

ESTUDIO DE LOS MODELOS PROBABILÍSTICOS
DE ÍNDICES DE EXPOSICIÓN A PLAGUICIDAS.
APLICACIÓN EN TRABAJADORES RURALES
DE LA PROVINCIA DE CÓRDOBA.

CANDIDATA: OLGA ESTELA PADRÓ

DIRECTORA DE TESIS: DRA. MARÍA INÉS STIMOLO

UNIVERSIDAD NACIONAL DE CÓRDOBA
MAESTRÍA EN ESTADÍSTICA APLICADA

CÓRDOBA

2020



Estudio de los modelos probabilísticos de índices de exposición a plaguicidas.
Aplicación en trabajadores rurales de la provincia de Córdoba. by Olga Estela Padró is
licensed under a [Creative Commons Reconocimiento-NoComercial-SinObraDerivada
4.0 Internacional License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Dedicatorias

A mis amados hijos, los que iluminan mi vida cada día.

A mis padres, por tanto amor recibido.

A mis queridos hermanos, cuñados y sobrinos, por tanto compartido.

A Carlos, mi compañero incondicional.

Agradecimientos

A mi Directora de Tesis Dra. María Inés Stimolo, le agradezco su infinito apoyo y dedicación.

Un especial agradecimiento a la Dra. María del Pilar Díaz por sus generosos aportes y valiosas orientaciones, al Grupo de Epidemiología Ambiental del Cáncer en Córdoba, Facultad de Ciencias Médicas, UNC (GEACC) por brindarme amplio acceso a su trabajo, en particular a la Dra. María Josefina Lantieri por compartir su experticia en la materia.

También agradezco al Dr. Víctor Leiva por brindarme la información necesaria para conseguir las rutinas de R de las nuevas distribuciones.

Al Dr. Víctor Leiva por brindarme las rutinas en R de las distribuciones GBS.

A la Dra. Mariana Funes por compartir excelente material sobre indicadores compuestos.

A la Dra. Patricia Caro, quien generosamente me brindara bibliografía del método kernel.

A Juan Manuel Luzuriaga, por ayudarme en las búsquedas de material.

Al Lic. Pablo Ortiz, por asistirme con antiguas versiones del lenguaje R.

Al Lic. Maximiliano Iglesias, por su aporte sobre transformación de variables.

A la memoria de quien fuera mi profesora de Estadística, Dra. Hebe Goldenersh le agradezco que me alentara en mi vocación docente, siempre tendrá mi admiración y respeto.

A la Facultad de Ciencias Económicas de la U.N.C., en particular a la Dra. Margarita Díaz y demás profesores integrantes del Instituto de Estadística y Demografía, siempre dispuestos a compartir sus conocimientos y experiencia.

Resumen

Dado que la agricultura, es una de las principales actividades productivas de nuestra provincia, es necesario que se preste especial atención al uso de los plaguicidas y su efecto en la salud de los agroaplicadores. Una manera de medir las consecuencias de la aplicación de este tipo de sustancias, es a través del índice Nivel de Intensidad a la Exposición (IE), construido por el Grupo de Epidemiología Ambiental del Cáncer en Córdoba, Facultad de Ciencias Médicas, UNC (GEACC), para la población de trabajadores rurales de Córdoba.

A partir de una muestra de trabajadores rurales, se analizaron las características empíricas del índice para inferir sobre la distribución poblacional del mismo. En la búsqueda de una distribución adecuada para el índice de intensidad se tuvieron en cuenta los modelos de las variables de vida desde los más tradicionales (exponencial, Weibull, lognormal, gamma, Birnbaum-Saunders y Gaussiana inversa) hasta las nuevas distribuciones (Birnbaum-Saunders Generalizada e Inversa Gaussiana Tipo). En este trabajo, el mejor ajuste para el IE se logró con la distribución Weibull.

Una vez definido el mejor ajuste, se estimaron los percentiles y se delimitaron tres niveles de riesgo (bajo, medio y alto). En función a estos niveles de riesgo se estudiaron características sociodemográficas y de salud de los trabajadores dentro de cada grupo. Los factores que resultaron más significativos para determinar el riesgo de los aplicadores fueron: la antigüedad (mayor a 10 años) en las tareas de mezclar, aplicar o estar presente mientras se manipulan plaguicidas, no tener receta de ingeniero agrónomo, inyectar animales y la presencia de signos de irritación.

Palabras claves: Índice de nivel de intensidad a la exposición de plaguicidas – Bondad de ajuste – Variables de vida – Distribución Birnbaum-Saunders Generalizada – Distribución Inversa Gaussiana Tipo

Abstract

Since agriculture is one of the main productive activities in Córdoba province, Argentina, it is necessary that special attention be paid to the use of pesticides and their effect on the health of agro-applicators. One way to measure the consequences of the application of this type of substances is through the Index Level of Intensity at Exposure (IE), built by the Group of Environmental Epidemiology of Cancer in Córdoba, Faculty of Medical Sciences, UNC (GEACC), for the population of rural workers in Córdoba.

From a sample of rural workers, the empirical characteristics of the index were analyzed to infer about its population distribution. In the search for an adequate distribution for the intensity index, the models of the life variables were taken into account from the most traditional (exponential, Weibull, lognormal, gamma, Birnbaum-Saunders and inverse Gaussian) to the new distributions (Generalized Birnbaum-Saunders and Inverse Gaussian Type). In this work, the best fit for IE was achieved with the Weibull distribution.

Once the best fit was defined, the percentiles were estimated and three risk levels were defined (low, medium and high). Based on these risk levels, sociodemographic and health characteristics of the workers within each group were studied. The factors that were most significant to determine the risk of applicators were: seniority (greater than 10 years) in the tasks of mixing, applying or being present while handling pesticides, not having a prescription from an agronomist, injecting animals and the presence signs of irritation.

Keywords: Pesticide exposure intensity level index - Goodness of fit - Life variables -
Generalized Birnbaum-Saunders Distribution - Inverse Gaussian Generalized
Distribution

ÍNDICE

Dedicatorias	III
Agradecimientos	IV
Resumen	V
Abstract.....	VII
Capítulo 1	1
Introducción.....	1
1.1. Relevancia y Justificación.....	1
1.2. Hipótesis	9
1.3. Objetivos	9
Capítulo 2	11
Índice de Exposición a Plaguicida.....	11
2.1 Población de estudio	11
2.2 Construcción del Índice Nivel de Intensidad a la Exposición	12
Capítulo 3	15
Variables de Vida	15
3.1 Características de una variable de vida.....	15
3.2. Distribuciones de variables de vida	20
3.2.1 Distribuciones de vida más conocidas	22
3.2.2 Los nuevos modelos de Variables de Vida	43
3.3 Selección del Modelo de Ajuste	51

3.3.1 Estadísticos Cramér-von Mises y Anderson-Darling.....	53
3.3.2 Validación de la distribución elegida con los gráficos TTT	54
Capítulo 4	57
Aplicación a Trabajadores Rurales de la Provincia de Córdoba	57
4.1 Metodología para el ajuste	57
4.1.1. Análisis Gráfico.....	57
4.1.2 Criterios de Información	61
4.1.3 Validación de la distribución elegida con los gráficos TTT	63
4.2. Construcción de una Escala Cualitativa.....	69
4.3 Caracterización de los grupos de riesgo.....	72
Capítulo 5	79
Discusión y Conclusiones.....	79
Referencias	84
Anexo 1.....	90
Anexo 2.....	92

Capítulo 1

Introducción

En la primera parte de este capítulo se destaca la relevancia de realizar un ajuste a un índice de nivel de intensidad a la exposición de plaguicidas, debido al efecto de la aplicación de estos productos en la salud de los agroaplicadores. Luego se plantean las hipótesis y los objetivos de este trabajo.

1.1. Relevancia y Justificación

El estudio de los plaguicidas y su impacto en la salud de los trabajadores rurales es un tema que adquirió suma importancia en los últimos años debido al aumento de la superficie cultivada, con un sistema de producción que tiene incorporado el uso de los mismos. Particularmente en la provincia de Córdoba, considerando los cuatro cultivos principales (girasol, maíz, trigo y soja), entre 1996 y 2018 la superficie sembrada se duplicó, el incremento fue del 109,67% (Cuadro 1 y Figura 1).

El mayor crecimiento se debe a una expansión horizontal (por la cantidad de superficie sembrada) antes que por una mejora “vertical” (en la eficiencia de producción), lo que implica que se han incorporado nuevas áreas al cultivo, seguramente marginales y de mayor fragilidad, y/o se han desplazado áreas destinadas a otros cultivos o producciones (Ybran & Lacelli, 2016). Los cultivos de soja, el maíz y el trigo representan más del 90% de la superficie sembrada en todo el periodo. La tasa promedio de crecimiento anual de la superficie sembrada para la totalidad de estos cuatro cultivos fue del 3,59%. El cultivo con mayor proporción de superficie sembrada fue el de soja (con un pico del 81,67% en

la campaña 2009/2010), y en segundo lugar estuvo el maíz o el trigo hasta la campaña 2004/2005 a partir de la cual la superficie sembrada de maíz superó siempre a la de trigo. La superficie cultivada con maíz, creció durante todo el periodo, y presentó la mayor tasa de crecimiento con un promedio de 4,89% anual. En el caso de la soja creció en forma sostenida hasta la campaña 2009/2010 con algunas oscilaciones en las campañas siguientes, por lo que la tasa promedio de crecimiento de superficie sembrada alcanzó el 4,15% anual. Mientras que, la superficie sembrada de trigo solo presentó un crecimiento promedio del 2,68% anual. El cultivo de girasol que representaba el 10,39% del total de la superficie sembrada en la campaña 1996/97, alcanzó tan solo el 0.15% del total en la campaña 2017/2018, cediendo superficie al cultivo de maíz, soja y trigo.

Cuadro 1: Superficie sembrada por tipo de cultivo en la Provincia de Córdoba

Campaña	Superficie Sembrada Girasol	Superficie Sembrada Maíz	Superficie Sembrada Soja	Superficie Sembrada Trigo	Total sup. Sembrada
1996/97	422.100	1.029.500	1.901.300	709.200	4.062.100
1997/98	469.100	919.100	2.096.800	450.800	3.935.800
1998/99	557.800	734.050	2.564.600	625.450	4.481.900
1999/00	546.600	855.400	2.729.000	749.000	4.880.000
2000/01	322.950	862.750	3.151.500	867.400	5.204.600
2001/02	248.975	906.950	3.452.900	1.086.700	5.695.525
2002/03	244.870	923.570	3.564.352	1.004.250	5.737.042
2003/04	68.250	738.830	4.172.940	1.063.480	6.043.500
2004/05	101.830	1.027.970	3.981.145	1.155.525	6.266.470
2005/06	120.764	1.038.180	4.343.718	772.750	6.275.412
2006/07	102.500	1.151.650	4.477.882	909.700	6.641.732
2007/08	83.446	1.440.000	4.699.230	1.235.700	7.458.376
2008/09	63.200	1.146.200	5.196.748	754.180	7.160.328
2009/10	48.600	998.100	5.594.340	208.590	6.849.630
2010/11	43.200	1.142.900	5.054.390	545.550	6.786.040
2011/12	41.300	1.272.000	5.014.250	514.720	6.842.270
2012/13	40.700	1.907.350	5.349.312	487.881	7.785.243
2013/14	25.750	1.917.500	5.052.760	724.830	7.720.840
2014/15	30.800	1.879.800	5.413.330	1.141.000	8.464.930

Cuadro 1: Superficie sembrada por tipo de cultivo en la Provincia de Córdoba (continuación)

Campaña	Superficie Sembrada Girasol	Superficie Sembrada Maíz	Superficie Sembrada Soja	Superficie Sembrada Trigo	Total sup. Sembrada
2015/16	20.070	1.935.970	5.579.530	1.071.240	8.606.810
2016/17	18.470	2.614.910	4.871.202	1.580.050	9.084.632
2017/18	12.700	2.804.452	4.463.610	1.236.318	8.517.080
Tasa promedio de crecimiento en el periodo 1996/2018	-15,37%	4,89%	4,15%	2,68%	3,59%

Fuente: Dirección de Estimaciones agrícolas (Dirección de Estimaciones Agrícolas - Dirección Nacional de Agricultura , 2018).

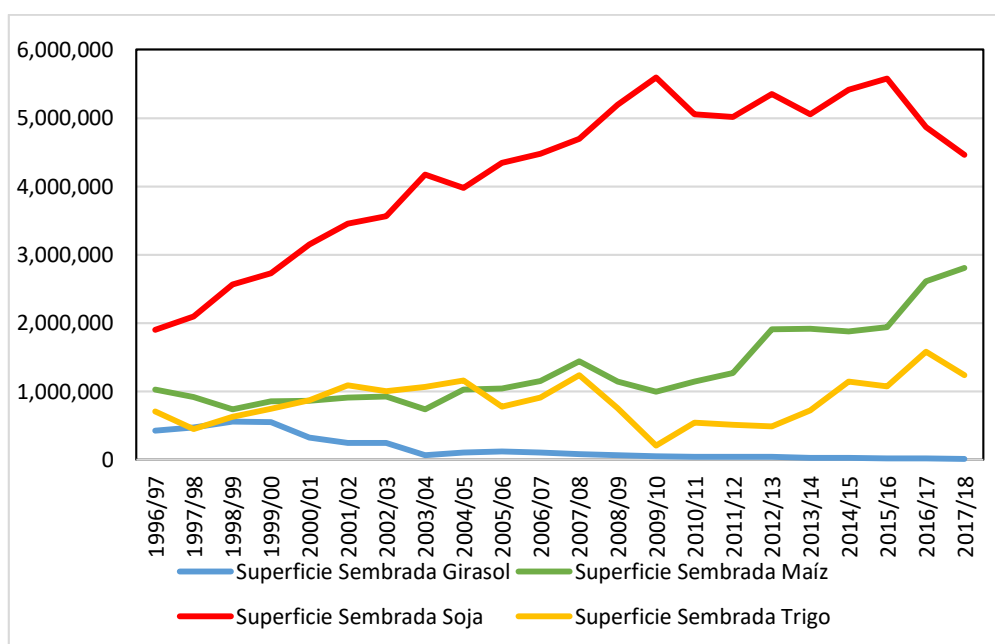


Figura 1. Evolución de la Superficie sembrada en la Provincia de Córdoba

Fuente: Dirección de Estimaciones agrícolas (Dirección de Estimaciones Agrícolas - Dirección Nacional de Agricultura , 2018).

Conjuntamente con el aumento de superficie sembrada, con un evidente predominio del cultivo de soja (Cuadro 1 y Figura 1) los nuevos sistemas de producción incorporan el uso de agroquímicos, por lo que los trabajadores rurales se encuentran aún más expuestos

a los efectos negativos de sus componentes tóxicos. Surge la necesidad de encontrar una herramienta que permita valorar o medir, de alguna manera, el riesgo a la exposición que experimenta un trabajador rural que está en contacto con este tipo de tóxicos de manera cotidiana. El origen del riesgo debido a los agentes químicos, tiene un efecto adverso en la salud. Por lo tanto, el riesgo es potencial y no una real posibilidad, es decir, se puede evaluar estadísticamente (Vilca, Sanhuenza, Leiva, & Christakos, 2010). Medir el riesgo por la exposición a plaguicidas en un individuo o en una población determinada es complejo, pues está influenciado por factores como la diversidad de productos comercializados, su uso indiscriminado, la multiplicidad de las fuentes de exposición y la variación en la intensidad y la duración de la exposición en un lapso de tiempo. Estos aspectos escapan al control del investigador y, generalmente, se carece de información completa y oportuna sobre ellos. Diversos estudios epidemiológicos han usado de forma combinada algunas herramientas de medición, como la historia de exposición, la evaluación de expertos, la monitorización ambiental y biológica, con el propósito de lograr una mayor precisión (Ramírez & Lacasaña, 2001).

Un indicador compuesto es una representación simplificada que busca resumir un concepto multidimensional en un índice simple (unidimensional) con base en un modelo conceptual subyacente. En términos técnicos, un indicador compuesto se define como función de dos o más variables, que conjuntamente miden características multidimensionales de los individuos en estudio (Schuschny & Soto, 2009). En el contexto del análisis de políticas los indicadores son útiles para identificar tendencias (OECD, 2008).

A partir de diferentes criterios se construyeron indicadores de riesgo a la exposición a plaguicidas como los de Dosemeci *et al* (2002) que, a partir del estudio prospectivo de cohorte de 58.000 aplicadores de Carolina del Norte y Iowa, desarrollaron dos algoritmos

para identificar distintos escenarios de exposición de los aplicadores. El algoritmo general usa características básicas como si aplica, si mezcla y repara los equipos, ponderadas por el equipo de protección personal empleado en las mencionadas actividades. El algoritmo detallado agrega a las variables anteriores especificaciones como, por ejemplo, si emplean un sistema cerrado para la mezcla (Dosemeci *et al*, 2002).

Otro antecedente es el índice construido por Palacios & Paz (2011), que con el fin de evaluar la sintomatología persistente en trabajadores agrícolas expuestos a plaguicidas órgano-fosforados, realizaron un estudio transversal analítico con 106 jornaleros en un campo agrícola en México, determinando el nivel de colinesterasa y los síntomas según la exposición (Palacios & Paz, 2011).

El Grupo de Epidemiología Ambiental del Cáncer en Córdoba, Facultad de Ciencias Médicas, UNC (GEACC¹), construyó un índice Nivel de Intensidad a la Exposición (IE) para la población de trabajadores rurales de Córdoba a partir de una adaptación del provisto por Docemesi *et al* en 2002 (Lantieri, Butinof, R.A., Stimolo, Blanco, & Diaz, 2011), cuyo detalle se presenta en el capítulo 2.

Estudios preliminares del índice IE, evaluado en un conjunto de trabajadores rurales, reportó empíricamente una distribución asimétrica a la derecha, con alta frecuencia de valores atípicos, de soporte positivo (Figura 2). En el análisis del IE se debe tener en cuenta que valores más altos del índice indican mayor riesgo.

¹ <http://geacc.webs.fcm.unc.edu.ar/>

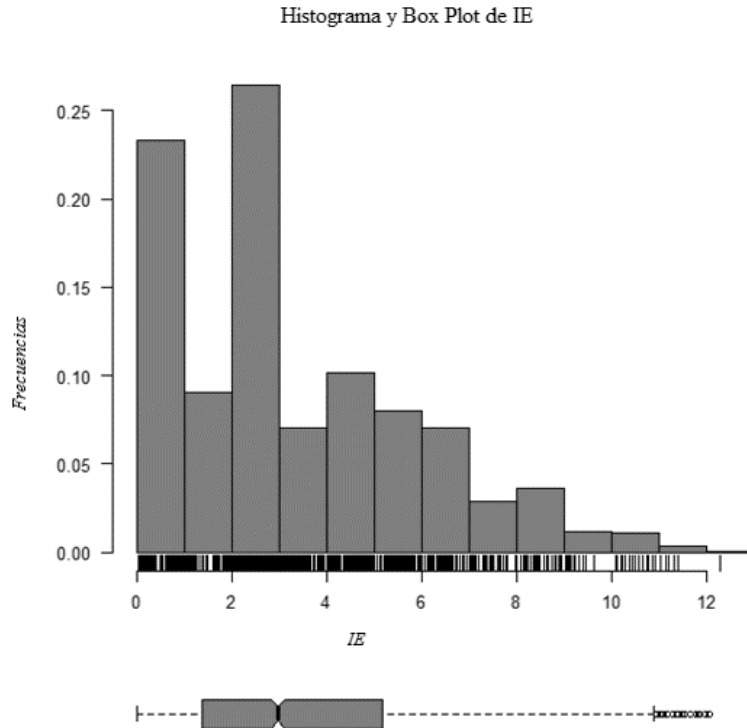


Figura 2. Distribución empírica del índice Nivel de Intensidad a la Exposición (IE) para la muestra de trabajadores rurales de Córdoba
Fuente: Realizado con lenguaje R paquete BS 1.0.

Estas características empíricas sugieren que este índice se distribuye como una variable de vida. Las variables de vida son continuas, asumen valores positivos y si bien en general representan, tiempos de vida de un organismo biológico o tiempos de duración de un componente electrónico o tiempos de espera hasta que un sistema falle (Marshall & Olkin, 2007); también pueden tener otras aplicaciones como medir la concentración de un contaminante (Leiva, Barros, Paula, & Saunders, 2010), siendo esta última aplicación de especial interés para el estudio del IE. Una distribución de vida (*life distribution*) es un modelo de probabilidad asociado con una variable de vida (Marshall & Olkin, 2007).

Una distribución de vida es, en general, asimétrica, unimodal, sesgada positivamente, de dos parámetros al menos y poseedora de momentos de cualquier orden.

Tradicionalmente se utilizaron para este tipo de variables aleatorias la distribución exponencial, Weibull, lognormal, gamma, Birnbaum-Saunders y la Gaussiana inversa. Las últimas cuatro distribuciones se caracterizan en modelar vida para fatiga (*fatigue life distributions*). La secuencia de exposición en un ciclo de tiempo a un contaminante genera una carga que desde el punto de vista estadístico se visualiza como una función unimodal continua definida en el intervalo de la unidad, en la cual, la fatiga corresponde al punto de deflexión que indica un daño estructural (grieta) que se acumula dentro de la muestra a lo largo del tiempo (Leiva, Barros, & Paula, 2008).

Los modelos de tiempos de vida arriba citados ajustan razonablemente bien a los datos en la parte central de la distribución, pero poseen desempeños pobres en las colas de las mismas. Ajustar las colas de la distribución se convierte en un aspecto muy importante en el análisis debido, por ejemplo, a los problemas de garantía (dosis letales) y mantenimiento (emergencias ambientales), que a menudo se encuentran en los percentiles altos o bajos de la distribución, lo cual es ilustrado en la Figura 3. En el capítulo 3 se aborda más detalladamente las características de distribuciones de variables de vida.

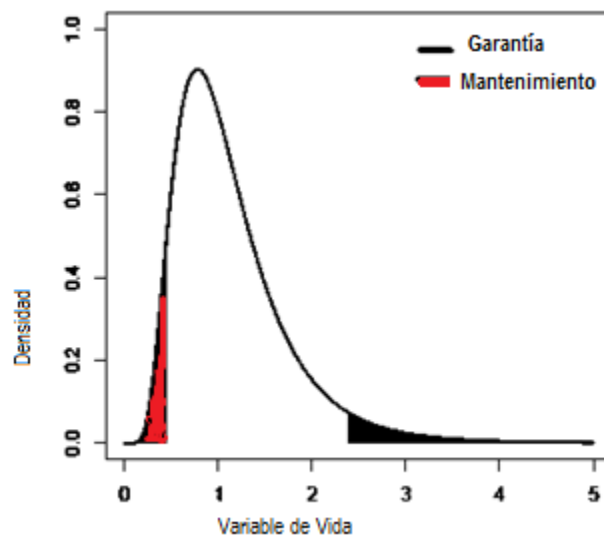


Figura 3. Modelos de daño acumulativo

Fuente: Modelos de daño acumulativo Birnbaum-Saunders. Simposio de Estadística Santa Marta, (Leiva, Barros, Paula, & Saunders, 2010).

Por lo tanto, se torna necesario considerar modelos de vida con mayor flexibilidad (Leiva, Barros, Paula, & Sanhuenza, 2008), como los modelos BSG (Birnbaum-Saunders Generalizada) y los modelos IGT (Inversa Gaussiana Tipo).

