.org.pe/orientacion.html>> [Consultada: 09 de setiembre de 2012].

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12: 171-178.

Azzalini, A. y Capitano, 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.

Dalla Valle, A. (2004). The Skew-Normal Distribution. In Genton, M. G., editor, *Skew-Elliptical Distributions and their Applications: a journey beyond normality*, capítulo 1, pp 3-24. USA: Editorial Chapman & Hall/CRC.

De Silva, N. (2006). *An Introduction to R: Examples for Actuaries. Guide. Actuarial Toolkit Working Party. Versión 0.1.*

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.

Ferreira, E. y Garín, M. (2010). *Estadística Actuarial: Modelos Estocásticos. Guía del curso de Estadística actuarial: modelos estocásticos*. Universidad del País Vasco /Euskal Herriko Unibertsitatea, Bilbao.

Fundación MAPFRE. Diccionario MAPFRE de Seguros.  
<<http://www.mapfre.com/wdiccionario/general/diccionario-mapfre-seguros.shtml>>  
[Consultada: 05 de diciembre de 2012].

Godoi, L. G. e Branco, M. D. (2007). A distribuição t-assimétrica univariada: propriedades e inferencia. Dissertação de Mestrado em Estatística. Universidade de São Paulo, Brasil.

Gomes, Priscila da Silva (2009). Distribuição normal assimétrica para dados de expressão gênica. Dissertação de Mestrado em Estatística. Universidade Federal de São Carlos, Brasil.

Macerau, Walkiria Maria de Oliveira (2012). Comparação das distribuições  $\alpha$ -estável, normal, t de student e Laplace assimétricas. Dissertação de Mestrado em Estatística. Universidade Federal de São Carlos, Brasil.

Reglamento de Tesis de la Universidad Nacional Agraria La Molina.  
[http://www.lamolina.edu.pe/Investigacion/web/pdf/Reglamento\\_de\\_Tesis.pdf](http://www.lamolina.edu.pe/Investigacion/web/pdf/Reglamento_de_Tesis.pdf)[Consultada: 2 de setiembre de 2012].

Ricci, Vito (2005). Fitting Distribution with R.

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

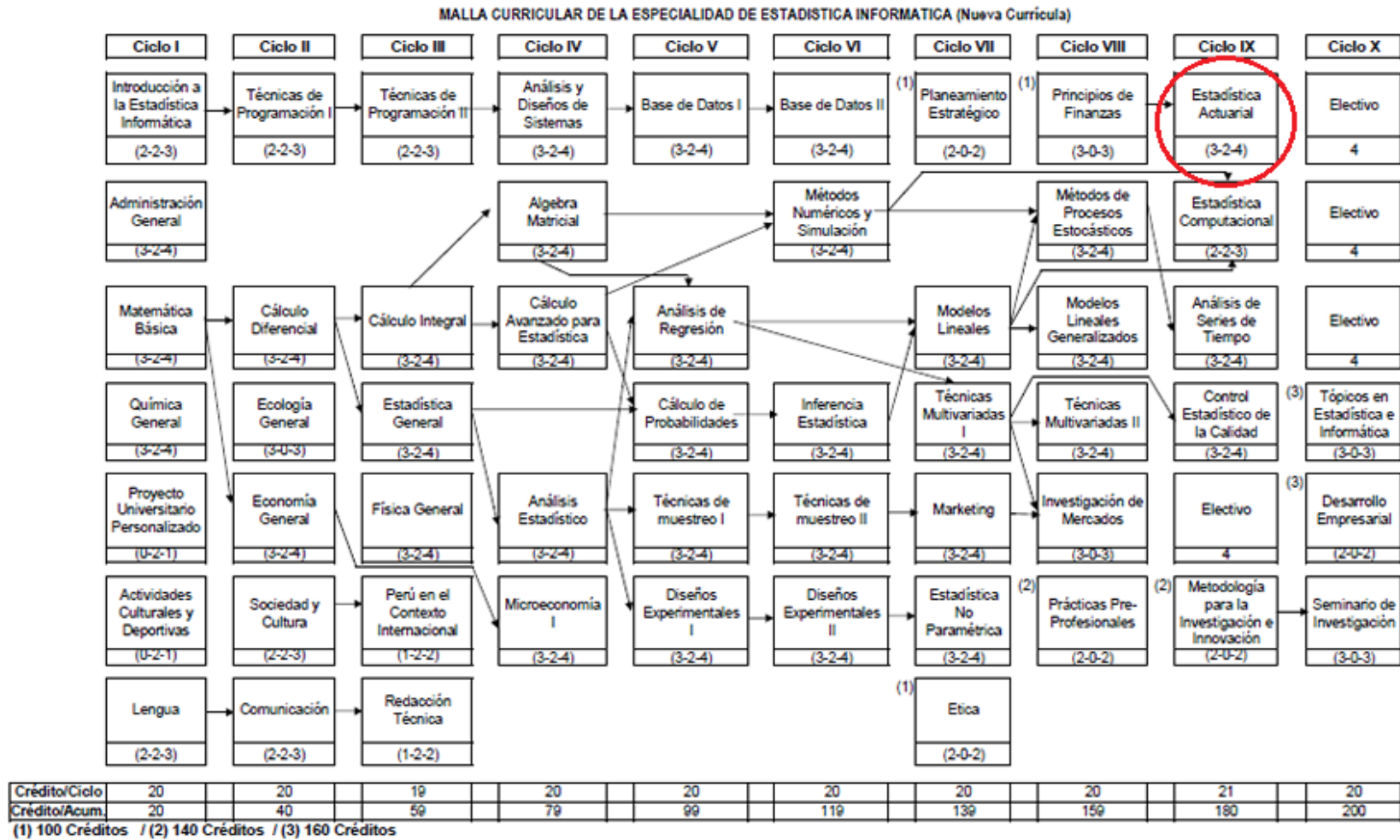
Superintendencia de Banca, Seguro y AFP. Preguntas Frecuentes.  
<[http://www.sbs.gob.pe/0/modulos/JER/JER\\_Interna.aspx?ARE=0&PFL=0&JER=2170](http://www.sbs.gob.pe/0/modulos/JER/JER_Interna.aspx?ARE=0&PFL=0&JER=2170)>  
[Consultada: 20 de octubre del 2012].

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

## **VIII. ANEXOS**

## ANEXO 1:

### Malla Curricular de la Especialidad de Estadística e Informática. Facultad de Economía y Planificación – UNALM. 2011.



## ANEXO 2:

### GLOSARIO

#### **Terminología Actuarial**

- Términos según la Superintendencia de Banca, Seguro y AFP (SBS):

#### **a. Seguro**

Un seguro es una actividad económico-financiera que transforma los riesgos de diversa naturaleza a los cuales están sometidos los patrimonios, en un gasto periódico determinado, que puede ser soportado con relativa facilidad por los tomadores del seguro.

#### **b. Contrato de seguro**

Es el acuerdo escrito a través del cual, una persona transfiere a una empresa de seguros un riesgo que la afecta (riesgo de incendio sobre una propiedad, robo, accidentes, enfermedades, etc.), a cambio del pago de una contraprestación en dinero, llamada *prima*.

#### **c. Póliza**

Una póliza de seguros es el documento en el que consta el contrato de seguro. Está compuesto por un condicionado general, un condicionado particular, y ocasionalmente por un condicionado especial o cláusulas especiales. Se considera que también forman parte del contrato de seguros los formularios y declaraciones de salud de otra índole que deba llenar el asegurado al momento de contratar el seguro.



**d. Siniestro**

Es la manifestación concreta del riesgo asegurado, a consecuencia de la cual, se producen daños cuya indemnización está garantizada por el contrato de seguro, motivando que la aseguradora satisfaga total o parcialmente al asegurado o a sus beneficiarios, el monto del capital asegurado.

- Términos según Saravia et al. (2008):

**e. Asegurado**

Es la persona física o jurídica titular del bien o interés asegurado que está expuesto al riesgo.

**f. Asegurador**

Es la entidad o compañía de seguros encargada de dar cobertura al asegurado ante la imprevisión de riesgos que supongan una pérdida.

- Término según el diccionario MAPFRE de seguros:

**g. Beneficiario**

Es la persona física o jurídica que recibe la prestación de servicios o la indemnización ante un siniestro. Generalmente coincide con el asegurado. Aunque, los seguros de vida constituyen un claro ejemplo en los que no siempre coincide asegurado y beneficiario.

**h. Tomador del seguro o contratante del seguro**

Es la persona física o jurídica que contrata el seguro y paga la prima. Generalmente coincide con el asegurado.

**i. Cobertura o indemnización**

Es la prestación monetaria o de servicios que efectúa el asegurador al beneficiario en caso de producirse un siniestro amparado por las condiciones de la póliza.

**j. Actuario**

Individuo que, en posesión de la titulación académica correspondiente, está capacitado para abordar las cuestiones técnicas (estadística, financieras, matemáticas, etc) relacionadas con las operaciones de seguros.

