

# Análisis estadístico para el cálculo de la UPC utilizando modelos lineales generalizados

Statistical analysis for the calculation of UPC using generalized linear models

Adriana Caballero.<sup>a</sup>  
adrianacaballero@usantotomas.edu.co

Wilmer Dario Pineda.<sup>b</sup>  
wilmerpineda@usantotomas.edu.co

## Resumen

La correcta estimación de una prima de seguros garantiza que en el momento de presentar algún siniestro la empresa aseguradora se encuentra en la capacidad de responder bajo las condiciones pactadas, el sistema de salud de Colombia se responde a este principio y es el Ministerio de Salud y Protección Social quien debe estimar el monto a reconocer a cada una de las EPS, garantizando la cobertura del riesgo en salud y el riesgo financiero, promoviendo la sostenibilidad financiera, y la atención integral de los afiliados; por lo tanto se hace necesario abordar modelos con capacidad de predicción, que permitan la distribución adecuada de los recursos, por lo cual se exploraran modelos lineales generalizados que respondan de manera adecuada a las necesidades del país, modelando la frecuencia y severidad de manera independiente, buscando con ello disminuir la posible selección de riesgos que se presente en el sistema.

**Palabras clave:** Clasificación de riesgo, Modelos Predictivos, Ajuste de riesgo, Prima de seguros, GLM, UPC.

## 1. Introducción

El sistema general de seguridad social en salud SGSSS se encuentra fundamentado desde la Ley 100 de 1993 en un esquema de aseguramiento público y colectivo, donde el seguro es obligatorio a través de la afiliación a dos regímenes, el contributivo en el cual se encuentran todas las personas con capacidad de pago, y el subsidiado en cual se encuentran las personas con mayor vulnerabilidad y sin capacidad de pago.

Como en todo esquema de seguros, se definen las coberturas a las cuales el asegurado tendrá acceso, para el caso de salud estas coberturas se especifican a través de tecnologías y servicios de salud, la prestación de las coberturas se garantiza por medio del reconocimiento económico de la prima que se denomina Unidad de Pago por Capitación.

Ahora bien, la salud según la organización mundial de la salud se define como "un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades", en este orden de ideas las coberturas se encuentran diseñadas para impactar en cualquier momento de la historia natural de la enfermedad, es decir cubren las etapas de promoción de la salud y prevención de la enfermedad, el diagnóstico oportuno, el tratamiento adecuado, la rehabilitación en aras de lograr nuevamente el equilibrio y la paliación.

Es así, como el estado ha delegado en las Empresas Promotoras de Servicios de Salud EPS la gestión del riesgo tanto en salud como financiero, estableciendo las coberturas y la prima a reconocer por cada uno de los afiliados, teniendo en cuenta que los riesgos son disímiles a lo largo del curso de vida de una persona por lo cual se han diseñado tres ajustadores de riesgo principales, la edad, el sexo y la ubicación geográfica, este último teniendo en cuenta que Colombia es un país tropical que implica mayores riesgos en salud en algunas regiones más que en otras.

Por lo expuesto anteriormente el Ministerio de Salud y Protección Social se encuentra en la constante búsqueda de mecanismos de protección financiera de los riesgos en salud, de manera tal que se logre un

<sup>a</sup>Estudiante

<sup>b</sup>Director proyecto de grado

acceso equitativo a las tecnologías y servicios de salud, con el fin de cumplir con el objetivo último de un seguro proteger financieramente a una persona en caso de presentar un siniestro que ponga en riesgo la estabilidad financiera de la persona y su núcleo familiar.

El objetivo del presente trabajo, es presentar alternativas para el cálculo de la Unidad de Pago por Capitalización, que permitan una mayor equidad en la distribución de recursos garantizando con ello el acceso y oportunidad en la atención de todas las personas afiliadas, incentivando la gestión integral del riesgo en salud por parte de las aseguradoras, llevándolas a ser compradores inteligentes, y reduciendo las posibles posiciones dominantes que se puedan presentar en el país.

El documento se abordará en el siguiente orden, en la primera parte se presentará el marco teórico base para el desarrollo del presente trabajo, el mismo se encuentra dividido en tres apartados, los antecedentes donde se presenta el porque de la necesidad de la estimación de una prima, los conceptos básicos y los modelos estadísticos propuestos, y por último la hipótesis que se busca comprobar con el desarrollo del trabajo. En una segunda parte se presentaran los objetivos a cumplir con la aplicación de los modelos propuestos, siguiendo la metodología con la cual se desarrollará el presente trabajo.

## 2. Marco teórico y revisión de literatura

El abordaje del marco teórico y la revisión de la literatura se hará teniendo en cuenta dos grandes apartados I. Antecedentes, II. Bases teóricas las cuales se trabajaran en dos secciones a. Conceptos de los seguros y Ajuste de riesgo y b. los modelos estadísticos a realizar, finalmente se planteara la hipótesis a elegir.

### 2.1. Antecedentes

Según lo expresado por [Arrow, 1981] el mercado sanitario como cualquier otro mercado que se desarrolle bajo la teoría económica, busca optimizar los recursos disponibles lo cual significa que no exista ninguna asignación de recursos en la producción de los diferentes bienes y servicios de salud que pueda mejorar la situación de todos los agentes del mercado al mismo tiempo, en este orden de ideas cada individuo trata de maximizar el valor esperado de su utilidad es decir busca mantener la salud y en caso de presentar ausencia de la misma estar protegido contra los riesgos en particular el de padecer una enfermedad de alto costo o catastrófica

Es así como en este mercado intervienen las leyes de oferta y demanda con una falla de mercado predominante y es la asimetría de información, toda vez que el paciente no tiene el conocimiento del médico, y el profesional de la salud se puede ver enfrentado al riesgo moral derivado del paciente al ocultar información relevante para el tratamiento médico.

En aras de lograr la equidad en el acceso a los servicios de salud a nivel mundial la [Organización Panamericana de la Salud, ] ha promovido la cobertura universal definida como los mecanismos de organización y financiamiento suficientes para cubrir a toda la población, este nuevo viraje de la política mundial de la salud sitúa en el centro a los pacientes y su atención, esta estrategia basa su propuesta en tres dimensiones los servicios de salud, los gastos directos, y la población, lo cual se puede observar en la figura 1.

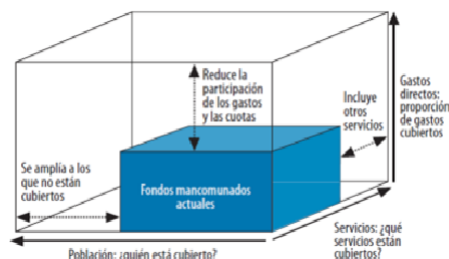


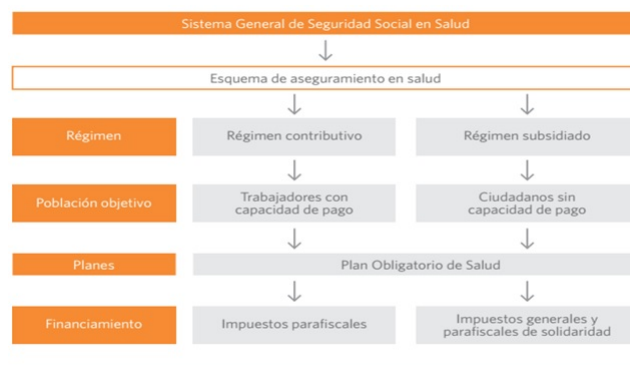
Figura 1: Informe Salud del Mundo 2010, OMS

Es así, como los diferentes sistemas de salud han venido trabajando en modelos económicos, estadísticos y actuariales que den respuesta a la asignación de recursos de manera adecuada, estos modelos se pueden agrupar en modelos retrospectivos y modelos prospectivos.

**Modelos retrospectivos:** En estos modelos no existe incentivos a la eficiencia pues se le reconoce al prestador aquello que gasta.

**Modelos prospectivos:** Son modelos que establecen una redistribución fija lo cual incentiva a la eficiencia de la prestador, con una producción adecuada y una calidad óptima en sus servicios, lo cual puede generar utilidades en el prestador.

Teniendo en cuenta lo anterior, el SGSSS de Colombia ha venido utilizando modelos prospectivos para la estimación de la prima, dado que es el estado quien define la cobertura de tecnologías y servicios de salud y su financiación (ver figura 2), es así como en Colombia el seguro es obligatorio y la duración es vitalicia bien sea a través del régimen contributivo o al régimen subsidiado.



Fuente: Adaptado de Giedion, Panopoulou y Gómez-Fraga (2009).

Figura 2: Sistema de Salud Colombiano  
[Giedion et al., 2014]

De acuerdo a [Lancheros, 2011] el cálculo de una prima de seguros han empleado dos métodos principales estos son el método de prima pura y método de razón de pérdida; la aplicación de cualquiera de ellos requiere tener en cuenta lo establecido en [de Salud de Protección Social, 2021]

**Período de experiencia:** Es el horizonte de tiempo en el cual tuvieron lugar las atenciones en salud que servirán como experiencia para la estimación de la(s) prima(s) pura(s) de riesgo.

**Período de recolección de la información (período de análisis):** Corresponde al espacio de tiempo durante el cual se recolecta la información.

**Período de aplicación o de pronóstico:** Es el período en el cual se darán las atenciones en salud a cargo de la UPC. Dicho período de aplicación corresponde en este estudio al tiempo comprendido entre el 1 de enero y el 31 de diciembre.

A continuación se presenta la definición de los métodos de estimación presentándose anteriormente.

- **Método de Prima Pura:** Se estiman las pérdidas totales y se calcula el número de expuestos correspondiente al período de experiencia. Posteriormente se proyecta teniendo en cuenta aspectos como la inflación. Por último la prima pura es cargada con un factor de gastos, así como un margen de riesgo y utilidad.
- **Razón de pérdida:** Este método de razón de pérdida es determinar en qué porcentaje debe ser aumentada o disminuida la tasa actual, para alcanzar la utilidad esperada.

## 2.2. Bases Teóricas

### 2.2.1. Conceptos de los seguros

Los sistemas de salud basados en el aseguramiento tienen como pilar el riesgo que se enfrenta una persona como consecuencia de la pérdida de salud o la necesidad de mantener el estado de salud a través de la promoción de la salud y prevención de la enfermedad, lo cual conlleva a realizar ajustes de riesgo los cuales tienen sus orígenes en los años 80 como un proceso estadístico que tiene en cuenta el estado de salud y el gasto en salud de los afiliados.

Los modelos prospectivos utilizados en Colombia y de acuerdo a la clasificación presentada por [Caballero Otálora, 2020], son aquellos basados en el gasto del pasado, dado que los mismos guardan una relación directa con el gasto futuro.

Los modelos de ajuste de riesgo actuarial, tienen su base en los modelos antes mencionados, y son útiles en los sistemas de salud basados en el aseguramiento donde interviene el cálculo de una prima la cual es un estimativo del valor de las obligaciones financieras adquiridas en un contrato de aseguramiento lo anterior expresado por [Ruiz Gómez, 2013]

Es así, como el mercado de seguros como cualquier otro bien o servicio que se encuentre en libre mercado fijará el precio que refleje para el productor los costos asociados al producto e incorpore un margen de beneficio, y para el comprador el precio que se está dispuesto a pagar por el mismo, en términos más generales la relación entre el precio, costo y utilidad se pueden expresar de la siguiente manera

$$\text{Precio} = \text{Costo} + \text{Utilidad}$$

Los seguros son la promesa de hacer algo si ocurren los eventos especificados en el futuro, por lo que el valor o tarificación de la póliza de seguros se debe fijar de manera diferente, por anterior y de acuerdo a [Li, 2022] las aseguradoras no ofrecen un bien o servicio directamente, sino que se hace una promesa (se establece un compromiso) de ofrecer un servicio en el futuro en caso de que un hecho se materialice sobre la persona o la propiedad del asegurado (cliente)

Según [Lancheros, 2011] "los seguros pueden ser definidos como un contrato entre el tomador de la póliza y la compañía de seguros, mediante el cual la compañía se obliga, a cambio de una suma de dinero (prima), al pago a un tercero (beneficiario) de una cantidad de dinero que usualmente tiene un límite (valor asegurado), si determinados eventos, tales como un choque o un incendio, ocurren dentro de un período específico de tiempo"

Ahora bien siguiendo a [Li, 2022] la ley de los grandes números es uno de los principios fundamentales sobre los que se sustenta el negocio asegurador y la técnica actuarial. La idea fundamental de este principio es que el riesgo relativo (Coeficiente de Variación) disminuye a medida que aumenta el tamaño de negocio entendido como el número de unidades de riesgo o contratos en la cartera de la compañía

Para comprender mejor que conlleva un sistema de seguros es necesario establecer los conceptos básicos, los cuales han sido tomados de [Werner and Modlin, 2010]:

**Prima:** Es el monto que el tomador paga por la cobertura de seguros.

**Expuestos:** Un expuesto es la unidad básica de riesgo que mide la exposición a una pérdida. La medida de exposición varía según el ramo de seguros al cual se le está fijando las tarifas, así mismo la elección de los expuestos debe tener en cuenta:

- Los expuestos deben relacionarse directamente con las pérdidas
- Los expuestos deben ser fáciles de medir de manera objetiva, evitando el riesgo moral.
- Los expuestos deben calcularse de manera consistente en el tiempo.

**Siniestros y Reclamaciones:** Cuando un riesgo cubierto por la póliza y este se materializa es decir se produce un siniestro, el tomador de la misma realiza la demanda correspondiente a la indemnización pactada, lo cual se conoce como reclamación.

Hay eventos que no son informados de manera inmediata por lo cual es necesario realizar una reserva que le permita a la compañía hacer frente a las reclamación en el momento en que esta se presente lo anterior se conoce como IBNR.

**Pérdidas:** Es la cantidad pagada al demandante de la póliza según los términos del seguro.

- El valor asegurado corresponde al monto máximo pagadero en caso de siniestro, lo cual ha quedado estimulado en el capsulado de la póliza.
- Los gastos para el ajuste y la liquidación de siniestros, así como para la emisión de pólizas o gastos de administración.

Teniendo en cuenta los conceptos presentados anteriormente y llevados al sistema general de seguridad social en salud de Colombia es el estado quien establece la prima a reconocer conocida como la Unidad de Pago por Capitación UPC, y es el mismo estado quien define las coberturas a financiar con la misma; es decir las tecnologías y servicios de salud que una persona tiene derecho con el fin de llevar a cabo cualquier evento de salud es decir que las mismas cubren desde la promoción y prevención de la salud hasta la paliación, pasando por el diagnóstico, el tratamiento y la rehabilitación

Al igual que las primas de los seguros, a ser la UPC la prima del SGSSS debe tener en cuenta para su estimación principios fundamentales, que de acuerdo a [Lancheros, 2011] son:

1. Las tarifas deben ser suficientes. para cubrir los costos de las reclamaciones mas los gastos y proveer un margen de utilidad, en el Sistema General de Seguridad Social en Salud, no existe margen de utilidad, se reconoce un porcentaje para la administración delegada.
2. Las tarifas deben estar directamente relacionadas con el riesgo a cubrir.
3. Las tarifas deben ser el producto de la utilización de la información estadística que cumpla exigencias de homogeneidad y representatividad.

Lo expuesto anteriormente se condensa en la ecuación fundamental de seguros la cual se presenta a continuación , la según [Werner and Modlin, 2010] la tarificación busca que la ecuación se encuentre balanceada. Es decir que las primas sean suficientes para cubrir todos los costos asociados con la transferencia de riesgo y obtenga un margen de utilidad:

$$Prima = \text{Perdidas} + \text{Gasto de Ajuste de Siniestros} + \text{Gastos de suscripción} + \text{Utilidad}$$

El balanceo de la ecuación requiere que la tarificación sea prospectiva, para ello se tiende a utilizar la información histórica para estimar los costos futuros que serán base en el cálculo de las primas, teniendo en cuenta que es necesario que se realicen ajustes que logren reflejar las experiencias futuras.

Para la tarificación hay dos razones fundamentales que deben tenerse en cuenta y que conforman la denominada prima pura.

**Frecuencia:** Es la tasa de las reclamaciones ocurridas y se expresa como sigue:

$$Frecuencia = \frac{\text{Número de reclamaciones}}{\text{Número de expuestos}}$$

**Severidad:** Es la medida del promedio del costo de las reclamaciones y es calculada como sigue

$$Severidad = \frac{\text{Pérdidas}}{\text{Número de reclamaciones}}$$

**Prima Pura:** Se refiere a los costos esperados, que corresponden a los siniestros. cuando se carga con factor de gastos de administración o de utilidad se convierte en la prima comercial, es decir en la prima que es cancelada por el tomador de la póliza. Por lo anterior Corresponde al costo promedio por expuesto, y su cálculo se realiza siguiendo la siguiente ecuación.

$$PrimaPura = \frac{\text{Pérdidas}}{\text{Número de expuestos}} = Frecuencia * Severidad$$

Los conceptos anteriores permiten presentan el contexto del presente trabajo, toda vez que el objetivo final será la estimación de la prima pura de manera diferente a la metodología actual utilizada por el Ministerio de Salud y Protección Social, con la aplicación de modelos lineales generalizados para cada uno de los componentes principales la frecuencia y la severidad.

### 2.2.2. Ajustadores de Riesgo

Ahora bien, los ajustadores de riesgo como su nombre lo indica buscan responder al comportamiento de la población expuesta, en los sistemas de salud basados en el aseguramiento es necesario reconocer las diferentes características que puedan afectar la salud de la población, en general esta clase de modelo identifica dos ajustadores principales la edad y la ubicación geográfica, dichos ajustadores se abordan a continuación.

#### 1. Grupos de Edad y Sexo

En Colombia, a partir de la Ley 100 de 1993, las aseguradoras (EPS) deben propender por una gestión adecuada del riesgo en salud, y en los últimos años esta función se ha realizado a partir del enfoque de curso de vida que de acuerdo a [Grupo Curso de Vida, 2015] aborda los momentos del continuo de la vida y reconoce que el desarrollo humano y los resultados en salud dependen de la interacción de diferentes factores a lo largo del curso de la vida, de experiencias acumulativas y situaciones presentes de cada individuo influenciadas por el contexto familiar, social, económico, ambiental y cultural; entendiendo que invertir en atenciones oportunas en cada generación repercutirá en las siguientes y que el mayor beneficio de un momento vital puede derivarse de intervenciones hechas en un período anterior.

Desde el punto de vista epidemiológico, el curso de la vida es el estudio a largo plazo de los efectos en la salud o la enfermedad de la exposición a riesgos físicos o sociales durante la gestación, la infancia, la adolescencia, la juventud y la vida adulta.

En general el enfoque de curso de vida se basa en el reconocimiento de las etapas de la vida como se presenta en la siguiente figura.

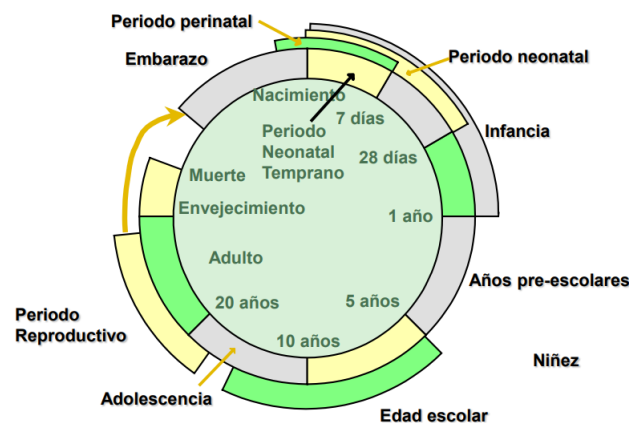


Figura 3: Enfoque Curso de Vida  
[Mecer, Raul, 2016]

Según [Mecer, Raul, 2016] este enfoque es un complejo interjuego de factores protectores de riesgo entre los que se encuentran:

- a) Biológicos
- b) Comportamentales
- c) Psicológicos
- d) Sociales

los cuales contribuyen los desenlaces en salud a lo largo del curso de vida, es por esto que la apuesta del Ministerio de Salud y Protección Social es al abordaje integral de la persona.

Este enfoque y siguiendo al [Mecer, Raul, 2016] ayuda a comprender que:

- Los beneficios máximos en un grupo de edad pueden derivarse de intervenciones realizadas en una edad más temprana
- Son necesarias intervenciones en varios momentos a lo largo del curso de vida para lograr mejores resultados y efectos acumulativos
- Existen riesgos y beneficios intergeneracionales que son esenciales a considerar en todos los grupos de edad

Teniendo en cuenta este enfoque y lo establecido en la normatividad vigente de Colombia en particular lo referido a la Resolución 3280 de 2018 donde reconoce el transitar de la persona dentro de sus entornos y por tanto las necesidades de salud que se deben reflejar en los recursos necesarios para la financiación de la demanda de servicios de salud, el ajustador de riesgo correspondiente a la edad reconocerá 17 grupos etarios así:

Tabla 1: Grupos edad

Grupo de edad	Sexo	Curso de Vida
Menor de un año	Ambos	Primera Infancia
1 a 5 Años	Ambos	Primera Infancia
6 a 11 Años	Ambos	Infancia
12 a 17 Años	Hombres	Adolescencia
12 a 17 Años	Mujeres	Adolescencia
18 a 28 Años	Hombres	Juventud
18 a 28 Años	Mujeres	Juventud
29 a 44 Años	Hombres	Adulthood
29 a 44 Años	Mujeres	Adulthood
45 a 49 Años	Ambos	Adulthood
50 a 54 Años	Ambos	Adulthood
55 a 59 Años	Ambos	Adulthood
60 a 64 Años	Ambos	Vejez
65 a 69 Años	Ambos	Vejez
70 a 74 Años	Ambos	Vejez
75 a 79 Años	Ambos	Vejez
Mayores de 80 años	Ambos	Vejez

Por lo anterior, los modelos estadísticos se analizarán para estos 17 grupos con el fin de estimar la frecuencia y la severidad ajustada al gasto en salud de la población que conforma cada uno de los grupos etarios propuestos, es importante tener en cuenta que esta agrupación reconoce 17 grupos de edad abiertos por sexo en 6 de ellos debido a la edad reproductiva de la mujer donde su demanda de servicios de salud puede ser mayor a la de un hombre y por tanto la prima debe reconocer y garantizar la financiación de todos los servicios y tecnologías en salud.

La UPC actual reconoce 14 grupos etarios, pero con el enfoque diferente pues son grupos de riesgo en salud y aunque su conformación se realiza teniendo en cuenta el gasto en los servicios de salud y algunas condiciones epidemiológicas propias no responden directamente al curso de vida.

## 2. Zonas Geográficas

Uno de los factores que influye en la salud de la población y en particular en Colombia, es el lugar de residencia, las condiciones geográficas pueden afectar el acceso a los servicios de salud por lo cual es necesario establecer recursos adicionales con el fin de garantizar la atención oportuna, es así como de acuerdo a [de salud y Protección Social, ] se reconocen 6 zonas las cuales se conforman teniendo en cuenta dimensiones que intervienen en la prestación de servicios de salud cada una de ellas se mide a través de indicadores que provienen de fuentes oficiales del país, los cuales se presentan a continuación:

Dimensión	Indicador	Fuente
Económico	Peso relativo municipal en el valor agregado nacional%	DANE- cuentas nacionales
Oferta de Servicios	Prestadores de servicios de Salud	MSPS-REPS
Condiciones de vida y acceso a servicios	IPM Total - Índice de pobreza multidimensional	DANE - Censo 2018
Urbano Regional	Población Rural % Rural	DANE
Urbano Regional	Distancia a la Capital (Departamento)	Google
Resultados Salud	Tasa Mortalidad General	SISPRO
Institucional	Nuevo IDF - Índice de desempeño fiscal	DNP

Tabla 2: Dimensiones e Indicadores para la zonificación para la UPC

Con los indicadores anteriores se realiza un análisis de componentes principales, con la finalidad de establecer las zonas que se tendrán en cuenta para el ajuste de riesgo, es así como se presentan 6 zonas frente a las 4 que actualmente reconocen UPC adicionales.

A continuación se presenta el numero de municipios que conforman cada una de las zonas propuestas así como la distribución geográficas de las mismas.

Tabla 3: Zonas

Zona 2022	Municipios	Participación
Normal con desempeño mayor	46	4.1
Normal con desempeño menor	376	33.2
Rural	506	44.7
Rural Disperso	201	17.8
Alejada	2	0.2
Bogotá	1	0.1

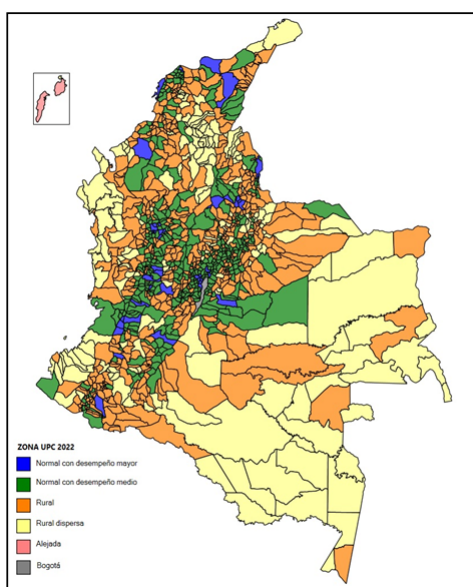


Figura 4: Mapa Nuevas Zonas [de salud y Protección Social, ]

A lo largo de los estudios de modelos de riesgo en salud en los sistemas de aseguramiento los dos ajustadores presentados son los mas comunes y básicos dado que su información es de baja manipulación por los agentes del sistema, por lo anterior son los que se desarrollarán en la aplicación de los modelos estadísticos propuestos.

### 2.2.3. Modelos estadísticos

De acuerdo a [Boucher et al., 2007] los datos de seguros suelen presentar una dispersión excesiva que puede deberse a la omisión de algunas variables de clasificación de seguros según el ramo que se este

analizando.

Según [Duncan, 2018], un modelo es una abstracción del mundo real que intenta capturar las características sobresalientes de los comportamientos humanos complejos en términos matemáticos y estadísticos simples, Un modelo por tanto es un conjunto de coeficientes que se pueden aplicar dentro de un entorno de producción (datos) para generar una predicción de algún resultado.

Los modelos estadísticos del presente trabajo se basan en los dos componentes principales de una prima de riesgo esto es la frecuencia y la severidad, en general y de acuerdo a [Frees, 2014], las compañías de seguro aceptan el pago de una prima a cambio se compromete a indemnizar al tomador de la póliza de la ocurrencia de un evento incierto.

Dado los dos componentes de la prima pura o de riesgo; la frecuencia se analiza teniendo en cuenta el aspecto proporcional a la exposición, y la severidad en términos de pérdida por siniestro o reclamaciones y por tanto no depende de la exposición.

En este sentido se han desarrollado modelos para dar respuesta a estas dos variables dentro de los que se encuentran los Modelos Lineales Generalizados que de acuerdo a [Goldburd and Guller, 2019] son un medio para modelar la relación entre una variable cuyo resultado se desea predecir y una o mas variables explicativas, en este mismo sentido el objetivo es explicar la mayor parte de la variabilidad usando los predictores seleccionados, es decir reducir lo mayor posible el componente aleatorio.

Estos modelos pueden ser utilizados en seguros para estimar la variable objetivo las cuales pueden ser:

1. Frecuencia de Reclamación
2. Frecuencia de Severidad
3. Prima Pura
4. Razón de pérdida

De acuerdo a [Goldburd and Guller, 2019] los modelos lineales generalizados se construyen teniendo en cuenta dos componentes así:

**Componente sistemático:** una porción de la variación de los resultados que esta relacionado con los valores de los predictores. La forma funcional por tanto se expresa como sigue

$$g(\mu_i) = \beta_0 + \beta_1 x_{i_1} + \beta_2 x_{i_2} + \dots \beta_p x_{i_p} \quad (1)$$

El lado izquierdo de la ecuación se denomina **función de enlace** y el lado derecho es llamado **predictor lineal**, la función de enlace por tanto y de acuerdo a [Duncan, 2018] conecta el componente sistemático con el componente aleatorio.

La función de enlace se encarga de linealizar la relación entre la variable dependiente y la(s) variable(s) independiente(s) mediante la transformación de la variable respuesta.

**Componente aleatorio:** es la parte que no se explica por los predictores del modelo es decir por causas ajenas al modelo. Este componente sigue la probabilidad de distribución que pertenece a la familia exponencial, y para cualquier tipo de riesgo se denota como

$$y_i \sim Exponencial(\mu_i, \phi) \quad (2)$$

Como se expreso anteriormente, este componente responde a la familia exponencial natural de forma [Duncan, 2018].

$$f_y(y_i) = exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i \phi) \right\} \quad (3)$$

Ahora bien, teniendo en cuenta que la prima pura de riesgo se compone de dos factores, cada uno tiene características propias que ocasiona que su modelaje deba hacerse por separado, a continuación se presentan las principales distribuciones que se utilizan en este tipo de modelos de tarificación:

## FRECUENCIA

Dentro de las distribuciones mas comunes para modelar la frecuencia, se encuentran;

- **Poisson:** modela el recuento de eventos que ocurren dentro de un período fijo, en este caso las reclamaciones que se pueden presentar. Aunque la distribución de Poisson suele ser una distribución discreta (definida sólo para valores) su implementación en un GLM le permite asumir valores fraccionarios también.

Esta característica es útil al modelar la frecuencia de reclamos, donde el recuento de reclamos se divide por un valor como la exposición o la prima, lo que da como resultado una variable objetivo no integral.

La varianza en esta distribución es  $V(\mu) = \mu$  significa que la varianza incrementa linealmente con la media, es decir que la varianza debe ser igual que a la media, sin embargo la frecuencia suele tener una variación mayor lo que se conoce como sobredispersión esto se presenta principalmente por el nivel de riesgo de los asegurados que no es posible explicar con la distribución de poisson, en esta situación es aconsejable utilizar un modelo de poisson sobredispersado que permite que la varianza tome cualquier valor.

- **Binomial Negativa:** Esta clase de distribución es una alternativa a la sobredispersión y matemáticamente se expresa así:

$$y \sim \text{Poisson}(\mu = \theta), \theta \sim \text{gamma}(\dots) \quad (4)$$

Lo anterior presenta por tanto la siguiente forma funcional

$$\text{NegBinomial}(y = k | r, \theta) = P(x = k | r, \theta) = \frac{\Gamma(r + k)}{\Gamma(r)k!} \theta^r (1 - \theta)^k \quad (5)$$

En palabras, el resultado  $y$  tiene una distribución de Poisson con una media de  $\theta$ , donde,  $\theta$  es en sí misma aleatoria y con distribución gamma. Esta combinación resulta que  $y$  siga una distribución binomial negativa, para esta distribución el parámetro  $\phi$  se restringe a 1.

En esta distribución aparece un tercer parámetro,  $k$  llamado de sobredispersión (overdispersed), la cual se presenta de acuerdo a [Andrews, 2021], cuando la varianza es mas grande que la media en caso que la varianza sea menor que la media se habla de baja dispersión (underdispersed), es así como la varianza es  $V(\mu) = \mu(1 + k\mu)$  entonces el parámetro  $k$  sirve para "inflar" la varianza por encima de la media, lo que sería la varianza implícita en la distribución de Poisson.

El fenómeno de la sobredispersión en la data es mas común cuando se utiliza la distribución de poisson.

La distribución binomial negativa por tanto indica la probabilidad de observar cualquier número de fallas antes de que se produzca un éxito.

De hecho, cuando  $k$  se acerca a cero, la distribución binomial negativa se acerca a Poisson.

Según lo expuesto por [Durbán, 2014] a principal diferencia entre la distribución de Poisson y la Binomial Negativa, es que, aunque ambas cuentan el número de veces que ocurre algo, en la distribución de Poisson se conoce cuantas veces no ocurrió, contrario sucede en la Binomial Negativa donde es posible conocerlo.

## SEVERIDAD

Las dos distribuciones mas comunes son Gamma y la LogNormal.

- **Gamma:** una distribución gamma está sesgada a la derecha, con un pico agudo y una larga cola a la derecha y tiene un límite inferior en cero. Como estas características tienden a exhibirse por distribuciones empíricas de severidad de reclamos, la gamma es un ajuste natural para modelar la severidad en un GLM. La función de varianza gamma es  $V(\mu) = \mu^2$  lo que significa que la varianza asumida por la severidad para cualquier reclamación en los modelos gamma es proporcional a la función exponencial de la media.
- **LogNormal:** una distribución LogNormal, es una distribución de probabilidad que define una variable aleatoria cuyo logaritmo sigue una distribución normal. Sigue por tanto la siguiente forma funcional.

$$X \sim \text{Lognormal}(\mu, \sigma^2)$$

Esta distribución solo puede ser utilizado en valores reales positivos, dado que el logaritmo no admite valores negativos, lo cual es una buena opción para modelar la severidad dado que no puede existir un valor promedio pagado negativo.

La función de varianza  $\text{Var}(X) = e^{\sigma^2} - 1) * e^{2\mu + \sigma^2}$

Una vez se han creado los dos modelos para frecuencia y severidad, y según lo establecido por [Gutiérrez Meléndez, 2022], es necesario otorgar valores a la frecuencia y severidad modeladas para cada una de nuestras pólizas. Tras hallar estos valores, para la frecuencia tendremos una ponderación para saber cuál es el número de siniestros de cada asegurado según la exposición de cada uno, al multiplicar cada una las frecuencias y severidades se obtiene la prima pura.

La prima pura se estimará para cada uno de los grupos etarios y de las zonas, dadas las características propias de los expuestos principalmente, los modelos presentados en este apartado son los más comunes y que se adaptan a las condiciones de la información disponible para el presente trabajo, Existen algunas distribuciones adicionales que por sus características no fue factible su utilización como lo es la distribución Weibull, así mismo se analizó modelos como:

Es importante tener en cuenta que dentro de los modelos analizados para la estimación del UPC se analizaron otras distribuciones como los son:

**Poisson cero - inflado:** Otra clase de modelos usados para la estimación de las frecuencias para la prima de riesgo son los llamados **Poisson cero - inflado**, según [Salinas-Rodríguez et al., 2009] es posible que el mecanismo aleatorio que dio origen a los datos de conteo en una distribución poisson de conteo general, muestre una mayor concentración para algún valor específico que puede ser el cero o cualquier otro valor positivo. Esto implica que dicho valor tiene una mayor probabilidad de ocurrencia que la especificada en una distribución Poisson o cualquier otra distribución.

Esta clase de modelos son considerados modelos probabilísticos mixtos, dado que su análisis se divide en dos partes:

- Una distribución Poisson
- Una distribución de toda la masa de ceros

**Hurdle:** De acuerdo [et al., 2009] Al igual que el modelo de regresión para datos de conteo cero-inflado, el modelo Hurdle asume la presencia de una mezcla de distribuciones, sólo que dicho modelo tiene una interpretación en dos partes.\* La primera se refiere a un modelo con variable de respuesta binaria y la segunda a un modelo de datos de conteo truncado-en-cero. En consecuencia, esta configuración en dos partes permite la interpretación de que los valores positivos se generan toda vez que el umbral (Hurdle) en cero se ha cruzado. Por consiguiente, la primera parte modela la probabilidad de que el umbral sea cruzado, mientras que la segunda modela el valor esperado de los valores positivos.

Tanto el modelo de Poisson Cero Inflados como los modelos Hurdle se evaluaron para su posible aplicación pero dadas las características de la información utilizada en el presente trabajo, se decidió no aplicarlos.

## 3. Metodología

### 3.1. Introducción a la base de datos

La base de información a utilizar extrae las variables de tres fuentes principales la primera corresponde al reporte que realizan las EPS al Ministerio de Salud y Protección Social correspondiente a los años 2018 y 2019, por la calidad de la misma se utilizará la información correspondiente al régimen contributivo, que de acuerdo a lo establecido en [de Salud de Protección Social, 2021] la información se agrega en seis tipos de registros y cada uno de ellos contiene variables generales referente a la prestación de los servicios y tecnologías en salud.

La segunda fuente de información corresponde a la base poblacional que contiene las personas por las que el gobierno reconoce el recurso de la UPC demanden o no servicios, y la tercera fuente de información es la referente a la afiliación es decir la Base de Datos Única de Afiliados BDUA.

Por lo anterior la base contiene las variables que se presentan a continuación.

Variables	Descripción
Grupo Etario	Agrupación por edad de la Población afiliada al Régimen Contributivo
Sexo	Hace referencia al sexo por Grupo Etario (Femenino o Masculino)
Reclamaciones	Registra el número de atenciones recibidas por la población de acuerdo de su Grupo Etario
Valor de las reclamaciones	Valor Total Reportado por los Servicios prestados por la población de cada grupo de edad.
Póliza	Valor girado por el SGSSS a las EPS por Cada Grupo Etario por concepto de la prima y la demanda inducida
Días compensados	Es la sumatoria de los días compensados por cada uno de los grupos etarios
Zona	Agrupación de los municipios del país de acuerdo a las características establecidas
Expuestos	Corresponde al Total de la Población afiliada agrupada por grupo etario
Severidad	Es la estimación de la severidad la cual es el cálculo entre el Valor Total y el número de registros

Tabla 4: Variables

Las características propias de la información se presentan en el Anexo 1, esta información no presenta datos faltantes dado que se trabaja con todas las prestaciones de tecnologías y servicios en salud prestados y reportados por las aseguradoras EPS, el sesgo que presenta se debe a que no se cuenta con información del régimen subsidiado, y por tanto se puede ver afectados los resultados en aquellas zonas donde la concentración de población de este régimen es mayor.

La información utilizada en el presente trabajo se presenta de manera agregada, toda vez que el detalle de la misma no es necesario para el fin ultimo de desarrollo de estimación del UPC utilizando modelos lineales generalizados.

## 3.2. Análisis Exploratorio de Datos

### 3.2.1. Zona Alejada

A continuación se presentan las estadísticas descriptivas para los años 2018 y 2019 para la zona alejada conformada por 2 municipios de Colombia en el Departamento de San Andrés y Providencia.

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	2.665	263.464.564	696.054.720	236	74.268
1st Qu.	4.667	546.744.003	1.065.990.773	605	95.036
Median	6.427	764.802.921	1.467.547.505	1.536	105.627
Mean	8.556	925.873.082	1.817.700.255	1.819	115.305
3rd Qu.	9.625	1.146.102.783	2.186.446.141	2.079	126.965
Max.	33.818	3.319.882.319	6.473.571.999	6.435	195.030

Figura 5: Estadísticos Descriptivos Zona Alejada Año 2018

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	6.304	388.533.573	732.747.423	254	53.551
1st Qu.	11.589	820.306.953	1.114.164.898	660	62.009
Median	14.908	1.170.297.374	1.637.079.664	1.559	67.107
Mean	20.264	1.386.639.852	1.935.452.969	1.864	69.423
3rd Qu.	23.499	1.603.662.362	2.362.559.877	2.075	73.758
Max.	72.216	4.847.666.176	6.947.647.024	6.818	93.636

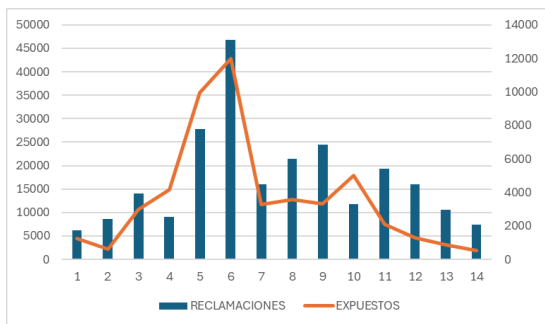
Figura 6: Estadísticos Descriptivos Zona Alejada Año 2019

A continuación se presenta el análisis visual de cada una de las variables explicativas en los modelos propuestos en cada uno de los años bajo estudio 2018 y 2019.

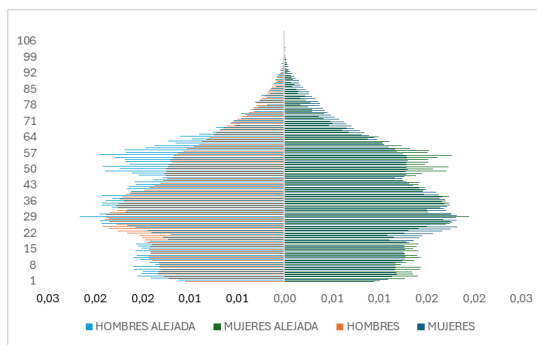
- Reclamaciones y Expuestos

Al analizar el número de reclamaciones es decir el numero de atenciones de la población de la zona alejada teniendo en cuenta los grupos etarios se presenta en ambos años un mínimo en los menores de un año del genero femenino, y el máximo en el grupo de 29 a 44 años del genero femenino, lo anterior derivado de la distribución poblacional.

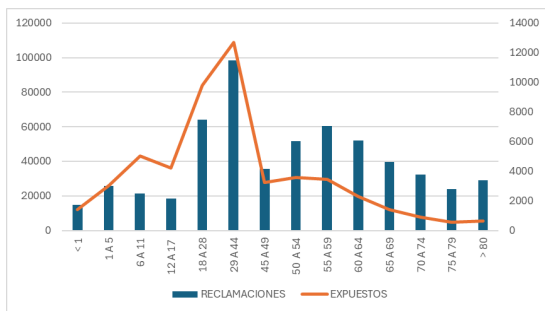
A continuación se presenta las gráficas de reclamaciones y expuestos de cada uno de los años, así mismo se presenta la pirámide poblacional de la zona alejada, comparada con la pirámide poblacional del país de las EPS de la selección, de cada uno de los años.



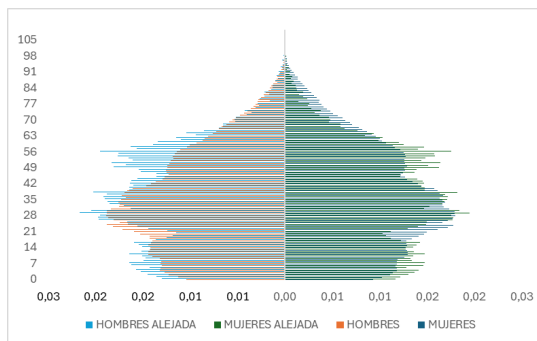
(a) Histograma Reclamaciones y Expuestos Zona Alejada Año 2018



(b) Pirámide Poblacional Zona Alejada Año 2018



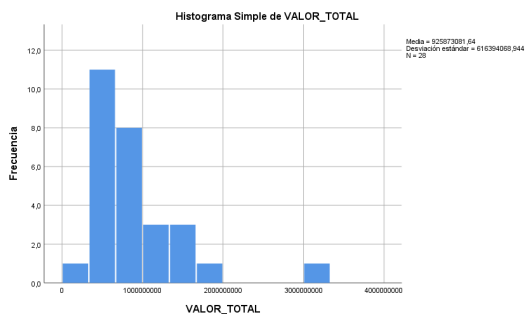
(a) Histograma Reclamaciones y Expuestos Zona Alejada Año 2019



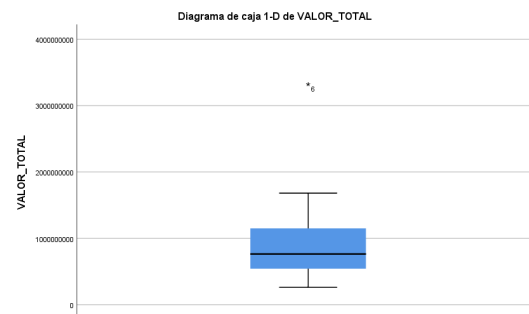
(b) Pirámide Poblacional Zona Alejada Año 2019

■ Valor Reclamaciones

El Valor de las reclamaciones corresponde a al valor reportado y reconocido por las aseguradoras para la prestación de las tecnologías o servicios de salud demandados por los afiliados en la zona alejada, en concordancia con las reclamaciones y los expuestos en los años se presenta el mayor valor reportado en el grupo de 29 a 44 a años y por tanto el valor mínimo reportado se observa en el grupo etario menores de un año.



(a) Histograma Valor Reclamaciones Zona Alejada Año 2018



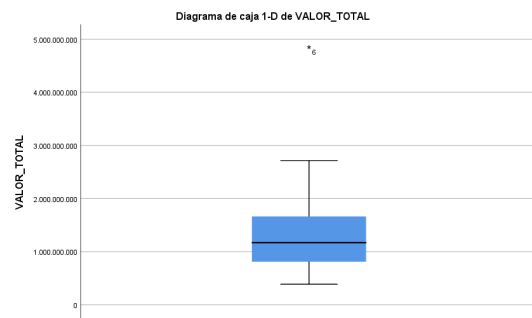
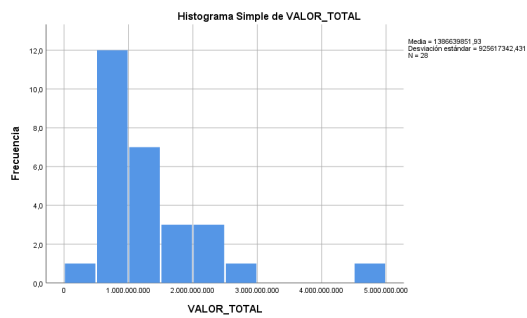
(b) Cajas y Bigotes Valor Reclamaciones Zona Alejada Año 2018

■ Póliza

El Valor de la póliza corresponde al reconocimiento realizado por el gobierno, por concepto de la prima es decir de la Unidad de Pago por Capitación UPC y el valor que se reconoce para la demanda inducida que hace referencia a las actividades de promoción de la salud y prevención de la enfermedad encaminadas a la educación.