En este trabajo se estudiará el desempeño del índice Nivel de Intensidad a la Exposición (IE), basándonos en una muestra de 2.210 aplicadores terrestres de la provincia de Córdoba. Así, se analiza la distribución empírica de los mismos y se propone un modelo de probabilidad que describa adecuadamente estas características. Se comenzará estudiando las distribuciones tradicionales para variables de vida; exponencial, gamma, lognormal, Weibull, Inversa Gaussiana y Birnbaum-Saunders. Luego, se considerarán modelos que debieran ser plausibles, en función de sus fundamentos teóricos y sus propiedades, como los empleados con éxito para datos relacionados con acumulación de algún tipo de daño, usados por ejemplo para la valoración de la concentración de contaminantes ambiental, es decir los modelos: BSG (Birnbaum-Saunders Generalizada) y los modelos IGT (Inversa Gaussiana Tipo). Una descripción exhaustiva de estos modelos probabilísticos se presenta en el capítulo 4. La selección de la distribución candidata será mediante el análisis del comportamiento de la tasa de riesgo empleando el gráfico TTT (*Total Time On Test*). Una vez elegido el modelo de probabilidad que presente el ajuste más satisfactorio, se obtendrán los percentiles que determinan los niveles de baja, media y alta exposición a plaguicidas. Un análisis descriptivo de los percentiles calculados permitirá caracterizar grupos en función a otras variables indagadas en los sujetos de la muestra. Los resultados se presentan en el capítulo 4 y la discusión, junto con las principales conclusiones en el Capítulo 5.

En relación a lo expuesto anteriormente, se plantean las siguientes hipótesis y objetivos:

1.2. Hipótesis

H1: Los modelos BSG y los modelos IGT permiten un ajuste adecuado para la descripción probabilística de los eventos relacionados con la medición de riesgo por exposición a plaguicidas.

H2: Existen diferencias significativas en las características socio-demográficas y el método de aplicación de los plaguicidas, por parte de los trabajadores de los grupos establecidos como de baja, media y alta exposición a plaguicidas.

1.3. Objetivos

Para poder verificar las hipótesis planteadas, se proponen los siguientes objetivos generales y específicos:

Objetivo general 1: Identificar la distribución del índice denominado: Nivel de intensidad a la Exposición (IE).

- Objetivo Específico 1.1: Comparar las performances de las distribuciones estudiadas usando distintas técnicas de bondad de ajuste.
- Objetivo Específico 1.2: Construir las curvas TTT correspondientes al índice IE, para determinar el comportamiento de la tasa de riesgo del índice y estudiar su comportamiento.

Objetivo general 2: Definir y caracterizar niveles alto, medio y bajo de riesgo de exposición a plaguicidas.

- Objetivo Específico 2.1: Obtener los percentiles que definen los niveles alto,

medio y bajo; de riesgo a exposición a plaguicidas, considerando la distribución elegida.

- Objetivo Específico 2.2: Determinar si existen diferencias significativas en las características socio-demográficas y el método de aplicación de los plaguicidas por parte de los trabajadores de los grupos establecidos como de baja, media y alta exposición a plaguicidas.

Capítulo 2

Índice de Exposición a Plaguicida

Como se mencionó en la Introducción se tomará en cuenta el índice Nivel de Intensidad a la Exposición (IE) elaborado por el Grupo de Epidemiología Ambiental del Cáncer en Córdoba, Facultad de Ciencias Médicas e INICSA (CONICET), UNC (GEACC²), basado en la versión construida por Docemesi *et al* en 2002 (Lantieri, Butinof, Stimolo, Blanco, & Diaz, 2011).

2.1 Población de estudio

Este proyecto utiliza información primaria, obtenida a través de muestreos polietápicos aleatorios de la población de agroaplicadores terrestres que, desde el año 2008, desarrollan sus actividades laborales en el territorio de la provincia de Córdoba. La información fue recogida por investigadores del GEACC (FCM, UNC) a partir de una encuesta autoadministrada entregada a cada uno de 2.209 trabajadores rurales de la provincia de Córdoba. Estas encuestas se administraron a los participantes de los Cursos (de carácter obligatorio anualmente) para la obtención de la licencia de Aplicadores de Plaguicidas, dictados desde el Ministerio de Agricultura, Alimentos y Ganadería de la Provincia de Córdoba. El instrumento utilizado es una adaptación, también llevada a cabo por el GEACC, del cuestionario AHS (Bonner & Alavanja, 2005) y consta de cinco

² <http://geacc.webs.fcm.unc.edu.ar/>

módulos, los cuales recogen datos demográficos, características sobre el uso de plaguicidas, información sobre cultivos, variables relacionadas con los hábitos de vida y salud del trabajador y su familia. La encuesta fue diseñada de manera tal que valora la posible exposición actual y pasada a plaguicidas y su intensidad.

2.2 Construcción del Índice Nivel de Intensidad a la Exposición

Los índices de exposición individual fueron obtenidos a partir de las siguientes variables consideradas: si el trabajador mezcla plaguicidas, si aplica, los métodos de aplicación que utiliza, si realiza tareas como rociar o sumergir animales, tratamiento de semillas o reparación de maquinarias, la distancia (en metros) de la vivienda al cultivo más cercano, y un índice de protección personal de los trabajadores elaborado por el GEACC, según el equipo de protección utilizado (Cuadro 2). Para mayores detalles consultar Lantieri *et al* (2011).

Cuadro 2: Niveles de protección según las medidas de protección utilizadas, GEACC (FCM, UNC)

Tipo de protección	Porcentaje de protección	Medida aplicada
EPP 0	0 %	Ninguna
EPP 1	20 %	Gafas de protección facial (protectores de cara o anteojos) Guantes de tela o cuero Otra ropa protectora como botas, casco o sombrero
EPP 2	30 %	Máscara o cartucho (careta antigas) Ropa desechable (ropa impermeable)
EPP 3	40 %	Guantes de goma químicamente resistentes

Fuente: Tabla creada por el GEACC.

Para el cálculo del índice se asignaron puntajes a los indicadores seleccionados (Cuadro

3), de este modo a cada indicador (tipo de actividad) le corresponde un determinado puntaje. Con el fin de dar más valor a los métodos de aplicación más utilizados en la región, a las actividades como la aplicación del tóxico con pulverizador de mochila, el tratamiento de semillas, sumergir o rociar los animales; se los ponderó por la proporción de aplicadores que realizan estas actividades sobre el total de encuestados, en el Cuadro 4 se detallan las puntuaciones correspondientes a estos últimos casos.

Cuadro 3: Puntuaciones asignadas por el GEACC a cada indicador de exposición individual

Actividad	Puntuación
Sólo está presente	0
Mezcla	9
Aplica con Máquina autopropulsada con filtro de carbón activado	0,1
Aplica con Máquina autopropulsada sin filtro de carbón activado	0,5
Aplica con Máquina de arrastre con tractor con cabina con filtro de carbón activado	0,1
Aplica con Máquina de arrastre con tractor sin cabina con filtro de carbón activado	0,5
Aplica con Máquina de arrastre con tractor sin cabina	1
Aplica con pulverizador de mochila	8
Tratamiento de semillas	1
Sumerge los animales	5
Rocía los animales	6
Repara maquinarias	2
EPP 0	1
EPP 1	0,8
EPP 2	0,7
EPP 3	0,6
EPP 1 y 2	0,5
EPP 1 y 3	0,4
EPP 2 y 3	0,3
EPP 1, 2 y 3	0,1
Distancia a la vivienda menor a 500 mts.	2
Distancia a la vivienda entre 501 a 1.500 mts.	1
Distancia a la vivienda de más de 1.500 mts.	0

Fuente: Tabla creada por el GEACC.

Cuadro 4: Puntuaciones asignadas por el GEACC a cada indicador de exposición individual, ponderadas por porcentaje de uso de cada práctica

Actividad	Puntuación Inicial	Puntuación Definitiva (después de la ponderación) ^a
Pulverizador de mochila	8	2,288
Tratamiento de semillas	1	0,342
Sumerge los animales	5	0,120
Rocía los animales	6	0,648

Fuente: Tabla creada por el GEACC.

^aPonderación realizada con la base depurada en marzo 2014.

A partir de los indicadores y sus ponderaciones, se definió un algoritmo para calcular el índice, Intensidad de Exposición (IE), según la expresión:

$$IE = (Mez * EPP) + \left(\sum \frac{Met * EPP}{Cant} \right) + (Repara * EPP) + DV$$

donde *Mez* representa el puntaje de mezcla, *EPP* (elementos de protección personal) el score de nivel de protección, *Met* el score de cada método de aplicación utilizado por el trabajador, *Cant* la cantidad de métodos de aplicación utilizados por cada trabajador, *DV* es el score para la distancia de la vivienda al cultivo más cercano.

En el IE se utiliza un modelo aditivo para mezcla, método de aplicación, repara y distancia de la vivienda al cultivo más cercano ya que constituyen factores condicionantes de la exposición individual independientes, mientras que se usa un modelo multiplicativo para el nivel de protección debido a que depende de los anteriores condicionantes independientes, excepto la distancia de la vivienda al cultivo más cercano (Dosemeci *et al*, 2002). Este índice, que asume valores positivos, es una combinación lineal de variables cuantitativas y cualitativas, lo que implica que el ajuste estadístico de la distribución que lo representa no resulte sencillo.

Capítulo 3

En este capítulo se aborda el tema variable de vida, en la primera parte se verá su definición y características. Luego se presentan las distribuciones de las variables de vida, comenzando con las más conocidas y finalizando con los nuevos modelos. En la última sección de este capítulo se exponen los lineamientos que se seguirán en la selección del modelo de ajuste realizada en el capítulo 4.

En todo el capítulo se sigue la simbología utilizada por el Dr. Víctor Leiva (Leiva, Barros, Paula, & Sanhuenza, 2008; Sanhuenza, Leiva, & Balakrishnan, 2008; Vilca, Sanhuenza, Leiva, & Christakos, 2010).

Los principales referentes que se tuvieron en cuenta para las distribuciones de vida más conocidas son los siguientes libros (Marshall & Olkin, 2007; Johnson, Kotz, & Balakrishnan, 1995; Saunders, 2007).

Variables de Vida

3.1 Características de una variable de vida.

Esta denominación hace referencia a variables continuas que asumen valores positivos. En general este tipo de variables centran su análisis en el tiempo de vida de un organismo biológico o la duración de un componente creado por el hombre. También se pueden presentar otros tipos de aplicaciones: el estudio de tiempos de duración o de vida como demoras en el tráfico, intervalos entre terremotos. En otras situaciones pueden surgir

variables referidas a magnitudes físicas, longitudes de grietas, resistencia de materiales, diámetro o altura de los árboles (Marshall & Olkin, 2007).

Es importante destacar que las variables de vida también pueden medir la concentración de un contaminante (Leiva, Barros, Paula, & Saunders, 2010) esta última aplicación es la que tomaremos en cuenta en la presente tesis.

Este tipo de variables se analizan en los estudios de confiabilidad. Un análisis de confiabilidad se refiere a la evaluación de la tasa de uso, deterioro o daño acumulativo de una estructura o sistema y la distribución de la vida útil de servicio subyacente, hasta que el sistema o estructura no pueda funcionar adecuadamente bajo condiciones de seguridad (Saunders, 2007). Se conoce que el sistema fallará, lo que no se sabe es cuándo, por eso la espera hasta que se produzca la falla es una variable aleatoria positiva (Saunders, 2007). Esta variable que representa el tiempo de vida o duración hasta que se produzca la interrupción del sistema, la llamaremos T y diremos que T sigue una distribución de vida con soporte $[0, \infty)$, siendo $f_T(\cdot)$ su función de densidad y $F_T(\cdot)$, su función de distribución. Luego,

$$F_T(t) = P(T \leq t) = \int_0^t f_T(x) dx; \quad \text{para } t \geq 0.$$

Función de Confiabilidad o de Sobrevida

La confiabilidad es la probabilidad de que un sistema trabaje satisfactoriamente durante un periodo determinado (Saunders, 2007). La función de sobrevida de T , también llamada de confiabilidad está dada por

$$R_T(t) = P(T > t) = 1 - F_T(t) = \int_t^{\infty} f_T(x)dx; \quad \text{para } t \geq 0.$$

Como se puede observar la función de sobrevivida es el complemento de la función de distribución y permite obtener la probabilidad de que un determinado organismo sobreviva o que un componente dure, más allá de un tiempo t .

Para cualquier variable aleatoria, la función de distribución F y la función de Confiabilidad R , siempre existe (Marshal & Olkin, 2007).

Tasa de Riesgo

La tasa de riesgo, recibe varias denominaciones como tasa de fallas, o en análisis actuarial se conoce como tasa de mortalidad ya que indica la probabilidad por año de que una persona de una edad determinada muera en el instante siguiente (Leiva, Soto, Cabrera, & Cabrera, 2011).

Una de las principales propiedades de la tasa de riesgo es que hace posible caracterizar el comportamiento de las distribuciones de vida. Distribuciones de vida con densidades similares tienen en algunos casos tasas de riesgo distintas. Esto se debe a que, en estos casos, el envejecimiento ocurre de diferentes maneras. Por lo tanto, la tasa de riesgo debe usarse para seleccionar o descartar familias de distribuciones de vida (Leiva, Soto, Cabrera, & Cabrera, 2011).

La tasa de riesgo se define como la tasa de falla instantánea o como la propensión a fallar en el tiempo $t+\Delta t$, dado que el elemento no ha fallado hasta el tiempo t , entonces la tasa de riesgo de una v.a. T es,

$$h_T(t) = \frac{f_T(t)}{R_T(t)} = \lim_{\Delta \rightarrow 0} \frac{P(t < T < t + \Delta t | T > t)}{\Delta}; \text{ para } t \geq 0.$$

Entonces,

$$\Delta h_T(t) \approx P(t < T < t + \Delta t | T > t).$$

En consecuencia $\Delta h_T(t)$ se puede considerar como la probabilidad condicional de que se produzca la muerte o falla en el siguiente incremento Δ de tiempo, dada la supervivencia hasta un tiempo t . Esta interpretación hace el concepto tasa de riesgo muy útil tanto en la teoría como en las aplicaciones (Marshall & Olkin, 2007).

Si se conoce la tasa de riesgo, a partir de la misma se pueden obtener las funciones de densidad, de distribución y de confiabilidad (ver Cuadro 5).

Cuadro 5: Relación entre las funciones de densidad, distribución, confiabilidad y tasa de riesgo

Expresado por	$f_T(t)$	$F_T(t)$	$R_T(t)$	$h_T(t)$
$f_T(t)$	-	$\frac{d}{dt}F_T(t)$	$-\frac{d}{dt}R_T(t)$	$h_T(t)\exp\left[-\int_0^t h_T(u)du\right]$
$F_T(t)$	$\int_0^t f_T(u)du$	-	$1-R_T(t)$	$1-\exp\left[-\int_0^t h_T(u)du\right]$
$R_T(t)$	$\int_t^\infty f_T(u)du$	$1-F_T(t)$	-	$\exp\left[-\int_0^t h_T(u)du\right]$
$h_T(t)$	$\frac{f_T(t)}{1-\int_0^t f_T(u)du}$	$\frac{dF_T(t)/dt}{1-F_T(t)}$	$-\frac{d}{dt}\ln(R_T(t))$	-

Fuente: (Leiva, Barros, & Paula, www.victorleiva.cl, 2008; Marshall & Olkin, 2007; Jhonson, Kotz, & Balakrishnan, 1995)

Función de Riesgo

Así como la función de densidad de una variable aleatoria se puede integrar para obtener su función de distribución, la tasa de riesgo se puede integrar para obtener la tasa de riesgo acumulativa o función de riesgo (Saunders, 2007).

La función de riesgo de una variable aleatoria T es

$$H_T(t) = \int_0^t h_T(u) du = -\log(R_T(t)); \quad t > 0.$$

Vida Media Residual

El tiempo medio hasta la falla, también llamado vida promedio, vida esperada, esperanza de vida o vida media, es otro descriptor importante en el análisis de vida (Marshall & Olkin, 2007). La media de la variable T se obtiene:

$$\mu = \mathbb{E}[T] = \int_0^{\infty} t f_T(t) dt = \int_0^{\infty} R_T(t) dt.$$

Además, es posible definir la vida media residual como el promedio de tiempo que le queda a un elemento que vivió hasta el momento x , que se obtiene:

$$\mu_x = \mathbb{E}[T - x | T > x] = \int_0^{\infty} R_T(t|x) dt = \frac{\int_x^{\infty} R_T(y) dy}{R_T(x)}; \quad x > 0.$$

Si se tratara del tiempo de vida de una persona de edad x , la vida media residual nos indicaría la esperanza de vida que le queda.

3.2. Distribuciones de variables de vida

El estudio del índice de intensidad a la exposición a plaguicidas que este trabajo presenta como aplicación, se abordó desde el enfoque de variables de vida, ya que su comportamiento es similar a aquellas que contemplan la concentración de contaminantes en el ambiente o la acumulación de daño, tienen asimetría derecha, presentan varios datos atípicos y comienzan desde cero, ya que los niveles de estas concentraciones son siempre positivos. Dadas las características mencionadas en algunos casos se transforman los datos ambientales para eliminar la asimetría utilizando las transformaciones de Box-Cox y luego se ajusta un modelo normal (Vilca, Sanhuenza, Leiva, & Christakos, 2010). A pesar del uso común de algunas transformaciones, se ha demostrado que una transformación incorrecta reduce la potencia del estudio. A veces, aunque una transformación se realice adecuadamente, el problema de la interpretación no se resuelve. Una forma alternativa de evitar la transformación de los datos ambientales es a través del ajuste a una distribución adecuada (Vilca, Sanhuenza, Leiva, & Christakos, 2010). Sin embargo, algunos de estos modelos no tienen argumentos teóricos para problemas de contaminación ambiental y sólo se basan en un ajuste empírico (Vilca, Sanhuenza, Leiva, & Christakos, 2010).

La distribución lognormal (LN) es un modelo con asimetría derecha y valores positivos que ha sido frecuentemente utilizado para describir datos ambientales. Esto se debe a sus argumentos teóricos relacionados con procesos acumulativos y su relación con la distribución normal. Sin embargo, las concentraciones de contaminantes varían dependiendo de factores tales como fuente, meteorología local y topografía local. Así, la distribución real de estas concentraciones no se comporta siempre de acuerdo con el

modelo LN, especialmente en el nivel superior, ya que este modelo tiene colas cortas. Durante las últimas dos décadas, otro modelo con asimetría positiva que ha recibido gran atención es la distribución Birnbaum-Saunders, sin embargo, en este tipo de modelos, las estimaciones de los parámetros suelen ser sensibles a las observaciones atípicas, teniendo así que prestar más atención a los nuevos modelos como la distribución Birnbaum-Saunders Generalizada (BSG). Estos últimos tienen argumentos teóricos apropiados y una gran flexibilidad para modelar la curtosis incluyendo modelos con estimaciones de parámetros robustas frente a la presencia de valores atípicos, especialmente en los niveles superiores de la distribución. Luego, la distribución de GBS surge como un buen modelo para describir, por ejemplo, la concentración de contaminantes del aire (Leiva, Barros, Paula, & Sanhuenza, 2008).

Algo similar a lo que ocurre con la distribución Birnbaum-Saunders sucede con la distribución Inversa Gaussiana, modelo que por sus propiedades y su relación con la distribución normal ha sido ampliamente utilizado, sin embargo es bien sabido que la normal produce estimadores a menudo sensibles a los datos atípicos. Esta sensibilidad se puede transferir a cualquier distribución relacionada con la normal y sus estimaciones de parámetros también serán sensibles a datos atípicos. Por tal motivo Sanhuenza, Leiva y Balakrishnan proponen el nuevo modelo IGT (Inversa Gaussiana Tipo). Este modelo se basa en las distribuciones univariadas de contorno elíptico. El modelo IGT se caracteriza por tener curtosis más flexible que la distribución inversa Gaussiana clásica, permitiendo distintos grados de curtosis y de asimetría. Este nuevo modelo brinda estimaciones de parámetros robustas frente a la existencia de datos atípicos (Sanhuenza, Leiva, & Balakrishnan, 2008).

Podemos resumir lo anterior concluyendo que la importancia de los nuevos modelos, IGT y BSG, radica en que, a diferencia de las distribuciones tradicionales, permiten un mejor

ajuste en las colas de la distribución, esto es importante cuando interesan los percentiles superiores, como en el caso del índice IE.

3.2.1 Distribuciones de vida más conocidas

Las distribuciones de vida más conocidas son la distribución Exponencial, Weibull y las reconocidas por Saunders (Saunders, 2007) como distribuciones de la fatiga y daño acumulativo, estos modelos son: Gamma, Inversa Gaussiana, Birnbaum-Saunders y Lognormal. A continuación, se darán las características de cada una de ellas.

3.2.1.1 Exponencial

El modelo exponencial es una distribución muy utilizada en ingeniería de confiabilidad. Debido a su simplicidad, ha sido ampliamente empleado, incluso en casos para los que no se aplica.

Si una variable aleatoria tiene distribución exponencial con parámetro β , esto es $T \sim \text{Exp}(\beta)$. La función de densidad de T es:

$$f_T(t) = \frac{1}{\beta} \exp\left(-\frac{t}{\beta}\right); \quad t > 0, \beta > 0.$$

Esta distribución presenta asimetría derecha, la misma se atenúa a medida que aumenta la esperanza (Figura 4-a).

La función de distribución de T se expresa:

$$F_T(t) = 1 - \exp\left(-\frac{t}{\beta}\right); \quad t > 0.$$

La media y la varianza de T son:

$$\mathbb{E}[T] = \beta,$$

$$\text{Var}[T] = \beta^2.$$

La función de sobrevivencia y la tasa de riesgo se definen como:

$$R_T(t) = \exp\left(-\frac{t}{\beta}\right); \quad t > 0,$$

$$h_T(t) = \frac{1}{\beta}; \quad t > 0,$$

A medida que la esperanza aumenta, la tasa de riesgo disminuye (Figura 4-d).

La distribución exponencial se utiliza para describir unidades que tienen una tasa de falla constante, por ejemplo cuando se estudia la duración de chips de computadoras, los que no envejecen con el tiempo (Leiva, Soto, Cabrera, & Cabrera, 2011).

