



MODELO DE REGRESIÓN LOG-GAMMA GENERAL APLICADO A DATOS CENSURADOS

Trabajo de titulación para optar al título profesional de:
Ingeniero Estadístico

Presentado por:
Nicole Aguilar Manríquez

Profesora guía:
Dra. Claudia Navarro Villarroel

Valparaíso, Chile, Marzo 2017

AGRADECIMIENTOS

Agradezco a cada una de las personas que han estado presentes durante este gran proceso de mi formación profesional, principalmente a mi familia, por brindarme su apoyo incondicional en todo momento; a mis profesores, los cuales me guiaron y orientaron durante este período; a mi novio, quien a lo largo de este proceso confió en mí, otorgándome todo el apoyo necesario, y por último a mis amigos, los cuales siempre me alentaron durante esta gran etapa.

Agradezco a mi profesora guía Claudia Navarro Villarroel, quien fue la persona que me orientó y me brindó toda su ayuda durante este ciclo. Fue la persona que siempre tuvo la mejor disposición para guiarme a lo largo de este período. Agradezco además su motivación incondicional, y por sobre todo, su capacidad de formarme académicamente para enfrentar los retos de la vida.

Agradezco a cada una de las personas que forman parte del Instituto de Estadística. Gracias por todo el aprendizaje, sabiduría y enseñanzas que me entregaron a lo largo de todos los años de mi formación. Fueron y serán, un pilar fundamental en mi vida.

Finalmente agradezco al personal del CESFAM Juan Bautista Bravo Vega, en especial a la Directora Sra. Rita González Schiappacasse, a la Enfermera María Álvarez Quezada y a la Técnico en Enfermería Maryori Sánchez Moyano. Gracias por todas las gestiones, por la colaboración y por la disposición que tuvieron independiente de sus labores.

Dedico este trabajo de titulación a mi hijo Bruno y a mis padres Cecilia y Damián, quienes fueron las personas que me alentaron y confiaron en mí para llevar a cabo esta gran meta. Gracias por todo el amor y apoyo incondicional que brindaron en cada momento de mi formación profesional.

Gracias por la comprensión y paciencia durante este proceso.

Gracias a todos, por ser parte de esto.

RESUMEN

Al realizar un análisis, es frecuente hallar situaciones donde existen observaciones incompletas de períodos que transcurren desde un tiempo inicial, hasta un tiempo final; esto se debe principalmente a la censura de los datos. La censura corresponde a un mecanismo que dificulta la observación completa del tiempo de seguimiento de un individuo, asociado a un determinado evento de interés.

El objetivo principal de este trabajo, es aplicar el modelo de regresión log-gamma general a un conjunto de datos censurados. Los cuales hacen referencia a los resultados de la cobertura de acuerdo a la campaña de vacunación contra la influenza del año 2016, correspondiente a las mujeres embarazadas a partir de las 13 semanas de gestación, inscritas en el CESFAM JBBV. Específicamente, aquellas observaciones son descritas mediante el modelo aleatorio de censura por la derecha, donde la duración del estudio corresponde a los meses en que se extiende la campaña; siendo el evento de interés, la inmunización del usuario.

De acuerdo a los análisis estadísticos realizados, se tiene que la censura está presente en el 50% de los casos. La mayor parte de ella, se produce debido a que las embarazadas no se inmunizaron durante la campaña; mientras que el porcentaje restante, corresponde a mujeres que sufrieron un aborto, o bien que se trasladaron de CESFAM antes que recibieran la vacuna antigripal. Por otro lado, al considerar el ajuste del modelo, los resultados indican que la edad gestacional y la cantidad de inmunizaciones contra la influenza recibidas anteriormente, no afectan al tiempo de seguimiento de las gestantes; mientras que la edad de aquellas mujeres, sí afecta a la variable respuesta.

Finalmente, se concluye que el modelo de regresión sí es adecuado para el conjunto de datos utilizado. Por esta razón, los resultados son una buena aproximación de la realidad de la información analizada. Es decir, aquellos resultados, reflejan la realidad de lo sucedido con respecto a la campaña de vacunación contra la influenza, correspondiente al año 2016 del CESFAM Juan Bautista Bravo Vega.

ABREVIATURAS Y NOTACIÓN

A continuación se presentan las abreviaturas y notaciones que se utilizarán a lo largo de este trabajo.

1EPV	Estimador ponderado de verosimilitud de un paso
AFT	Tiempo de fallo acelerado (del inglés <i>accelerated failure time</i>)
CESFAM	Centro de Salud Familiar
EMV	Estimador de máxima verosimilitud
EPV	Estimador ponderado de verosimilitud
GG	Gamma generalizada
i.i.d.	Independientes e idénticamente distribuidas
INE	Instituto Nacional de Estadísticas
JBBV	Juan Bautista Bravo Vega
LGG	Log-gamma generalizada
MINSAL	Ministerio de Salud
OMS	Organización Mundial de la Salud
PAI	Programa Ampliado de Inmunizaciones
PNI	Programa Nacional de Inmunizaciones
RAYEN	Red Asistencial y Enlace Nacional
RNI	Registro Nacional de Inmunizaciones
RUN	Rol Único Nacional
SRCeI	Servicio de Registro Civil e Identificación
VIH	Virus de la inmunodeficiencia humana

Índice general

AGRADECIMIENTOS	2
RESUMEN	3
ABREVIATURAS	4
Lista de figuras	7
Lista de tablas	8
OBJETIVOS	9
PRELIMINARES	10
1. INTRODUCCIÓN	17
1.1. Datos censurados	18
1.2. Modelos de regresión	23
2. MODELO DE REGRESIÓN LOG-GAMMA GENERAL	25
2.1. Introducción	25
2.1.1. Historia	25
2.2. Distribución gamma	26
2.3. Distribución gamma generalizada	27
2.4. Distribución log-gamma generalizada	28
2.5. Estimación de parámetros de la distribución LGG	32
2.5.1. Estimador de máxima verosimilitud	32
2.5.2. Estimador ponderado de verosimilitud	35
2.6. Modelo de regresión log-gamma general	36

2.7.	Estimación de parámetros del modelo de regresión	37
2.7.1.	Estimador ponderado de verosimilitud	37
3.	APLICACIÓN	38
3.1.	Introducción	38
3.1.1.	Antecedentes generales	38
3.2.	Base de datos	39
3.2.1.	Administración de datos	40
3.2.2.	Variables	40
3.3.	Análisis exploratorio	41
3.4.	Aplicación del modelo de regresión log-gamma general	44
3.4.1.	Análisis de residuos	45
	CONCLUSIONES FINALES	47
	Referencias	48
	APÉNDICE	52
A.1.	Código Stata	52
A.2.	Código R-project	63
A.3.	Certificado Directora CESFAM JBBV	66

Índice de figuras

1.	Cobertura Nacional Influenza año 2015. Departamento de Vacunas e Inmunizaciones (2016).	15
2.1.	Distribuciones asociadas a la familia LGG. Adaptado de Agostinelli et al. (2016).	30
3.1.	Distribución del tiempo de seguimiento y del log tiempo de seguimiento ajustadas a sus respectivas densidades.	43
3.2.	Distribución del log tiempo de seguimiento ajustada a una distribución LGG.	43
3.3.	Diagnóstico del modelo.	45

Índice de tablas

2.1. Distribuciones asociadas a la distribución GG.	28
2.2. Distribuciones asociadas a la distribución LGG.	29
3.1. Descripción de los tipos de datos a utilizar.	41
3.2. Resumen descriptivo de las variables de interés.	41
3.3. Resumen estadístico del logaritmo de la variable respuesta.	42
3.4. Estimación de parámetros del modelo de regresión log-gamma general.	44

OBJETIVOS

Los objetivos de este trabajo de titulación se presentan a continuación.

Objetivo general

El principal objetivo de este trabajo es aplicar el modelo de regresión log-gamma general, a los resultados de la cobertura de acuerdo a la campaña de vacunación contra la influenza del año 2016, correspondiente a las mujeres embarazadas a partir de las 13 semanas de gestación, pertenecientes al CESFAM Juan Bautista Bravo Vega de la ciudad de Villa Alemana.

Objetivos específicos

- (i) Definir el tipo de censura a utilizar para la aplicación del modelo.
- (ii) Analizar y estudiar el modelo de regresión log-gamma general, su estructura y sus componentes.

PRELIMINARES

Actualmente existe un gran acuerdo con respecto a la importancia del sistema de salud en el país. Tanto en Chile como en países extranjeros, se tiene la necesidad de lograr determinados estándares para mejorar la calidad de salud de la población, del mismo modo, verificar si con los recursos que se invierten, se obtienen resultados esperados.

En Chile, enfermedades como la rubéola, el sarampión, la influenza, entre otras. Se encuentran totalmente controladas debido a los efectivos programas y campañas de vacunación desarrollados por el Ministerio de Salud (MINSAL). Iniciativa que ha permitido a través del tiempo, señalar al país como un modelo para Latinoamérica en esta área. Afortunadamente, es posible evitar el contagio de enfermedades inmunoprevenibles¹ importantes para la salud, mediante el uso de vacunas.

Las vacunas son esenciales para proteger a la población de diversas enfermedades. Éstas, son instrumentos médicos que contienen sustancias biológicas que activan la producción de anticuerpos, generando inmunidad contra una enfermedad en particular (OMS, 2016c). Actualmente las vacunas son administradas a nivel mundial, evitando entre dos y tres millones de fallecimientos cada año. Sin embargo, los programas de inmunización impulsados en cada país con el fin de prevenir diversas enfermedades, han permanecido constantes en los últimos años (OMS, 2016a).

Chile dispone de un Programa Ampliado de Inmunizaciones (PAI) desde el año 1978. Sus importantes crecimientos a nivel nacional lograron dejar atrás su denominación en el año 2009, adquiriendo el nombre de Programa Nacional de Inmunizaciones (PNI) (Departamento de Vacunas e Inmunizaciones, 2012). El actual programa tiene por objetivo evitar la morbilidad y mortalidad por causa de patologías importantes para la salud pública, en

¹Las enfermedades inmunoprevenibles se definen como aquellas afecciones prevenibles mediante el uso de vacunas. Enfermedades tales como tuberculosis, sarampión, rubéola, hepatitis, rotavirus e influenza, son algunas de las principales afecciones inmunoprevenibles (OMS, 2016a).

conformidad con el desarrollo biotecnológico y evidencia científica a través de diversos tipos de vacunas. Estas inmunizaciones destinadas a la población nacional, se encuentran definidas en el PNI de acuerdo a las especificaciones proporcionadas por la Organización Mundial de la Salud (OMS) (Departamento de Vacunas e Inmunizaciones, 2016c).

Las vacunas que abarca el PNI están clasificadas en tres categorías: vacunas por programas, vacunas por campañas y vacunas por eventos y/o condiciones especiales. Las vacunas por programas se administran continuamente de acuerdo a determinaciones científicas, acorde a un calendario de inmunización definido anualmente. Éstas, son destinadas a diversas poblaciones principalmente infantiles y escolares. Por otro lado las vacunas por campañas se aplican generalmente de forma anual durante un determinado intervalo de tiempo. Éstas, son dirigidas a numerosas poblaciones propensas a desarrollar complicaciones, en el caso de adquirir enfermedades causadas por diversos virus presentes en el país. Por último las vacunas por eventos y/o condiciones especiales, se administran por causa de un suceso imprevisto, o bien por indicación profesional debido a una determinada enfermedad de riesgo, como es el caso de la hepatitis (Departamento de Vacunas e Inmunizaciones, 2012).

Además de las inmunizaciones anteriormente mencionadas, el país cuenta con otras vacunas complementarias al PNI que se aplican de forma particular y sólo con certificación médica. En el caso de vacunas para viajeros, éstas se utilizan para no adquirir el posible contagio de enfermedades que en Chile ya están retiradas y/o erradicadas, como es el caso de los virus de la viruela, el sarampión y la poliomielitis. A diferencia de las vacunas complementarias, las vacunas incluidas en el PNI son obligatorias para todas las poblaciones definidas anualmente de acuerdo a cada tipo de inmunización. Estas vacunas son de carácter gratuito y se encuentran constituidas con bienes públicos.

Entre los acontecimientos más relevantes de acuerdo a inmunizaciones en Chile, se destaca la erradicación de la viruela en el año 1950, la eliminación del virus de la poliomielitis en el año 1975 y la eliminación del sarampión en el año 1992. Además de la implementación de vacunas como las nombradas anteriormente comprendidas en el PNI. Tanto las vacunas de campañas como las vacunas programadas, han permitido evitar enfermedades contagiosas contribuyendo a la disminución de la mortalidad en el país. Sin embargo, la implementación de la vacuna contra la influenza en Chile en el año 1982, ha logrado prevenir esta enfermedad estacional causada por distintos tipos de virus que se originan todos los inviernos en el mundo (Departamento de Vacunas e Inmunizaciones, 2016b). Por esta razón, cada año la OMS analiza los agentes infecciosos que se presentan en cada hemisferio, proyectando en relación a lo sucedido en la temporada invernal del hemisferio norte, las diferentes clases de virus de la influenza que serán más significativas en el hemisferio sur. La OMS además proporciona a cada laboratorio, las indicaciones para la fabricación de las inmunizaciones a utilizar en cada campaña de vacunación contra la influenza (MINSAL, 2016).

La influenza o gripe estacional, es una enfermedad respiratoria que ataca anualmente durante el período invernal a la población nacional. Actualmente miles de personas fallecen cada año en el mundo debido a esta infección viral. Sin embargo durante el año 2015, 44 personas fallecieron en el país por causa de esta patología (Departamento de Vacunas e Inmunizaciones, 2016a). La gripe estacional puede afectar a personas de cualquier edad, incluso es confundida con otros estados infecciosos que también atacan al sistema respiratorio.

Los diversos virus que ocasionan esta afección se encuentran clasificados en tres tipos: A, B y C. No obstante los virus tipo A y B, son los encargados de originar epidemias de infecciones respiratorias contagiosas que se presentan habitualmente todos los inviernos a nivel mundial. A diferencia de los graves problemas de salud que generan ambos tipos de virus mencionados anteriormente, el virus de la influenza tipo C, ocasiona frecuentemente leves enfermedades que afectan al sistema respiratorio (URMC, 2016). Por esta razón, las campañas de inmunización impulsadas anualmente para prevenir la gripe, incluyen exclusivamente vacunas que contengan cepas de los virus tipo A y B. Entre los diversos subtipos del virus tipo A, actualmente circulan en la especie humana los virus subtipos A (H1N1) y A (H3N2) (OMS, 2016b).

Al igual que un resfriado común, los virus de la influenza se transmiten rápidamente a través manos, objetos o superficies contaminadas. Las partículas infectadas que expulsa una persona contagiada al toser y/o estornudar, pueden ser transmitidas espontáneamente a personas sanas de cualquier edad. Sin embargo estos virus se propagan con facilidad en hogares, escuelas, centros de salud y ciudades. A diferencia de un resfrío habitual, sus síntomas son significativamente mayores, afectando primordialmente a la garganta, nariz, bronquios y en ocasiones a los pulmones. La gripe se presenta de forma repentina y se caracteriza por el inicio de fiebre elevada (sobre 38 °C), dolor de garganta, tos y malestar general (OMS, 2016b). Estos síntomas frecuentemente desaparecen alrededor de siete días. No obstante las personas propensas a sufrir complicaciones en el caso de adquirir esta patología, pueden manifestar problemas de salud que conllevan ocasionalmente a la muerte. Dentro de las personas propensas a desarrollar estas dificultades, se encuentran los niños entre 6 meses y 5 años de edad, las embarazadas con más de 13 semanas de gestación, los adultos mayores de 65 años de edad y las personas con enfermedades crónicas² (MINSAL, 2015).

La forma más segura de prevenir esta infección es a través de la administración de la vacuna antigripal. Su eficacia varía dependiendo de cada organismo, según la edad y las condiciones de salud en que se encuentre cada paciente (URMC, 2016). Su eficacia además varía dependiendo de los constantes cambios que presentan los virus cada año en el mundo.

²Las enfermedades crónicas son aquellas afecciones de larga duración y generalmente de lento progreso. Enfermedades tales como cardíacas, renales, sanguíneas, vasculares, hepáticas y metabólicas, son algunas de las principales patologías crónicas (OMS, 2016b).

Por esta razón, la OMS sugiere aplicar anualmente la inmunización contra la influenza a los siguientes grupos de riesgo, según el orden de prioridad que se indica a continuación (OMS, 2016b).