### ANEXO 3

#### Decreto Supremo N° 009-97-SA, Reglamento de la Ley de Modernización de la Seguridad Social en Salud.

"ANEXO 5

sctr

ACTIVIDADES COMPRENDIDAS EN EL SEGURO COMPLEMENTARIO DE TRABAJO DE RIESGO

(\* De conformidad con la Segunda Disposición Final y Transitoria de la Ley 28806, publicada el 22 julio 2006, se establece que los Inspectores Auxiliares, Inspectores del Trabajo y Supervisores-Inspectores del Trabajo están comprendidos dentro de los alcances del presente Anexo como actividades de alto riesgo los servicios prestados por dichos servidores públicos. Para tales efectos, es de aplicación el artículo 19 de la Ley N° 26790 y las demás normas legales sobre la materia.

CONCORDANCIAS: ACUERDO N° 41-14-ESSALUD-99  
Ley N° 28081  
Ley N° 28806, Segunda Disp. Final y Trans. (Ley General de Inspección del Trabajo)

| CIU REV. 3 | CIU REV. 2 | ACTIVIDAD                 |
|------------|------------|---------------------------|
|            |            | 122 EXTRACCION DE MADERA. |
| 0200 3     | 1220 00    | Extracción de madera.     |

Mayor información en:

<[http://www.mintra.gob.pe/archivos/file/normasLegales/DS\\_009\\_1997\\_SA.pdf](http://www.mintra.gob.pe/archivos/file/normasLegales/DS_009_1997_SA.pdf)>[Consulta da: 06 de enero del 2013].

## ANEXO 4:

### Códigos en R para la creación de la Figura 2: Tipos de asimetrías presentados por una distribución

**#Instalar los siguientes paquetes del CRAN mirror: e1071, sn y MASS**

```
library(sn)
x<-seq(-4,1,0.1); y<-seq(-3,3,0.1); z<-seq(-1,4,0.1)
fx<-dsn(x,0,1,-5); fy<-dsn(y,0,1,0); fz<-dsn(z,0,1,5)
par(mfrow=c(1,3));
plot(x,fx,type="l",main="(a)");
plot(y,fy,type="l",main="(b)");
plot(z,fz,type="l",main="(c)");
```

## ANEXO 5:

### Códigos en R para la creación de la Figura 3: Gráfico de la función de distribución *skew-normal* estándar con valores diferentes de $\lambda$

```
par(mfrow=c(1,1))

z<-seq(-20,20,0.1)
fz<-dsn(z,location=0,scale=1,shape=0)
plot(z,fz,type="l", xlim=c(-4,4),
ylim=c(0,0.8),col=1,lwd=2,lty=2)

fz<-dsn(z,location=0,scale=1,shape=-10)
points(z,fz,type="l",col=2,lwd=2,lty=2)

fz<-dsn(z,location=0,scale=1,shape=10)
points(z,fz,type="l",col=3,lwd=2,lty=2)

not<-c("lambda=0","lambda=-10","lambda=10")
legend(1.3,0.8,not,col=c(1,2,3),lty=c(2,2,2))
```

## ANEXO 6:

### Códigos en R para la creación de la Figura 4: Gráficos de las funciones de densidad de la distribución skew-t:

(a) Parámetro de asimetría fijo. (b) Parámetro de curtosis fijo.

```
par(mfrow=c(2,1))
#parametro de asimetria fijo
x<-seq(-20,20,0.1)
fx<-dst(x,location=0,scale=1,shape=11,df=1)
plot(x,fx,type="l", xlim=c(-1,4), ylim=c(0,1),col=1,main="(a)
Parámetro de Asimetría fijo")
fx<-dst(x,location=0,scale=1,shape=11,df=3)
points(x,fx,type="l",col=2)
fx<-dst(x,location=0,scale=1,shape=11,df=20)
points(x,fx,type="l",col=3)
not<-c("St(1,11)", "St(3,11)", "St(20,11)")
legend(2,1,not,col=1:3,lty=1)
#parametro de curtosis fijo
x<-seq(-20,20,0.1)
fx<-dst(x,location=0,scale=1,shape=0,df=7)
plot(x,fx,type="l", xlim=c(-2,3), ylim=c(0,1),col=1,main="(b)
Parámetro de Curtosis fijo")
fx<-dst(x,location=0,scale=1,shape=3,df=7)
points(x,fx,type="l",col=2)
fx<-dst(x,location=0,scale=1,shape=11,df=7)
points(x,fx,type="l",col=3)
fx<-dst(x,location=0,scale=1,shape=52,df=7)
points(x,fx,type="l",col=4)
not<-c("St(7,0)", "St(7,3)", "St(7,11)", "St(7,52)")
legend(1.5,1,not,col=1:4,lty=1)
```

## ANEXO 7:

### Códigos en R para la obtención de los Figura 5: Gráficos descriptivos del coste de los siniestros para los grupos de Invalidez (izquierda) y Sobrevivencia (derecha).

```
library(e1071)
stat<-c("Nº of observations", "E (X)", "St.Dev (X)", "Skewness
(X)", "Kurtosis (X)", "Minimum", "Maximum", "99% Quantile (Value
at Risk)", "E(X|X>Value at Risk)")

# Cambiar la carpeta desde donde se leen los archivos:
ARCHIVO -->CAMBIAR DIR...

#### Datos Originales Invalidez ####
data1<-read.table("invalidezPT.txt", header=T)
str(data1)
summary(data1)
Inv<-data1[,1]/1
str(Inv)
summary(Inv)
sd(Inv)
skewness(Inv)
kurtosis(Inv)
quantile(Inv,0.99)
mean(Inv[Inv>=quantile(Inv,0.99)])
#hist(Inv, breaks=50, main="Histograma de Invalidez")
#qqnorm(Inv); qqline(Inv, col = 2)
Inv_stat<-
round(c(length(Inv), mean(Inv), sd(Inv), skewness(Inv), kurtosis(
Inv), min(Inv), max(Inv), quantile(Inv,0.99), mean(Inv[Inv>=quant
ile(Inv,0.99)])), 2)

#### Datos Originales Sobrevivencia ####
```

```

data2<-read.table("sobrevivenciaPP.txt", header=T)
str(data2)
summary(data2)
Sob<-data2[,1]/1
str(Sob)
summary(Sob)
sd(Sob)
skewness(Sob)
kurtosis(Sob)
quantile(Sob,0.99)
mean(Sob[Sob>=quantile(Sob,0.99)])
#hist(Sob,breaks=50,main=" Histograma de Supervivencia")
#qqnorm(Sob); qqline(Sob, col = 2)
Sob_stat<-
round(c(length(Sob),mean(Sob),sd(Sob),skewness(Sob),kurtosis(
Sob),min(Sob),max(Sob),quantile(Sob,0.99),mean(Sob[Sob>=quant
ile(Sob,0.99)])),2)

```

**# Cuadro 2: Estadísticas descriptivas del coste de los siniestros para los grupos de Invalidez y Supervivencia**

```

table1<-
data.frame(Invalidez=Inv_stat,Sobrevivencia=Sob_stat);
rownames(table1)<-stat;table1

```

**# Figura 5: Gráficos descriptivos del coste de los siniestros para los grupos de Invalidez (izquierda) y Supervivencia (derecha)**

```

par(mfrow=c(2,2)); hist(Inv, breaks=50, main="Histograma de
Invalidez"); hist(Sob, breaks=50, main="Histograma de
Supervivencia"); qqnorm(Inv); qqline(Inv, col = 2);
qqnorm(Sob); qqline(Sob, col=2)

```



## ANEXO 8:

### Códigos en R para la obtención de la Figura 6: Gráficos descriptivos del logaritmo del coste de los siniestros para los grupos de Invalidez (izquierda) y Sobrevivencia (derecha).

```
#### Log Datos logInv ####
logInv<-log(Inv)
str(logInv)
summary(logInv)
sd(logInv)
skewness(logInv)
kurtosis(logInv)
quantile(logInv,0.99)
mean(logInv [logInv>=quantile(logInv,0.99)])
#hist(logInv, breaks=50)
#qqnorm(logInv); qqline(logInv, col = 2)
logInv_stat<-
round(c(length(logInv),mean(logInv),sd(logInv),skewness(logInv),
kurtosis(logInv),min(logInv),max(logInv),quantile(logInv,0.99),
mean(logInv[logInv>=quantile(logInv,0.99)])),2)

#### Log Datos logSob ####
logSob<-log(Sob)
str(logSob)
summary(logSob)
sd(logSob)
skewness(logSob)
kurtosis(logSob)
quantile(logSob,0.99)
mean(logSob[logSob>=quantile(logSob,0.99)])
#hist(logSob, breaks=50)
#qqnorm(logSob); qqline(logSob, col = 2)
```

```
logSob_stat<-
round(c(length(logSob),mean(logSob),sd(logSob),skewness(logSob),kurtosis(logSob),min(logSob),max(logSob),quantile(logSob,0.99),mean(logSob[logSob>=quantile(logSob,0.99)])),2)
```

**# Cuadro 3: Estadísticas descriptivas del logaritmo del coste de los siniestros para los grupos de Invalidez y Sobrevivencia**

```
table2<-
data.frame(Invalidez=logInv_stat,Sobrevivencia=logSob_stat);
rownames(table2)<-stat; table2
```

**# Figura 6: Gráficos descriptivos del logaritmo del coste de los siniestros para los grupos de Invalidez (izquierda) y Sobrevivencia (derecha)**

```
par(mfrow=c(2,2)); hist(logInv, breaks=50, main="Histograma de Log Invalidez"); hist(logSob, breaks=50, main="Histograma de Log Sobrevivencia"); qqnorm(logInv); qqline(logInv, col = 2); qqnorm(logSob); qqline(logSob, col=2)
```