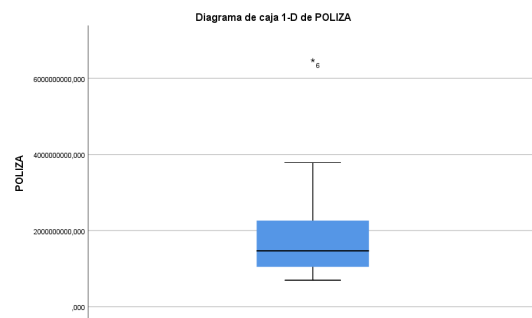
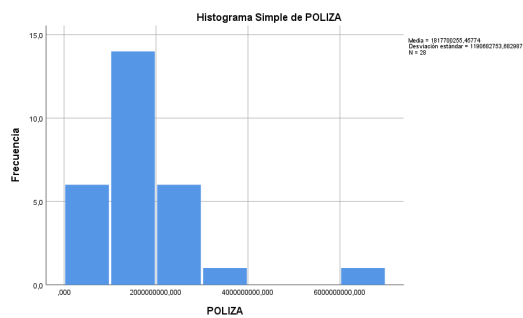
El valor máximo para los dos años se encuentra en grupo de 29 a 44 años con valores de prima promedio de de \$1.005.994 y \$1.019.015,40 respectivamente.

Así mismo el valor mínimo se presenta en el grupo etario de 6 a 11 año con valor de prima promedio de \$291.725, y de \$ 297.502 respectivamente.

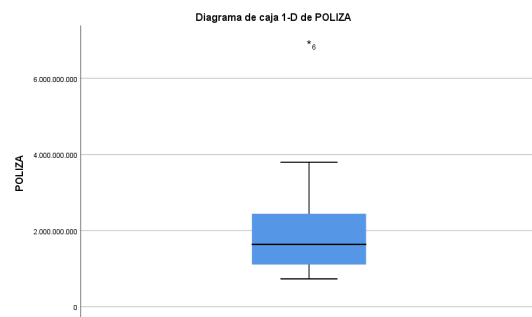
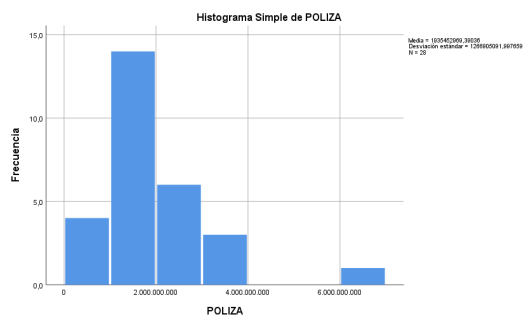


(a) Histograma Valor Reclamaciones Zona Alejada Año 2019 (b) Cajas y Bigotes Valor Reclamaciones Zona Alejada Año 2019

En las gráficas que a continuación se presentan, el histograma de la distribución de variable así como el gráfico de cajas y bigotes para cada uno de los años en estudio.



(a) Histograma de Póliza Zona Alejada Año 2018 (b) Cajas y Bigotes Póliza Zona Alejada Año 2018



(a) Histograma de Póliza Zona Alejada Año 2019 (b) Cajas y Bigotes Póliza Zona Alejada Año 2019

### 3.2.2. Zona Bogotá

A continuación se presentan las estadísticas descriptivas para los años 2018 y 2019 para la zona Bogotá, dado que es la capital del país y por lo tanto con mayor densidad poblacional se convierte en una zona de análisis.

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	1.067.009	72.334.390.862	70.540.526.152	43.064	37.345
1st Qu.	2.024.989	105.288.603.112	111.605.137.093	83.893	44.101
Median	3.063.449	175.495.994.796	192.726.584.235	173.707	49.313
Mean	3.790.845	186.857.115.009	212.120.654.391	260.999	52.836
3rd Qu.	4.921.143	217.015.006.839	258.596.439.217	283.451	59.701
Max.	12.886.604	557.424.229.110	748.904.429.857	950.733	93.837

Figura 13: Estadísticos Descriptivos Zona Bogotá Año 2018

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	972.510	72.815.116.190	72.349.033.535	45.364	37.957
1st Qu.	2.085.797	99.812.520.657	115.389.392.971	89.709	42.813
Median	3.664.418	189.700.843.978	207.008.361.783	175.712	46.110
Mean	4.215.088	200.602.166.950	228.222.488.736	262.683	50.733
3rd Qu.	5.552.319	240.843.531.304	281.824.275.966	272.379	54.908
Max.	14.013.013	620.645.972.917	802.373.631.727	973.360	91.640

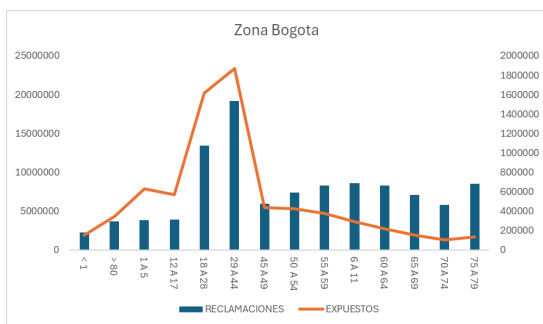
Figura 14: Estadísticos Descriptivos Zona Bogotá Año 2019

A continuación se presenta el análisis visual de cada una de las variables para cada uno de los años bajo estudio.

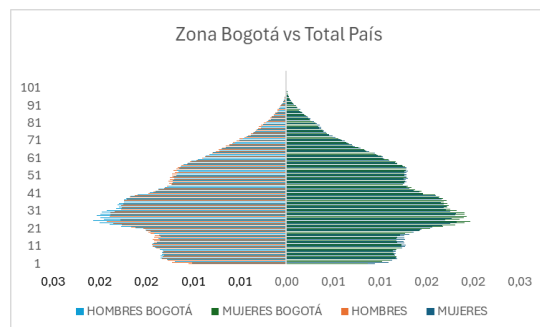
■ Reclamaciones y Expuestos

Al analizar el número de reclamaciones es decir el número de atenciones de la población de la zona Bogotá, teniendo en cuenta los grupos etarios se presenta en ambos años un mínimo en los menores de un año del genero femenino, y las mayoría de reclamaciones en el grupo etario de 29 a 44 años del mismo genero, lo cual se relaciona con la distribución poblacional, es importante mencionar que el menor numero de expuestos en ambos años se presenta en el grupo etario de 75 a 79 años.

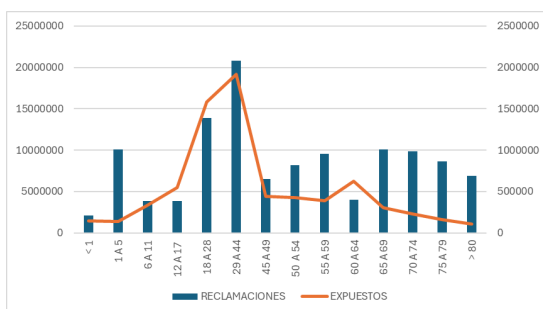
A continuación se presenta las gráficas de reclamaciones y expuestos de cada uno de los años, así mismo se presenta la pirámide poblacional de la zona alejada, comparada con la pirámide poblacional del país de las EPS de la selección, de cada uno de los años.



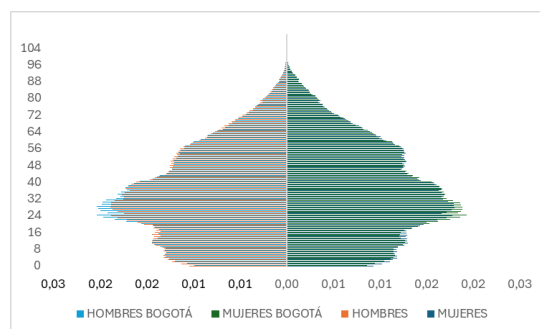
(a) Histograma Reclamaciones y Expuestos Zona Bogotá Año 2018



(b) Pirámide Poblacional Zona Bogotá Año 2018



(a) Histograma Reclamaciones y Expuestos Zona Bogotá Año 2019

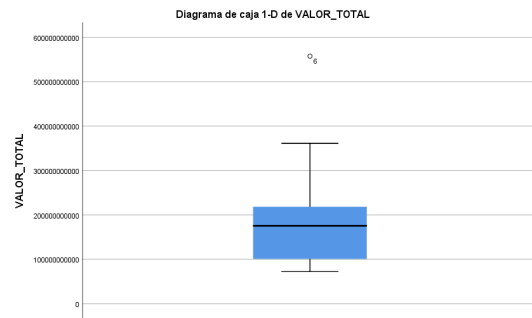
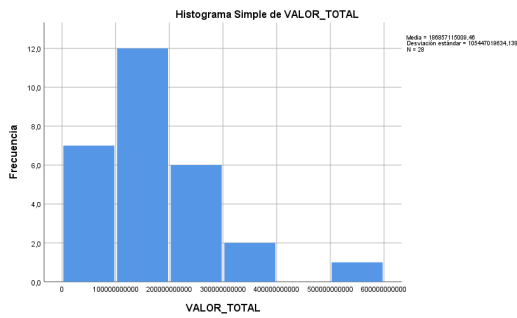


(b) Pirámide Poblacional Zona Bogotá Año 2019

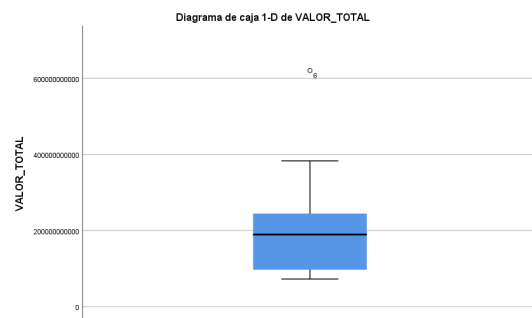
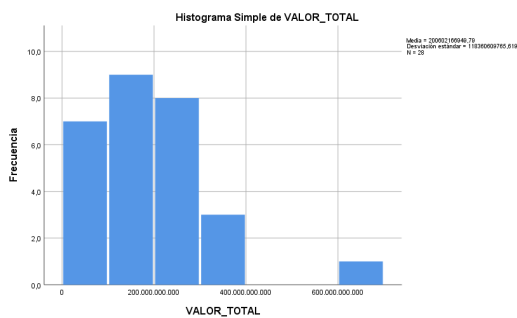
■ Valor Reclamaciones

El Valor de las reclamaciones corresponde a al valor reportado y reconocido por las aseguradoras para la prestación de las tecnologías o servicios de salud demandados por los afiliados en la Zona Bogotá, en concordancia con las reclamaciones y los expuestos en los años se presenta el mayor valor reportado en el grupo de 29 a 44 años del genero femenino y el valor mínimo se presenta en el grupo de 6 a 11 años del genero femenino, lo anterior se relaciona con las reclamaciones y los expuestos.

A continuación se presenta el histograma del valor de las reclamaciones y un gráfico de cajas y bigotes que determina lo expresado anteriormente.



(a) Histograma Valor Reclamaciones Zona Bogotá Año 2018 (b) Cajas y Bigotes Valor Reclamaciones Zona Bogotá Año 2018



(a) Histograma Valor Reclamaciones Zona Bogotá Año 2019 (b) Cajas y Bigotes Valor Reclamaciones Zona Bogotá Año 2019

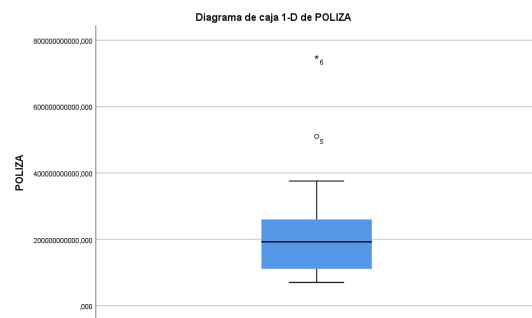
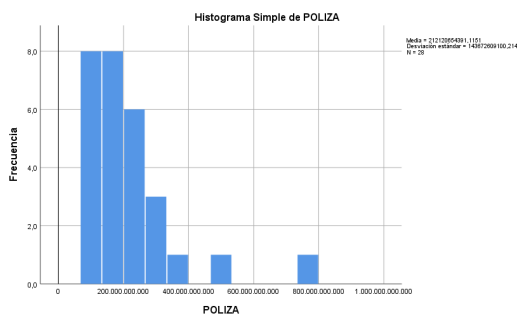
■ Póliza

El Valor de la póliza corresponde al reconocimiento realizado por el gobierno, por concepto de la prima es decir de la Unidad de Pago por Capitación UPC y el valor que se reconoce para la demanda inducida que hace referencia a las actividades de promoción de la salud y prevención de la enfermedad encaminadas a la educación.

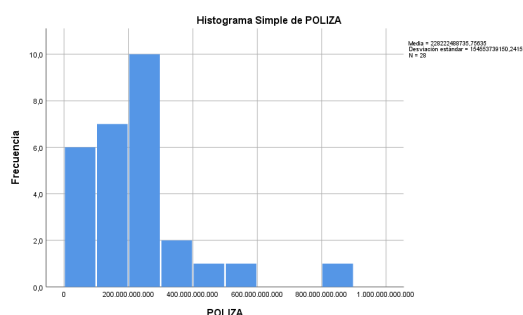
El valor máximo para los dos años se encuentra en grupo de 29 a 44 años con valores de prima promedio de de\$787.712 y \$824.333,89. respectivamente.

Así mismo el valor mínimo se presenta en el grupo etario de 6 a 11 año con valor de prima promedio de \$228.006, y de \$235.619,32 respectivamente.

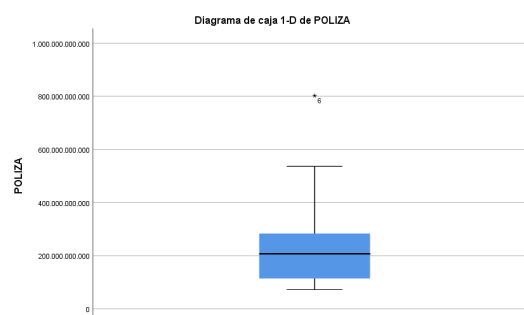
En las gráficas que a continuación se presentan, el histograma de la distribución de variable así como el gráfico de cajas y bigotes, para cada uno de los años bajo estudio.



(a) Histograma de Póliza Zona Bogotá Año 2018 (b) Cajas y Bigotes Póliza Zona Bogotá Año 2018



(a) Histograma de Póliza Zona Bogotá Año 2019



(b) Cajas y Bigotes Póliza Zona Bogotá Año 2019

### 3.2.3. Zona Normal Mayor

A continuación se presentan las estadísticas descriptivas para la zona Normal Mayor para los años 2018 y 2019, a esta zona pertenecen 32 municipios del país.

ESTADISTICOS	RECLAMACIONES	VALOR TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
<b>Min.</b>	1.811.986	120.743.001.102	115.046.811.444	86.979	39.376
<b>1st Qu.</b>	3.370.936	157.525.221.008	187.711.584.197	168.795	43.762
<b>Median</b>	6.145.486	323.855.935.078	363.621.579.179	305.117	50.088
<b>Mean</b>	6.919.727	335.414.412.767	380.524.833.651	455.785	50.792
<b>3rd Qu.</b>	9.247.950	409.763.736.874	447.198.415.909	473.814	56.130
<b>Max.</b>	22.398.185	1.003.321.914.310	1.258.739.565.410	1.636.932	75.798

Figura 21: Estadísticos Descriptivos Zona Normal Mayor Año 2018

ESTADISTICOS	RECLAMACIONES	VALOR TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
<b>Min.</b>	1.753.334	114.296.899.659	119.237.210.309	90.416	34.829
<b>1st Qu.</b>	3.768.272	160.264.448.394	198.556.487.495	178.661	39.871
<b>Median</b>	7.290.446	348.197.458.353	393.605.838.725	318.709	45.468
<b>Mean</b>	8.044.375	352.514.457.431	408.492.090.826	464.481	45.854
<b>3rd Qu.</b>	10.623.496	425.412.298.638	475.988.121.843	467.652	50.078
<b>Max.</b>	25.620.655	1.042.118.126.180	1.348.614.875.990	1.699.219	69.528

Figura 22: Estadísticos Descriptivos Zona Normal Mayor Año 2019

A continuación se presenta el análisis visual de cada una de las variables para los años 2018 y 2019.

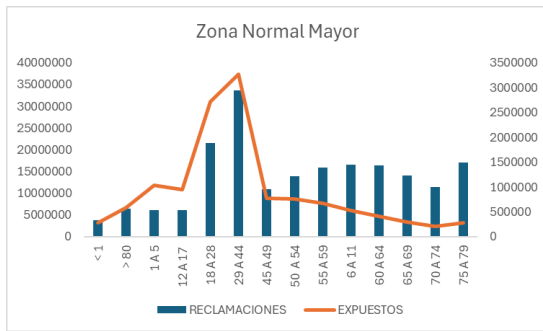
#### ■ Reclamaciones y Expuestos

Al analizar el número de reclamaciones es decir el número de atenciones de la población de la Zona Normal Mayor teniendo en cuenta los grupos etarios se presenta en ambos años un mínimo en los menores de un año del género femenino, y el máximo en el grupo de 29 a 44 años del género femenino, lo anterior derivado de la distribución poblacional, ahora bien el mayor número de expuestos independientemente de si tuvieron o no reclamaciones se encuentra en el grupo de 29 a 44 años del género masculino para los dos años de estudio, y el menor se encuentra en grupo etario de 75 a 79 años del género masculino, en el número de atenciones no presenta una relación directa con el número de expuestos, lo anterior puede derivarse las enfermedades propias de los municipios que conforman esta zona, es necesario realizar un análisis más exhaustivo.

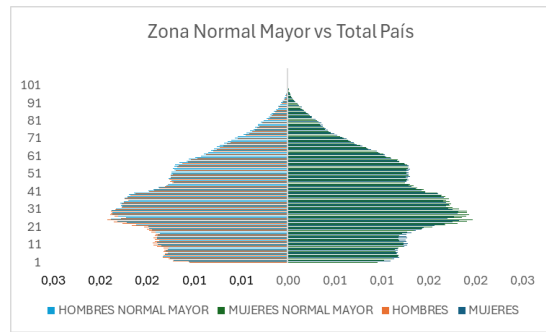
A continuación se presenta las gráficas de reclamaciones y expuestos de cada uno de los años, así mismo se presenta la pirámide poblacional de la zona alejada, comparada con la pirámide poblacional del país de las EPS de la selección, de cada uno de los años.

#### ■ Valor Reclamaciones

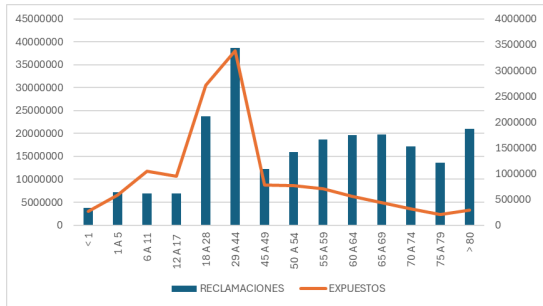
El Valor de las reclamaciones corresponde a al valor reportado y reconocido por las aseguradoras para la prestación de las tecnologías o servicios de salud demandados por los afiliados en la Zona



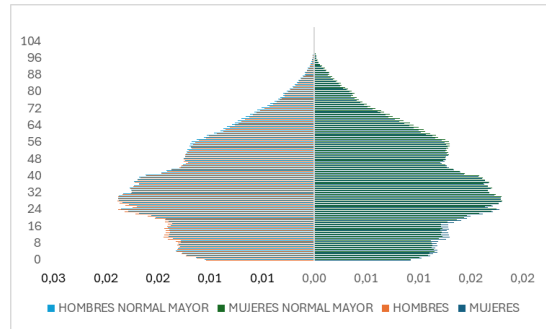
(a) Histograma Reclamaciones y Expuestos Zona Normal Mayor Año 2018



(b) Pirámide Poblacional Zona Normal Mayor Año 2018



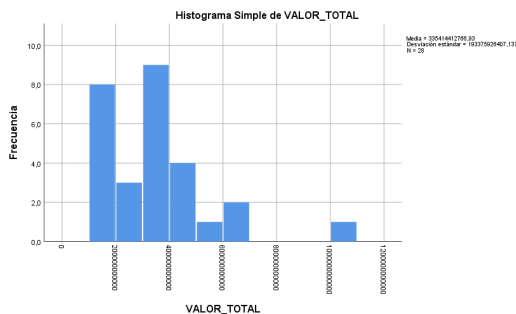
(a) Histograma Reclamaciones y Expuestos Zona Normal Mayor Año 2019



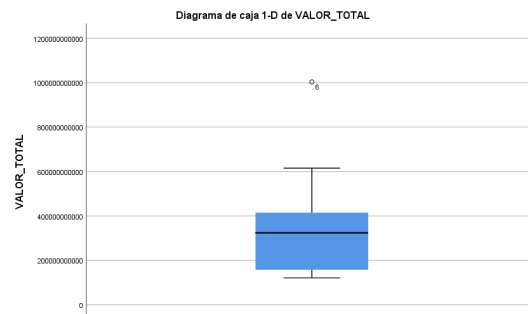
(b) Pirámide Poblacional Zona Normal Mayor Año 2019

Normal Mayor, el valor máximo para los años 2018 y 2019 se registra en el grupo etario de 29 a 44 años del genero femenino, y el valor menor se registra en el grupo de 6 a 11 en el 2018 y en grupo menor de un año en el año 2019, lo anterior evidencia la dinámica poblacional que se presenta en los municipios que conforman la Zona Normal Mayor, así mismo se puede inferir que el grupo de 75 a 79 años se presentan menos reclamaciones pero las mismas pueden ser mas costosas entre algunas razones que se podrían explorar es el perfil epidemiológico de estas edades.

A continuación se presenta el histograma y el diagrama de cajas y bigotes de la distribución de la variable de valor de reclamaciones en la Zona Normal Mayor para los años 2018 y 2019.



(a) Histograma Valor Reclamaciones Zona Normal Mayor Año 2018

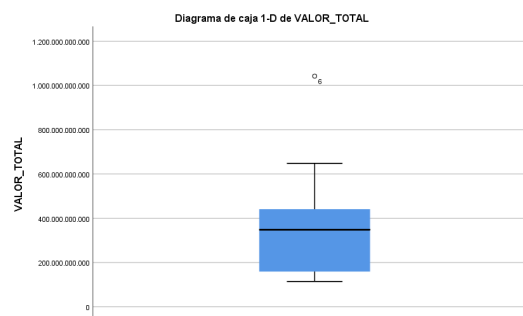
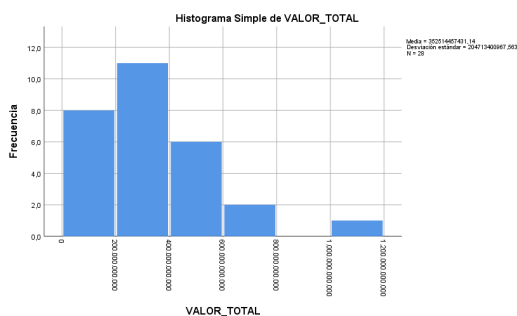


(b) Cajas y Bigotes Valor Reclamaciones Zona Normal Mayor Año 2018

- **Póliza El Valor** de la póliza corresponde al reconocimiento realizado por el gobierno, por concepto de la prima es decir de la Unidad de Pago por Capitación UPC y el valor que se reconoce para la demanda inducida que hace referencia a las actividades de promoción de la salud y prevención de la enfermedad encaminadas a la educación.

El valor máximo para los dos años se encuentra en grupo de 29 a 44 años con valores de prima promedio de \$774.177,81 y \$800.889,17 respectivamente.

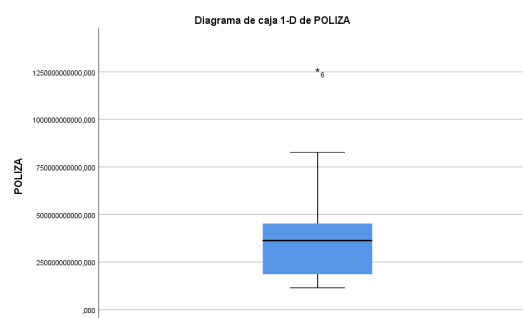
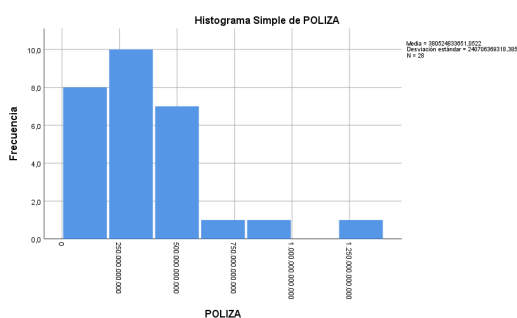
Así mismo el valor mínimo se presenta en el grupo etario de 6 a 11 año con valor de prima promedio



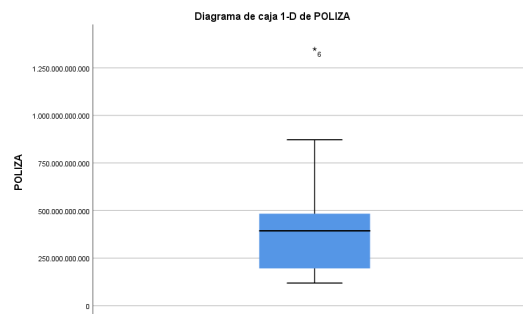
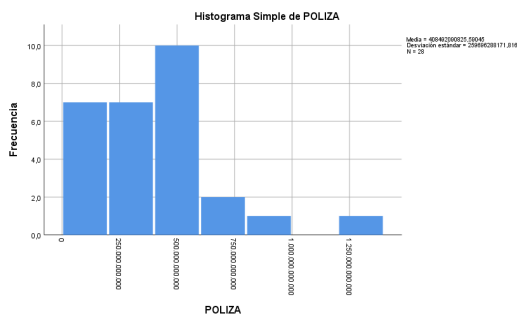
(a) Histograma Valor Reclamaciones Zona Normal Ma- (b) Cajas y Bigotes Valor Reclamaciones Zona Normal Mayor Año 2019

de \$226.474,11, y de \$233.364,21 respectivamente.

En las gráficas que a continuación se presentan, el histograma de la distribución de variable así como el gráfico de cajas y bigotes para cada uno de los años en estudio.



(a) Histograma de Póliza Zona Normal Mayor Año 2018 (b) Cajas y Bigotes Póliza Zona Normal Mayor Año 2018



(a) Histograma de Póliza Zona Normal Mayor Año 2019 (b) Cajas y Bigotes Póliza Zona Normal Mayor Año 2019

### 3.2.4. Zona Normal Medio

A continuación se presentan las estadísticas descriptivas para los años 2018 y 2019 para la zona Normal Medio, a esta zona pertenecen 215 municipios del país.

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
<b>Min.</b>	614.507	39.180.412.087	45.236.435.400	31.242	36.771
<b>1st Qu.</b>	1.223.671	59.302.010.029	69.518.480.733	55.591	44.921
<b>Median</b>	1.750.218	97.590.309.932	115.361.787.348	118.215	50.615
<b>Mean</b>	2.048.981	102.683.059.570	124.449.556.794	177.128	52.038
<b>3rd Qu.</b>	2.489.810	116.232.745.894	137.971.924.386	212.311	59.275
<b>Max.</b>	6.943.938	323.057.850.846	416.572.487.112	652.461	79.614

Figura 29: Estadísticos Descriptivos Zona Normal Medio Año 2018

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	620.228	34.796.110.183	46.161.534.772	33.009	30.961
1st Qu.	1.387.327	53.807.984.797	72.611.555.603	55.949	40.491
Median	2.081.632	105.443.886.168	126.823.905.288	124.993	44.926
Mean	2.380.329	105.380.265.638	132.824.350.774	181.972	45.445
3rd Qu.	3.018.284	123.392.267.488	149.083.024.235	208.827	52.929
Max.	8.010.504	322.988.139.807	440.557.827.770	685.646	64.845

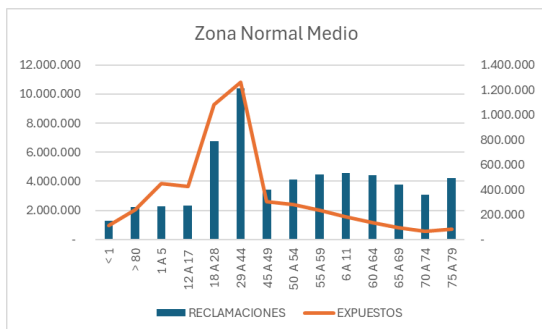
Figura 30: Estadísticos Descriptivos Zona Normal Medio Año 2019

A continuación se presenta el análisis visual de cada una de las variables para cada uno de los años 2018 y 2019.

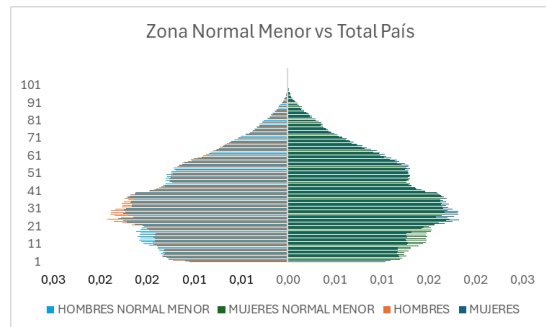
■ Reclamaciones y Expuestos

Al analizar el número de reclamaciones es decir el numero de atenciones de la población de la Zona Normal Medio teniendo en cuenta los grupos etarios se presenta en ambos años un mínimo en los menores de un año del genero femenino, y el máximo en el grupo de 29 a 44 año del genero femenino, lo anterior derivado de la distribución poblacional y la concentración de expuestos en este grupo de etario, ahora bien el numero mínimo de expuestos se presenta el grupo etario de 75 a 79 años del genero masculino, en los años 2018 y 2019, lo anterior infiere que el grupo de 75 a 79 años del genero masculino presenta menos población pero mayor numero de reclamaciones comparado con el grupo menor de un año, derivado probablemente enfermedades crónicas.

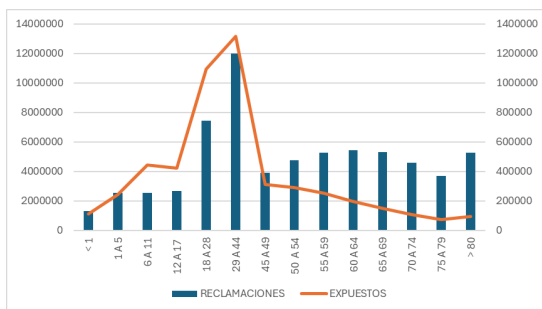
A continuación se presenta las gráficas de reclamaciones y expuestos de cada uno de los años, así mismo se presenta la pirámide poblacional de la Zona Normal Menor, comparada con la pirámide poblacional del país de las EPS de la selección, de cada uno de los años.



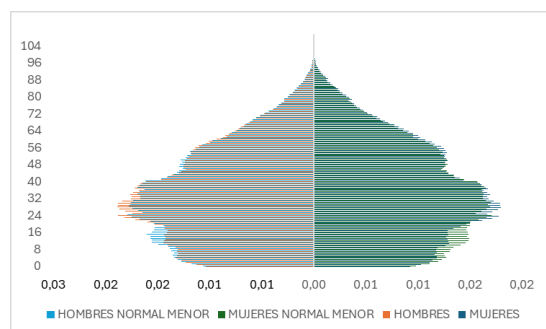
(a) Histograma Reclamaciones y Expuestos Zona Normal Medio Año 2018



(b) Pirámide Poblacional Zona Normal Medio Año 2018



(a) Histograma Reclamaciones y Expuestos Zona Normal Medio Año 2019

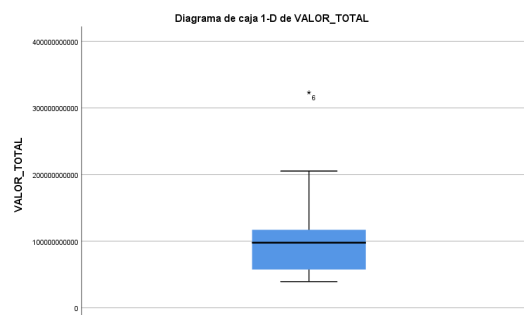
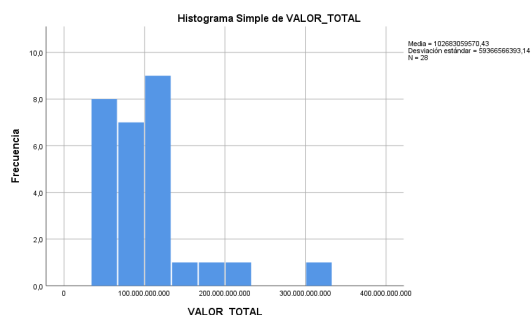


(b) Pirámide Poblacional Zona Normal Medio Año 2019

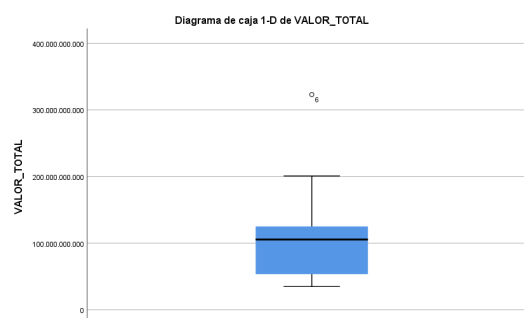
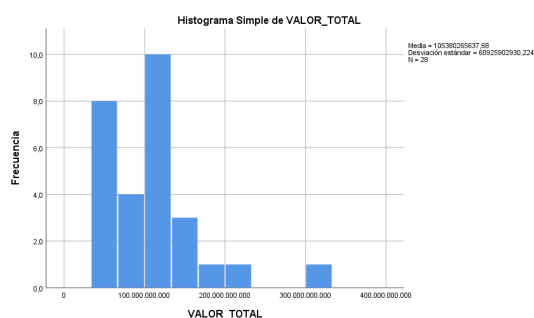
■ Valor Reclamaciones

El Valor de las reclamaciones corresponde a al valor reportado y reconocido por las aseguradoras para la prestación de las tecnologías o servicios de salud demandados por los afiliados en la Zona

Normal Menor, en concordancia con las reclamaciones y los expuestos en los años se presenta el mayor valor reportado en el grupo de 29 a 44 años del genero femenino, sin embargo el valor mínimo se presenta en el grupo etario de 1 a 5 años del genero femenino para el año 2018 y en el grupo etario menor de un año para el año 2019, lo cual presenta el dinamismo que se puede observar en los municipios que conforman esta zona. A continuación se presenta el histograma de la distribución del valor de las reclamaciones y el diagrama de cajas y bigotes para esta variable.



(a) Histograma Valor Reclamaciones Zona Normal Medio Año 2018 (b) Cajas y Bigotes Valor Reclamaciones Zona Normal Medio Año 2018



(a) Histograma Valor Reclamaciones Zona Normal Medio Año 2019 (b) Cajas y Bigotes Valor Reclamaciones Zona Normal Medio Año 2019

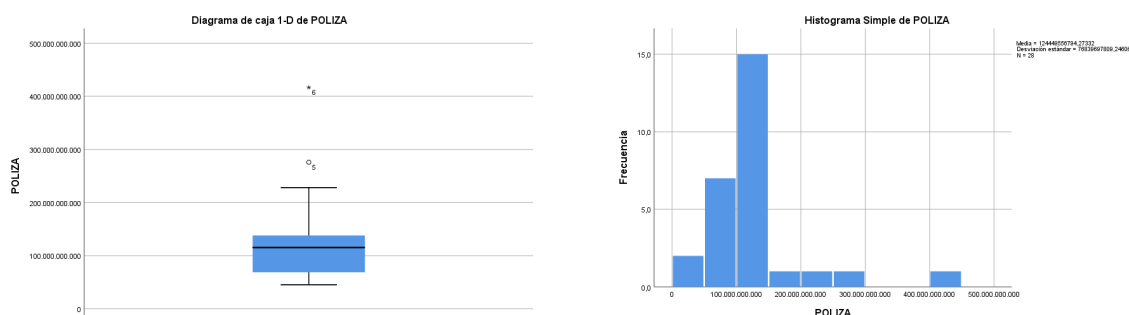
## ■ Póliza

El Valor de la póliza corresponde al reconocimiento realizado por el gobierno, por concepto de la prima es decir de la Unidad de Pago por Capitación UPC y el valor que se reconoce para la demanda inducida que hace referencia a las actividades de promoción de la salud y prevención de la enfermedad encaminadas a la educación.

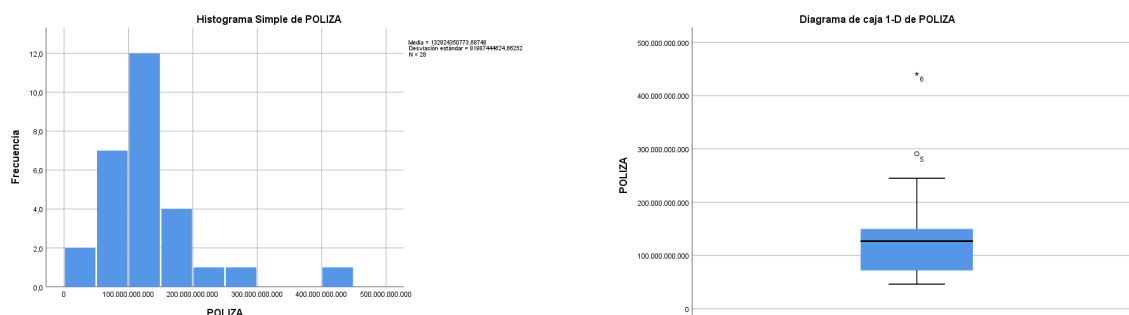
El valor máximo para los dos años se encuentra en grupo de 29 a 44 años del genero femenino con valores de prima promedio de de \$687.146 y \$697.876,91 respectivamente.

Así mismo el valor mínimo se presenta en el grupo etario de 6 a 11 año con valor de prima promedio de \$205.031, y de \$212.449,8 respectivamente.

En las gráficas que a continuación se presentan, el histograma de la distribución de variable así como el gráfico de cajas y bigotes para cada uno de los años en estudio.



(a) Histograma de Póliza Zona Normal Medio Año 2018 (b) Cajas y Bigotes Póliza Zona Normal Medio Año 2018



(a) Histograma de Póliza Zona Normal Medio Año 2019 (b) Cajas y Bigotes Póliza Zona Normal Medio Año 2019

### 3.2.5. Zona Rural

A continuación se presentan las estadísticas descriptivas para los años 2018 y 2019 para la zona Rural, a esta zona pertenecen 641 municipios del país.

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	144.270	10.683.021.507	15.237.400.258	9.543	45.172
1st Qu.	269.840	16.641.433.906	23.302.458.462	19.172	55.120
Median	364.956	25.018.149.358	36.110.582.756	41.762	61.267
Mean	432.590	26.545.394.534	40.828.282.990	67.833	63.139
3rd Qu.	480.459	28.980.624.599	41.234.416.834	74.171	70.455
Max.	1.563.960	89.969.916.412	148.098.580.469	304.056	90.591

Figura 37: Estadísticos Descriptivos Zona Rural Año 2018

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	165.020	8.565.168.454	15.502.642.520	9.764	30.399
1st Qu.	337.755	13.927.626.088	24.573.498.553	19.734	43.247
Median	465.826	24.602.162.026	38.611.910.854	43.042	47.744
Mean	527.129	25.024.977.685	43.768.717.130	69.789	48.084
3rd Qu.	582.537	28.135.831.196	44.660.605.643	72.664	54.917
Max.	1.881.912	81.877.093.444	156.993.123.118	323.318	60.744

Figura 38: Estadísticos Descriptivos Zona Rural Año 2019

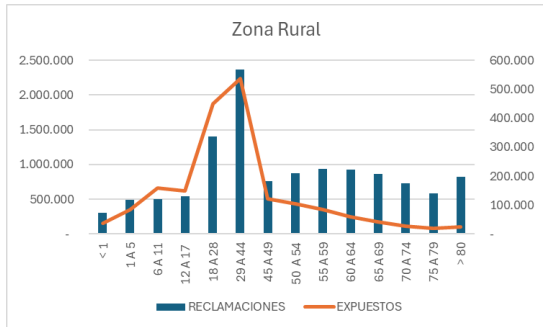
A continuación se presenta el análisis visual de cada una de las variables para los años 2018 y 2019.

#### ■ Reclamaciones y Expuestos

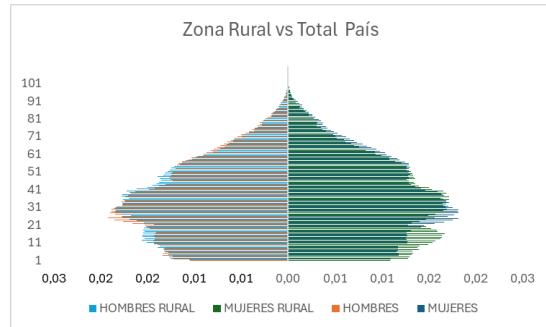
Al analizar el número de reclamaciones es decir el numero de atenciones de la población de la Zona Rural teniendo en cuenta los grupos etarios se presenta en ambos años un mínimo en los menores de un año del genero femenino, y el máximo en el grupo de 29 a 44 años del genero femenino lo anterior

derivado de la distribución poblacional, al analizar los expuestos de esta zona se identifica que el grupo etario del genero masculino de 75 a 79 años presenta el menor numero de expuestos pero no el menor numero de reclamaciones es decir que este grupo etario en los dos años de estudio presenta mayores reclamaciones, así mismo se observa que el mayor numero de expuestos se presenta en el grupo de 29a 44 años del genero masculino.

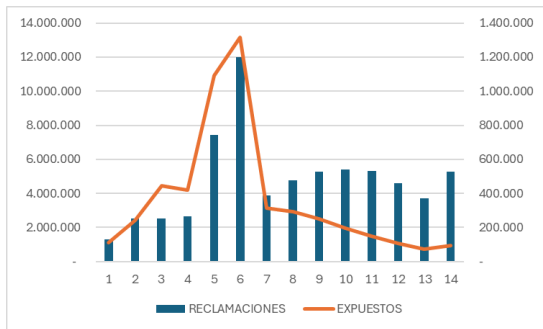
A continuación se presenta las gráficas de reclamaciones y expuestos de cada uno de los años, así mismo se presenta la pirámide poblacional de la Zona Rural, comparada con la pirámide poblacional del país de las EPS de la selección, de cada uno de los años.



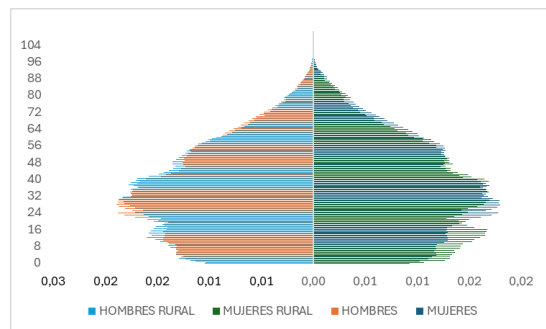
(a) Histograma Reclamaciones y Expuestos Zona Rural Año 2018



(b) Pirámide Poblacional Zona Rural Año 2018



(a) Histograma Reclamaciones y Expuestos Zona Rural Año 2019



(b) Pirámide Poblacional Zona Rural Año 2019

■ Valor Reclamaciones

El Valor de las reclamaciones corresponde a al valor reportado y reconocido por las aseguradoras para la prestación de las tecnologías o servicios de salud demandados por los afiliados en la Zona Rural, en concordancia con las reclamaciones y los expuestos en los años se presenta el mayor valor reportado en el grupo de 29 a 44 años del genero femenino en los dos años de estudio, sin embargo el valor mínimo para el año 2018 se presenta en el grupo etario de 1 a 5 años y en año 2019 el mínimo se presenta en el grupo etario de menores de un año, de lo anterior se evidencia la dinámica en la población en particular de los niños en los municipios que conforman esta zona.

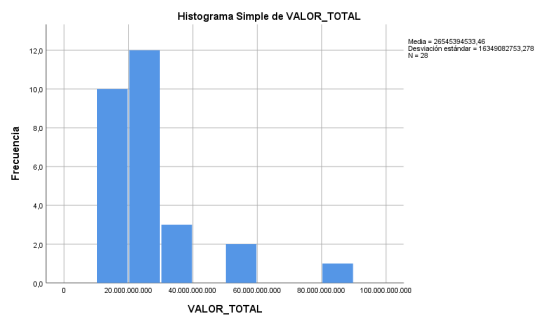
A continuación se presenta el histograma del valor de la reclamaciones y el diagrama de cajas y bigotes para la ZOna Rural.