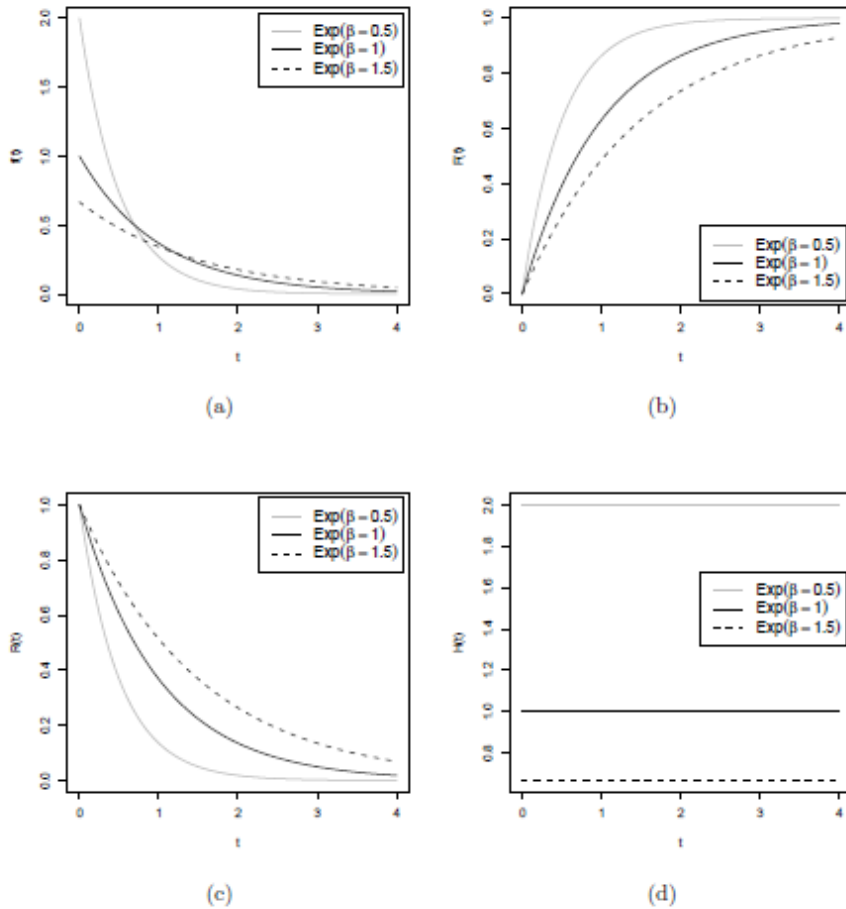


Figura 4: Distribución exponencial con distintos parámetros. (a) Función de densidad. (b) Función de distribución. (c) Función de sobrevivida. (d) tasa de riesgo
 Fuente: (Marshall & Olkin, 2007; Leiva, Barros, & Paula, www.victorleiva.cl, 2008).

Si $T \sim \text{Exp}(1)$ entonces la transformación lineal:

$$Y = \theta + \beta T; \quad -\infty < \theta < \infty, \quad \beta > 0,$$

extiende la familia exponencial a las distribuciones con función de distribución,

$$F_Y(y) = 1 - \exp\left(-\frac{[y - \theta]}{\beta}\right); \quad y \geq \theta,$$

la correspondiente función de densidad de Y es:

$$f_Y(y) = \frac{1}{\beta} \exp\left(-\frac{[y - \theta]}{\beta}\right); \quad y \geq \theta.$$

En este caso se utiliza la notación $Y \sim Exp(\theta, \beta)$ y los parámetros θ y β corresponden a la garantía o umbral (locación) y parámetro de escala respectivamente.

Una distribución exponencial con parámetros de locación θ y de escala β se denomina estándar cuando $\theta=0$ y $\beta=1$ ³.

3.2.1.2 Weibull

El modelo Weibull se utiliza para problemas de resistencia de materiales y es ampliamente citado como distribución de vida. La distribución de Weibull fue desarrollada por Walody Weibull en el año 1951. Un antecedente de esta distribución fue la desarrollada el año 1933 por Rosen P. y Rammles E. para describir "las leyes que rigen la finura del polvo de carbón", por lo que en ingeniería de materiales, la distribución de Weibull se la suele llamar distribución Rosen-Rammles. La distribución Weibull también se conoce con el nombre de distribución Fréchet, por Maurice Fréchet y en la literatura rusa, la se conoce como la distribución de Gnedenko, nombre debido a Boris Vladimirovich Gnedenko. Aunque la distribución Weibull no es un modelo de fatiga, su autor ha trabajado mucho en este tema (Leiva, Barros, & Paula, www.victorleiva.cl, 2008).

Desde el punto de vista probabilístico esta distribución corresponde a una variable aleatoria que es una transformación de potencia de otra variable aleatoria con distribución exponencial (Marshall & Olkin, 2007). Por lo tanto, el parámetro del modelo exponencial es el parámetro de forma de la distribución Weibull y el parámetro de potencia es el

³ Otras características de la distribución exponencial se pueden observar en Johnson, Kotz, & Balakrishnan (1994, pp. 494-556), Evans, Hastings, & Peacock, (2000, pp.77-81), y Marshall & Olkin (2007, pp. 291-360).

parámetro de escala de esta distribución ⁴.

El modelo de Weibull es una distribución de vida muy útil ya que corresponde a una familia que cubre una amplia gama de distribuciones que abarcan desde el modelo normal al modelo exponencial. La distribución de Weibull ha sido aplicada en diferentes áreas por ejemplo, en fatiga, resistencia de materiales, investigación genética, efectos de ozono en la agricultura, control de calidad, precipitaciones de lluvia, análisis de confiabilidad, y velocidad del viento, entre otros. Todos ellos muestran la utilidad de esta distribución.

Dada una variable $T \sim Weibull(b, c)$ su función de densidad es:

$$f_T(t) = \frac{ct^{c-1}}{b^c} \exp\left(-\left[\frac{t}{b}\right]^c\right); t > 0, b > 0, c > 0,$$

donde b es el parámetro de forma y c el parámetro de escala.

La función de distribución de T se expresa:

$$F_T(t) = 1 - \exp\left(-\left[\frac{t}{b}\right]^c\right); t > 0.$$

La media y la varianza de T son ⁵:

$$E(T) = b\Gamma\left(\frac{c+1}{c}\right),$$

$$\text{Var}(T) = b^2 \left[\Gamma\left(\frac{c+2}{c}\right) - \left[\Gamma\left(\frac{c+1}{c}\right) \right]^2 \right].$$

La función de sobrevivida y la tasa de riesgo de t se expresan:

⁴ Para más detalles sobre un parámetro de potencia y diferentes tipos de parámetros estadísticos, ver (Marshall & Olkin, 2007).

⁵ Dada $\alpha > 0$, se define la función Gamma o función factorial como $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$ Propiedades: Si $\alpha >$

$1, \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1);$ Si $\alpha \in \mathbb{N}, \Gamma(\alpha) = (\alpha - 1)!$.

$$R_T(t) = \exp\left(-\left[\frac{t}{b}\right]^c\right); t > 0,$$

$$h_T(t) = \frac{ct^{c-1}}{b^c}; t > 0.$$

Algunas características de la distribución Weibull son las siguientes:

- (i) La distribución Weibull es estándar cuando $b = 1$.
- (ii) Si el parámetro c pertenece al intervalo $(1; 2,6)$, la densidad de la distribución de Weibull tiene asimetría positiva. Si $c > 3,7$, entonces la densidad tiene asimetría negativa. Mientras que, si c pertenece al intervalo $(2,6; 3,7)$, la distribución Weibull es similar a la del modelo de distribución normal
- (iii) Cuando el parámetro $c = 1$, la distribución Weibull es una distribución exponencial de media b (Por lo tanto su tasa de riesgo será constante como se observa en la Figura 5-d).
- (iv) Cuando el parámetro $c > 1$, la distribución Weibull pertenece a la familia con tasa de riesgo creciente (Figura 5-d).
- (v) Cuando el parámetro $c < 1$, la distribución Weibull pertenece a la familia con tasa de riesgo decreciente ⁶ (Figura 5-d).

⁶ Se pueden ver más características de la distribución Weibull en (Leiva, Barros, & Paula, www.victorleiva.cl, 2008).

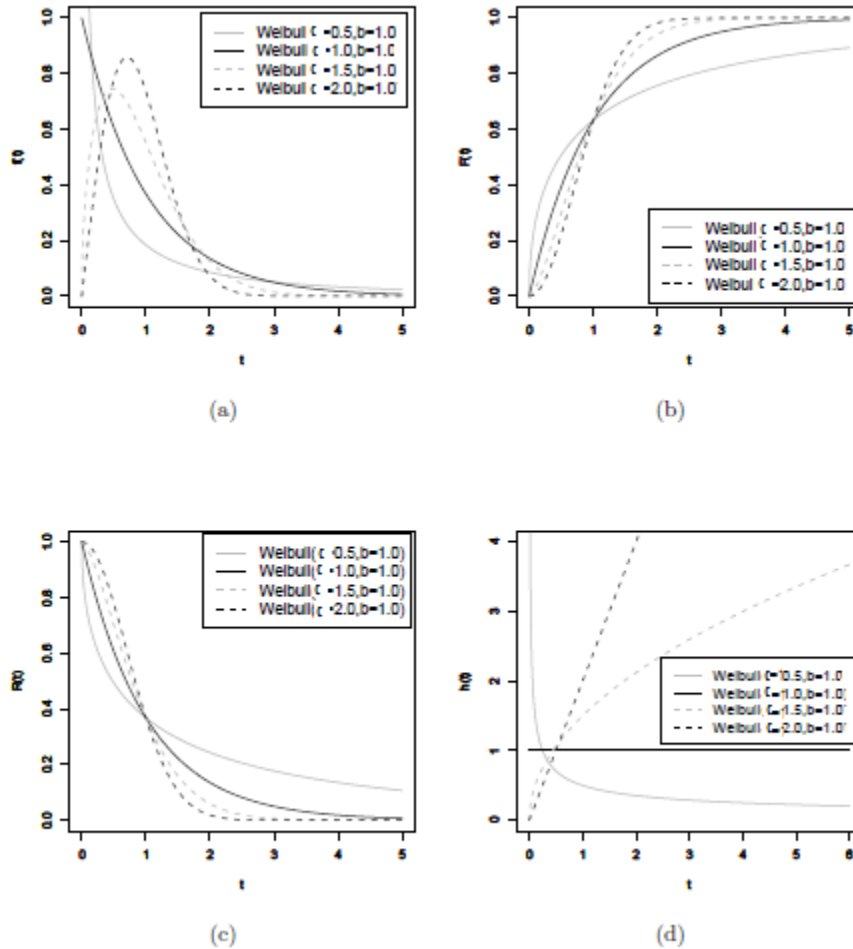


Figura 5: Distribución Weibull con distintos parámetros. (a) Función de densidad. (b) Función de distribución. (c) Función de sobrevivida. (d) Tasa de riesgo
Fuente: (Marshal & Olkin, 2007; Leiva, Barros, & Paula, www.victorleiva.cl, 2008).

3.2.1.3 Distribuciones de vida de la fatiga y daño acumulativo.

En esta sección abordamos algunas distribuciones de vida, que son reconocidas como vida de fatiga y distribuciones de daño acumulativo (Saunders, 2007).

La fatiga ha sido identificada como una causa importante de falla en estructuras de ingeniería. Cuando se estudia el efecto de vibraciones en cierto tipo de materiales o estructuras, lo primero que se considera es el mecanismo que provoca la fatiga de estos materiales, para comprender este proceso de fatiga se deben incorporar los conceptos de

grieta, ciclo, fatiga y carga. Considere un tipo de material expuesto a una secuencia de cargas durante un periodo de tiempo que llamamos ciclo. Cada ciclo produce sobre el material un desplazamiento o desvío (fatiga) por la aplicación de estas fuerzas o cargas, lo que le impone un estrés. En cada ciclo, se aplica la misma secuencia de cargas que en el ciclo anterior. La fatiga corresponde a un daño estructural (grieta), que se está acumulando dentro de cierto material a lo largo del tiempo ⁷.

Los modelos de vida de fatiga son: Gamma, Lognormal, Gaussiana inversa y Birnbaum-Saunders.

Gamma

El modelo gamma surge cuando la ruptura o falla de cierto material es causada por un ciclo de carga específico (shock) que ocurre según un proceso homogéneo de Poisson. La vida de fatiga, en este caso, viene dada por la suma de variables aleatorias, que representan el tiempo de ocurrencia entre cargas consecutivas. Ya que los shocks suceden de acuerdo con un proceso homogéneo de Poisson, el tiempo de ocurrencia de un shock sigue una distribución exponencial. Específicamente si consideramos un material expuesto a una serie de ciclos de carga, los cuales ocurren según un proceso homogéneo de Poisson de tasa λ , el tiempo entre los ciclos consecutivos, llamados T_1, T_2, \dots , son variables aleatorias independientes e idénticamente distribuidas con distribución exponencial de parámetro λ .

Si el material falla exactamente en el k -ésimo ciclo, entonces la vida de fatiga del material

⁷ Para más información sobre este proceso (Marshall & Olkin, 2007) (Saunders, 2007).

en el k -ésimo ciclo de carga, que denominamos T , sigue una distribución gamma de parámetros k y λ , esto es:

$$T = T_1 + \dots + T_k \sim \text{Gamma}(k, \lambda); \quad k \in \mathbb{N}, \lambda > 0.$$

La función de densidad de T es:

$$f_T(t) = \frac{\lambda}{\Gamma(k)} [\lambda t]^{k-1} \exp(-\lambda t); \quad t > 0, \lambda > 0, k \in \mathbb{N},$$

donde λ es la tasa de shock y $\Gamma(\cdot)$ es la función gamma.

La función de distribución de T se expresa,

$$F_T(t) = 1 - \sum_{x=0}^{k-1} \frac{[\lambda t]^x}{x!} \exp(-\lambda t); \quad t > 0.$$

La media y la varianza de T son:

$$E(T) = \frac{k}{\lambda},$$

$$\text{Var}(T) = \frac{k}{\lambda^2}.$$

La función de sobrevivencia y la tasa de riesgo son:

$$R_T(t) = \sum_{x=0}^{k-1} \frac{[\lambda t]^x}{x!} \exp(-\lambda t); \quad t > 0,$$

$$h_T(t) = \frac{\lambda [\lambda t]^{k-1}}{\Gamma(k) \sum_{x=0}^{k-1} \frac{[\lambda t]^x}{x!}}; \quad t > 0.$$

La distribución exponencial es un caso particular de la Gamma, cuando $k=1$:

$$\text{Exp}(\lambda) \equiv \text{Gamma}(1, \lambda); \text{ para } \lambda > 0$$

Algunas características de la distribución Gamma⁸ son las siguientes:

- (i) La función de densidad es completamente monótona, log convexa y decreciente, para $0 < k < 1$; log cóncava y unimodal, para $k > 1$ (Fig. 6-a).
- (ii) La función de supervivencia es log cóncava; para $k > 1$, log convexa, para $k < 1$ (Fig. 6-c).
- (iii) La función de distribución es log cóncava, para todo k (Fig. 6-b).
- (iv) La tasa de riesgo es creciente, para $k > 1$; es la constante λ , para $k = 1$; es decreciente $0 < k < 1$ (Fig. 6-d).

⁸ Se puede ampliar en (Marshall & Olkin, 2007).

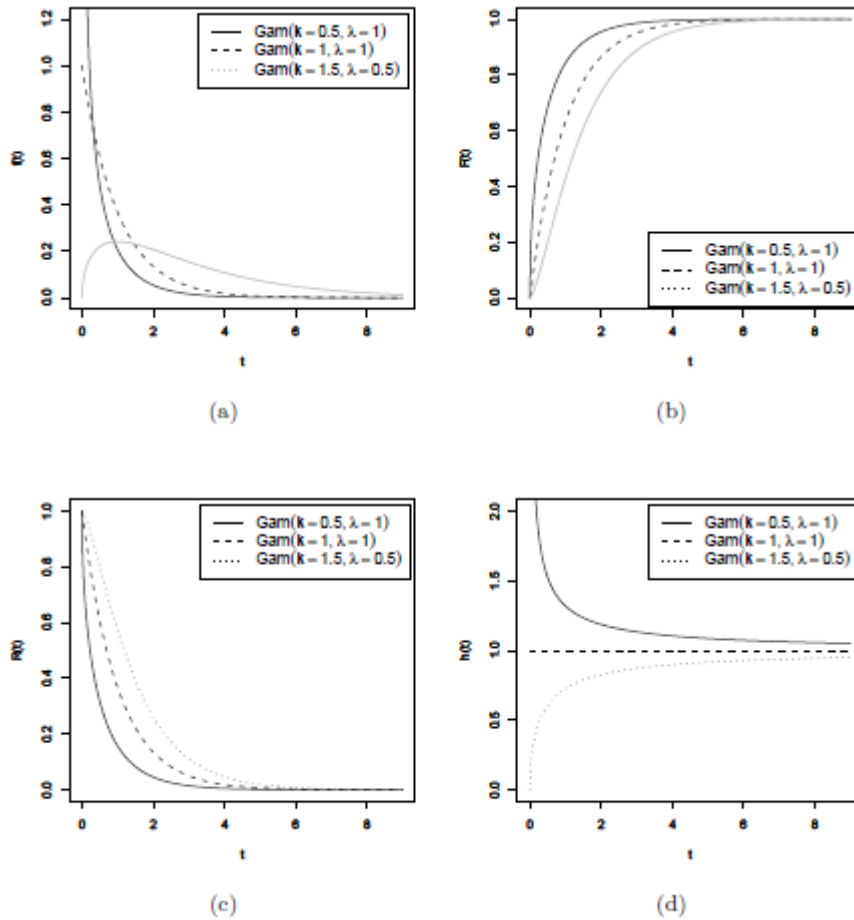


Figura 6: Distribución Gamma con distintos parámetros. (a) Función de densidad. (b) Función de distribución. (c) Función de supervida. (d) Tasa de riesgo.
 Fuente: (Marshal & Olkin, 2007; Leiva, Barros, & Paula, www.victorleiva.cl, 2008)

Lognormal

La distribución normal no es apropiada para variables de vida por dos motivos, el primero se debe a que la distribución normal admite valores negativos lo cual no es posible en variables de vida; la segunda razón es que la distribución normal es simétrica y las distribuciones de vida generalmente son asimétricas hacia la derecha. El modelo lognormal resulta de una de las transformaciones más simples de una variable normal que da lugar a una distribución asimétrica hacia la derecha que admite sólo valores positivos.

Sean T_1, T_2, \dots, T_k una serie de variables aleatorias que denotan los tamaños de una grieta dominante después de sucesivos ciclos de estrés. El crecimiento de la grieta durante el i -ésimo ciclo, es decir $T_i - T_{i-1}$ es proporcional al tamaño de la grieta T_{i-1} para $i=1, 2, \dots, k$, es decir que el material falla cuando la grieta alcanza el tamaño T_k . Específicamente, si,

$$\Delta T_i = T_i - T_{i-1}, \quad i = 1, 2, \dots, k,$$

entonces,

$$\Delta T_i = p_i T_{i-1},$$

donde p_i es una constante de proporcionalidad y corresponde a una variable aleatoria.

Así la suma desde el primer ciclo hasta el ciclo k -ésimo está dada por:

$$\begin{aligned} \sum_{i=1}^k p_i &= \sum_{i=1}^k \frac{\Delta T_i}{T_{i-1}}, \quad \text{ahora, si } \Delta T_i \rightarrow 0 \text{ y } k \rightarrow \infty, \text{ entonces,} \\ \sum_{i=1}^k p_i &= \int_{T_0}^{T_k} \frac{1}{T} dT = \log(T_k) - \log(T_0). \end{aligned}$$

Por el teorema central del límite $\sum_{i=1}^k p_i$ converge a una distribución normal, así

$Y = \log(T)$ asintóticamente sigue una distribución normal también. Si $Y = \log(T)$ sigue una distribución normal con media ν y varianza τ^2 , entonces la variable aleatoria T sigue una distribución lognormal con parámetros de posición y escala, ν y τ respectivamente.

La función de densidad de T es:

$$f_T(t) = \frac{1}{t\tau\sqrt{2\pi}} \exp\left(-\frac{[\log(t) - \nu]^2}{2\tau^2}\right); \quad t > 0, \nu \in \mathbb{R}, \tau > 0.$$

La función de distribución de T es:

$$F_T(t) = \Phi\left(\frac{\log(t) - \nu}{\tau}\right); \quad t > 0,$$

donde Φ es la función de distribución de una variable aleatoria normal estándar.

La media y la varianza son:

$$E(T) = \exp\left(\frac{\nu + \tau^2}{2}\right),$$

$$Var(T) = \exp(2\nu)[\exp(2\tau^2) - \exp(\tau^2)].$$

Las funciones de sobrevivida y tasa de riesgo son:

$$R_T(t) = \Phi\left(\frac{\nu - \log(t)}{\tau}\right); t > 0,$$

donde Φ es la función de distribución de una variable normal estándar,

$$h(t) = \frac{\phi([\log(t) - \nu]/\tau)}{\Phi([\nu - \log(t)]/\tau)}; t > 0,$$

donde ϕ es la función de densidad de una variable normal estándar.

Algunas características de la distribución Lognormal⁹ son las siguientes:

- (i) La función de densidad es unimodal (Fig. 7-a).
- (ii) La tasa de riesgo es unimodal (Fig. 7-d).

⁹ Se puede ampliar en (Marshall & Olkin, 2007).

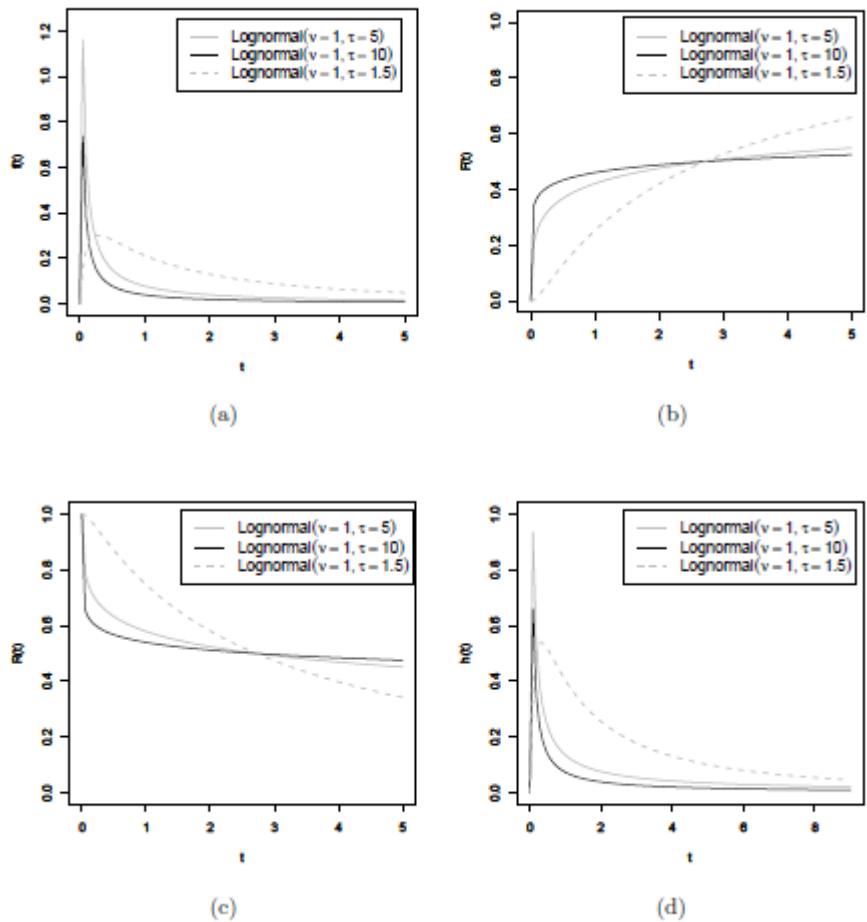


Figura 7: Distribución lognormal con distintos parámetros. (a)Función de densidad. (b)Función de distribución. (c)Función de supervida. (d)Tasa de riesgo
Fuente: (Marshal & Olkin, 2007; Leiva, Barros, & Paula, www.victorleiva.cl, 2008).

Gaussiana Inversa

La distribución gaussiana inversa tiene su origen en el proceso Wiener. El proceso de Wiener es un proceso estocástico de tiempo continuo que se caracteriza por tener tres propiedades importantes: 1. Es un proceso de Markov. Lo que significa que la distribución de probabilidad de todos los valores futuros del proceso depende únicamente de su valor actual, no siendo afectada por sus valores pasados, ni por ninguna otra información actual.