## ANEXO 9:

**Cuadro 8: Parámetros estimados de las distribuciones *exponencial, normal, logística, log-normal y weibull***

|                           | <b>Invalidéz</b> | <b>Sobrevivencia</b> | <b>logInvalidéz</b> | <b>logSobrevivencia</b> |
|---------------------------|------------------|----------------------|---------------------|-------------------------|
| <b><i>Exponencial</i></b> |                  |                      |                     |                         |
| Rate                      | 3.8e-04          | 6.0e-04              | 0.131               | 0.139                   |
| <b><i>Normal</i></b>      |                  |                      |                     |                         |
| Media                     | 2.6e+03          | 1.7e+03              | 7.645               | 7.201                   |
| Desviación Estándar       | 1.7e+03          | 1.3e+03              | 0.728               | 0.632                   |
| <b><i>Logística</i></b>   |                  |                      |                     |                         |
| Locación                  | 2.5e+03          | 1.4e+03              | 7.685               | 7.175                   |
| Escala                    | 9.6e+02          | 5.9e+02              | 0.428               | 0.362                   |
| <b><i>Log-Normal</i></b>  |                  |                      |                     |                         |
| Media Log                 | 7.6e+00          | 7.2e+00              | 2.029               | 1.970                   |
| Desviación Estándar Log   | 7.3e-01          | 6.3e-01              | 0.098               | 0.087                   |
| <b><i>Weibull</i></b>     |                  |                      |                     |                         |
| Forma                     | 1.7e+00          | 1.5e+00              | 12.773              | 11.423                  |
| Escala                    | 3.0e+03          | 1.8e+03              | 7.966               | 7.497                   |

Fuente: Elaboración propia

## ANEXO 10:

### Códigos en R para la creación de los Cuadros 4, 5, 6, 7 y 8

#### **# Ajuste de la Distribución, Comparación del Modelo y Bondad de Ajuste**

```
library(MASS)
mod<-c("skew-normal", "skew-
t", "exponencial", "geometrica", "cauchy", "gamma", "normal", "logi
stica", "log-normal", "t(no-central)", "weibull")
mod_ks<-c("Valor critico", "skew-normal", "skew-
t", "normal", "exponencial", "log-normal", "logistica", "weibull")
mod_VR<-c("skew-normal", "skew-t", "normal", "exponencial", "log-
normal", "logistica", "weibull", "empirico")
sim<-
c("FALSE", "FALSE", "FALSE", "FALSE", "TRUE", "FALSE", "TRUE", "TRUE
", "FALSE", "TRUE", "FALSE"); npar<-c(3, 4, 1, 1, 2, 2, 2, 2, 2, 3, 2)
a<-matrix(0, nrow=200, ncol=1);
VR.sn.Inv<-matrix(0, nrow=200, ncol=1); TVR.sn.Inv<-
matrix(0, nrow=200, ncol=1); VR.sn.Sob<-
matrix(0, nrow=200, ncol=1); TVR.sn.Sob<-
matrix(0, nrow=200, ncol=1)
VR.st.Inv<-matrix(0, nrow=200, ncol=1); TVR.st.Inv<-
matrix(0, nrow=200, ncol=1); VR.st.Sob<-
matrix(0, nrow=200, ncol=1); TVR.st.Sob<-
matrix(0, nrow=200, ncol=1)
VR.norm.Inv<-matrix(0, nrow=200, ncol=1); TVR.norm.Inv<-
matrix(0, nrow=200, ncol=1); VR.norm.Sob<-
matrix(0, nrow=200, ncol=1); TVR.norm.Sob<-
matrix(0, nrow=200, ncol=1)
VR.empi.Inv<-matrix(0, nrow=200, ncol=1); TVR.empi.Inv<-
matrix(0, nrow=200, ncol=1); VR.empi.Sob<-
```

```

matrix(0,nrow=200,ncol=1);      TVR.empi.Sob<-
matrix(0,nrow=200,ncol=1);

#Invalidez

#Ajuste de la Distribución - Invalidez
a1.sn<-sn.mle(y=Inv,plot.it=FALSE,trace=FALSE);
  sn_par<-
cp.to.dp(c(a1.sn$cp[1],a1.sn$cp[2],a1.sn$cp[3]));
  sn_llh<-a1.sn$logL
a2.st<-st.mle(y=Inv,trace=FALSE);          st_par<-
a2.st$dp;                                st_llh<-a2.st$logL
#a2.st<-st.mle(y=Inv,fixed.df=4,trace=FALSE);
  st_par<-a2.st$dp;
  st_llh<-a2.st$logL
#a3<-
stepAIC.ghyp(Inv,dist=c("ghyp","VG","NIG","hyp","gauss"),symm
etric=NULL,control=list(maxit=500),silent=TRUE);
  hyp_dis<-a3$fit.table
a4.exp<-fitdistr(Inv,"exponential");          exp_par<-
a4.exp$estimate;                            exp_llh<-
a4.exp$loglik
a4.geo<-fitdistr(Inv,"geometric");          geo_par<-
a4.geo$estimate;                            geo_llh<-
a4.geo$loglik
a4.cau<-fitdistr(Inv,"cauchy");            cau_par<-
a4.cau$estimate;                            cau_llh<-
a4.cau$loglik
a4.gam<-fitdistr(Inv,"gamma");              gam_par<-
a4.gam$estimate;                            gam_llh<-
a4.gam$loglik
a4.nor<-fitdistr(Inv,"normal");            nor_par<-
a4.nor$estimate;                            nor_llh<-
a4.nor$loglik

```

```

a4.logi<-fitdistr(Inv,"logistic");          logi_par<-
a4.logi$estimate;                          logi_llh<-
a4.logi$loglik
a4.logn<-fitdistr(Inv,"log-normal");      logn_par<-
a4.logn$estimate;                          logn_llh<-
a4.logn$loglik
a4.t<-fitdistr(Inv,"t");                  t_par<-a4.t$estimate;
                                           t_llh<-a4.t$loglik
a4.wei<-fitdistr(Inv,"weibull");          wei_par<-
a4.wei$estimate;                          wei_llh<-
a4.wei$loglik
Inv_par<-c(sn_par,st_par); Inv_par
Inv_par2<-c(exp_par,nor_par,logi_par,logn_par,wei_par);
Inv_par2

#Comparación del Modelo - Invalidez
#llh_Inv<-
c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,gam_llh,nor_llh,logi_
llh,logn_llh,t_llh,wei_llh)
llh_Inv<-c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,-
9999,nor_llh,logi_llh,logn_llh,-9999,wei_llh)
aic_Inv<-(-2)*(llh_Inv)+2*npar
tab_Inv<-
data.frame(Modelo=mod,Simetria=sim,n_parametros=npar,Log_Vero
similitud=llh_Inv,AIC=aic_Inv)
max_llh_Inv<-subset(tab_Inv,tab_Inv[,4]==max(tab_Inv[,4]));
max_llh_Inv
min_aic_Inv<-subset(tab_Inv,tab_Inv[,5]==min(tab_Inv[,5]));
min_aic_Inv

#Bondad de Ajuste - Invalidez
valor_critico<-1.36/(length(Inv)^0.5)

```

```

ks_sn<-
ks.test(Inv,"psn",location=sn_par[1],scale=sn_par[2],shape=sn
_par[3])
ks_st<-
ks.test(Inv,"pst",location=st_par[1],scale=st_par[2],shape=st
_par[3],df=st_par[4])
ks_norm<-ks.test(Inv,"pnorm",m=nor_par[1],sd=nor_par[2])
ks_exp<-ks.test(Inv,"pexp",rate=exp_par[1])
ks_logn<-
ks.test(Inv,"plnorm",meanlog=logn_par[1],sdlog=logn_par[2])
ks_logi<-
ks.test(Inv,"plogis",location=logi_par[1],scale=logi_par[2])
ks_wei<-
ks.test(Inv,"pweibull",shape=wei_par[1],scale=wei_par[2])
ks_Inv<-
c(valor_critico,ks_sn$statistic,ks_st$statistic,ks_norm$stati
stic,ks_exp$statistic,ks_logn$statistic,ks_logi$statistic,ks_
wei$statistic)

#Simulación de números aleatorios y Determinación del Valor
en Riesgo y el Valor en Riesgo Condicional - Invalidez
#a2.st<-st.mle(y=Inv,fixed.df=4,trace=FALSE);
  st_par<-a2.st$dp;
  st_llh<-a2.st$logL
n_sn<-
rsn(n=1000000,location=sn_par[1],scale=sn_par[2],shape=sn_par
[3])
n_st<-
rst(n=1000000,location=st_par[1],scale=st_par[2],shape=st_par
[3],df=st_par[4])
n_norm<-rnorm(n=1000000,m=nor_par[1],sd=nor_par[2])
n_exp<-rexp(n=1000000,rate=exp_par[1])
n_logn<-
rlnorm(n=1000000,meanlog=logn_par[1],sdlog=logn_par[2])