■ Póliza

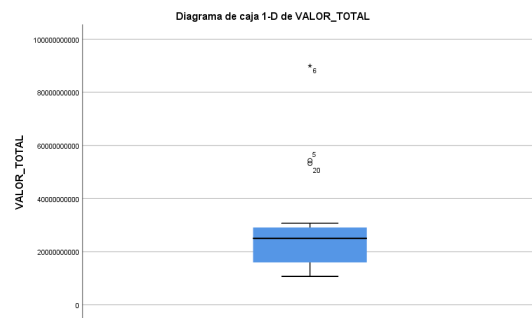
El Valor de la póliza corresponde al reconocimiento realizado por el gobierno, por concepto de la prima es decir de la Unidad de Pago por Capitación UPC y el valor que se reconoce para la demanda inducida que hace referencia a las actividades de promoción de la salud y prevención de la enfermedad encaminadas a la educación.

El valor máximo para los dos años se encuentra en grupo de 29 a 44 años del genero femenino con valores de prima promedio de de \$636.356y \$655.755,54 respectivamente.

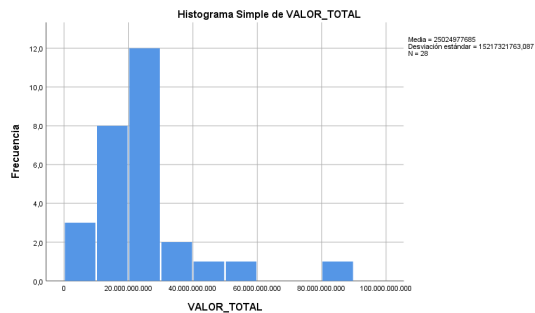
Así mismo el valor mínimo se presenta en el grupo etario de 6 a 11 años del genero femenino con valor de prima promedio de \$195.777 y de \$208.472,53 respectivamente.



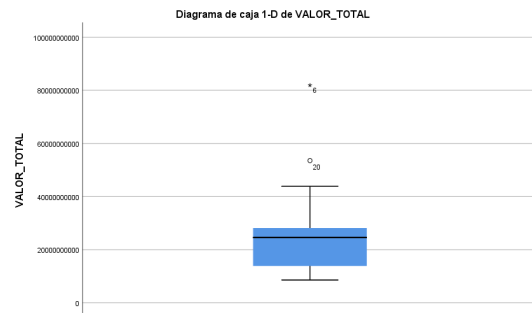
(a) Histograma Valor Reclamaciones Zona Rural Año 2018



(b) Cajas y Bigotes Valor Reclamaciones Zona Rural Año 2018

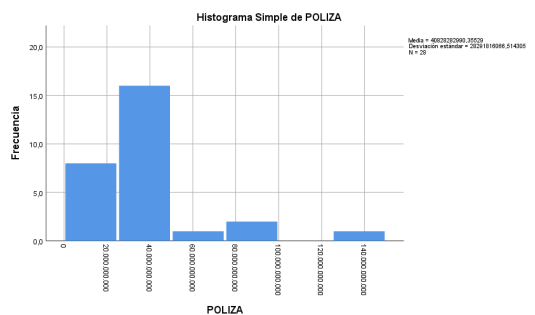


(a) Histograma Valor Reclamaciones Zona Rural Año 2019

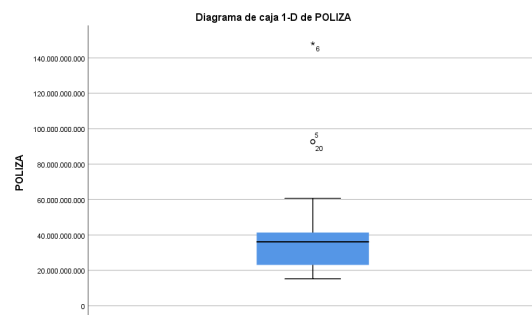


(b) Cajas y Bigotes Valor Reclamaciones Zona Rural Año 2019

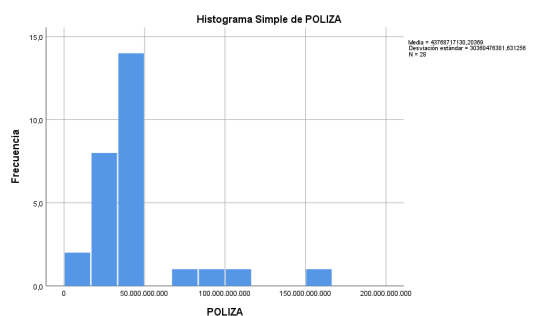
En las gráficas que a continuación se presentan, el histograma de la distribución de variable así como el gráfico de cajas y bigotes para cada uno de los años en estudio.



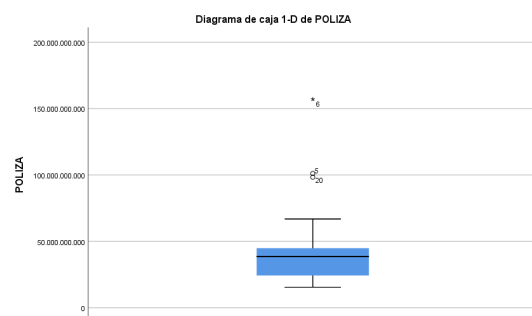
(a) Histograma de Póliza Zona Rural Año 2018



(b) Cajas y Bigotes Póliza Zona Rural Año 2018



(a) Histograma de Póliza Zona Rural Año 2019



(b) Cajas y Bigotes Póliza Zona Rural Año 2019

### 3.2.6. Zona Rural Disperso

A continuación se presentan las estadísticas descriptivas para los años 2018 y 2019 para la zona Rural Disperso, a esta zona pertenecen 241 municipios del país.

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	15.190	666.733.040	1.459.217.312	703	30.820
1st Qu.	21.562	1.140.019.078	2.194.119.029	1.836	45.447
Median	25.300	1.751.059.006	2.863.168.939	4.172	55.220
Mean	35.313	1.908.126.512	4.312.310.326	8.730	58.373
3rd Qu.	35.022	1.915.575.975	4.226.610.663	7.464	69.731
Max.	153.686	7.154.791.600	19.439.979.582	47.047	94.898

Figura 45: Estadísticos Descriptivos Zona Rural Disperso Año 2018

ESTADISTICOS	RECLAMACIONES	VALOR_TOTAL	POLIZA	EXPUESTOS	COSTO_MEDIO
Min.	20.053	728.804.464	1.534.208.820	677	23.090
1st Qu.	29.122	1.134.938.696	2.503.150.468	2.049	39.126
Median	32.765	1.710.947.083	3.185.050.569	4.846	45.593
Mean	43.264	1.874.312.294	4.728.627.913	9.398	45.036
3rd Qu.	44.281	2.087.157.246	4.636.646.509	7.884	54.781
Max.	176.510	6.990.022.011	20.754.784.725	50.811	60.818

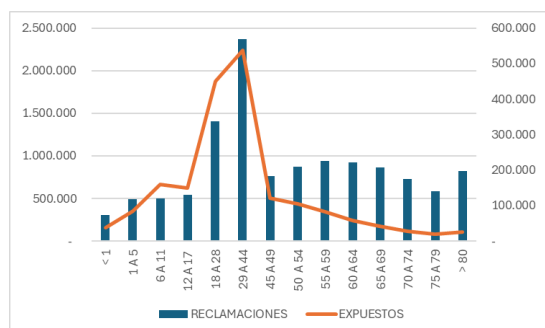
Figura 46: Estadísticos Descriptivos Zona Rural Disperso Año 2019

A continuación se presenta el análisis visual de cada una de las variables para los años 2018 y 2019

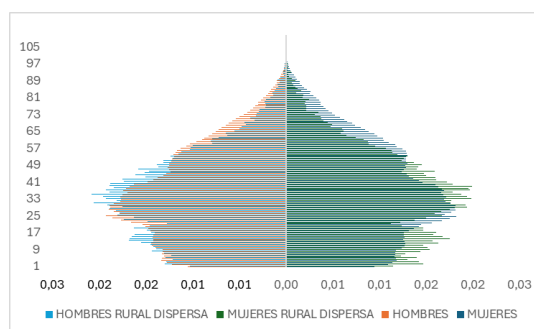
#### ■ Reclamaciones y Expuestos

Al analizar el número de reclamaciones es decir el número de atenciones de la población de la Zona Rural Disperso teniendo en cuenta los grupos etarios se presenta en ambos años el máximo en el grupo de 29 a 44 años del género femenino, lo anterior derivado de la distribución poblacional, el mínimo de reclamaciones para el año 2018 se encuentra en el grupo de 75 a 79 años del género masculino, para el año 2019 el mínimo de reclamaciones se encuentra en los menores de un año, lo anterior es evidencia de la dinámica de población de los municipios de esta zona. Por su parte el mayor número de expuestos en los dos años se ubica el grupo de etario de 29 a 44 años del género masculino y el menor número en los dos años se observa en el grupo etario de 75 a 79 años del género femenino.

A continuación se presenta las gráficas de reclamaciones y expuestos de cada uno de los años, así mismo se presenta la pirámide poblacional de la Zona Rural Disperso, comparada con la pirámide poblacional del país de las EPS de la selección, de cada uno de los años.

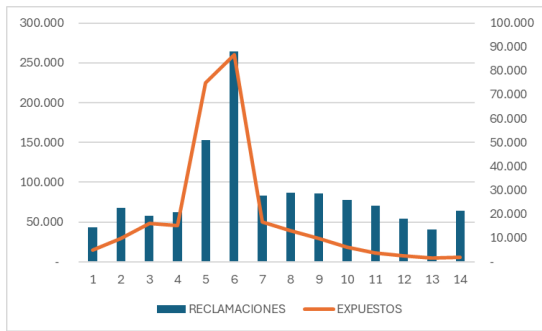


(a) Histograma Reclamaciones y Expuestos Zona Rural Disperso Año 2018

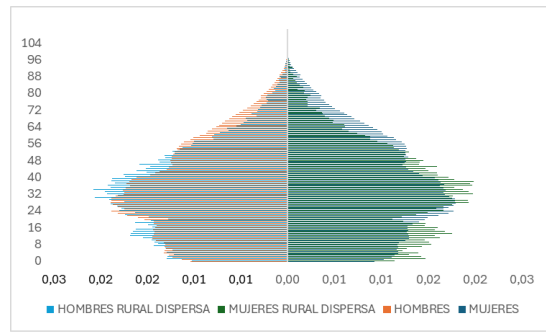


(b) Pirámide Poblacional Zona Rural Disperso Año 2018

#### ■ Valor Reclamaciones



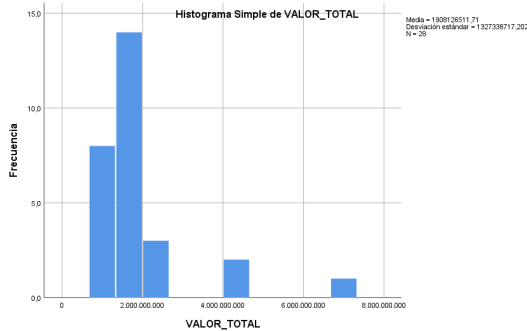
(a) Histograma Reclamaciones y Expuestos Zona Rural Disperso Año 2019



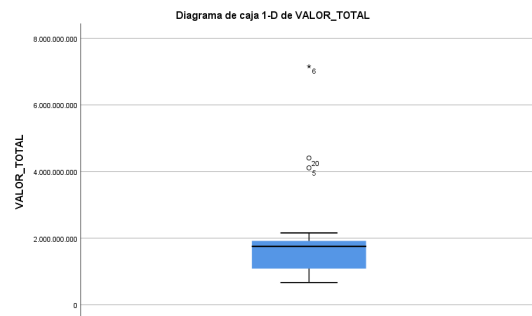
(b) Pirámide Poblacional Zona Rural Disperso Año 2019

El Valor de las reclamaciones corresponde a al valor reportado y reconocido por las aseguradoras para la prestación de las tecnologías o servicios de salud demandados por los afiliados en la Zona Rural Disperso, en concordancia con las reclamaciones y los expuestos en los años se presenta el mayor valor reportado en el grupo de 29 a 44 años del genero femenino y el valor mínimo reportado se observa el grupo etario de 6 a 11 años del genero femenino para los años 2018 y 2019.

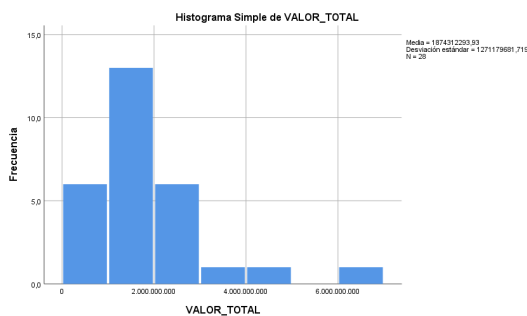
A continuación se presenta el histograma del valor de las reclamaciones y el diagrama de cajas y bigotes de esta variable.



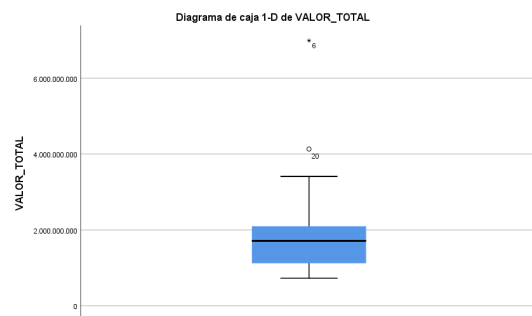
(a) Histograma Valor Reclamaciones Zona Rural Disperso Año 2018



(b) Cajas y Bigotes Valor Reclamaciones Zona Rural Disperso Año 2018



(a) Histograma Valor Reclamaciones Zona Rural Disperso Año 2019



(b) Cajas y Bigotes Valor Reclamaciones Zona Rural Disperso Año 2019

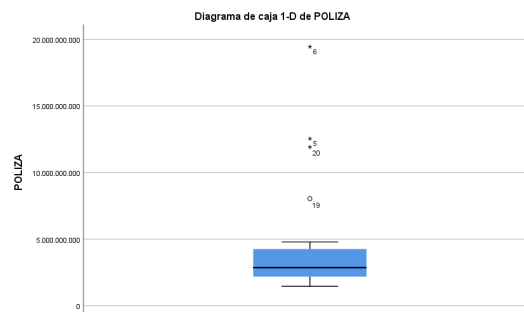
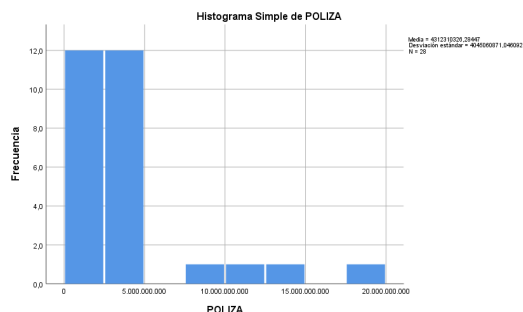
■ **Póliza**

El Valor de la póliza corresponde al reconocimiento realizado por el gobierno, por concepto de la prima es decir de la Unidad de Pago por Capitación UPC y el valor que se reconoce para la demanda inducida que hace referencia a las actividades de promoción de la salud y prevención de la enfermedad encaminadas a la educación.

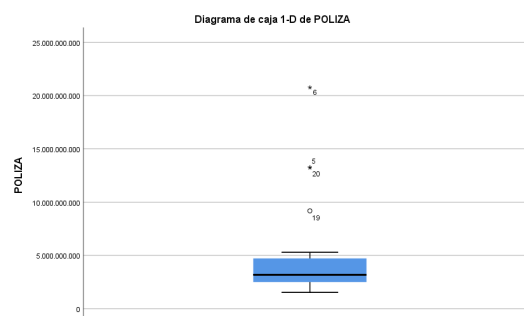
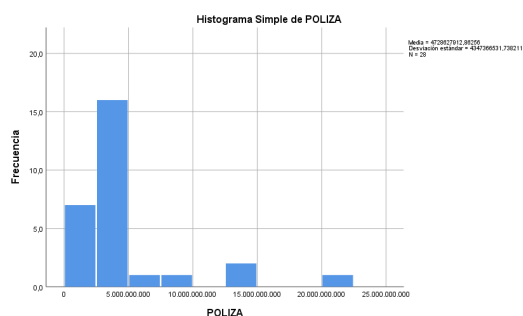
El valor máximo para los dos años se encuentra en grupo de 29 a 44 años del genero femenino con valores de prima promedio de \$569.402,76 respectivamente.

Así mismo el valor mínimo se presenta en el grupo etario de 6 a 11 año con valor de prima promedio de \$195.579,32, y de \$194.968,72 respectivamente.

En las gráficas que a continuación se presentan, el histograma de la distribución de variable así como el gráfico de cajas y bigotes para los años 2018 y 2019.



(a) Histograma de Póliza Zona Rural Disperso Año 2018 (b) Cajas y Bigotes Póliza Zona Rural Disperso Año 2018



(a) Histograma de Póliza Zona Rural Disperso Año 2019 (b) Cajas y Bigotes Póliza Zona Rural Disperso Año 2019

### 3.3. Modelos Aplicados

Los modelos aplicados para la estimación de la prima pura se divide en dos modelos de frecuencias y modelos de severidad a continuación se presenta el desarrollo de cada uno de ellos.

#### 3.3.1. Modelos de Frecuencia

- Zona Alejada

La aplicación de los modelos que den cuenta de la proyección de las reclamaciones (frecuencia uso), parte del análisis de la denominada frecuencia empírica este análisis se realiza con el fin de observar su comportamiento el cual se presenta a continuación, las dos distribuciones más comunes para la frecuencia desde los seguros son la Poisson y la Binominal negativa por lo que se decide realizar el análisis de las distribuciones empíricas con estas dos clases de distribuciones, lo cual se puede observar en los siguientes gráficos para el año 2018 y 2019 respectivamente.

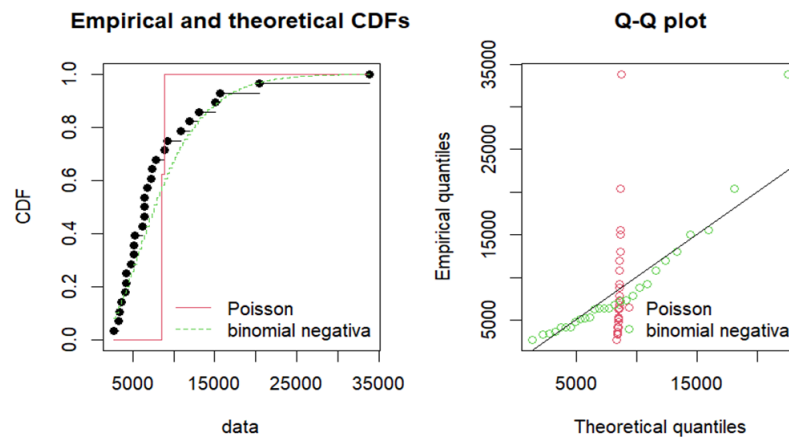


Figura 53: Frecuencia empírica Zona Alejada Año 2018

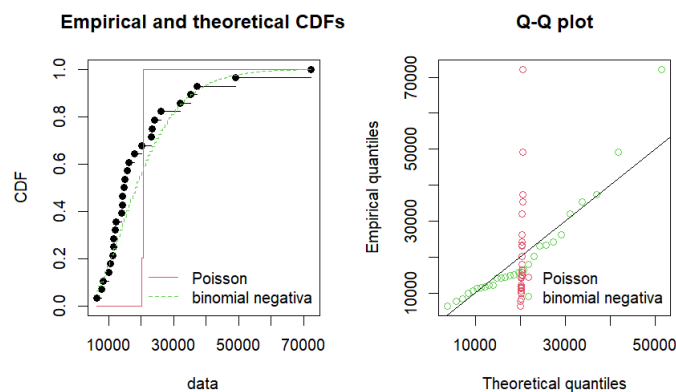


Figura 54: Frecuencia empírica Zona Alejada Año 2019

Gráficamente la distribución que mejor ajusta en los dos años es la binominal negativa, en aras de comprobar lo anterior se realiza el modelo teniendo en cuenta la distribución de poisson.

#### ■ Zona Bogotá

La aplicación de los modelos que den cuenta de la proyección de las reclamaciones (frecuencia uso), parte del análisis de la denominada frecuencia empírica este análisis se realiza con le fin de observar su comportamiento el cual se presenta a continuación, las dos distribuciones mas comunes para la frecuencia desde los seguros son la Poisson y la Binominal negativa por lo que se decide realizar el análisis de las distribuciones empíricas con estas dos clases de distribuciones para el año 2018 y 2019, lo cual se puede observar en los siguientes gráficos.

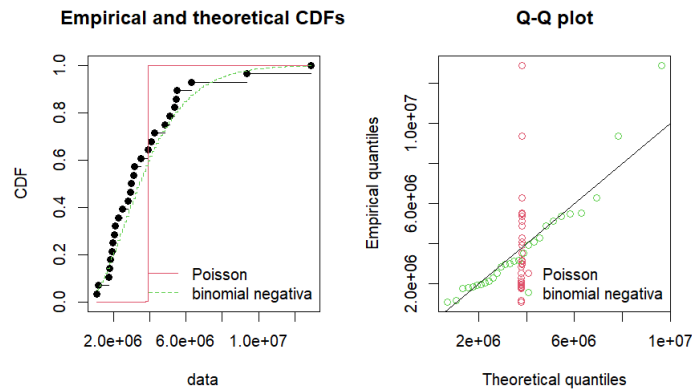


Figura 55: Frecuencia empírica Zona Bogotá Año 2018

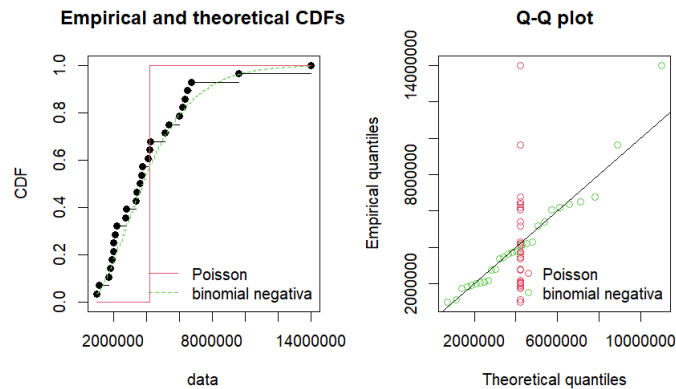


Figura 56: Frecuencia empírica Zona Bogotá Año 2019

Gráficamente la distribución que mejor ajusta para los dos años es la binominal negativa, en aras de comprobar lo anterior se realiza el modelo teniendo en cuenta la distribución de poisson.

- Zona Normal Mayor

La aplicación de los modelos que den cuenta de la proyección de las reclamaciones (frecuencia uso), parte del análisis de la denominada frecuencia empírica este análisis se realiza con le fin de observar su comportamiento el cual se presenta a continuación, las dos distribuciones mas comunes para la frecuencia desde los seguros son la Poisson y la Binominal negativa por lo que se decide realizar el análisis de las distribuciones empíricas con estas dos clases de distribuciones para 2018 y 2019, lo cual se puede observar en los siguientes gráficos.

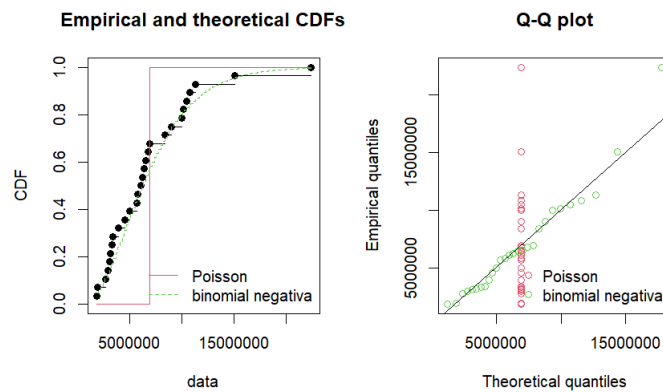


Figura 57: Frecuencia empírica Zona Normal Mayor Año 2018

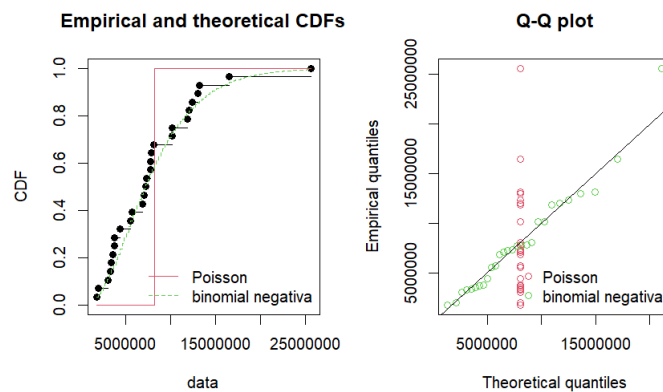


Figura 58: Frecuencia empírica Zona Normal Mayor Año 2019

Gráficamente la distribución que mejor ajusta es la binominal negativa para los dos años, en aras de comprobar lo anterior se realiza el modelo teniendo en cuenta la distribución de poisson.

#### ■ Zona Normal Medio

La aplicación de los modelos que den cuenta de la proyección de las reclamaciones (frecuencia uso), parte del análisis de la denominada frecuencia empírica este análisis se realiza con le fin de observar su comportamiento el cual se presenta a continuación, las dos distribuciones mas comunes para la frecuencia desde los seguros son la Poisson y la Binominal negativa por lo que se decide realizar el análisis de las distribuciones empíricas con estas dos clases de distribuciones para el 2018 y 2019, lo cual se puede observar en los siguientes gráficos.

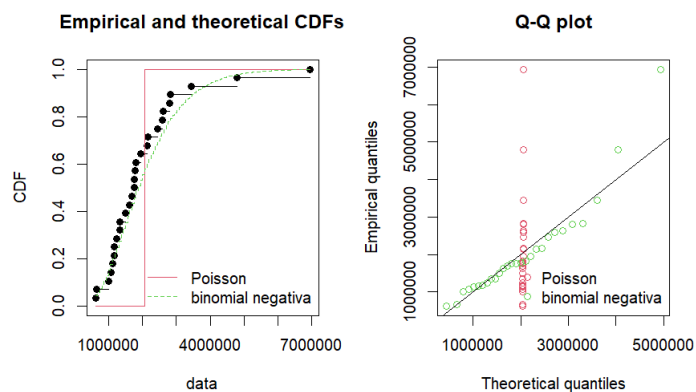


Figura 59: Frecuencia empírica Zona Normal Medio Año 2018

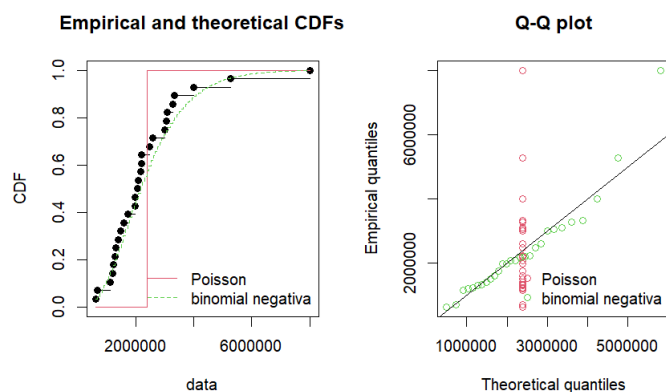


Figura 60: Frecuencia empírica Zona Normal Medio Año 2019

Gráficamente la distribución que mejor ajusta es la binominal negativa aunque presenta algunos puntos distantes del ajuste binomial para los dos años, en aras de comprobar lo anterior se realiza el modelo teniendo en cuenta la distribución de poisson.

#### ■ Zona Rural

La aplicación de los modelos que den cuenta de la proyección de las reclamaciones (frecuencia uso), parte del análisis de la denominada frecuencia empírica este análisis se realiza con el fin de observar su comportamiento el cual se presenta a continuación, las dos distribuciones más comunes para la frecuencia desde los seguros son la Poisson y la Binominal negativa por lo que se decide realizar el análisis de las distribuciones empíricas con estas dos clases de distribuciones para los años 2018 y 2019, lo cual se puede observar en los siguientes gráficos.

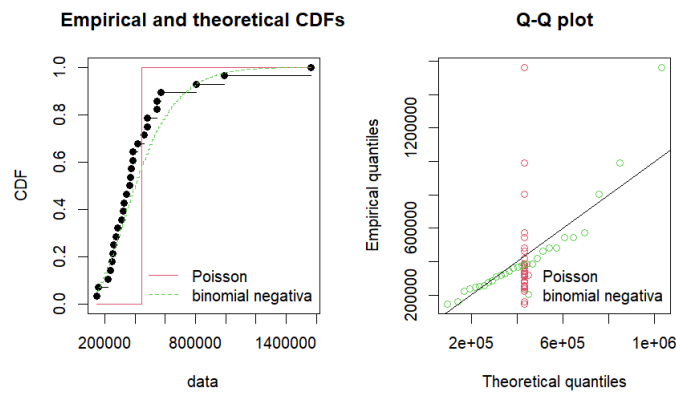


Figura 61: Frecuencia empírica Zona Rural Año 2018

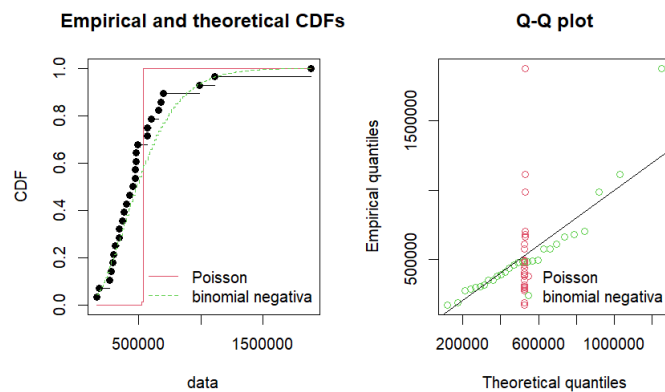


Figura 62: Frecuencia empírica Zona Rural Año 2019

Gráficamente la distribución que mejor ajusta es la binomial negativa aunque presenta algunos puntos distantes del ajuste binomial negativa para los dos años, en aras de comprobar lo anterior se realiza el modelo teniendo en cuenta la distribución de poisson.

#### ■ Zona Rural Disperso

La aplicación de los modelos que den cuenta de la proyección de las reclamaciones (frecuencia uso), parte del análisis de la denominada frecuencia empírica este análisis se realiza con le fin de observar su comportamiento el cual se presenta a continuación, las dos distribuciones mas comunes para la frecuencia desde los seguros son la Poisson y la Binominal negativa por lo que se decide realizar el análisis de las distribuciones empíricas con estas dos clases de distribuciones para los años 2018 y 2019, lo cual se puede observar en los siguientes gráficos.

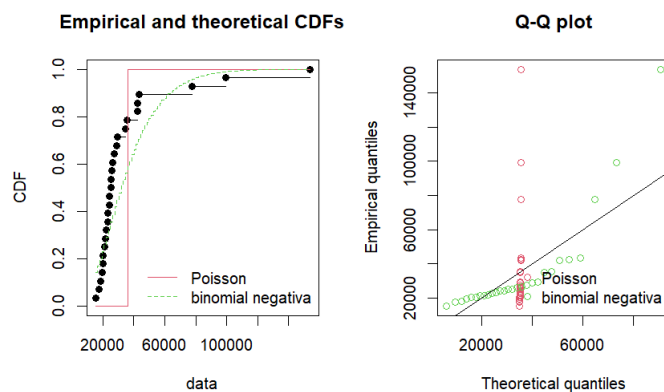


Figura 63: Frecuencia empírica Zona Rural Disperso Año 2018

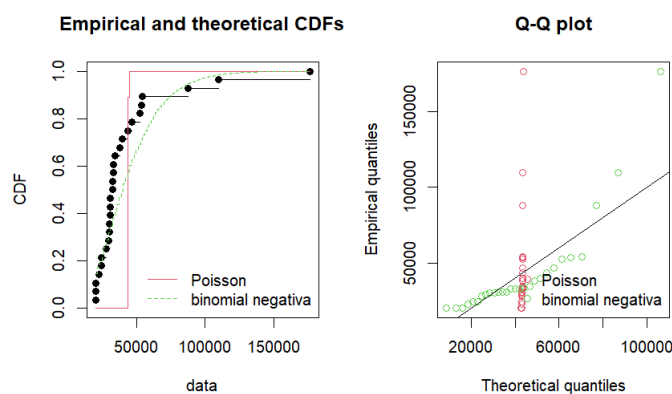


Figura 64: Frecuencia empírica Zona Rural Disperso Año 2019

Gráficamente la distribución que mejor ajusta es la binomial negativa aunque presenta algunos puntos distantes del ajuste binomial negativa para los dos años, en aras de comprobar lo anterior se realiza el modelo teniendo en cuenta la distribución de poisson.

### 3.3.2. Modelos de Severidad

- Zona Alejada Entendida como el costo promedio de las atenciones de acuerdo a lo establecido en la revisión de la literatura dentro de las distribuciones mas utilizadas se encuentra la Gamma y la Lognormal

A continuación se presentan las distribución empírica la cual brinda información gráfica sobre el posible modelo a ajustar para los años 2018 y 2019.

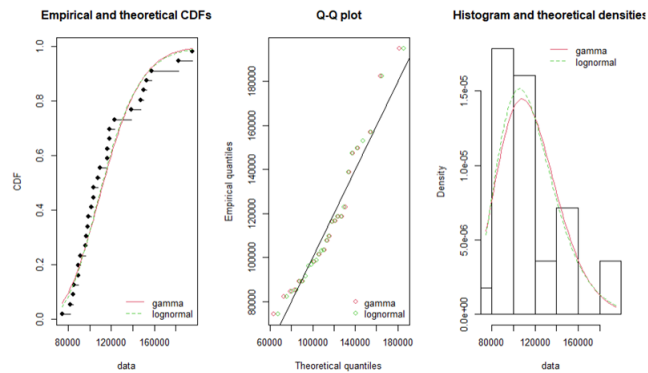


Figura 65: Severidad empírica Zoan Alejada Año 2018

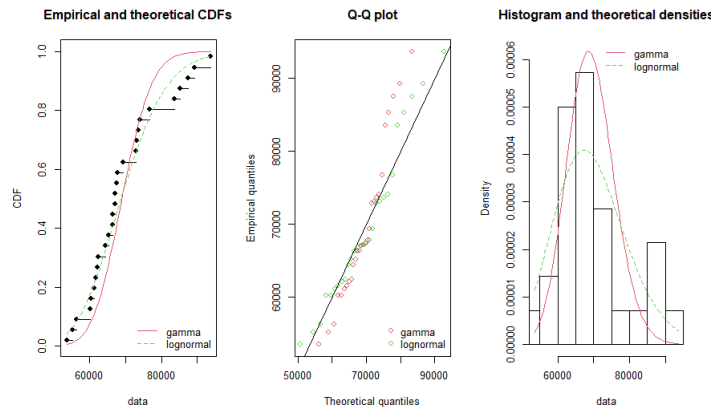


Figura 66: Severidad empírica Zona Alejada Año 2019

Las distribuciones gamma y lognormal se ajustan a los datos del año 2018 por lo tanto se realiza el modelo utilizando cada una de estas distribuciones y teniendo como variables explicativas el grupo etario y el sexo.

Ahora bien para el año 2019 el modelo que se ajusta es la gamma por lo cual el modelo se realiza teniendo en cuenta como variables explicativas el grupo etario y el sexo.

- Zona Bogotá Entendida como el costo promedio de las atenciones de acuerdo a lo establecido en la revisión de la literatura dentro de las distribuciones mas utilizadas se encuentra la Gamma y la Lognormal

A continuación se presentan las distribución empírica la cual brinda información gráfica sobre el posible modelo a ajustar para los años 2018 y 2019.

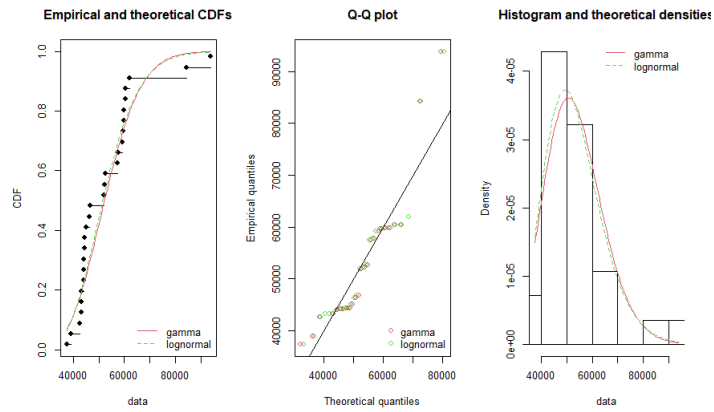


Figura 67: Severidad empírica Zona Bogotá Año 2018

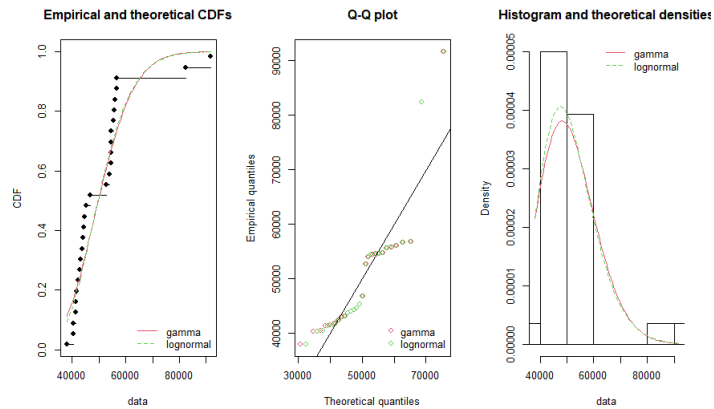


Figura 68: Severidad empírica Zona Bogotá Año 2019

Las distribuciones gamma y lognormal se ajustan a los datos por lo tanto se realiza el modelo utilizando cada una de estas distribuciones y teniendo como variables explicativas el grupo etario y el sexo para los dos años.

- Zona Normal Mayor Entendida como el costo promedio de las atenciones de acuerdo a lo establecido en la revisión de la literatura dentro de las distribuciones mas utilizadas se encuentra la Gamma y la Lognormal

A continuación se presentan las distribución empírica la cual brinda información gráfica sobre el posible modelo a ajustar para los años 2018 y 2019.

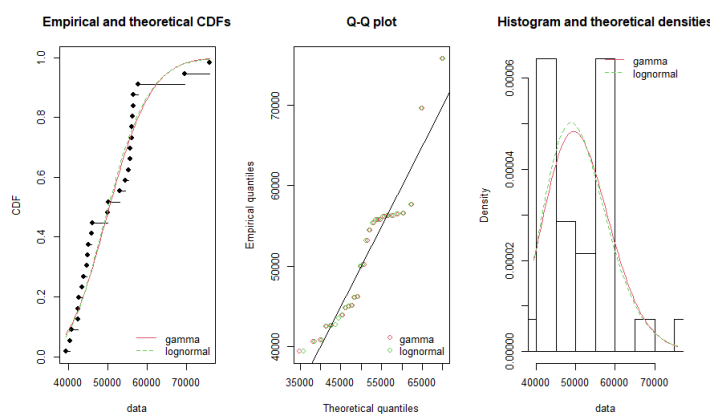


Figura 69: Severidad empírica Zona Normal Mayor Año 2018

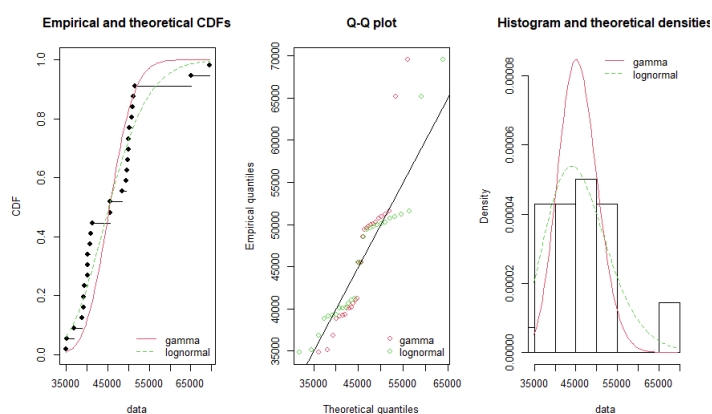


Figura 70: Severidad empírica Zona Normal Mayor Año 2019

Las distribuciones gamma y lognormal se ajustan a los datos para el año 2018 por lo tanto se realiza el modelo utilizando cada una de estas distribuciones y teniendo como variables explicativas el grupo etario y el sexo.

En cuanto al año 2019 la distribución que más se ajusta es la lognormal por lo cual se realiza el modelo teniendo en cuenta esta distribución teniendo como variables explicativas el grupo etario y el sexo.

- **Zona Normal Medio** Entendida como el costo promedio de las atenciones de acuerdo a lo establecido en la revisión de la literatura dentro de las distribuciones más utilizadas se encuentra la Gamma y la Lognormal

A continuación se presentan las distribución empírica la cual brinda información gráfica sobre el posible modelo a ajustar para los años 2018 y 2019.

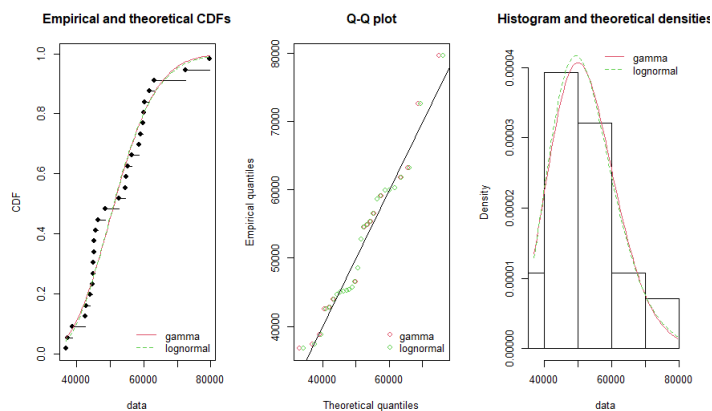


Figura 71: Severidad empírica Zona Normal Medio Año 2018

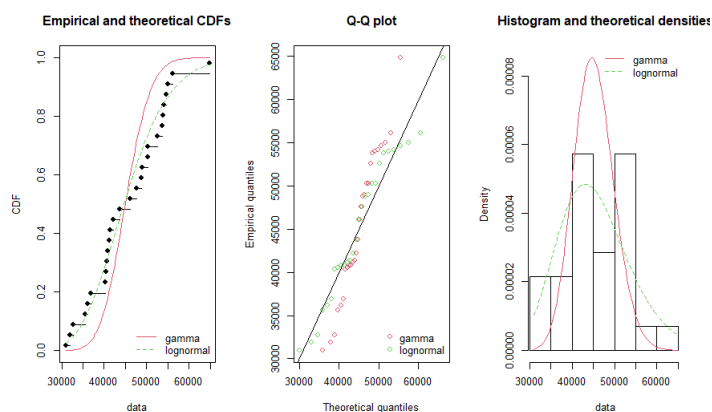


Figura 72: Severidad empírica Zona Normal Medio Año 2019

Las distribuciones gamma y lognormal se ajustan a los datos para el año 2018 por lo tanto se realiza el modelo utilizando cada una de estas distribuciones y teniendo como variables explicativas el grupo etario y el sexo.

Para el año 2019, la distribución que mejor se ajusta es la lognormal por lo cual el modelo se realiza con a partir de esta distribución y teniendo como variables explicativas el grupo etario y el sexo.

- Zona Rural Entendida como el costo promedio de las atenciones de acuerdo a lo establecido en la revisión de la literatura dentro de las distribuciones mas utilizadas se encuentra la Gamma y la Lognormal

A continuación se presentan las distribución empírica la cual brinda información gráfica sobre el posible modelo a ajustar para los años 2018 y 2019.