Por tanto, el valor actual del proceso es la única información necesaria para realizar la mejor estimación de su valor futuro. 2. Tiene incrementos independientes. Lo que significa que la distribución de probabilidad de los cambios en el proceso en cualquier intervalo temporal es independiente de la de cualquier otro intervalo. Así pues, si la variable aleatoria z sigue un proceso de Wiener sus variaciones (Δz) para cualesquiera dos pequeños intervalos de tiempo (Δt) son independientes. 3. Las variaciones en el proceso (Δz) producidas en un intervalo finito de tiempo (Δt) se distribuyen normalmente, con una varianza que aumenta linealmente con el tamaño del intervalo temporal. Por tanto, $\Delta z = \varepsilon \sqrt{\Delta t}$ (donde ε es una variable aleatoria del tipo $\phi[0,1]$).

Por lo tanto como distribución de vida de la fatiga, el modelo IG se puede considerar suponiendo un daño acumulativo durante el intervalo $[0, t]$ que provoca fatiga, lo cual es representado por el proceso Wiener $\{B(t): 0 \leq t \leq \infty\}$, con una función media $\mu(t)$ y una constante de difusión δ^2 . Si determinado material falla cuando $B(t)$ excede un valor crítico o umbral llamado C , entonces la vida de la fatiga $T \equiv \inf \{t: B(t) > C\}$, el cual corresponde al primer instante en que se supera al punto C . Así la distribución de vida de fatiga de T es la distribución del tiempo del primer paso del proceso Wiener y se corresponde con el modelo gaussiano inverso¹⁰.

Si una variable aleatoria T sigue la distribución gaussiana inversa con media $\mu > 0$ y parámetro de escala $\lambda > 0$, entonces la notación es $T \sim IG(\mu, \lambda)$ en este caso la densidad viene dada por:

¹⁰Se puede ampliar este tema en (Leiva, Barros, & Paula, www.victorleiva.cl, 2008).

$$f_T(t) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(-\frac{\lambda}{2\mu^2} \frac{[t-\mu]^2}{t}\right); \quad t > 0, \mu > 0, \lambda > 0.$$

La función de distribución de T es:

$$F_T(t) = \Phi\left(\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}}\right]\right) + \Phi\left(\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} + \sqrt{\frac{\mu}{t}}\right]\right) \exp\left(\frac{2\lambda}{\mu}\right); \quad t > 0.$$

La esperanza y la varianza de T son:

$$E(T) = \mu,$$

$$Var(T) = \frac{\mu^3}{\lambda}.$$

La función de sobrevivencia y la tasa de riesgo son:

$$R_T(t) = \Phi\left(-\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}}\right]\right) - \Phi\left(\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} + \sqrt{\frac{\mu}{t}}\right]\right) \exp\left(\frac{2\lambda}{\mu}\right); \quad t > 0,$$

$$h_T(t) = \frac{\sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(-\frac{\lambda}{2\mu^2} \frac{[t-\mu]^2}{t}\right)}{\Phi\left(-\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}}\right]\right) - \Phi\left(\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} + \sqrt{\frac{\mu}{t}}\right]\right) \exp\left(\frac{2\lambda}{\mu}\right)}; \quad t > 0.$$

La tasa de riesgo de la gaussiana inversa comienza en cero y aumenta hasta un único máximo, luego disminuye hasta un valor límite positivo distinto de cero¹¹ (Figura 8-d).

¹¹ Se puede ver la demostración en (Marshall & Olkin, 2007).

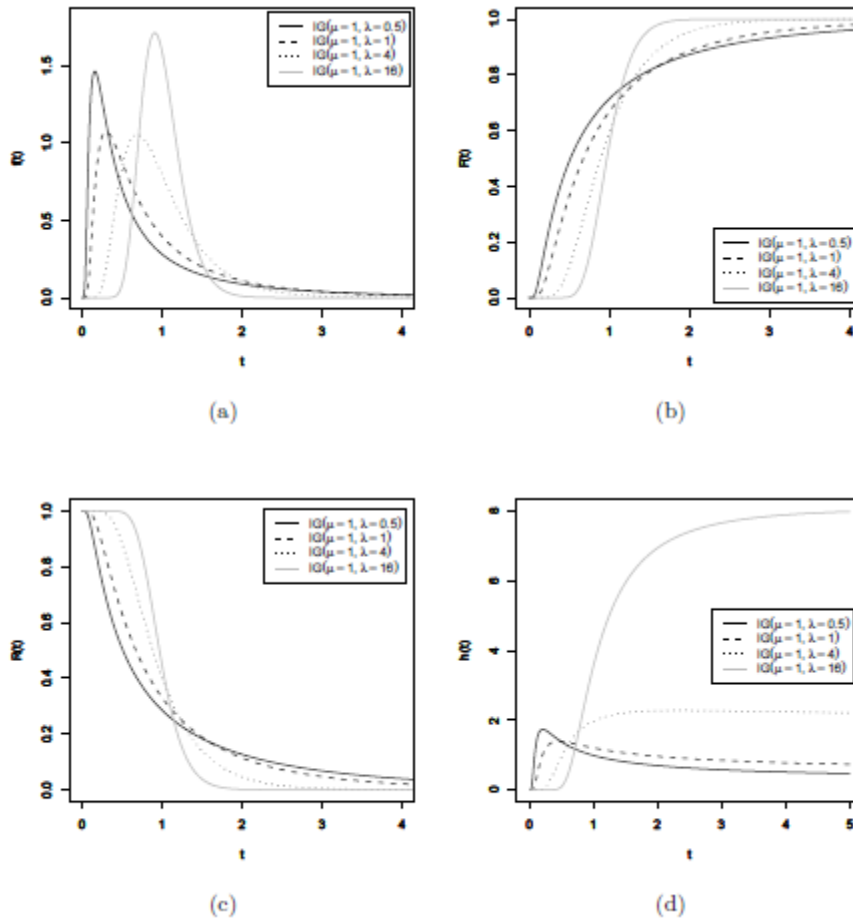


Figura 8: Distribución Gaussiana Inversa con distintos parámetros. (a) Función de densidad. (b) Función de distribución. (c) Función de sobrevivida. (d) Tasa de riesgo
Fuente (Leiva, Barros, & Paula, www.victorleiva.cl, 2008)

Birnbaum-Saunders

Birnbaum y Saunders (1969) introdujeron la distribución que lleva sus nombres con el propósito de modelar la vida de la fatiga de metales sujetos a estrés periódico, por eso también se la conoce como distribución de vida de la fatiga. Birnbaum y Saunders aplican el teorema central del límite al proceso de crecimiento de la grieta y aproximan

el número de ciclos de tensión hasta la falla a una variable aleatoria continua (Marshall & Olkin, 2007), es decir, el número de ciclos hasta la falla fue reemplazado por el tiempo.

El proceso de la fatiga

El proceso de fatiga de materiales consta de las siguientes tres etapas:

- i) Comienzo de una fisura imperceptible.
- ii) Crecimiento y propagación de la fisura producto del estrés cíclico y la tensión.
- iii) Ruptura o falla del material debido a fatiga, evento puntual que no ocupa tiempo.

Justificación física

Considere un espécimen de material expuesto a una secuencia de m cargas cíclicas, digamos $\{l_i, i = 1, \dots, m\}$. El esquema de carga es el siguiente:

$$\begin{array}{ccc}
 l_1, l_2, \dots, l_m & l_{m+1}, l_{m+2}, \dots, l_{2m} & l_{jm+1}, l_{jm+2}, \dots, l_{jm+m} \\
 \text{ciclo 1} & \text{ciclo 2} & \text{ciclo } j + 1
 \end{array}$$

donde $l_{jm+i} = l_{km+i}$, para $j \neq k$.

Bajo los siguientes supuestos:

- *Un espécimen de un material está expuesto a cargas cíclicas que producen desgaste.
- *La ruptura (falla) ocurre cuando el tamaño de la grieta excede un umbral denotado por w .
- *La secuencia de cargas son las mismas de un ciclo a otro.

*El crecimiento de la grieta, X_i , debido a la carga l_i durante el ciclo j -ésimo es una variable aleatoria.

*El tamaño total de la grieta, Y_j , debido al ciclo j -ésimo es también una variable aleatoria con media μ y varianza σ^2 .

*Las variables aleatorias Y_i son independientes (Leiva, Barros, Paula, & Saunders, 2010).

La extensión de la grieta debido al ciclo $(j + 1)$ -ésimo de carga es

$$Y_{j+1} = X_{jm+1} + X_{jm+2} + \dots + X_{jm+m}; \quad j, m = 0, 1, 2, \dots$$

Bajo una aplicación repetitiva de n ciclos de cargas, el tamaño acumulado de la grieta al

final del ciclo n -ésimo está dado por $Z_n = \sum_{i=1}^n Y_i$.

Entonces, basados en la expresión anterior, los dos últimos supuestos estadísticos y en el teorema del límite central, es posible establecer que Z_n converge en distribución a una normal con media $n\mu$ y varianza $n\sigma^2$, cuando $n \rightarrow \infty$ (Leiva, Barros, Paula, & Saunders, 2010).

Sea N el número de ciclos requeridos hasta la falla. La función de distribución de N , basada en el teorema de probabilidad total, es

$$\begin{aligned} P(N \leq n) &= P(N \leq n, Z_n > w) + P(N \leq n, Z_n \leq w) \\ &= P(Z_n > w) + P(N \leq n, Z_n \leq w), \end{aligned}$$

donde w es el umbral dado en el segundo supuesto. Note que

$P(N \leq n, Z_n \leq w) > 0$, ya que $Z_n \sim N(n\mu, n\sigma^2)$, lo cual puede despreciarse tal que

$$P(N \leq n) \approx P(Z_n > w), \text{ así}$$

$$P(N \leq n) \approx P\left(\frac{Z_n - n\mu}{\sigma\sqrt{n}} > \frac{w - n\mu}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{\sqrt{w\mu}}{\sigma} \left[\sqrt{\frac{n}{w/\mu}} - \sqrt{\frac{w/\mu}{n}}\right]\right);$$

donde $\Phi(\cdot)$ es la función de distribución de la $N(0,1)$.

Parametrización

Birnbaum & Saunders usaron la ecuación anterior para definir una distribución estadística continua idealizando la variable aleatoria discreta N mediante la variable aleatoria continua T , es decir el número de ciclos hasta la falla fue reemplazado por el tiempo hasta que la falla ocurre (Marshall & Olkin, 2007). Así, tomando

$$\alpha = \frac{\sigma}{\sqrt{w\mu}}, \quad \beta = \frac{w}{\mu}, \quad a(t; \alpha, \beta) = \frac{1}{\alpha} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right],$$

se obtiene la función de distribución de la BS para el tiempo de vida de la fatiga con parámetros de forma α y de escala β ,

$$F_T(t) = \Phi(a_t(\alpha, \beta)); \quad t > 0, \alpha > 0, \beta > 0,$$

significa que se puede decir que la variable aleatoria T sigue la distribución Birbaum-Saunders con parámetros de forma y escala, $\alpha > 0$ y $\beta > 0$, respectivamente, $T \sim BS(\alpha, \beta)$; esto se puede escribir,

$$T = \beta \left[\frac{\alpha}{2} Z + \sqrt{\left[\frac{\alpha}{2} Z \right]^2 + 1} \right]^2,$$

donde Z es una variable aleatoria que sigue la distribución normal estándar tal que,

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim N(0,1).$$

La densidad de T es

$$f_T(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2}\left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right) \frac{t^{-\frac{3}{2}}[t + \beta]}{2\alpha\sqrt{\beta}}; \quad t > 0.$$

La función de distribución es

$$F_T(t) = \Phi\left(\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right); \quad t > 0.$$

La media y la varianza de T son

$$E(T) = \beta \left[1 + \frac{\alpha^2}{2}\right]$$

y

$$\text{Var}(T) = \beta^2 \alpha^2 \left[\frac{5}{4} \alpha^2 + 1\right],$$

respectivamente.

La función de sobrevivencia de la BS es

$$R_T(t) = \Phi(-a(t; \alpha, \beta)); \quad t > 0.$$

La tasa de falla de la BS se obtiene

$$h_T(t) = \frac{\phi\left(\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right) t^{-3/2}[t + \beta]}{\Phi\left(\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right) 2\alpha\sqrt{\beta}}; \quad t > 0.$$

Algunas características de la distribución Birbaum-Saunders son las siguientes:

- (i) La función de supervivencia decrece a medida que α aumenta (Figura 9-a).
- (ii) La tasa de riesgo tiende a ser creciente a medida que α decrece (Figura 9-b).

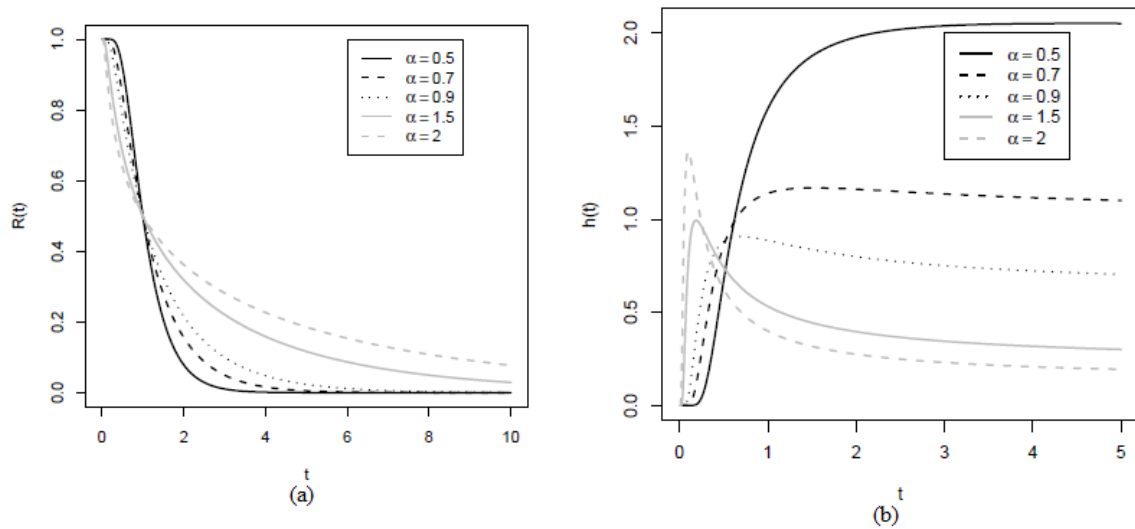


Figura 9: Distribución Birbaum &Saunders con distintos parámetros. (a) Función de supervivencia. (b) Tasa de riesgo
Fuente: (Leiva, Soto, Cabrera, & Cabrera, 2011).

3.2.2 Los nuevos modelos de Variables de Vida

3.2.2.1 Distribución Birbaum-Saunders Generalizada

Una distribución de vida altamente flexible que admite diferentes grados de curtosis y asimetría, que además posee unimodalidad y bimodalidad es la distribución Birbaum-Saunders Generalizada (BSG) propuesta por Díaz-García & Leiva (2002); Díaz-García

& Leiva (2005). Las distribuciones BSG producen modelos cuyos parámetros estimados, suelen ser robustos cuando hay datos atípicos, como lo demostraron Barros (2007), Leiva, Barros, Paula y Sanhueza (2008), y Barros, Paula y Leiva (2008).

Además, como se ha mencionado, hay argumentos teóricos que justifican el uso de la distribución BS en situaciones que implican daños acumulativos, lo cual ocurre comúnmente en estudios de ingeniería, medioambientales y médicos, estos mismos argumentos pueden servir para convalidar el uso de las distribuciones BSG (Leiva, Barros, Paula, & Sanhuenza, 2008).

Se parte de la variable aleatoria T que tiene distribución Birnbaum-Saunders:

$$T \sim BS(\alpha, \beta), \quad \alpha \text{ es el parámetro de forma y } \beta \text{ de escala.}$$

Como se mencionó, una variable aleatoria T que sigue la distribución BS permite la siguiente representación estocástica

$$T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\frac{\alpha^2 Z^2}{2} + 1} \right]^2; \quad \text{donde } \alpha > 0, \beta > 0, Z \sim N(0,1),$$

entonces Z se puede representar estocásticamente en términos de T,

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim N(0,1).$$

Díaz, García y Leiva propusieron la generalización de la BS, basándose en la familia de distribuciones de contornos elípticos (CE) unidimensionales que corresponden a todas las distribuciones simétricas en reales que se caracterizan por un parámetro de posición, un parámetro de escala y una función generadora de densidades o kernel g, es decir:

$$X \sim CE(\mu, \sigma^2; g)$$

siendo la normal estándar una esférica en reales:

$$Z \sim CE(0,1; g) \equiv S(g)$$

Por lo tanto:

$$Z = \frac{1}{\alpha} \left[\left(\frac{T}{\beta} \right)^{1/2} - \left(\frac{\beta}{T} \right)^{1/2} \right] \sim s(g)$$

Así, se logra generalizar la distribución Birnbaum-Saunders a partir de una distribución Elíptica (ver Anexo 1).

$$T = \beta \left[\frac{\alpha}{2} Z + \sqrt{\left(\frac{\alpha}{2} Z \right)^2 + 1} \right]^2 \sim BSG(\alpha, \beta; g) \quad \alpha, \beta > 0$$

La densidad de la BSG es

$$f_T(t) = cg \left(\frac{1}{\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2 \right] \right) \frac{t^{-3/2} [t + \beta]}{2\alpha\beta^{1/2}}; \quad t > 0,$$

donde c es la constante de normalización y $g(\cdot)$ es el kernel de la distribución $Z \sim S(g)$

Cabe destacar que la distribución BSG obtenida a partir del kernel de una distribución normal estándar coincide con la distribución Birnbaum-Saunders clásica (Leiva, Barros, Paula, & Saunders, 2010).

La distribución BSG puede ser obtenida desde los kernels (Figura 10-a):

- Cauchy.
- Laplace o exponencial doble
- Logística
- T-Student
- Exponencial potencia
- Normal (BS clásica).

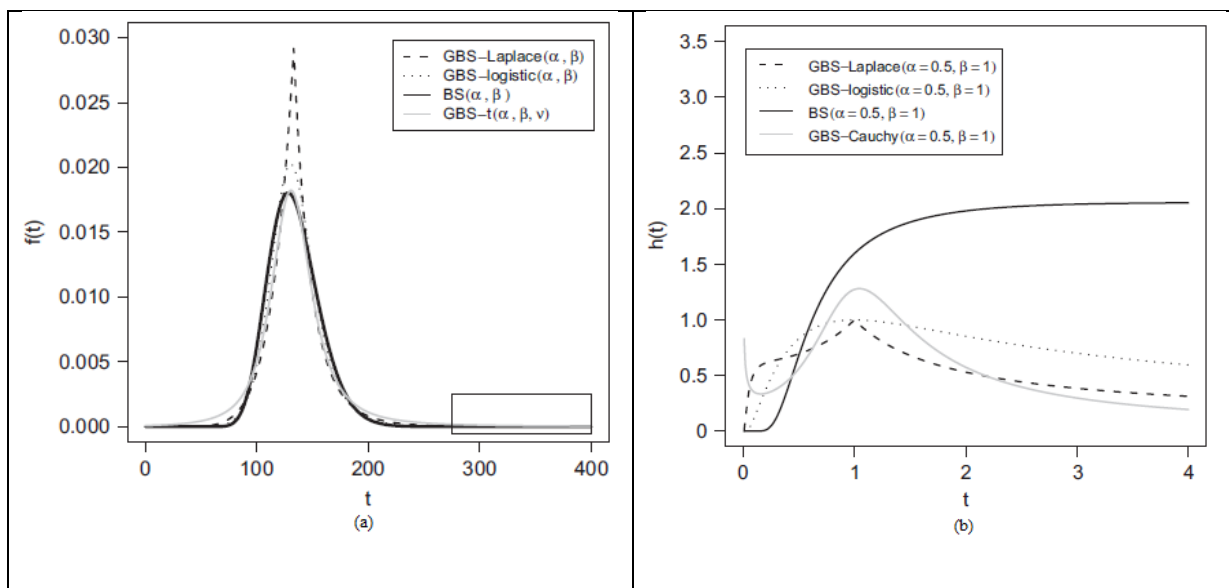


Figura 10: Distribución BSG para los kernels y parámetros indicados. (a) Función de densidad. (b) Tasa de riesgo

Fuente: Leiva, Riquelme, Balakrishnan, & Sanhuenza, 2008.

Algunas propiedades de la BSG son:

1) $cT \sim \text{BSG}(\alpha, c\beta; g), \quad c > 0.$

2) $\frac{1}{T} \sim \text{BSG}\left(\alpha, \frac{1}{\beta}; g\right).$

3) La función de distribución de T es $F_T(t) = F_Z(a(t; \alpha, \beta))$ donde $F_Z(\cdot)$ es la función de distribución de

$$Z \sim S(g) \text{ y } a(t; \alpha, \beta) = \frac{1}{\alpha} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right].$$

4) La función de cuantil de T es

$$t(q) = \left[\frac{\beta}{4} \right] \left[\alpha Z_q + \sqrt{\alpha^2 Z_q^2 + 4} \right]^2,$$

donde Z_q es el cuantil q -ésimo de $Z \sim S(g)$.

Si $q = 0.5$, entonces $t(0.5) = \beta$ es decir que β es también la mediana de la distribución BSG.

La tasa de riesgo de T es:

$$h_T(t) = \frac{f_Z(a_t)}{F_Z(-a_t)} A_t; \quad t > 0, \quad 0 < F_Z(\cdot) < 1,$$

$$\text{donde } A_t = \frac{d}{dt} a_t = \frac{t^{-\frac{3}{2}} [t + \beta]}{2\alpha\sqrt{\beta}}.$$

La tasa de riesgo de la distribución BSG crece hasta un punto de cambio, donde permanece estable durante un periodo y luego decrece (Leiva, Riquelme, Balakrishnan, & Sanhuenza, 2008), es decir que tiene forma de bañera invertida *^-shaped* (Figura 10-b).

$$E(T) = \frac{\beta}{2} [2 + u_1 \alpha^2],$$

donde $u_1 = E[U^1]$ con $U = Z^2 \sim G_{X^2}(1; g)$, siendo G_{X^2} es una variable Chi cuadrado generalizada.

$$V(T) = \frac{\beta^2 \alpha^2}{4} [4u_1 + [2u_2 - u_1^2] \alpha^2]$$

donde $u_2 = E[U^2]$.

3.2.2.2 Distribución Inversa Gaussiana Tipo

Antonio Sanhueza, Víctor Leiva y N. Balakrishnan (Sanhuenza, Leiva, & Balakrishnan, 2008) presentaron nuevos modelos de distribuciones relacionados con la distribución inversa Gaussiana, esta nueva distribución es llamada distribución inversa Gaussiana tipo (IGT) y se basa en la familia de distribuciones univariadas de contorno elíptico, que contienen todas las distribuciones simétricas en reales y se caracterizan por tener curtosis más flexible que la distribución inversa Gaussiana clásica, permitiendo distintos grados de curtosis y de asimetría.

De esta manera, la distribución IGT presenta curtosis más flexible que la distribución inversa Gaussiana y sus estimaciones de parámetros son más robustas frente a la presencia de datos atípicos.