- Mujeres embarazadas en cualquier etapa del embarazo.
- Niños entre 6 meses y 5 años de edad.
- Adultos mayores de 65 años de edad.
- Personas con enfermedades crónicas.
- Trabajadores del área de la salud.

Frente a este escenario, en Chile se impulsó nuevamente la campaña nacional de inmunización contra la influenza correspondiente al año 2016, donde su finalidad es vacunar a diversos grupos que puedan desarrollar dificultades por causa de esta patología. Debido a que los virus de la gripe, son los encargados de producir graves afecciones –principalmente respiratorias– que permiten desencadenar complicaciones severas e incluso la muerte. El Departamento de Vacunas e Inmunizaciones del Ministerio de Salud, adquirió las vacunas contra la influenza para uso adulto y pediátrico. Ambas cumplen con las especificaciones de la OMS de acuerdo a las cepas de los virus del año 2016 para el hemisferio sur (MINSAL, 2016). Esta campaña de inmunización está indicada a toda la población nacional, y de forma obligatoria a los grupos de población objetivo³ definidos anualmente por el Ministerio de Salud.

En un principio la vacuna antigripal estaba destinada a los grupos tradicionales de alto riesgo. Los cuales estaban conformados por adultos mayores de 65 años de edad, personal de salud y pacientes con enfermedades crónicas. Sin embargo en la medida que se dispuso mayor información con respecto a la identificación de nuevos grupos de riesgo, considerando los beneficios proporcionados por la vacuna, en el año 2004 se estableció la inmunización a mujeres embarazadas y en el año 2006, se incorporó la vacunación a los lactantes sanos entre 6 y 24 meses de edad. Por esta razón, Chile ha logrado ser uno de los países con las mejores coberturas⁴ en campañas de inmunización contra la influenza en Latinoamérica (Vega-Briceño, Abarca y Sánchez, 2006). Hecho que sin lugar a dudas, ha repercutido a través del tiempo en el país.

³Los grupos de población objetivo son aquellos conjuntos de personas que pueden presentar un mayor riesgo de sufrir complicaciones causadas por los virus de la influenza (MINSAL, 2016).

⁴La cobertura se calcula como el cociente entre vacunas administradas y vacunas programadas, expresado en términos porcentuales. Las inmunizaciones programadas son establecidas y cuantificadas anualmente para cada grupo de población objetivo (Departamento de Vacunas e Inmunizaciones, 2012).

Actualmente para el año 2016, el Ministerio de Salud estableció de forma obligatoria la vacunación a los grupos de población objetivo definidos a continuación (Decreto N° 73, 2016).

- Embarazadas a partir de las 13 semanas de gestación.
- Niñas y niños entre 6 meses y 5 años 11 meses 29 días.
- Adultos mayores de 65 años de edad.
- Trabajadores de plantas avícolas y de criaderos de cerdo.
- Trabajadores, voluntarios y estudiantes en práctica que desempeñen sus labores en: establecimientos de salud públicos, establecimientos de salud privados e instituciones que cuenten con unidades de atención de urgencias y/o servicios de hospitalización.
- Pacientes entre 6 y 64 años que presenten alguna de las siguientes enfermedades crónicas:
 - Diabetes.
 - Hipertensión arterial en tratamiento farmacológico.
 - Enfermedades pulmonares crónicas, tales como asma bronquial, enfermedad pulmonar obstructiva crónica, fibrosis quística y fibrosis pulmonar de cualquier causa.
 - Cardiopatías, tales como congénitas, reumáticas, isquémicas y miocardiopatías de cualquier causa.
 - Enfermedades neuromusculares congénitas y adquiridas que causan trastornos de la deglución o del manejo de secreciones del sistema respiratorio.
 - Obesidad mórbida.
 - Insuficiencia renal en etapa 4 o mayor.
 - Insuficiencia renal en diálisis.
 - Insuficiencia hepática crónica.
 - Enfermedades autoinmunes como lupus, escleroderma, artritis reumatoidea, enfermedad de Crohn, etc.
 - Cáncer en tratamiento con radioterapia, quimioterapia, terapias hormonales o medidas paliativas de cualquier tipo.
 - Infección por VIH (virus de la inmunodeficiencia humana).
 - Inmunodeficiencias congénitas o adquiridas.

Sin embargo, además de los grupos de población objetivo definidos anualmente, hace un par de años se decidió inmunizar a diversas poblaciones pertenecientes a otras prioridades. Las poblaciones que conforman esta categoría son las siguientes:

- Personas en situación de calle.
- Gendarmería.
- Escuelas institucionales de las Fuerzas Armadas.
- Establecimientos penitenciarios.
- Profesores.

Es preciso señalar que para cada campaña de inmunización contra la influenza, el Ministerio de Salud tiene por objetivo alcanzar el 90 % de cobertura a nivel nacional, en cada uno de los grupos mencionados anteriormente, logrando administrar todas las vacunas programadas por cada Centro de Salud Familiar (CESFAM) a lo largo del país. En el año 2015, la cobertura promedio alcanzada a nivel nacional fue de un 95.9 % (Departamento de Vacunas e Inmunizaciones, 2016a). Sin embargo los resultados obtenidos por cada grupo de población objetivo se presentan a continuación.

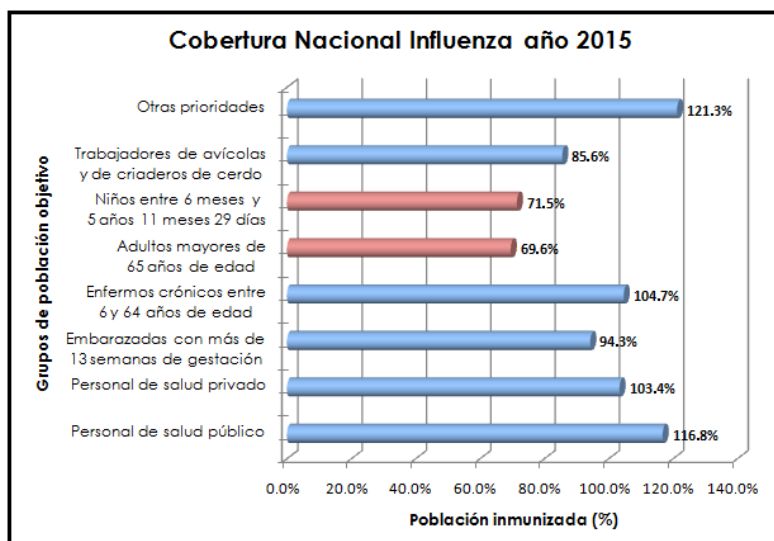


Figura 1: Cobertura Nacional Influenza año 2015. Departamento de Vacunas e Inmunizaciones (2016). Recuperado de http://estadisticas.ssosorno.cl/old_page/RNI/Lineamientos-campa%C3%B1a-Influenza-2016.pdf

Las coberturas de vacunación en niños y en adultos mayores correspondientes al año 2015, disminuyeron en comparación con años anteriores. Ambas fueron considerablemente menores que las obtenidas por los demás grupos de población objetivo (Departamento de Vacunas e Inmunizaciones, 2016a). No obstante las coberturas alcanzadas en otras prioridades, personal de salud, embarazadas y enfermos crónicos, fueron mayores al 90 %, es decir, en aquellos grupos se logró la meta propuesta. Se espera que para el año 2016, las coberturas por cada categoría sean superiores a los resultados obtenidos en años anteriores.

Cabe destacar que la población objetivo a inmunizar en el año 2015, correspondió aproximadamente al 30 % de la población total del país (Departamento de Vacunas e Inmunizaciones, 2016a). Sin embargo, la cantidad de personas a vacunar por cada grupo es calculada anualmente para cada campaña de inmunización. Para el año 2016, las poblaciones fueron obtenidas de la siguiente forma:

- Las poblaciones de niños y adultos mayores se estimaron según proyecciones demográficas durante el período 2013-2020. Éstas fueron calculadas y proporcionadas por el Instituto Nacional de Estadísticas (INE).
- Las poblaciones de trabajadores de plantas avícolas y de criaderos de cerdo, pacientes crónicos y trabajadores del área de la salud (público y privado), se calcularon de acuerdo a la cantidad de inmunizados durante el año 2015, según el Registro Nacional de Inmunizaciones (RNI).
- La población de embarazadas fue calculada de acuerdo al promedio de nacidos vivos en el país durante el período 2011-2015, según el Servicio de Registro Civil e Identificación (SRCeI).

En un comienzo la campaña de vacunación contra la influenza correspondiente al año 2016, se implementaría a partir del 14 de marzo y se extendería hasta el 13 de mayo, o bien hasta que se cumpliera el objetivo del 90 % de cobertura en el país (Departamento de Vacunas e Inmunizaciones, 2016a). Sin embargo, debido a que no se originó un incremento significativo de influenza que identifica la época invernal en Chile, el Ministerio de Salud decretó prolongar la campaña de vacunación para los grupos de población objetivo hasta el 30 de junio (MINSAL, 2016). Insistiendo principalmente en inmunizar a mujeres embarazadas a partir de las 13 semanas de gestación, a niños entre 6 meses y 5 años 11 meses 29 días, a personas mayores de 65 años y a pacientes con afecciones crónicas.

Finalmente, es importante destacar la necesidad de inmunizar anualmente a las poblaciones que sean más propensas a sufrir dificultades producidas por la influenza. Al igual que años anteriores, se espera que la cobertura de la vacuna antigripal a nivel nacional supere el 90 %, logrando administrar todas las vacunas programadas por cada región, comuna y CESFAM a lo largo del país.

INTRODUCCIÓN

Ocurre a menudo que al realizar algún tipo de análisis se cuente con toda la información disponible, pero al momento de implementar un estudio enfocado a tiempos de vida, sucede en reiteradas ocasiones que los datos u observaciones presentan uno de los problemas más frecuentes; que estén incompletos o que presenten incertidumbre con respecto a cuándo ocurrió el evento de interés fijado por el investigador. Este conjunto de observaciones es denominado “datos censurados”. La censura se origina cuando existen ocasiones donde no se conoce exactamente el tiempo de vida de una determinada observación, sino que sólo se sabe, que la muerte o falla ocurre dentro de un intervalo de tiempo (Flores, 2011).

Actualmente es posible encontrar una gran variedad de datos, sin embargo el hecho de no conocer la causa por la cual se pierde información con respecto a las observaciones, ocasiona inseguridad e incertidumbre tanto en el proceso de recolección de datos, como también al investigador del estudio. Este problema se origina principalmente en diversos análisis enfocados a tiempos de vida, relacionados al área de la medicina e ingeniería.

En este trabajo de titulación se analizarán datos de vida, específicamente observaciones censuradas pertenecientes al área de la salud de acuerdo a un determinado período de tiempo. Los datos de vida, corresponden frecuentemente a tiempos de falla o de supervivencia de un grupo de individuos en seguimiento y se clasifican en dos categorías: datos completos, los cuales corresponden a un conjunto de observaciones que se presentan en su totalidad, es decir, cuando toda la información está disponible; y datos censurados, los cuales corresponden a un conjunto de observaciones incompletas que se presentan cuando existe información que está perdida durante un intervalo de tiempo.

1.1 Datos censurados

En la vida cotidiana, es frecuente hallar situaciones donde existen observaciones incompletas de períodos que transcurren desde un tiempo inicial, hasta un tiempo final. Esto se debe principalmente a la censura de los datos. La censura corresponde a un mecanismo que dificulta la observación completa del tiempo de seguimiento de un determinado individuo. Ésta se presenta habitualmente en diversas aplicaciones y se encuentra clasificada en tres categorías: censura tipo I (censura por tiempo), censura tipo II (censura por fallas) y censura tipo III (censura aleatoria). La censura tipo I es aquella que se presenta cuando se observa a los individuos hasta un tiempo definido. La censura tipo II se manifiesta cuando se observa a los individuos hasta que ocurra un número establecido de eventos, o bien hasta que se produzca una cantidad determinada de fallas. Sin embargo la censura aleatoria, se origina cuando los individuos u observaciones presentan acontecimientos independientes al evento de interés, causando –en algunas ocasiones– su eliminación del estudio (López, 2011).

Habitualmente los tres tipos de censura se encuentran relacionados con diversas áreas enfocadas a la industria, economía, psicología y medicina (Godoy, 2009). No obstante la censura tipo I y la censura aleatoria, son frecuentemente las más empleadas en numerosos estudios relacionados con las ciencias de la salud. Este fenómeno es producido por otras categorías de censura que se encuentran clasificadas en tres grupos: censura por la derecha, la cual se presenta si en la última observación de un conjunto de datos, aún no ha ocurrido el evento de interés fijado por el investigador; censura por la izquierda, la cual se manifiesta si al momento de realizar la primera observación de un conjunto de datos, ya ha ocurrido aquel suceso; y censura por intervalo, la cual se presenta si el evento ocurre en un determinado período de tiempo (López, 2011).

En otras palabras, los datos censurados hacen referencia a algún tipo de información desconocida en situaciones donde la variable de interés, corresponde al tiempo de vida de los individuos. Generalmente se dice que las observaciones se encuentran censuradas, cuando poseen información parcial sobre su existencia durante un seguimiento específico. Es decir, cuando no se conoce con certeza el tiempo de supervivencia de una observación en un lapso definido (Flores, 2011).

En diversas situaciones, sucede que al realizar estudios enfocados al análisis de supervivencia se encuentre con este tipo de datos, debido a que principalmente se trabaja con tiempos de vida o de falla –frecuentemente censurados– de una determinada área en particular. El análisis de supervivencia se define como un conjunto de técnicas estadísticas que se utilizan para analizar observaciones, donde la variable de interés constantemente es el tiempo que transcurre desde un instante inicial, hasta que acontece un determinado evento de interés o instante final (Flores, 2011).

El tiempo de supervivencia corresponde al período que transcurre desde la fecha de inicio o ingreso al estudio de una observación, hasta el momento en que ocurre un evento, fecha o tiempo en que finaliza el estudio (Godoy, 2009). Es decir, es el tiempo desde que se produce un suceso (nacimiento de un paciente, comienzo de un tratamiento o inicio de alguna actividad), hasta el acontecimiento de un evento de interés (muerte de un paciente, fin de un tratamiento o término de alguna actividad) (Flores, 2011). En general estos lapsos que transcurren entre dos episodios, son denominados tiempos de supervivencia, tiempos de vida o tiempos de falla, dependiendo de la aplicación a realizar.

En un análisis de supervivencia existen dos funciones de gran interés: la función de supervivencia y la función de riesgo. La función de supervivencia corresponde a la probabilidad de que una observación o individuo sobreviva a un tiempo establecido (Hernández, 2010). Es decir, se define como la probabilidad de que un dispositivo falle (experimente el evento de interés), posterior al tiempo delimitado en el estudio (López, 2011). Sin embargo la función de riesgo, corresponde a la probabilidad condicional de que un individuo muera en un intervalo de tiempo posterior al período de seguimiento, dado que sobrevivió a este período de observación (Flores, 2011).

Por ejemplo en diversos estudios que involucran datos de vida, es posible observar durante días, un conjunto de maquinas, dispositivos o piezas en funcionamiento. La variable a estudiar habitualmente es el tiempo de falla de las observaciones, siendo el evento de interés, el momento exacto en que las máquinas o piezas en funcionamiento caducan. No obstante, existen ocasiones en que no es posible observar el momento exacto en que ocurre aquel evento (Hernández, 2010). Algunos ejemplos de aplicaciones que involucran este tipo de datos son los siguientes:

- Área médica: Tiempo de vida de un paciente con una determinada enfermedad.
- Área industrial: Tiempo de falla de alguna máquina o dispositivo electrónico.
- Área psicológica: Tiempo para realizar una evaluación psicológica.
- Área económica: Tiempo de cesantía de una persona económicamente activa.
- Área demográfica: Tiempo (duración) de un matrimonio.

Sin embargo, este tipo de análisis es utilizado principalmente en diversas especialidades relacionadas al área de la medicina e ingeniería. En numerosos estudios clínicos, los investigadores tienen por objetivo analizar los tiempos de vida, y verificar si las características clínicas de los individuos, están directamente relacionadas con el tiempo de supervivencia de cada observación (Flores, 2011).