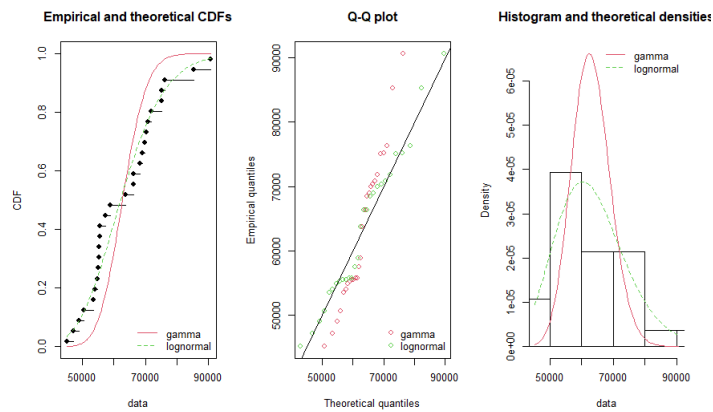


Figura 73: Severidad empírica Zona Rural Año 2018

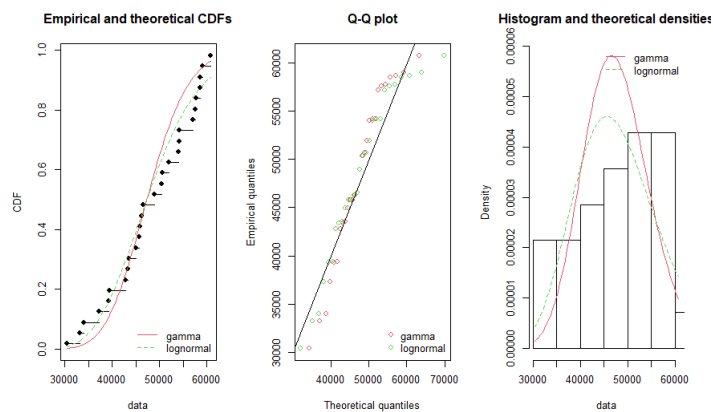


Figura 74: Severidad empírica Zona Rural Año 2019

La distribución Gamma no ajusta a los datos del año 2018, a diferencia de la distribución lognormal por lo que se decide realizar el modelo con esta distribución teniendo como variables explicativas el grupo etario y el sexo.

Para el año 2019, la distribución que mejor se ajusta es la lognormal por lo cual el modelo se realiza con a partir de esta distribución y teniendo como variables explicativas el grupo etario y el sexo.

- **Zona Rural Disperso** Entendida como el costo promedio de las atenciones de acuerdo a lo establecido en la revisión de la literatura dentro de las distribuciones más utilizadas se encuentra la Gamma y la Lognormal

A continuación se presentan las distribución empírica la cual brinda información gráfica sobre el posible modelo a ajustar para los años 2018 y 2019.

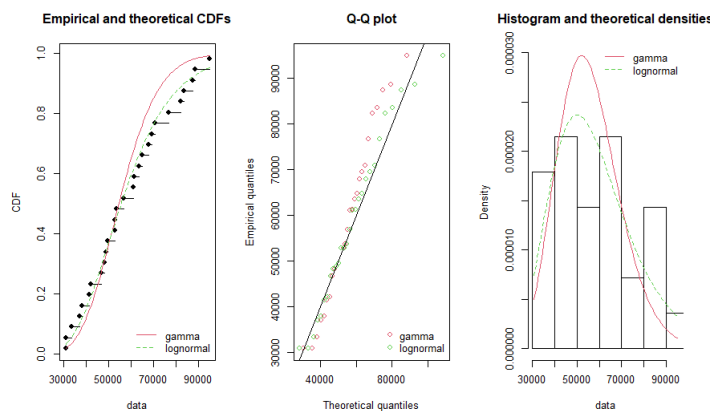


Figura 75: Severidad empírica Zona Rural Disperso Año 2018

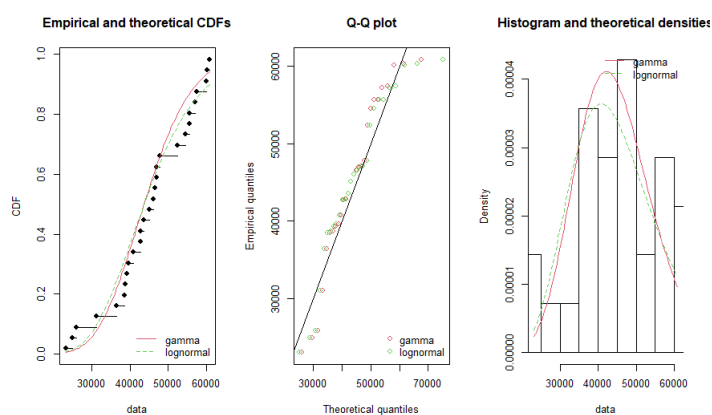


Figura 76: Severidad empírica Zona Rural Disperso Año 2019

La distribución Gamma no ajusta a los datos para el año 2018, a diferencia de la distribución lognormal por lo que se decide realizar el modelo con esta distribución teniendo como variables explicativas el grupo etario y el sexo.

Para el año 2019 las distribuciones gamma y lognormal se ajustan a los datos por lo tanto se realiza el modelo utilizando cada una de estas distribuciones y teniendo como variables explicativas el grupo etario y el sexo.

## 4. Aplicación y Resultados de los Modelos

### 4.1. Modelos de Frecuencias

#### 4.1.1. Zona Alejada

Como se evidencio en el apartado anterior la distribución binomial negativa presenta un mejor ajuste que la distribución poisson, la aplicación de este modelo se puede observar en el Anexo 2, ahora bien el modelo poisson presenta sobre dispersión lo cual ratifica le necesidad de aplicar el modelo binomial negativa.

##### ■ GLM Binomial Negativa

Dado al resultado de la prueba de sobredispersion, se considera que el modelo poisson no presenta un buen ajuste para ninguno de los dos años por lo que se decide realizar el modelo utilizando la distribución binominal negativa cuyo resultado se presenta continuación.

```

Overdispersion test
data: fqpoi_A
z = 5.0856, p-value = 0.0000001833
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
257.293

```

Figura 77: Overdispersion Zona Alejada Año 2018

```

Overdispersion test
data: fqpoi_A_2019
z = 4.2281, p-value = 0.00001178
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
827.2121

```

Figura 78: Overdispersion Zona Alejada Año 2019

```

call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
SEXO + offset(log(EXPUESTOS)), data = BASE_2018_A, link = log,
init.theta = 31.95517805)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.76612    0.13010  13.575 < 2e-16 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.57025    0.17750   3.213 0.001315 **
GRUPO_ETAREO_RUTA_AGREGADO11  0.89628    0.17753   5.049 4.45e-07 ***
GRUPO_ETAREO_RUTA_AGREGADO12  0.87367    0.17762   4.919 8.71e-07 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.01899    0.17773   5.733 9.85e-09 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.05202    0.17770   5.920 3.21e-09 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.05409    0.17754  -0.305 0.760642
GRUPO_ETAREO_RUTA_AGREGADO3  -0.75254    0.17758  -4.238 2.26e-05 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.85035    0.17766  -4.786 1.70e-06 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.66522    0.17746  -3.749 0.000178 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.32796    0.17741  -1.849 0.064514 .
GRUPO_ETAREO_RUTA_AGREGADO7  -0.08297    0.17754  -0.467 0.640238
GRUPO_ETAREO_RUTA_AGREGADO8   0.10804    0.17749   0.609 0.542725
GRUPO_ETAREO_RUTA_AGREGADO9   0.34313    0.17747   1.933 0.053178 .
SEXOM                -0.29532    0.06703  -4.406 1.06e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(31.9552) family taken to be 1

Null deviance: 383.808  on 27  degrees of freedom
Residual deviance: 28.131  on 13  degrees of freedom
AIC: 511.1

Number of Fisher Scoring iterations: 1

Theta: 31.96
Std. Err.: 8.54

2 x log-likelihood: -479.102

```

Figura 79: Modelo Binomial Negativa Zona Alejada Año 2018

Así mismo se realiza la prueba de Vuong que permite verificar cual es el mejor modelo

Estadístico	Vuong z-statistic	HA	p-value
Raw	-4.82054	model2 > model1	0.00000071585
AIC-corrected	-4.82054	model2 > model1	0.00000071585
BIC-corrected	-4.82054	model2 > model1	0.00000071585

Tabla 5: Prueba de Vuong Zona Alejada Año 2018

Lo anterior evidencia que el modelo 2 que corresponde a la binomial negativa es mejor que el modelo poisson, por lo cual se presenta a continuación los residuales del modelo binomial negativa para el año 2018.

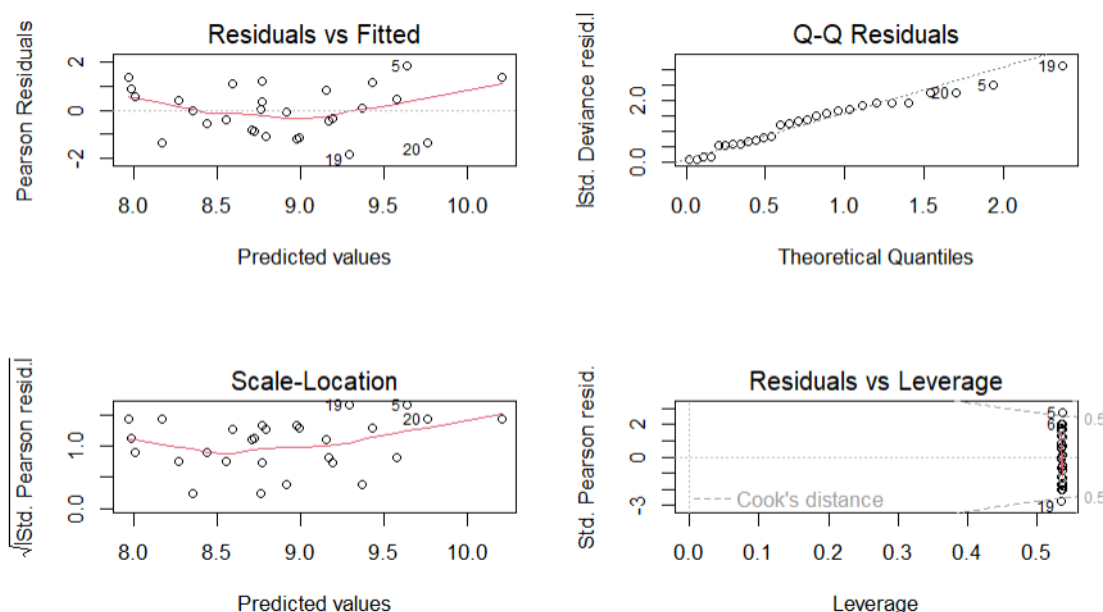


Figura 80: Residuales Modelo Binomial Negativa Zona Alejada Año 2018

Como se observa en la figura número 80, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 6 y figura 81, presentando puntos influyentes en las observaciones 5, 13, 14 y 19, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
5	1.9430638	0.5360080	0.57922638
13	-0.5436499	0.5364517	0.05048325
14	-0.8931103	0.5369205	0.12808369
19	-2.3622304	0.5356088	0.57829754

Tabla 6: Puntos Influyentes Zona Alejada Año 2018

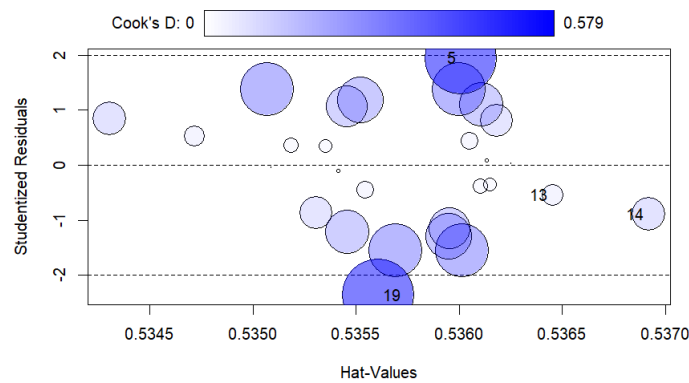


Figura 81: Puntos Influyentes Zona Alejada Año 2018

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la frecuencia esperada para el año 2018, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	4,940015785	5,06675723
Entre 1 y 5 Años	Ambos	4,694620782	4,800857429
Entre 6 y 11 Años	Ambos	2,340586944	2,386468467
Entre 12 y 17 Años	Femenino	2,327860937	2,498696696
Entre 12 y 17 Años	Masculino	1,9871673	1,859763701
Entre 18 y 28 Años	Femenino	4,001565864	3,006870631
Entre 18 y 28 Años	Masculino	1,497012157	2,237994256
Entre 29 y 44 Años	Femenino	5,255322455	4,212897632
Entre 28 y 44 Años	Masculino	2,359247876	3,13563231
Entre 45 y 49 Años	Ambos	4,900735294	4,745718349
Entre 50 y 54 Años	Ambos	5,991348032	5,741577578
Entre 55 y 59 Años	Ambos	7,362650602	7,244732459
Entre 60 y 64 Años	Ambos	9,161213845	9,125836342
Entre 65 y 69 Años	Ambos	12,44883721	12,61763419
Entre 70 y 74 Años	Ambos	12,33488372	12,31886615
Entre 75 y 79 Años	Ambos	13,97748593	14,27383736
Mayores de 80 Años	Ambos	14,24834437	15,07257679

Tabla 7: Frecuencias Observadas y Estimadas Zona Alejada Año 2018

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada, para el año 2018

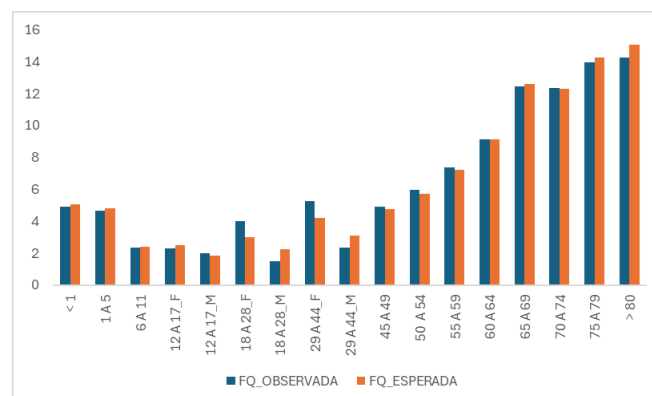


Figura 82: Frecuencia Observada Vs Frecuencia Estimada Zona Alejada Año 2018

El modelo para el año 2018 presenta un error porcentual absoluto medio (MAPE) del 9.78 % lo cual se presenta un ajuste aceptable, así mismo al calcular el pseudo R2 este es de 0.92.

A continuación se presentan los residuales del modelo binomial negativa para el año 2019.

Call:

```
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
SEXO + offset(log(EXPUESTOS)), data = BASE_2019_A, link = log,
init.theta = 23.77799672)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	2.51508	0.15033	16.731	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO10	0.71389	0.20529	3.478	0.000506	***
GRUPO_ETAREO_RUTA_AGREGADO11	0.94861	0.20530	4.621	0.000003826690303	***
GRUPO_ETAREO_RUTA_AGREGADO12	1.19813	0.20531	5.836	0.000000005359962	***
GRUPO_ETAREO_RUTA_AGREGADO13	1.40434	0.20534	6.839	0.0000000000007970	***
GRUPO_ETAREO_RUTA_AGREGADO14	1.50173	0.20532	7.314	0.000000000000259	***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.21789	0.20533	-1.061	0.288626	
GRUPO_ETAREO_RUTA_AGREGADO3	-0.92353	0.20535	-4.497	0.000006881709634	***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.91818	0.20537	-4.471	0.000007792393971	***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.57277	0.20528	-2.790	0.005268	**
GRUPO_ETAREO_RUTA_AGREGADO6	-0.41127	0.20526	-2.004	0.045111	*
GRUPO_ETAREO_RUTA_AGREGADO7	-0.03457	0.20531	-0.168	0.866272	
GRUPO_ETAREO_RUTA_AGREGADO8	0.23653	0.20529	1.152	0.249253	
GRUPO_ETAREO_RUTA_AGREGADO9	0.45456	0.20528	2.214	0.026803	*
SEXOM	-0.26466	0.07757	-3.412	0.000645	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(23.778) family taken to be 1)

Null deviance: 429.591 on 27 degrees of freedom  
Residual deviance: 28.191 on 13 degrees of freedom  
AIC: 568.78

Number of Fisher Scoring iterations: 1

Theta: 23.78  
Std. Err.: 6.32

2 x log-likelihood: -536.782



Figura 83: Modelo Binomial Negativa Zona Alejada Año 2019

La mayoría de variables son significativas al 95 %, dado que los dos modelos presentan significancia en la variables se realiza la prueba de vuong, cuyos resultados se presentan continuación:

Estadístico	Vuong z-statistic	HA	p-value
Raw	-4.955335	model2 > model1	0.00000036103
AIC-corrected	-4.955335	model2 > model1	0.00000036103
BIC-corrected	-4.955335	model2 > model1	0.00000036103

Tabla 8: Prueba de Vuong Zona Alejada Año 2019

Como se puede evidenciar al Modelo 2 es mejor que el modelo 1 en este caso es el modelo que corresponde a la distribución binomial negativa, por lo anterior se presentan los residuales correspondientes al año

2019.

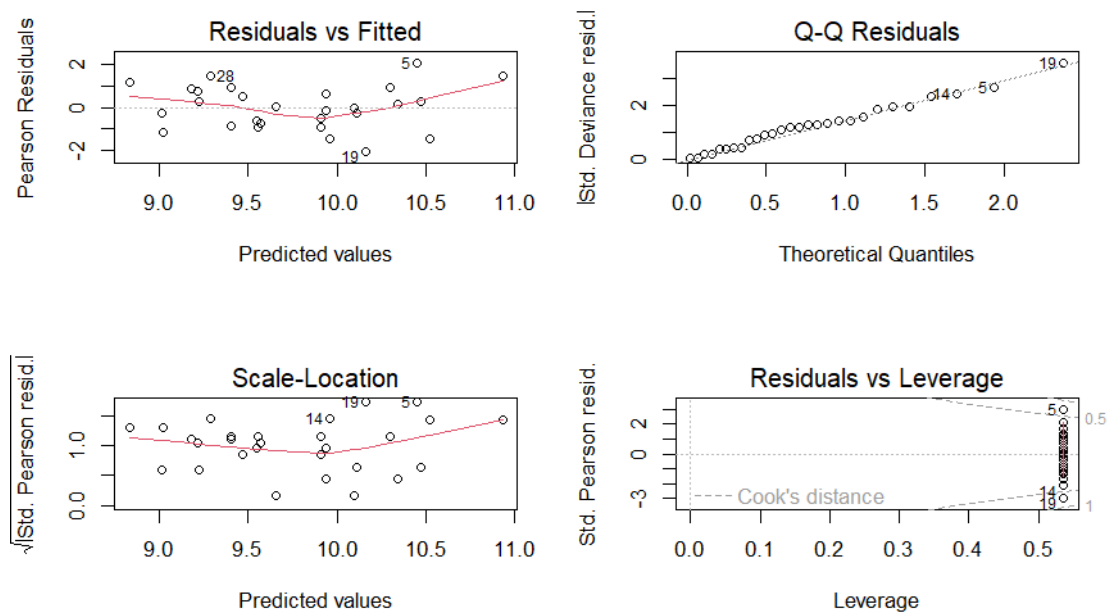


Figura 84: Residuales Modelo Binomial Negativa Zona Alejada Año 2019

Como se observa en la figura 84 año 2019, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta en tabla 9 y figura 85, presentando puntos influyentes en las observaciones 5, 6, 14 y 19, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
5	2.1432849	0.5357954	0.69075044
12	-0.4943344	0.5358684	0.04182692
14	-1.6553413	0.5359585	0.35227211
19	-2.8761792	0.5356870	0.69044955

Tabla 9: Puntos Influyentes Zona Alejada Año 2019

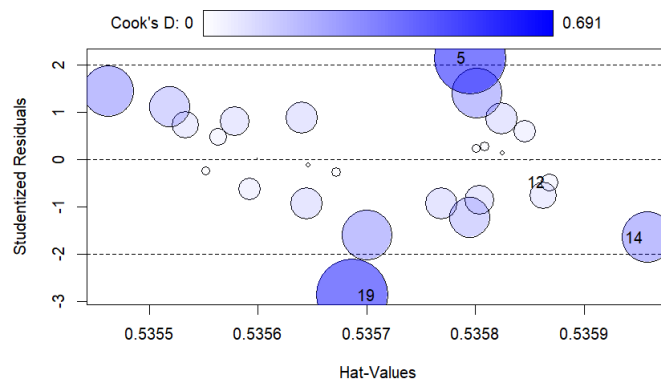


Figura 85: Puntos Influyentes Zona Alejada Año 2019

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la frecuencia esperada, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	10,63354932	10,87475877
Entre 1 y 5 Años	Ambos	8,600864074	8,724850161
Entre 6 y 11 Años	Ambos	4,245631454	4,327856796
Entre 12 y 17 Años	Femenino	5,180528376	4,937701428
Entre 12 y 17 Años	Masculino	3,603042877	3,789503414
Entre 18 y 28 Años	Femenino	9,895753673	6,974873003
Entre 18 y 28 Años	Masculino	3,110766046	5,352957331
Entre 29 y 44 Años	Femenino	10,59196245	8,197317783
Entre 28 y 44 Años	Masculino	4,452939178	6,291138534
Entre 45 y 49 Años	Ambos	10,90479115	10,64353405
Entre 50 y 54 Años	Ambos	14,46959933	13,98662912
Entre 55 y 59 Años	Ambos	17,50723798	17,330998
Entre 60 y 64 Años	Ambos	22,65259034	22,52830015
Entre 65 y 69 Años	Ambos	28,51971326	28,5437368
Entre 70 y 74 Años	Ambos	35,82150776	36,64265354
Entre 75 y 79 Años	Ambos	43,54862385	44,78332012
Mayores de 80 Años	Ambos	45,51104101	50,35117086

Tabla 10: Frecuencias Observadas y Estimadas Zona Alejada Año 2019

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada.

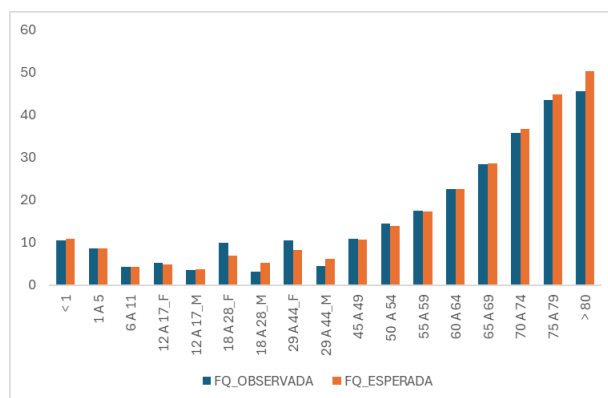


Figura 86: Frecuencia Observada Vs Frecuencia Estimada Zona Alejada Año 2019

El modelo presenta un error porcentual absoluto medio (MAPE) del 12.01 % lo cual presenta un ajuste aceptable; así mismo, al calcular el pseudo R<sup>2</sup>, este es de 0.88.

#### 4.1.2. Zona Bogotá

Como se evidenció en el apartado anterior, la distribución binomial negativa presenta un mejor ajuste para la Zona Bogotá que la distribución Poisson. La aplicación de este modelo se puede observar en el Anexo 2. Ahora bien, el modelo Poisson presenta sobredispersión, lo cual ratifica la necesidad de aplicar el modelo binomial negativa.

```
Overdispersion test
data: fqpoi_BOG
z = 4.806, p-value = 0.00000077
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
70576.01
```

Figura 87: Overdispersion Zona Bogotá Año 2018

```
Overdispersion test
data: fqpoi_BOG_2019
z = 4.9525, p-value = 0.0000003662
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
79611.08
```

Figura 88: Overdispersion Zona Bogotá Año 2019

#### ■ GLM Binomial Negativa

Dado al resultado de la prueba de sobredispersión, se considera que el modelo Poisson no presenta un buen ajuste para ninguno de los dos años por lo que se decide realizar el modelo utilizando la distribución binomial negativa cuyo resultado se presenta continuación, la mayoría de variables son significativas al 95 % con excepción de los grupos 7, 8 y 9; sin embargo dado el objetivo del presente trabajo no es posible realizar el modelo sin estos grupos de edad.

```

Call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2018_BOG, link = log,
      init.theta = 53.40383174)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.84089   0.10016  28.364 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.65227   0.13684   4.767   0.000018737904 ***
GRUPO_ETAREO_RUTA_AGREGADO11  0.92940   0.13684   6.792   0.0000000000111 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.15797   0.13684   8.462 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.33405   0.13684   9.749 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.47177   0.13684  10.755 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.31325   0.13684  -2.289     0.02207 *
GRUPO_ETAREO_RUTA_AGREGADO3  -0.88564   0.13684  -6.472   0.0000000000967 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.78541   0.13684  -5.740   0.0000000094949 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.62635   0.13684  -4.577   0.0000047123777 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.41353   0.13684  -3.022     0.00251 **
GRUPO_ETAREO_RUTA_AGREGADO7  -0.12874   0.13684  -0.941     0.34682
GRUPO_ETAREO_RUTA_AGREGADO8   0.11616   0.13684   0.849     0.39594
GRUPO_ETAREO_RUTA_AGREGADO9   0.35732   0.13684   2.611     0.00902 **
SEXOM                    -0.28541   0.05172  -5.518   0.0000000342576 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(53.4038) family taken to be 1)

Null deviance: 918.135 on 27 degrees of freedom
Residual deviance: 28.087 on 13 degrees of freedom
AIC: 838.79

Number of Fisher Scoring iterations: 1

      Theta: 53.4
    Std. Err.: 14.2

2 x log-likelihood: -806.794

```

Figura 89: Modelo Binomial Negativa Zona Bogotá Año 2018

A continuación se presentan los residuales del modelo binomial negativa.

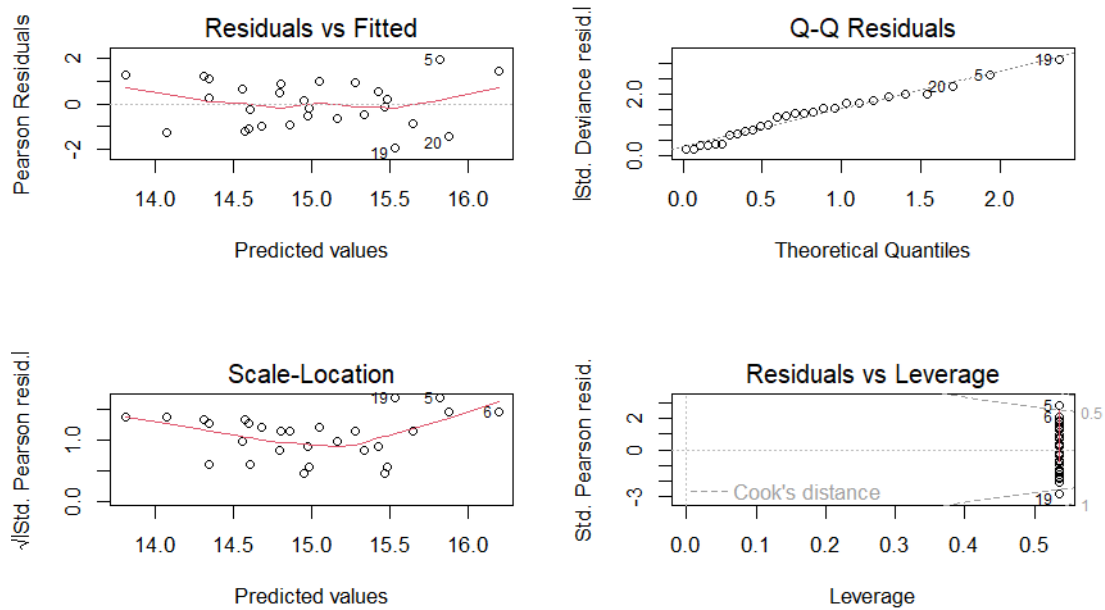


Figura 90: Residuales Modelo Binomial Negativa Zona Alejada Año 2018

Como se observa en la figura número 90, en el gráfico de residuales frente a las predicciones no presenta patrón en los datos, lo cual se corrobora en el gráfico de nivel de dispersión, dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla número 11 y figura 91, presentando puntos influyentes en las observaciones 5,13,14 y 19, sin embargo, dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
5	2.0541652	0.5357152	0.61690432
13	-0.6549518	0.5357169	0.07231088
14	-0.8933492	0.5357171	0.12904267
19	-2.4076381	0.5357141	0.61690158

Tabla 11: Puntos Influyentes Zona Bogotá Año 2018

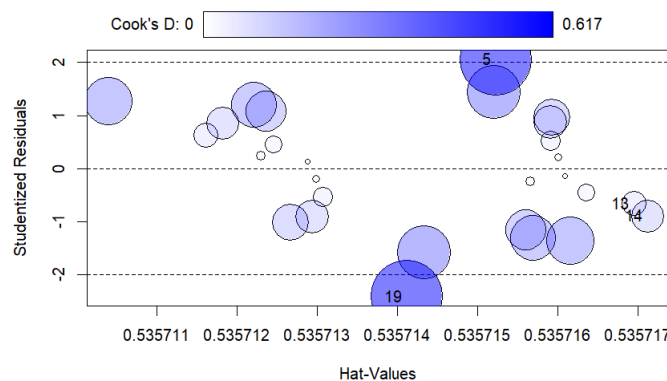


Figura 91: Puntos Influyentes Zona Bogotá Año 2018

Ahora bien, teniendo en cuenta el análisis de los residuales presentados, se realiza la proyección de la frecuencia esperada, cuyos resultados se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	14,63769673	14,98050919
Entre 1 y 5 Años	Ambos	10,71194172	10,95133973
Entre 6 y 11 Años	Ambos	6,057156566	6,173606276
Entre 12 y 17 Años	Femenino	7,539085404	7,810542814
Entre 12 y 17 Años	Masculino	6,075314387	5,87125591
Entre 18 y 28 Años	Femenino	11,5750615	9,15711801
Entre 18 y 28 Años	Masculino	5,065894314	6,883488705
Entre 29 y 44 Años	Femenino	13,55438804	11,32886365
Entre 28 y 44 Años	Masculino	6,84305961	8,516009611
Entre 45 y 49 Años	Ambos	13,60673444	13,2736456
Entre 50 y 54 Años	Ambos	17,44043426	17,01012008
Entre 55 y 59 Años	Ambos	22,06264622	21,70344295
Entre 60 y 64 Años	Ambos	29,44768951	29,24216148
Entre 65 y 69 Años	Ambos	38,43204136	38,62471011
Entre 70 y 74 Años	Ambos	47,81220891	48,63085013
Entre 75 y 79 Años	Ambos	56,69465865	58,23922266
Mayores de 80 Años	Ambos	64,63140339	67,92465273

Tabla 12: Frecuencias Observadas y Estimadas Zona Bogotá Año 2018

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Bogotá.

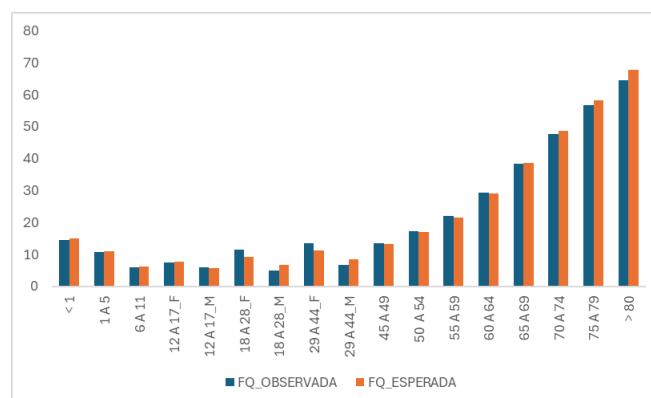


Figura 92: Frecuencia Observada Vs Frecuencia Estimada Zona Bogota Año 2018

El modelo presenta un error porcentual absoluto medio (MAPE) del 7.55 % lo cual presenta un ajuste aceptable; así mismo, al calcular el pseudo R2, este es de 0.94

A continuación se presentan los resultados del modelo usando la distribución binomial negativa para el año 2019, la mayoría de variables son significativas el 95 %, con excepción del grupo etario 2, 7 y 8, sin embargo no es posible la realización del modelo sin estos grupos etarios

```

Call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2019_BOG, link = log,
      init.theta = 50.86356706)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)    2.80856   0.10263  27.366 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.79282   0.14022   5.654  0.0000000156584973 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.07456   0.14022   7.664  0.00000000000000181 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.31014   0.14022   9.344 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.49315   0.14022  10.649 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.61384   0.14022  11.510 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.23133   0.14022  -1.650    0.098980 .
GRUPO_ETAREO_RUTA_AGREGADO3  -0.81288   0.14022  -5.797  0.0000000067416809 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.73178   0.14022  -5.219  0.0000001800006593 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.54509   0.14022  -3.887    0.000101 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.32711   0.14022  -2.333    0.019653 *
GRUPO_ETAREO_RUTA_AGREGADO7  -0.01992   0.14022  -0.142    0.887039 .
GRUPO_ETAREO_RUTA_AGREGADO8   0.24303   0.14022   1.733    0.083059 .
GRUPO_ETAREO_RUTA_AGREGADO9   0.48986   0.14022   3.494    0.000477 ***
SEXOM                -0.28194   0.05300  -5.320  0.0000001037716527 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(50.8636) family taken to be 1)

Null deviance: 943.934  on 27  degrees of freedom
Residual deviance:  28.092  on 13  degrees of freedom
AIC: 845.33

Number of Fisher Scoring iterations: 1

      Theta:  50.9
    Std. Err.:  13.5

2 x log-likelihood:  -813.328

```

Figura 93: Modelo Binomial Negativa Zona Bogotá 2019

A continuación se presentan los residuales del modelo binomial negativa para el año 2019

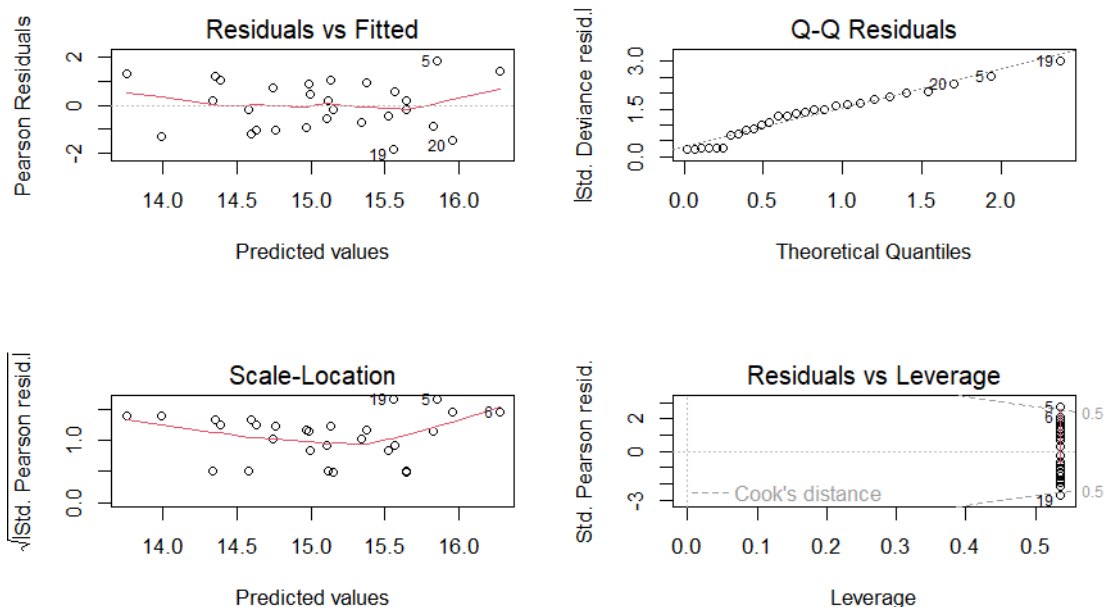


Figura 94: Residuales Modelo Binomial Negativa Zona Alejada Año 2019

Como se observa en la figura 94, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 13 y figura 95, presentando puntos influyentes en las observaciones 5, 13, 14 y 19, sin embargo, dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
5	1.9612860	0.5357151	0.57431375
13	-0.7046078	0.5357164	0.08297074
14	-0.9061060	0.5357166	0.13229505
19	-2.2781393	0.5357141	0.57431131

Tabla 13: Puntos Influyentes Zona Bogotá Año 2019

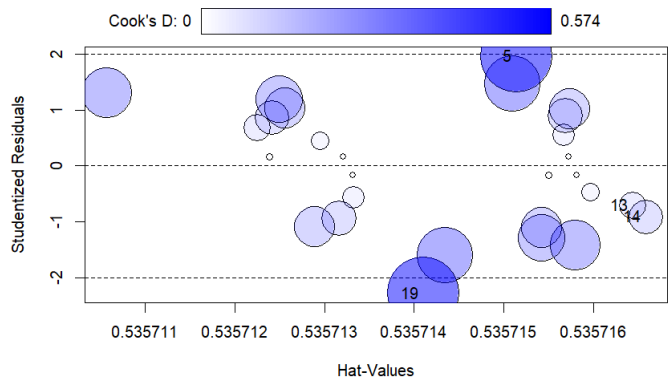


Figura 95: Puntos Influyentes Zona Bogotá Año 2019

Ahora bien, teniendo en cuenta el análisis de los residuales presentados se realiza la proyección de la frecuencia esperada, cuyos resultados para el año 2019 se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	14,18907443	14,49997949
Entre 1 y 5 Años	Ambos	11,27578552	11,5093166
Entre 6 y 11 Años	Ambos	6,319898103	6,435540118
Entre 12 y 17 Años	Femenino	7,779917836	7,978678788
Entre 12 y 17 Años	Masculino	6,168376071	6,01844647
Entre 18 y 28 Años	Femenino	12,12679161	9,6163685
Entre 18 y 28 Años	Masculino	5,360125378	7,253782311
Entre 29 y 44 Años	Femenino	14,39653674	11,95850302
Entre 28 y 44 Años	Masculino	7,181441087	9,020492268
Entre 45 y 49 Años	Ambos	14,73641141	14,35064748
Entre 50 y 54 Años	Ambos	19,2100163	18,72013222
Entre 55 y 59 Años	Ambos	24,45571245	24,02732382
Entre 60 y 64 Años	Ambos	32,80717289	32,62099855
Entre 65 y 69 Años	Ambos	43,04443499	43,31487472
Entre 70 y 74 Años	Ambos	53,92149327	54,87393328
Entre 75 y 79 Años	Ambos	64,26700254	66,16333981
Mayores de 80 Años	Ambos	72,07421165	75,83997976

Tabla 14: Frecuencias Observadas y Estimadas Zona Bogotá Año 2019

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Bogotá Año 2019.

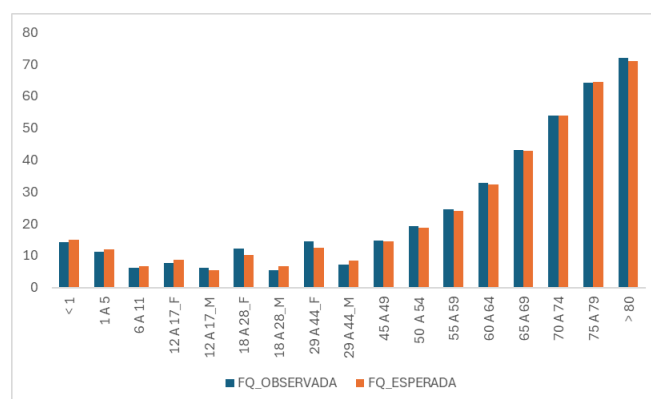


Figura 96: Frecuencia Observada Vs Frecuencia Estimada Zona Bogotá Año 2019

El modelo para el año 2019 presenta un error porcentual absoluto medio (MAPE) del 7.51 % lo cual presenta un ajuste aceptable; así mismo al calcular el pseudo R2 este es de 0.95

#### 4.1.3. Zona Normal Mayor

Como se evidencio en el apartado anterior, la distribución binomial negativa presenta un mejor ajuste para la Zona Normal Mayor que la distribución poisson, la aplicación de este modelo se puede observar en el Anexo 2, ahora bien el modelo poisson presenta sobredispersión en los dos años, lo cual ratifica la necesidad de aplicar el modelo binomial negativa, lo cual se presenta a continuación.

```

Overdispersion test
data: fgp01_NME
z = 4.6657, p-value = 0.000001538
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
50543.79

```

Figura 97: Overdispersion Zona Normal Mayor Año 2018

```

Overdispersion test
data: fpoi_NM_2019
z = 4.5626, p-value = 0.00002526
alternative hypothesis: true alpha is greater than 0
sample estimates:
  alpha
149397.1

```

Figura 98: Overdispersion Zona Normal Mayor Año 2019

#### ■ GLM Binomial Negativa

Dado al resultado de la prueba de sobredispersión, se considera que el modelo poisson no presenta un buen ajuste en ninguno de los dos años por lo que se decide realizar el modelo utilizando la distribución binomial negativa cuyo resultado se presenta continuación, la mayoría de variables son significativas al 95 % con excepción del grupo etario 2, 7, 8 sin embargo, no es posible establecer el modelo sin estos grupos etarios.

```

Call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO + offset(log(EXPUESTOS)), data = BASE_2018_NM, link = log,
  init.theta = 52.05396104)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.76361    0.10145  27.242 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.78448    0.13860   5.660  0.0000000151494151 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.04942    0.13860   7.571  0.0000000000000369 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.24618    0.13860   8.991 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.39556    0.13860  10.069 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.50460    0.13860  10.855 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.20719    0.13860  -1.495    0.134961
GRUPO_ETAREO_RUTA_AGREGADO3  -0.83119    0.13860  -5.997  0.0000000020118621 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.76827    0.13860  -5.543  0.0000000297473001 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.59130    0.13860  -4.266  0.0000198910782273 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.32097    0.13860  -2.316    0.020572 *
GRUPO_ETAREO_RUTA_AGREGADO7  -0.01619    0.13860  -0.117    0.906999
GRUPO_ETAREO_RUTA_AGREGADO8   0.25773    0.13860   1.859    0.062965 .
GRUPO_ETAREO_RUTA_AGREGADO9   0.50744    0.13860   3.661    0.000251 ***
SEXOM                 -0.29003    0.05239  -5.536  0.0000000309011115 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(52.054) family taken to be 1)

Null deviance: 899.10 on 27 degrees of freedom
Residual deviance: 28.09 on 13 degrees of freedom
AIC: 872.89

Number of Fisher Scoring iterations: 1

Theta: 52.1
Std. Err.: 13.9

2 x log-likelihood: -840.892

```

Figura 99: Modelo Binomial Negativa Zona Normal Mayor Año 2018

A continuación se presentan los residuales del modelo binomial negativa

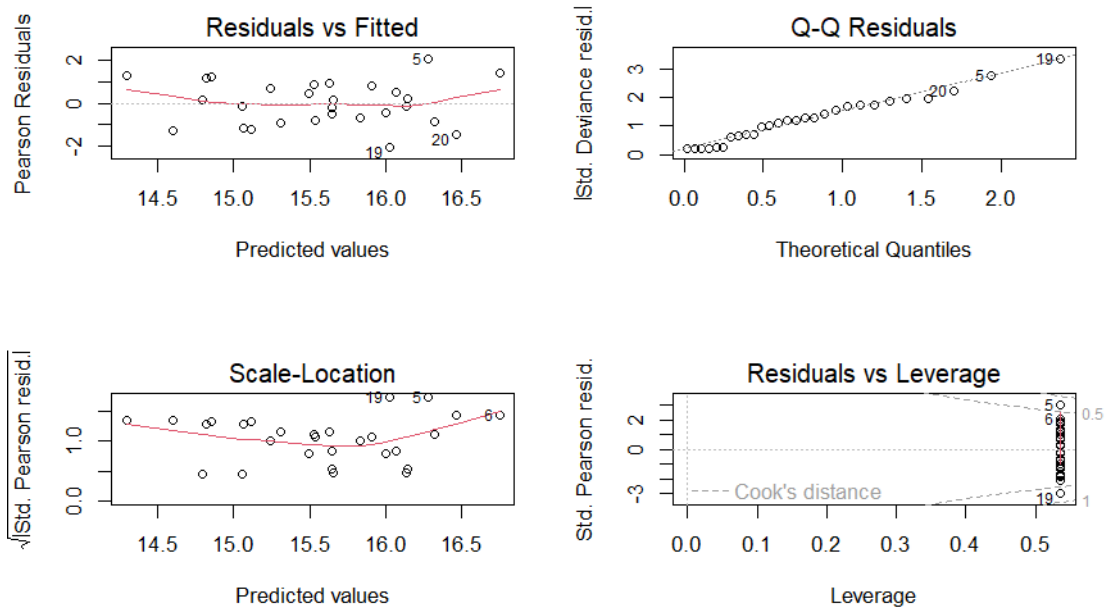


Figura 100: Residuales Modelo Binomial Negativa Zona Normal Mayor Año 2018

Como se observa en la figura número 100, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla número 95 y figura 262, presentando puntos influyentes en las observaciones 1,5,13 y 19 , sin embargo, dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	-1.3448317	0.5357156	0.26402956
5	2.2241588	0.5357148	0.69869006
13	-0.6777157	0.5357156	0.07711998
19	-2.6862098	0.5357142	0.69868841

Tabla 15: Puntos Influyentes Zona Normal Mayor Año 2018

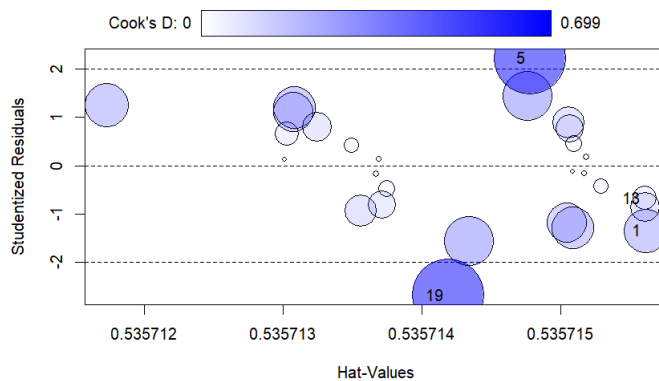


Figura 101: Puntos Influyentes Zona Normal Mayor Año 2018

Ahora bien, teniendo en cuenta el análisis de los residuales presentados se realiza la proyección de la frecuencia esperada, cuyos resultados se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	13,50907222	13,87297451
Entre 1 y 5 Años	Ambos	10,99726612	11,2483084
Entre 6 y 11 Años	Ambos	5,902002292	6,020078274
Entre 12 y 17 Años	Femenino	7,218939832	7,354666837
Entre 12 y 17 Años	Masculino	5,60462334	5,503066322
Entre 18 y 28 Años	Femenino	11,27715305	8,778525138
Entre 18 y 28 Años	Masculino	4,698876819	6,568456072
Entre 29 y 44 Años	Femenino	13,77582639	11,50333857
Entre 28 y 44 Años	Masculino	6,90690328	8,607274329
Entre 45 y 49 Años	Ambos	13,94848694	13,6700743
Entre 50 y 54 Años	Ambos	18,41079766	18,04391906
Entre 55 y 59 Años	Ambos	23,56949412	23,24630761
Entre 60 y 64 Años	Ambos	31,01496768	30,82137123
Entre 65 y 69 Años	Ambos	39,93318797	40,13784769
Entre 70 y 74 Años	Ambos	48,20508079	48,94810156
Entre 75 y 79 Años	Ambos	55,60432173	57,14443519
Mayores de 80 Años	Ambos	61,77905086	64,59195982

Tabla 16: Frecuencias Observadas y Estimadas Zona Normal Mayor Año 2018

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Normal Mayor.

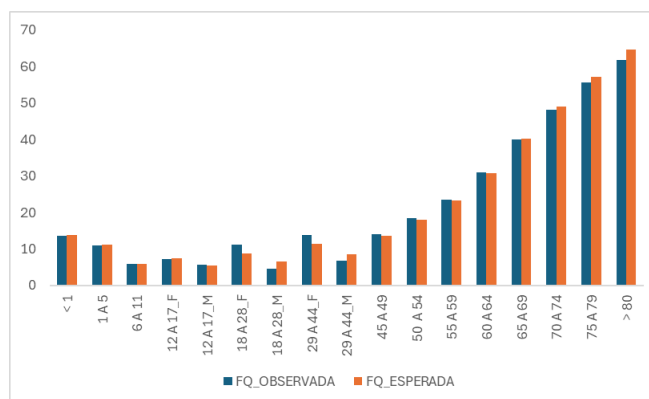


Figura 102: Frecuencia Observada Vs Frecuencia Estimada Zona Normal Mayor Año 2018

El modelo presenta un error porcentual absoluto medio (MAPE) del 7.59 % lo cual se presenta un ajuste aceptable; así mismo al calcular el pseudo R2 este es de 0.94

A continuación se presenta el análisis del modelo binomial negativa para el año 2019.