Dada la variable aleatoria T que tiene distribución Inversa Gaussiana con media μ y parámetro de escala λ su función de densidad y su función de distribución son:

$$f_T(t) = \phi \left(\sqrt{\frac{\lambda}{\mu}} \left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}} \right] \right) \sqrt{\frac{\lambda}{t^3}} ;$$

donde $t > 0, \mu > 0, \lambda > 0,$

la función de distribución de T es:

$$F_T(t) = \Phi \left(\sqrt{\frac{\lambda}{\mu}} \left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}} \right] \right) + \Phi \left(\sqrt{\frac{\lambda}{\mu}} \left[\sqrt{\frac{t}{\mu}} + \sqrt{\frac{\mu}{t}} \right] \right) e^{\left(\frac{2\lambda}{\mu}\right)}; t > 0.$$

En las funciones anteriores se observa que existe una relación entre la distribución inversa Gaussiana con la normal estándar, sin embargo, se conoce que la normal produce

estimadores a menudo sensibles a los datos atípicos. Esta sensibilidad se puede transferir a cualquier distribución relacionada con la normal y sus estimaciones de parámetros también serán sensibles a datos atípicos (Sanhuenza, Leiva, & Balakrishnan, 2008).

En este sentido Sanhuenza, Leiva y Balakrishnan proponen la IGT que se basa en las distribuciones univariadas de contorno elíptico, lo cual es posible en virtud de la relación con la normal estándar.

Dada la variable aleatoria T que tiene distribución inversa gaussiana tipo, esto se expresa, $T \sim IGT(\mu, \lambda; g)$, siendo su función de densidad:

$$f_T(t) = cg(K_t) \frac{\sqrt{\lambda}}{\sqrt{t^3}};$$

c constante de normalización, g kernel de la distribución esférica asociada,

$$K_t = \left(\frac{\lambda}{\mu}\right) \left(\frac{t}{\mu} + \frac{\mu}{t} - 2\right); \quad t, \mu, \lambda > 0.$$

La IGT se puede generar desde los kernels Normal, Cauchy, Laplace, Logística y t de Student.

La IGT obtenida a partir del kernel normal estándar es la IG, así vemos que la inversa gaussiana clásica, que se basa en la distribución normal, es un caso particular de la IGT, es decir que las propiedades de la inversa gaussiana clásica se pueden extender al caso generalizado ¹².

La función de distribución de la IGT es:

$$F_T(t) = F\left(\sqrt{\frac{\lambda}{\mu}} \left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}} \right]\right) + \int_{\sqrt{\frac{\lambda}{\mu}} \left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}} \right]}^{\infty} cg\left(u^2 - \frac{4\lambda}{\mu}\right) du;$$

donde $F(\cdot)$ denota la función de distribución de la distribución esférica asociada.

¹² Para ver más sobre la IGT ver (Sanhuenza, Leiva, & Balakrishnan, 2008)

La función de sobrevivencia de la IGT es:

$$R_T(t) = F\left(-\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}}\right]\right) - \int_{\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}}\right]}^{\infty} \text{cg}\left(u^2 - \frac{4\lambda}{\mu}\right) du;$$

la tasa de riesgo de la IGT es

$$h(t) = \frac{\text{cg}(K_t) \frac{\sqrt{\lambda}}{\sqrt{t^3}}}{F\left(-\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}}\right]\right) - \int_{\sqrt{\frac{\lambda}{\mu}}\left[\sqrt{\frac{t}{\mu}} - \sqrt{\frac{\mu}{t}}\right]}^{\infty} \text{cg}\left(u^2 - \frac{4\lambda}{\mu}\right) du}.$$

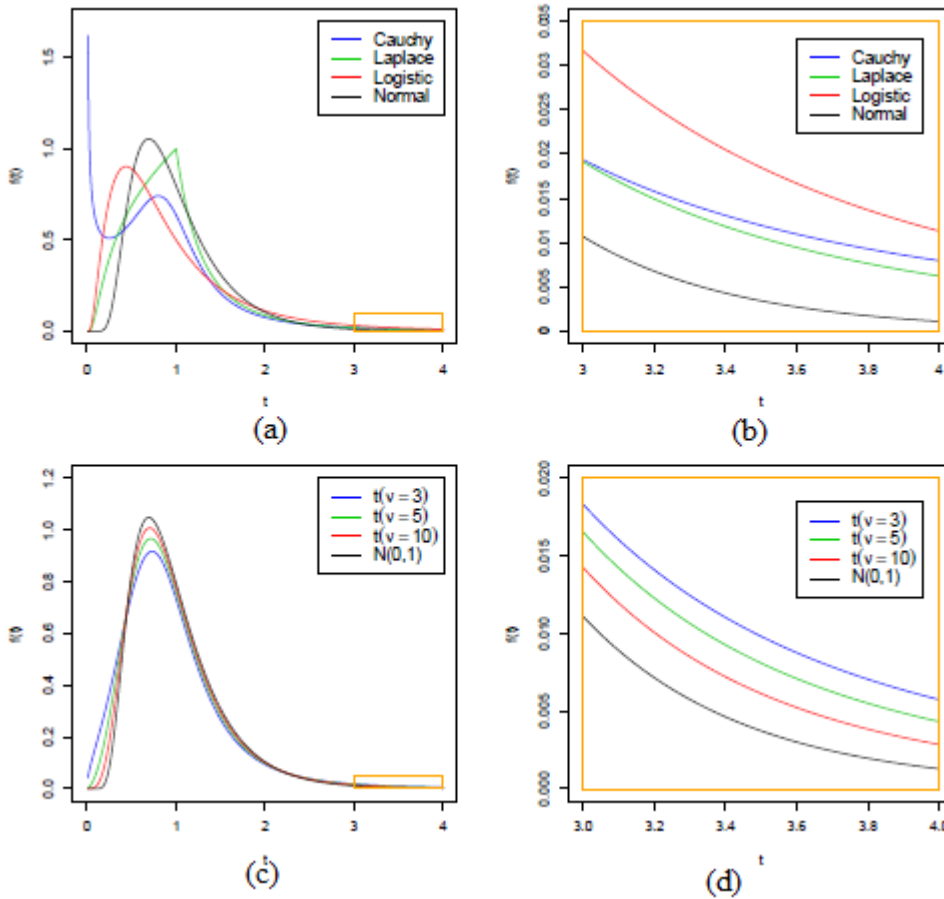


Figura 11: Distribución Inversa Gaussiana Tipo, función de densidad para $\mu=1$, $\lambda=4$ y kernels indicados. (a) Kernels Cauchy, Laplace, Logística y Normal. (b) Zoom de cola derecha del gráfico a. (c) Kernels t-Student con ν grados de libertad y Normal. (d) Zoom de cola derecha del gráfico c

Fuente: (Leiva, Hernández, & Sanhuenza, An R Package for a General Class of Inverse, 2008).

En la figura 11 se observa el comportamiento de la función de densidad de la distribución IGT, obtenida desde distintos kernels. La IGT obtenida desde los kernels Laplace, Logística y t de Student; presentan mayor curtosis que la clásica IG (Figuras 11- a y c). La curtosis, de la distribución IGT obtenida desde el kernel t de Student, es flexible dependiendo de los grados de libertad (Figura 11-c); lo cual hace que este último modelo, realice estimaciones de parámetros robustas a la presencia de datos atípicos.

$$E(T) = \mu$$

$$Var(T) = \frac{\mu^2}{\lambda/\mu} u_1$$

donde $u_1 = E[U^1]$ con $U = Z^2 \sim G_{\chi^2}(1: g)$, siendo G_{χ^2} es una variable Chi cuadrado generalizada.

3.3 Selección del Modelo de Ajuste

Las distribuciones de variables de vida elegidas para estudiar el comportamiento de la distribución del índice de exposición; serán las tradicionales: exponencial, Weibull, lognormal, gamma, Birnbaum-Saunders y Gaussiana inversa, así como aquellas introducidas como generalizaciones; los modelos BSG (Birnbaum-Saunders Generalizada) y los modelos IGT (Inversa Gaussiana Tipo).

Para determinar el mejor ajuste en las colas de la distribución, en este trabajo se inicia con métodos gráficos del análisis exploratorio; como el histograma, el gráfico que coteja la función de distribución acumulativa empírica con las funciones de distribución teóricas, los gráficos Q-Q plot y los P-P plot, que pueden ser muy informativos en algunos

casos (Delignette-Muller & Dutang, 2015). Estos últimos dos tipos de gráficos permiten observar cuál es la distribución que brinda un mejor ajuste en la cola derecha de la distribución empírica, especialmente si esta cola es importante en el uso de la distribución, como es en el contexto de la evaluación del riesgo (Delignette-Muller & Dutang, 2015). Luego se emplean los criterios de información de Akaike (AIC), el Bayesiano de Schwarz (BIC) y el de Hannan-Quinn (HQIC) para la selección del modelo más apropiado. Para aplicar estos criterios, se estimarán los parámetros correspondientes a cada distribución, empleando el método de máxima verosimilitud. Los criterios se obtienen de las siguientes expresiones:

$$AIC = -\frac{l(\hat{\theta})}{n} + \frac{p}{n}$$

$$BIC = -\frac{l(\hat{\theta})}{n} + \frac{p \log(n)}{2n}$$

$$HQIC = -\frac{l(\hat{\theta})}{n} + \frac{p \log(\log(n))}{n}$$

donde $l(\hat{\theta})$ es la función de log-verosimilitud para el parámetro θ asociado al modelo, n es el tamaño de la muestra y p es la dimensión del espacio paramétrico.

Los tres criterios se basan en el Principio de Parsimonia, es decir que penalizan la componente de la función de verosimilitud cuando el modelo se hace más complejo, por ejemplo, por mayor cantidad de parámetros (Vilca, Sanhuenza, Leiva, & Christakos, 2010). Por lo tanto, el modelo con el menor valor obtenido para el criterio de información, resultará el mejor para la descripción del comportamiento del índice.

3.3.1 Estadísticos Cramér-von Mises y Anderson-Darling.

Se completará la evaluación de los criterios de información con dos estadísticos que se utilizan en el caso de las distribuciones asimétricas ya que en la mayoría de éstas aparecen los parámetros de forma, lo cual complica los test de bondad de ajuste (Chen & Balakrishnan, 1995). Por eso se proponen estos tests que utilizan pocos puntos críticos y se basan en la función de distribución empírica, esta es:

$$F_n(x) = n^{-1} \sum_{i=1}^n I(x_i \leq x),$$

donde $I(a \leq b) = 1$ si $a \leq b$ y $I(a \leq b) = 0$ si $a > b$; para x real.

Cualquier estadístico que mide la diferencia entre F_n y F (función de distribución teórica ajustada), es considerado un estadístico EDF (función de distribución empírica). Se elige usar dos estadísticos EDF, Cramér-von Mises y Anderson-Darling. Ambos estadísticos de bondad de ajuste tienen como objetivo medir la distancia entre lo acumulado por la función de distribución empírica y la función de distribución teórica (Gemai Chen, N. Balakrishnan 1995).

Dado $\hat{\theta}$ estimador de θ luego el estadístico de Cramér-von Mises es

$$W^2 = n \int_{-\infty}^{+\infty} \{F_n(x) - F(x; \hat{\theta})\}^2 dF(x; \hat{\theta}),$$

y el de Anderson-Darling es

$$A^2 = n \int_{-\infty}^{+\infty} \frac{\{F_n(x) - F(x; \hat{\theta})\}^2}{\{F(x; \hat{\theta})(1 - F(x; \hat{\theta}))\}} dF(x; \hat{\theta}).$$

El estadístico con menor valor representará al mejor ajuste.

3.3.2 Validación de la distribución elegida con los gráficos TTT

Por último, se validará la distribución elegida con los gráficos TTT.

Una variable de vida T con función de densidad $f(t)$ y función de distribución $F(t)$, también se puede caracterizar por su función de supervivencia $R(t)$ y por su tasa de riesgo $h(t)$. Una propiedad interesante de la tasa de riesgo es que hace posible caracterizar el comportamiento de las distribuciones de vida. Diversos modelos de probabilidad que poseen densidades con formas similares, en algunos casos, tienen tasas de riesgo distintas, por lo que es muy importante tener en cuenta la tasa de riesgo para seleccionar o descartar distribuciones (Leiva, Soto, Cabrera, & Cabrera, 2011). Estimar la tasa de riesgo es difícil. Una herramienta que permite determinar la forma de la tasa de riesgo es el gráfico TTT (para la construcción de este gráfico ver Anexo 2), en inglés *total time on test*. Por medio de este gráfico se puede detectar el tipo de tasa de riesgo que tiene el índice IE (Aarset, 1987).

En la figura 4 es posible observar diferentes formas teóricas para el gráfico TTT, las cuales están asociadas con la tasa de riesgo respectiva. Si este gráfico arroja una curva cóncava, la tasa de riesgo es creciente. Por el contrario, si esta curva es convexa, la tasa de riesgo es decreciente. Ahora bien, si el gráfico TTT produce una curva que es primero cóncava y luego convexa, la tasa de riesgo tiene forma de bañera invertida \cap -shaped. Si esta curva es convexa y luego cóncava, la tasa de riesgo tiene forma de bañera \cup -shaped. Por último, si el gráfico TTT produce una línea recta, la tasa de riesgo es constante, como es el caso de la distribución exponencial. Así, el gráfico TTT nos puede dar una indicación del tipo de distribución asociada con los datos (Leiva, Soto, Cabrera, & Cabrera, 2011).

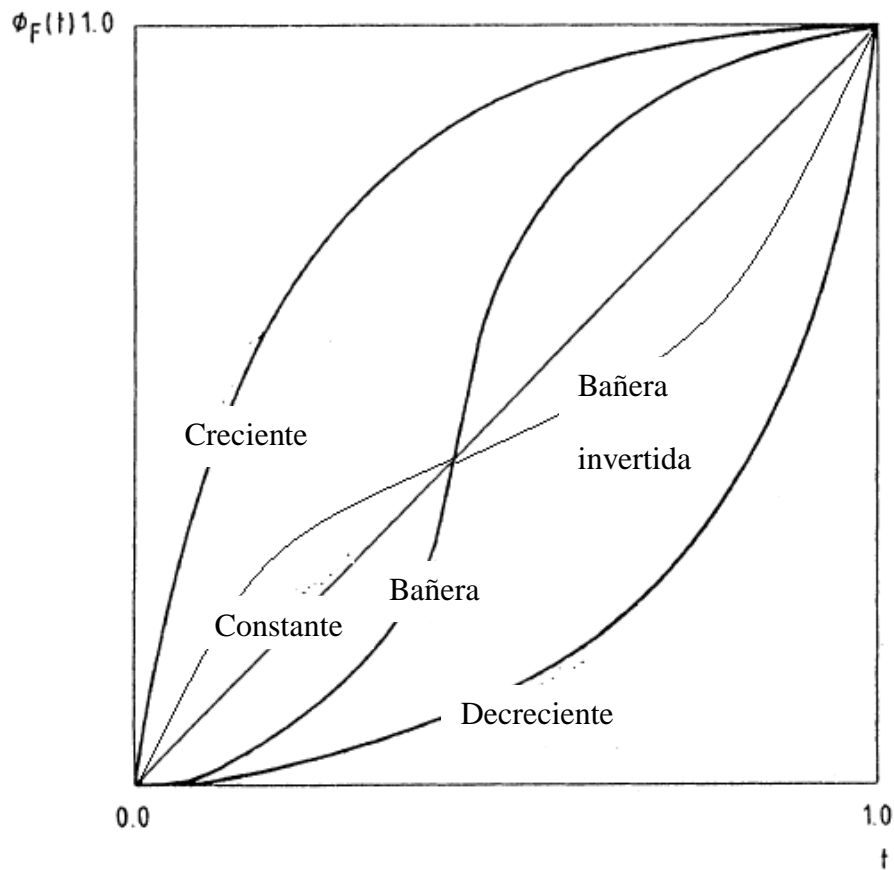


Figura 12: Transformación TTT escalada de distribuciones con los tipos de tasa de riesgo según cada curva
 Fuente: (Aarset, 1987).

De esta manera, así como el gráfico Q-Q plot es útil para aproximar la distribución, el gráfico TTT, nos puede indicar el tipo de distribución asociada con los datos (Vilca, Sanhuenza, Leiva, & Christakos, 2010).

Si la tasa de riesgo es creciente, *IHR increasing hazard rate*, la probabilidad de supervivencia decrece con el tiempo, por eso esta clase puede ser adecuada para modelar tiempos de vida donde se presenta desgaste o envejecimiento (Vilca, Sanhuenza, Leiva 2010). En otros casos es conveniente emplear la clase U-shaped HR, conocida como *bathtub curve*, como en el caso en que se tiene que modelar la mortalidad humana, ya que la tasa de riesgo en los primeros años de vida decrece, volviéndose estable durante cierto periodo luego del cual la tasa de riesgo es creciente (Vilca, Sanhuenza, Leiva 2010). Cuando se

analizan datos de fatiga o de contaminación ambiental con frecuencia es adecuado considerar la clase de tasa de riesgo de forma de bañera invertida, porque la tasa de riesgo crece hasta un punto a partir del cual se mantiene estable y luego decrece. La clase de tasas de riesgo decreciente, *DHR decreasing hazard rate* puede aplicarse en casos de sobrevivencia después de una cirugía exitosa en la cual al comienzo existe un riesgo inicial de infección, que luego decrece a medida que el paciente se recupera (Vilca, Sanhuenza, Leiva 2010). La tasa de riesgo constante se da cuando el modelo adecuado es el exponencial lo cual suele presentarse en los tiempos de vida de los chips de computadora (Vilca, Sanhuenza, Leiva 2010).

Las distribuciones Gamma y Weibull admiten tasas de riesgo creciente y decreciente.

La tasa de riesgo de la distribución Lognormal tiene un comportamiento no monótono ya que aumenta hasta su punto de cambio y luego disminuye a cero, es decir que tiene la forma de bañera invertida \cap -shaped. Las tasas de riesgo de las distribuciones Birbaum-

Sanders e Inversa Gaussiana, tienen un comportamiento similar a la tasa de riesgo de la Lognormal, pero disminuyen hasta un valor constante positivo distinto de cero (Vilca, Sanhuenza, Leiva, & Christakos, 2010).

Se construirá la curva TTT del índice IE, y luego se cotejará con las curvas TTT correspondientes a cada modelo de distribución. Para obtener las curvas TTT de cada modelo de distribución, se simularán observaciones de cada modelo con los estimadores máximo verosímiles obtenidos para cada distribución: gamma, Weibull, lognormal, exponencial, Birnbaum-Saunders, Gaussiana inversa, los modelos BSG (Birnbaum-Saunders Generalizada) y los modelos IGT (Inversa Gaussiana Tipo).

Capítulo 4

Aplicación a Trabajadores Rurales de la Provincia de Córdoba

En este capítulo se presentan los resultados del ajuste para el Índice de intensidad a la exposición a los plaguicidas que afecta a los agroaplicadores de la Provincia de Córdoba, cuya construcción se expuso en el capítulo 2. En la primera parte de esta sección, se presenta la selección de la distribución que ajusta a los datos utilizando la metodología presentada en el capítulo 3, a partir de la cual se construye una escala de riesgo y finalmente, se analizan las características de los grupos determinados en función de la escala lograda.

4.1 Metodología para el ajuste

En este punto se muestran los resultados de las estrategias de ajuste definidas en el apartado 3.5.

4.1.1. Análisis Gráfico

En primer lugar, se realiza un ajuste gráfico de la distribución empírica del índice a las distribuciones de vida del capítulo 3 que se resumen en las Figuras 13, 14 y 15.

En la Figura 13, se representa el (a) histograma de la distribución empírica del Índice de Intensidad con los ajustes de las distribuciones y la (c) función de distribución empírica del Índice de Intensidad y de las distribuciones teóricas; se complementan con los (b)

gráficos Q-Q plot, que representan los cuantiles teóricos *versus* los cuantiles empíricos, y (d) P-P plot, el cual representa las probabilidades teóricas *versus* las probabilidades empíricas. El Q-Q plot pone énfasis en la falta de ajuste en las colas de la distribución mientras que el P-P plot enfatiza la falta de ajuste en el centro de la distribución.

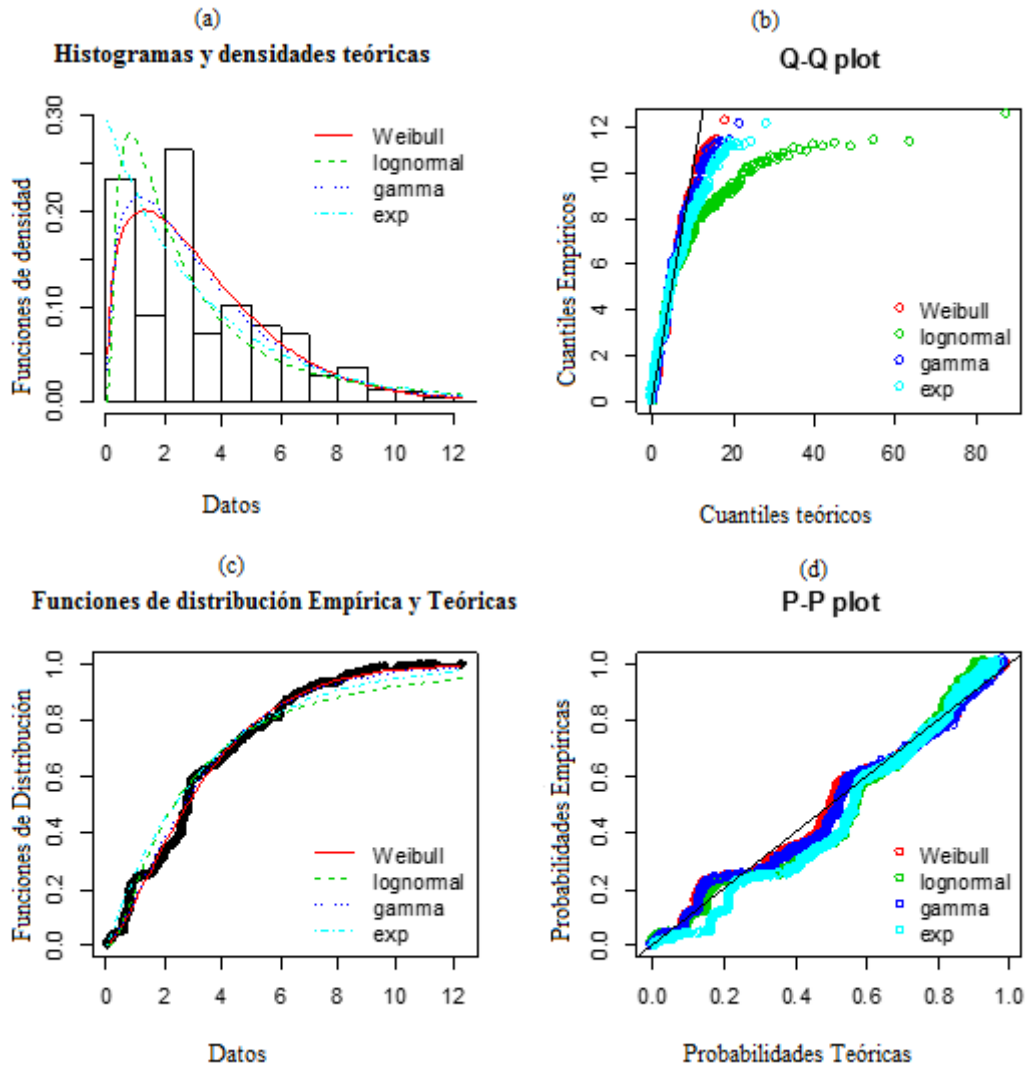


Figura 13: Análisis Gráfico de los datos del Índice de Intensidad *versus* distribuciones tradicionales

Fuente: Construcción propia con el paquete de lenguaje R “fitdistrplus”.