```

```

n_logi<-
rlogis(n=1000000,location=logi_par[1],scale=logi_par[2])
n_wei<-rweibull(n=1000000,shape=wei_par[1],scale=wei_par[2])
Inv_VR<-
c(quantile(n_sn,0.99),quantile(n_st,0.99),quantile(n_norm,0.99),
quantile(n_exp,0.99),quantile(n_logn,0.99),quantile(n_logi,0.99),
quantile(n_wei,0.99),quantile(Inv,0.99))
Inv_TVR<-
c(mean(n_sn[n_sn>=quantile(n_sn,0.99)]),mean(n_st[n_st>=quantile(n_st,0.99)]),
mean(n_norm[n_norm>=quantile(n_norm,0.99)]),
mean(n_exp[n_exp>=quantile(n_exp,0.99)]),mean(n_logn[n_logn>=
quantile(n_logn,0.99)]),mean(n_logi[n_logi>=quantile(n_logi,0.99)]),
mean(n_wei[n_wei>=quantile(n_wei,0.99)]),mean(Inv[Inv>=quantile(Inv,0.99)]))

#Sobrevivencia

#Ajuste de la Distribución - Sobrevivencia
a1.sn<-sn.mle(y=Sob,plot.it=FALSE,trace=FALSE);
  sn_par<-
cp.to.dp(c(a1.sn$cp[1],a1.sn$cp[2],a1.sn$cp[3]));
  sn_llh<-a1.sn$logL
a2.st<-st.mle(y=Sob,trace=FALSE);          st_par<-
a2.st$dp;          st_llh<-a2.st$logL
#a2.st<-st.mle(y=Sob,fixed.df=4,trace=FALSE);
  st_par<-a2.st$dp;
  st_llh<-a2.st$logL
a4.exp<-fitdistr(Sob,"exponential");          exp_par<-
a4.exp$estimate;          exp_llh<-
a4.exp$loglik
a4.geo<-fitdistr(Sob,"geometric");          geo_par<-
a4.geo$estimate;          geo_llh<-
a4.geo$loglik

```



```

a4.cau<-fitdistr(Sob,"cauchy");
a4.cau$estimate;
a4.cau$loglik
a4.gam<-fitdistr(Sob,"gamma");
a4.gam$estimate;
a4.gam$loglik
a4.nor<-fitdistr(Sob,"normal");
a4.nor$estimate;
a4.nor$loglik
a4.logi<-fitdistr(Sob,"logistic");
a4.logi$estimate;
a4.logi$loglik
a4.logn<-fitdistr(Sob,"log-normal");
a4.logn$estimate;
a4.logn$loglik
a4.t<-fitdistr(Sob,"t");
a4.wei<-fitdistr(Sob,"weibull");
a4.wei$estimate;
a4.wei$loglik
Sob_par<-c(sn_par,st_par); Sob_par
Sob_par2<-c(exp_par,nor_par,logi_par,logn_par,wei_par);
Sob_par2

#Comparación del Modelo - Sobrevivencia
llh_Sob<-
c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,gam_llh,nor_llh,logi_
llh,logn_llh,t_llh,wei_llh)
#llh_Sob<-c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,-
9999,nor_llh,logi_llh,logn_llh,t_llh,wei_llh)
aic_Sob<-(-2)*(llh_Sob)+2*npar
tab_Sob<-
data.frame(Modelo=mod,Simetria=sim,n_parametros=npar,Log_Vero
similitud=llh_Sob,AIC=aic_Sob)

```

```

max_llh_Sob<-subset(tab_Sob,tab_Sob[,4]==max(tab_Sob[,4]));
  max_llh_Sob
min_aic_Sob<-subset(tab_Sob,tab_Sob[,5]==min(tab_Sob[,5]));
  min_aic_Sob

#Bondad de Ajuste - Supervivencia
valor_critico<-1.36/(length(Sob)^0.5)
ks_sn<-
ks.test(Sob,"psn",location=sn_par[1],scale=sn_par[2],shape=sn
_par[3])
ks_st<-
ks.test(Sob,"pst",location=st_par[1],scale=st_par[2],shape=st
_par[3],df=st_par[4])
ks_norm<-ks.test(Sob,"pnorm",m=nor_par[1],sd=nor_par[2])
ks_exp<-ks.test(Sob,"pexp",rate=exp_par[1])
ks_logn<-
ks.test(Sob,"plnorm",meanlog=logn_par[1],sdlog=logn_par[2])
ks_logi<-
ks.test(Sob,"plogis",location=logi_par[1],scale=logi_par[2])
ks_wei<-
ks.test(Sob,"pweibull",shape=wei_par[1],scale=wei_par[2])
ks_Sob<-
c(valor_critico,ks_sn$statistic,ks_st$statistic,ks_norm$stati
stic,ks_exp$statistic,ks_logn$statistic,ks_logi$statistic,ks_
wei$statistic)

#Simulación de números aleatorios y Determinación del Valor
en Riesgo y el Valor en Riesgo Condicional - Supervivencia
a2.st<-st.mle(y=Sob,fixed.df=4,trace=FALSE);      st_par<-
a2.st$dp;          st_llh<-a2.st$logL
n_sn<-
rsn(n=1000000,location=sn_par[1],scale=sn_par[2],shape=sn_par
[3])

```

```

n_st<-
rst(n=1000000,location=st_par[1],scale=st_par[2],shape=st_par
[3],df=st_par[4])
n_norm<-rnorm(n=1000000,m=nor_par[1],sd=nor_par[2])
n_exp<-rexp(n=1000000,rate=exp_par[1])
n_logn<-
rlnorm(n=1000000,meanlog=logn_par[1],sdlog=logn_par[2])
n_logi<-
rlogis(n=1000000,location=logi_par[1],scale=logi_par[2])
n_wei<-rweibull(n=1000000,shape=wei_par[1],scale=wei_par[2])
Sob_VR<-
c(quantile(n_sn,0.99),quantile(n_st,0.99),quantile(n_norm,0.9
9),quantile(n_exp,0.99),quantile(n_logn,0.99),quantile(n_logi
,0.99),quantile(n_wei,0.99),quantile(Sob,0.99))
Sob_TVR<-
c(mean(n_sn[n_sn>=quantile(n_sn,0.99)]),mean(n_st[n_st>=quant
ile(n_st,0.99)]),mean(n_norm[n_norm>=quantile(n_norm,0.99)]),
mean(n_exp[n_exp>=quantile(n_exp,0.99)]),mean(n_logn[n_logn>=
quantile(n_logn,0.99)]),mean(n_logi[n_logi>=quantile(n_logi,0
.99)]),mean(n_wei[n_wei>=quantile(n_wei,0.99)]),mean(Sob[Sob>
=quantile(Sob,0.99)]))

```

## **#LogInvaliddez**

```

#Ajuste de la Distribución - LogInvaliddez
a1.sn<-sn.mle(y=logInv,plot.it=FALSE,trace=FALSE);
  sn_par<-
cp.to.dp(c(a1.sn$cp[1],a1.sn$cp[2],a1.sn$cp[3]));
  sn_llh<-a1.sn$logL
a2.st<-st.mle(y=logInv,trace=FALSE);          st_par<-
a2.st$dp;          st_llh<-a2.st$logL
#a2.st<-st.mle(y=logInv,fixed.df=4,trace=FALSE); st_par<-
a2.st$dp;          st_llh<-a2.st$logL

```

```

#a3<-
stepAIC.ghyp(logInv,dist=c("ghyp","VG","NIG","hyp","gauss"),s
ymmetric=NULL,control=list(maxit=500),silent=TRUE);
    hyp_dis<-a3$fit.table
a4.exp<-fitdistr(logInv,"exponential");           exp_par<-
a4.exp$estimate;                                exp_llh<-
a4.exp$loglik
a4.geo<-fitdistr(logInv,"geometric");           geo_par<-
a4.geo$estimate;                                geo_llh<-
a4.geo$loglik
a4.cau<-fitdistr(logInv,"cauchy");             cau_par<-
a4.cau$estimate;                                cau_llh<-
a4.cau$loglik
a4.gam<-fitdistr(logInv,"gamma");             gam_par<-
a4.gam$estimate;                                gam_llh<-
a4.gam$loglik
a4.nor<-fitdistr(logInv,"normal");           nor_par<-
a4.nor$estimate;                                nor_llh<-
a4.nor$loglik
a4.logi<-fitdistr(logInv,"logistic");         logi_par<-
a4.logi$estimate;                                logi_llh<-
a4.logi$loglik
a4.logn<-fitdistr(logInv,"log-normal");       logn_par<-
a4.logn$estimate;                                logn_llh<-
a4.logn$loglik
a4.t<-fitdistr(logInv,"t");                 t_par<-
a4.t$estimate;                                t_llh<-
a4.t$loglik
a4.wei<-fitdistr(logInv,"weibull");         wei_par<-
a4.wei$estimate;                                wei_llh<-
a4.wei$loglik
logInv_par<-c(sn_par,st_par); logInv_par
logInv_par2<-c(exp_par,nor_par,logi_par,logn_par,wei_par);
logInv_par2