```
Call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
SEXO + offset(log(EXPUESTOS)), data = BASE_2019_NM, link = log,
init.theta = 51.61822443)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)    2.77163    0.10188  27.206 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.89746    0.13919   6.448  0.000000000113 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.16897    0.13919   8.399 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.37190    0.13919   9.856 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.53267    0.13919  11.012 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.64767    0.13919  11.838 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.12351    0.13919  -0.887    0.374883
GRUPO_ETAREO_RUTA_AGREGADO3  -0.73896    0.13919  -5.309  0.00000110198 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.65506    0.13919  -4.706  0.000002523054 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.50285    0.13919  -3.613    0.000303 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.23283    0.13919  -1.673    0.094369 .
GRUPO_ETAREO_RUTA_AGREGADO7   0.08804    0.13919   0.633    0.527054
GRUPO_ETAREO_RUTA_AGREGADO8   0.36617    0.13919   2.631    0.008520 **
GRUPO_ETAREO_RUTA_AGREGADO9   0.61695    0.13919   4.433  0.000009314400 ***
SEXOM             -0.27818    0.05261  -5.288  0.00000123848 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(51.6182) family taken to be 1)

Null deviance: 934.08  on 27  degrees of freedom
Residual deviance:  28.09  on 13  degrees of freedom
AIC: 881.01

Number of Fisher Scoring iterations: 1

Theta: 51.6
Std. Err.: 13.8

2 x log-likelihood: -849.006
```

Figura 103: Modelo Binomial Negativa Zona Normal Mayor Año 2019

A continuación se presentan los residuales del modelo binomial negativa

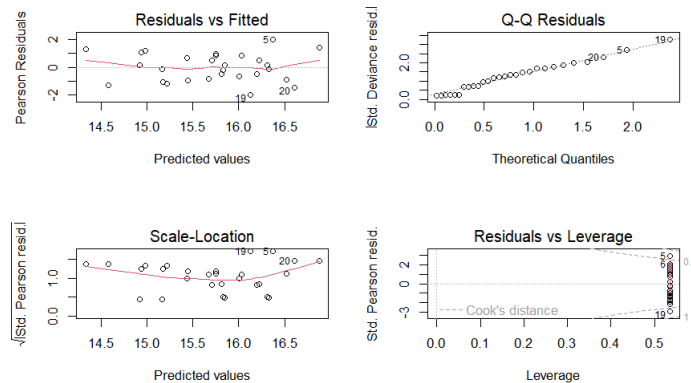


Figura 104: Residuales Modelo Binomial Negativa Zona Normal Mayor Año 2019

Como se observa en la figura 104, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta en tabla 47 y figura 105, presentando puntos influyentes en las observaciones 5,13, 14 y 19. sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
5	2.1530565	0.5357147	0.66498557
13	-0.6602358	0.5357154	0.07338839
14	-0.8777023	0.5357154	0.12485345
19	-2.5717217	0.5357142	0.66498449

Tabla 17: Puntos Influyentes Zona Normal Mayor Año 2019

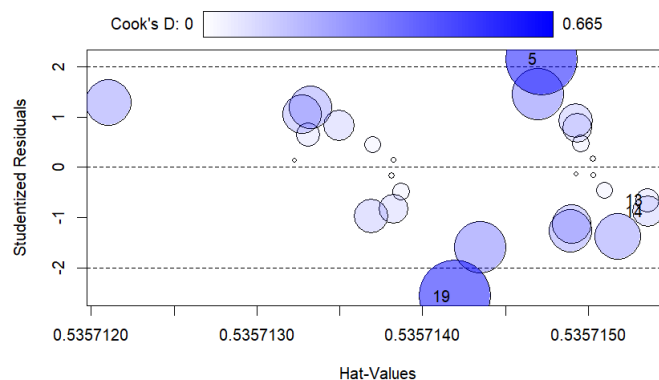


Figura 105: Puntos Influyentes Zona Normal Mayor Año 2019

Ahora bien, teniendo en cuenta el análisis de los residuales presentados se realiza la proyección de la frecuencia esperada, cuyos resultados se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	13,70525129	14,00901086
Entre 1 y 5 Años	Ambos	12,13087589	12,37599654
Entre 6 y 11 Años	Ambos	6,569127886	6,687559423
Entre 12 y 17 Años	Femenino	8,144195654	8,302629657
Entre 12 y 17 Años	Masculino	6,406407133	6,286446258
Entre 18 y 28 Años	Femenino	12,36334884	9,667558982
Entre 18 y 28 Años	Masculino	5,278764774	7,319920615
Entre 29 y 44 Años	Femenino	15,21509629	12,66442424
Entre 28 y 44 Años	Masculino	7,65775983	9,589036926
Entre 45 y 49 Años	Ambos	15,67398537	15,36471116
Entre 50 y 54 Años	Ambos	20,76762477	20,35967053
Entre 55 y 59 Años	Ambos	26,62156618	26,25981272
Entre 60 y 64 Años	Ambos	35,12788219	34,92867248
Entre 65 y 69 Años	Ambos	45,59727769	45,84037457
Entre 70 y 74 Años	Ambos	55,29389945	56,18603756
Entre 75 y 79 Años	Ambos	64,62295897	66,31939447
Mayores de 80 Años	Ambos	72,04519273	75,40292146

Tabla 18: Frecuencias Observadas y Estimadas Zona Normal Mayor Año 2019

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Normal Mayor.

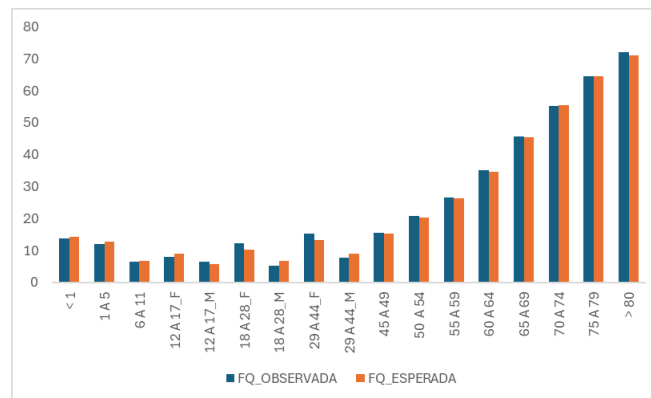


Figura 106: Frecuencia Observada Vs Frecuencia Estimada Zona Normal Mayor Año 2019

El modelo presenta un error porcentual absoluto medio (MAPE) del 7.50 % lo cual se presenta un ajuste aceptable; así mismo al calcular el pseudo R2 este es de 0.95

#### 4.1.4. Zona Normal Medio

Como se evidencio en el apartado anterior, la distribución binomial negativa presenta un mejor ajuste para la Zona Normal Medio que la distribución poisson, la aplicación de este modelo se puede observar en el Anexo 2, ahora bien, el modelo poisson presenta sobredispersión en los dos años, lo cual ratifica la necesidad de aplicar el modelo con distribución de binomial negativa, lo cual se presenta a continuación.

```
Overdispersion test
data: fqpoi_NME
z = 4.6657, p-value = 0.000001538
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
50543.79
```

Figura 107: Overdispersion Zona Normal Medio Año 2018

```
Overdispersion test
data: fqpoi_NME_2019
z = 4.8018, p-value = 0.0000007861
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
59309.11
```

Figura 108: Overdispersion Zona Normal Medio Año 2019

#### ■ GLM Binomial Negativa

Dado al resultado de la prueba de sobredispersión, se considera que el modelo poisson no presenta un buen ajuste por lo que se decide realizar el modelo utilizando la distribución binomial negativa para los dos años, los resultados y análisis para el año 2018 se presentan a continuación.

```

Call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2018_NME, link = log,
      init.theta = 41.0641457)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.62956    0.11422  23.022 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.74129    0.15605   4.750  0.0000020319966662 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.00025    0.15606   6.410  0.0000000001459327 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.20532    0.15606   7.724  0.00000000000000113 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.36348    0.15606   8.737 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.44113    0.15606   9.235 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.21029    0.15606  -1.348      0.1778
GRUPO_ETAREO_RUTA_AGREGADO3  -0.81235    0.15606  -5.205  0.0000001934875770 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.76304    0.15606  -4.890  0.0000010105690389 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.67169    0.15605  -4.304  0.0000167611918683 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.38458    0.15605  -2.464      0.0137 *
GRUPO_ETAREO_RUTA_AGREGADO7  -0.07413    0.15606  -0.475      0.6348
GRUPO_ETAREO_RUTA_AGREGADO8   0.20114    0.15606   1.289      0.1974
GRUPO_ETAREO_RUTA_AGREGADO9   0.45703    0.15605   2.929      0.0034 **
SEXOM                    -0.34332    0.05898  -5.821  0.0000000058634291 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(41.0641) family taken to be 1)

Null deviance: 704.564  on 27  degrees of freedom
Residual deviance:  28.114  on 13  degrees of freedom
AIC: 813.19

Number of Fisher Scoring iterations: 1

      Theta:  41.1
    Std. Err.:  10.9

2 x log-likelihood:  -781.193

```

Figura 109: Modelo Binomial Negativa Zona Normal Medio Año 2018

La mayoría de las variables son significativas con una confianza del 95%, dado que los dos modelos presentan significancia en las variables se realiza la prueba de Vuong, cuyos resultados se presentan continuación.

Estadístico	Vuong z-statistic	HA	p-value
Raw	-4.509826	model2 / $\hat{\mu}$ model1	0.000003244
AIC-corrected	-4.509826	model2 / $\hat{\mu}$ model1	0.000003244
BIC-corrected	-4.509826	model2 / $\hat{\mu}$ model1	0.000003244

Tabla 19: Prueba de Vuong Zona Normal Medio Año 2018

Lo anterior evidencia que el modelo 2 que corresponde a la binomial negativa es mejor que el modelo poisson.

Con base a lo anterior, se presentan los residuales del modelo binomial negativa.

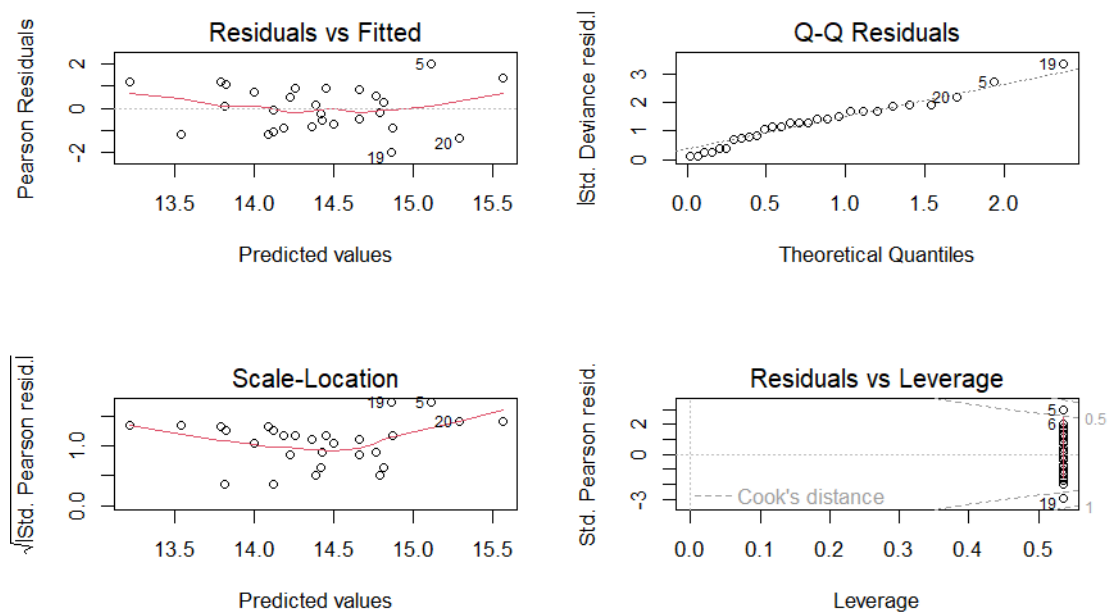


Figura 110: Residuales Modelo Binomial Negativa Zona Normal Medio Año 2018

Como se observa en la figura 110, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla número 95 y figura número 262, presentando puntos influyentes en las observaciones 1,5,13 y 19, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	-1.2966103	0.5357177	0.24638668
5	2.1606717	0.5357155	0.67654109
13	-0.7521988	0.5357174	0.09346342
19	-2.6562378	0.5357140	0.67653798

Tabla 20: Puntos Influyentes Zona Normal Medio Año 2018

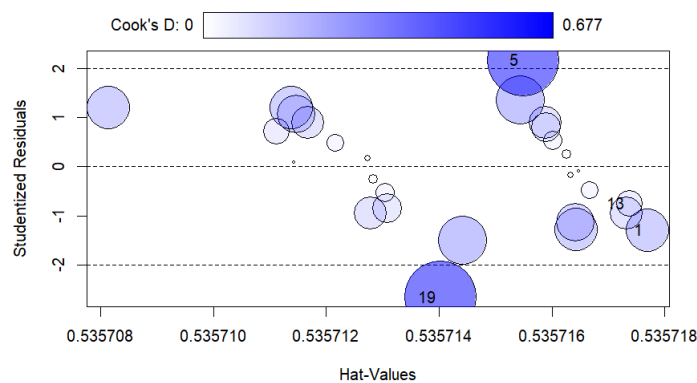


Figura 111: Puntos Influyentes Zona Normal Medio Año 2018

Ahora bien, teniendo en cuenta el análisis de los residuales presentados se realiza la proyección de la frecuencia esperada, cuyos resultados se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	11,47188218	11,83180792
Entre 1 y 5 Años	Ambos	9,301505393	9,568527471
Entre 6 y 11 Años	Ambos	5,108568093	5,241925231
Entre 12 y 17 Años	Femenino	6,374022147	6,465724929
Entre 12 y 17 Años	Masculino	4,6519336	4,586877588
Entre 18 y 28 Años	Femenino	9,318202763	7,084247071
Entre 18 y 28 Años	Masculino	3,44085993	5,025666027
Entre 29 y 44 Años	Femenino	11,45420175	9,440184133
Entre 28 y 44 Años	Masculino	5,268224461	6,697001418
Entre 45 y 49 Años	Ambos	11,14544896	10,93517186
Entre 50 y 54 Años	Ambos	14,70740097	14,43408616
Entre 55 y 59 Años	Ambos	18,97902127	18,71288113
Entre 60 y 64 Años	Ambos	25,19410079	24,98939216
Entre 65 y 69 Años	Ambos	32,22125488	32,39837649
Entre 70 y 74 Años	Ambos	39,21155628	39,88375874
Entre 75 y 79 Años	Ambos	45,61087848	46,96166731
Mayores de 80 Años	Ambos	49,03463133	51,23699178

Tabla 21: Frecuencias Observadas y Estimadas Zona Normal Medio Año 2018

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Normal Medio.

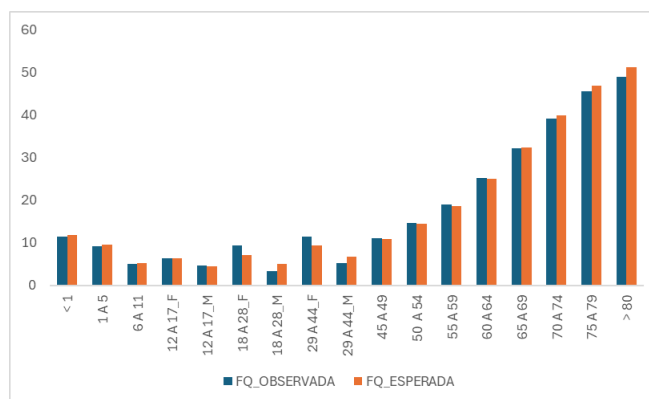


Figura 112: Frecuencia Observada Vs Frecuencia Estimada Zona Normal Medio Año 2018

El modelo presenta un error porcentual absoluto medio (MAPE) del 8,34 % lo cual se presenta un ajuste aceptable; así mismo al calcular el pseudo R<sup>2</sup> este es de 0,92

Los resultados de la aplicación del modelo binomial negativa para el año 2019 se presentan a continuación.

```

call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2019_NME, link = log,
      init.theta = 40.04513622)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.668280   0.115666  23.069 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.801631   0.158028   5.073 0.000000392160498058 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.068255   0.158028   6.760 0.000000000013806806 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.273171   0.158028   8.057 0.000000000000000784 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.442446   0.158028   9.128 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.541361   0.158028   9.754 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.147633   0.158028  -0.934      0.35019
GRUPO_ETAREO_RUTA_AGREGADO3  -0.739518   0.158028  -4.680 0.000002873649048197 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.674643   0.158028  -4.269 0.000019624080244062 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.623609   0.158028  -3.946 0.000079399462006002 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.325148   0.158027  -2.058      0.03963 *
GRUPO_ETAREO_RUTA_AGREGADO7  -0.006889   0.158028  -0.044      0.96523
GRUPO_ETAREO_RUTA_AGREGADO8   0.263386   0.158028   1.667      0.09557 .
GRUPO_ETAREO_RUTA_AGREGADO9   0.514467   0.158028   3.256      0.00113 **
SEXOM                  -0.334828   0.059728  -5.606 0.000000020723980543 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(40.0451) family taken to be 1)

Null deviance: 701.555 on 27 degrees of freedom
Residual deviance: 28.117 on 13 degrees of freedom
AIC: 821.94

Number of Fisher Scoring iterations: 1

      Theta: 40.0
    Std. Err.: 10.7

2 x log-likelihood: -789.941

```

Figura 113: Modelo Binomial Negativa Zona Normal Medio Año 2019

Las variables son significativas con una confianza del 95 %, dado que los dos modelos poisson y binomial negativo presentan significancia en la variables se realiza la prueba de vuong, cuyos resultados se presentan continuación.

Estadístico	Vuong z-statistic	HA	p-value
Raw	-4.062179	model2 model1	0.000024308
AIC-corrected	-4.062179	model2 model1	0.000024308
BIC-corrected	-4.062179	model2 model1	0.000024308

Tabla 22: Prueba de Vuong Zona Normal Medio Año 2019

Lo anterior evidencia que el modelo 2 que corresponde a la distribución binomial negativa es mejor que el modelo con distribución poisson.

Con base a lo anterior, se presentan los residuales del modelo binomial negativa.

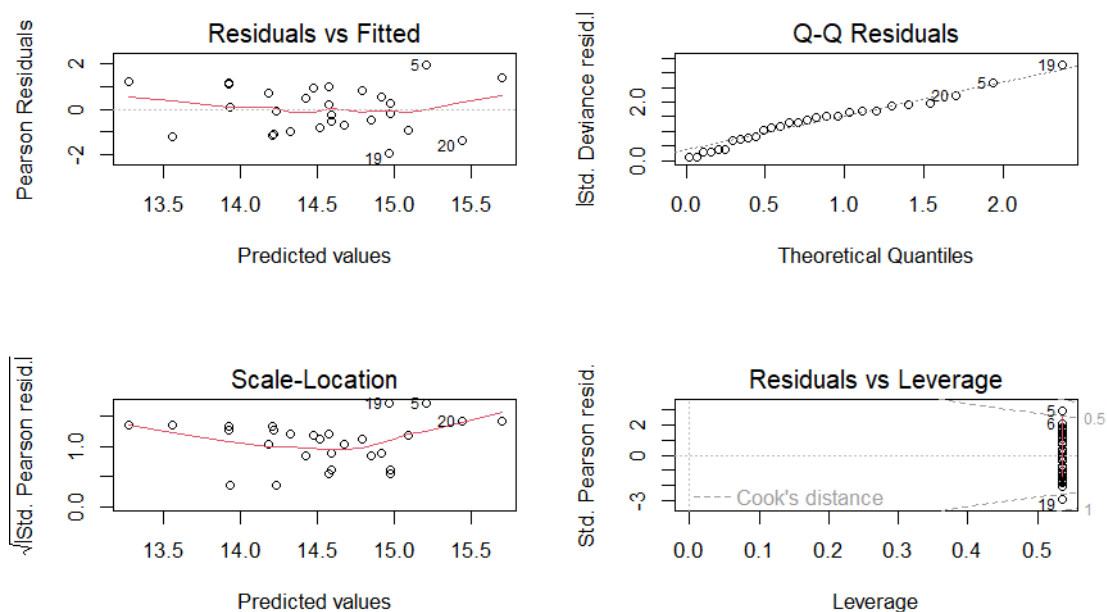


Figura 114: Residuales Modelo Binomial Negativa Zona Normal Medio Año 2019

Como se observa en la figura número 114 en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 23 y figura 115, presentando puntos influyentes en las observaciones 1,5, 13 y 19, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	-1.3142469	0.5357169	0.25174628
5	2.0972441	0.5357153	0.64657420
13	-0.7323927	0.5357168	0.08888066
19	-2.5548933	0.5357141	0.64657170

Tabla 23: Puntos Influyentes Zona Normal Medio Año 2019

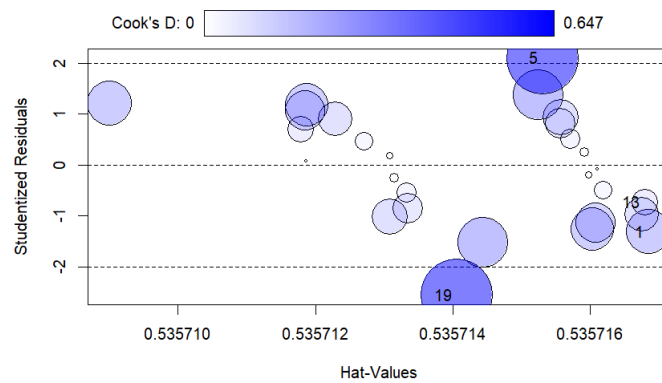


Figura 115: Puntos Influyentes Zona Normal Medio Año 2019

Ahora bien, teniendo en cuenta el análisis de los residuales presentados se realiza la proyección de la frecuencia esperada, cuyos resultados para el año 2019 se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	11,9742917	12,31013727
Entre 1 y 5 Años	Ambos	10,33946485	10,62408868
Entre 6 y 11 Años	Ambos	5,733694048	5,881374812
Entre 12 y 17 Años	Femenino	7,244180201	7,342190408
Entre 12 y 17 Años	Masculino	5,323173459	5,253050303
Entre 18 y 28 Años	Femenino	10,13867762	7,726614495
Entre 18 y 28 Años	Masculino	3,802348155	5,528090714
Entre 29 y 44 Años	Femenino	12,68924397	10,41379827
Entre 28 y 44 Años	Masculino	5,822669424	7,45066569
Entre 45 y 49 Años	Ambos	12,43393976	12,19641853
Entre 50 y 54 Años	Ambos	16,30147324	16,01066767
Entre 55 y 59 Años	Ambos	20,95495456	20,66848629
Entre 60 y 64 Años	Ambos	27,90561999	27,68663933
Entre 65 y 69 Años	Ambos	35,96260116	36,19554964
Entre 70 y 74 Años	Ambos	43,76864775	44,49502632
Entre 75 y 79 Años	Ambos	51,51583281	52,99645782
Mayores de 80 Años	Ambos	56,46180266	59,12786761

Tabla 24: Frecuencias Observadas y Estimadas Zona Normal Medio Año 2019

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Normal Medio para el año 2019.

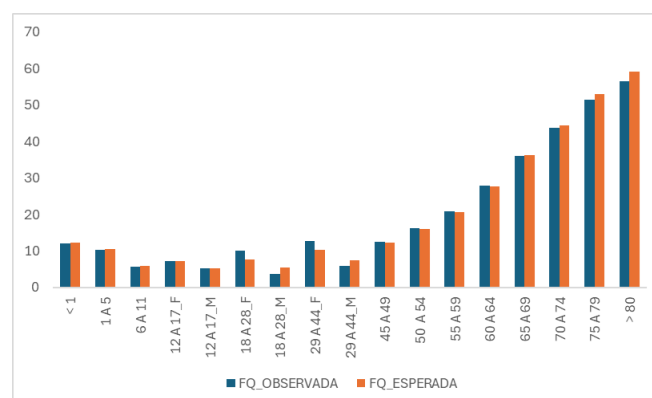


Figura 116: Frecuencia Observada Vs Frecuencia Estimada Zona Normal Medio Año 2019

El modelo presenta un error porcentual absoluto medio (MAPE) del 8,33 % lo cual se presenta un ajuste aceptable; así mismo al calcular el pseudo R<sup>2</sup> este es de 0.92

#### 4.1.5. Zona Rural

Como se evidencio en el apartado anterior la distribución binomial negativa presenta un mejor ajuste para la Zona Rural que la distribución poisson, la aplicación de este modelo se puede observar en el Anexo 2, ahora bien el modelo poisson presenta sobredispersión en los dos años, lo cual ratifica la necesidad de aplicar el modelo binomial negativa, lo cual se presenta a continuación.

```
Overdispersion test
data: fgpoi_RU
z = 5.453, p-value = 0.0000002476
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
15996.65
```

Figura 117: Overdispersion Zona Rural Año 2018

```
Overdispersion test
data: fgpoi_RU_2019
z = 5.4609, p-value = 0.0000002369
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
20300.46
```

Figura 118: Overdispersion Zona Rural Año 2019

- GLM Binomial Negativa

Dado al resultado de la prueba de sobredispersión, se considera que el modelo poisson no presenta un buen ajuste por lo que se decide realizar el modelo utilizando la distribución binomial negativa cuyo resultado para el año 2018 se presenta continuación.

```

all:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2018_RU, link = log,
      init.theta = 27.32677083)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)    2.30360    0.14003  16.451 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.64432    0.19131   3.368    0.000757 ***
GRUPO_ETAREO_RUTA_AGREGADO11  0.94464    0.19131   4.938    0.000000789993063 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.16931    0.19131   6.112    0.000000000982700 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.33309    0.19131   6.968    0.000000000003210 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.41576    0.19131   7.400    0.000000000000136 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.29154    0.19131  -1.524    0.127523
GRUPO_ETAREO_RUTA_AGREGADO3  -0.93359    0.19131  -4.880    0.000001061005046 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.82214    0.19131  -4.297    0.000017277326778 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.96762    0.19131  -5.058    0.000000423807475 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.62270    0.19131  -3.255    0.001134 **
GRUPO_ETAREO_RUTA_AGREGADO7  -0.25418    0.19131  -1.329    0.183974
GRUPO_ETAREO_RUTA_AGREGADO8   0.02193    0.19131   0.115    0.908738
GRUPO_ETAREO_RUTA_AGREGADO9   0.32522    0.19131   1.700    0.089135 .
SEXOM                -0.42086    0.07231  -5.821    0.000000005867069 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(27.3268) family taken to be 1)

Null deviance: 538.394 on 27 degrees of freedom
Residual deviance: 28.171 on 13 degrees of freedom
AIC: 737.88

Number of Fisher Scoring iterations: 1

      Theta: 27.33
    Std. Err.: 7.26

2 x log-likelihood: -705.884

```

Figura 119: Modelo Binomial Negativa Zona Rural Año 2018

Las variables son significativas con una confianza del 95 %, dado que los dos modelos presentan significancia en la variables se realiza la prueba de vuong, cuyos resultados se presentan continuación.

Estadístico	Vuong z-statistic	HA	p-value
Raw	-2.286657	model2 /¿model1	0.011108
AIC-corrected	-2.286657	model2 /¿model1	0.011108
BIC-corrected	-2.286657	model2 /¿model1	0.011108

Tabla 25: Prueba de Vuong Zona Rural Año 2018

Lo anterior evidencia que el modelo 2 que corresponde a la distribución binomial negativa es mejor que el modelo cuya distribución es poisson.

Con base a lo anterior se presentan los residuales del modelo binomial negativa

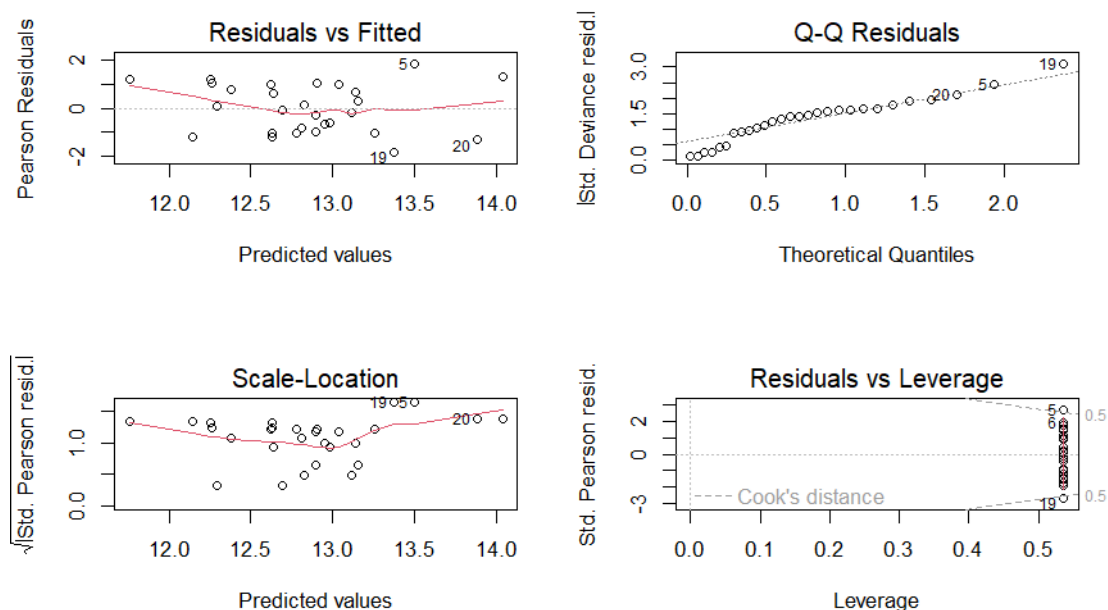


Figura 120: Residuales Modelo Binomial Negativa Zona Rural Año 2018

Como se observa en la figura número 120, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla número 26 y figura número 121, presentando puntos influyentes en las observaciones 1,5, 14 y 19, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	-1.307578	0.5357266	0.2456431
5	1.895649	0.5357169	0.5607421
14	-1.046241	0.5357245	0.1682592
19	-2.328617	0.5357146	0.5607386

Tabla 26: Puntos Influyentes Zona Rural Año 2018

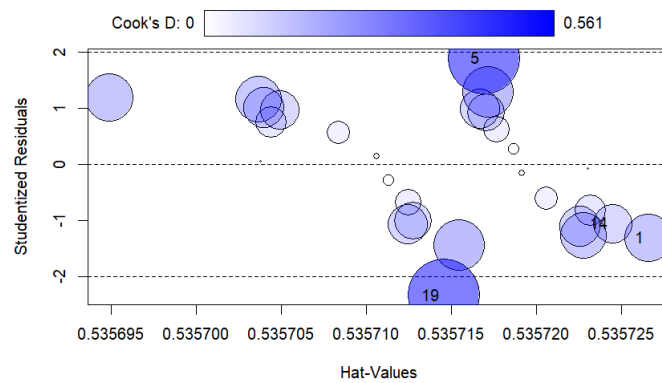


Figura 121: Puntos Influyentes Zona Rural Año 2018

Ahora bien, teniendo en cuenta el análisis de los residuales presentados, se realiza la proyección de la frecuencia esperada, cuyos resultados se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	7,894117647	8,159177166
Entre 1 y 5 Años	Ambos	5,905332599	6,100542513
Entre 6 y 11 Años	Ambos	3,123404575	3,214646502
Entre 12 y 17 Años	Femenino	4,342069246	4,228367405
Entre 12 y 17 Años	Masculino	2,925708518	3,006411451
Entre 18 y 28 Años	Femenino	5,142384966	4,054291277
Entre 18 y 28 Años	Masculino	1,618283757	2,192985431
Entre 29 y 44 Años	Femenino	6,720090749	5,726152896
Entre 28 y 44 Años	Masculino	2,639487463	3,179743368
Entre 45 y 49 Años	Ambos	6,31097385	6,204242406
Entre 50 y 54 Años	Ambos	8,341791031	8,204422811
Entre 55 y 59 Años	Ambos	11,33479816	11,21768517
Entre 60 y 64 Años	Ambos	15,64332474	15,57701843
Entre 65 y 69 Años	Ambos	20,93754233	21,10235762
Entre 70 y 74 Años	Ambos	25,97322319	26,40251732
Entre 75 y 79 Años	Ambos	30,45937223	30,93592832
Mayores de 80 Años	Ambos	32,84206324	33,03680823

Tabla 27: Frecuencias Observadas y Estimadas Zona Rural Año 2018

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Rural.

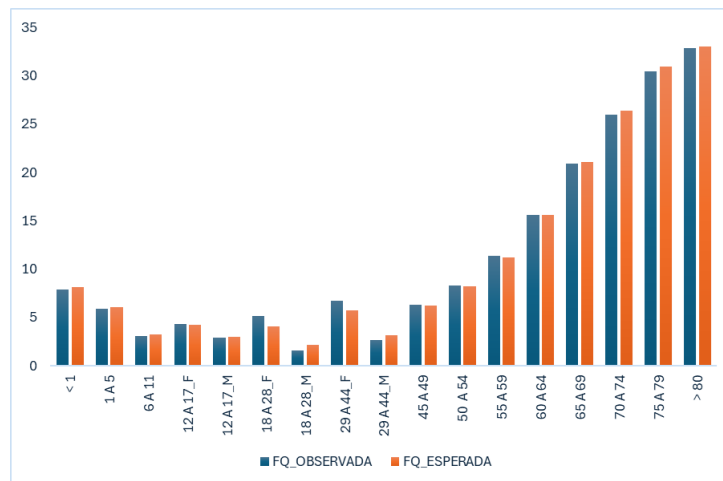


Figura 122: Frecuencia Observada Vs Frecuencia Estimada Zona Rural Año 2018

El modelo para el año 2018 presenta un error porcentual absoluto medio (MAPE) del 6.84% lo cual presenta un ajuste aceptable; así mismo, al calcular el pseudo R<sup>2</sup>, este es de 0.87.

A continuación se presenta el análisis y resultados del modelo utilizando la distribución binomial negativa para el año 2019.

```

Call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2019_RU, link = log,
      init.theta = 26.07794117)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)    2.43674    0.14334  17.000 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.68446    0.19583   3.495    0.000474 ***
GRUPO_ETAREO_RUTA_AGREGADO11  0.98486    0.19583   5.029  0.0000004928166910 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.23155    0.19583   6.289  0.0000000003199205 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.40214    0.19583   7.160  0.0000000000008075 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.50460    0.19583   7.683  0.0000000000000155 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.26326    0.19583  -1.344    0.178856
GRUPO_ETAREO_RUTA_AGREGADO3  -0.84796    0.19583  -4.330  0.0000149116314606 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.75321    0.19583  -3.846    0.000120 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -1.00224    0.19583  -5.118  0.0000003090206097 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.60349    0.19583  -3.082    0.002058 **
GRUPO_ETAREO_RUTA_AGREGADO7  -0.22372    0.19583  -1.142    0.253283
GRUPO_ETAREO_RUTA_AGREGADO8   0.05934    0.19583   0.303    0.761896
GRUPO_ETAREO_RUTA_AGREGADO9   0.35619    0.19583   1.819    0.068934 .
SEXOM                -0.41901    0.07402  -5.661  0.0000000150512881 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(26.0779) family taken to be 1)

Null deviance: 533.262  on 27  degrees of freedom
Residual deviance:  28.179  on 13  degrees of freedom
AIC: 750.41

Number of Fisher Scoring iterations: 1

      Theta:  26.08
    Std. Err.:  6.93

2 x log-likelihood:  -718.409

```

Figura 123: Modelo Binomial Negativa Zona Rural Año 2019

La mayoría de variables son significativas con una confianza del 95 %, dado que los dos modelos presentan significancia en las variables se realiza la prueba de "vuong", cuyos resultados para el año 2019 se presentan a continuación.

Estadístico	Vuong z-statistic	HA	p-value
Raw	-4.786439	model2 /¿model1	0.00000084883
AIC-corrected	-4.786439	model2 /¿model1	0.00000084883
BIC-corrected	-4.786439	model2 /¿model1	0.00000084883

Tabla 28: Prueba de Vuong Zona Rural Año 2019

Lo anterior evidencia que el modelo 2, que corresponde a la distribución binomial negativa, es mejor que el modelo cuya distribución es Poisson.

Con base a lo anterior, se presentan los residuales del modelo binomial negativa.

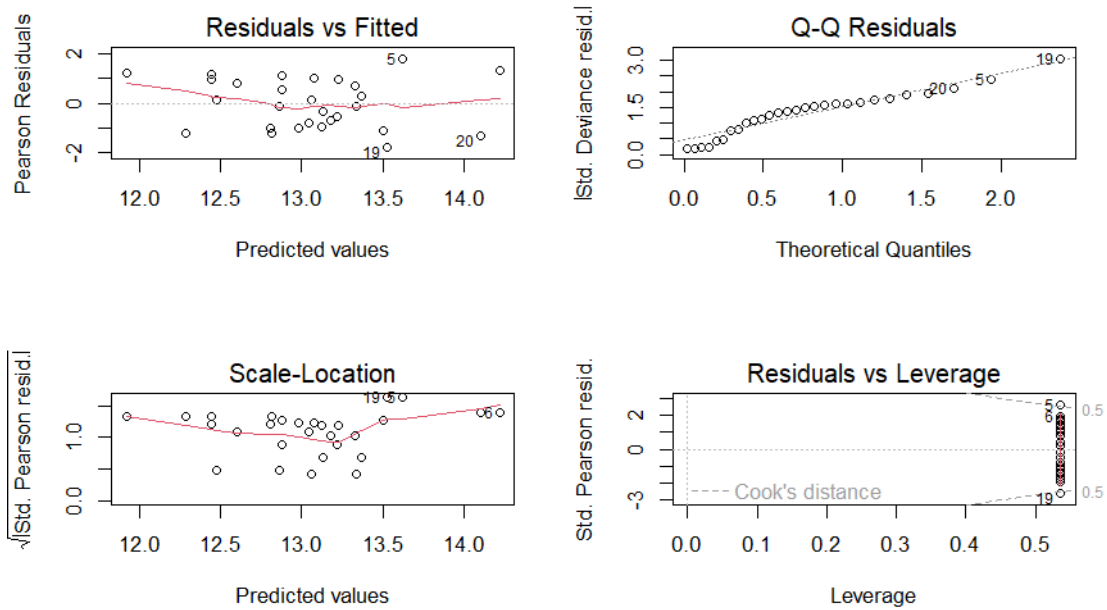


Figura 124: Residuales Modelo Binomial Negativa Zona Rural Año 2019

Como se observa en la figura 124, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 29 y figura 125, presentando puntos influyentes en las observaciones 1, 5, 14 y 19 , sin embargo, dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	-1.309221	0.5357236	0.2455999
5	1.845814	0.5357161	0.5378066
14	-1.163756	0.5357219	0.2017963
19	-2.257126	0.5357146	0.5378032

Tabla 29: Puntos Influyentes Zona Rural Año 2019

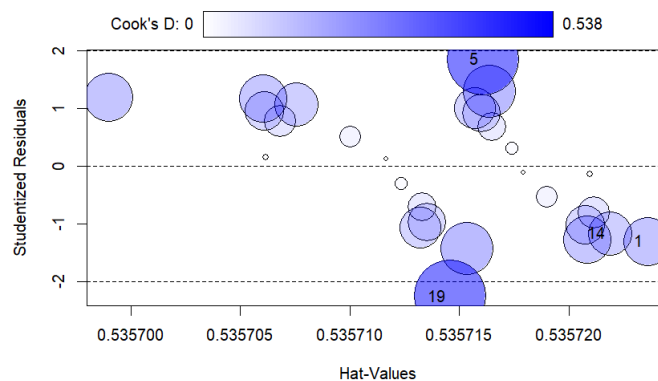


Figura 125: Puntos Influyentes Zona Rural Año 2019

Ahora bien, teniendo en cuenta el análisis de los residuales presentados se realiza la proyección de la frecuencia esperada, cuyos resultados para el año 2019 se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	9,020003081	9,425790751
Entre 1 y 5 Años	Ambos	6,936781943	7,249369892
Entre 6 y 11 Años	Ambos	3,895882257	4,036756237
Entre 12 y 17 Años	Femenino	5,224464236	5,384514165
Entre 12 y 17 Años	Masculino	3,646667206	3,541398907
Entre 18 y 28 Años	Femenino	5,678486998	4,197538736
Entre 18 y 28 Años	Masculino	1,786700459	2,760724299
Entre 29 y 44 Años	Femenino	7,860689701	6,254184629
Entre 28 y 44 Años	Masculino	3,056780012	4,113381808
Entre 45 y 49 Años	Ambos	7,394475031	7,3256377
Entre 50 y 54 Años	Ambos	9,834298165	9,736954563
Entre 55 y 59 Años	Ambos	13,30117539	13,16232009
Entre 60 y 64 Años	Ambos	18,57226669	18,43973378
Entre 65 y 69 Años	Ambos	24,92678514	25,01095698
Entre 70 y 74 Años	Ambos	31,65225404	32,22278406
Entre 75 y 79 Años	Ambos	37,27912646	38,58495814
Mayores de 80 Años	Ambos	340,72912392	43,57565831

Tabla 30: Frecuencias Observadas y Estimadas Zona Rural Año 2019

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Rural para el año 2019.

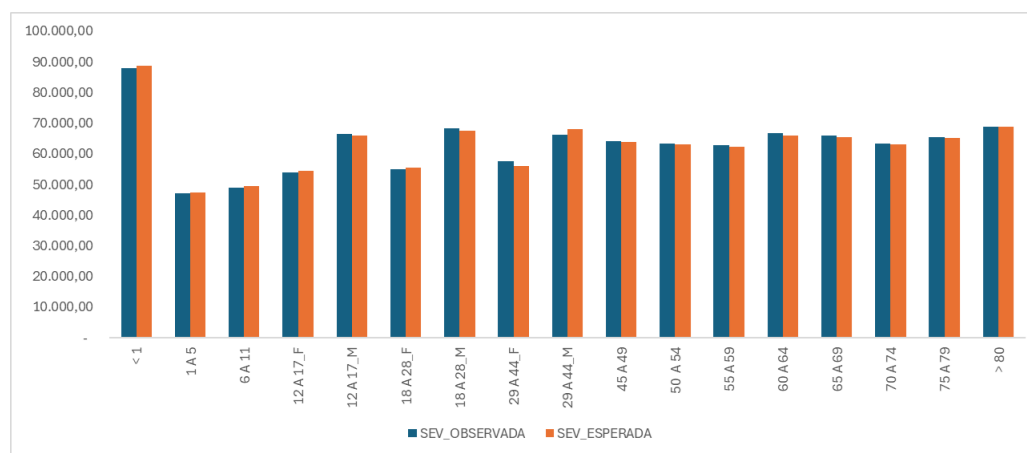


Figura 126: Frecuencia Observada Vs Frecuencia Estimada Zona Rural Año 2019

El modelo presenta un error porcentual absoluto medio (MAPE) del 10.02% lo cual presenta un ajuste aceptable; así mismo al calcular el pseudo R2 este es de 0.86

#### 4.1.6. Zona Rural Disperso

La distribución binomial negativa presenta un mejor ajuste para la Zona Rural Dispersa que la distribución poisson, la aplicación de este modelo se puede observar en el Anexo 2, ahora bien el modelo poisson presenta sobredispersión en los dos años, lo cual ratifica la necesidad de aplicar el modelo binomial negativa, lo cual se presenta a continuación.