Se observa así que ninguna de las cuatro distribuciones propuestas describe correctamente el centro de la distribución de los datos, no obstante se prefieren las distribuciones

Weibull y Gamma, por su mejor descripción de la cola derecha de la distribución empírica, especialmente porque esta tendencia hacia la derecha es relevante en el contexto de evaluación de riesgos.

En la Figura 14 se presentan los gráficos Q-Q plots de las nuevas distribuciones, Birnbaum-Saunders generalizadas generadas desde los kernels Normal, Laplace, Logística y t de Student; y en la Figura 15 los Q-Q plot de las nuevas distribuciones inversas gaussianas tipo, generadas desde los kernels normal, Laplace, Logística y t de Student¹³.

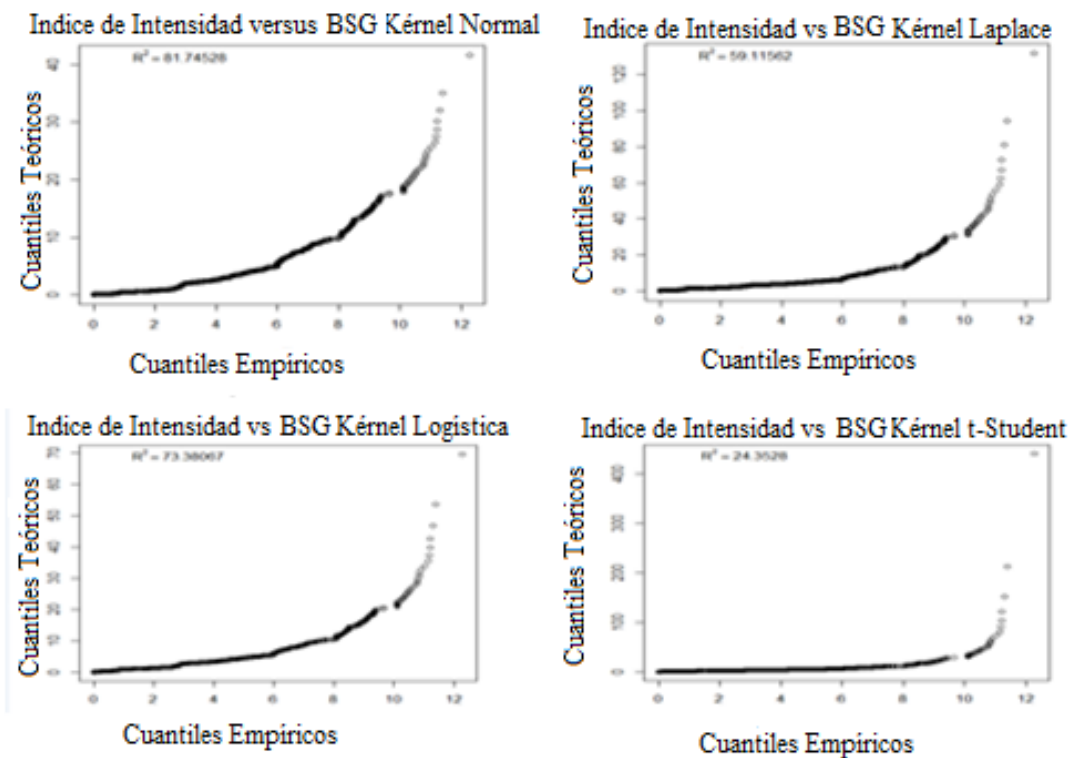


Figura 14: Q-Q Plots de los datos del Índice de Intensidad versus distribuciones BSG
Fuente: Construcción propia con el paquete de lenguaje R “gbs”.

¹³ Los gráficos de la Figura 14 y 15 fueron realizados con el paquete R “gbs” Version 1.0. (Barros, Leiva, & Paula, 2008) y los de la Figura 15 se realizaron con el paquete R Package “ig” Version 1.2 (Leiva, Hernández, & Sanhuenza, Paquete de lenguaje R “ig”, 2008). Estos paquetes solo tienen implementado la gráfica de QQ plot.

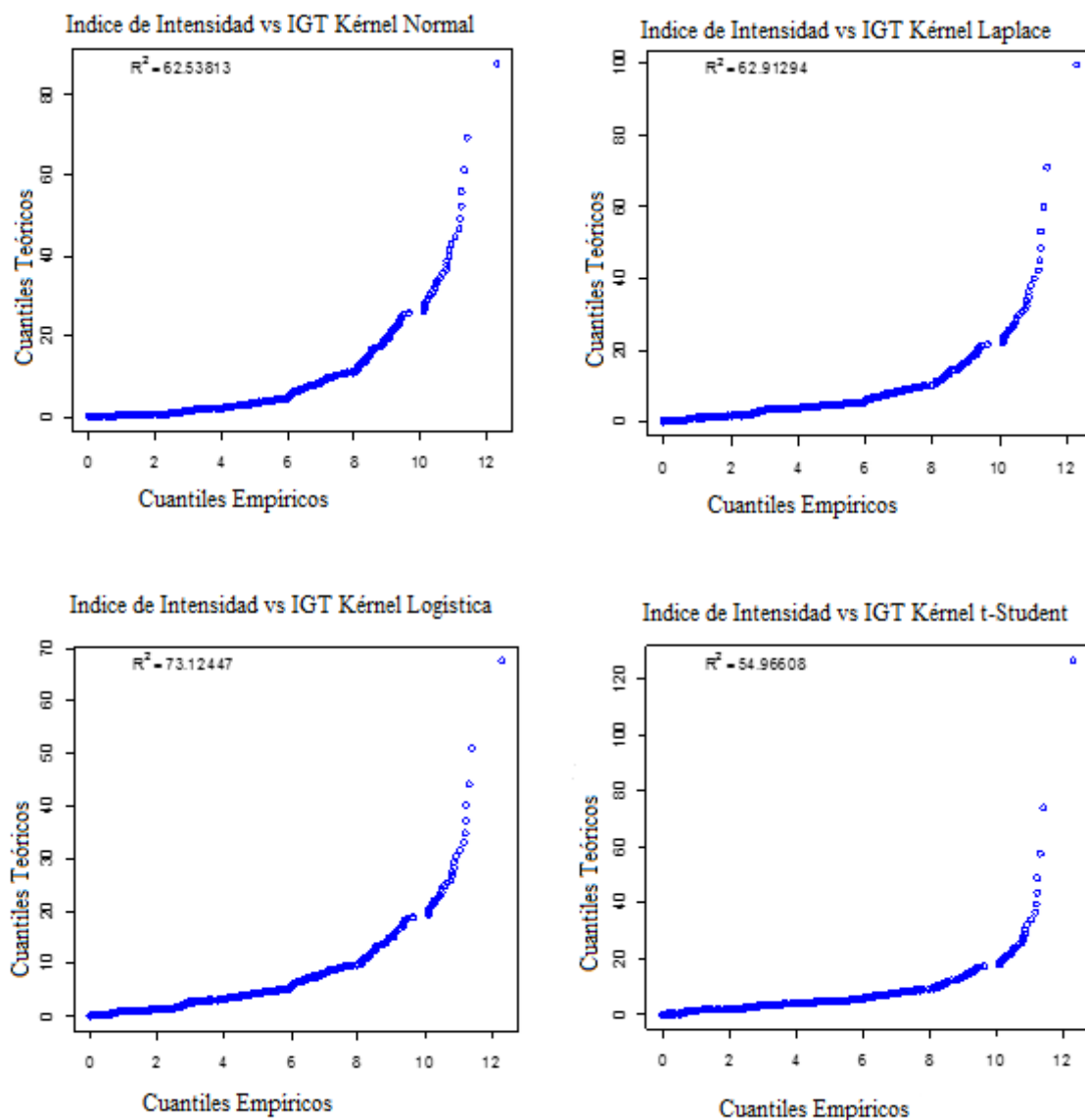


Figura 15: Q-Q Plots de los datos del Índice de Intensidad versus distribuciones IGT
Fuente: Construcción propia con el paquete de lenguaje R “ig”.

Observando los gráficos de la Figura 14 y 15 se concluye que ninguna de las BSG, ni de las IGT resultan en ajustes satisfactorios para las distribuciones empíricas de los datos.

4.1.2 Criterios de Información

En segundo lugar, se calcularon los criterios de información HQIC, AIC y BIC¹⁴ para todas las distribuciones tanto las tradicionales (Exponencial, Lognormal, Weibull, Inversa Gaussiana, Birnbaum-Saunders, Gamma) como las nuevas (GBS e IGT). Los resultados (Cuadro 6) muestran que el mejor ajuste para los datos del índice de intensidad a la exposición lo brinda la distribución Weibull (HQIC= 2,161, AIC= 2,160; BIC= 2,163), seguida por la distribución Gamma (HQIC= 2,173, AIC= 2,172; BIC= 2,175).

Cuadro 6: Criterios de Información para las distribuciones de vida más conocidas

Distribución	Gamma	Exponencial	Lognormal	Gaussiana Inversa	Birbaum Saunders	Weibull
$l(\theta)$	-4592,070	-4686,990	-4869,527	-5691,144	-5329,858	-4566,571
p	2	1	2	2	2	2
n	2115	2115	2115	2115	2115	2115
HQIC	2,173	2,217	2,304	2,693	2,522	2,161
AIC	2,172	2,217	2,303	2,692	2,521	2,160
BIC	2,175	2,218	2,306	2,695	2,524	2,163

El cuadro 7 muestra los resultados luego de utilizar la distribución BSG obtenida desde kernel t-Student (HQIC= 2,254, AIC= 2,252; BIC= 2,256), superando en el ajuste a las distribuciones clásicas Lognormal, Gaussiana Inversa y Birnbaum-Saunders. Lo mismo ocurre con la IGT generada con kernel t-Student (HQIC= 2,197, AIC= 2,196; BIC= 2,199, ver Cuadro 8) que considerando las nuevas distribuciones IGT, logra un mejor ajuste que

¹⁴ Ver detalle de criterios de información en punto 3.5.

las clásicas distribuciones Exponencial, Lognormal, Gaussiana Inversa y Birnbaum-Saunders.

Cuadro 7: Criterio de Información para las nuevas distribuciones BSG

Distribución	BSG kernel Normal	BSG kernel Laplace	BSG kernel Logística	BSG kernel t-Student
$l(\theta)$	-5329,858	-4857,106	-4942,868	-4760,090
p	2	2	2	3
n	2115	2115	2115	2115
HQIC	2,522	2,298	2,339	2,254
AIC	2,521	2,297	2,338	2,252
BIC	2,524	2,301	2,341	2,256

Cuadro 8: Criterio de Información para las nuevas distribuciones IGT

Distribución	IGT kernel Normal	IGT kernel Laplace	IGT kernel Logística	IGT kernel t-Student
$l(\theta)$	-5691,144	-4826,096	-4981,707	-4640,726
p	2	2	2	3
n	2115	2115	2115	2115
HQIC	2,693	2,284	2,357	2,197
AIC	2,692	2,283	2,356	2,196
BIC	2,695	2,286	2,359	2,199

De la comparación de todos los ajustes, considerando tanto las Figuras presentadas como los valores obtenidos para los criterios de información, se considerarán únicamente las distribuciones Weibull y Gamma, por su buen ajuste. Dado que estas dos distribuciones tienen la misma cantidad de parámetros y poseen parámetros de forma, para definir la elección entre ellas se utilizaron los estadísticos Cramér-von Mises y Anderson-Darling que son muy utilizados en el caso de las distribuciones asimétricas con parámetros de forma (Ver capítulo 3 punto 3.3.1). Al dar más peso a las colas de distribución, el

estadístico de Anderson-Darling es de especial interés cuando importan las colas de una distribución como sucede en la evaluación de riesgos (Delignette-Muller & Dutang, 2015).

El estadístico de Cramer-von Mises, no considera la complejidad del modelo (es decir, el número de parámetros) lo cual no es un problema en este caso ya que se comparan dos distribuciones que tienen la misma cantidad de parámetros como Weibull y Gamma (Delignette-Muller & Dutang, 2015).

Finalmente, la distribución que mejor describe el comportamiento de los datos del Índice de Intensidad a la Exposición es una distribución Weibull (con parámetro de escala mayor que uno) ya que la misma presenta los menores valores de los estadísticos Cramer-von Mises y Anderson-Darling (Cuadro 9).

Cuadro 9: Estadísticos Cramer-von Mises y Anderson-Darling

Estadísticos	Distribuciones	
	Gamma	Weibull
Cramer-von Mises	3,0780	2,3304
Anderson-Darling	17,2081	13,8067

4.1.3 Validación de la distribución elegida con los gráficos TTT

Como se mencionó en el punto 3.5.2 es muy importante tener en cuenta la tasa de riesgo para seleccionar o descartar distribuciones de vida y los gráficos TTT constituyen una herramienta para determinar la forma de la tasa de riesgo. En las Figuras 16 a 19 se presentan las representaciones del gráfico TTT de los datos del índice de intensidad a la exposición en contraste con los gráficos TTT de las distribuciones de vida clásicas, de los

nuevos modelos GBS e IGT, respectivamente.

Los TTT Plots de las distribuciones clásicas Lognormal, Inversa Gaussiana, Exponencial, Gamma, Weibull y Birnbaum-Saunders (Figura 16); se construyeron a partir de los parámetros que surgieron de ajustar los datos del índice de Intensidad con estas distribuciones.

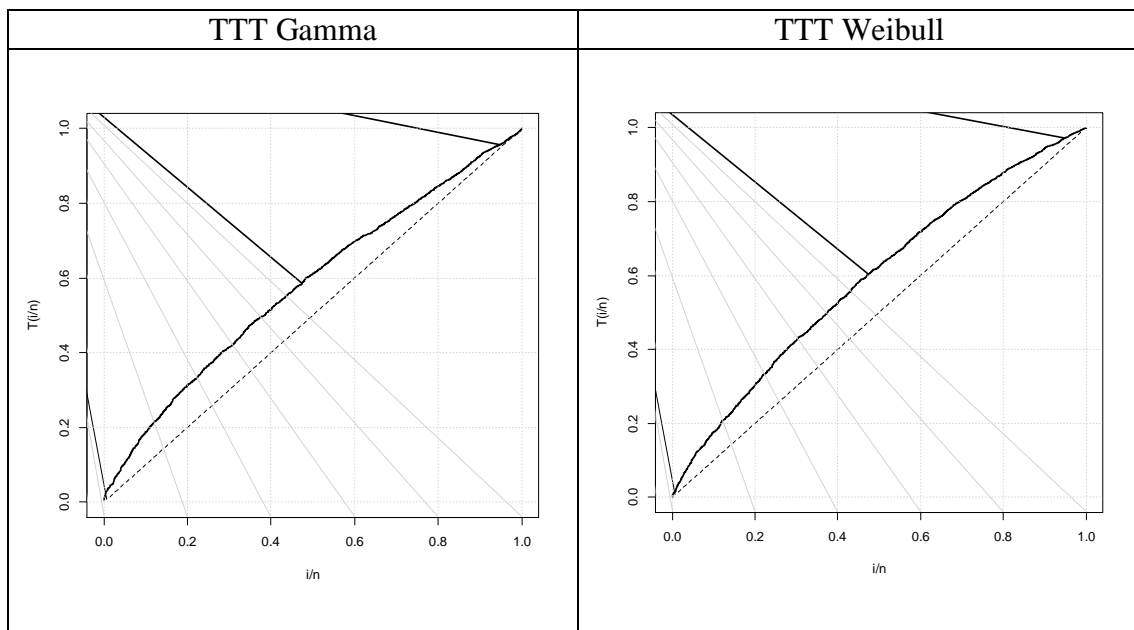


Figura 16: TTT Plots de las distribuciones clásicas

Fuente: Construcción propia con el paquete de lenguaje R “AdequacyModel”.

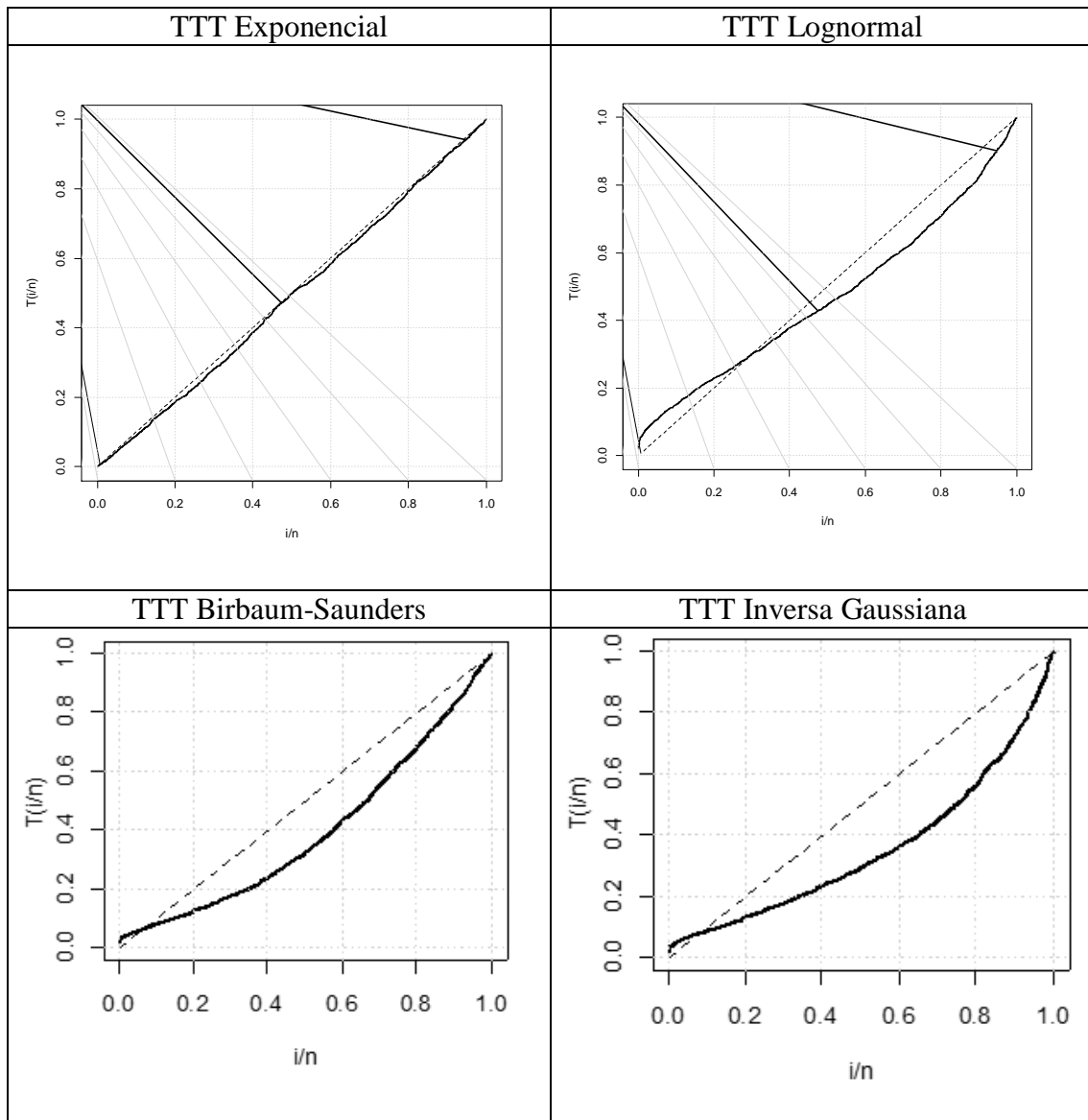


Figura 16: TTT Plots de las distribuciones clásicas (continuación)

Fuente: Construcción propia con el paquete de lenguaje R “AdequacyModel”.

Como se mencionó en el punto 3.5.2 la distribución Exponencial presenta una tasa de riesgo constante. En las distribuciones Gamma y Weibull la tasa de riesgo es creciente. La distribución Lognormal presenta una tasa de riesgo de bañera inversa (aumenta hasta su punto de cambio y luego disminuye). En las distribuciones Gaussiana Inversa y Birnbaum-Saunders las tasas de riesgo son prácticamente decrecientes, presentando al inicio un tramo muy reducido de tasa de riesgo creciente (Figura 16).

Para los TTT Plots de las nuevas distribuciones se procedió de la misma manera que con las distribuciones clásicas, con la salvedad de que las estimaciones de los parámetros se hicieron, en el caso de las distribuciones GBS con el paquete R “gbs” (Barros, Leiva, & Paula, 2008); para las distribuciones IGT las estimaciones de los parámetros se hicieron con el paquete R “ig” (Leiva, Hernández, & Sanhuenza, Paquete de lenguaje R “ig”, 2008).

La tasa de riesgo para las distribuciones BSG Kernel Normal, Logística (Figura 17 a y c) e IGT Kernel Normal (Figura 18 a) tienen un comportamiento muy similar a las de la Gaussiana Inversa y Birnbaum-Saunders. Un comportamiento muy diferente presenta la tasa de riesgo de la distribución BSG Kernel Laplace que es predominantemente creciente (Figura 17 b), y la BSG kernel t de Student que es totalmente decreciente (Figura 17 d).

Las distribuciones IGT kernel Laplace e IGT kernel t de Student presentan tasas de riesgo de bañera invertida similar al observado en la Lognormal, predominando en la de kernel Laplace el tramo decreciente y en la de kernel t de Student el tramo creciente (Figura 18 b y d). La tasa de riesgo de la IGT de kernel Logística es creciente en todo su trayecto con cambios en su comportamiento (Figura 18 c).