Algunos ejemplos enfocados al área de la salud que involucran este tipo de datos según Flores (2011) son los siguientes:

- Tiempo transcurrido desde el nacimiento de un paciente hasta el diagnóstico de una determinada enfermedad.
- Tiempo que transcurre desde la hospitalización de un paciente hasta su alta médica.
- Tiempo transcurrido desde la aparición de una determinada enfermedad, hasta la muerte o recuperación del individuo.

En este tipo de estudios, es frecuente que exista un alto porcentaje de observaciones censuradas. Sin embargo este inconveniente se origina principalmente en diversos análisis de mantenimiento industrial y en ensayos clínicos (Flores, 2011). Por ejemplo, si un individuo abandona el estudio por alguna causa ajena; si el seguimiento de un dispositivo se pierde y no se tiene conocimiento de esta observación; o si el estudio termina antes que aparezca el evento de interés (López, 2011).

De acuerdo a lo señalado anteriormente con respecto a los diversos tipos de observaciones censuradas, Godoy (2009) y Hernández (2007) definen detalladamente estas observaciones clasificadas de acuerdo a los siguientes tipos:

- (i) Censura tipo I: Este tipo de censura se origina cuando los individuos son observados hasta un tiempo definido, es decir, cuando las observaciones tienen un seguimiento durante un intervalo de tiempo. En numerosos análisis, se debe fijar el tiempo de observación para que los sujetos presenten el evento de interés. Sin embargo los individuos que no experimentan aquel evento al término de este período, corresponden a observaciones censuradas. En reiteradas ocasiones, debido a razones de presupuesto o de diseño del estudio, el investigador decide terminar el seguimiento antes de que todos los sujetos experimenten la falla. En este caso, el tiempo de vida de las observaciones será un valor constante, es decir, corresponderá al tiempo desde que se inicia el seguimiento, hasta el momento en que el estudio finaliza. No obstante, se tendrá en cuenta que aquellos sujetos no presentaron el evento (Juez y Díez, 1997).
- (ii) Censura tipo II: Este tipo de censura se presenta cuando los datos son observados hasta que ocurra una cantidad determinada de fallas. Sin embargo los individuos que no experimentan el evento de interés al completarse las primeras r fallas, corresponden a observaciones censuradas. En esta ocasión, el investigador decide extender la observación de los individuos en el estudio, hasta que ocurran r de las n posibles fallas ($r \leq n$). Este acontecimiento se produce generalmente en pruebas de duración de máquinas. Por ejemplo, todos los dispositivos se colocan a funcionar al mismo tiempo y la prueba termina cuando los primeros r de los n equipos se deterioran.

Cabe destacar que la determinación del tiempo para la censura tipo I y el número de fallas para la censura tipo II, deben ser fijados previo al experimento y no durante el período de observación. En ambas ocasiones, la censura es controlada por el investigador del estudio.

(iii) Censura tipo III: Este tipo de censura se presenta cuando el inicio del seguimiento de las observaciones no comienza en un instante determinado. Es decir, se fija la duración del estudio y los individuos se incorporan en diferentes momentos a lo largo de aquel período (Cordón y García, 2014). Su característica fundamental se basa en que durante el transcurso de la investigación, algunas observaciones presentan otros acontecimientos independientes al evento de interés, provocando su eliminación del estudio. Este fenómeno se presenta sin el control del investigador. Sin embargo ésta puede surgir debido a la pérdida del seguimiento de un determinado individuo, o bien porque el estudio ha finalizado. Las situaciones que producen este problema son las siguientes:

- En el caso de que no se experimente el evento de interés hasta el momento en que finaliza el estudio (sólo si este período es finito).
- En el caso de que ocurra otro episodio que dificulte llevar a cabo el evento que se desea analizar.
- En el caso de que un individuo abandone el estudio por alguna causa ajena.

En resumen, este inconveniente corresponde a una generalización de la censura tipo I, con la diferencia que en lugar de tener una variable constante (en el caso de que las observaciones no presentaron el evento de interés), ésta incorpora una variable C que representa el tiempo censurado de cada observación (Juez y Díez, 1997).

Además de los tipos de censura detallados anteriormente, existen otras tres categorías de censura que ocasionan este fenómeno, las cuales se clasifican de la siguiente forma:

(a) Censura por la derecha

Esta censura se origina cuando ha finalizado el período de seguimiento de una determinada observación, no obstante el evento de interés ocurre después de este momento. En otras palabras, la censura por la derecha se presenta cuando no se conoce el instante exacto en que ocurrió aquel evento, sólo se sabe, que se ha experimentado después del tiempo de observación del individuo. Frecuentemente, esta censura se produce en numerosos análisis clínicos.

(b) Censura por la izquierda

Esta censura se presenta cuando el evento de interés ocurre antes de que la observación se incluya en el estudio. Es decir, la censura por la izquierda se origina cuando se desconoce el momento preciso en que ocurrió aquel evento, sólo se sabe, que ha sucedido antes del período de seguimiento.

(c) Censura por intervalo o censura doble

Esta censura se presenta cuando los datos se encuentran censurados tanto por la derecha como por la izquierda. Es decir, se produce cuando no es posible observar el momento exacto en que ocurre el evento de interés, sólo se sabe, que se ha experimentado dentro de un intervalo de tiempo. Este fenómeno se origina en diversos estudios donde los sujetos no son analizados continuamente, sin embargo el seguimiento se realiza de forma periódica, y solamente se sabe que aquel suceso ocurrió entre dos períodos de revisión. En general, esta censura se presenta en estudios longitudinales.

En resumen, es importante recalcar lo que sucede cuando la falla no es observada en los tres tipos de censura mencionados anteriormente (tipo I, tipo II y tipo III). Evidentemente, en estos casos predomina la censura por la derecha, dado que se supone que el evento de interés debe manifestarse después del seguimiento (Hernández, 2007). De forma análoga, eventualmente en estos casos puede presentarse censura por la izquierda o censura por intervalo (Juez y Díez, 1997).

En particular, en el presente estudio se trabajará bajo el modelo aleatorio de censura por la derecha, el cual está definido según Jácome (2005) y Hernández (2010).

Sea Y la variable aleatoria de interés que representa el tiempo de vida, con función de densidad $f(\cdot)$ y función de distribución $F(\cdot)$ continua. El modelo de censura aleatoria por la derecha está representado por el par (Y^*, δ) y se define de la siguiente forma:

$$Y^* = \min\{Y, C\} \quad \text{y} \quad \delta = 1_{\{Y \leq C\}} \begin{cases} 1 & \text{si } Y \leq C \\ 0 & \text{si } Y > C \end{cases} \quad (1.1)$$

donde δ corresponde a la variable que indica la censura y C corresponde al tiempo censurado con función de distribución $J(\cdot)$ continua.

Específicamente, cada valor del conjunto de datos será un par (y_i^*, δ_i) , con $\delta_i = 1$ cuando el valor observado de Y^* corresponda al tiempo de vida de un individuo, y $\delta_i = 0$ cuando el valor observado de Y^* corresponda a un tiempo de censura. Es decir, cuando en algunas situaciones no se ha observado el tiempo de supervivencia de un individuo, dado que un acontecimiento aleatorio e independiente al evento de interés ha ocurrido, interrumpiendo el período de seguimiento de la observación hasta aquel evento. En esta ocasión, la información que se observará acerca de la variable de interés, tomará un valor superior al tiempo observado. En el caso de que el período de observación del estudio sea finito y no sea posible observar el momento exacto en que ocurrió el evento de interés, la variable de censura δ será utilizada de forma idéntica, proporcionando la misma finalidad.

En particular, la variable Y^* representa el tiempo transcurrido desde la entrada al estudio de un individuo hasta el término de su seguimiento; la variable Y corresponde al tiempo que transcurre desde la entrada al estudio de un sujeto, hasta el momento en que acontece el evento de interés; y la variable C , representa el tiempo transcurrido desde que un individuo se incluye en el estudio, hasta la pérdida de su seguimiento. Sin embargo en el caso de que no ocurra el evento de interés, el tiempo de censura se extenderá hasta el término del estudio.

En otras palabras, se define que para cada i -ésimo individuo ($i = 1, 2, 3, \dots, n$) existe un tiempo de vida y_i y un tiempo de censura c_i , mutuamente independientes no observados directamente. Se supone además que los vectores aleatorios definidos por (y_i, c_i) son al mismo tiempo independientes. Los n pares de datos son observados de la forma (y_i^*, δ_i) donde $y_i^* = \min\{y_i, c_i\}$ y $\delta_i = 1_{\{y_i \leq c_i\}}$. Entonces, si la i -ésima observación falla se tendrá que $\delta_i = 1$, y por lo tanto y_i^* corresponderá al tiempo de vida de un sujeto ($y_i^* = y_i$). De lo contrario si la i -ésima observación está censurada se tendrá que $\delta_i = 0$, y por consiguiente y_i^* corresponderá al tiempo de censura de un individuo ($y_i^* = c_i$) (Hernández, 2010).

En conclusión, de acuerdo a lo señalado con respecto a los diversos tipos de datos censurados, es posible conocer y analizar lo que sucede en ocasiones cuando existe información que está perdida por algún motivo en particular. Además mediante modelos de regresión, se podrá analizar a través de alguna distribución enfocada a tiempos de vida estos datos, siendo de interés también conocer su comportamiento.

1.2 Modelos de regresión

Los modelos de regresión se utilizan para explicar una variable respuesta en función de una o más variables explicativas. Estas herramientas son útiles para modelar observaciones relacionadas a diversos ámbitos, tales como medicina, psicología, economía, entre otros. Por ejemplo en un proceso químico, el rendimiento del producto analizado está vinculado directamente con la temperatura de operación (Montgomery, Peck y Vining, 2004).

Generalmente estas funciones matemáticas, son apropiadas para estudiar situaciones donde se cuenta con numerosas observaciones que poseen diversas variables. Sin embargo uno de los principales análisis de regresión, es determinar si las variables en conjunto causan algún tipo de efecto a la denominada variable respuesta.

Existen diversos modelos de regresión, donde cada uno presenta su particularidad de modo de tener su propia función. Entre los más utilizados se destacan los siguientes tipos: modelos lineales, cuadráticos, polinomiales, logarítmicos, generalizados, mixtos, entre otros. Sin embargo cada uno tiene su propia estructura, componentes y sus respectivos supuestos.

Cabe destacar que existen diversos tipos de modelos que se han propuesto en la literatura para describir observaciones relacionadas a tiempos de vida. Entre los más utilizados se destaca el modelo exponencial, gamma, Weibull y log-normal (Celis, 2008). No obstante las distribuciones gamma generalizada (GG) y log-gamma generalizada (LGG), son algunas de las más empleadas y beneficiosas para modelar conjuntos de datos relacionados a tiempos de falla.

El modelo de regresión log-gamma general es útil para describir observaciones enfocadas a tiempos de vida, particularmente para modelar datos con algún tipo de censura. Este modelo se origina a partir de la distribución LGG y es considerado un mecanismo eficaz para realizar estudios de supervivencia.

El objetivo general de este trabajo de titulación, es aplicar el modelo de regresión log-gamma general a un conjunto de datos reales censurados del área de la salud. Estos datos corresponden a resultados de la cobertura de acuerdo a la campaña nacional de vacunación contra la influenza del año 2016, de un determinado CESFAM del país. Sin embargo para realizar este proceso, se definirán previamente los aspectos teóricos, especificando además el tipo de censura a utilizar. Al término de este análisis, se espera obtener resultados fiables de acuerdo a esta aplicación, con el objetivo de conocer el verdadero comportamiento de las observaciones a estudiar.

La hipótesis que sustenta este trabajo, se basa principalmente en que aquel modelo de regresión describe con eficacia tiempos de vida o de falla, presentando una determinada cantidad de datos censurados. En general, se considera factible realizar estudios que demuestren cómo es posible mediante diversos modelos de regresión, describir y modelar numerosos conjuntos de observaciones con algún porcentaje de censura. Estudios de esta índole, frecuentemente son replicables en otros tipos de investigaciones enfocadas en la estimación de datos censurados, sin importar el área al cual pertenecen.

Finalmente, cabe destacar que a través de la distribución LGG es posible describir observaciones relacionadas a tiempos de vida, generando además interés en estudiar el comportamiento y la utilidad de aquella distribución. Asimismo por medio de observaciones censuradas, se logra conocer a través de un modelo de regresión lo que sucede cuando existe información que está perdida en un determinado período, utilizando alguna distribución relacionada con la temática de interés.

En el siguiente capítulo, se dará a conocer el modelo de regresión log-gamma general definiendo su estructura y sus componentes. Posteriormente, aquel modelo se aplicará a resultados de la cobertura de acuerdo a la campaña de vacunación contra la influenza correspondiente al año 2016, del CESFAM Juan Bautista Bravo Vega.

MODELO DE REGRESIÓN LOG-GAMMA GENERAL

En este capítulo se presentará el origen del modelo de regresión log-gamma general, describiendo su estructura, sus componentes y la estimación de los parámetros del modelo bajo un porcentaje de observaciones censuradas. Previamente a esto, se presentarán las funciones de densidad de las distribuciones que generan en forma cronológica el modelo de regresión.

2.1 Introducción

El modelo de regresión log-gamma general está formado a través de la distribución log-gamma generalizada. Principalmente esta función se obtiene a partir de la distribución gamma generalizada, siendo esta última, una extensión del denominado modelo gamma.

2.1.1 Historia

La función gamma se introdujo en el año 1729 junto a numerosos y destacados trabajos de Leonhard Euler (1707-1783), con el propósito de resolver diversos tipos de problemas matemáticos y analíticos. Su afán por solucionar una gran variedad de sucesiones y series infinitas, lo llevó a crear dicha función. Posteriormente, Adrien-Marie Legendre (1752-1833) propuso en el año 1814, denotar aquella función con el nombre “gamma” (Γ) (Vásquez, 1999).

La función gamma es estudiada y utilizada ampliamente, siendo ésta, una herramienta reconocida en el mundo matemático por su gran utilidad en diversos contextos. Su capacidad para resolver un sinnúmero de problemas teóricos, matemáticos y analíticos, demuestra la creciente importancia desde su origen.

Esta función es definida según Vásquez (1999) y se presenta de la siguiente forma:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} \exp\{-t\} dt, \quad z > 0. \quad (2.1)$$

Donde algunas de sus propiedades más importantes se señalan a continuación.

- (i) $\Gamma(n) = (n - 1)!$, si n es un número entero positivo.
- (ii) $\Gamma(z + 1) = z\Gamma(z)$, ($\forall z > 0$).
- (iii) $\Gamma(1) = \int_0^{\infty} \exp\{-t\} dt = 1$.
- (iv) $\Gamma(n + 1) = n!$, si n es un número entero positivo.

2.2 Distribución gamma

La distribución gamma (Casella & Berger, 2002) es empleada en diversos ámbitos tales como: análisis de supervivencia, teoría de colas, entre otros. Ésta, es utilizada desde su aparición, proporcionando importantes aplicaciones en tiempos de espera y teoría de confiabilidad (Walpole, Myres, R. y Myres, S., 1999).

La función de densidad de esta distribución se define de la siguiente forma:

Sea T una variable aleatoria que sigue una distribución gamma, su función de densidad está dada por:

$$f_T(t; k, \alpha) = \frac{1}{\alpha^k \Gamma(k)} (t)^{k-1} \exp\left\{-\frac{t}{\alpha}\right\}, \quad (2.2)$$

donde $t, k, \alpha > 0$.

Además, $\Gamma(\cdot)$ corresponde a la función gamma, k corresponde al parámetro de forma y α corresponde al parámetro de escala.

2.3 Distribución gamma generalizada

A través de la distribución gamma y con la finalidad de obtener numerosos e importantes modelos probabilísticos, se origina la denominada distribución gamma generalizada (Prentice, 1974). La cual fue introducida por primera vez en el trabajo de L. Amoroso en el año 1925 y retomada por Stacy E.W. en el año 1962 (Gutiérrez, 2010). No obstante, diversos autores definen en sus investigaciones que este importante modelo, fue presentado e introducido por Stacy E.W. en el año 1962 y ampliado por Prentice en el año 1974 (Ortega, Cancho & Paula, 2009).

La función de densidad de esta distribución se encuentra formulada y definida de la siguiente manera:

Sea T una variable aleatoria que sigue una distribución gamma generalizada. Su función de densidad con parámetros $(\alpha, \tau, k)^\top$ está dada por:

$$f_T(t; \alpha, \tau, k) = \frac{\tau}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k - 1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\}, \quad (2.3)$$

$$f_T(t; \alpha, \tau, k) = \frac{\tau}{\alpha^{\tau k} \Gamma(k)} (t)^{\tau k - 1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\tau\right\}, \quad (2.4)$$

donde $t, \alpha, \tau, k > 0$.