```

```

#Comparación del Modelo - logInvalidez
#llh_logInv<-
c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,gam_llh,nor_llh,logi_
llh,logn_llh,t_llh,wei_llh)
llh_logInv<-
c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,gam_llh,nor_llh,logi_
llh,logn_llh,-9999,wei_llh)
aic_logInv<-(-2)*(llh_logInv)+2*npar
tab_logInv<-
data.frame(Modelo=mod,Simetria=sim,n_parametros=npar,Log_Vero
similitud=llh_logInv,AIC=aic_logInv)
max_llh_logInv<-
subset(tab_logInv,tab_logInv[,4]==max(tab_logInv[,4]));
      max_llh_logInv
min_aic_logInv<-
subset(tab_logInv,tab_logInv[,5]==min(tab_logInv[,5]));
      min_aic_logInv

#Bondad de Ajuste - logInvalidez
valor_critico<-1.36/(length(logInv)^0.5)
ks_sn<-
ks.test(logInv,"psn",location=sn_par[1],scale=sn_par[2],shape
=sn_par[3])
ks_st<-
ks.test(logInv,"pst",location=st_par[1],scale=st_par[2],shape
=st_par[3],df=st_par[4])
ks_norm<-ks.test(logInv,"pnorm",m=nor_par[1],sd=nor_par[2])
ks_exp<-ks.test(logInv,"pexp",rate=exp_par[1])
ks_logn<-
ks.test(logInv,"plnorm",meanlog=logn_par[1],sdlog=logn_par[2]
)

```

```

ks_logi<-
ks.test(logInv,"plogis",location=logi_par[1],scale=logi_par[2]
)
ks_wei<-
ks.test(logInv,"pweibull",shape=wei_par[1],scale=wei_par[2])
ks_logInv<-
c(valor_critico,ks_sn$statistic,ks_st$statistic,ks_norm$statistic,ks_exp$statistic,ks_logn$statistic,ks_logi$statistic,ks_wei$statistic)

#Simulación de números aleatorios y Determinación del Valor en Riesgo y el Valor en Riesgo Condicional - logInvalidez
#a2.st<-st.mle(y=logInv,fixed.df=4,trace=FALSE); st_par<-
a2.st$dp; st_llh<-a2.st$logL
n_sn<-a5.sn.Inv<-
rsn(n=1000000,location=sn_par[1],scale=sn_par[2],shape=sn_par[3])
n_st<-a5.st.Inv<-
rst(n=1000000,location=st_par[1],scale=st_par[2],shape=st_par[3],df=st_par[4])
n_norm<-a5.norm.Inv<-
rnorm(n=1000000,m=nor_par[1],sd=nor_par[2])
n_exp<-rexp(n=1000000,rate=exp_par[1])
n_logn<-
rlnorm(n=1000000,meanlog=logn_par[1],sdlog=logn_par[2])
n_logi<-
rlogis(n=1000000,location=logi_par[1],scale=logi_par[2])
n_wei<-rweibull(n=1000000,shape=wei_par[1],scale=wei_par[2])
logInv_VR<-
c(quantile(n_sn,0.99),quantile(n_st,0.99),quantile(n_norm,0.99),quantile(n_exp,0.99),quantile(n_logn,0.99),quantile(n_logi,0.99),quantile(n_wei,0.99),quantile(logInv,0.99))
logInv_TVR<-
c(mean(n_sn[n_sn>=quantile(n_sn,0.99)]),mean(n_st[n_st>=quant

```

```

ile(n_st,0.99)]),mean(n_norm[n_norm>=quantile(n_norm,0.99)]),
mean(n_exp[n_exp>=quantile(n_exp,0.99)]),mean(n_logn[n_logn>=
quantile(n_logn,0.99)]),mean(n_logi[n_logi>=quantile(n_logi,0
.99)]),mean(n_wei[n_wei>=quantile(n_wei,0.99)]),mean(logInv[l
ogInv>=quantile(logInv,0.99)]))

```

## **#logSobrevivencia**

```

#Ajuste de la Distribución - logSobrevivencia

```

```

a1.sn<-sn.mle(y=logSob,plot.it=FALSE,trace=FALSE);

```

```

    sn_par<-

```

```

cp.to.dp(c(a1.sn$cp[1],a1.sn$cp[2],a1.sn$cp[3]));

```

```

    sn_llh<-a1.sn$logL

```

```

a2.st<-st.mle(y=logSob,trace=FALSE);

```

```

    st_par<-

```

```

a2.st$dp;

```

```

    st_llh<-a2.st$logL

```

```

#a2.st<-st.mle(y=logSob,fixed.df=4,trace=FALSE); st_par<-

```

```

a2.st$dp;

```

```

    st_llh<-a2.st$logL

```

```

#a3<-

```

```

stepAIC.ghyp(logSob,dist=c("ghyp","VG","NIG","hyp","gauss"),s

```

```

ymmetric=NULL,control=list(maxit=500),silent=TRUE);

```

```

    hyp_dis<-a3$fit.table

```

```

a4.exp<-fitdistr(logSob,"exponential");

```

```

    exp_par<-

```

```

a4.exp$estimate;

```

```

    exp_llh<-

```

```

a4.exp$loglik

```

```

a4.geo<-fitdistr(logSob,"geometric");

```

```

    geo_par<-

```

```

a4.geo$estimate;

```

```

    geo_llh<-

```

```

a4.geo$loglik

```

```

a4.cau<-fitdistr(logSob,"cauchy");

```

```

    cau_par<-

```

```

a4.cau$estimate;

```

```

    cau_llh<-

```

```

a4.cau$loglik

```

```

a4.gam<-fitdistr(logSob,"gamma");

```

```

    gam_par<-

```

```

a4.gam$estimate;

```

```

    gam_llh<-

```

```

a4.gam$loglik

```

```

a4.nor<-fitdistr(logSob,"normal");
a4.nor$estimate;
a4.nor$loglik
a4.logi<-fitdistr(logSob,"logistic");
a4.logi$estimate;
a4.logi$loglik
a4.logn<-fitdistr(logSob,"log-normal");
a4.logn$estimate;
a4.logn$loglik
a4.t<-fitdistr(logSob,"t");
a4.t$estimate;
a4.t$loglik
a4.wei<-fitdistr(logSob,"weibull");
a4.wei$estimate;
a4.wei$loglik
logSob_par<-c(sn_par,st_par); logSob_par
logSob_par2<-c(exp_par,nor_par,logi_par,logn_par,wei_par);
logSob_par2

#Comparación del Modelo - logSobrevivencia
#llh_logSob<-
c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,gam_llh,nor_llh,logi_
llh,logn_llh,t_llh,wei_llh)
llh_logSob<-
c(sn_llh,st_llh,exp_llh,geo_llh,cau_llh,gam_llh,nor_llh,logi_
llh,logn_llh,-9999,wei_llh)
aic_logSob<-(-2)*(llh_logSob)+2*npar
tab_logSob<-
data.frame(Modelo=mod,Simetria=sim,n_parametros=npar,Log_Vero
similitud=llh_logSob,AIC=aic_logSob)
max_llh_logSob<-
subset(tab_logSob,tab_logSob[,4]==max(tab_logSob[,4]));
max_llh_logSob

```



```

min_aic_logSob<-
subset(tab_logSob,tab_logSob[,5]==min(tab_logSob[,5]));
      min_aic_logSob

#Bondad de Ajuste - Sobrevivencia
valor_critico<-1.36/(length(logSob)^0.5)
ks_sn<-
ks.test(logSob,"psn",location=sn_par[1],scale=sn_par[2],shape
=sn_par[3])
ks_st<-
ks.test(logSob,"pst",location=st_par[1],scale=st_par[2],shape
=st_par[3],df=st_par[4])
ks_norm<-ks.test(logSob,"pnorm",m=nor_par[1],sd=nor_par[2])
ks_exp<-ks.test(logSob,"pexp",rate=exp_par[1])
ks_logn<-
ks.test(logSob,"plnorm",meanlog=logn_par[1],sdlog=logn_par[2]
)
ks_logi<-
ks.test(logSob,"plogis",location=logi_par[1],scale=logi_par[2]
)
ks_wei<-
ks.test(logSob,"pweibull",shape=wei_par[1],scale=wei_par[2])
ks_logSob<-
c(valor_critico,ks_sn$statistic,ks_st$statistic,ks_norm$stati
stic,ks_exp$statistic,ks_logn$statistic,ks_logi$statistic,ks_
wei$statistic)

#Simulación de números aleatorios y Determinación del Valor
en Riesgo y el Valor en Riesgo Condicional- logSobrevivencia
#a2.st<-st.mle(y=logSob,fixed.df=4,trace=FALSE); st_par<-
a2.st$dp;          st_llh<-a2.st$logL
n_sn<-a5.sn.Sob<-
rsn(n=1000000,location=sn_par[1],scale=sn_par[2],shape=sn_par
[3])