```
Overdispersion test
data: fqqoi_RUD
z = 4.7784, p-value = 0.000008834
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
1662.642
```

Figura 127: Overdispersion Zona Rural Disperso Año 2018

```
Overdispersion test
data: fqqoi_RUD_2019
z = 4.7918, p-value = 0.000008265
alternative hypothesis: true alpha is greater than 0
sample estimates:
alpha
2196.103
```

Figura 128: Overdispersion Zona Rural Disperso Año 2019

- GLM Binomial Negativa

Dado al resultado de la prueba de sobredispersion, se considera que el modelo poisson no presenta un buen ajuste para ninguno de los dos años bajo estudio por lo que se decide realizar el modelo utilizando la distribución binominal negativa cuyo resultado para el año 2018 se presenta continuación.

```

Call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2018_RUD, link = log,
      init.theta = 22.21732817)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)    2.42082    0.15537  15.581 < 0.000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.16949    0.21226   0.798    0.424585
GRUPO_ETAREO_RUTA_AGREGADO11  0.50612    0.21227   2.384    0.017110 *
GRUPO_ETAREO_RUTA_AGREGADO12  0.80452    0.21227   3.790    0.000151 ***
GRUPO_ETAREO_RUTA_AGREGADO13  0.94989    0.21229   4.474    0.000007660679186 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.07267    0.21227   5.053    0.000000434292270 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.31015    0.21226  -1.461    0.143969
GRUPO_ETAREO_RUTA_AGREGADO3  -1.02133    0.21227  -4.811    0.000001498134293 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.94833    0.21227  -4.468    0.000007909893553 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -1.52922    0.21224  -7.205    0.0000000000000579 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -1.13320    0.21223  -5.339    0.000000093204712 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.69645    0.21225  -3.281    0.001034 **
GRUPO_ETAREO_RUTA_AGREGADO8  -0.43158    0.21225  -2.033    0.042022 *
GRUPO_ETAREO_RUTA_AGREGADO9  -0.06712    0.21225  -0.316    0.751836
SEXOM              -0.50974    0.08022  -6.354    0.000000000209435 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(22.2173) family taken to be 1)

Null deviance: 420.625  on 27  degrees of freedom
Residual deviance:  28.206  on 13  degrees of freedom
AIC: 601.5

Number of Fisher Scoring iterations: 1

      Theta:  22.22
   Std. Err.:  5.90

2 x log-likelihood:  -569.50
    
```

Figura 129: Modelo Binomial Negativa Zona Rural Disperso Año 2018

Las variables son significativas con una confianza del 95%, dado que los dos modelos presentan significancia en la variables se realiza la prueba de vuong, cuyos resultados se presentan continuación.

Estadístico	Vuong z-statistic	HA	p-value
Raw	-4.598753	model2 model1	0.0000021251
AIC-corrected	-4.598753	model2 model1	0.0000021251
BIC-corrected	-4.598753	model2 model1	0.0000021251

Tabla 31: Prueba de Vuong Zona Rural Disperso Año 2018

Lo anterior evidencia que el modelo 2, que corresponde a la distribución binomial negativa, es mejor que el modelo Poisson.

Con base a lo anterior, se presentan los residuales del modelo binomial negativa.

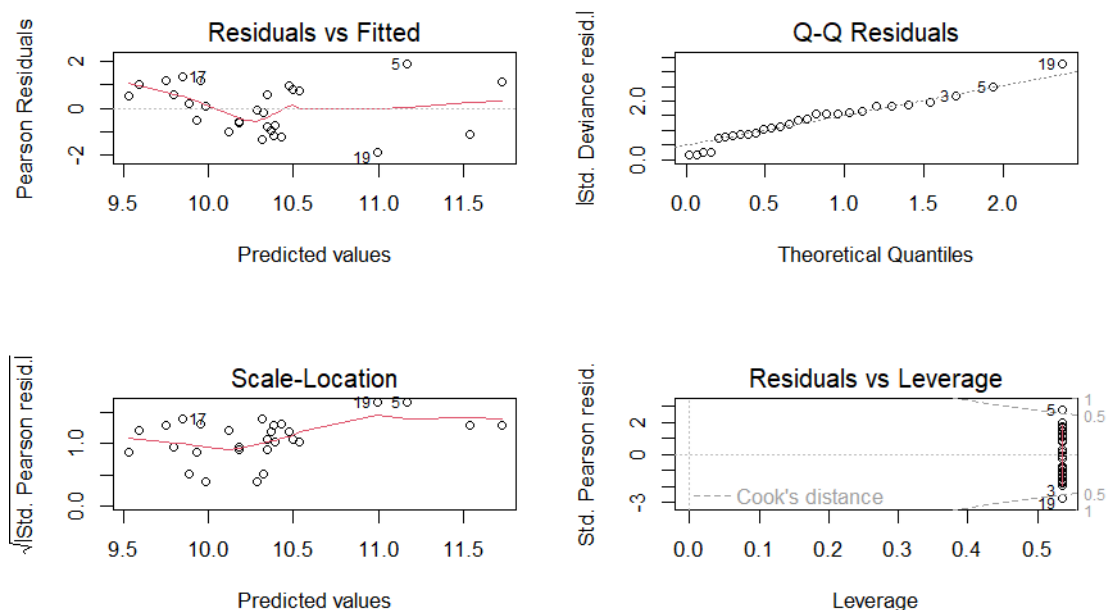


Figura 130: Residuales Modelo Binomial Negativa Zona Rural Disperso Año 2018

Como se observa en la figura número 130, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla número 59 y figura número 131, presentando puntos influyentes en las observaciones 1,5, 14 y 19 , sin embargo, dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	-1.039721	0.5358460	0.1652621
5	1.943328	0.5357451	0.5917593
14	-1.238085	0.5358493	0.2223812
19	-2.491373	0.5357176	0.5916939

Tabla 32: Puntos Influyentes Zona Rural Disperso Año 2018

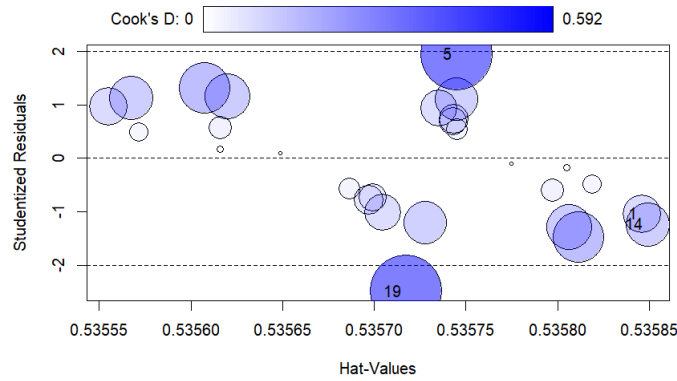


Figura 131: Puntos Influyentes Zona Rural Disperso Año 2018

Ahora bien, teniendo en cuenta el análisis de los residuales presentado se realiza la proyección de la frecuencia esperada, cuyos resultados se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	8,535550459	9,031448206
Entre 1 y 5 Años	Ambos	6,1869831	6,579798037
Entre 6 y 11 Años	Ambos	3,015951162	3,227237979
Entre 12 y 17 Años	Femenino	4,194285714	4,360059292
Entre 12 y 17 Años	Masculino	2,718490768	2,618878036
Entre 18 y 28 Años	Femenino	3,417011447	2,439023039
Entre 18 y 28 Años	Masculino	0,877538462	1,4650039
Entre 29 y 44 Años	Femenino	4,50150845	3,624144663
Entre 28 y 44 Años	Masculino	1,649839522	2,176849492
Entre 45 y 49 Años	Ambos	4,343062902	4,290076254
Entre 50 y 54 Años	Ambos	5,612626012	5,555369419
Entre 55 y 59 Años	Ambos	8,139974779	8,047413855
Entre 60 y 64 Años	Ambos	10,31414357	10,21447097
Entre 65 y 69 Años	Ambos	14,51830664	14,56487279
Entre 70 y 74 Años	Ambos	19,34109221	19,81406451
Entre 75 y 79 Años	Ambos	22,48454301	22,96858025
Mayores de 80 Años	Ambos	24,72402597	26,73644283

Tabla 33: Frecuencias Observadas y Estimadas Zona Rural Disperso Año 2018

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Rural Disperso.

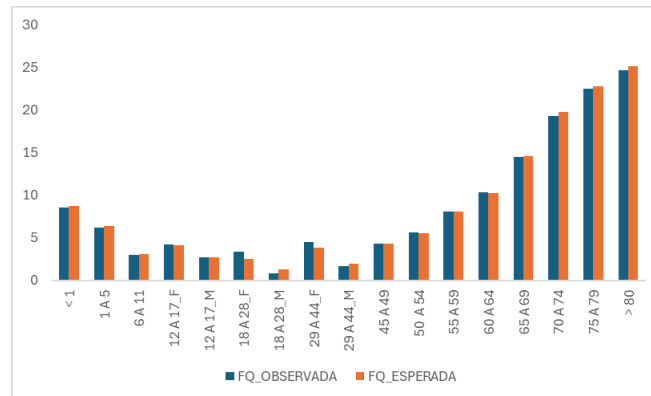


Figura 132: Frecuencia Observada Vs Frecuencia Estimada Zona Rural Dispersa Año 2018

El modelo para el año 2018 presenta un error porcentual absoluto medio (MAPE) del 11,24 %, lo cual presenta un ajuste aceptable; así mismo al calcular el pseudo R2 este es de 0.87

A continuación se presenta el análisis del modelo usando la distribución binomial negativa para el año 2019.

```

call:
glm.nb(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO +
      SEXO + offset(log(EXPUESTOS)), data = BASE_2019_RUD, link = log,
      init.theta = 20.42599864)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.43471    0.16202  15.027 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.36326    0.22135   1.641    0.100772
GRUPO_ETAREO_RUTA_AGREGADO11  0.73297    0.22135   3.311    0.000928 ***
GRUPO_ETAREO_RUTA_AGREGADO12  0.94061    0.22136   4.249    0.0000214488618 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.13896    0.22137   5.145    0.0000002675071 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.38286    0.22135   6.247    0.0000000004174 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.19660    0.22135  -0.888    0.374452
GRUPO_ETAREO_RUTA_AGREGADO3  -0.86874    0.22136  -3.925    0.0000868581373 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.77570    0.22135  -3.504    0.000458 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -1.46418    0.22133  -6.615    0.0000000000371 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -1.06686    0.22133  -4.820    0.0000014333580 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.55536    0.22134  -2.509    0.012106 *
GRUPO_ETAREO_RUTA_AGREGADO8  -0.27770    0.22134  -1.255    0.209611
GRUPO_ETAREO_RUTA_AGREGADO9   0.02829    0.22134   0.128    0.898312
SEXOM                    -0.52725    0.08365  -6.303    0.0000000002926 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(20.426) family taken to be 1)

Null deviance: 426.785  on 27  degrees of freedom
Residual deviance: 28.226  on 13  degrees of freedom
AIC: 616.48

Number of Fisher Scoring iterations: 1

      Theta: 20.43
    Std. Err.: 5.42

2 x log-likelihood: -584.482

```

Figura 133: Modelo Binomial Negativa Zona Rural Disperso 2019

Las mayoría de las variables son significativas con una confianza del 95 %, dado que los dos modelos presentan significancia en las variables, se realiza la prueba de vung, cuyos resultados se presentan continuación.

Estadístico	Vuong z-statistic	HA	p-value
Raw	5.363845	model2 model1	0.000000040734
AIC-corrected	5.363845	model2 model1	0.000000040734
BIC-corrected	5.363845	model2 model1	0.000000040734

Tabla 34: Prueba de Vuong Año 2019

Lo anterior evidencia que el modelo 2 que corresponde a la distribución binomial negativa es mejor que

el modelo cuya distribución es poisson.

Con base a lo anterior se presentan los residuales del modelo con distribución binomial negativa

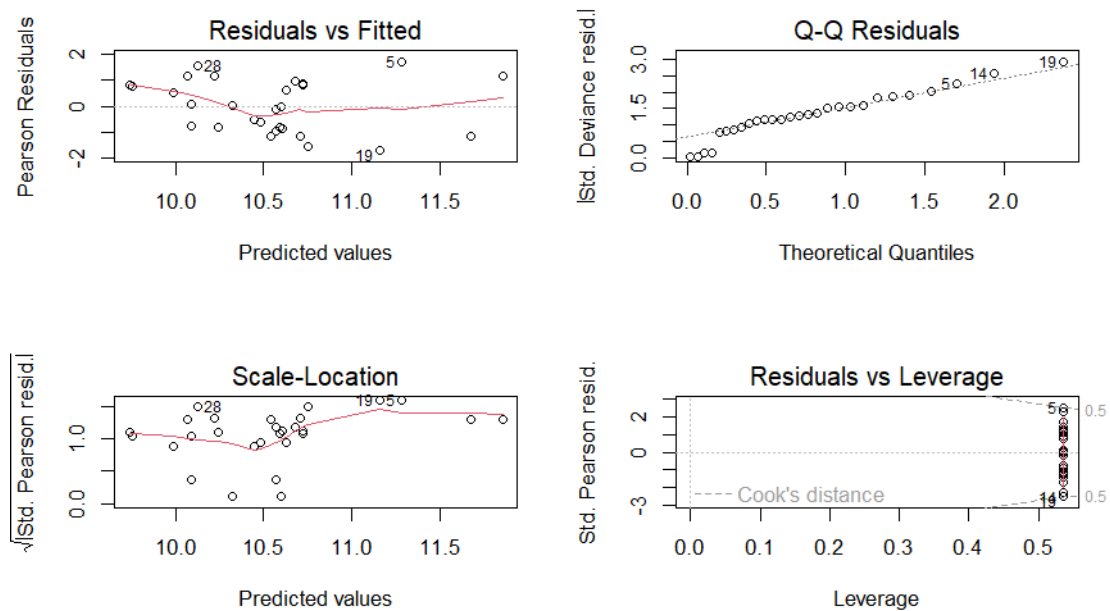


Figura 134: Residuales Modelo Binomial Negativa Zona Rural Dispersa Año 2019

Como se observa en la figura 134, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 35 y figura 135, presentando puntos influyentes en las observaciones 1,5,14 y 19 , sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	-0.8430216	0.5358113	0.1134981
5	1.7263746	0.5357340	0.4869108
14	-1.7982797	0.5358018	0.3916610
19	-2.1233309	0.5357177	0.4868789

Tabla 35: Puntos Influyentes Zona Rural Disperso Año 2019

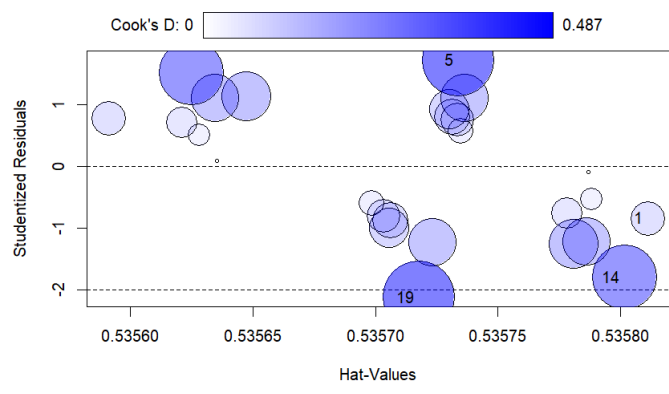


Figura 135: Puntos Influyentes Zona Rural Disperso Año 2019

Ahora bien, teniendo en cuenta el análisis de los residuales presentados, se realiza la proyección de la frecuencia esperada, cuyos resultados para el año 2019 se presentan a continuación.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	8,637406885	8,803944339
Entre 1 y 5 Años	Ambos	6,956050758	7,200371707
Entre 6 y 11 Años	Ambos	3,557278794	3,683416372
Entre 12 y 17 Años	Femenino	5,140582839	4,91416068
Entre 12 y 17 Años	Masculino	3,168109575	3,320438418
Entre 18 y 28 Años	Femenino	3,640520273	2,826357144
Entre 18 y 28 Años	Masculino	0,966847753	1,337997329
Entre 29 y 44 Años	Femenino	4,926731236	4,187976941
Entre 28 y 44 Años	Masculino	1,727618035	2,083303275
Entre 45 y 49 Años	Ambos	4,985177216	4,89420852
Entre 50 y 54 Años	Ambos	6,584231295	6,48966563
Entre 55 y 59 Años	Ambos	8,974396336	8,850698136
Entre 60 y 64 Años	Ambos	12,54008711	12,45735484
Entre 65 y 69 Años	Ambos	18,26440062	18,40413912
Entre 70 y 74 Años	Ambos	22,41635841	22,68068543
Entre 75 y 79 Años	Ambos	26,88435829	27,77426685
Mayores de 80 Años	Ambos	32,85509473	34,05061419

Tabla 36: Frecuencias Observadas y Estimadas Zona Rural Disperso Año 2019

A continuación se presenta gráficamente la frecuencia observada y la frecuencia estimada de la Zona Rural Disperso para el año 2019.

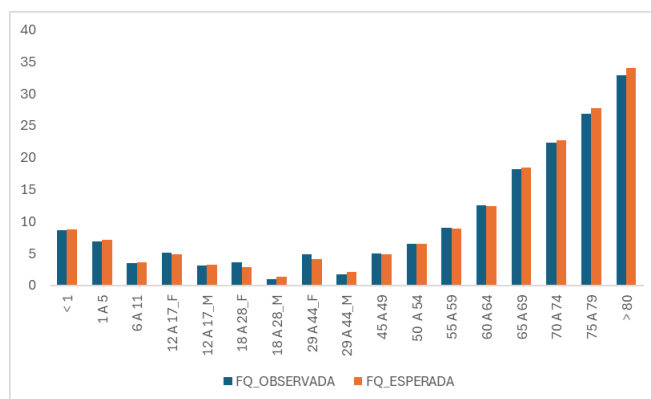


Figura 136: Frecuencia Observada Vs Frecuencia Estimada Zona Rural Disperso Año 2019

El modelo para el año 2019 presenta un error porcentual absoluto medio (MAPE) del 7.57% lo cual presenta un ajuste aceptable; así mismo al calcular el pseudo R2 este es de 0.84

## 4.2. Modelos de Severidad

### 4.2.1. Zona Alejada

- GLM Gamma

A continuación se presentan los resultados de la aplicación del modelo Gamma, para modelar la severidad de la zona alejada para el año 2018

Call:

```
glm(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO,
     family = Gamma(link = "log"), data = BASE_2018_A)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	11.538454	0.105350	109.525	< 2e-16	***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.134878	0.143936	-0.937	0.3658	
GRUPO_ETAREO_RUTA_AGREGADO11	-0.122293	0.143936	-0.850	0.4109	
GRUPO_ETAREO_RUTA_AGREGADO12	0.205575	0.143936	1.428	0.1768	
GRUPO_ETAREO_RUTA_AGREGADO13	0.051138	0.143936	0.355	0.7281	
GRUPO_ETAREO_RUTA_AGREGADO14	0.003675	0.143936	0.026	0.9800	
GRUPO_ETAREO_RUTA_AGREGADO2	-0.242860	0.143936	-1.687	0.1154	
GRUPO_ETAREO_RUTA_AGREGADO3	-0.279005	0.143936	-1.938	0.0746	.
GRUPO_ETAREO_RUTA_AGREGADO4	-0.249730	0.143936	-1.735	0.1064	
GRUPO_ETAREO_RUTA_AGREGADO5	-0.068853	0.143936	-0.478	0.6403	
GRUPO_ETAREO_RUTA_AGREGADO6	-0.127484	0.143936	-0.886	0.3919	
GRUPO_ETAREO_RUTA_AGREGADO7	-0.172620	0.143936	-1.199	0.2518	
GRUPO_ETAREO_RUTA_AGREGADO8	-0.108090	0.143936	-0.751	0.4661	
GRUPO_ETAREO_RUTA_AGREGADO9	-0.024006	0.143936	-0.167	0.8701	
SEXOM	0.363412	0.054403	6.680	1.52e-05	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.02071743)

Null deviance: 1.71951 on 27 degrees of freedom  
 Residual deviance: 0.27195 on 13 degrees of freedom  
 AIC: 632.72

Number of Fisher Scoring iterations: 4

Figura 137: Modelo Gamma para Severidad Zona Alejada Año 2018

En este modelo la variable sexo es significativa pero las variables de grupo etario no, sin embargo dada la finalidad el presente trabajo de estimar la prima pura para cada grupo etario no es dable pensar en no tenerlas en cuenta, pues es como si se el valor de las atenciones fuera el mismo para todos los grupos lo cual se sabe que no es así, pues las tecnologías y servicios de salud son usadas con mayor intensidad en los grupos de mayor edad.

A continuación se presenta los residuales.

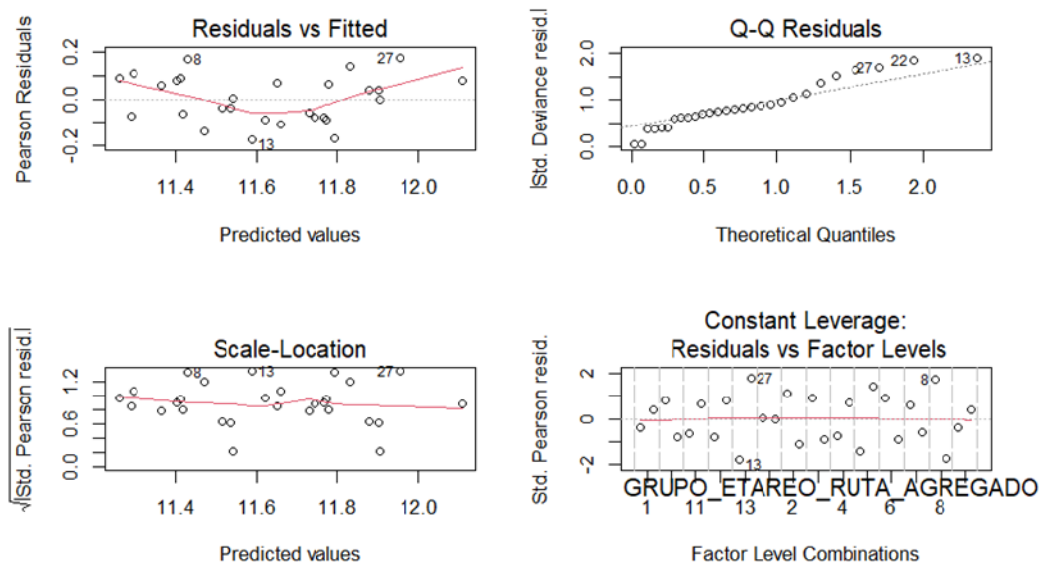


Figura 138: Residuales Modelo Gamma para Severidad Zona Alejada Año 2018

Como se observa en la figura 138, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 37 y figura 139, presentando puntos influyentes en las observaciones 4,5,13,22 y 27 sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
4	-0.7117511	0.5357143	0.03993432
5	-1.5511112	0.5357143	0.15934093
13	-2.0629237	0.5357143	0.24448520
22	-1.9969270	0.5357143	0.23357455
27	1.8798323	0.5357143	0.24448464

Tabla 37: Puntos Influyentes Zona Alejada Año 2018

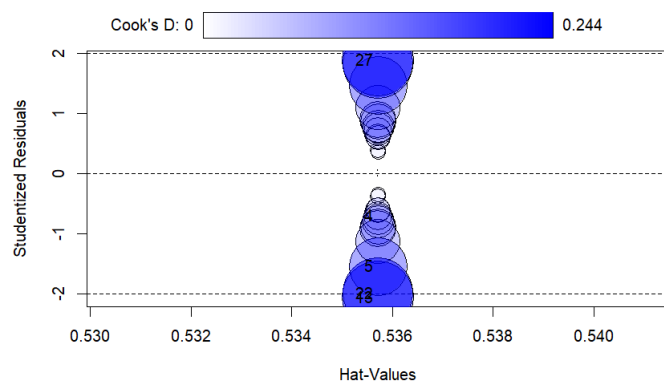


Figura 139: Puntos Influyentes Zona Alejada Año 2018

A continuación se presentan los resultados para el año 2018.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	127.095,44	126.075,01
Entre 1 y 5 Años	Ambos	96.493,35	98.869,42
Entre 6 y 11 Años	Ambos	93.487,05	95.419,82
Entre 12 y 17 Años	Femenino	74.268,17	79.915,41
Entre 12 y 17 Años	Masculino	123.058,62	114.936,61
Entre 18 y 28 Años	Femenino	82.243,14	95.760,15
Entre 18 y 28 Años	Masculino	157.165,51	137.724,98
Entre 29 y 44 Años	Femenino	98.169,09	90.307,01
Entre 28 y 44 Años	Masculino	118.574,65	129.882,11
Entre 45 y 49 Años	Ambos	103.155,69	103.821,86
Entre 50 y 54 Años	Ambos	108.732,10	110.814,73
Entre 55 y 59 Años	Ambos	121.527,53	120.920,72
Entre 60 y 64 Años	Ambos	106.853,72	107.728,22
Entre 65 y 69 Años	Ambos	110.161,65	109.372,33
Entre 70 y 74 Años	Ambos	153.517,36	152.066,23
Entre 75 y 79 Años	Ambos	132.522,21	129.983,42
Mayores de 80 Años	Ambos	120.616,36	120.593,66

Tabla 38: Severidad Observada y Estimada Zona Alejada Año 2018

A continuación se presenta gráficamente la severidad observada y la severidad estimada para el año 2018.

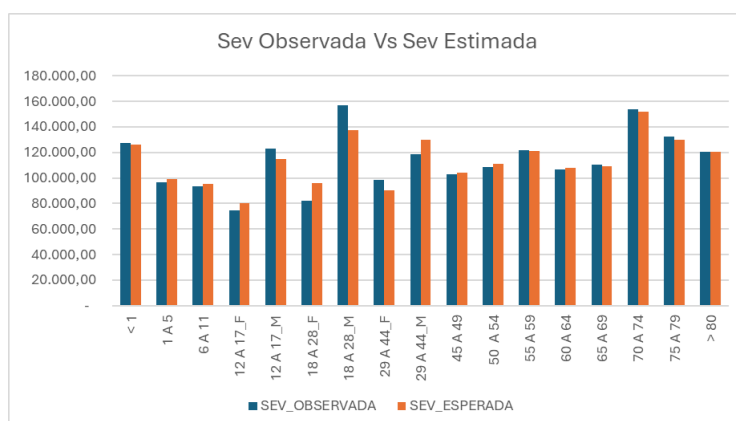


Figura 140: Severidad Observada Vs Severidad Estimada Zona Alejada Año 2018

El modelo presenta un error porcentual absoluto medio (MAPE) del 4.81 % lo cual se presenta un ajuste aceptable y un pseudo R2 de 0.84, no se realiza el modelo lognormal para el año 2018.

A continuación se presentan los resultados de la aplicación del modelo Gamma, para modelar la severidad de la zona alejada para el año 2019.

```

call:
glm(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO,
     family = Gamma(link = "log"), data = BASE_2019_A)

Coefficients:
                Estimate Std. Error t value      Pr(>|t|)
(Intercept)      10.97446    0.08265 132.784 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.05038    0.11292   0.446    0.66282
GRUPO_ETAREO_RUTA_AGREGADO11  0.12055    0.11292   1.068    0.30512
GRUPO_ETAREO_RUTA_AGREGADO12  0.04164    0.11292   0.369    0.71827
GRUPO_ETAREO_RUTA_AGREGADO13  0.27800    0.11292   2.462    0.02857 *
GRUPO_ETAREO_RUTA_AGREGADO14 -0.05393    0.11292  -0.478    0.64084
GRUPO_ETAREO_RUTA_AGREGADO2   0.07012    0.11292   0.621    0.54536
GRUPO_ETAREO_RUTA_AGREGADO3   0.06455    0.11292   0.572    0.57734
GRUPO_ETAREO_RUTA_AGREGADO4  -0.05580    0.11292  -0.494    0.62944
GRUPO_ETAREO_RUTA_AGREGADO5   0.09329    0.11292   0.826    0.42364
GRUPO_ETAREO_RUTA_AGREGADO6   0.19578    0.11292   1.734    0.10659
GRUPO_ETAREO_RUTA_AGREGADO7   0.10266    0.11292   0.909    0.37984
GRUPO_ETAREO_RUTA_AGREGADO8   0.20831    0.11292   1.845    0.08797 .
GRUPO_ETAREO_RUTA_AGREGADO9   0.11447    0.11292   1.014    0.32922
SEXOM                    0.15621    0.04268   3.660    0.00288 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.01275101)

Null deviance: 0.58973  on 27  degrees of freedom
Residual deviance: 0.16694  on 13  degrees of freedom
AIC: 591.76

Number of Fisher Scoring iterations: 4

```

Figura 141: Modelo Gamma para Severidad Zona Alejada Año 2019

En este modelo la variable sexo es significativa pero las variables de grupo etario no con excepción del grupo etario 13, sin embargo dada la finalidad el presente trabajo de estimar la prima pura para cada grupo etario no es dable pensar en no tenerlas en cuenta, pues es como si se el valor de las atenciones fuera el mismo para todos los grupos lo cual se sabe que no es así, pues las tecnologías y servicios de salud son usadas con mayor intensidad en los grupos de mayor edad.

A continuación se presenta los residuales.

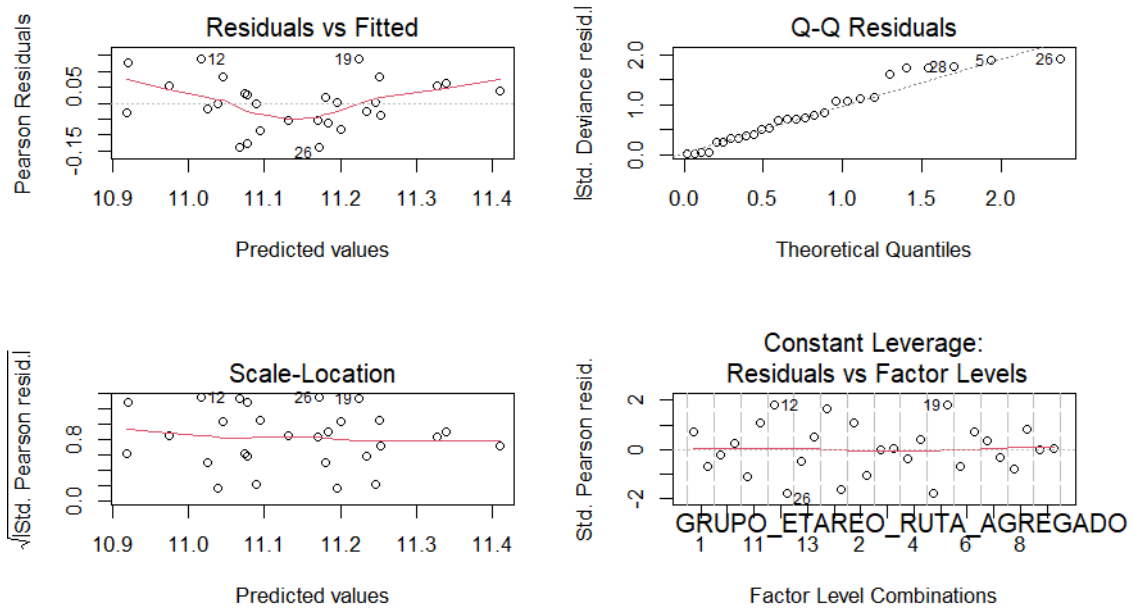


Figura 142: Residuales Modelo Gamma para Severidad Zona Alejada Año 2019

Como se observa en la figura 142, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 39 y figura 143, presentando puntos influyentes en las observaciones 4,5, 12 y 26 sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
4	-0.3749697	0.5357143	0.01155277
5	-2.0679366	0.5357143	0.24906367
12	1.9470141	0.5357143	0.25451972
26	-2.1001570	0.5357143	0.25452019

Tabla 39: Puntos Influyentes Zona Alejada Año 2019

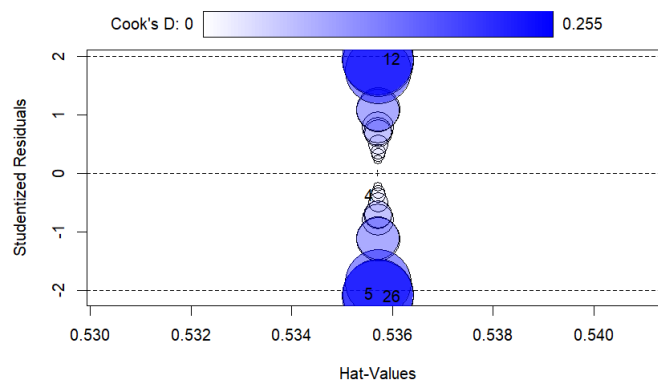


Figura 143: Puntos Influyentes Zona Alejada Año 2019

A continuación se presentan los resultados para el año 2019.

Descripción Grupo Etario	Sexo	Fq Observada	Fq Estimada
Menor de un año	Ambos	63.074,99	63.486,54
Entre 1 y 5 Años	Ambos	67.434,13	68.193,25
Entre 6 y 11 Años	Ambos	67.646,79	67.633,62
Entre 12 y 17 Años	Femenino	53.551,03	55.196,90
Entre 12 y 17 Años	Masculino	66.453,15	64.529,02
Entre 18 y 28 Años	Femenino	55.200,44	64.071,05
Entre 18 y 28 Años	Masculino	85.273,90	74.903,52
Entre 29 y 44 Años	Femenino	67.127,31	70.986,19
Entre 28 y 44 Años	Masculino	87.499,11	82.987,80
Entre 45 y 49 Años	Ambos	69.775,91	69.805,74
Entre 50 y 54 Años	Ambos	77.628,69	77.564,71
Entre 55 y 59 Años	Ambos	70.710,01	70.702,61
Entre 60 y 64 Años	Ambos	66.204,11	66.196,33
Entre 65 y 69 Años	Ambos	70.872,07	70.925,43
Entre 70 y 74 Años	Ambos	65.632,49	65.533,24
Entre 75 y 79 Años	Ambos	83.391,03	83.285,79
Mayores de 80 Años	Ambos	59.977,20	59.045,72

Tabla 40: Severidad Observada y Estimada Zona Alejada Año 2019

A continuación se presenta gráficamente la severidad observada y la severidad estimada.

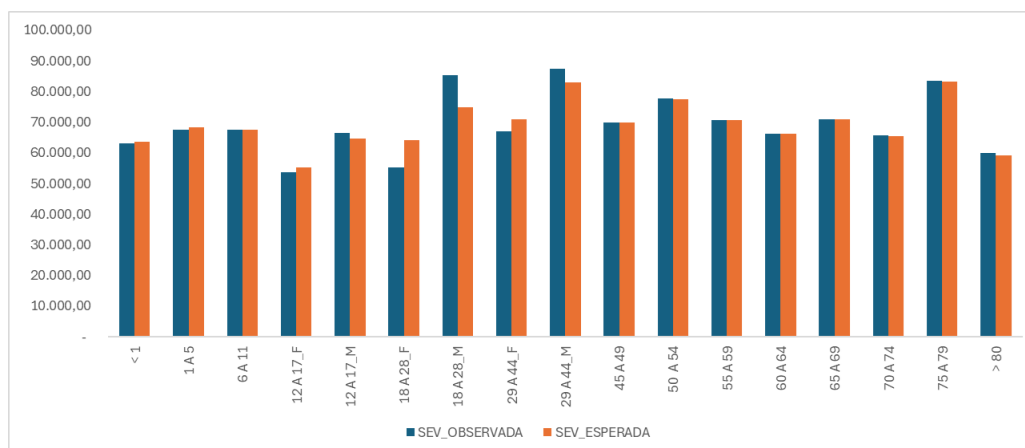


Figura 144: Severidad Observada Vs Severidad Estimada Zona Alejada Año 2019

El modelo presenta un error porcentual absoluto medio (MAPE) del 2.87 % lo cual se presenta un ajuste aceptable y un pseudo R2 de 0.71, teniendo en cuenta lo presentado en cuanto a la severidad empirica no se realiza el modelo lognormal dado que los datos no presentan un ajuste usando dicha distribución.

#### 4.2.2. Zona Bogotá

Tanto para el año 2018 como para el año 2019, el modelo gamma presenta un buen ajuste sin embargo el modelo lognormal se ajusta mejor dados los analisis de MAPE, por lo anterior se presenta a continuación el desarrollo del modelo lognormal, el desarrollo del modelo gamma se presnta en el Anexo 2.

##### ■ GLM LogNormal

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la zona Bogotá para el año 2018

```

GAMLSS-RS iteration 1: Global Deviance = 508.9953
GAMLSS-RS iteration 2: Global Deviance = 508.9953
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2018_BOG)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.422062	0.002684	902.27	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.049426	0.003741	-13.21	0.000000016434 ***
GRUPO_ETAREO_RUTA_AGREGADO11	-0.050307	0.003743	-13.44	0.000000013533 ***
GRUPO_ETAREO_RUTA_AGREGADO12	-0.049234	0.003741	-13.16	0.000000017154 ***
GRUPO_ETAREO_RUTA_AGREGADO13	-0.046495	0.003735	-12.45	0.000000032116 ***
GRUPO_ETAREO_RUTA_AGREGADO14	-0.040243	0.003723	-10.81	0.000000153531 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.061771	0.003765	-16.41	0.000000001390 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.069231	0.003780	-18.31	0.000000000389 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.056744	0.003755	-15.11	0.000000003574 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.063323	0.003768	-16.80	0.000000001054 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.052042	0.003746	-13.89	0.000000009312 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.049598	0.003741	-13.26	0.000000015819 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.050620	0.003743	-13.52	0.000000012640 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.047777	0.003738	-12.78	0.000000023849 ***
SEXOM	0.022103	0.001448	15.27	0.000000003180 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
sigma link function: log
sigma coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.1805	0.1336	-23.8	0.0000000000181 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance: 508.9953
      AIC: 540.9953
      SBC: 562.3106
*****

```

Figura 145: Modelo LogNormal para Severidad Zona Bogotá Año 2018

El modelo LogNormal presenta todas las variables significativas al 95 % por lo cual se presenta a continuación los residuales

A continuación se presentan los residuales

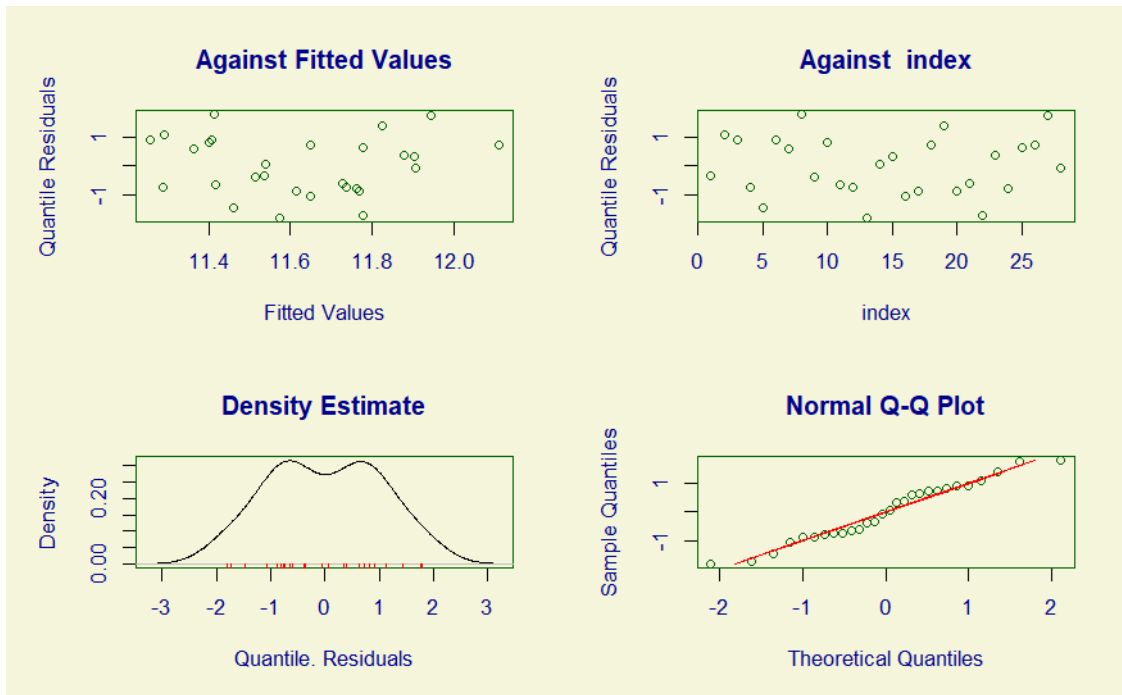


Figura 146: Residuales Modelo LogNormal para Severidad Zona Bogotá Año 2018

Como se observa en la figura 146, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 41 y figura 147, presentando puntos influyentes en las observaciones 2,15,16 y 28 , sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
2	2.1668669	0.5245203	0.26887622
15	-1.9214714	0.5495561	0.24877776
16	-2.1668669	0.5450810	0.29204449
28	-0.7602403	0.5465783	0.04800579

Tabla 41: Puntos Influyentes LogNormal Zona Bogotá Año 2018

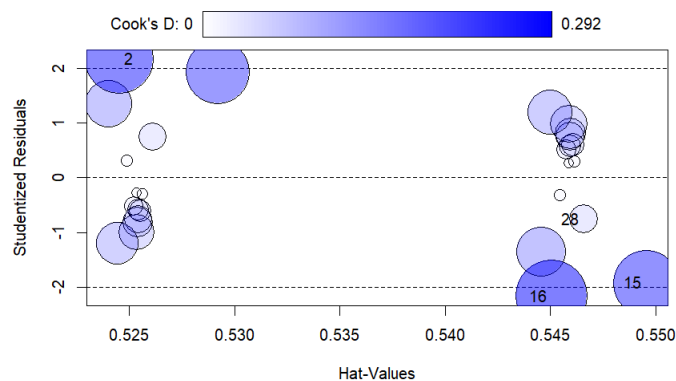


Figura 147: Puntos Influyentes Zona Bogotá Año 2018

En la tabla a continuación se presenta los valores observados y los valores estimados aplicando el modelo LogNormal para la zona Bogotá, para el año 2018

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	89.127,36	89.706,31
Entre 1 y 5 Años	Ambos	45.003,01	45.285,58
Entre 6 y 11 Años	Ambos	41.624,43	41.842,28
Entre 12 y 17 Años	Femenino	42.665,36	42.083,91
Entre 12 y 17 Años	Masculino	52.678,03	53.389,84
Entre 18 y 28 Años	Femenino	37.345,39	39.245,94
Entre 18 y 28 Años	Masculino	52.185,05	49.711,81
Entre 29 y 44 Años	Femenino	43.256,10	44.249,63
Entre 28 y 44 Años	Masculino	57.462,68	56.200,37
Entre 45 y 49 Años	Ambos	51.403,98	51.304,78
Entre 50 y 54 Años	Ambos	50.632,97	50.604,34
Entre 55 y 59 Años	Ambos	52.128,39	52.077,71
Entre 60 y 64 Años	Ambos	51.046,17	51.012,78
Entre 65 y 69 Años	Ambos	50.506,69	50.477,82
Entre 70 y 74 Años	Ambos	50.982,54	50.988,69
Entre 75 y 79 Años	Ambos	52.304,90	52.334,61
Mayores de 80 Años	Ambos	55.554,41	55.221,95

Tabla 42: Severidad Observada y Estimada Modelo LogNormal Zona Bogotá Año 2018

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Bogotá, para el año 2018.