Además, $\Gamma(\cdot)$ corresponde a la función gamma, α corresponde al parámetro de escala y τ, k corresponden a los parámetros de forma.

Estos parámetros se encuentran definidos según Cantú, Villaseñor y Arnold (2000). Sin embargo, es importante destacar que esta función es empleada en múltiples trabajos utilizando distintas parametrizaciones, con el objetivo de facilitar diversos análisis.

La distribución gamma generalizada es considerada como un modelo flexible, debido a que ésta vincula numerosas distribuciones de probabilidad fundamentales en el área de la estadística. No obstante, mediante su capacidad de ser un modelo que incluye un tercer parámetro τ , es posible obtener diversas distribuciones cuando a sus parámetros se le asignan valores establecidos. Las principales distribuciones asociadas a este modelo, son definidas por Hashimoto, Ortega, Cancho y Cordeiro (2013) y se detallan en la tabla 2.1.

Parámetros	Distribución
$k = \tau = 1$	Exponencial
$\tau = 1$	Gamma
$k = 1$	Weibull
$k \rightarrow \infty$	Log-normal

Tabla 2.1: Distribuciones asociadas a la distribución GG.

2.4 Distribución log-gamma generalizada

La distribución log-gamma generalizada definida por Hashimoto et al. (2013), se origina a partir del modelo gamma generalizado presentado anteriormente. Esta distribución logarítmica surge por medio de una modificación de sus parámetros, proporcionando un nuevo modelo para la familia gamma.

Uno de los fundamentales problemas que se presentan al estudiar esta distribución, consiste en la estimación del parámetro de forma. Así, mediante una transformación se crea este importante modelo; siendo su principal objetivo, verificar si diversos conjuntos de observaciones poseen un comportamiento log-gamma generalizado (Gutiérrez, 2010).

A través de esta distribución y mediante su capacidad para modelar datos, es posible describir grupos de observaciones relacionados a tiempos de vida, tiempos de supervivencia o tiempos de falla (dependiendo de la aplicación a realizar). Por esta razón, este modelo probabilístico es considerado como una familia flexible de distribuciones útiles, para analizar conjuntos de datos en diferentes áreas de la ciencia y tecnología (Agostinelli, Locatelli, Marazzi & Yohai, 2015). Siendo además eficaz, para describir observaciones con algún porcentaje de censura.

La función de densidad de esta distribución es representada y definida de la siguiente forma:

Sea T una variable aleatoria que sigue una distribución GG. La variable aleatoria $Y = \log(T)$, sigue una distribución LGG con una función de densidad parametrizada en términos de:

$$\begin{aligned}\mu &= \log(\alpha) + \tau^{-1} \log(\lambda^{-2}), \\ \sigma &= (\tau\sqrt{k})^{-1}, \\ \lambda &= (\sqrt{k})^{-1}.\end{aligned}$$

Luego, la función de densidad de la distribución log-gamma generalizada se define de la siguiente manera:

$$f_Y(y; \lambda, \sigma, \mu) = \begin{cases} \frac{|\lambda|(\lambda^{-2})^{\lambda-2}}{\sigma\Gamma(\lambda^{-2})} \exp \left\{ \lambda^{-2} \left[\left(\frac{y-\mu}{\sigma} \right) \lambda - \exp \left\{ \left(\frac{y-\mu}{\sigma} \right) \lambda \right\} \right] \right\} & \text{si } \lambda \neq 0, \\ \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right\} & \text{si } \lambda = 0, \end{cases} \quad (2.5)$$

donde $-\infty < y < \infty$, $-\infty < \lambda < \infty$, $-\infty < \mu < \infty$ y $\sigma > 0$.

Además, $\Gamma(\cdot)$ corresponde a la función gamma, λ corresponde al parámetro de forma, σ corresponde al parámetro de escala y μ corresponde al parámetro de localización.

Por otro lado, sea U una variable aleatoria estandarizada y definida por:

$$U = \frac{Y - \mu}{\sigma}. \quad (2.6)$$

Luego U sigue una distribución log-gamma generalizada, donde su función de densidad puede ser expresada como:

$$f_U(u; \lambda, \sigma, \mu) = \begin{cases} \frac{|\lambda|(\lambda^{-2})^{\lambda-2}}{\Gamma(\lambda^{-2})} \exp \{ \lambda^{-1}u - \lambda^{-2} \exp \{ \lambda u \} \} & \text{si } \lambda \neq 0, \\ \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{u^2}{2} \right\} & \text{si } \lambda = 0. \end{cases} \quad (2.7)$$

Por otra parte, de acuerdo a lo señalado anteriormente con respecto a la flexibilidad que posee este modelo, las principales distribuciones asociadas a esta función, cuando a los parámetros se le asignan valores establecidos según Agostinelli, Marazzi, Yohai y Randriamiharisoa (2016), se detallan en la tabla 2.2.

Parámetros	Distribución
$\lambda = \sigma = 1$	Log-exponencial
$\lambda = \sigma$	Log-gamma
$\lambda = 1$	Log-Weibull
$\lambda = 0$	Normal

Tabla 2.2: Distribuciones asociadas a la distribución LGG.

A continuación en la figura 2.1, se presenta la gráfica de las funciones de densidad de las distribuciones vinculadas a la familia log-gamma generalizada, detalladas en la tabla 2.2. Esta representación gráfica, se realizó mediante datos simulados a partir de una distribución log-gamma generalizada de parámetros $\lambda = 2$ y $\sigma = 2$. Sin embargo, para generar las distribuciones restantes se utilizó $\lambda = 0$ y $\sigma = 1$ para la distribución normal, $\lambda = 1$ y $\sigma = 2$ para la distribución log-Weibull, y por último $\lambda = 1$ y $\sigma = 1$ para la distribución log-exponencial. Los valores de los parámetros fueron establecidos por Agostinelli et al. (2016).

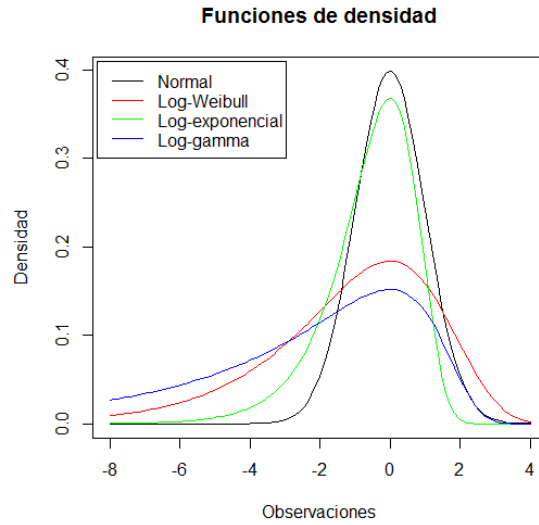


Figura 2.1: Distribuciones asociadas a la familia LGG. Adaptado de Agostinelli et al. (2016).

A pesar de que las funciones de densidad de las variables U e Y presentadas en las expresiones (2.7) y (2.5) respectivamente, se encuentran definidas según Hashimoto et al. (2013); la escritura de la parametrización de ambas relaciones serán establecidas según Agostinelli et al. (2015), quedando reformuladas de la siguiente manera:

$$f_U(u; \lambda, \sigma, \mu) = f_\lambda(u), \quad (2.8)$$

$$f_Y(y; \lambda, \sigma, \mu) = f_\theta(y) = \frac{1}{\sigma} f_\lambda \left(\frac{y - \mu}{\sigma} \right), \quad (2.9)$$

donde $\theta = (\lambda, \sigma, \mu)$.

Mediante el siguiente teorema definido por Hashimoto et al. (2013), se presentarán los momentos de la distribución log-gamma generalizada. De este modo, se darán a conocer algunas de las características más importantes de la distribución, como por ejemplo: tendencia, dispersión, asimetría y curtosis.

Teorema 2.4.1. Si $Y \sim \text{LGG}(\lambda, \sigma, \mu)$, el r -ésimo momento de Y para $\lambda > 0$ está dado por:

$$\mu'_r = E(Y^r) = \frac{1}{\Gamma(w)} \sum_{j=0}^r \binom{r}{j} \left[2\sigma \log(w^{-1/2})^{w^{1/2}} + \mu \right]^{r-j} \sigma^j w^{j/2} \Gamma^{(j)}(w), \quad (2.10)$$

donde $w = \lambda^{-2}$, $\Gamma^{(j)}(w) = \frac{\partial^j \Gamma(w)}{\partial w^j}$.

El r -ésimo momento de la distribución log-gamma generalizada se define como sigue:

$$\mu'_r = \int_{-\infty}^{\infty} \frac{y^r \lambda}{\sigma \Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp \left\{ \lambda^{-2} \left[\left(\frac{y - \mu}{\sigma} \right) \lambda - \exp \left\{ \left(\frac{y - \mu}{\sigma} \right) \lambda \right\} \right] \right\} dy. \quad (2.11)$$

Al ajustar $x = \lambda^{-2} \exp \left\{ \lambda \left(\frac{y - \mu}{\sigma} \right) \right\}$, μ'_r se reduce y adquiere la siguiente forma:

$$\mu'_r = \frac{1}{\Gamma(\lambda^{-2})} \int_0^{\infty} \left\{ \frac{\sigma}{\lambda} [\log(x) + 2 \log(\lambda)] + \mu \right\}^r x^{\lambda^{-2}-1} \exp \{-x\} dx. \quad (2.12)$$

Entonces, al aplicar la expansión binomial en la ecuación (2.12) y luego al utilizar operaciones algebraicas, se tiene lo siguiente:

$$\mu'_r = \frac{1}{\Gamma(\lambda^{-2})} \sum_{j=0}^r \binom{r}{j} [2\sigma \log(\lambda^{1/\lambda}) + \mu]^{r-j} \left(\frac{\sigma}{\lambda} \right)^j \int_0^{\infty} x^{\lambda^{-2}-1} \exp \{-x\} [\log(x)]^j dx, \quad (2.13)$$

donde la integral de la expresión (2.13) puede ser definida como:

$$\int_0^{\infty} x^{\lambda^{-2}-1} \exp \{-x\} [\log(x)]^j dx = \frac{\partial^j \Gamma(\lambda^{-2})}{\partial (\lambda^{-2})^j}. \quad (2.14)$$

Posteriormente, al sustituir el resultado en la ecuación (2.13) y al reemplazar $w = \lambda^{-2}$, es posible obtener el siguiente término:

$$\mu'_r = \frac{1}{\Gamma(w)} \sum_{j=0}^r \binom{r}{j} \left[2\sigma \log(w^{-1/2})^{w^{1/2}} + \mu \right]^{r-j} \sigma^j w^{j/2} \Gamma^{(j)}(w). \quad (2.15)$$

Por lo tanto, para $\lambda > 0$ se define:

$$(i) \ E(Y) = \mu + \sigma \lambda^{-1} [\phi(\lambda^{-2}) - \log(\lambda^{-2})],$$

$$(ii) \ \text{Var}(Y) = \lambda^{-2} \sigma^2 \phi'(\lambda^{-2}),$$

donde $\phi(\cdot)$ y $\phi'(\cdot)$ corresponden a las funciones digamma y trigamma respectivamente.

2.5 Estimación de parámetros de la distribución LGG

La estimación de los parámetros con respecto al modelo de regresión y a la distribución LGG, se realizará a través del estimador ponderado de verosimilitud (para el caso de datos censurados), definido por Agostinelli et al. (2015) y establecido adicionalmente por Agostinelli, Locatelli, Marazzi y Yohai (2017).

Sea y_1, \dots, y_n n muestras i.i.d. (independientes e idénticamente distribuidas), con función de distribución acumulada $F_{\theta_0}(y)$, donde $\theta_0 = (\lambda_0, \sigma_0, \mu_0)$. Para estimar el parámetro θ_0 , se consideran observaciones censuradas por la derecha, donde los verdaderos valores de y_1, \dots, y_n no son observados directamente. Sin embargo se sabe que $y_i^* = \min(y_i, c_i)$ ($i = 1, \dots, n$), corresponden a datos completamente observados definidos en la función 1.1, donde c_1, \dots, c_n son denominados tiempos censurados i.i.d.

Se establece adicionalmente el indicador de censura detallado en el capítulo 1, donde $\delta_i = 1$ si $y_i^* = y_i$ y $\delta_i = 0$ si $y_i^* = c_i$. Considerando además que $z_i = (y_i^*, \delta_i)$, y G_n corresponde a la función de distribución empírica basada en (z_1, \dots, z_n) .

2.5.1 Estimador de máxima verosimilitud

El estimador de máxima verosimilitud (EMV) de acuerdo a los parámetros de la distribución LGG (para el caso de observaciones censuradas), puede ser escrito como sigue.

Sea $S_{\theta}(y)$ la función de supervivencia definida como:

$$S_{\theta}(y) = 1 - F_{\theta}(y) = 1 - F_{\lambda} \left(\frac{y - \mu}{\sigma} \right). \quad (2.16)$$

Entonces, la función negativa de log-verosimilitud está dada por:

$$- \sum_{i=1}^n [\delta_i \log f_{\theta}(y_i^*) + (1 - \delta_i) \log S_{\theta}(y_i^*)]. \quad (2.17)$$

En ausencia de censura, se consideran las funciones *score* $d = (d_1, d_2, d_3)^\top$, las cuales son establecidas de la siguiente forma:

$$d_1(y, \theta) = -\frac{\partial}{\partial \mu} \log f_\theta(y) = \frac{1}{\sigma} \xi_\lambda(u), \quad (2.18)$$

$$d_2(y, \theta) = -\frac{\partial}{\partial \sigma} \log f_\theta(y) = \frac{1}{\sigma} (\xi_\lambda(u)u + 1), \quad (2.19)$$

$$d_3(y, \theta) = -\frac{\partial}{\partial \lambda} \log f_\theta(y) = \psi_\lambda(u), \quad (2.20)$$

donde $u = \frac{y - \mu}{\sigma}$, $\dot{F}_\theta(y) = \dot{F}_\lambda(u) = \frac{dF_\lambda(u)}{d\lambda}$,

$$\xi_\lambda(u) = \frac{f'_\lambda(u)}{f_\lambda(u)} = \frac{(1 - \exp(\lambda u))}{\lambda},$$

$$\psi_\lambda(u) = -\frac{\partial}{\partial \lambda} \log f_\lambda(u) = \frac{1}{\lambda^3} (2\zeta(\lambda) - \lambda^2 + \lambda u - \exp(\lambda u)(2 - \lambda u))$$

y $\zeta(\lambda) = -2\log(\lambda) - \phi(\lambda^{-2}) + 1$. Considerando además que $\phi(\cdot)$ corresponde a la función digamma.

Adicionalmente, se especifican aquellas tres expresiones:

$$s_1(y, \theta) = -\frac{\partial}{\partial \mu} \log S_\theta(y) = -\frac{1}{\sigma} \frac{f_\lambda(u)}{S_\lambda(u)}, \quad (2.21)$$

$$s_2(y, \theta) = -\frac{\partial}{\partial \sigma} \log S_\theta(y) = -\frac{1}{\sigma} \frac{f_\lambda(u)u}{S_\lambda(u)}, \quad (2.22)$$

$$s_3(y, \theta) = -\frac{\partial}{\partial \lambda} \log S_\theta(y) = \frac{\dot{F}_\lambda(u)}{S_\lambda(u)}. \quad (2.23)$$

Por lo tanto, las funciones *score* para el caso de datos censurados se definen como:

$$v_k(y, \delta, \theta) = \delta d_k(y, \theta) + (1 - \delta) s_k(y, \theta), \quad k = 1, 2, 3. \quad (2.24)$$

Finalmente, el EMV de θ está dado por el siguiente sistema de ecuaciones:

$$E_{Gn}(v(y, \delta, \theta)) = \frac{1}{n} \sum_{i=1}^n v(y_i^*, \delta_i, \theta) = 0, \quad (2.25)$$

donde $v = (v_1, v_2, v_3)^\top$ corresponde al vector de funciones *score*.

Frente a este escenario, es posible demostrar que:

$$-\nabla_{\theta} \log S_{\theta}(y_i^*) = E_{\theta} (d(y, \theta) | y > y_i^*), \quad (2.26)$$

donde ∇_{θ} indica la diferenciación con respecto al parámetro θ .