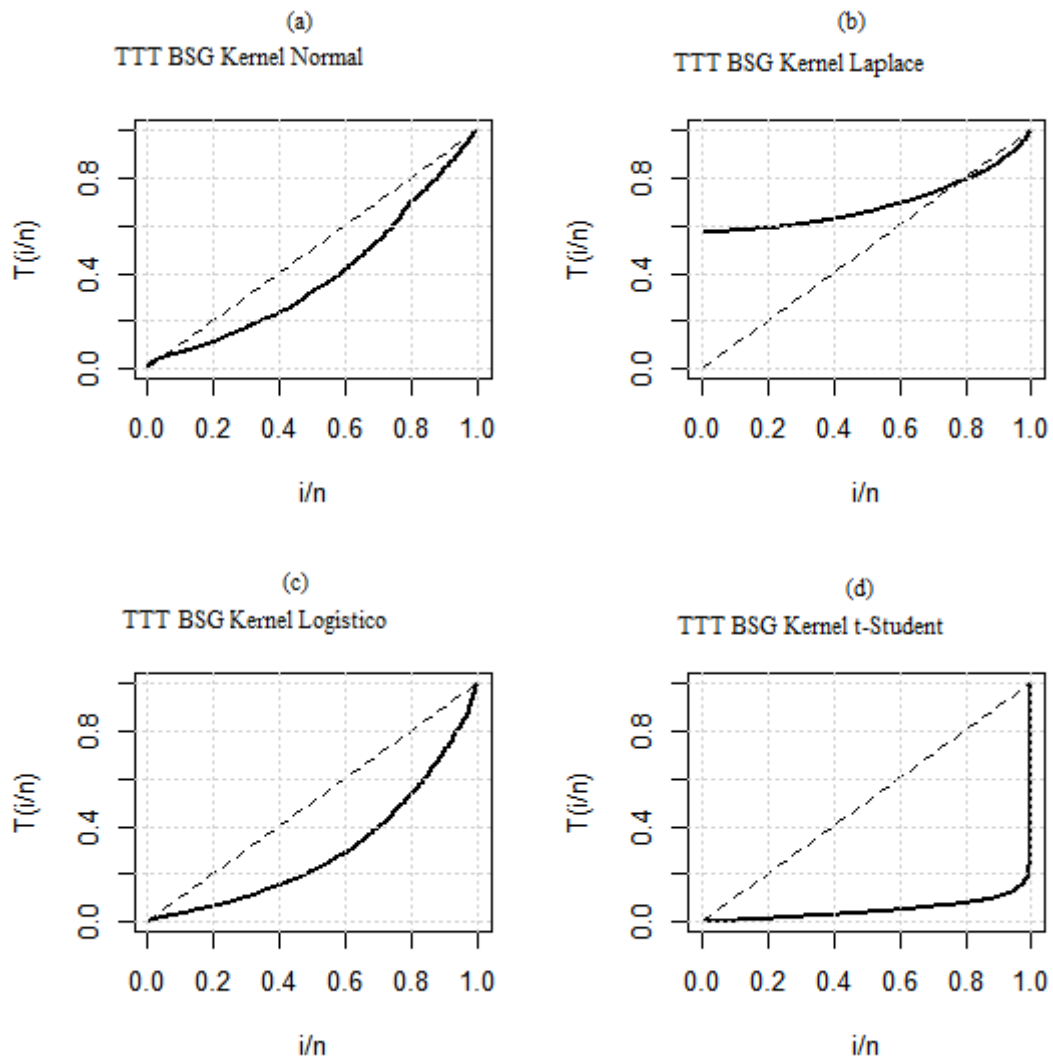


Figura 17: TTT Plots de las nuevas distribuciones BSG
Fuente: Construcción propia con el paquete de lenguaje R “AdequacyModel”.

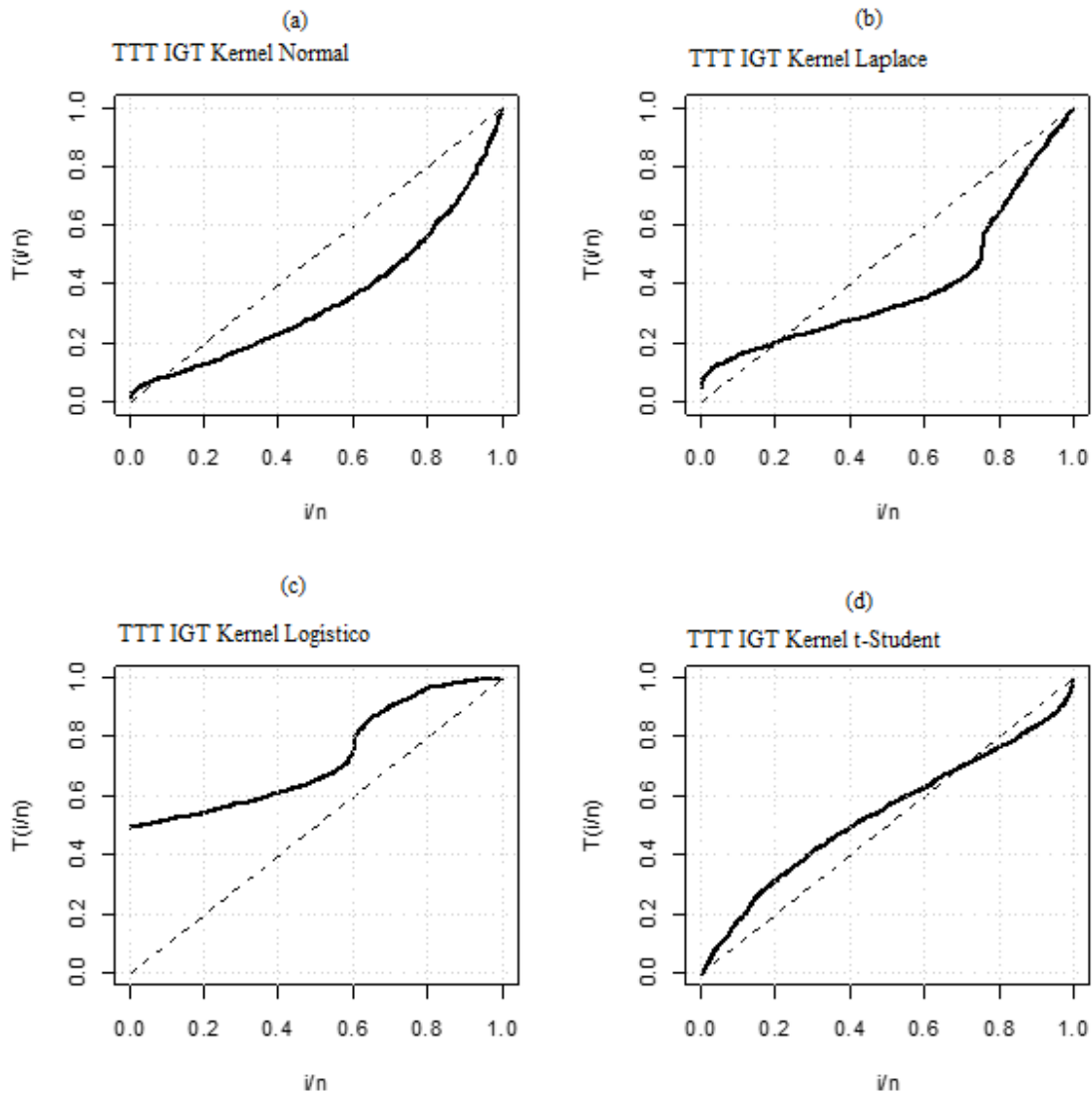


Figura 18: TTT Plots de las nuevas distribuciones IGT
Fuente: Construcción propia con el paquete de lenguaje R “AdequacyModel”

En la Figura 19 se observa que el comportamiento de la tasa de riesgo de los datos es creciente, lo cual valida la elección del modelo Weibull, que se caracteriza por tener una tasa de riesgo creciente cuando su parámetro de escala es mayor a uno. en la estimación inicial del IE el parámetro de escala es 3,66, lo que confirma la elección de este modelo. Esto indica que un mayor valor en el índice de exposición a los plaguicidas implica una mayor propensión al riesgo de exposición.

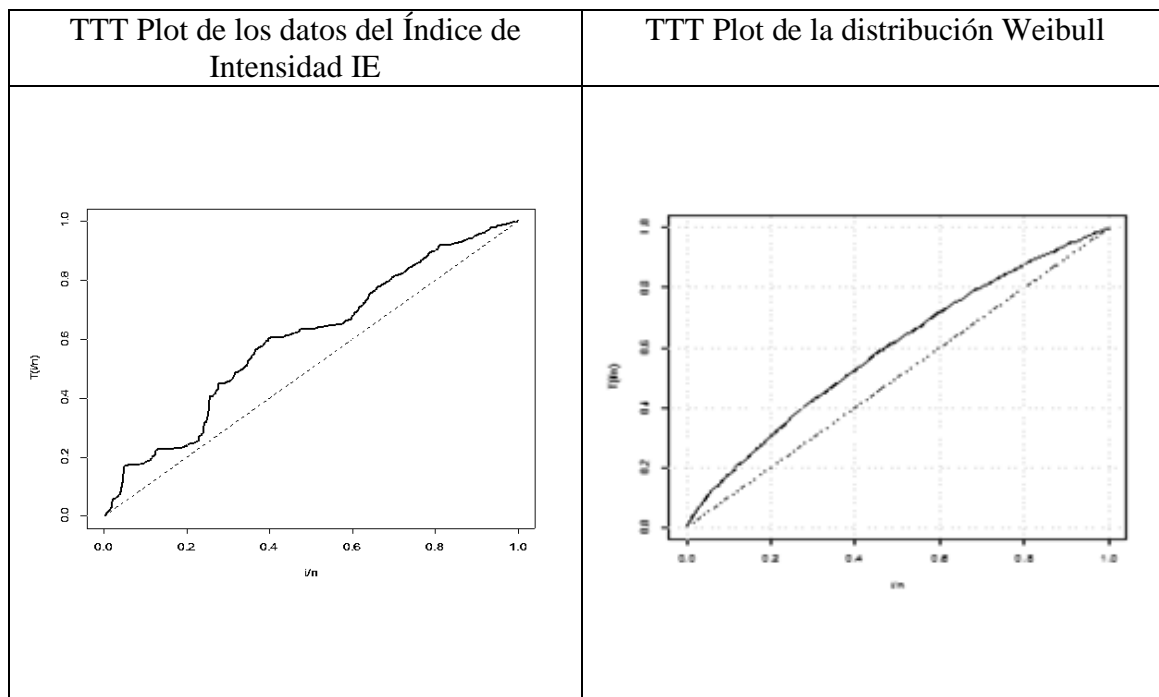


Figura 19: Comparación del TTT Plot de los datos del Índice de Intensidad con el TTT Plot de la Weibull

Fuente: Construcción propia con el paquete de lenguaje R “Adequacy Model”.

4.2. Construcción de una Escala Cualitativa

Para construir una escala cualitativa de niveles de riesgo se estimarán los percentiles del índice IE a partir del modelo Weibull seleccionado. Con este fin, se seleccionaron a 30000 muestras bootstrap del Índice de exposición a los plaguicidas calculados sobre el grupo de agroaplicadores, las que se ajustaron con la distribución Weibull. En cada ajuste se determinaron los parámetros de forma y escala y los percentiles de la distribución.

En el cuadro 10 se presentan el promedio de las estimaciones de los parámetros de forma (1,336) y escala (3,66) ajustados para las 30.000 muestras, sus desviaciones estándar, y sus intervalos de confianza. Ambos parámetros presentan muy baja dispersión. Los límites del intervalo de confianza del parámetro de escala [3,537; 3,783] se corresponden con una tasa de riesgo creciente.

Cuadro 10: Estimación de los parámetros de la distribución Weibull

Parámetros de la distribución Weibull	Estimadores	Desviación Estándar	Límite Inferior	Límite Superior	Amplitud del intervalo
Forma (b)	1,336	0,024	1,29	1,382	0,092
Escala (c)	3,66	0,063	3,537	3,783	0,246

En el cuadro 11 se detallan los percentiles promedio estimados y su desviación estándar, a partir de los cuales se construyeron los intervalos de confianza de la siguiente manera:

$$LIC = \hat{p} - Z_{\frac{\alpha}{2}} \sigma / \sqrt{n} \quad \text{y} \quad LSC = \hat{p} + Z_{1-\frac{\alpha}{2}} \sigma / \sqrt{n}$$

donde Z corresponde a un nivel de confianza del 95%, \hat{p} es el estimador del percentil, σ / \sqrt{n} es el error estándar del estimador. La menor amplitud de los intervalos de confianza ocurre entre los percentiles 10 y 20 y entre los percentiles 40 y 55. Mientras que el intervalo del percentil 25% y los calculados a partir del percentil 60, son menos precisos.

Cuadro 11: Percentiles estimados con medidas resumen (media, desviación estándar e intervalos de confianza)

Percentiles	Media	Desviación Estándar	Límite Inferior	Límite Superior	Amplitud del intervalo
1%	0,06	0,012	0,033	0,079	0,046
5%	0,55	0,087	0,376	0,715	0,339
10%	0,64	0,016	0,612	0,674	0,062
15%	0,81	0,002	0,806	0,814	0,008
20%	0,86	0,022	0,821	0,909	0,088
25%	1,36	0,218	0,93	1,786	0,856
30%	1,85	0,051	1,749	1,948	0,199
35%	2,2	0,105	1,992	2,404	0,412
40%	2,57	0,051	2,474	2,673	0,199
45%	2,68	0,031	2,62	2,742	0,122
50%	2,81	0,009	2,794	2,831	0,037
55%	2,88	0,02	2,844	2,924	0,08
60%	3,15	0,131	2,896	3,41	0,514
65%	3,88	0,141	3,6	4,152	0,552
70%	4,39	0,097	4,204	4,582	0,378
75%	4,91	0,112	4,691	5,131	0,44
80%	5,68	0,15	5,389	5,977	0,588
85%	6,1	0,052	5,997	6,2	0,203
90%	6,84	0,141	6,559	7,113	0,554

Cuadro 11: Percentiles estimados con medidas resumen (media, desviación estándar e intervalos de confianza) Continuación.

Percentiles	Media	Desviación Estándar	Límite Inferior	Límite Superior	Amplitud del intervalo
95%	8,24	0,106	8,032	8,449	0,417
99%	10,29	0,225	9,848	10,732	0,884

A continuación, en el Cuadro 12 se comparan los percentiles de la distribución empírica con los de la distribución estimada, en los que se observa que la estimación es buena en todo el tramo de la distribución, sólo a partir del percentil 99 se sobreestima el riesgo, lo cual no es relevante ya que corresponde a los agroaplicadores de mayor riesgo.

Cuadro 12: Comparación de los percentiles de la distribución empírica y de la distribución de ajuste

Percentiles	Distribución empírica	Distribución ajustada (Weibull)	Niveles de Riesgo
1%	0,05	0,06	Bajo
5%	0,6	0,55	
10%	0,64	0,64	
15%	0,81	0,81	
20%	0,86	0,86	
25%	1,3	1,36	
30%	1,83	1,85	Medio
35%	2,21	2,2	
40%	2,6	2,57	
45%	2,68	2,68	
50%	2,81	2,81	
55%	2,89	2,88	
60%	3,14	3,15	
65%	3,89	3,88	
70%	4,43	4,39	
75%	4,9	4,91	
80%	5,66	5,68	Alto
85%	6,1	6,1	
90%	6,86	6,84	
95%	8,29	8,24	
99%	10,29	10,29	

Se definieron los puntos de cortes en dicha tabla de percentiles a los fines de conformar la escala del Índice de Intensidad a la Exposición, a partir del estudio de las distribuciones. Con dichos puntos se definieron los grupos de exposición que clasificarán a los agroaplicadores en baja, media y alta exposición (riesgo bajo, medio o alto, respectivamente), establecido por el equipo de expertos de GEACC (Grupo de Epidemiología Ambiental del Cáncer en Córdoba, Facultad de Ciencias Médicas, UNC). De esta manera los cortes se definen en los percentiles estimados 25% y 70%, con los valores del índice 1,36 y 4,39; respectivamente. Finalmente, las categorías obtenidas son hasta el valor del índice de 1,36, BAJO RIESGO, desde este valor hasta 4,39, RIESGO MEDIO y a partir de 4,39, ALTO RIESGO.

4.3 Caracterización de los grupos de riesgo

A partir de la definición de los grupos de riesgo de agroaplicadores (Bajo, Medio y Alto), se analizaron diversas variables sociodemográficas, de estado y valoración de la salud, así como también sobre las características o modalidad de la tarea que llevan a cabo como agroaplicadores, con el fin de identificar, en el caso de que existan, diferencias entre los grupos de exposición. Se llevó a cabo un análisis descriptivo exhaustivo, obteniendo como principales resultados que el promedio de edad de los agroaplicadores es igual a 36 (D.E. 11, 17) años con una dispersión relativa similar en todos los grupos (C.V. Bajo: 0,3066, C.V. Medio: 0,3052, C.V. Alto: 0,3091). En el Cuadro 13 se observa que en el grupo de bajo riesgo se encuentran los agroaplicadores de menor edad (Mínimo 14 años) mientras que los agroaplicadores de mayor edad se presentan en los grupos de riesgo

medio y alto (Máximo 88 y 87 años respectivamente). En relación al nivel de educación, en todos los grupos, la mayor proporción de trabajadores cursó hasta el nivel secundario completo o incompleto y respecto al estado civil predominan los trabajadores casados o en concubinato.

Cuadro 13: Estadística descriptiva de variables sociodemográficas

Estadística Descriptiva IE		Índice de Intensidad a la Exposición		
		Bajo	Medio	Alto
Socio Demografía				
Edad (años)	Media	36,66	36,27	36,36
	DS	11,24	11,07	11,24
	Mínimo	14	17	16
	Máximo	76	88	87
Nivel de Educación ^a	Primario completo o Incompleto	40,08	40,04	42,29
	Secundario completo o incompleto	42,41	47,72	44,67
	Terc.-Univ. completo o incompleto	17,51	12,24	13,04
Estado Civil ^a	Soltero, separado o viudo	44,74	43,76	45,96
	Casado o concubino	55,26	56,24	54,04

^a Porcentaje de las categorías dentro de cada grupo.

Con respecto a la estadística descriptiva de las variables de salud (cuadro 14), en los tres grupos de riesgo predominan los trabajadores con dolor de cabeza o vértigo, los que nunca o raramente fueron hospitalizados como así también los que presentan con frecuencia o casi siempre signos de irritación.

En el grupo de Riesgo Medio predominan los trabajadores que respondieron que nunca o raramente tienen cansancio excesivo o se sienten nerviosos o deprimidos.

En los grupos de Bajo y Alto Riesgo hay mayor proporción de trabajadores que presentan mayor frecuencia (A veces, con frecuencia, casi siempre) de cansancio excesivo, o que se sienten nerviosos o deprimidos, siendo en ambos casos un poco mayor la proporción en el grupo alto.

Cuadro 14: Estadística descriptiva de variables de salud

Variable	Categorías	Bajo	Medio	Alto
Tiene dolores de cabeza o vértigos ^a	Nunca o raramente	46,62	49,47	42,33
	A veces, con frecuencia, casi siempre	53,38	50,53	57,67
Está cansado excesivamente ^a	Nunca o raramente	47,37	52,67	42,33
	A veces, con frecuencia, casi siempre	52,63	47,33	57,67
Se siente nervioso o deprimido ^a	Nunca o raramente	48,31	53,73	43,26
	A veces, con frecuencia, casi siempre	51,69	46,27	56,74
Fue hospitalizado ^a	Nunca o raramente	75,56	79,85	69,61
	A veces, con frecuencia, casi siempre	24,44	20,15	30,39
Tiene signos de irritación ^a	Nunca o raramente	39,29	44,03	32,56
	A veces, con frecuencia, casi siempre	60,71	55,97	67,44

^a Porcentaje de las categorías dentro de cada grupo.

En la estadística descriptiva de las variables de modalidad de la tarea (cuadro 15), prevalecen en todos los grupos: los que utilizan receta del ingeniero agrónomo con la indicación de la dosis de plaguicida, los que no inyectan animales y los que mezclaron, aplicaron o estuvieron presentes cuando se manipulaban plaguicidas en un periodo menor o igual a 10 años.

Cuadro 15: Estadística descriptiva de variables de modalidad de la tarea

Variable	Categorías	Bajo	Medio	Alto
Años que mezcló, aplicó o estuvo presente ^a	Menor o igual a 10	72,18	72,81	66,82
	Más de 10 años	27,82	27,19	33,18
Inyecta animales ^a	Sí	9,21	15,61	18,59
	No	90,79	84,39	81,41
Utiliza receta de ingeniero agrónomo ^a	Sí	83,96	87,88	82,16
	No	16,04	12,13	17,84

^a Porcentaje de las categorías dentro de cada grupo.

A partir de la estimación de un modelo de regresión logística multinomial se compararon los grupos tomando como referencia el grupo de riesgo Alto. A partir del modelo con todas las variables detalladas en los cuadros 16,17 y 18, se eliminaron en una primera instancia las variables con nivel de significación superior al 25% (edad, dolor de cabeza, hospital, nervioso o depresión y cansancio). Se analizaron modelos alternativos

comparando el cambio en la verosimilitud, los criterios BIC y AIC. El modelo seleccionado incluye las variables Años que mezcló, aplicó o estuvo presente cuando se manipulaban plaguicidas, Inyecta animales, Utiliza receta de ingeniero agrónomo y presenta signos de irritación, cuya expresión se presenta a continuación,

$$\ln\left(\frac{P_{ij}}{P_{i3}}\right) = \beta_{0j} + \beta_{1j}\text{años mezcla} + \beta_{2j}\text{inyecta} + \beta_{3j}\text{receta ingeniero} + \beta_{4j}\text{irritación},$$

para $j = 1, 2,$

donde,

j niveles de riesgo de intensidad a la exposición a plaguicidas (1. Bajo riesgo, 2. Riesgo medio, 3. Alto riesgo),

añosmezcla: Años que mezcló, aplicó o estuvo presente (0. Más de 10 años
1. Menos de 10 años),

inyecta: Inyecta animales (0. Sí inyecta animales, 1. No inyecta animales),

recetaingeniero: Utiliza receta de ingeniero agrónomo (0. Sí utiliza receta de ingeniero agrónomo; 1. No utiliza receta de ingeniero agrónomo),

irritación: Tiene signos de irritación (0. A veces, con frecuencia, casi siempre; 1. Nunca o raramente).

En el cuadro 16 se presentan los resultados del modelo.

Cuadro 16: Coeficientes estimados del modelo

Niveles de Riesgo	Variable	Exp(β)	Desviación Estándar	Significación	Intervalo de Confianza 95% para Exp(β)	
					Límite Inferior	Límite Superior
Bajo Riesgo	añosmezcla	1,2355	0,1857	0,159	0,9202	1,6590
	inyecta	2,2652	0,4923	0,000	1,4794	3,4684
	recetaingeniero	0,8976	0,1672	0,562	0,6230	1,2933
	irritación	1,2480	0,1744	0,113	0,9488	1,6414
	constante	0,3464	0,0807	0,000	0,2193	0,5471
Medio Riesgo	añosmezcla	1,2521	0,1634	0,085	0,9694	1,6172
	inyecta	1,3738	0,2273	0,055	0,9932	1,9002
	recetaingeniero	0,6979	0,1178	0,033	0,5012	0,9717
	irritación	1,4543	0,1774	0,002	1,1449	1,8472
	constante	0,9374	0,1678	0,720	0,6586	1,3341

Cuadro 17: Coeficientes estimados de los modelos Inicial y Final

Modelo	Observaciones	Logverosimilitud	gl	AIC	BIC
Inicial	1.649	-1.734,04	16	3.500,08	3.586,60
Final	1.651	-1.739,69	10	3.499,38	3.553,47

De los resultados presentados en el cuadro 16 se concluye que:

- Para la variable “Años que mezcló, aplicó o estuvo presente mientras se manipuló plaguicidas” los que aplican menos de 10 años tienen una chance 1,25 veces mayor (*p-value*: 0,085) de estar en el grupo de riesgo medio que en el grupo de alto riesgo, que los que aplicaron más de 10 años.

- Los aplicadores que no inyectan animales tienen una chance 2,26 veces de pertenecer (*p-value*: 0,000) al grupo de bajo riesgo que al de riesgo alto que los que inyectan animales. Esta chance disminuye a 1,37 veces en el grupo de riesgo medio respecto al de riesgo alto (*p-value*: 0,055).
- Los aplicadores que aplican sin receta de ingeniero agrónomo tienen una chance de un 30,21% menor (*p-value*: 0,033) de estar en el grupo de riesgo medio que en el de alto riesgo, que los que tienen receta.
- Los aplicadores que nunca o raramente tengan signos de irritación tienen una chance 1,45 veces mayor (*p-value*: 0,002) de estar en el grupo de riesgo medio que en el de riesgo alto, que los que presentan signos irritativos frecuentemente.

En resumen, los factores que resultan significativos para determinar el riesgo de los aplicadores fueron: no tener receta de ingeniero agrónomo, inyectar animales, tener una antigüedad mayor a 10 años en las tareas de mezclar, aplicar o estar presente mientras se manipulan plaguicidas y la presencia de signos de irritación (Cuadro 18).