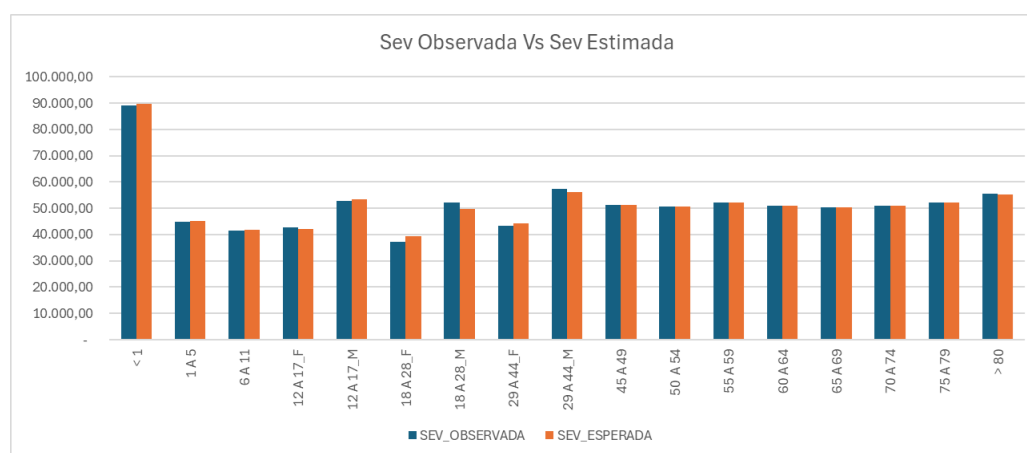


Figura 148: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Bogotá Año 2018

El MAPE presentado es de 1.17% siendo mejor que el presentado por el modelo Gamma, y un Pseudo R2 de 0.96 similar al presentado por la distribución Gamma.

Ahora la elección del mejor modelo se hará a partir del análisis del MAPE identificando aquel que tenga el menor que en este caso es el modelo LogNormal, dado que los dos modelos ajustan de manera aceptable.

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la zona Bogotá para el año 2019

```

GAMLSS-RS iteration 1: Global Deviance = 501.414
GAMLSS-RS iteration 2: Global Deviance = 501.414
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2019_BOG)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.420859	0.002442	991.34	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.052080	0.003408	-15.28	0.000000003142 ***
GRUPO_ETAREO_RUTA_AGREGADO11	-0.053954	0.003411	-15.81	0.000000002118 ***
GRUPO_ETAREO_RUTA_AGREGADO12	-0.054384	0.003412	-15.94	0.000000001939 ***
GRUPO_ETAREO_RUTA_AGREGADO13	-0.051706	0.003407	-15.17	0.000000003405 ***
GRUPO_ETAREO_RUTA_AGREGADO14	-0.048830	0.003402	-14.35	0.000000006432 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.065994	0.003433	-19.22	0.000000000222 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.064205	0.003430	-18.72	0.000000000302 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.055952	0.003415	-16.38	0.000000001411 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.058917	0.003420	-17.23	0.000000000792 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.049537	0.003404	-14.55	0.000000005484 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.049150	0.003403	-14.44	0.000000005983 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.050571	0.003405	-14.85	0.000000004359 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.052624	0.003409	-15.44	0.000000002799 ***
SEXOM	0.020400	0.001319	15.46	0.000000002742 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
sigma link function: log
sigma coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.2771	0.1336	-24.52	0.0000000000128 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance:    501.414
                   AIC:    533.414
                   SBC:    554.7292
*****

```

Figura 149: Modelo LogNormal para Severidad Zona Bogotá Año 2019

El modelo LogNormal presenta todas las variables significativas al 95 % por lo cual se presenta a continuación los residuales

A continuación se presentan los residuales

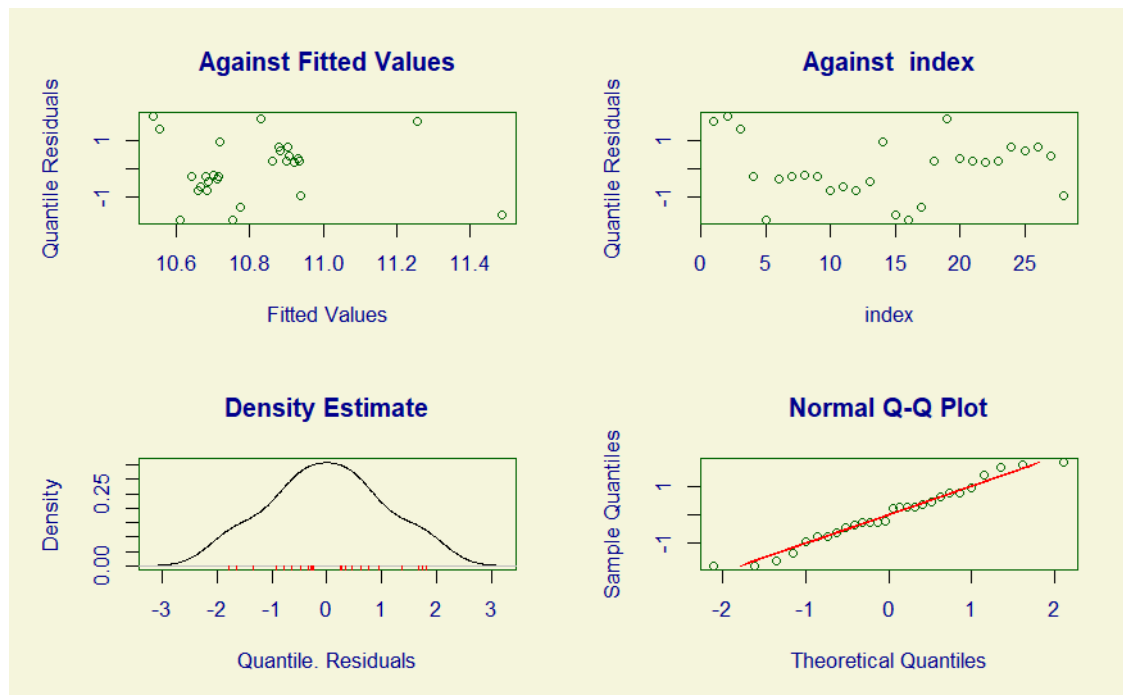


Figura 150: Residuales Modelo LogNormal para Severidad Zona Bogotá Año 2019

Como se observa en la figura 150, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 43 y figura 151, presentando puntos influyentes en las observaciones 2,15, 16, 19 y 28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
2	1.9915661	0.5251205	0.23807393
15	-1.8014385	0.5488900	0.22447152
16	-1.9915661	0.5441059	0.25695422
19	1.9450430	0.5445892	0.24841701
28	-0.9285385	0.5452900	0.06966752

Tabla 43: Puntos Influyentes LogNormal Zona Bogotá Año 2019

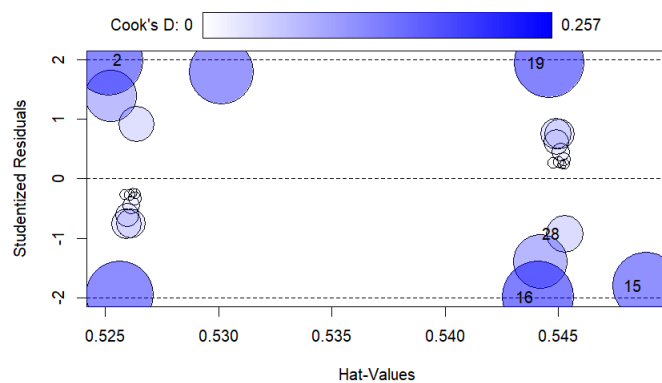


Figura 151: Puntos Influyentes LogNormal Zona Bogotá Año 2019

En la tabla a continuación se presenta los valores observados y los valores estimados aplicando el modelo LogNormal para la zona Bogotá, para el año 2019

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	87.103,19	87.637,88
Entre 1 y 5 Años	Ambos	42.094,76	42.339,74
Entre 6 y 11 Años	Ambos	42.944,88	43.147,91
Entre 12 y 17 Años	Femenino	41.457,88	41.900,42
Entre 12 y 17 Años	Masculino	52.722,94	52.177,29
Entre 18 y 28 Años	Femenino	37.956,76	40.600,49
Entre 18 y 28 Años	Masculino	53.971,45	50.525,71
Entre 29 y 44 Años	Femenino	44.290,69	44.870,94
Entre 28 y 44 Años	Masculino	56.673,49	55.955,32
Entre 45 y 49 Años	Ambos	50.406,70	50.376,74
Entre 50 y 54 Años	Ambos	49.512,33	49.494,69
Entre 55 y 59 Años	Ambos	48.313,30	48.302,98
Entre 60 y 64 Años	Ambos	48.489,93	48.477,99
Entre 65 y 69 Años	Ambos	47.431,01	47.439,35
Entre 70 y 74 Años	Ambos	47.167,59	47.183,80
Entre 75 y 79 Años	Ambos	48.365,48	48.405,67
Mayores de 80 Años	Ambos	49.621,38	49.284,32

Tabla 44: Severidad Observada y Estimada Modelo LogNormal Zona Bogotá Año 2019

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Bogotá, para el año 2019.

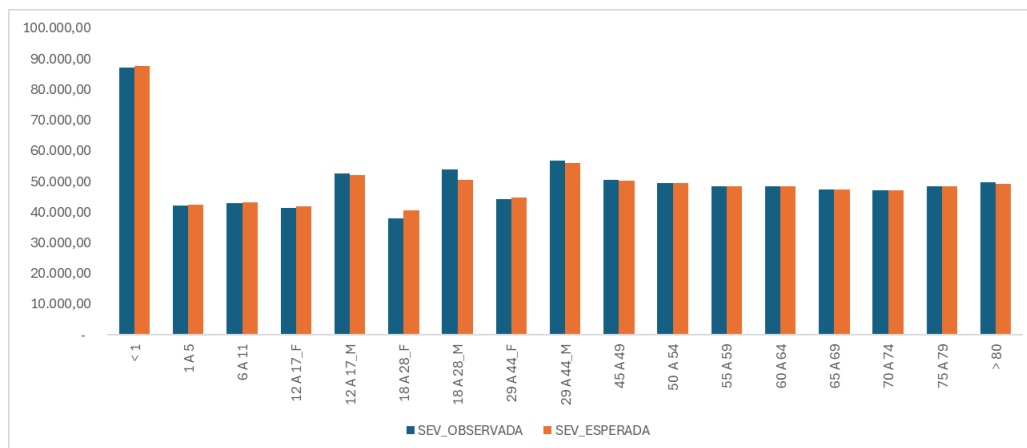


Figura 152: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Bogotá Año 2019

El MAPE presentado es de 1.21 % siendo mejor que el presentado por el modelo Gamma, y un Pseudo R2 de 0.96 similar al presentado por la distribución Gamma.

Ahora la elección del mejor modelo se hará a partir del análisis del MAPE identificando aquel que tenga el menor que en este caso es el modelo LogNormal, dado que los dos modelos ajustan de manera aceptable.

#### 4.2.3. Zona Normal Mayor

Tanto para el año 2018 y 2019 el modelo seleccionado para modelar la severidad es el lognormal, para el año 2018 el modelo gamma presenta un buen ajuste y su desarrollo se presenta en el Anexo 2; para el año 2019 no se desarrolla el modelo gamma por no presentar un buen ajuste en las distribuciones empíricas realizadas en el apartado anterior.

- GLM LogNormal

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la Zona Normal Mayor para el año 2018

```

GAMLSS-RS iteration 1: Global Deviance = 491.0079
GAMLSS-RS iteration 2: Global Deviance = 491.0079
*****
Family: c("LOGNO", "Log Normal")

Call:  gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
            SEXO, family = LOGNO(mu.link = "log"), data = BASE_2018_NM)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.405642	0.002037	1181.153	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.036144	0.002823	-12.805	0.000000023389 ***
GRUPO_ETAREO_RUTA_AGREGADO11	-0.036729	0.002824	-13.008	0.000000019590 ***
GRUPO_ETAREO_RUTA_AGREGADO12	-0.035990	0.002822	-12.751	0.000000024515 ***
GRUPO_ETAREO_RUTA_AGREGADO13	-0.033213	0.002818	-11.784	0.000000059169 ***
GRUPO_ETAREO_RUTA_AGREGADO14	-0.028097	0.002811	-9.995	0.000000359918 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.051190	0.002845	-17.993	0.000000000477 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.043320	0.002833	-15.290	0.000000003121 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.030953	0.002815	-10.995	0.000000127292 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.040760	0.002829	-14.406	0.000000006165 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.035065	0.002821	-12.430	0.000000032648 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.032660	0.002818	-11.592	0.000000071059 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.033362	0.002819	-11.836	0.000000056341 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.034658	0.002821	-12.288	0.000000037117 ***
SEXOM	0.019157	0.001083	17.682	0.00000000585 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.4732	0.1336	-25.99	0.000000000000643 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
Residual Deg. of Freedom: 12
at cycle: 2

Global Deviance: 491.0079
AIC: 523.0079
SBC: 544.3232
*****

```

Figura 153: Modelo LogNormal para Severidad Zona Normal Mayor Año 2018

El modelo LogNormal presenta todas las variables significativas al 95 % por lo cual se presenta a continuación los residuales

A continuación se presentan los residuales

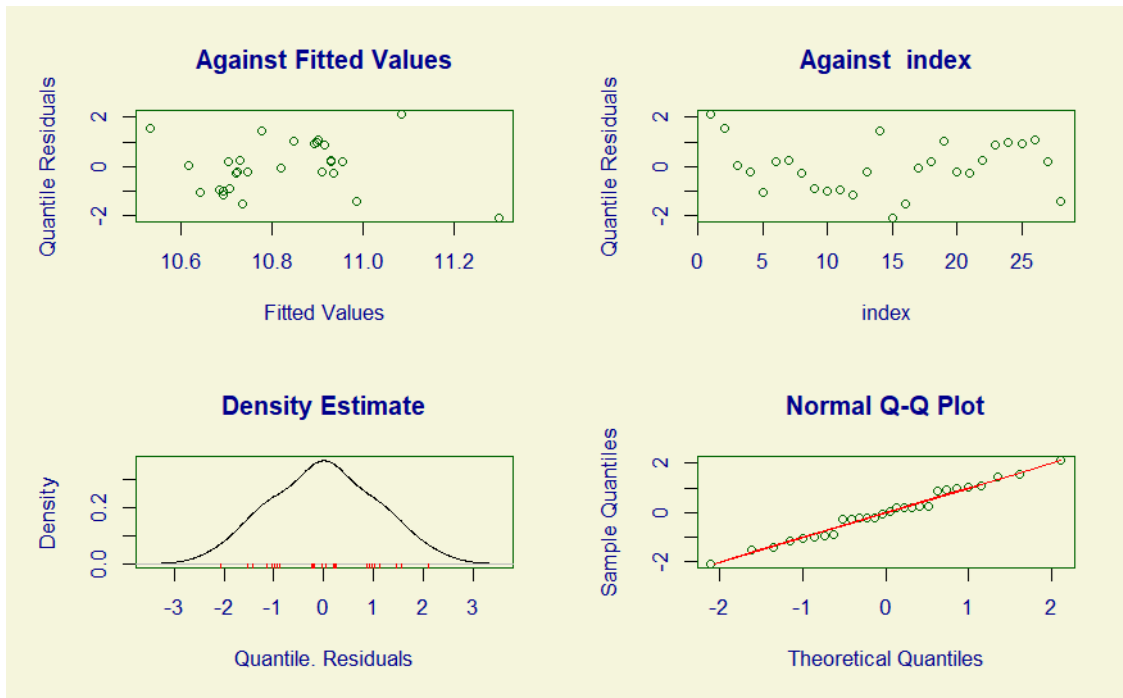


Figura 154: Residuales Modelo LogNormal para Severidad Zona Normal Mayor Año 2018

Como se observa en la figura 158, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 45 y figura 155, presentando puntos influyentes en las observaciones 1, 15 y 28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	2.469445	0.5293517	0.3284462
15	-2.469445	0.5470427	0.3526796
28	-1.508150	0.5449954	0.1654081

Tabla 45: Puntos Influyentes LogNormal Zona Normal Mayor Año 2018

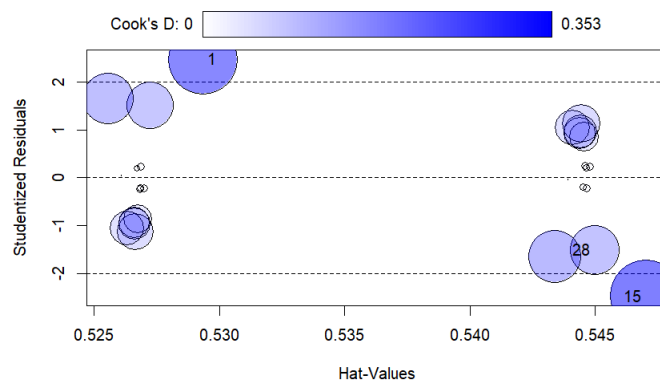


Figura 155: Puntos Influyentes LogNormal Zona Normal Mayor Año 2018

En la tabla a continuación se presenta los valores observados y los valores estimados aplicando el modelo LogNormal para la Zona Normal Mayor para el año 2018

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	72.694,02	72.972,60
Entre 1 y 5 Años	Ambos	41.636,72	41.796,26
Entre 6 y 11 Años	Ambos	45.493,11	45.499,66
Entre 12 y 17 Años	Femenino	46.186,65	46.521,37
Entre 12 y 17 Años	Masculino	57.677,67	57.270,53
Entre 18 y 28 Años	Femenino	40.554,01	41.889,09
Entre 18 y 28 Años	Masculino	53.124,99	51.463,41
Entre 29 y 44 Años	Femenino	44.794,79	44.514,25
Entre 28 y 44 Años	Masculino	54.416,50	54.752,92
Entre 45 y 49 Años	Ambos	50.829,78	50.858,00
Entre 50 y 54 Años	Ambos	50.359,71	50.338,47
Entre 55 y 59 Años	Ambos	49.560,49	49.506,91
Entre 60 y 64 Años	Ambos	48.548,90	48.540,13
Entre 65 y 69 Años	Ambos	48.279,52	48.263,67
Entre 70 y 74 Años	Ambos	48.594,26	48.590,34
Entre 75 y 79 Años	Ambos	49.852,47	49.869,50
Mayores de 80 Años	Ambos	52.589,09	52.181,76

Tabla 46: Severidad Observada y Estimada Modelo LogNormal Zona Normal Mayor Año 2018

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Normal Mayor para el año 2018.

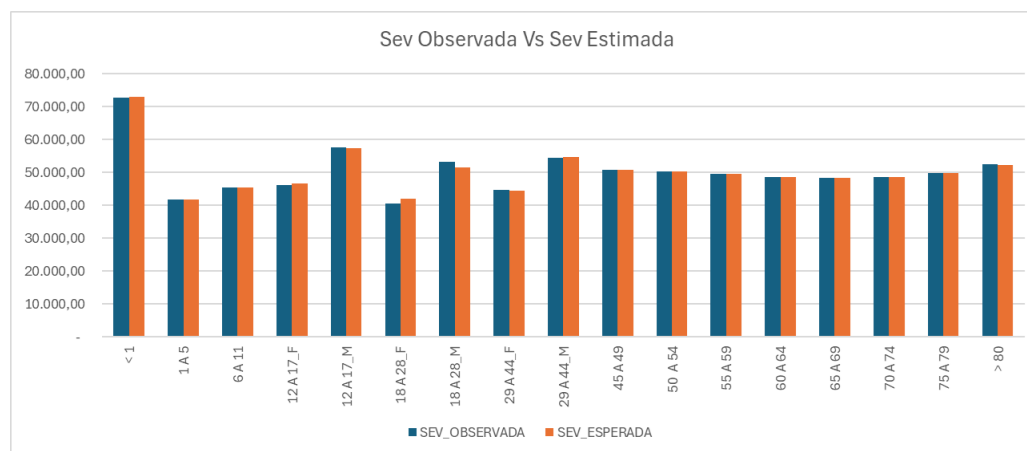


Figura 156: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Normal Mayor Año 2018

El MAPE presentado es de 0.64% siendo mejor que el presentado por el modelo Gamma, y un Pseudo R2 de 0.96 similar al presentado con la distribución GAMMA.

Ahora la elección del mejor modelo se hará a partir del análisis del MAPE identificando aquel que tenga el menor que en este caso es el modelo LogNormal, dado que los dos modelos ajustan de manera aceptable.

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la Zona Normal Mayor para el año 2019

```

GAMLSS-RS iteration 1: Global Deviance = 487.731
GAMLSS-RS iteration 2: Global Deviance = 487.731
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2019_NM)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.398677	0.002145	1118.19	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.037084	0.002974	-12.47	0.000000031443 ***
GRUPO_ETAREO_RUTA_AGREGADO11	-0.038106	0.002975	-12.81	0.000000023319 ***
GRUPO_ETAREO_RUTA_AGREGADO12	-0.038368	0.002976	-12.89	0.000000021618 ***
GRUPO_ETAREO_RUTA_AGREGADO13	-0.035334	0.002971	-11.89	0.000000053397 ***
GRUPO_ETAREO_RUTA_AGREGADO14	-0.030049	0.002963	-10.14	0.000000307343 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.054417	0.003001	-18.14	0.000000000436 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.045762	0.002987	-15.32	0.000000003051 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.037977	0.002975	-12.77	0.000000024202 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.044886	0.002986	-15.03	0.000000003787 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.036841	0.002973	-12.39	0.000000033802 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.036888	0.002973	-12.41	0.000000033327 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.037701	0.002974	-12.68	0.000000026224 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.037391	0.002974	-12.57	0.000000028719 ***
SEXOM	0.019418	0.001144	16.98	0.000000000937 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
sigma link function: log
sigma coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.4284	0.1336	-25.66	0.000000000000749 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance: 487.731
      AIC: 519.731
      SBC: 541.0462
*****

```

Figura 157: Modelo LogNormal para Severidad Zona Normal Mayor Año 2019

El modelo LogNormal presenta todas las variables significativas al 95 % por lo cual se presenta a continuación los residuales

A continuación se presentan los residuales

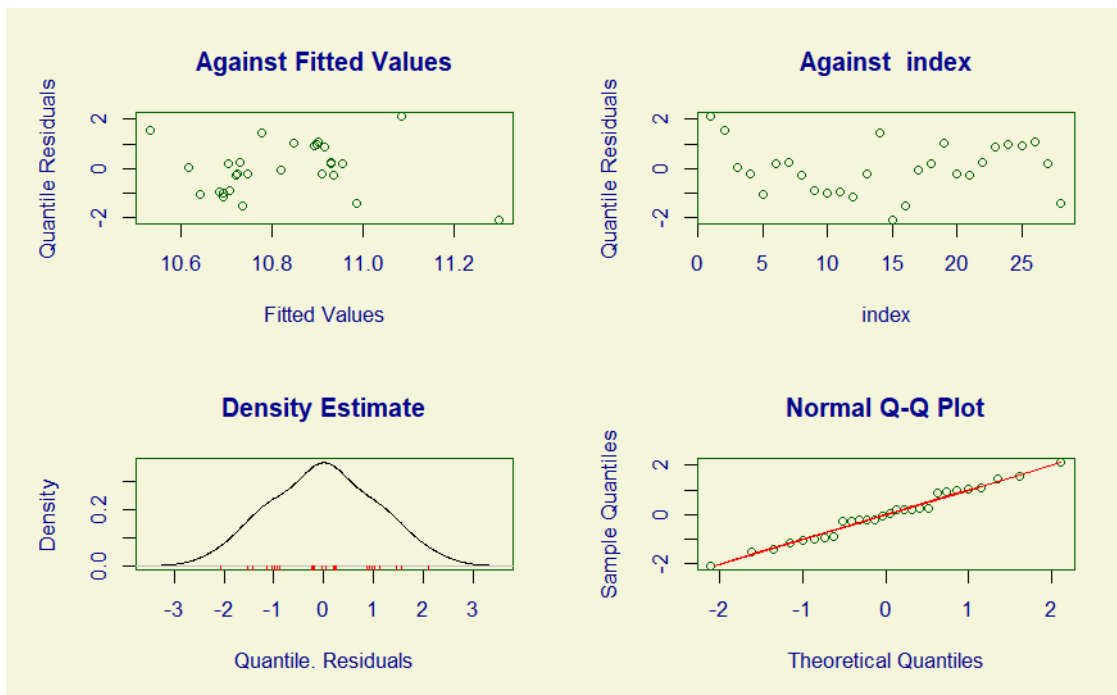


Figura 158: Residuales Modelo LogNormal para Severidad Zona Normal Mayor Año 2018

De acuerdo a la figura 158 el modelo se ajusta de manera aceptable, no se presenta tendencia y los datos se ajustan a la diagonal en el gráfico de QQPLOT, sin embargo se presentan puntos influyentes, con el fin de detectar dichos puntos se presenta la tabla 47 y figura 159, presentando puntos influyentes en las observaciones 1,15 y 28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	2.956504	0.5294455	0.4109533
15	-2.956504	0.5473695	0.4416904
28	-1.318903	0.5451728	0.1315208

Tabla 47: Frecuencias Observadas y Estimadas Zona Normal Mayor Año 2019

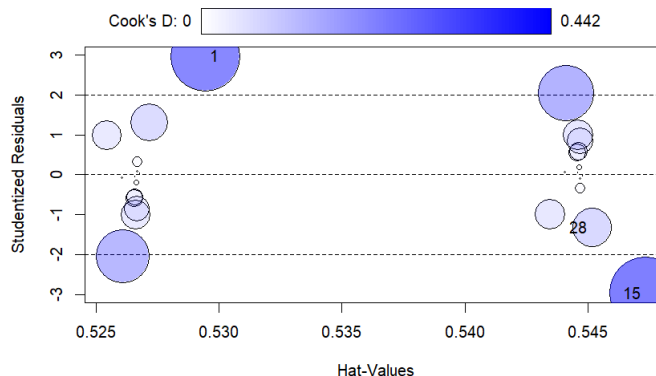


Figura 159: Puntos Influyentes Zona Normal Mayor Año 2019

En la tabla a continuación se presenta los valores observados y los valores estimados aplicando el modelo LogNormal para la Zona Normal Mayor para el año 2019

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	67.396,99	67.796,31
Entre 1 y 5 Años	Ambos	37.495,62	37.613,42
Entre 6 y 11 Años	Ambos	41.236,14	41.224,32
Entre 12 y 17 Años	Femenino	38.786,70	40.069,08
Entre 12 y 17 Años	Masculino	50.923,07	49.324,16
Entre 18 y 28 Años	Femenino	35.089,88	37.249,31
Entre 18 y 28 Años	Masculino	48.549,49	45.787,52
Entre 29 y 44 Años	Femenino	40.674,92	40.554,95
Entre 28 y 44 Años	Masculino	49.789,61	49.934,06
Entre 45 y 49 Años	Ambos	45.120,93	45.158,54
Entre 50 y 54 Años	Ambos	44.657,75	44.653,58
Entre 55 y 59 Años	Ambos	44.685,73	44.676,64
Entre 60 y 64 Años	Ambos	44.671,43	44.661,02
Entre 65 y 69 Años	Ambos	44.164,88	44.163,57
Entre 70 y 74 Años	Ambos	44.017,41	44.020,07
Entre 75 y 79 Años	Ambos	45.299,67	45.297,48
Mayores de 80 Años	Ambos	47.782,92	47.442,38

Tabla 48: Severidad Observada y Estimada Modelo LogNormal Zona Normal Mayor Año 2019

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Normal Mayor para el año 2019.

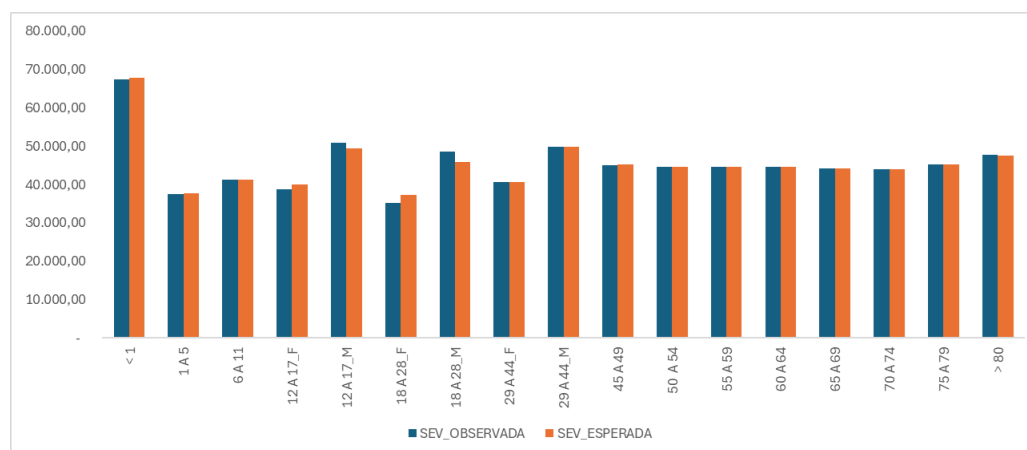


Figura 160: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Normal Mayor Año 2019

El MAPE presentado es de 1.21 % y un Pseudo R2 de 0.96 lo cual presenta un buen ajuste para el año 2019.

#### 4.2.4. Zona Normal Medio

Para los años de estudio 2018 y 2019 el modelo seleccionado para modelar la severidad es el modelo Lognormal, para el año 2018 el modelo de distribución gamma presentaba un buen ajuste por lo que su desarrollo se presenta en el Anexo 2; para el año 2019 no se presenta el desarrollo dado que empíricamente los datos no evidenciaba un buen ajuste.

- GLM LogNormal

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la Zona Normal Medio para el año 2018.

```

GAMLSS-RS iteration 1: Global Deviance = 506.3522
GAMLSS-RS iteration 2: Global Deviance = 506.3522
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2018_NME)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.409813	0.002618	920.526	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.034344	0.003624	-9.476	0.000000638606 ***
GRUPO_ETAREO_RUTA_AGREGADO11	-0.035415	0.003626	-9.766	0.000000462197 ***
GRUPO_ETAREO_RUTA_AGREGADO12	-0.033842	0.003623	-9.340	0.000000745203 ***
GRUPO_ETAREO_RUTA_AGREGADO13	-0.028590	0.003614	-7.912	0.000004210294 ***
GRUPO_ETAREO_RUTA_AGREGADO14	-0.024287	0.003606	-6.736	0.000020860078 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.064315	0.003682	-17.469	0.000000000673 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.059153	0.003672	-16.111	0.000000001713 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.039731	0.003634	-10.932	0.000000135663 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.040804	0.003636	-11.221	0.000000101798 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.036628	0.003629	-10.094	0.000000323583 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.037206	0.003630	-10.251	0.000000273926 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.035237	0.003626	-9.718	0.000000487418 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.035578	0.003627	-9.810	0.000000440283 ***
SEXOM	0.018955	0.001396	13.582	0.000000012028 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.2182	0.1336	-24.08	0.0000000000158 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance:    506.3522
      AIC:          538.3522
      SBC:          559.6674
*****

```

Figura 161: Modelo LogNormal para Severidad Zona Normal Medio Año 2018

El modelo LogNormal presenta todas las variables significativas al 95 % por lo cual se presenta a continuación los residuales

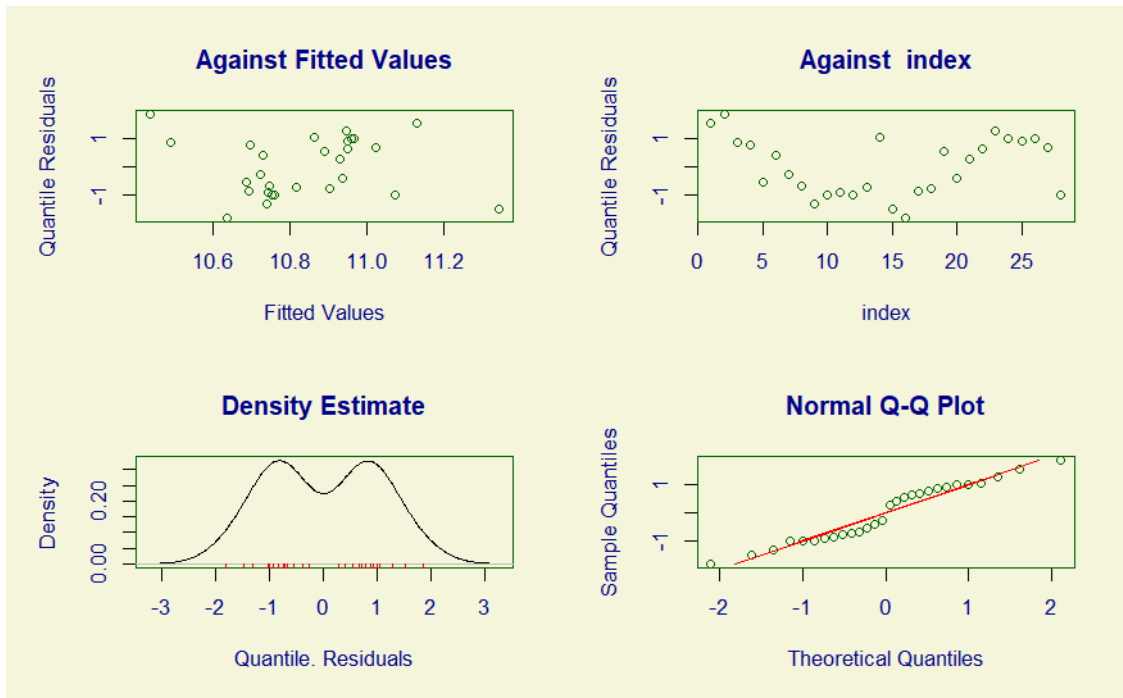


Figura 162: Residuales Modelo LogNormal para Severidad Zona Normal Medio Año 2018

Como se observa en la figura 162, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 49 y figura 163, presentando puntos influyentes en las observaciones 2,15,16,28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
2	2.036286	0.5249027	0.24589425
15	-1.605937	0.5471200	0.18521663
16	-2.036286	0.5425764	0.26399428
28	-1.037609	0.5453353	0.08558467

Tabla 49: Puntos Influyentes Distribución LogNormal Zona Normal Medio Año 2018

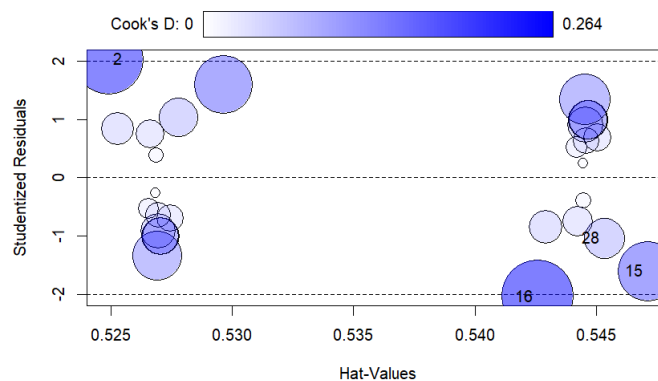


Figura 163: untos Influyentes Distribución LogNormal Zona Normal Medio Año 2018

En la tabla a continuación se presenta los valores observados y los valores estimados aplicando el modelo LogNormal para la zona normal medio Año 2018.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	76.152,84	76.508,31
Entre 1 y 5 Años	Ambos	37.796,98	38.007,58
Entre 6 y 11 Años	Ambos	40.004,89	40.134,37
Entre 12 y 17 Años	Femenino	45.667,95	44.279,15
Entre 12 y 17 Años	Masculino	52.716,60	54.338,52
Entre 18 y 28 Años	Femenino	42.821,71	43.773,96
Entre 18 y 28 Años	Masculino	54.878,41	53.706,77
Entre 29 y 44 Años	Femenino	46.523,72	45.776,19
Entre 29 y 44 Años	Masculino	55.325,05	56.211,41
Entre 45 y 49 Años	Ambos	50.945,37	50.870,99
Entre 50 y 54 Años	Ambos	52.075,69	51.897,31
Entre 55 y 59 Años	Ambos	51.914,89	51.589,75
Entre 60 y 64 Años	Ambos	52.307,24	52.128,97
Entre 65 y 69 Años	Ambos	51.656,23	51.503,89
Entre 70 y 74 Años	Ambos	52.439,16	52.302,84
Entre 75 y 79 Años	Ambos	55.251,59	55.205,92
Mayores de 80 Años	Ambos	57.675,06	57.526,16

Tabla 50: Severidad Observada y Estimada Modelo LogNormal Zona Normal Medio Año 2018

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Normal Medio.

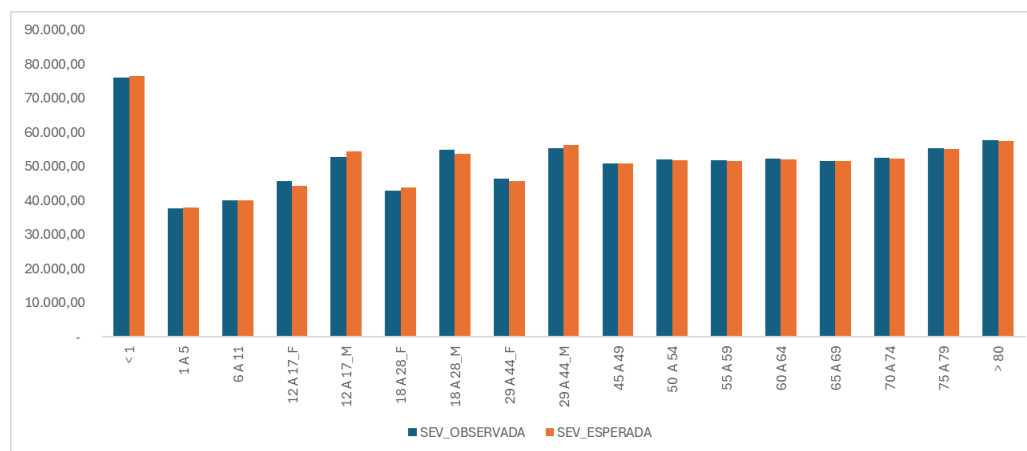


Figura 164: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Normal Medio Año 2018

El MAPE presentado es de 1.022 % siendo mejor que el presentado por el modelo Gamma, el Pseudo R2 este se ubica en 0.95 similar al presentado con el modelo Gamma.

Ahora la elección del mejor modelo se hará a partir del análisis del MAPE identificando aquel que tenga el menor que en este caso es el modelo LogNormal, dado que los dos modelos ajustan de manera aceptable.

Para el año 2019 se aplica el modelo utilizando la distribución LogNormal y a continuación se presentan los resultados para modelar la severidad de la Zona Normal Medio

```

GAMLSS-RS iteration 1: Global Deviance = 488.7367
GAMLSS-RS iteration 2: Global Deviance = 488.7367
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2019_NME)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.388710   0.002230 1070.957 < 2e-16 ***
GRUPO_ETAREO_RUTA_AGREGADO10 -0.021990   0.003071  -7.161 1.15e-05 ***
GRUPO_ETAREO_RUTA_AGREGADO11 -0.022377   0.003071  -7.286 9.66e-06 ***
GRUPO_ETAREO_RUTA_AGREGADO12 -0.021571   0.003070  -7.026 1.38e-05 ***
GRUPO_ETAREO_RUTA_AGREGADO13 -0.018937   0.003066  -6.177 4.75e-05 ***
GRUPO_ETAREO_RUTA_AGREGADO14 -0.015780   0.003061  -5.155 0.000239 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.059842   0.003132 -19.108 2.37e-10 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.052528   0.003120 -16.838 1.03e-09 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.036044   0.003093 -11.654 6.70e-08 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.034046   0.003090 -11.019 1.24e-07 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.026882   0.003078  -8.733 1.52e-06 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.027552   0.003079  -8.947 1.17e-06 ***
GRUPO_ETAREO_RUTA_AGREGADO8  -0.026646   0.003078  -8.657 1.66e-06 ***
GRUPO_ETAREO_RUTA_AGREGADO9  -0.024221   0.003074  -7.879 4.39e-06 ***
SEXOM                0.019547   0.001180  16.565 1.24e-09 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
sigma link function: log
sigma coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.3984     0.1336 -25.43 8.31e-12 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance: 488.7367
              AIC: 520.7367
              SBC: 542.052
*****

```

Figura 165: Modelo LogNormal para Severidad Zona Normal Medio Año 2019

El modelo LogNormal presenta todas las variables significativas al 95 % por lo cual se presenta a continuación los residuales

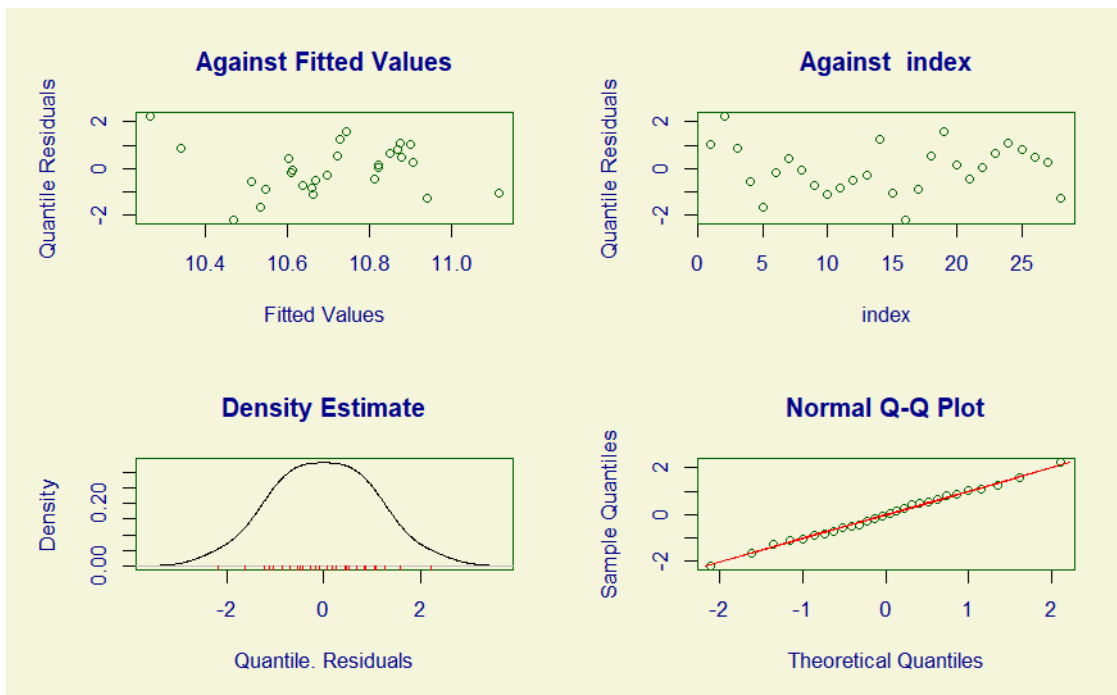


Figura 166: Residuales Modelo LogNormal para Severidad Zona Normal Medio Año 2019

Como se observa en la figura 166, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 51 y figura 167, presentando puntos influyentes en las observaciones 2,15,16,28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
2	2.651965	0.5243622	0.35305083
15	-1.059121	0.5467713	0.08938004
16	-2.651965	0.5425979	0.37989372
28	-1.279674	0.5456219	0.12496502

Tabla 51: Puntos Influyentes Distribución LogNormal Zona Normal Medio Año 2019

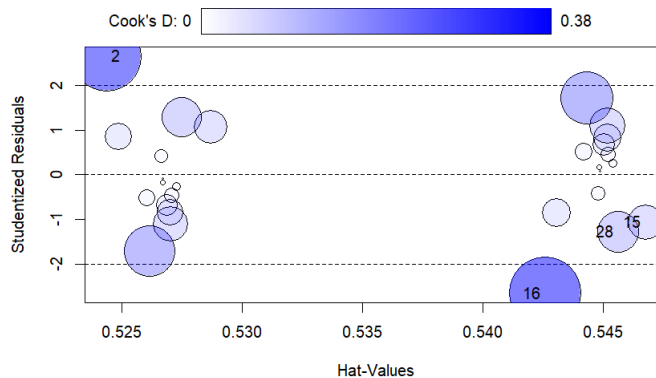


Figura 167: untos Influyentes Distribución LogNormal Zona Normal Medio Año 2018

En la tabla a continuación se presenta los valores observados y los valores estimados aplicando el modelo LogNormal para la zona normal medio Año 2019.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	60.589,07	60.815,54
Entre 1 y 5 Años	Ambos	31.870,30	32.055,68
Entre 6 y 11 Años	Ambos	34.483,60	34.582,14
Entre 12 y 17 Años	Femenino	36.151,86	36.811,29
Entre 12 y 17 Años	Masculino	46.111,39	45.301,21
Entre 18 y 28 Años	Femenino	35.611,00	37.593,70
Entre 18 y 28 Años	Masculino	48.808,95	46.283,28
Entre 29 y 44 Años	Femenino	40.320,58	40.551,76
Entre 29 y 44 Años	Masculino	50.280,90	49.999,78
Entre 45 y 49 Años	Ambos	45.059,48	45.142,36
Entre 50 y 54 Años	Ambos	45.545,70	45.531,09
Entre 55 y 59 Años	Ambos	46.725,55	46.608,16
Entre 60 y 64 Años	Ambos	47.736,01	47.584,63
Entre 65 y 69 Años	Ambos	47.443,35	47.345,48
Entre 70 y 74 Años	Ambos	47.755,90	47.711,94
Entre 75 y 79 Años	Ambos	48.920,90	48.911,65
Mayores de 80 Años	Ambos	50.407,66	50.262,35

Tabla 52: Severidad Observada y Estimada Modelo LogNormal Zona Normal Medio Año 2019

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Normal Medio para el año 2019.