En consecuencia, una expresión alternativa para las ecuaciones de verosimilitud puede ser expresada de la siguiente manera:

$$\frac{1}{n} \sum_{i=1}^n \delta_i d(y_i^*, \theta) + (1 - \delta_i) E_{\theta} (d(y, \theta) | y > y_i^*) = 0. \quad (2.27)$$

Por otra parte, Locatelli, Marazzi y Yohai (2011) citados por Agostinelli et al. (2015), establecen la función de distribución acumulada semiempírica de y para θ , la cual está representada como:

$$H_{n,\theta}(z) = \frac{1}{n} \sum_{i=1}^n E_{\theta} [I(y \leq z) | y_i^*, \delta_i], \quad (2.28)$$

o de manera equivalente:

$$H_{n,\theta}(z) = \frac{1}{n} \sum_{i=1}^n \delta_i I(y_i \leq z) + \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \frac{[F_{\theta}(z) - F_{\theta}(y_i^*)]^+}{1 - F_{\theta}(y_i^*)}. \quad (2.29)$$

Por lo tanto, cuando no existen observaciones censuradas, $H_{n,\theta}(z)$ coincide con la tradicional función de distribución acumulada empírica. Por consiguiente, si $\tilde{\theta}$ es un estimador consistente de θ_0 , entonces $H_{n,\tilde{\theta}}(z)$ es un estimador consistente de $F_{\theta_0}(z)$; y para cualquier función medible $h(y)$, se tiene que:

$$\lim_{n \rightarrow \infty} E_{n,\tilde{\theta}} [h(y)] = E_{\theta_0} [h(y)], \quad (2.30)$$

donde $E_{n,\theta}$ indica la expectación bajo $H_{n,\theta}$.

Finalmente, es posible determinar otra expresión que hace referencia a las ecuaciones de verosimilitud, definida tal como se indica:

$$E_{n,\theta} (d(y, \theta)) = 0. \quad (2.31)$$

Adicionalmente se considera que:

$$M(\theta) = E(v(y, \delta, \theta)v(y, \delta, \theta)^\top) \text{ y } G(\theta) = E(\nabla_\theta v(y, \delta, \theta)), \quad (2.32)$$

donde la matriz de covarianza asintótica del EMV se presenta de la siguiente forma:

$$\Sigma(\theta_0) = G(\theta_0)^{-1}M(\theta_0)G(\theta_0)^{-\top}. \quad (2.33)$$

2.5.2 Estimador ponderado de verosimilitud

Markatou, Basu y Lindsay (1998) citados por Agostinelli et al. (2015), introdujeron los estimadores ponderados de verosimilitud para el caso de datos no censurados en modelos continuos. Sin embargo Agostinelli et al. (2015), definen el estimador ponderado de verosimilitud (EPV) para el caso de observaciones censuradas, tal como se indica a continuación.

Primordialmente, se asume que $w = (y, \theta)$ corresponde a una función peso y $\tilde{\theta}$ es un estimador altamente robusto y consistente, pero no necesariamente eficiente de θ_0 .

Por lo tanto, se define el estimador ponderado de verosimilitud para el caso de datos censurados, como una solución del siguiente sistema de ecuaciones:

$$u(y^*, \delta, \theta) = \sum_{i=1}^n w(y_i^*, \theta)v(y_i^*, \delta_i, \theta) = 0, \quad (2.34)$$

donde $v(y_i^*, \delta_i, \theta)$ corresponde a la expresión definida en la ecuación 2.25.

Entonces cuando $w(y^*, \theta) \equiv 1$, la expresión 2.34 coincide exactamente con el sistema de ecuaciones 2.25.

Por otra parte, Markatou et al. (1998) citados por Agostinelli et al. (2015), establecen aquella función peso, tal como sigue:

$$w(y, \theta) = \min\left(1, \frac{[A(r_p(y, \theta)) + 1]^+}{r_p(y, \theta) + 1}\right), \quad (2.35)$$

donde $[x]^+ = \max(0, x)$ y $r_p(y, \theta)$ corresponde a una variación conveniente del residuo de Pearson, que mide la similitud entre los datos observados y el modelo propuesto; considerando además Lindsay (1994) citado por Agostinelli et al. (2015), quien establece que $A(\cdot)$ corresponde a la función del ajuste residual.

Cabe destacar que para definir los residuos de Pearson, se debe considerar que:

Sea $H_{n,\theta}$ la función de distribución acumulada semiparamétrica especificada en la expresión 2.29. Sea además $f_{n,\theta}^*(y) = \int k(y, t, h) dH_{n,\theta}$, la densidad de *kernel* semiparamétrica del estimador de $f_\theta(y)$, con ancho de banda h ; y por último, sea $f_\theta^*(y) = \int k(y, t, h) f_\theta(t) dt$, el correspondiente modelo de densidad suavizado.

Por lo tanto, los residuos de Pearson se definen de la siguiente forma:

$$r_p(y, \theta) = \frac{[f_{n,\theta}^*(y) - f_\theta^*(y)]}{f_\theta^*(y)}. \quad (2.36)$$

Finalmente, Markatou et al. (1998) y además Agostinelli, Locatelli, Marazzi y Yohai (2014) citados por Agostinelli et al. (2015), definen el estimador ponderado de verosimilitud de un paso (1EPV). El cual se obtiene al aplicar una iteración del método de Newton-Raphson en la ecuación 2.34. Sin embargo aquel estimador está dado por:

$$\hat{\theta} = \tilde{\theta} - J^{-1}u(y^*, \delta, \tilde{\theta}), \quad (2.37)$$

donde $J = \sum_{j=1}^n w(y_j^*, \tilde{\theta}) \nabla_{\theta} v(y_j^*, \delta_j, \theta)|_{\theta=\tilde{\theta}}$ es la matriz Jacobiana de las funciones definidas de acuerdo a las ecuaciones de las estimaciones, y ∇_{θ} corresponde a la diferenciación con respecto al parámetro θ .

2.6 Modelo de regresión log-gamma general

El modelo de regresión de tiempo de fallo acelerado (AFT) está definido según Agostinelli et al. (2017), el cual se presenta a continuación.

Sea $\mathbf{x} = (x_1, \dots, x_p)^\top$ el vector de variables explicativas asociadas a la variable respuesta y^{**} . Basado en la densidad de la distribución LGG, es posible construir un modelo de regresión que enlaza la variable respuesta y_i^* y el vector de variables explicativas \mathbf{x}_i , donde $\mathbf{x}_i \in \mathbb{R}^p$ e y_i^{**} satisfacen lo siguiente:

$$y_i^{**} = \mu_0 + \beta_0^\top \mathbf{x}_i + \sigma_0 u_i, \quad i = 1, \dots, n. \quad (2.38)$$

Usualmente y_i^{**} , representa el tiempo de supervivencia o de censura en escala logarítmica para la i -ésima unidad experimental; y $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$, representa al vector de covariables del modelo para la i -ésima unidad experimental. Los parámetros $\beta_0^\top = (\beta_1, \dots, \beta_p)$, σ_0 y el

intercepto μ_0 , corresponden a los parámetros desconocidos de aquel modelo que deben ser estimados. Sin embargo los u_i ($1 \leq i \leq n$), son denominados errores aleatorios, los cuales son i.i.d y además son independientes de \mathbf{x}_i . Cabe destacar que aquellos errores siguen una distribución $f_{(0,1,\lambda_0)}$, donde λ_0 es el parámetro de forma, y f_θ está definida en la expresión 2.7.

2.7 Estimación de parámetros del modelo de regresión

2.7.1 Estimador ponderado de verosimilitud

Para estimar los parámetros del modelo de regresión definido en la expresión 2.38, se establece el estimador ponderado de verosimilitud completamente iterado, de acuerdo a las soluciones de las siguientes ecuaciones:

$$u(\mathbf{X}, \mathbf{y}^*, \delta, \gamma) = \sum_{j=1}^n \check{w}(\mathbf{x}_j, \mathbf{y}_j^*, \gamma) v(\mathbf{x}_j, \mathbf{y}_j^*, \delta_j, \gamma) = 0, \quad (2.39)$$

donde $\gamma = (\theta, \beta)$ y $\check{w}(\mathbf{x}_j, \mathbf{y}_j^*, \gamma) = w(\mathbf{y}_j^* - \mathbf{x}_j^\top \beta, \theta)$. Considerando además que w está definida en la expresión 2.35.

Entonces, el estimador ponderado de verosimilitud de un paso (para el caso de observaciones censuradas), está dado por:

$$\hat{\gamma} = \tilde{\gamma} - J^{-1} u(\mathbf{X}, \mathbf{y}^*, \delta, \tilde{\gamma}), \quad (2.40)$$

donde $J = \sum_{i=1}^n \check{w}(\mathbf{x}_i, \mathbf{y}_i, \tilde{\gamma}) \nabla_\gamma v(\mathbf{x}_i, \mathbf{y}_i^*, \tilde{\gamma})|_{\gamma=\tilde{\gamma}}$ es una matriz Jacobiana $(3+p) \times (3+p)$.

Finalmente, se destaca que las ponderaciones son establecidas de acuerdo a los residuos de Pearson; comparando $F_{\tilde{\theta}^*}$ con el estimador de $H_{n, \tilde{\theta}^*}$ señalado en la expresión 2.29. No obstante aquellas observaciones se determinan de la siguiente forma:

$$(\omega_i^*, \delta_i) = (\mathbf{y}_i^* - \mathbf{x}_i^\top \bar{\beta}, \delta_i), \quad 1 \leq i \leq n. \quad (2.41)$$

En resumen, cabe señalar que aquel capítulo fue definido por Agostinelli et al. (2015), y actualizado adicionalmente por Agostinelli et al. (2017); considerando que algunas notaciones fueron establecidas según Agostinelli et al. (2016).

En el siguiente capítulo, se realizará la aplicación del modelo propuesto, utilizando un conjunto de observaciones censuradas aleatoriamente por la derecha. Sin embargo, los parámetros de aquel modelo y de la distribución LGG, serán estimados a través del 1EPV.

APLICACIÓN

3.1 Introducción

En este capítulo se presentará el ajuste del modelo de regresión log-gamma general, al utilizar los resultados de la cobertura de acuerdo a la campaña de vacunación contra la influenza del año 2016, correspondiente a las mujeres embarazadas a partir de las 13 semanas de gestación, pertenecientes al CESFAM JBBV. Previamente a esto, se presentará un análisis exploratorio para conocer el comportamiento de las observaciones utilizadas.

3.1.1 Antecedentes generales

Anteriormente, se mencionó que el presente estudio estaría basado de acuerdo al modelo de censura aleatoria por la derecha definido en la expresión 1.1 del capítulo 1, el cual representa al conjunto de datos censurados que se utilizarán para llevar a cabo la aplicación del modelo de regresión.

El concepto “cobertura vacunal”, hace referencia a la proporción de individuos de una determinada localidad que reciben alguna inmunización específica (OMS, 2016a). Sin embargo en este análisis, la cobertura está relacionada directamente con las personas inscritas en este CESFAM que debían ser inmunizadas contra la influenza; y con aquellos usuarios que se les administró la vacuna. En otras palabras, este término corresponde al cociente entre los individuos pertenecientes a este centro de salud que recibieron la vacuna antigripal; y aquellos usuarios inscritos en aquel establecimiento que debían ser inmunizados, expresado en términos porcentuales.

Generalmente en este tipo de campañas de vacunación, no es posible inmunizar a la totalidad de personas establecidas durante los meses en que éstas se extienden. Esto se debe principalmente a diversos inconvenientes que manifiestan los individuos durante aquel período.

Durante el año 2016, la campaña preventiva se extendió desde el 14 de marzo hasta el 30 de junio; mientras que en aquel centro de salud, se comenzó a inmunizar a los individuos desde el 10 de marzo. Se tiene conocimiento que diversos usuarios no fueron inoculados en el tiempo establecido, sin embargo aquellas personas podían recibir la inmunización después de estas fechas, dependiendo de la cantidad de vacunas disponibles.

Frente a este escenario, se concluye efectivamente que la información analizada posee un alto porcentaje de observaciones censuradas. No obstante en estudios de esta índole, generalmente se debe fijar el período de seguimiento de las observaciones, considerando además el evento de interés. Específicamente, en este análisis, la duración del estudio correspondió a los meses en que se extendió la campaña (112 días); siendo el evento de interés, la inmunización del usuario.

Por consiguiente, de acuerdo a lo establecido con respecto a la duración del estudio es posible definir lo siguiente: en el caso de que un individuo sea inmunizado, esta observación corresponderá a un dato completo; mientras que en el caso de que un individuo no haya sido inmunizado, esta observación corresponderá a un dato censurado.

En esta última oportunidad, la vacuna podría administrarse solamente después de que ha finalizado la campaña; sin embargo, sólo se sabrá, que aquel acontecimiento ocurrirá posterior al 30 de junio. Es decir, evidentemente en este caso se produce el fenómeno de censura por la derecha. Además, es posible corroborar la existencia de censura aleatoria, debido a que diversos usuarios presentaron otros acontecimientos independientes a la inmunización, evitando la administración de la vacuna.

3.2 Base de datos

Para llevar a cabo la aplicación del estudio se utilizaron tres conjuntos de bases de datos. El primer grupo corresponde a los registros de las personas inmunizadas contra la influenza de acuerdo al año 2016 del CESFAM Juan Bautista Bravo Vega; el segundo grupo, corresponde a los registros de los usuarios inscritos en este centro de atención primaria; y el tercer grupo, corresponde a los registros de las mujeres embarazadas pertenecientes a este establecimiento. Específicamente, los tres conjuntos fueron obtenidos por medio de aquel

centro de salud asistencial, a través del Registro Nacional de Inmunizaciones¹ (RNI), mediante la Red Asistencial y Enlace Nacional² (RAYEN) y a través del libro de registro de maternidad, respectivamente.

Las bases de datos correspondientes al RNI están compuestas por variables como: nombre de la persona, Rol Único Nacional (RUN), edad, sexo, comuna, fecha de administración de la vacuna antigripal, rechazo a la inmunización, entre otras. En cambio la información correspondiente a la RAYEN está constituida por numerosas variables, donde algunas son las siguientes: nombre del usuario, RUN, fecha de nacimiento, edad, sexo, fecha de inscripción al centro de salud, etc. Sin embargo, cabe destacar que las bases de datos correspondientes al libro de registro de maternidad se presentan en forma física, donde se incluyen variables como: nombre de la persona, edad, edad gestacional, estado nutricional, resultados de exámenes de salud, entre otras. Por lo tanto, para obtener información precisa de acuerdo a la cobertura de la campaña de inmunización, se efectuó un exhaustivo proceso de administración de datos.

3.2.1 Administración de datos

Inicialmente se realizó una limpieza de las bases de datos a través del software estadístico Stata (2011), donde principalmente se verificó que las personas inmunizadas de acuerdo al RNI, correspondieran a los usuarios inscritos en la RAYEN y en el libro de registro de maternidad; corroborando además que los datos de estos individuos, coincidieran con determinados antecedentes válidos. Posteriormente después de comprobar esta información, modificar registros y variables, entre otros procedimientos; se contabilizó un total de 144 observaciones. Estas observaciones, corresponden a las mujeres embarazadas pertenecientes a este CESFAM que debían ser inmunizadas contra la influenza, enfatizando a aquellas que se les administró la vacuna antigripal.

3.2.2 Variables

Las variables consideradas en el estudio se definen a continuación.

i) Variable respuesta:

- Y^* : Tiempo transcurrido desde la entrada al estudio de la persona hasta el término de su seguimiento (días).

¹El Registro Nacional de Inmunizaciones corresponde a un sistema informático que tiene por finalidad registrar las vacunas aplicadas de acuerdo al PNI (DEIS, 2016).

²La Red Asistencial y Enlace Nacional corresponde a un mecanismo de información, que tiene por objetivo apoyar la gestión clínica y administrativa de diversos centros de salud (SAYDEX, 2016).

ii) Variables explicativas:

- X_1 : Edad en años que posee la persona al comienzo de la campaña de vacunación.
- X_2 : Edad gestacional en semanas que posee la embarazada al comienzo de la campaña de vacunación.
- X_3 : Cantidad de inmunizaciones contra la influenza recibidas entre los años 2010 y 2015.

Además, como se trabajó con un conjunto de datos censurados, se incluyó una variable de censura (la cual indica si la observación está completa o censurada).