```

```

n_st<-a5.st.Sob<-
rst(n=1000000,location=st_par[1],scale=st_par[2],shape=st_par
[3],df=st_par[4])
n_norm<-a5.norm.Sob<-
rnorm(n=1000000,m=nor_par[1],sd=nor_par[2])
n_exp<-rexp(n=1000000,rate=exp_par[1])
n_logn<-
rlnorm(n=1000000,meanlog=logn_par[1],sdlog=logn_par[2])
n_logi<-
rlogis(n=1000000,location=logi_par[1],scale=logi_par[2])
n_wei<-rweibull(n=1000000,shape=wei_par[1],scale=wei_par[2])

logSob_VR<-
c(quantile(n_sn,0.99),quantile(n_st,0.99),quantile(n_norm,0.9
9),quantile(n_exp,0.99),quantile(n_logn,0.99),quantile(n_logi
,0.99),quantile(n_wei,0.99),quantile(logSob,0.99))
logSob_TVR<-
c(mean(n_sn[n_sn>=quantile(n_sn,0.99)]),mean(n_st[n_st>=quant
ile(n_st,0.99)]),mean(n_norm[n_norm>=quantile(n_norm,0.99)]),
mean(n_exp[n_exp>=quantile(n_exp,0.99)]),mean(n_logn[n_logn>=
quantile(n_logn,0.99)]),mean(n_logi[n_logi>=quantile(n_logi,0
.99)]),mean(n_wei[n_wei>=quantile(n_wei,0.99)]),mean(logSob[1
ogSob>=quantile(logSob,0.99)]))

# Cuadro 4: Parámetros estimados de las distribuciones skew-
normal y skew-t
table3<-
data.frame(Invalidez=Inv_par,Sobrevivencia=Sob_par,logInvalid
ez=logInv_par,logSobrevivencia=logSob_par);
rownames(table3)<-c(paste("skew-
normal",names(sn_par)),paste("skew-student",names(st_par)));
table3

```

```
# Cuadro 8 - Anexo 8: Parámetros estimados de las distribuciones exponencial, normal, logística, log-normal y weibull
```

```
cuadro8<-
data.frame(Invalidez=Inv_par2,Sobrevivencia=Sob_par2,logInvalidez=logInv_par2,logSobrevivencia=logSob_par2);
rownames(cuadro8)<-
c(paste("exponencial",names(exp_par)),paste("normal",names(norm_par)),paste("logistica",names(logi_par)),paste("log-normal",names(logn_par)),paste("weibull",names(wei_par)));
cuadro8
```

```
# Cuadro 5: Criterio del Logaritmo de la función de Verosimilitud y Criterio de Información de Akaike (AIC)
```

```
table4<-
data.frame(Simetria=sim,N_par=npar,llh_Inv,llh_Sob,llh_logInv,llh_logSob,aic_Inv,aic_Sob,aic_logInv,aic_logSob);
rownames(table4)<-mod; table4
#table4[order(aic_Inv),(ncol(table4)-3):ncol(table4)];
table4[order(aic_Sob),(ncol(table4)-3):ncol(table4)];
table4[order(aic_logInv),(ncol(table4)-3):ncol(table4)];
table4[order(aic_logSob),(ncol(table4)-3):ncol(table4)]
```

```
# Cuadro 6: Prueba de bondad de ajuste Kolmogorov Smirnov
```

```
table5<-
round(data.frame(ks_Inv,ks_Sob,ks_logInv,ks_logSob),4);
rownames(table5)<-mod_ks; table5
#table5[order(ks_Inv),]; table5[order(ks_Sob),];
table5[order(ks_logInv),]; table5[order(ks_logSob),]
```

```
# Cuadro 7: Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) usando el percentil99%
```

```
table6.VR<-
data.frame(Invalidez=Inv_VR,Sobrevivencia=Sob_VR,logInvalidez
```

```
=logInv_VR,logSobrevivencia=logSob_VR);  
rownames(table6.VR)=mod_VR  
table6.TVR<-  
data.frame(Invalidez=Inv_TVR,Sobrevivencia=Sob_TVR,logInvalid  
ez=logInv_TVR,logSobrevivencia=logSob_TVR);  
rownames(table6.TVR)=mod_VR  
table6<-round(rbind(VR=table6.VR,TVR=table6.TVR),2); table6
```

## ANEXO 11:

**Tabla del Test de Kolmogorov-Smirnov sobre Bondad de Ajuste**

| <i>n</i> | <i>Nivel de significación <math>\alpha</math></i> |             |             |             |             |              |              |              |
|----------|---------------------------------------------------|-------------|-------------|-------------|-------------|--------------|--------------|--------------|
|          | <b>0.20</b>                                       | <b>0.10</b> | <b>0.05</b> | <b>0.02</b> | <b>0.01</b> | <b>0.005</b> | <b>0.002</b> | <b>0.001</b> |
| 1        | 0.90000                                           | 0.95000     | 0.97500     | 0.99000     | 0.99500     | 0.99750      | 0.99900      | 0.99950      |
| 2        | 0.68337                                           | 0.77639     | 0.84189     | 0.90000     | 0.92929     | 0.95000      | 0.96838      | 0.97764      |
| 3        | 0.56481                                           | 0.63604     | 0.70760     | 0.78456     | 0.82900     | 0.86428      | 0.90000      | 0.92065      |
| 4        | 0.49265                                           | 0.56522     | 0.62394     | 0.68887     | 0.73424     | 0.77639      | 0.82217      | 0.85047      |
| 5        | 0.44698                                           | 0.50945     | 0.56328     | 0.62718     | 0.66853     | 0.70543      | 0.75000      | 0.78137      |
| 6        | 0.41037                                           | 0.46799     | 0.51926     | 0.57741     | 0.61661     | 0.65287      | 0.69571      | 0.72479      |
| 7        | 0.38148                                           | 0.43607     | 0.48342     | 0.53844     | 0.57581     | 0.60975      | 0.65071      | 0.67930      |
| 8        | 0.35831                                           | 0.40962     | 0.45427     | 0.50654     | 0.54179     | 0.57429      | 0.61368      | 0.64098      |
| 9        | 0.33910                                           | 0.38746     | 0.43001     | 0.47960     | 0.51332     | 0.54443      | 0.58210      | 0.60846      |
| 10       | 0.32260                                           | 0.36866     | 0.40925     | 0.45562     | 0.48893     | 0.51872      | 0.55500      | 0.58042      |
| 11       | 0.30829                                           | 0.35242     | 0.39122     | 0.43670     | 0.46770     | 0.49539      | 0.53135      | 0.55588      |
| 12       | 0.29577                                           | 0.33815     | 0.37543     | 0.41918     | 0.44905     | 0.47672      | 0.51047      | 0.53422      |
| 13       | 0.28470                                           | 0.32549     | 0.36143     | 0.40362     | 0.43247     | 0.45921      | 0.49189      | 0.51490      |
| 14       | 0.27481                                           | 0.31417     | 0.34890     | 0.38970     | 0.41762     | 0.44352      | 0.47520      | 0.49753      |
| 15       | 0.26589                                           | 0.30397     | 0.33750     | 0.37713     | 0.40420     | 0.42934      | 0.45611      | 0.48182      |
| 16       | 0.25778                                           | 0.29472     | 0.32733     | 0.36571     | 0.39201     | 0.41644      | 0.44637      | 0.46750      |
| 17       | 0.25039                                           | 0.28627     | 0.31796     | 0.35528     | 0.38086     | 0.40464      | 0.43380      | 0.45540      |
| 18       | 0.24360                                           | 0.27851     | 0.30936     | 0.34569     | 0.37062     | 0.39380      | 0.42224      | 0.44234      |
| 19       | 0.23735                                           | 0.27136     | 0.30143     | 0.33685     | 0.36117     | 0.38379      | 0.41156      | 0.43119      |
| 20       | 0.23156                                           | 0.26473     | 0.29408     | 0.32866     | 0.35241     | 0.37451      | 0.40165      | 0.42085      |
| 21       | 0.22517                                           | 0.25858     | 0.28724     | 0.32104     | 0.34426     | 0.36588      | 0.39243      | 0.41122      |
| 22       | 0.22115                                           | 0.25283     | 0.28087     | 0.31394     | 0.33666     | 0.35782      | 0.38382      | 0.40223      |
| 23       | 0.21646                                           | 0.24746     | 0.27491     | 0.30728     | 0.32954     | 0.35027      | 0.37575      | 0.39380      |
| 24       | 0.21205                                           | 0.24242     | 0.26931     | 0.30104     | 0.32286     | 0.34318      | 0.36787      | 0.38588      |
| 25       | 0.20790                                           | 0.23768     | 0.26404     | 0.29518     | 0.31657     | 0.33651      | 0.36104      | 0.37743      |
| 26       | 0.20399                                           | 0.23320     | 0.25908     | 0.28962     | 0.30963     | 0.33022      | 0.35431      | 0.37139      |
| 27       | 0.20030                                           | 0.22898     | 0.25438     | 0.28438     | 0.30502     | 0.32425      | 0.34794      | 0.36473      |
| 28       | 0.19680                                           | 0.22497     | 0.24993     | 0.27942     | 0.29971     | 0.31862      | 0.34190      | 0.35842      |
| 29       | 0.19348                                           | 0.22117     | 0.24571     | 0.27471     | 0.29466     | 0.31327      | 0.33617      | 0.35242      |
| 30       | 0.19032                                           | 0.21756     | 0.24170     | 0.27023     | 0.28986     | 0.30818      | 0.33072      | 0.34672      |
| 31       | 0.18732                                           | 0.21412     | 0.23788     | 0.26596     | 0.28529     | 0.30333      | 0.32553      | 0.34129      |
| 32       | 0.18445                                           | 0.21085     | 0.23424     | 0.26189     | 0.28094     | 0.29870      | 0.32058      | 0.33611      |
| 33       | 0.18171                                           | 0.20771     | 0.23076     | 0.25801     | 0.27577     | 0.29428      | 0.31584      | 0.33115      |
| 34       | 0.17909                                           | 0.21472     | 0.22743     | 0.25429     | 0.27271     | 0.29005      | 0.31131      | 0.32641      |
| 35       | 0.17659                                           | 0.20185     | 0.22425     | 0.25073     | 0.26897     | 0.28600      | 0.30597      | 0.32187      |
| 36       | 0.17418                                           | 0.19910     | 0.22119     | 0.24732     | 0.26532     | 0.28211      | 0.30281      | 0.31751      |
| 37       | 0.17188                                           | 0.19646     | 0.21826     | 0.24404     | 0.26180     | 0.27838      | 0.29882      | 0.31333      |
| 38       | 0.16966                                           | 0.19392     | 0.21544     | 0.24089     | 0.25843     | 0.27483      | 0.29498      | 0.30931      |
| 39       | 0.16753                                           | 0.19148     | 0.21273     | 0.23785     | 0.25518     | 0.27135      | 0.29125      | 0.30544      |