Cuadro 18: Resumen de variables significativas del Modelo Multinomial

Modalidad de la tarea	Índice de Intensidad a la Exposición		
	BAJO	MEDIO	ALTO
Menos de 10 años que mezcló, aplicó o estuvo presente	1,2355 [0,9202; 1,6590]	1,2521** [0,9694; 1,6172]	1
No Inyecta animales	2,2652* [1,4794; 3,4684]	1,3738** [0,9932; 1,9002]	1
No Utiliza receta de ingeniero agrónomo	0,8976 [0,6230; 1,2933]	0,6979* [0,5012; 0,9717]	1
Nunca o raramente tiene signos de irritación	1,2480 [0,9488; 1,6414]	1,4543* [1,1449; 1,8472]	1

* Diferencias significativas al 5%.

** Diferencias significativas al 10 %.

Capítulo 5

Discusión y Conclusiones

Esta tesis tuvo como principal propósito encontrar una distribución teórica que ajuste el índice Nivel de Intensidad a la Exposición (IE) diseñado para medir la exposición a plaguicidas en trabajadores rurales que, como otros indicadores definidos en este contexto (Dosemeci *et al*, 2002) se calcula como una combinación lineal de variables observadas. La distribución estadística de estas combinaciones lineales de variables aleatorias, no son sencillas de modelar y la literatura estadística sobre este tema resulta escasa.

Aunque no se encontraron trabajos que ajusten la distribución de índices iguales o similares al que se presenta en esta tesis, se tomaron como antecedentes algunos estudios sobre contaminación ambiental. Algunos de los trabajos considerados miden la concentración de sustancias en el aire (Leiva, Barros, Paula, & Sanhueza, 2008; Leiva, Vilca, Balakrishnan, & Sanhueza, 2010; Ferreira, Gomes, & Leiva, 2012; Leiva, Ferreira, Gomes, & Lillo, 2015); otros miden el oxígeno disuelto en el agua (Vilca, Sanhuenza, Leiva, & Christakos, 2010). Si bien estos trabajos no son comparables con el estudio del índice IE, se puede destacar que su distribución empírica es similar a la de la concentración de contaminantes en el aire (dióxido de azufre) en Santiago de Chile (Leiva, Barros, Paula, & Sanhuenza, 2008). En este último caso el mejor ajuste se logró con el modelo de la nueva distribución Birnbaum-Saunders Generalizada con kernel Laplace; a diferencia del índice de intensidad a la exposición, que presenta el mejor ajuste con la distribución Weibull.

La distribución empírica del IE presenta asimetría derecha, con alta frecuencia de valores atípicos, de soporte positivo; características que sugieren que este índice se distribuye como una variable de vida. Las variables de vida son continuas, asumen valores positivos y si bien en general representan, tiempos de vida de un organismo biológico o tiempos de duración de un componente electrónico o tiempos de espera hasta que un sistema falle (Marshall & Olkin, 2007); también pueden tener otras aplicaciones como medir la concentración de un contaminante (Leiva, 2010), siendo esta última aplicación la que se consideró de especial interés para el estudio del ajuste adecuado del IE.

En la búsqueda de una distribución adecuada para el índice de intensidad se tuvieron en cuenta los modelos de las variables de vida desde los más tradicionales (exponencial, Weibull, lognormal, gamma, Birnbaum-Saunders y Gaussiana inversa) hasta las nuevas distribuciones (Birnbaum-Saunders Generalizada e Inversa Gaussiana Tipo).

Las nuevas distribuciones GBS e IGT, fueron especialmente consideradas en esta tesis; tanto por los argumentos teóricos que justifican el uso de estas distribuciones en situaciones que impliquen daños acumulativos (Leiva, Barros, Paula, & Sanhuenza, 2008) como también sus características: distribuciones altamente flexibles, que admiten diferentes grados de curtosis y asimetría, cuya estimación de parámetros suelen ser robustas cuando hay datos atípicos.

En este trabajo se propuso una secuencia de análisis para encontrar el mejor ajuste, a partir de un análisis gráfico de la distribución, se ajustaron los datos con las distribuciones recomendadas para este tipo de variables, y con el objeto de seleccionar uno de estos ajustes se utilizaron los criterios de información y una validación final con los gráficos TTT. Esta secuencia propuesta para encontrar la distribución de un índice con las características mencionadas constituye el primer aporte de esta tesis ya que puede ser de utilidad para estudios de índices similares.

En el estudio de la concentración de contaminantes en el aire (dióxido de azufre) en Santiago de Chile (Leiva, Barros, Paula, & Sanhuenza, 2008) el mejor ajuste se obtuvo con el modelo de la nueva distribución Birnbaum-Saunders Generalizada con kernel Laplace. En este trabajo, las distribuciones Gamma y Weibull resultaron un buen ajuste para el IE, sin embargo, como ambas presentan la misma cantidad de parámetros fue necesario compararlas utilizando los estadísticos de Cramér-von Mises y Anderson-Darling que son muy utilizados en asimétricas y no tienen en cuenta la cantidad de parámetros. De esta comparación, el mejor ajuste se logró con la distribución Weibull, aunque en relación con la primera hipótesis planteada en esta tesis, cabe destacar que los nuevos modelos GBS e IGT, con kernel t-Student ajustaron mejor que cuatro de las distribuciones tradicionales Exponencial, Lognormal, Gaussiana Inversa y Birnbaum-Saunders. La distribución Weibull fue finalmente validada analizando el comportamiento de la tasa de riesgo a partir de los gráficos TTT. Otro aporte de esta tesis es una revisión completa y exhaustiva de todas las distribuciones consideradas para el ajuste, presentadas en el capítulo 3.

Una vez definido el mejor ajuste, se estimaron los percentiles utilizando un remuestreo Bootstrap, y a partir de los percentiles estimados se delimitaron tres niveles de riesgo (bajo, medio y alto). En función a estos niveles de riesgo se estudiaron características sociodemográficas y de salud de los trabajadores dentro de cada grupo. Se estimó un modelo de regresión logística multinomial en el que los factores que resultaron más significativos para determinar el riesgo de los aplicadores son: la antigüedad (mayor a 10 años) en las tareas de mezclar, aplicar o estar presente mientras se manipulan plaguicidas, no tener receta de ingeniero agrónomo, inyectar animales y la presencia de signos de irritación. Este último análisis permite realizar estudios predictivos ya que, dada las características de un nuevo trabajador, se puede determinar a qué grupo de riesgo

pertenece y adoptar las medidas preventivas necesarias. Este aporte contribuye a la toma de conciencia del peligro al que están expuestos estos trabajadores, y de esta manera se promueve la implementación de políticas públicas de prevención.

Cuando surgieron las nuevas distribuciones GBS e IGT hubo un desarrollo de software estadístico, en 2007 se crea el paquete para la distribución Birbaum-Saunders en lenguaje R (Leiva, Hernández, & Riquelme, 2007) luego en 2008 surgen los paquetes en lenguaje R para las distribuciones GBS (Barros, Leiva, & Paula, 2008) e IGT (Leiva, Hernández, & Sanhuenza, 2008). Estos tres paquetes son los únicos disponibles para ajustar estas distribuciones nuevas y nunca fueron actualizados para poder emplearlos con las nuevas versiones del lenguaje R, lo que hizo necesario recrear entornos de versiones anteriores del aplicativo para poder ejecutarlas. Esto insumió un tiempo y esfuerzo adicional en el análisis.

La falta de actualización y disponibilidad de estas rutinas resulta una limitación en su aplicación y que el código no haya sido testeado debidamente. En este sentido es posible indicar la actualización o la integración de estas distribuciones en paquetes más accesibles como una posible línea futura de investigación.

El estudio de un índice de nivel de intensidad a la exposición de plaguicidas para los agroaplicadores, y la estimación de los percentiles que permiten determinar niveles de riesgo de exposición es un trabajo inédito en la provincia de Córdoba y en nuestro país. Este puede ser el punto inicial de nuevos trabajos donde se puedan ir perfeccionando el cálculo del índice y su ajuste estadístico. Algunas líneas de análisis que se pueden continuar a partir de este trabajo es ajustar la distribución del índice utilizando métodos de interpolación (s-splines), o modelos de aprendizaje automático (redes neuronales).

Dado que la forma de la distribución de una variable de vida se puede componer de más de una forma básica, una alternativa es contemplar la distribución de una mezcla de

variables, en este sentido otra posibilidad es ajustar el IE a una mezcla finita de variables Weibull, donde las variables Weibull participan con ponderaciones positivas. En este caso se deben estimar los parámetros de las variables que componen la mezcla y el peso de las componentes en la mezcla (Bucar, Nagode, & Fadjiga, 2004).

Teniendo en cuenta que se conocen las localidades a las que pertenecen los agroaplicadores, en otra línea de análisis se puede realizar un ajuste que considere la distribución espacial del IE en la provincia, de manera de examinar de determinar puntos de corte para el grupo de alto riesgo teniendo en cuenta la distribución espacial y considerar en este caso políticas de regulación de aplicación de plaguicidas diferenciadas (Goodking, Coggins, & Marshall, 2014).

También se puede replicar este estudio en otras provincias y extender el trabajo a todo el país.

Referencias

- Aarset, M. V. (1987). How to identify a bathtub shaped hazard rate? *IEEE Trans Reliab* 36, 106-108.
- Alavanja, M., Samanic, C. ., Lubin, J., Tarone, R., Lynch, C., Knott, C., y otros. (2003). Use of Agricultural Pesticides and Prostate Cancer Risk in the Agricultural Health Study Cohort. *American Journal of Epidemiolog*, 800–814.
- Balakrishnan, N., Leiva, V., & López, J. (2007). Acceptance Sampling Plans from Truncated Life Tests Based on the Generalized Birnbaum–Saunders Distribution. *Journal Communications in Statictis -Simulation and Computation* 36, 643-656.
- Barros, M., Leiva, V., & Paula, G. (6 de Mayo de 2008). Paquete de lenguaje R "gbs".
- Birnbaum, Z., & Saunders, S. (1969a). A new family of life distributions. *J. Appl. Prob.*, (6): 319-327.
- Bonner, M., & Alavanja, M. (2005). The Agricultural Health Study biomarker workshop on cancer etiology. Introduction: overview of study design, results, and goals of workshop. *Journal of Biochemical and Molecular Toxicology*, 19(3):169-171.
- Bucar, T., Nagode, M., & Fadjdiga, M. (2004). Reliability approximation using finite Weibull mixture distributions. *Reliability Engineering and System Safety*, 241–251.
- Chen, G., & Balakrishnan, N. (1995). A General Purpose Apppproximate Goodness-of-Fit Test. *Journal of quality technology*, 154.

- Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, Volume 64, Issue 4.
- Díaz, M. d., Butinof, M., Lantieri, M., R., F., Blanco, M., & Stimolo, M. (2007). *GEACC (Grupo de Epidemiología Ambiental del Cáncer en Córdoba)*. Obtenido de geacc.webs.fcm.unc.edu.ar/lineas-de-investigacion/
- Díaz-García, J. A., & Leiva-Sánchez, V. (2002). A new family of life distributions based on Birnbaum-Saunders distribution. *PE/CIMAT, Comunicación Técnica* No I-02-17.
- Dirección de Estimaciones Agrícolas - Dirección Nacional de Agricultura . (2018). *Estimaciones Agrícolas*. Obtenido de Ministerio de Agricultura, Ganadería y Pesca, Presidencia de la Nación: <http://datosestimaciones.magyp.gob.ar/>
- Dosemeci, M., Alavaja, M., A.S., R., Mage, D., S., H. Z., Rothman, N., y otros. (2002). A Quantitative Approach for Estimating Exposure to Pesticides in the Agricultural Health Study. *The Annals of Occupational Hygiene*, 46 (2): 245-260.
- Ferreira, M., Gomes, M. I., & Leiva, V. (2012). On An Extreme Value Version of the Birbaum-Saunders Distribution. *Revstat Statistical Journal* 10, 181-210.
- Gómez Carmona, O., & Garcés N., L. P. (2004). INFERENCIA ESTADÍSTICA UTILIZANDO REDES NEURONALES ARTIFICIALES. *Scientia Et Technica*, 31-36.
- Goodking, A. L., Coggins, J. S., & Marshall, J. D. (2014). A Spatial Model of Air Pollution: The Impact. *Journal of the Association of Environmental and Resource Economists*, 451-479.
- Hoppin, J., Umbach, D., London, S. J., Lynch, C., Alavanja, M., & Sandler, D. (2006). Pesticides Associated With Wheeze Among Commercial Pesticide Applicators

- in the Agricultural Health Study. *American Journal of Epidemiology*, 1129-1137.
- Jhonson, N., Kotz, S., & Balakrishnan, N. (1995). *Continuous Univariate Distributions Volume 2*. New York; Chichester;Brisbane;Toronto; Singapore: Jhon Wiley & Sons Inc.
- Lange, K., Little, J., & Taylor, M. (1989). Robust statistical modelling using the t distribution. *J. Am. Stat. Soc.*, 84:881-896.
- Lantieri, M., Butinof, M., R.A., F., Stimolo, M., Blanco, M., & Diaz, M. D. (2011). *Work Practices, Exposure Assessment and Geographical Analysis of Pesticide Applicators in Argentina*. Córdoba: Margarita Stoytcheva.
- Leiva, V. (11 al 15 de Agosto de 2010). Modelos de Daño Acumulativo y su Implementación en Código R. *Simposio de Estadística 20 años*. Santa Marta , Colombia.
- Leiva, V., Barros, M., & Paula, G. A. (Diciembre de 2008). *www.victorleiva.cl*.
Obtenido de
http://www.victorleiva.cl/archivos/books/leiva_barros_paula_2009.pdf.old.pdf
- Leiva, V., Barros, M., Paula, G. A., & Sanhueza, A. (2008). Generalized Birbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics 19*, 235-249.
- Leiva, V., Barros, M., Paula, G., & Sanhuenza, A. (2008). Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. . En *Environmetrics, 19 (3)* (págs. 235-249).
- Leiva, V., Barros, M., Paula, G., & Sanhuenza, A. (2008). Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. En *Environmetrics, 19 (3)* (págs. 235-249).

- Leiva, V., Barros, M., Paula, G., & Saunders, S. (11 al 15 de Agosto de 2010). Modelos de Daño Acumulativo Birbaum-Saunders y su Implementación en Código R. *Simposio de Estadística 20 años*. Santa Marta, Colombia.
- Leiva, V., Ferreira, M., Gomes, M. I., & Lillo, C. (2015). Extreme Value Birbaum-Saunders Regression Models Applied to Environmental Data. *Stochastic Environmental Research and Risk Assessment*, 1069-6.
- Leiva, V., Hernández, H., & Riquelme, M. (3 de Septiembre de 2007). *Package for the Birbaum-Saunders distribution*. Obtenido de <http://www.r-project.org>
- Leiva, V., Hernández, H., & Sanhuenza, A. (Junio de 2008). *An R Package for a General Class of Inverse*. Obtenido de Journal of Statistical Software: <http://www.jstatsoft.org/>
- Leiva, V., Hernández, H., & Sanhuenza, A. (20 de Abril de 2008). Paquete de lenguaje R "ig".
- Leiva, V., Riquelme, M., Balakrishnan, N., & Sanhuenza, A. (2008). Lifetime analysis based on the Generalized Birbaum-Saunders distribution. *Computational Statistics & Data Analysis* 52, 2079-2097.
- Leiva, V., Soto, G., Cabrera, E., & Cabrera, G. (2011). Nuevas cartas de control basadas en la distribución. *Revista Colombiana de Estadística*, 34 (1) ., 147 a 176.
- Leiva, V., Vilca, F., Balakrishnan, N., & Sanhueza, A. (2010). A Skewed Sinh-Normal Distribution and Its Properties and Application to Air Pollution. *Communications in Statistics-Theory and Methods* 39, 426-443.
- Marshall, A., & Olkin, I. (2007). *Life Distributions*. New York: Springer.
- Ntzani, E., Chondrogiorgi, M., Ntritsos, G., Evangelou, E., & Tzoulaki, I. (2013). *Literature review on epidemiological studies linking exposure to pesticides and*

health effects. Obtenido de European Food Safety Authority:

www.efsa.europa.eu/publications

- OECD, Organisation for Economic Co-operation and Development. (2008). Handbook on Constructing Composite Indicators. *Handbook on Constructing Composite Indicators*. The OECD member countries are: Australia, Austria, Belgium, Canada, the Czech Republic,.
- Palacios, M. E., & Paz, M. P. (2011). Sintomatología persistente en trabajadores agrícolas expuestos a plaguicidas órgano-fosforados. *Revista Facultad Nacional de Salud Pública Medellín*, 29(2):153-162.
- Ramírez, J., & Lacasaña, M. (2001). Plaguicidas: clasificación, uso y medición de la exposición. *Arch Prev Riesgos Labor.*, 4(2):67-75.
- Rusiecki, J., Hou, L., Lee, W., Blair, A., Dosemeci, M., Lubin, J., y otros. (2006). Cancer incidence among pesticide applicators exposed to metolachlor in the Agricultural Health Study. *International Journal of cancer*, 3118-3123.
- Sanhuenza, A., Leiva, V., & Balakrishnan, N. (2008). A new class of inverse Gaussian type distributions. *Metrika*, 31-49.
- Saunders, S. C. (2007). *Reliability, life testing and prediction of services lives*. New York: Springer.
- Schuschny, A., & Soto, H. (Mayo de 2009). Guía metodológica Diseño de indicadores compuestos de desarrollo sostenible. *Documento de Proyecto*. Santiago de Chile: Naciones Unidas.
- Silva, C., Alvarado, S., Montaña, R., & Pérez, P. (2003). Modelamiento de la contaminación atmosférica. *BIOMATEMÁTICA XIII*, 113–127.

Vilca, F., Sanhuenza, A., Leiva, V., & Christakos, G. (2010). An extended Birnbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. *Stoch Environ Res Risk Assess* 24, 771-782.

Vilca, F., Sanhueza, A., Leiva, V., & Christakos, G. (2010). An extended Birbaum-Saunders model and its application in the study of environmental quality in Santiago, Chile. *Stochastic Environmental Research and Risk Assessment* 24, 771-782.

Ybran, R. G., & Lacelli, G. A. (14 de Marzo de 2016). *Instituto Nacional de Tecnología Agropecuaria*. Obtenido de <http://inta.gob.ar/documentos/informe-estadistico-del-mercado-de-la-soja>

Anexo 1

La familia de distribuciones de contornos elípticos es una clase de distribuciones multivariadas cuyas densidades tienen la misma forma elíptica que la distribución normal. Sin embargo, ellas también incluyen distribuciones cuyas colas son más y menos pesadas que la distribución normal, siendo esta última un caso particular de esta familia (Leiva, Barros, Paula, & Saunders, 2010).

Una variable X con distribución elíptica se caracteriza por sus parámetros de posición μ , de escala σ^2 y una función g generadora de densidades o kernel. La notación usada en este caso es $X \sim EC(\mu, \sigma^2; g)$. La función de densidad de X es

$$f_X(x) = c \frac{1}{\sqrt{\sigma^2}} g\left(\frac{[x-\mu]^2}{\sigma^2}\right); x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma^2 > 0,$$

donde c es una constante de normalización, tal que $\int_{-\infty}^{\infty} g(Z^2) dZ = \frac{1}{c}$

Para el caso unidimensional las distribuciones elípticas corresponden a todas las distribuciones simétricas en \mathbb{R} .

En este trabajo se consideran distribuciones esféricas en \mathbb{R} es decir:

$Z = \frac{x-\mu}{\sigma} \sim EC(0, 1; g) \equiv S(g)$ Sea $Z \sim S(g)$. Entonces la densidad simétrica es

$$f(Z) = cg(Z^2) \quad Z \in \mathbb{R} \text{ donde } g(\cdot) \text{ es el kernel de la densidad de } Z \text{ y la función de}$$

distribución simétrica está dada por $F(Z) = P(Z \leq z) = 1 - F(-Z) \quad Z \in \mathbb{R}$ (Leiva,

Barros, Paula, & Saunders, 2010).

Cuadro 19: kernel (g) y constantes de normalización (c) para las distribuciones indicadas.

Distribución	Notación	c	$g(Z^2), Z \in \mathbb{R}$
Kotz tipo	$KT(q,r,s)$	$\frac{sr^{[2q-1]/[2s]}}{\Gamma\left(\frac{2q-1}{2s}\right)}$	$Z^{2[q-1]} \exp(-rZ^{2s}), q > \frac{1}{2}, r > 0, s > 0$
Logística	$\text{Logistic}(0,1)$	1	$\frac{[\exp(Z)]}{[1 + \exp(Z)]^2}$
Normal	$N(0,1)$	$\frac{1}{\sqrt{2\pi}}$	$\exp\left(-\frac{1}{2}Z^2\right)$
t-Student	t_v	$\frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma(v+2)\sqrt{v\pi}}$	$\left[1 + \frac{Z^2}{v}\right]^{-[v+1]/2}, v > 0$

Fuente: (Leiva, Barros, & Paula, www.victorleiva.cl, 2008)

La distribución estándar Laplace es un caso particular de la distribución KT cuando $q=1$, $r=1/2$, y $s=1/2$, esto es

$$Laplace(0,1) \equiv KT(1,1/2,1/2).$$

Anexo 2

Estimar la tasa de riesgo es difícil. Una herramienta que permite determinar la forma de la tasa de riesgo es el gráfico TTT, en inglés *total time on test* (Leiva, Soto, Cabrera, & Cabrera, 2011).

Específicamente, si T es una variable aleatoria positiva, con función de distribución $F(t)$ y función de sobrevivida $R(t)$ sus funciones TTT y TTT escalada son:

$$\text{Transformación TTT} \quad H_F^{-1}(t) = \int_0^{F^{-1}(t)} R(u) du \quad \text{donde } F^{-1}(t) \text{ es la inversa de } F(t)$$

$$\text{Transformación TTT escalada} \quad \phi_F(t) = \frac{H_F^{-1}(t)}{H_F^{-1}(1)} \quad \text{para } 0 \leq t \leq 1$$

Para una muestra aleatoria t_1, t_2, \dots, t_N de tamaño N

$$\text{La transformación TTT empírica está dada por } H_N^{-1}\left(\frac{r}{N}\right) = \int_0^{F_N^{-1}(r/N)} R_N(u) du$$

para $r = 1, 2, \dots, N$ y donde F_N es la función de distribución empírica

$$\text{Ahora } H_N^{-1}\left(\frac{r}{N}\right) \text{ se puede aproximar empíricamente por } \sum_{i=1}^r T_{N:i} + (N-r)T_{N:r}$$

Por lo tanto la Transformación TTT escalada empírica es :

$$\phi_N\left(\frac{r}{N}\right) = \frac{H_N^{-1}\left(\frac{r}{N}\right)}{H_N^{-1}(1)} = \frac{\left[\sum_{i=1}^r T_{N:i} + (N-r)T_{N:r} \right]}{\sum_{i=1}^N T(i)}$$

Para $0 \leq t \leq 1$ $\phi_N(t)$ se define por interpolación lineal

donde T_i es una muestra aleatoria $i = 1, \dots, N$

donde $T_{N:i}$ es una muestra ordenada $i = 1, \dots, N$

En el gráfico TTT los puntos consecutivos $\left(\frac{r}{N}, \phi_N\left(\frac{r}{N}\right)\right)$ $r = 0, \dots, N$

están conectados por líneas rectas.