3.3 Análisis exploratorio

A continuación, se presentan los resultados obtenidos a través de tablas de resumen y gráficos estadísticos entregados por el software R-project (2016).

Datos	Observaciones	Porcentaje (%)
No censurados	72	50.00
Censurados por pérdida de seguimiento	5	3.47
Censurados por no inmunización	67	46.53
Total	144	100.00

Tabla 3.1: Descripción de los tipos de datos a utilizar.

En la tabla 3.1 se detalla la cantidad de observaciones censuradas y no censuradas presentes en este estudio. Específicamente, los resultados indican que de las 144 observaciones, 72 corresponden a datos completos; es decir evidentemente se destaca que la censura está presente en el 50 % de los casos. La mayor parte de ella (46.53 %), se produjo debido a que las embarazadas no se inmunizaron durante los meses en que se extendió la campaña; mientras que el 3.47 % restante, corresponde a mujeres que sufrieron un aborto, o bien que se trasladaron de CESFAM antes que recibieran la vacuna antigripal.

Variable	Media	Desviación estándar	Mediana	Mínimo	Máximo	Coef. de variación
Tiempo de seguimiento (días)	43.29	35.27	33.00	1.00	112.00	0.81
Edad (años)	28.13	6.48	28.00	13.00	45.00	0.23
Edad gestacional (semanas)	10.86	7.78	10.86	0.00	24.00	0.72
Inmunizaciones influenza	0.72	1.12	0.00	0.00	6.00	1.56
Censura	0.50	0.50	0.50	0.00	1.00	1.00

Tabla 3.2: Resumen descriptivo de las variables de interés.

En la tabla 3.2 se presenta una descripción de las variables a estudiar, en la cual se muestran medidas de tendencia central, así como también medidas de dispersión.

Estos resultados indican que el tiempo promedio desde el inicio del seguimiento de las embarazadas (día en que comenzó la campaña de vacunación, instante en que ingresaron al CESFAM, o día en que cumplieron 13 semanas de gestación); específicamente hasta el término de sus respectivos seguimientos, es aproximadamente 43 días. Considerando además que el 46.53 % de aquellas mujeres, no recibieron la vacuna antigripal hasta el 30 de junio.

Por otro lado, se visualiza que las gestantes tienen en promedio 28 años de edad indicando, a la vez, que presentan un embarazo de 11 semanas de gestación al inicio de la campaña preventiva. Adicionalmente, estas mujeres en promedio fueron inmunizadas contra la influenza estacional una vez, aproximadamente entre los años 2010 y 2015.

Al seguir con estos resultados, se destaca que las observaciones clasificadas por cada variable (a excepción de las edades), presentan un significativo porcentaje de variación. Esto se debe principalmente a que sólo la edad (en años) de las embarazadas, fluctúan de forma homogénea en el rango en que éstas se encuentran. No obstante las dispersiones del resto de las variables, se producen debido a la gran cantidad de datos censurados presentes en el estudio.

Cabe destacar que la variable respuesta debe transformarse a la escala logarítmica, sin embargo en la tabla 3.3, se presenta un resumen descriptivo del logaritmo del tiempo de seguimiento; donde se comprueba que la variación de las observaciones se estabiliza, y aquellos datos poseen un mayor grado de homogeneidad. Esto, es corroborado a través de la discrepancia que existe entre los valores de los coeficientes de variación de la variable respuesta y de su transformación.

Como se muestra en la tabla 3.2, el valor de aquel coeficiente de acuerdo al tiempo de seguimiento (en días), es de 0.81; mientras que la tabla 3.3, indica que el tiempo de seguimiento (en escala logarítmica), posee un coeficiente de variación de 0.34.

Variable	Media	Desviación estándar	Mediana	Mínimo	Máximo	Coef. de variación
Logaritmo tiempo de seguimiento	3.31	1.13	3.50	0.00	4.72	0.34

Tabla 3.3: Resumen estadístico del logaritmo de la variable respuesta.

Para efectos de comparación, en la figura 3.1 se muestran dos histogramas que hacen referencia a las distribuciones del tiempo de seguimiento y del logaritmo tiempo de seguimiento, ajustadas a sus respectivas densidades.

Debido al alto porcentaje de censura, se observa que la primera distribución no decrece completamente a medida que el valor de la variable aumenta. Por esta razón si se considera una menor cantidad de observaciones censuradas, sería posible obtener una distribución asimétrica positiva conocida, como es el caso del modelo gamma. Por lo tanto al aplicar logaritmo a aquellos tiempos de seguimiento, se visualiza que estos datos evidentemente son asimétricos hacia la izquierda, teniendo en cuenta la posibilidad de que aquellas observaciones sigan una distribución log-gamma generalizada.

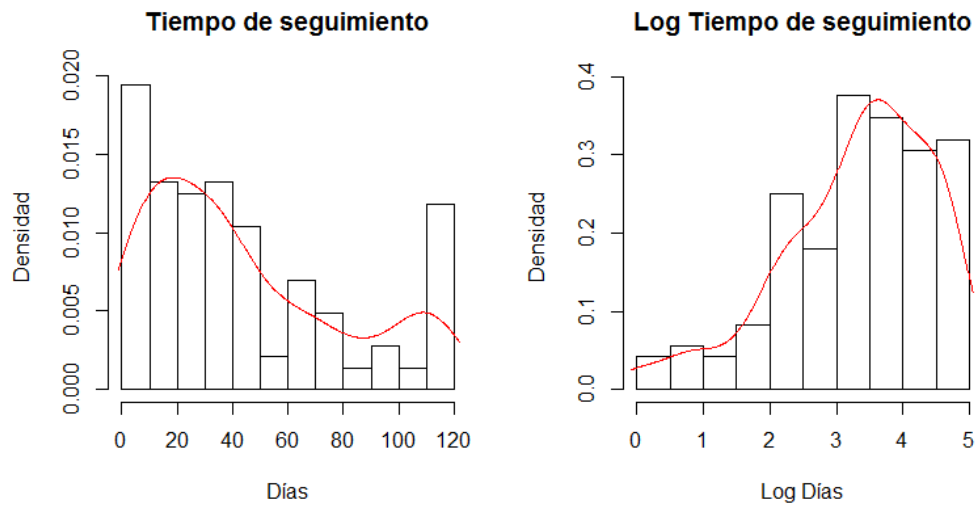


Figura 3.1: Distribución del tiempo de seguimiento y del log tiempo de seguimiento ajustadas a sus respectivas densidades.

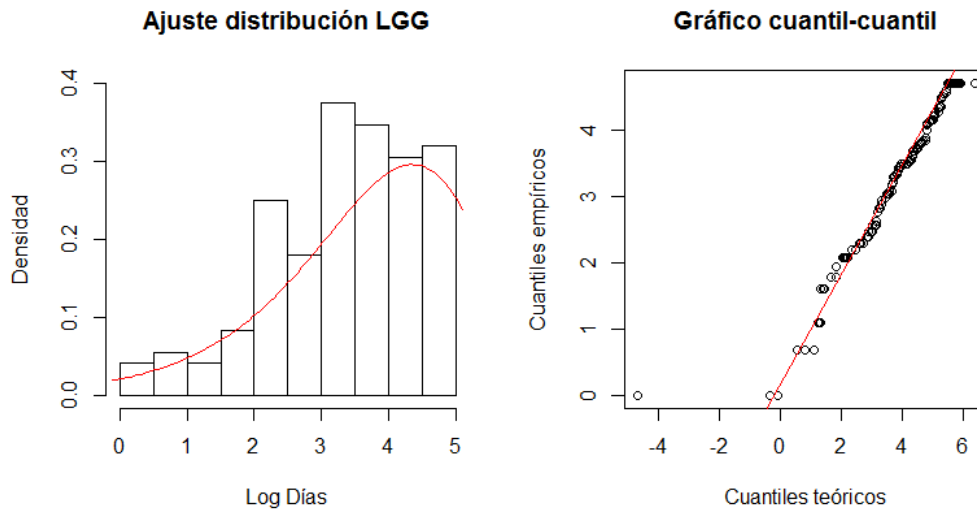


Figura 3.2: Distribución del log tiempo de seguimiento ajustada a una distribución LGG.

Para verificar si la variable respuesta sigue una distribución LGG, se realizó el ajuste de aquella variable a la respectiva distribución. Como se muestra en la figura 3.2, los datos aparentemente se ajustan al modelo log-gamma generalizado. No obstante, al proyectar la curva de densidad de la distribución teórica, es posible concluir que los tiempos de seguimiento (en escala logarítmica), poseen un comportamiento log-gamma generalizado. Esto puede ser corroborado a través del gráfico cuantil-cuantil; el cual indica claramente que, a pesar de que las observaciones en estudio presentan una mínima desviación en los extremos, la mayoría de los datos se ajustan considerablemente a la distribución teórica.

3.4 Aplicación del modelo de regresión log-gamma general

El modelo de regresión propuesto en la expresión 2.38 del capítulo 2, se implementó a través del software estadístico R-project (2016). Al ser aplicado a las observaciones analizadas, el modelo puede ser definido como sigue:

$$y_i^{**} = \mu_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma_0 u_i, \quad i = 1, \dots, 144. \quad (3.1)$$

A continuación, se presentan los resultados obtenidos mediante la función `loggamma-censlmrob` del paquete `robustloggamma` (2016). Estos resultados, corresponden a la estimación de los parámetros del modelo, considerando a la vez sus respectivos residuos.

Parámetro	Estimación	Error estándar	valor- p
$\hat{\mu}_0 \equiv \hat{\mu}_{\text{Intercepto}}$	6.0314	0.8178	<0.0010
$\hat{\beta}_1 \equiv \hat{\beta}_{\text{Edad}}$	-0.0579	0.0216	0.0074
$\hat{\beta}_2 \equiv \hat{\beta}_{\text{Edad gestacional}}$	-0.0145	0.0197	0.4609
$\hat{\beta}_3 \equiv \hat{\beta}_{\text{Inmunizaciones influenza}}$	-0.1193	0.1216	0.3266
$\hat{\lambda}$	0.1437	0.4881	0.7685
$\hat{\sigma}$	1.3812	0.2040	-

Tabla 3.4: Estimación de parámetros del modelo de regresión log-gamma general.

En primer lugar, es de suma importancia mencionar que los coeficientes del modelo presentados en la tabla 3.4, se interpretan mediante sus $\exp(\hat{\beta})$, de la forma $1 - \exp(\hat{\beta})$, expresados en porcentaje.

En esta ocasión carece de sentido interpretar el intercepto por sí solo, debido a que la edad en años de las gestantes nunca será 0. Lo cual implica en absoluto, que los valores de las variables explicativas serán igual a 0 simultáneamente.

Por otro lado, al considerar el criterio del valor- p , es posible concluir que la variable edad afecta de forma significativa al tiempo de seguimiento. Es decir, existe suficiente evidencia estadística con un valor- $p = 0.0074$ para rechazar la hipótesis nula. Por lo que se deduce que la variable edad, sí afecta de forma significativa a la variable respuesta.

Por el contrario, las variables edad gestacional y cantidad de inmunizaciones contra la influenza recibidas anteriormente, no afectan al tiempo de seguimiento de aquellas mujeres. Es decir, mediante el criterio del valor- p , se deduce que existe suficiente evidencia estadística para no rechazar la hipótesis nula (H_0). Esto indica claramente que aquellas covariables no causan efecto en el modelo.

Frente a este escenario, es posible concluir (de acuerdo a la variable edad) que, por cada año que se incrementa, el tiempo de seguimiento de las embarazadas disminuye en un 5.63%; mientras que para el resto de las covariables, carece de sentido realizar sus interpretaciones, puesto que éstas evidentemente no causan efecto alguno a la variable dependiente.

3.4.1 Análisis de residuos

Como se definió anteriormente, los residuos del modelo deben seguir una distribución log-gamma generalizada con parámetros $\mu = 0$, $\sigma = 1$ y $\lambda = \hat{\lambda}_0$. Sin embargo en la figura 3.3, se presentan aquellos residuos, considerando además el ajuste a una distribución LGG de parámetros $\mu = 0$, $\sigma = 1$ y $\lambda = 1.01$.

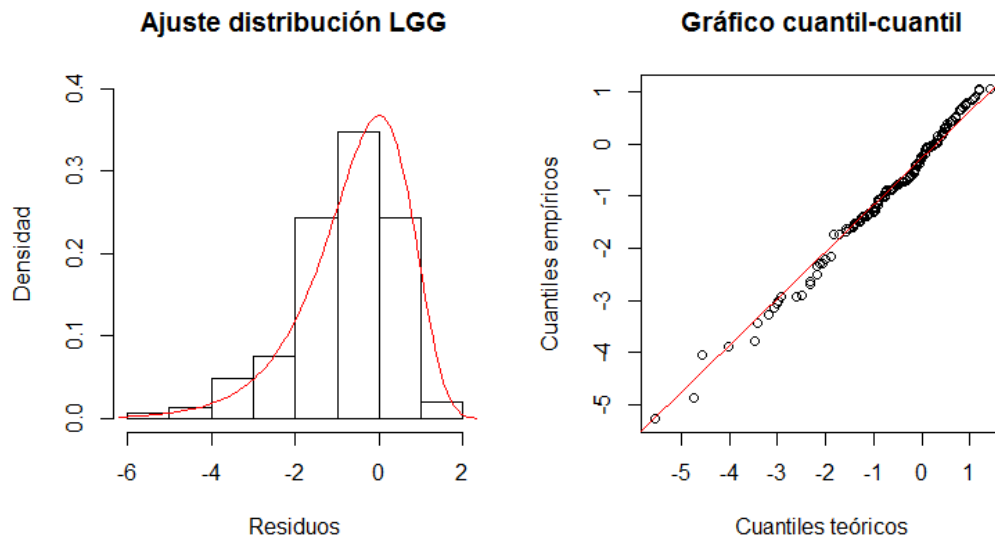


Figura 3.3: Diagnóstico del modelo.

El diagnóstico del modelo, indica que los residuos aparentemente presentan un buen ajuste a la distribución teórica, observando que la mayoría de los valores son representativos de la recta. Esto puede ser corroborado a través del gráfico cuantil-cuantil; el cual muestra que a pesar de que algunos residuos presentan una variación mínima en los extremos, el resto de los valores, sí se ajusta a una distribución LGG de parámetros $\mu = 0$, $\sigma = 1$ y $\lambda = 1.01$.

Por otra parte, es posible concluir de acuerdo a los análisis y resultados obtenidos, que el modelo de regresión log-gamma general sí es adecuado para el conjunto de datos utilizado, considerando la gran cantidad de observaciones censuradas presentes en este estudio.

Por lo tanto, si las observaciones presentan un ajuste apropiado con respecto a la distribución teórica, las estimaciones obtenidas también son una buena aproximación de la realidad de la información analizada. Por esta razón, los resultados y conclusiones son entregados con la finalidad de demostrar la correcta implementación del modelo, el cual puede ser de gran utilidad en otros casos.

En la siguiente sección, se detallarán las conclusiones finales de este estudio. Éstas, son completamente significativas, pues verifican y determinan la importancia que existe cuando se presenta información que está perdida por algún motivo en particular.

La implementación completa de este capítulo se presenta en el apéndice.

CONCLUSIONES FINALES

En este trabajo se ha presentado la aplicación del modelo de regresión log-gamma general, al utilizar los resultados de la cobertura de acuerdo a la campaña de vacunación contra la influenza estacional del año 2016, correspondiente a las mujeres embarazadas con más de 13 semanas de gestación, pertenecientes al CESFAM Juan Bautista Bravo Vega, de la ciudad de Villa Alemana.

De acuerdo a los análisis estadísticos realizados, se puede concluir que la censura está presente en el 50% de los casos. Ésta se presentó debido a que las gestantes no se inmunizaron durante la campaña; y además porque diversas embarazadas tuvieron un aborto, o bien porque fueron trasladadas de CESFAM, antes que recibieran la vacuna antigripal.

En relación al ajuste del modelo, los resultados indican que la edad de las mujeres embarazadas, sí afecta a la variable respuesta; mientras que la edad gestacional y la cantidad de inmunizaciones contra la influenza recibidas anteriormente, no afectan al tiempo de seguimiento de las gestantes.