...continuación

| <i>n</i>                | <i>Nivel de significación <math>\alpha</math></i> |             |             |             |             |              |              |              |
|-------------------------|---------------------------------------------------|-------------|-------------|-------------|-------------|--------------|--------------|--------------|
|                         | <b>0.20</b>                                       | <b>0.10</b> | <b>0.05</b> | <b>0.02</b> | <b>0.01</b> | <b>0.005</b> | <b>0.002</b> | <b>0.001</b> |
| <b>40</b>               | 0.16547                                           | 0.18913     | 0.21012     | 0.23494     | 0.25205     | 0.26803      | 0.28772      | 0.30171      |
| <b>41</b>               | 0.16349                                           | 0.18687     | 0.20760     | 0.23213     | 0.24904     | 0.26482      | 0.28429      | 0.29811      |
| <b>42</b>               | 0.16158                                           | 0.18468     | 0.20517     | 0.22941     | 0.24613     | 0.26173      | 0.28097      | 0.29465      |
| <b>43</b>               | 0.15974                                           | 0.18257     | 0.20283     | 0.22679     | 0.24332     | 0.25875      | 0.27778      | 0.29130      |
| <b>44</b>               | 0.15795                                           | 0.18051     | 0.20056     | 0.22426     | 0.24060     | 0.25587      | 0.27468      | 0.28806      |
| <b>45</b>               | 0.15623                                           | 0.17856     | 0.19837     | 0.22181     | 0.23798     | 0.25308      | 0.27169      | 0.28493      |
| <b>46</b>               | 0.15457                                           | 0.17665     | 0.19625     | 0.21944     | 0.23544     | 0.25038      | 0.26880      | 0.28190      |
| <b>47</b>               | 0.15295                                           | 0.17481     | 0.19420     | 0.21715     | 0.23298     | 0.24776      | 0.26600      | 0.27896      |
| <b>48</b>               | 0.15139                                           | 0.17301     | 0.19221     | 0.21493     | 0.23059     | 0.24523      | 0.26328      | 0.27611      |
| <b>49</b>               | 0.14987                                           | 0.17128     | 0.19028     | 0.21281     | 0.22832     | 0.24281      | 0.26069      | 0.27339      |
| <b>50</b>               | 0.14840                                           | 0.16959     | 0.18841     | 0.21068     | 0.22604     | 0.24039      | 0.25809      | 0.27067      |
| <b><i>n</i> &gt; 50</b> | <b>1.07</b>                                       | <b>1.22</b> | <b>1.36</b> | <b>1.52</b> | <b>1.63</b> | <b>1.73</b>  | <b>1.85</b>  | <b>1.95</b>  |
|                         | <i>n</i>                                          | <i>n</i>    | <i>n</i>    | <i>n</i>    | <i>n</i>    | <i>n</i>     | <i>n</i>     | <i>n</i>     |

## ANEXO 12:

### Códigos en R para la obtención de la Figura 7: Valor en Riesgo (VaR) y Valor en Riesgo Condicional (TVaR) con percentiles del 90% al 99.9%

```
# Simulación del VaR y TVaR para percentiles variantes

for (i in 1:200) { a[i]=0.8995+i/2000; VR.sn.Inv[i]<-
quantile(a5.sn.Inv,a[i]); TVR.sn.Inv[i]<-
mean(a5.sn.Inv[a5.sn.Inv>=quantile(a5.sn.Inv,a[i])])}
for (i in 1:200) { a[i]=0.8995+i/2000; VR.st.Inv[i]<-
quantile(a5.st.Inv,a[i]); TVR.st.Inv[i]<-
mean(a5.st.Inv[a5.st.Inv>=quantile(a5.st.Inv,a[i])])}
for (i in 1:200) { a[i]=0.8995+i/2000; VR.norm.Inv[i]<-
quantile(a5.norm.Inv,a[i]); TVR.norm.Inv[i]<-
mean(a5.norm.Inv[a5.norm.Inv>=quantile(a5.norm.Inv,a[i])])}
for (i in 1:200) { a[i]=0.8995+i/2000; VR.empi.Inv[i]<-
quantile(logInv,a[i]); TVR.empi.Inv[i]<-
mean(logInv[logInv>=quantile(logInv,a[i])])}
for (i in 1:200) { a[i]=0.8995+i/2000; VR.sn.Sob[i]<-
quantile(a5.sn.Sob,a[i]); TVR.sn.Sob[i]<-
mean(a5.sn.Sob[a5.sn.Sob>=quantile(a5.sn.Sob,a[i])])}
for (i in 1:200) { a[i]=0.8995+i/2000; VR.st.Sob[i]<-
quantile(a5.st.Sob,a[i]); TVR.st.Sob[i]<-
mean(a5.st.Sob[a5.st.Sob>=quantile(a5.st.Sob,a[i])])}
for (i in 1:200) { a[i]=0.8995+i/2000; VR.norm.Sob[i]<-
quantile(a5.norm.Sob,a[i]); TVR.norm.Sob[i]<-
mean(a5.norm.Sob[a5.norm.Sob>=quantile(a5.norm.Sob,a[i])])}
for (i in 1:200) { a[i]=0.8995+i/2000; VR.empi.Sob[i]<-
quantile(logSob,a[i]); TVR.empi.Sob[i]<-
mean(logSob[logSob>=quantile(logSob,a[i])])}
```

```

# Figura 7: Valor en Riesgo y Valor en Riesgo Condicional
para niveles de confianza variantes
par(mfrow=c(2,2)); not2<-c("Empírica", "Skew-normal", "Skew-
t", "Normal")
g1<-plot(a,VR.empi.Inv,col=1,type="l",lty=3,lwd=3,main="a)
logInvalidez",xlab="Percentil",ylab="VaR");
  points(a,VR.sn.Inv,col=2,type="l",lwd=2);
  points(a,VR.st.Inv,col=4,type="l",lty=2,lwd=2);
  points(a,VR.norm.Inv,col=8,type="l",lwd=2)
g2<-plot(a,VR.empi.Sob,col=1,type="l",lty=3,lwd=3,main="b)
logSobrevivencia",xlab="Percentil",ylab="VaR");
  points(a,VR.sn.Sob,col=2,type="l",lwd=2);
  points(a,VR.st.Sob,col=4,type="l",lty=2,lwd=2);
  points(a,VR.norm.Sob,col=8,type="l",lwd=2)
g3<-plot(a,TVR.empi.Inv,col=1,type="l",lty=3,lwd=3,main="c)
logInvalidez",xlab="Percentil",ylab="TVaR");
  points(a,TVR.sn.Inv,col=2,type="l",lwd=2);
  points(a,TVR.st.Inv,col=4,type="l",lty=2,lwd=2);
  points(a,TVR.norm.Inv,col=8,type="l",lwd=2)
g4<-plot(a,TVR.empi.Sob,col=1,type="l",lty=3,lwd=3,main="d)
logSobrevivencia",xlab="Percentil",ylab="TVaR");
  points(a,TVR.sn.Sob,col=2,type="l",lwd=2);
  points(a,TVR.st.Sob,col=4,type="l",lty=2,lwd=2);
  points(a,TVR.norm.Sob,col=8,type="l",lwd=2);
legend(0.90,8.9,not2,lty=c(3,1,2,1),col=c(1,2,4,8),
lwd=c(3,2,2,2))

```



## ANEXO 12:

### Datos de los 2 conjuntos de grupos utilizados en la investigación

| <b>1) Coste (S/.) del Grupo con Cobertura por Sobrevivencia</b> |         |         |         |         |
|-----------------------------------------------------------------|---------|---------|---------|---------|
| 2712.24                                                         | 2090.67 | 3559.51 | 1888.43 | 714.09  |
| 522.56                                                          | 2519.79 | 988.57  | 670.43  | 465.12  |
| 2130.41                                                         | 1790.71 | 530.28  | 2133.34 | 1667.29 |
| 1121.89                                                         | 3190.33 | 895.84  | 322.09  | 6420.73 |
| 1067.53                                                         | 889.26  | 752.55  | 4273.39 | 736.66  |
| 985.8                                                           | 1020.91 | 4974.99 | 800.89  | 787.51  |
| 880.53                                                          | 1004.91 | 900.05  | 6732.56 | 764.67  |
| 1034.05                                                         | 868.79  | 534.4   | 722.31  | 1598.09 |
| 1341.66                                                         | 360.21  | 2695.05 | 1492.2  | 1134.92 |
| 794.77                                                          | 846.11  | 3486.57 | 951.64  | 1510.51 |
| 755.37                                                          | 7299.21 | 905.5   | 1975.92 | 2271.71 |
| 348.81                                                          | 604.5   | 2266.46 | 3744.02 | 1259.64 |
| 1267.13                                                         | 594     | 1342.26 | 1057.82 | 707.75  |
| 801.21                                                          | 816.47  | 2050.1  | 2431.46 | 1209.52 |
| 1203.81                                                         | 1786.87 | 650     | 1646.16 | 1500.06 |
| 1078.84                                                         | 2054.17 | 738.17  | 1466.22 | 3259.54 |
| 753.87                                                          | 655.82  | 786.7   | 2322.09 | 1132.84 |
| 2113.24                                                         | 1741.79 | 1167.45 | 1433.94 | 800     |
| 932.62                                                          | 1261.18 | 1028.96 | 2878.66 | 4573.3  |
| 786.31                                                          | 1814.03 | 647.44  | 1963.88 | 2327.2  |
| 1860.84                                                         | 807.54  | 1147.56 | 3205.06 | 1339.14 |
| 924.86                                                          | 681.63  | 1119.91 | 3266.22 | 4418.97 |
| 1422.86                                                         | 530     | 1018.88 | 898.19  | 1929.9  |
| 1636.65                                                         | 1562.37 | 700.84  | 2262.14 | 1654.17 |
| 1394.19                                                         | 2293.03 | 1279.09 | 2679.98 | 1426.76 |
| 6081.7                                                          | 2516.62 | 843.12  | 1047.06 | 850.8   |
| 1054.31                                                         | 1243.61 | 1272.7  | 1504.91 | 2186.52 |
| 481.74                                                          | 2131.34 | 976.73  | 782.07  | 1066.92 |
| 970.55                                                          | 1806.41 | 737.06  | 1190.93 | 360.22  |
| 772.94                                                          | 3323.62 | 2366.52 | 1262.51 | 652.83  |
| 457.42                                                          | 2362.16 | 2397.07 | 1525.56 | 538.5   |
| 1741.39                                                         | 781.86  | 2483.45 | 840.18  | 2626.34 |
| 2213.09                                                         | 960     | 2470.84 | 973.38  | 1038.9  |
| 1320.41                                                         | 1574.78 | 614.06  | 2208.38 | 1881.92 |
| 3430                                                            | 1193.77 | 1749.74 | 671.38  | 2011.29 |
| 5672.07                                                         | 1081.49 | 1686.82 | 413.15  | 2213.33 |
| 827.01                                                          | 1839.38 | 1088.04 | 1568.03 | 1305.57 |
| 982.38                                                          | 1928.76 | 1062.46 | 832.78  |         |
| 3229.84                                                         | 1063.66 | 1352.96 | 2717.9  |         |
| 5654.96                                                         | 1778.42 | 6712.64 | 460     |         |

...continuación

| <b>2) Coste (S/.) del Grupo con Cobertura por Invalidez</b> |         |         |         |         |         |         |         |
|-------------------------------------------------------------|---------|---------|---------|---------|---------|---------|---------|
| 1051.39                                                     | 667.81  | 2721.03 | 3412.17 | 1696.68 | 4382.69 | 1026.9  | 4058.33 |
| 1146.12                                                     | 2090.54 | 282.33  | 3647.82 | 3371.95 | 4655.04 | 724.39  | 563.66  |
| 460.74                                                      | 1087.76 | 5922.37 | 4606.77 | 3442.56 | 916.67  | 3645.34 | 554.14  |
| 1087.23                                                     | 2802.25 | 5381.33 | 2241.91 | 1430.1  | 2753.92 | 3109.59 | 3076.79 |
| 2346.27                                                     | 1412.12 | 6315.4  | 4553.72 | 3994.62 | 1388.57 | 870.66  | 1069.94 |
| 776.57                                                      | 1421.92 | 2706.17 | 516.9   | 1774.46 | 1130.42 | 3744.39 | 1224.37 |
| 1153.2                                                      | 1466.52 | 1458.92 | 6849.53 | 3881.9  | 5167.62 | 485.66  | 4649.84 |
| 411.95                                                      | 6845.86 | 4498.45 | 1729.65 | 4813.04 | 5652.44 | 427.37  | 1432.8  |
| 2241.58                                                     | 1212.29 | 1899.08 | 1566.66 | 5956.33 | 2631.88 | 1591.26 | 2082.39 |
| 2388.36                                                     | 4047.07 | 1713.95 | 986.82  | 3377.06 | 4397.48 | 3189.21 | 4493.92 |
| 599.44                                                      | 685.67  | 3191.28 | 3415.94 | 3357.55 | 5665.94 | 958.93  | 4995.22 |
| 6376.52                                                     | 2049.14 | 2188.4  | 2471.9  | 2018    | 6120.85 | 1744.52 | 1973.46 |
| 5739.82                                                     | 2259    | 3487.01 | 784.93  | 1418.11 | 4893.16 | 4742.1  | 1023.38 |
| 5072.99                                                     | 1075.68 | 1905.38 | 570.33  | 708.47  | 3041.74 | 3782.76 | 782.99  |
| 5600.65                                                     | 366     | 2044.68 | 2638.88 | 870.65  | 1885.06 | 3427.99 | 669.29  |
| 1769.56                                                     | 6450.15 | 1585.31 | 3135.85 | 4488.79 | 5592.52 | 744.07  | 2304.72 |
| 1260.01                                                     | 996.02  | 1324.62 | 4974.6  | 4644.22 | 1639.81 | 2669.38 | 582.05  |
| 2265.58                                                     | 4161.6  | 1648.08 | 1400.91 | 286.83  | 4881.82 | 2441.24 | 1046.16 |
| 361.09                                                      | 1438.59 | 887.79  | 3812.61 | 2922    | 2357.53 | 1258.3  | 2920.03 |
| 616.6                                                       | 1060.23 | 1342.2  | 2684.25 | 2831.12 | 1475.43 | 5235.89 | 4617.48 |
| 6705.16                                                     | 1465.08 | 2317.46 | 2396.75 | 2420.1  | 2573.35 | 2418.46 | 4174.65 |
| 694.02                                                      | 3130.74 | 5675.7  | 4487.36 | 1808.81 | 1775.56 | 2787.66 | 473.92  |
| 958.54                                                      | 7066.17 | 5804.25 | 1111.32 | 2451.28 | 2072.35 | 3522.84 | 465.03  |
| 910.05                                                      | 903.14  | 4968.78 | 1934.89 | 2415.2  | 1949.47 | 2754.87 | 4594.96 |
| 1868.95                                                     | 2188.67 | 3064.55 | 804.17  | 4016.56 | 1980.81 | 4954.04 | 2348.86 |
| 1936.13                                                     | 1732.72 | 5308.92 | 2684.59 | 1620.15 | 1939.93 | 3468.42 | 3121.86 |
| 683.57                                                      | 6321.57 | 2033.53 | 6784.44 | 2131.55 | 1114.86 | 3698.11 | 2526.51 |
| 753.09                                                      | 2257.14 | 2555.11 | 4724.8  | 4769.48 | 1113.2  | 451.83  | 715.98  |
| 5519.62                                                     | 1311.02 | 1241.94 | 942.71  | 2585.55 | 1251.19 | 3712.71 | 3326.9  |
| 1170.83                                                     | 699.74  | 4724.53 | 1190.98 | 3715.96 | 4983.23 | 1019.53 | 2773.18 |
| 640.86                                                      | 2209.14 | 5153.87 | 1681.07 | 5155.77 | 2136.28 | 2996.02 | 3443.54 |
| 443                                                         | 1412.69 | 2174.05 | 5093.52 | 1742.07 | 1142.75 | 3397.04 | 1897.63 |
| 1517.67                                                     | 1427.7  | 1776.54 | 3005.59 | 4089.28 | 2689.32 | 3792.9  | 2282.93 |
| 1649.46                                                     | 3015.63 | 1286.33 | 1924.15 | 1503.82 | 3106.33 | 2156.44 | 3779.87 |
| 770.6                                                       | 1131.32 | 5123.56 | 1966.81 | 4547.72 | 4980.14 | 5250.07 | 3645.53 |
| 1747.94                                                     | 2663.69 | 588.47  | 4261.23 | 4284.4  | 2064.56 | 3486.6  | 640.54  |
| 1310.17                                                     | 894.9   | 2556.22 | 4328.49 | 5777.5  | 4695.85 | 675.06  | 1790.42 |
| 1144.2                                                      | 931.39  | 1283.17 | 1636.44 | 793.36  | 1429.92 | 3374.67 | 3421.93 |
| 551.48                                                      | 3684.77 | 1298.48 | 520.67  | 4414.63 | 3324.04 | 3673.94 | 1779.54 |
| 840.39                                                      | 1069.7  | 5354.54 | 4207.5  | 3261.29 | 5644.57 | 3177.44 |         |
| 2840.84                                                     | 1644.55 | 3194.32 | 3278.23 | 6334.85 | 4505.3  | 4326.8  |         |