Por otro lado, se concluye que el modelo de regresión sí es adecuado para el conjunto de datos utilizado. Por esta razón, los resultados son una buena aproximación de la realidad de la información analizada. Es decir, aquellas conclusiones, reflejan verídicamente lo sucedido con respecto a la campaña de vacunación contra la influenza, correspondiente al año 2016 del CESFAM Juan Bautista Bravo Vega. Considerando además, que estos resultados son entregados con el fin de demostrar la correcta implementación de este método.

Finalmente, cabe destacar que el modelo de regresión fue aplicado a un conjunto de observaciones con un significativo porcentaje de censura. Por este motivo, es de suma importancia recalcar la existencia de modelos estadísticos que sean capaces de describir este tipo de datos; teniendo en cuenta que aquellas observaciones se presentan habitualmente en diversos ámbitos, por lo que es inevitable ignorarlas.

Referencias

- Agostinelli, C., Marazzi, A. & Yohai, V. J. (2014). Robust estimators of the generalized log-gamma distribution. *Technometrics*, 56(1), 92-101.
- Agostinelli, C., Locatelli, I., Marazzi, A. & Yohai, V. J. (2015). Robust estimators of accelerated failure time regression with generalized log-gamma errors. Presentado. Recuperado de <https://arxiv.org/pdf/1512.01473v1.pdf>
- Agostinelli, C., Marazzi, A., Yohai, V. J. & Randriamiharisoa, A. (2016). Robust Estimation of the Generalized Loggamma Model: The R Package robustloggamma. *Journal of Statistical Software*, 70(7), 1-21.
- Agostinelli, C., Locatelli, I., Marazzi, A. & Yohai, V. J. (2017). Robust estimators of accelerated failure time regression with generalized log-gamma errors. *Computational Statistics and Data Analysis*, 107, 92-106.
- Cantú, M., Villaseñor, J. A. & Arnold, B. C. (2000). A parametric regression model for possibly censored lifetime data. *Agrociencia*, 34(4), 453-465.
- Casella, G. & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Pacific Grove, USA: Duxbury Press.
- Celis, D. (2008). *Algunas técnicas numéricas aplicadas en la solución de las ecuaciones de verosimilitud para obtener estimaciones de los parámetros de las distribuciones Weibull, gama y gama generalizada* (Tesis de pregrado). Universidad Autónoma Chapingo, Chapingo, México.
- Cordón, E. y García, F. (2014). *Tratamiento de la supervivencia con datos de duración: Aplicaciones con Excel y STATA*. Huelva, España: Universidad de Huelva.
- Decreto N° 73 Dispone vacunación obligatoria contra influenza para el año 2016 a grupo de población del país que indica. Ministerio de Salud, Santiago, Chile, 25 de febrero de 2016.

- DEIS (2016). Registro Nacional de Inmunizaciones RNI. Recuperado de <http://www.deis.cl/registro-nacional-de-inmunizaciones-rni/>.
- Departamento de Vacunas e Inmunizaciones (2012). Informe final Programa Nacional de Inmunizaciones (PNI). Recuperado de http://www.dipres.gob.cl/595/articles-139781_informe_final.pdf
- Departamento de Vacunas e Inmunizaciones (2016a). Campaña Nacional de Vacunación contra la Influenza en Chile. Recuperado de http://estadisticas.ssosorno.cl/old_page/RNI/Lineamientos-campa%C3%B1a-Influenza-2016.pdf
- Departamento de Vacunas e Inmunizaciones (2016b). Línea de tiempo Chile. Recuperado de <http://vacunas.minsal.cl/conozcanos/historia/linea-de-tiempo-chile/>.
- Departamento de Vacunas e Inmunizaciones (2016c). Obligatoriedad de la vacunación. Recuperado de <http://vacunas.minsal.cl/informacion-a-la-comunidad/obligatoriedad-de-la-vacunacion/>.
- Flores, C. (2011). *Modelo de Regresión de Cox con Métodos Flexibles en Pacientes con Linfoma No Hodgkin* (Tesis de maestría). Universidad Politécnica de Cataluña, Cataluña, España.
- Godoy, A. (2009). *Introducción al Análisis de Supervivencia con R* (Tesis de pregrado). Universidad Nacional Autónoma de México, México.
- Gutiérrez, E. (2010). *Pruebas de bondad de ajuste de bootstrap para la distribución log-gamma generalizada* (Tesis doctoral). Colegio de Postgraduados, Montecillo, México.
- Hashimoto, E. M., Ortega, E. M. M., Cancho, V. G. & Cordeiro, G. M. (2013). On estimation and diagnostics analysis in log-generalized gamma regression model for interval-censored data. *Statistics*, 47(2), 379-398.
- Hernández, M. (2007). (Ed.). *Epidemiología: Diseño y análisis de estudios*. México: Editorial Médica Panamericana.
- Hernández, A. (2010). *Análisis estadístico de datos de tiempos de fallo en R* (Tesis de maestría). Universidad de Granada, Granada, España.
- Jácome, M. (2005). *Estimación presuavizada de las funciones de densidad y distribución con datos censurados* (Tesis doctoral). Universidad de Coruña, Coruña, España.
- Juez, P. y Díez, F. (1997). *Probabilidad y estadística en medicina: Aplicaciones en la práctica clínica y en la gestión sanitaria*. Madrid, España: Díaz de Santos.

- Lindsay, B. J. (1994). Efficiency versus robustness: The case for minimum hellinger distance and related methods. *The Annals of Statistics*, 22(2), 1081-1114.
- Locatelli, I., Marazzi, A. & Yohai, V. J. (2011). Robust accelerated failure time regression. *Computational Statistics & Data Analysis*, 55(1), 874-887.
- López, B. (2011). *Estimador presuavizado de Kaplan-Meier con indicador de censura perdido aleatoriamente* (Tesis de maestría). Universidad de Coruña, Coruña, España.
- Markatou, M., Basu, A. & Lindsay, B. G. (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, 93(442), 740-750.
- MINSAL (2015). Influenza. Recuperado de <http://www.minsal.cl/influenza/>.
- MINSAL (2016). Campaña de vacunación contra la influenza 2016. Recuperado de <http://web.minsal.cl/vacunacion-influenza-2016/>.
- Montgomery, D. C., Peck, E. A. y Vining, G. G. (2004). *Introducción al análisis de regresión lineal*. México: CECSA.
- OMS (2016a). Cobertura vacunal. Recuperado de <http://www.who.int/mediacentre/factsheets/fs378/es/>.
- OMS (2016b). Gripe (estacional). Recuperado de <http://www.who.int/mediacentre/factsheets/fs211/es/>.
- OMS (2016c). Vacunas. Recuperado de <http://www.who.int/topics/vaccines/es/>.
- Ortega, E. M. M., Cancho, V. G. & Paula, G. A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis*, 15(1), 79-106.
- Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, 61(3), 539-544.
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Disponible en: <http://www.R-project.org/>.
- Robustloggamma package (2016). Package ‘robustloggamma’. Recuperado de <https://cran.r-project.org/web/packages/robustloggamma/robustloggamma.pdf>
- SAYDEX (2016). Ficha clínica electrónica para atención primaria. Recuperado de <https://www.saydex.cl/ficha-clinica-electronica-para-atencion-primaria/>.
- StataCorp (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP. Disponible en: <https://www.stata.com/>.

- URMC (2016). Influenza (Gripa). Recuperado de <https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=85&contentid=P03718>
- Vásquez, L. (1999). *Sobre la Función Gamma (Γ)* (Tesis de pregrado). Universidad Nacional de Colombia, Medellín, Colombia.
- Vega-Briceno, L., Abarca, K. y Sánchez, I. (2006). Vacuna anti-influenza en niños: Estado del arte. *Revista Chilena de Infectología*, 23(2), 164-169.
- Walpole, R. E., Myres, R. H. y Myres, S. L. (1999). *Probabilidad y estadística para ingenieros* (6ª ed.). México: Prentice-Hall Hispanoamericana.

APÉNDICE

A.1 Código Stata

```
*****
***** Administración de datos *****
*****

*****
***** BD Influenza 2016 RNI Marzo - Junio CESFAM JBBV *****
*****

clear all
version 12

cd "C:\Datos CESFAM JBBV\BD RNI Campaña Influenza 2016 Marzo - Junio\BD RNI Campaña
Influenza"

*****
*** BD Marzo 2016 ***
*****

forvalues i = 1(1)15 {
import excel using "Mes de Marzo 2016", sheet("Hoja" `i') firstrow clear
tostring NoAdministrada, replace
save marzo`i'.dta, replace
}

use marzo1.dta, clear
forvalues i = 2(1)15 {
append using marzo`i'.dta
}

save marzo2016001.dta, replace
```

```

*****
*** BD Abril 2016 ***
*****

cd "C:\Datos CESFAM JBBV\BD RNI Campaña Influenza 2016 Marzo - Junio\BD RNI Campaña
Influenza"

forvalues i = 1(1)21 {
import excel using "Mes de Abril 2016", sheet("Hoja" `i') firstrow clear
tostring NoAdministrada, replace
save abril`i'.dta, replace
}

use abril1.dta, clear
forvalues i = 2(1)21 {
append using abril`i'.dta
}

save abril2016001.dta, replace

*****
*** BD Mayo 2016 ***
*****

cd "C:\Datos CESFAM JBBV\BD RNI Campaña Influenza 2016 Marzo - Junio\BD RNI Campaña
Influenza"

forvalues i = 1(1)22 {
import excel using "Mes de Mayo 2016", sheet("Hoja" `i') firstrow clear
tostring NoAdministrada, replace
save mayo`i'.dta, replace
}

use mayo1.dta, clear
forvalues i = 2(1)22 {
append using mayo`i'.dta
}

save mayo2016001.dta, replace

*****
*** BD Junio 2016 ***
*****

cd "C:\Datos CESFAM JBBV\BD RNI Campaña Influenza 2016 Marzo - Junio\BD RNI Campaña
Influenza"

```

```

forvalues i = 1(1)22 {
import excel using "Mes de Junio 2016", sheet("Hoja" `i') firstrow clear
tostring NoAdministrada, replace
save junio`i'.dta, replace
}

use junio1.dta, clear
forvalues i = 2(1)22 {
append using junio`i'.dta
}

save junio2016001.dta, replace

*****
*** Unir BD Marzo - Junio ***
*****

use marzo2016001.dta, clear

append using abril2016001.dta
append using mayo2016001.dta
append using junio2016001.dta

drop if FechadeAdministracin == ""
replace NoAdministrada = "" if NoAdministrada == "."
keep if NoAdministrada == ""

drop if Dosis == "2° dosis 0.25mL"
drop if Dosis == "2° dosis 0.5mL"

replace Persona = "RENATO MEZA" if Identificacin == "7096372" & Lote == "K01R"
replace Identificacin = "17275547" if Persona == "RENATO MEZA"
replace dv = "k" if Persona == "RENATO MEZA"
replace Sexo = "Mujer" if Identificacin == "16327450"

drop if Identificacin == "23652195" & dv == "8" & FechadeAdministracin == "22-04-2016"

duplicates drop

save marzo_junio001.dta, replace

*****
***** Embarazadas a partir de las 13 semanas de gestación RNI *****
*****

keep if Elegiblepor == "Embarazadas, a partir de la 13ª semana de gestación"

```

```

replace dv = "K" if dv == "k"
egen Rut = concat(Identificacin dv)

/* Inmunizadas pertenecientes al CESFAM (no embarazadas) */

drop if Rut == "150677785"
drop if Rut == "169721076"
drop if Rut == "221678273"

/* Inmunizadas no pertenecientes al CESFAM (embarazadas) */

drop if Rut == "163533405"
drop if Rut == "177899461"
drop if Rut == "143160734"
drop if Rut == "138514110"
drop if Rut == "15781382K"
drop if Rut == "136341324"

sort Rut
rename Rut RUT

save "C:\Datos CESFAM JBBV\BD RNI Campaña Influenza 2016 Marzo - Junio\BD Embarazadas\
bdrniembarazadas1301.dta", replace

*****
***** BD Usuarios LIBRO DE MATERNIDAD - RAYEN CESFAM JBBV *****
*****

*****
***** BD LIBRO DE MATERNIDAD *****
*****

cd "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos\BD LIBRO DE MATERNIDAD Usuarios
Inscritos"

import excel using "BD Embarazadas Junio 2015 - Junio 2016 CESFAM JBBV",
sheet("Embarazadas Junio 2015-2016") firstrow clear

generate dias_ges_total = (Sem_ges*7) + Dias_ges
generate dias_fin_campaña_a_ultimocontrol = mdy(06,30,2016) - Ultimo_control
generate dias_in_campaña_a_ultimocontrol = mdy(03,10,2016) - Ultimo_control

generate dias_ges_fin_campaña = dias_ges_total + dias_fin_campaña_a_ultimocontrol
generate dias_ges_in_campaña = dias_ges_total + dias_in_campaña_a_ultimocontrol

generate sem_ges_fin_campaña = dias_ges_fin_campaña/7
generate sem_ges_in_campaña = dias_ges_in_campaña/7

```

```

generate cfin = "13-40sem" if (dias_ges_fin_campaña <= 280 & dias_ges_fin_campaña >= 91)
generate cin  = "13-40sem" if (dias_ges_in_campaña <= 280 & dias_ges_in_campaña >= 91)
drop if cfin == "" & cin == ""

/* Aborto antes de la campaña de influenza */

drop if Rut == "177901989"
drop if Rut == "21864936k"
drop if Rut == "198469270"
drop if Rut == "156600687"
drop if Rut == "132280452"
drop if Rut == "106740453"
drop if Rut == "146051081"
drop if Rut == "163328526"
drop if Rut == "158253240"
drop if Rut == "164883396" & Fecha_inscripcion == mdy(08,04,2015)

/* Aborto antes de las 13 semanas de gestación */

drop if Rut == "172754945"
drop if Rut == "157283359"
drop if Rut == "13431558k"
drop if Rut == "124355397"
drop if Rut == "19790920k"
drop if Rut == "120567039"
drop if Rut == "160350970"

/* Traslado antes de la campaña de influenza */

drop if Rut == "194346808"
drop if Rut == "188413064"
drop if Rut == "175920935"
drop if Rut == "18842287k"
drop if Rut == "179445565"
drop if Rut == "164672026"
drop if Rut == "174160031"
drop if Rut == "184486059"
drop if Rut == "15467974k"
drop if Rut == "186304012"
drop if Rut == "162325507"

/* Aborto durante la campaña de influenza */

generate suceso_indep      = ""
generate suceso_indep_fecha = .

replace suceso_indep      = "Aborto"          if Rut == "139878582"
replace suceso_indep_fecha = mdy(06,05,2016) if Rut == "139878582"

```

```

format suceso_indep_fecha %tdnn/dd/CCYY

/* Traslado durante la campaña de influenza */

replace suceso_indep_fecha = mdy(06,07,2016) if Rut == "185827429"
replace suceso_indep_fecha = mdy(06,24,2016) if Rut == "189984960"
replace suceso_indep_fecha = mdy(05,22,2016) if Rut == "182378437"
replace suceso_indep_fecha = mdy(03,28,2016) if Rut == "126268777"
replace suceso_indep_fecha = mdy(03,18,2016) if Rut == "188420656"
replace suceso_indep_fecha = mdy(04,07,2016) if Rut == "160570733"
replace suceso_indep_fecha = mdy(04,04,2016) if Rut == "187867681"
replace suceso_indep_fecha = mdy(05,12,2016) if Rut == "172745113"
replace suceso_indep_fecha = mdy(04,04,2016) if Rut == "179455641"

replace suceso_indep = "Traslado" if suceso_indep != "Aborto" & suceso_indep_fecha != .

generate rut01 = substr(Rut,1,8)
generate rut02 = substr(Rut,9,1)

replace rut02 = "K" if rut02 == "k"
generate RUT1 = rut01 + rut02

drop Rut rut01 rut02
rename RUT1 RUT

/* Suceso dependiente de la inmunización */

drop if dias_ges_in_campaña > 168
sort RUT

save "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos\BD Embarazadas\
bdlibromatembarazadas1301.dta", replace

*****
***** BD RAYEN *****
*****

cd "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos\BD RAYEN Usuarios Inscritos"

import excel using "Listado de usuarios inscritos en CESFAM Juan Bautista Bravo
al 26-JULIO-2016", sheet("Hoja1") firstrow clear

keep if ESTADO == "Activo"
sort RUT

save "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos\BD Embarazadas\
bdrayenembarazadas1301.dta", replace

```

```
*****  
***** MATERNIDAD - RAYEN Embarazadas a partir de las 13 semanas de gestación *****  
*****
```

```
*****  
** Unir BD LIBRO DE MATERNIDAD - RAYEN **  
*****
```

```
cd "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos\BD Embarazadas"
```

```
merge RUT using bdlbromatembarazadas1301.dta
```

```
replace RUT = "13074799K" if RUT == "13074799k"  
replace RUT = "15102372K" if RUT == "15102372k"  
replace RUT = "16107912K" if RUT == "16107912k"  
replace RUT = "16970578K" if RUT == "16970578k"  
replace RUT = "24019970K" if RUT == "24019970k"
```

```
replace FECHA_DE_NACIMIENTO = mdy(08,19,1976) if RUT == "13074799K"  
replace FECHA_INSCRIPCION = mdy(10,22,2008) if RUT == "13074799K"
```

```
replace FECHA_DE_NACIMIENTO = mdy(06,07,1983) if RUT == "15102372K"  
replace FECHA_INSCRIPCION = mdy(07,23,2010) if RUT == "15102372K"
```

```
replace FECHA_DE_NACIMIENTO = mdy(05,19,1986) if RUT == "16107912K"  
replace FECHA_INSCRIPCION = mdy(07,30,2013) if RUT == "16107912K"
```

```
replace FECHA_DE_NACIMIENTO = mdy(06,14,1988) if RUT == "16970578K"  
replace FECHA_INSCRIPCION = mdy(07,09,2010) if RUT == "16970578K"
```

```
replace FECHA_DE_NACIMIENTO = mdy(07,01,1981) if RUT == "24019970K"  
replace FECHA_INSCRIPCION = mdy(03,15,2012) if RUT == "24019970K"
```

```
drop if RUT == "13074799K" & Sem_ges == .  
drop if RUT == "15102372K" & Sem_ges == .  
drop if RUT == "16107912K" & Sem_ges == .  
drop if RUT == "16970578K" & Sem_ges == .  
drop if RUT == "24019970K" & Sem_ges == .  
drop if _merge == 1
```

```
replace FECHA_DE_NACIMIENTO = mdy(03,23,1993) if RUT == "182378437"  
replace FECHA_INSCRIPCION = mdy(02,04,2016) if RUT == "182378437"
```

```
replace FECHA_DE_NACIMIENTO = mdy(05,18,1995) if RUT == "185827429"  
replace FECHA_INSCRIPCION = mdy(04,22,2016) if RUT == "185827429"
```

```
replace FECHA_DE_NACIMIENTO = mdy(05,28,1994) if RUT == "187867681"  
replace FECHA_INSCRIPCION = mdy(12,14,2015) if RUT == "187867681"
```

```

replace FECHA_DE_NACIMIENTO = mdy(06,19,1994) if RUT == "188420656"
replace FECHA_INSCRIPCION    = mdy(01,12,2016) if RUT == "188420656"

replace FECHA_DE_NACIMIENTO = mdy(09,04,1995) if RUT == "189984960"
replace FECHA_INSCRIPCION    = mdy(01,25,2016) if RUT == "189984960"

drop _merge
sort RUT

save bdlbromaternalrayen01.dta, replace

*****
** Unir BD LIBRO DE MATERNIDAD (RAYEN) - RNI **
*****

*****
***** BD LIBRO DE MATERNIDAD (RAYEN) *****
*****

cd "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos\BD Embarazadas"

use bdlbromaternalrayen01.dta, clear

save "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos vs RNI Campaña Influenza 2016\BD
Embarazadas\bdlibromaternalrayen02.dta", replace

*****
***** BD RNI *****
*****

cd "C:\Datos CESFAM JBBV\BD RNI Campaña Influenza 2016 Marzo - Junio\BD Embarazadas"

use bdrniembarazadas1301.dta, clear

save "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos vs RNI Campaña Influenza 2016\BD
Embarazadas\bdrniembarazadas1302.dta.dta", replace

cd "C:\Datos CESFAM JBBV\BD RAYEN Usuarios Inscritos vs RNI Campaña Influenza 2016\BD
Embarazadas"

merge    RUT using bdlbromaternalrayen02.dta
drop if  _merge == 1
drop    _merge

replace suceso_indep          = "" if RUT == "189984960"
replace suceso_indep_fecha = .  if RUT == "189984960"

```

```

*****
***** Total embarazadas a partir de las 13 semanas de gestación *****
*****

*****
*** Definir, clasificar y modificar variables de interés ***
*****

*****
** Variable seguimiento **
*****

generate Fecha_Administracion = date(FechadeAdministracin,"DMY")
format Fecha_Administracion %tdnn/dd/CCYY

/* Fecha inicio seguimiento */

generate fecha_in_seg = .
replace fecha_in_seg = mdy(03,10,2016) if FECHA_INSCRIPCION <= mdy(03,10,2016) &
dias_ges_in_campaña >= 91

generate dias91 = 91 - dias_ges_in_campaña
generate fecha91 = mdy(03,10,2016) + dias91
format fecha_in_seg fecha91 %tdnn/dd/CCYY

replace fecha_in_seg = fecha91 if FECHA_INSCRIPCION <= mdy(03,10,2016) &
dias_ges_in_campaña < 91

generate dias_insc = FECHA_INSCRIPCION - mdy(03,10,2016)
generate dias_ges_insc = dias_insc + dias_ges_in_campaña

replace fecha_in_seg = FECHA_INSCRIPCION if FECHA_INSCRIPCION > mdy(03,10,2016) &
dias_ges_insc >= 91

replace fecha_in_seg = fecha91 if FECHA_INSCRIPCION > mdy(03,10,2016) &
dias_ges_insc < 91

/* Fecha fin seguimiento */

generate fecha_fin_seg = .
replace fecha_fin_seg = mdy(06,30,2016) if Fecha_Administracion == . &
suceso_indep == ""

format fecha_fin_seg %tdnn/dd/CCYY

replace fecha_fin_seg = suceso_indep_fecha if suceso_indep == "Aborto"
replace fecha_fin_seg = suceso_indep_fecha if suceso_indep == "Traslado"
replace fecha_fin_seg = Fecha_Administracion if Fecha_Administracion != .

```

```

generate T1 = Fecha_Administracion - fecha_in_seg
generate C1 = mdy(06,30,2016) - fecha_in_seg if T1 != .

generate C2 = suceso_indep_fecha - fecha_in_seg
generate T2 = mdy(07,01,2016) - fecha_in_seg if C2 != .

generate T3 = mdy(07,01,2016) - fecha_in_seg if T1 == . & C2 == .
generate C3 = mdy(06,30,2016) - fecha_in_seg if T3 != .

generate Ti = T1
replace Ti = T2 if T2 != .
replace Ti = T3 if T3 != .

generate Ci = C1
replace Ci = C2 if C2 != .
replace Ci = C3 if C3 != .

generate Dias_Seg = min(Ti,Ci)
generate LOG_Dias_Seg = log(Dias_Seg)

generate Cat_Dato = 1 if Fecha_Administracion != .
replace Cat_Dato = 2 if suceso_indep_fecha != .
replace Cat_Dato = 0 if Fecha_Administracion == . & suceso_indep_fecha == .

keep FECHA_DE_NACIMIENTO LOG_Dias_Seg Dias_Seg dias_ges_in_campaña
Immunizacionesinfluenza Ti Ci Cat_Dato

*****
** Variable edad **
*****

generate Edad1 = (mdy(03,10,2016) - FECHA_DE_NACIMIENTO)/365.25
generate Edad = trunc(Edad1)

*****
** Variable edad gestacional **
*****

rename dias_ges_in_campaña Edad_Ges
replace Edad_Ges = 0 if Edad_Ges < 0
replace Edad_Ges = Edad_Ges/7

*****
** Variable censura **
*****

generate Censura = 1 if Ti <= Ci
replace Censura = 0 if Ti > Ci

```

```
*****  
** Variable inmunización **  
*****
```

```
rename Inmunizacionesinfluenza Inmun_Influenza
```

```
keep LOG_Dias_Seg Dias_Seg Edad Edad_Ges Inmun_Influenza Censura Cat_Dato  
order LOG_Dias_Seg Dias_Seg Edad Edad_Ges Inmun_Influenza Censura Cat_Dato
```

```
save bdLMRNI.dta, replace
```

```
*****  
***** Análisis exploratorio de datos *****  
*****
```

```
summarize Dias_Seg  
summarize LOG_Dias_Seg  
summarize Edad  
summarize Edad_Ges
```

```
summarize Inmun_Influenza  
tabulate Inmun_Influenza
```

```
summarize Censura  
tabulate Censura
```

```
summarize Cat_Dato  
tabulate Cat_Dato
```

```
tabstat Dias_Seg, stats(count, mean, median, sd, var, cv, p25, p75 min, max)  
tabstat LOG_Dias_Seg, stats(count, mean, median, sd, var, cv, p25, p75 min, max)  
tabstat Edad, stats(count, mean, median, sd, var, cv, p25, p75 min, max)  
tabstat Edad_Ges, stats(count, mean, median, sd, var, cv, p25, p75 min, max)  
tabstat Inmun_Influenza, stats(count, mean, median, sd, var, cv, p25, p75 min, max)  
tabstat Censura, stats(count, mean, median, sd, var, cv, p25, p75 min, max)
```

```
tabstat Dias_Seg, stats(count, mean, sd) by(Censura)  
tabstat LOG_Dias_Seg, stats(count, mean, sd) by(Censura)  
tabstat Edad, stats(count, mean, sd) by(Censura)  
tabstat Edad_Ges, stats(count, mean, sd) by(Censura)  
tabstat Inmun_Influenza, stats(count, mean, sd) by(Censura)
```

```
tabstat Dias_Seg, stats(count, mean, sd) by(Inmun_Influenza)  
tabstat LOG_Dias_Seg, stats(count, mean, sd) by(Inmun_Influenza)  
tabstat Edad, stats(count, mean, sd) by(Inmun_Influenza)  
tabstat Edad_Ges, stats(count, mean, sd) by(Inmun_Influenza)  
tabstat Censura, stats(count, mean, sd) by(Inmun_Influenza)
```

A.2 Código R-project

```
install.packages("foreign")
library(foreign)

install.packages("pastecs")
library(pastecs)

install.packages("psych")
library(psych)

install.packages("robustloggamma")
library(robustloggamma)

Influenza <- read.dta("C:/Datos CESFAM JBBV/BD RAYEN Usuarios Inscritos vs RNI Campaña
  Influenza 2016/BD Embarazadas/bdLMRNI.dta")

attach(Influenza)
names(Influenza)
class(Influenza)

#### Funciones de densidad asociadas a la distribución LGG ####

plot(function(x) dloggamma(x, mu = 0, sigma = 1, lambda = 0), from = -8, to = 4, col =
  "black", main = "Funciones de densidad", ylab = "Densidad", xlab = "Observaciones")

plot(function(x) dloggamma(x, mu = 0, sigma = 2, lambda = 1), from = -8, to = 4, col =
  "red", add = TRUE)

plot(function(x) dloggamma(x, mu = 0, sigma = 1, lambda = 1), from = -8, to = 4, col =
  "green", add = TRUE)

plot(function(x) dloggamma(x, mu = 0, sigma = 2, lambda = 2), from = -8, to = 4, col =
  "blue", add = TRUE)

legend(legend = c("Normal", "Log-Weibull", "Log-exponencial", "Log-gamma"), col =
  c("black", "red", "green", "blue"), inset = 0.01, lty = 1, "topleft")

#####
##### Análisis exploratorio de datos #####
#####

#### Resumen descriptivo ####

summary(Influenza)
stat.desc(Influenza)
describe(Influenza)
```

```

#### Histogramas variables de interés ####

### Variable respuesta ###

par(mfrow = c(1,2))

hist(Dias_Seg, main = "Tiempo de seguimiento", xlab = "Días", ylab =
"Densidad", freq = F)

lines(density(Dias_Seg, from = -1.3, to = 122), lwd = 1, col = "red")

hist(LOG_Dias_Seg, main = "Log Tiempo de seguimiento", xlab = "Log Días", ylab =
"Densidad", ylim = c(0,0.39), breaks = 10, freq = F)

lines(density(LOG_Dias_Seg, from = -0.09, to = 5.05), lwd = 1, col = "red")

## Ajuste LGG ##

set.seed(027)
par(mfrow = c(1,2))

Log_DiasSeg <- sort(LOG_Dias_Seg)
Par_Log_DiasSeg <- loggammacensrob(Log_DiasSeg, method = "oneWL", delta = Censura)

# Histograma ajuste LGG #

hist(Log_DiasSeg, main = "Ajuste distribución LGG", xlab = "Log Días", ylab =
"Densidad", ylim = c(0,0.4), freq = F)

plot(function(Log_DiasSeg) dloggamma(Log_DiasSeg, mu = Par_Log_DiasSeg$mu, sigma =
Par_Log_DiasSeg$sigma, lambda = Par_Log_DiasSeg$lambda), from = -0.1, to = 5.1, col =
"red", add = TRUE)

# Q-Q plot #

n <- length(Par_Log_DiasSeg$data)

M_LGG <- rloggamma(n, mu = Par_Log_DiasSeg$mu, sigma = Par_Log_DiasSeg$sigma, lambda =
Par_Log_DiasSeg$lambda)

qqplot(M_LGG, Log_DiasSeg, main = "Gráfico cuantil-cuantil", xlab =
"Cuantiles teóricos", ylab = "Cuantiles empíricos")

C_Log_DiasSeg <- quantile(Log_DiasSeg, c(0.25,0.75))
C_LGG <- qlongamma(c(0.25, 0.75), mu = Par_Log_DiasSeg$mu, sigma =
Par_Log_DiasSeg$sigma, lambda = Par_Log_DiasSeg$lambda)

b <- (C_Log_DiasSeg[2] - C_Log_DiasSeg[1]) / (C_LGG[2] - C_LGG[1])

```

```

a <- C_Log_DiasSeg[1] - C_LGG[1]*b
abline(a,b, col = "red")

#####
##### Aplicación modelo log-gamma general #####
#####

#### Ajuste modelo LGG ####

ModeloLGG <- loggammacenslmrob(LOG_Dias_Seg ~ Edad + Inmun_Influenza + Edad_Ges, delta =
  Censura, method = "oneWL", data = Influenza)

AjusteLGG <- summary(ModeloLGG)

#### Residuos modelo LGG ####

ResidualsLGG <- residuals(AjusteLGG)
ResiduosLGG <- sort(ResidualsLGG)

Par_residuos <- loggammarob(ResiduosLGG, method = "oneWL")

#### Histogramas residuos ####

par(mfrow = c(1,2))

hist(ResiduosLGG, main = "Ajuste distribución LGG", xlab = "Residuos", ylab =
  "Densidad", ylim = c(0,0.4), freq = F)

plot(function(ResiduosLGG) dloggamma(ResiduosLGG, mu = 0, sigma = 1, lambda =
  Par_residuos$lambda), from = -6.2, to = 2.3, col = "red", add = TRUE)

#### Q-Q plot ####

set.seed(016)
n <- length(Par_residuos$data)

MR_LGG <- rloggamma(n, mu = 0, sigma = 1, lambda = Par_residuos$lambda)

qqplot(MR_LGG, ResiduosLGG, main = "Gráfico cuantil-cuantil", xlab =
  "Cuantiles teóricos", ylab = "Cuantiles empíricos")

C_Residuos <- quantile(ResiduosLGG, c(0.25,0.75))
CR_LGG <- qloggamma(c(0.25, 0.75), mu = 0, sigma = 1, lambda = Par_residuos$lambda)

b <- (C_Residuos[2] - C_Residuos[1]) / (CR_LGG[2] - CR_LGG[1])
a <- C_Residuos[1] - CR_LGG[1]*b
abline(a,b, col = "red")

```

A.3 Certificado Directora CESFAM JBBV

Villa Alemana, 05 de Abril de 2017.

Señores

Universidad Valparaíso

Presente

Como directora del CESFAM Juan Bautista Bravo Vega proporciono a Nicole Aguilar Manríquez, información sobre Influenza año 2016 (específicamente embarazadas con más de 13 semanas de gestación), para realizar la aplicación de su trabajo de titulación, el cual tiene como título "Modelo de regresión Log-gama general aplicado a datos censurados", todo esto autorizado por el Departamento de Salud.

Atentamente.



Dña. GONZALEZ SCHIPPACASSE
DIRECTORA
CESFAM JBBV