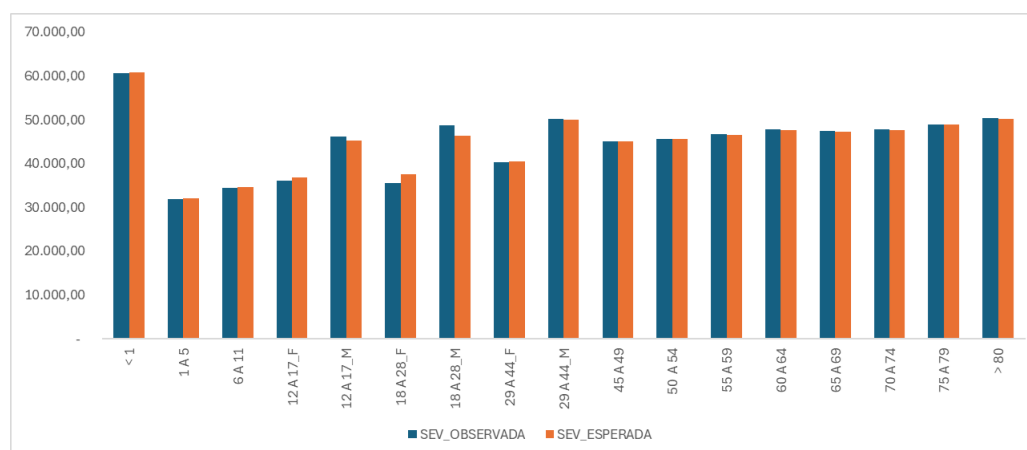


Figura 168: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Normal Medio Año 2019

El MAPE presentado es de 1.063 % , así mismo el Pseudo R2 este se ubica en 0.96 similar al presentado con el modelo Gamma.

#### 4.2.5. Zona Rural

De acuerdo a las distribuciones empíricas presentadas para las severidades para los años 2018 y 2019, la distribución que se ajusta a los datos es la lognormal por lo cual se realiza la estimación de la severidad utilizando dicha distribución.

- GLM LogNormal

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la Zona Rural para el año 2018

```

GAMLSS-RS iteration 1: Global Deviance = 516.4591
GAMLSS-RS iteration 2: Global Deviance = 516.4591
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2018_RU)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.423271	0.002543	953.064	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.026664	0.003508	-7.602	0.00000631709 ***
GRUPO_ETAREO_RUTA_AGREGADO11	-0.027371	0.003509	-7.800	0.00000486520 ***
GRUPO_ETAREO_RUTA_AGREGADO12	-0.030287	0.003514	-8.618	0.00000174065 ***
GRUPO_ETAREO_RUTA_AGREGADO13	-0.026767	0.003508	-7.630	0.00000608106 ***
GRUPO_ETAREO_RUTA_AGREGADO14	-0.021270	0.003498	-6.081	0.00005496757 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.056380	0.003562	-15.827	0.00000000210 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.052823	0.003556	-14.856	0.00000000434 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.034224	0.003521	-9.719	0.00000048695 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.032154	0.003518	-9.141	0.00000093671 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.031469	0.003516	-8.949	0.00000117181 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.030212	0.003514	-8.597	0.00000178530 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.031030	0.003516	-8.826	0.00000135563 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.032180	0.003518	-9.148	0.00000092888 ***
SEXOM	0.017680	0.001349	13.105	0.00000001801 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.2341	0.1336	-24.2	0.0000000000149 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance: 516.4591
AIC: 548.4591
SBC: 569.7744
*****

```

Figura 169: Modelo LogNormal para Severidad Zona Rural Año 2018

El modelo LogNormal presenta todas las variables significativas al 95% por lo cual se presenta a continuación los residuales

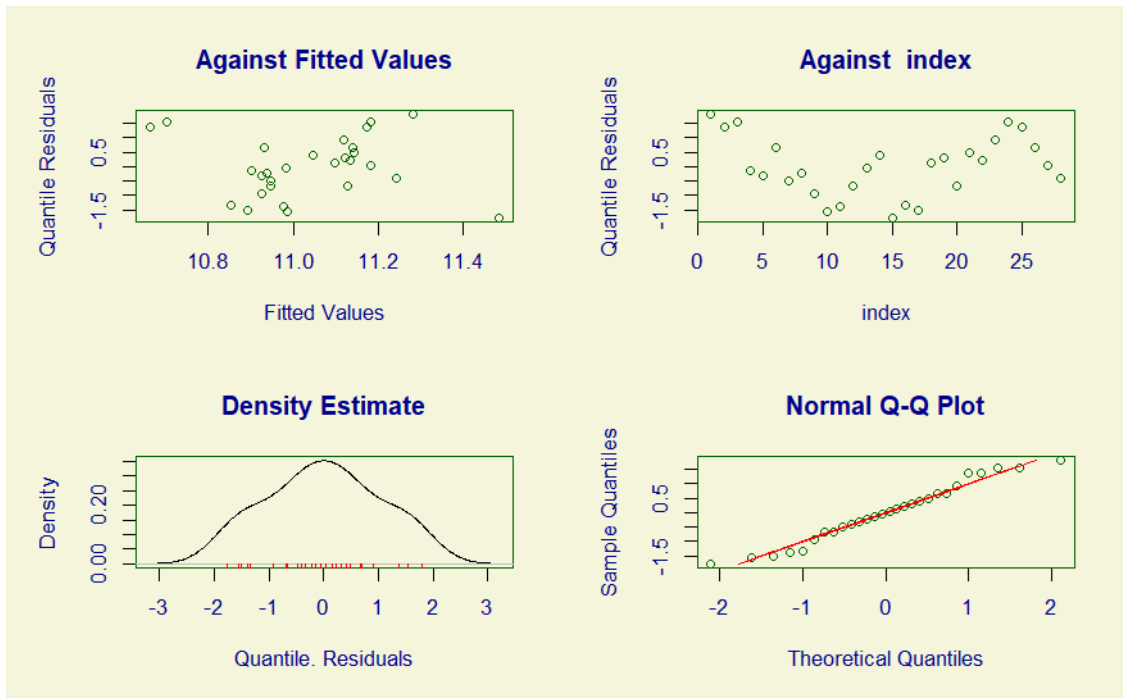


Figura 170: Residuales Modelo LogNormal para Severidad Zona Rural Año 2018

Como se observa en la figura 170, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta en la tabla 53 y en la figura 171, presentando puntos influyentes en las observaciones 1,15 y 28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	1.9858285	0.5298125	0.24154750
15	-1.9858285	0.5461478	0.25795699
28	-0.3987218	0.5445940	0.01355093

Tabla 53: Puntos Influyentes Distribución Lognormal Zona Rural Año 2018

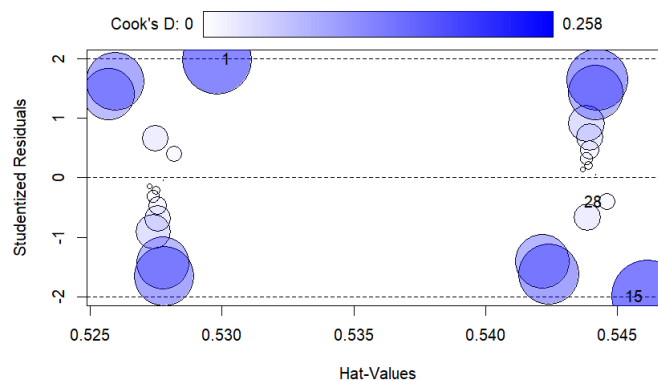


Figura 171: Puntos Influyentes Distribución Lognormal Zona Rural Año 2018

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la severidad esperada para la zona rural año 2018, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	87.986,41	88.445,79
Entre 1 y 5 Años	Ambos	47.168,59	47.370,91
Entre 6 y 11 Años	Ambos	49.004,02	49.242,99
Entre 12 y 17 Años	Femenino	54.034,10	54.344,65
Entre 12 y 17 Años	Masculino	66.384,18	66.011,45
Entre 18 y 28 Años	Femenino	54.863,55	55.586,04
Entre 18 y 28 Años	Masculino	68.420,38	67.546,56
Entre 29 y 44 Años	Femenino	57.526,99	56.003,77
Entre 29 y 44 Años	Masculino	66.292,23	68.063,27
Entre 45 y 49 Años	Ambos	64.088,21	63.797,87
Entre 50 y 54 Años	Ambos	63.293,03	63.169,11
Entre 55 y 59 Años	Ambos	62.736,24	62.241,91
Entre 60 y 64 Años	Ambos	66.718,42	65.926,76
Entre 65 y 69 Años	Ambos	65.970,12	65.330,42
Entre 70 y 74 Años	Ambos	63.297,05	63.063,96
Entre 75 y 79 Años	Ambos	65.315,89	65.305,70
Mayores de 80 Años	Ambos	68.777,73	68.751,13

Tabla 54: Severidad Observada y Estimada Modelo LogNormal Zona Rural Año 2018

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Rural.

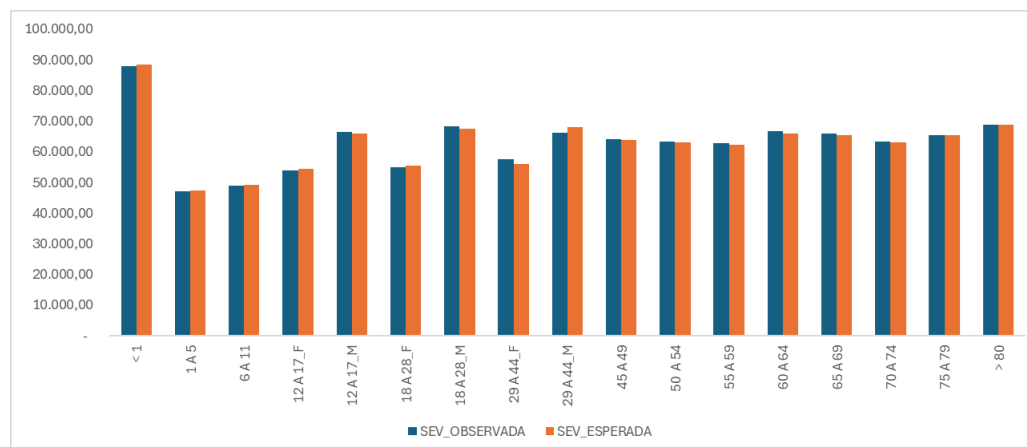


Figura 172: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Rural Año 2018

El MAPE presentado es de 0.85 % y Pseudo R2 que presenta es de 0.94

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la Zona Rural para el año 2019.

```

GAMLSS-RS iteration 1: Global Deviance = 484.3893
GAMLSS-RS iteration 2: Global Deviance = 484.3893
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2019_RU)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.381936   0.001960 1215.079 < 2e-16 ***
GRUPO_ETAREO_RUTA_AGREGADO10 -0.006259   0.002680  -2.336 0.037676 *
GRUPO_ETAREO_RUTA_AGREGADO11 -0.007298   0.002681  -2.722 0.018537 *
GRUPO_ETAREO_RUTA_AGREGADO12 -0.006955   0.002681  -2.595 0.023448 *
GRUPO_ETAREO_RUTA_AGREGADO13 -0.008213   0.002682  -3.062 0.009859 **
GRUPO_ETAREO_RUTA_AGREGADO14 -0.008132   0.002682  -3.032 0.010431 *
GRUPO_ETAREO_RUTA_AGREGADO2  -0.052340   0.002744 -19.075 2.42e-10 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.041061   0.002728 -15.053 3.73e-09 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.023601   0.002703  -8.731 1.52e-06 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.021388   0.002700  -7.921 4.16e-06 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.013361   0.002689  -4.968 0.000326 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.013479   0.002689  -5.012 0.000303 ***
GRUPO_ETAREO_RUTA_AGREGADO8  -0.011133   0.002686  -4.145 0.001360 **
GRUPO_ETAREO_RUTA_AGREGADO9  -0.009358   0.002684  -3.487 0.004489 **
SEXOM                    0.020052   0.001026  19.548 1.82e-10 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
sigma link function: log
sigma coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.5332    0.1336  -26.44 5.25e-12 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance: 484.3893
              AIC: 516.3893
              SBC: 537.7046
*****

```

Figura 173: Modelo LogNormal para Severidad Zona Rural Año 2019

El modelo LogNormal presenta todas las variables significativas al 95 % por lo cual se presenta a continuación los residuales

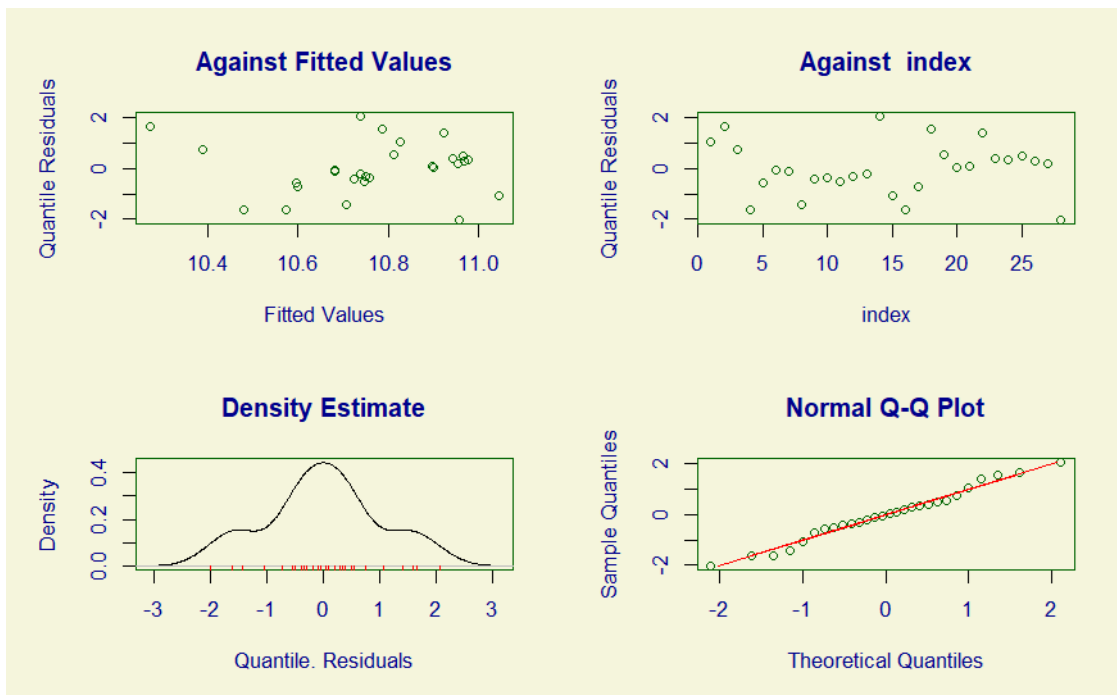


Figura 174: Residuales Modelo LogNormal para Severidad Zona Rural Año 2019

Como se observa en la figura 174, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta en la tabla 55 y en la figura 175, presentando puntos influyentes en las observaciones 14,15,24 y 28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
14	2.3706528	0.5269616	0.307939077
15	-1.0681813	0.5461397	0.090551322
24	0.3235922	0.5456905	0.009005094
28	-2.3706528	0.5455571	0.331851009

Tabla 55: Puntos Influyentes Distribución Lognormal Zona Rural Año 2019

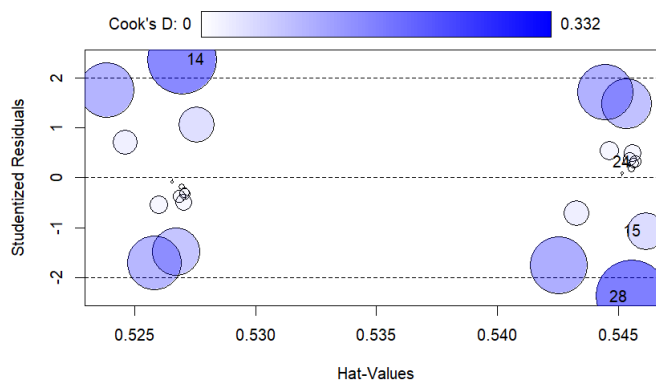


Figura 175: Puntos Influyentes Distribución Lognormal Zona Rural Año 2019

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la severidad esperada para la zona rural año 2019, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	56.443,00	56.636,53
Entre 1 y 5 Años	Ambos	32.247,79	32.391,97
Entre 6 y 11 Años	Ambos	36.371,21	36.457,83
Entre 12 y 17 Años	Femenino	37.284,87	39.078,47
Entre 12 y 17 Años	Masculino	50.692,68	48.411,15
Entre 18 y 28 Años	Femenino	39.362,01	40.004,82
Entre 18 y 28 Años	Masculino	50.375,76	49.582,25
Entre 29 y 44 Años	Femenino	43.507,40	43.571,50
Entre 29 y 44 Años	Masculino	54.174,43	54.096,31
Entre 45 y 49 Años	Ambos	49.654,22	49.620,29
Entre 50 y 54 Años	Ambos	51.427,16	50.859,64
Entre 55 y 59 Años	Ambos	51.865,41	51.730,23
Entre 60 y 64 Años	Ambos	53.363,67	53.263,33
Entre 65 y 69 Años	Ambos	52.679,40	52.545,66
Entre 70 y 74 Años	Ambos	52.625,44	52.560,52
Entre 75 y 79 Años	Ambos	51.615,12	51.587,60
Mayores de 80 Años	Ambos	51.209,46	51.107,55

Tabla 56: Severidad Observada y Estimada Modelo LogNormal Zona Rural Año 2019

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Rural.

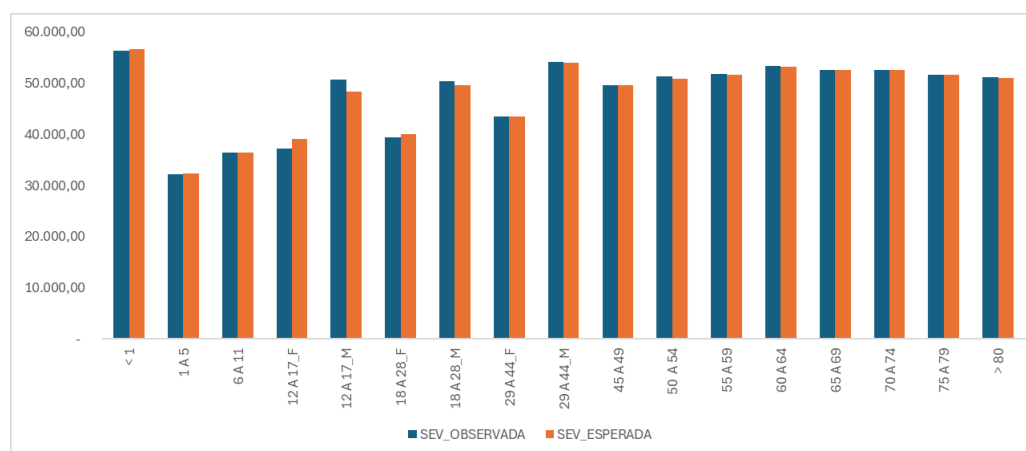


Figura 176: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Rural Año 2019

El MAPE presentado es de 0.94% y Pseudo R2 que presenta es de 0.97

#### 4.2.6. Zona Rural Disperso

GLM Gamma Dada la gráfica de severidad empírica para el año 2018 se observa que el modelo Gamma no ajusta a los datos por lo cual no se realiza este modelo para este año A continuación se presentan los resultados de la aplicación del modelo Gamma, para modelar la severidad de la Zona Rural Disperso para el año 2019, el desarrollo del modelo lognormal se presenta en el Anexo 2.

```

Call:
glm(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO,
     family = Gamma(link = "log"), data = BASE_2019_RUD)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      10.624312   0.134674  78.889 < 2e-16 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.111486   0.184000   0.606  0.55501
GRUPO_ETAREO_RUTA_AGREGADO11  0.125894   0.184000   0.684  0.50586
GRUPO_ETAREO_RUTA_AGREGADO12  0.045859   0.184000   0.249  0.80707
GRUPO_ETAREO_RUTA_AGREGADO13  0.153491   0.184000   0.834  0.41924
GRUPO_ETAREO_RUTA_AGREGADO14  0.165045   0.184000   0.897  0.38604
GRUPO_ETAREO_RUTA_AGREGADO2   -0.623936   0.184000  -3.391  0.00483 **
GRUPO_ETAREO_RUTA_AGREGADO3   -0.311741   0.184000  -1.694  0.11403
GRUPO_ETAREO_RUTA_AGREGADO4    0.126662   0.184000   0.688  0.50331
GRUPO_ETAREO_RUTA_AGREGADO5   -0.163382   0.184000  -0.888  0.39070
GRUPO_ETAREO_RUTA_AGREGADO6   -0.037836   0.184000  -0.206  0.84027
GRUPO_ETAREO_RUTA_AGREGADO7   -0.002788   0.184000  -0.015  0.98814
GRUPO_ETAREO_RUTA_AGREGADO8    0.004195   0.184000   0.023  0.98216
GRUPO_ETAREO_RUTA_AGREGADO9    0.146992   0.184000   0.799  0.43872
SEXOM                      0.172778   0.069545   2.484  0.02738 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.03385599)

Null deviance: 1.74693  on 27  degrees of freedom
Residual deviance: 0.45362  on 13  degrees of freedom
AIC: 594.41

Number of Fisher scoring iterations: 6

```

Figura 177: Modelo Gamma para Severidad Zona Rural Disperso Año 2019

En el modelo Gamma 3 de las variables son significativas al 95 %, sin embargo no es posible no tener en cuenta las variables no cumplen con esta característica dada la finalidad del presente trabajo de grado, por lo cual se analizan los residuales con el fin de realizar la proyección de severidad, para la zona rural disperso año 2019.

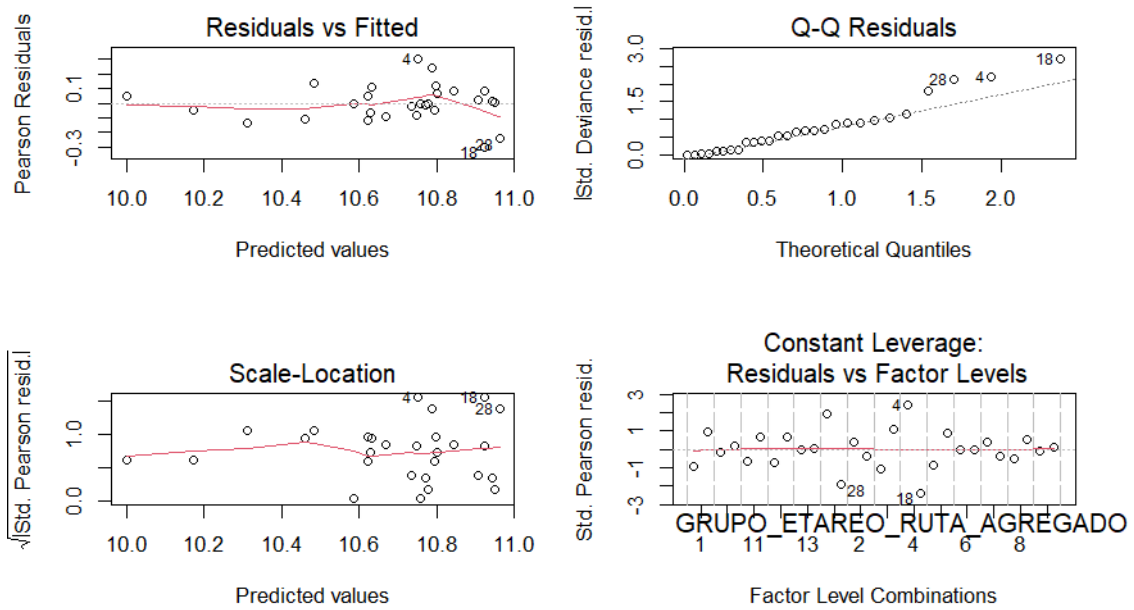


Figura 178: Residuales Modelo Gamma para Severidad Zona Rural Disperso Año 2019

Como se observa en la figura 178, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 57 y figura 179, presentando puntos influyentes en las observaciones 4, 5v y 18, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
4	2.7557053	0.5357143	0.4492507
5	-0.8745151	0.5357143	0.0593889
18	-3.6131683	0.5357143	0.4492518

Tabla 57: Puntos Influyentes Distribución Gamma Zona Rural Disperso Año 2019

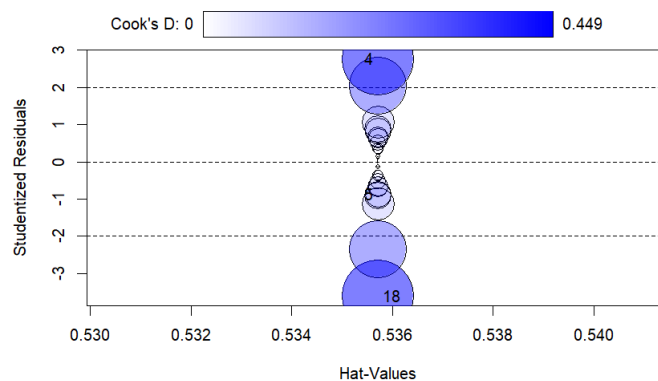


Figura 179: Puntos Influyentes Distribución Gamma Zona Rural Disperso Año 2019

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la severidad esperada para el año 2019, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	45.563,92	45.051,21
Entre 1 y 5 Años	Ambos	24.031,55	24.154,02
Entre 6 y 11 Años	Ambos	33.547,12	33.026,30
Entre 12 y 17 Años	Femenino	60.817,60	46.675,46
Entre 12 y 17 Años	Masculino	38.669,14	55.478,54
Entre 18 y 28 Años	Femenino	31.076,70	34.924,03
Entre 18 y 28 Años	Masculino	46.083,71	41.510,76
Entre 29 y 44 Años	Femenino	39.601,28	39.595,73
Entre 28 y 44 Años	Masculino	47.056,95	47.063,55
Entre 45 y 49 Años	Ambos	45.080,61	45.676,50
Entre 50 y 54 Años	Ambos	46.860,59	45.981,59
Entre 55 y 59 Años	Ambos	53.205,41	52.995,65
Entre 60 y 64 Años	Ambos	51.381,93	51.121,05
Entre 65 y 69 Años	Ambos	52.512,48	51.594,73
Entre 70 y 74 Años	Ambos	47.754,28	47.254,03
Entre 75 y 79 Años	Ambos	52.926,75	52.895,08
Mayores de 80 Años	Ambos	52.376,29	52.848,45

Tabla 58: Severidad Observada y Estimada Modelo Gamma Zona Rural Disperso Año 2019

A continuación se presenta gráficamente la severidad observada y la severidad estimada del modelo Gamma para el año 2019.

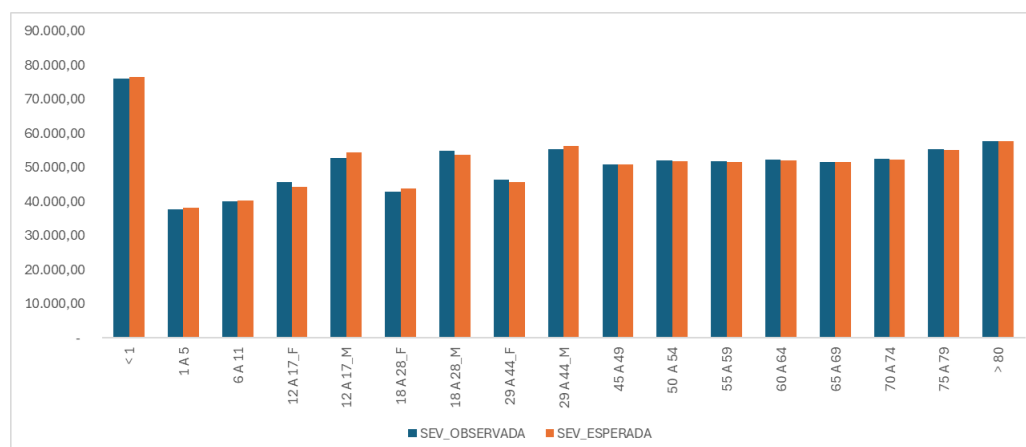


Figura 180: Severidad Observada Vs Severidad Estimada Zona Rural Disperso Año 2019

El modelo presenta un error porcentual absoluto medio (MAPE) del 5.88 % lo cual se presenta un ajuste aceptable, y un pseudo R2 de 0.74

De acuerdo a las distribuciones empíricas presentadas para modelar la severidad en el año 2018, la distribución gamma no presenta un ajuste para los datos analizados por lo cual se decide realizar el análisis para utilizando la distribución LogNormal, cuyo desarrollo se presenta a continuación.

#### ■ GLM LogNormal

A continuación se presentan los resultados de la aplicación del modelo LogNormal, para modelar la severidad de la Zona Rural Disperso 2018.

```

GAMLSS-RS iteration 1: Global Deviance = 543.477
GAMLSS-RS iteration 2: Global Deviance = 543.477
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2018_RUD)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.400901	0.004683	512.668	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.013048	0.006416	-2.034	0.064695 .
GRUPO_ETAREO_RUTA_AGREGADO11	-0.009266	0.006403	-1.447	0.173506
GRUPO_ETAREO_RUTA_AGREGADO12	0.008640	0.006346	1.361	0.198412
GRUPO_ETAREO_RUTA_AGREGADO13	-0.025336	0.006456	-3.924	0.002019 **
GRUPO_ETAREO_RUTA_AGREGADO14	0.005792	0.006355	0.911	0.380071
GRUPO_ETAREO_RUTA_AGREGADO2	-0.072161	0.006617	-10.906	0.0000001392 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.070941	0.006612	-10.729	0.0000001666 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.039465	0.006503	-6.068	0.0000559911 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.040563	0.006507	-6.234	0.0000435929 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.032644	0.006480	-5.037	0.000291 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.013852	0.006418	-2.158	0.051874 .
GRUPO_ETAREO_RUTA_AGREGADO8	-0.033597	0.006484	-5.182	0.000229 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.023701	0.006451	-3.674	0.003182 **
SEXOM	0.030930	0.002472	12.514	0.0000000303 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
sigma link function: log
sigma coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.6390	0.1336	-19.75	0.000000000162 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
Residual Deg. of Freedom: 12
at cycle: 2

Global Deviance: 543.477
AIC: 575.477
SBC: 596.7923
*****

```

Figura 181: Modelo LogNormal para Severidad Zona Rural Disperso Año 2018

El modelo LogNormal presenta 9 de las variables significativas al 95 %, sin embargo no es posible descartar alguno de los grupos etarios por lo cual se presenta a continuación los residuales

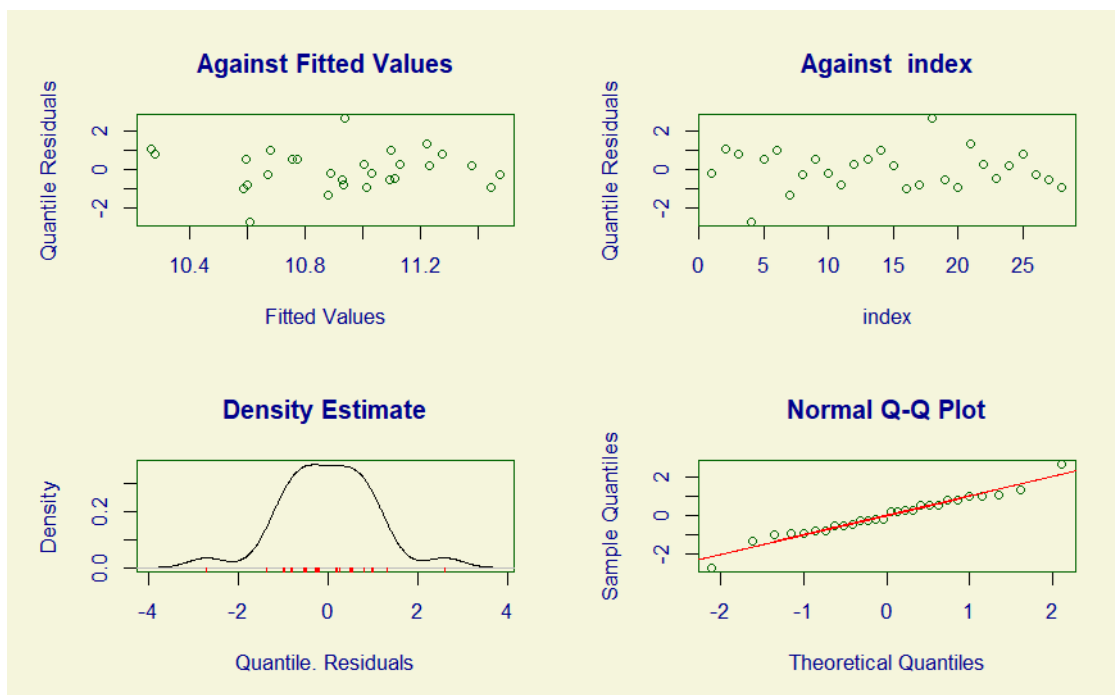


Figura 182: Residuales Modelo LogNormal para Severidad Zona Rural Disperso Año 2018

Como se observa en la figura numero 182, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla numero 59 y figura numero 183, presentando puntos influyentes en las observaciones 4,18,26 Y 28, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
4	-3.7467828	0.5203189	0.506841548
18	3.7467828	0.5490935	0.569003458
26	-0.2455946	0.5524900	0.005351221
28	-0.9697507	0.5522797	0.077692013

Tabla 59: Puntos Influyentes Distribución LogNormal Zona Rural Disperso 2018

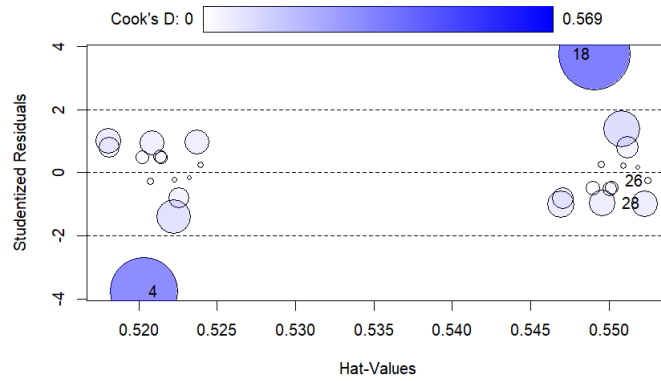


Figura 183: Puntos Influyentes Distribución LogNormal Zona Rural Disperso 2018

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la severidad esperada para el año 2018 en la zona rural disperso, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	74.717,21	74.573,29
Entre 1 y 5 Años	Ambos	33.953,46	34.258,27
Entre 6 y 11 Años	Ambos	34.453,24	34.724,65
Entre 12 y 17 Años	Femenino	33.321,20	40.383,58
Entre 12 y 17 Años	Masculino	67.895,44	56.350,66
Entre 18 y 28 Años	Femenino	41.455,58	39.916,08
Entre 18 y 28 Años	Masculino	53.672,11	55.677,95
Entre 29 y 44 Años	Femenino	46.554,61	43.423,81
Entre 29 y 44 Años	Masculino	56.767,32	60.731,27
Entre 45 y 49 Años	Ambos	68.253,26	65.933,33
Entre 50 y 54 Años	Ambos	53.643,03	53.270,81
Entre 55 y 59 Años	Ambos	58.520,92	59.142,31
Entre 60 y 64 Años	Ambos	66.805,08	66.459,85
Entre 65 y 69 Años	Ambos	69.627,68	68.513,65
Entre 70 y 74 Años	Ambos	82.959,84	83.278,10
Entre 75 y 79 Años	Ambos	56.518,55	56.922,02
Mayores de 80 Años	Ambos	78.558,83	78.898,25

Tabla 60: Severidad Observada y Estimada Modelo LogNormal Zona Rural Disperso Año 2018

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Rural Disperso.

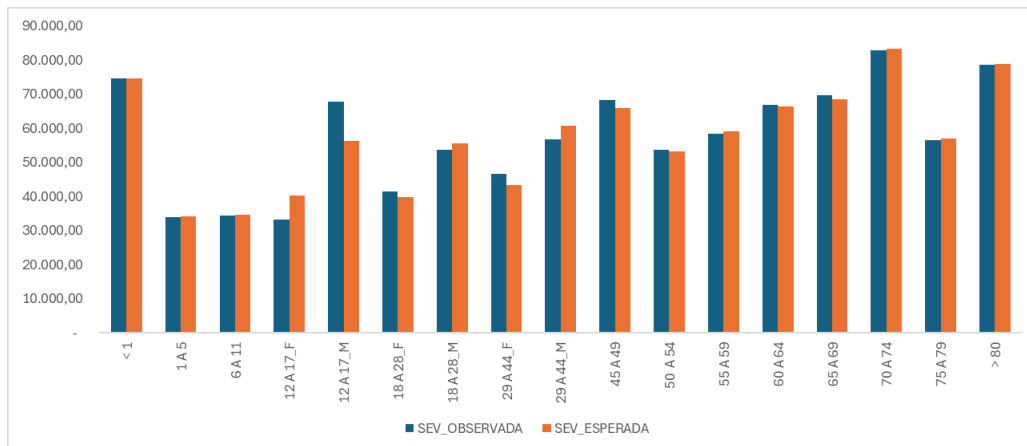


Figura 184: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Rural Disperso Año 2018

El MAPE presentado es de 4.11 % y un Pseudo R2 de 0.94.

### 4.3. Modelos de Prima Pura

Esta sección se abordará en dos partes, la primera presenta la prima pura analizada por cada una de las zonas, y la segunda parte presenta la prima pura establecida a partir de modelos donde se involucran todas las zonas como variable explicativa.

### 4.4. Prima Pura por Zona

#### 4.4.1. Zona Alejada

A continuación se presenta la prima pura para cada uno de los grupos etarios de la zona alejada para el año 2018, teniendo en cuenta que la misma se expresa como:

$$PrimaPura = \frac{\text{Pérdidas}}{\text{Número de expuestos}} = Frecuencia * Severidad$$

Teniendo en cuenta que la modelo seleccionado para la frecuencia sigue una distribución binomial negativa y para la severidad sigue una distribución LogNormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	5,06675723	126.117,94	639.008,99
Entre 1 y 5 Años	Ambos	4,800857429	98.082,19	470.878,60
Entre 6 y 11 Años	Ambos	2,386468467	94.851,25	226.359,52
Entre 12 y 17 Años	Femenino	2,498696696	79.955,68	199.784,99
Entre 12 y 17 Años	Masculino	1,859763701	114.565,79	213.065,29
Entre 18 y 28 Años	Femenino	3,006870631	94.929,21	285.439,84
Entre 18 y 28 Años	Masculino	2,237994256	136.766,80	306.083,31
Entre 29 y 44 Años	Femenino	4,212897632	89.842,05	378.495,36
Entre 28 y 44 Años	Masculino	3,13563231	129.210,67	405.157,15
Entre 45 y 49 Años	Ambos	4,745718349	103.528,50	491.317,10
Entre 50 y 54 Años	Ambos	5,741577578	108.890,39	625.202,62
Entre 55 y 59 Años	Ambos	7,244732459	120.924,50	876.065,67
Entre 60 y 64 Años	Ambos	9,125836342	107.251,25	978.757,33
Entre 65 y 69 Años	Ambos	12,61763419	109.268,57	1.378.710,80
Entre 70 y 74 Años	Ambos	12,31886615	151.895,00	1.871.174,21
Entre 75 y 79 Años	Ambos	14,27383736	128.369,47	1.832.324,98
Mayores de 80 Años	Ambos	15,07257679	120.569,45	1.817.292,25

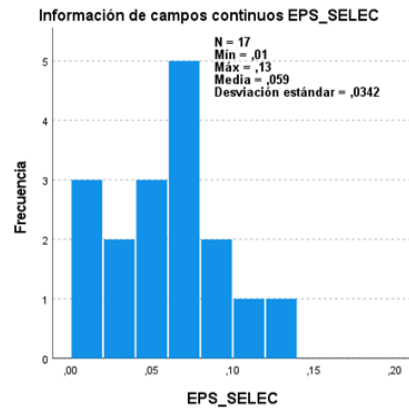
Tabla 61: Prima Pura Zona Alejada Año 2018

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones, para lo anterior compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

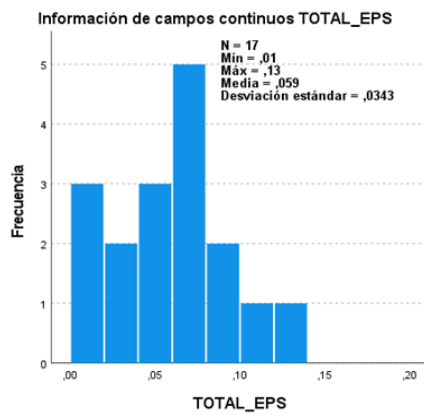
La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona alejada, adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.



(a) Prueba de Rangos Zona Alejada Año 2018

(b) EPS Selección Zona Alejada Año 2018



(c) Total EPS Zona Alejada Año 2018

Resumen de contrastes de hipótesis			
Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1 La mediana de diferencias entre EPS_SELEC y TOTAL_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	.653	Conserve la hipótesis nula.

a. El nivel de significación es de .050.  
 b. Se muestra la significancia asintótica.

Figura 185: Prueba Wilcoxon Zona Alejada Año 2018

A continuación se presenta la prima pura para cada uno de los grupos etarios de la zona alejada para el año 2019, teniendo en cuenta que la modelo seleccionado para la frecuencia sigue una distribución binomial negativa y para la severidad sigue una distribución LogNormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	10,87475877	63.486,54	690.400,83
Entre 1 y 5 Años	Ambos	8,724850161	68.193,25	594.975,93
Entre 6 y 11 Años	Ambos	4,327856796	67.633,62	292.708,62
Entre 12 y 17 Años	Femenino	4,937701428	55.196,90	272.545,80
Entre 12 y 17 Años	Masculino	3,789503414	64.529,02	244.532,93
Entre 18 y 28 Años	Femenino	6,974873003	64.071,05	446.887,47
Entre 18 y 28 Años	Masculino	5,352957331	74.903,52	400.955,37
Entre 29 y 44 Años	Femenino	8,197317783	70.986,19	581.896,39
Entre 29 y 44 Años	Masculino	6,291138534	82.987,80	522.087,77
Entre 45 y 49 Años	Ambos	10,64353405	69.805,74	742.979,73
Entre 50 y 54 Años	Ambos	13,96400469	77.564,71	1.083.113,95
Entre 55 y 59 Años	Ambos	17,330998	70.702,61	1.225.346,78
Entre 60 y 64 Años	Ambos	22,52830015	66.196,33	1.491.290,79
Entre 65 y 69 Años	Ambos	28,5437368	70.925,43	2.024.476,74
Entre 70 y 74 Años	Ambos	36,64265354	65.533,24	2.401.311,89
Entre 75 y 79 Años	Ambos	44,78332012	83.285,79	3.729.814,40
Mayores de 80 Años	Ambos	50,35117086	59.045,72	2.973.020,89

Tabla 62: Prima Pura Zona Alejada Año 2019

El valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares para lo anterior se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones, para lo anterior compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona alejada, adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.

Resumen de contrastes de hipótesis				
	Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1	La mediana de diferencias entre EPS_SELEC y TOTAL_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,653	Conserve la hipótesis nula.

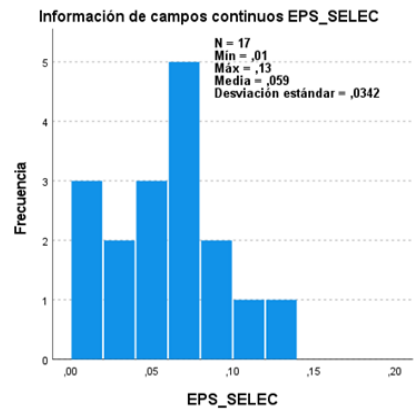
a. El nivel de significación es de ,050.  
b. Se muestra la significancia asintótica.

Figura 187: Prueba Wilcoxon Zona Alejada Año 2019

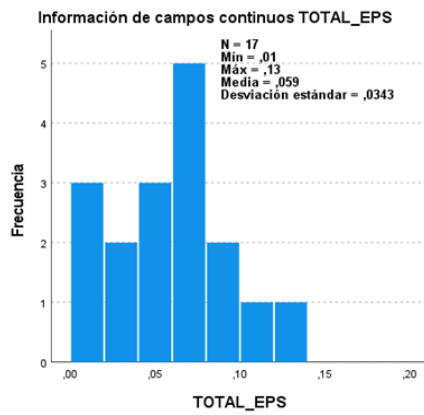
Teniendo como referencia la prima pura establecida para el año 2018 a través de la Resolución 5268 de 2017 y para el año 2019 a través de la Resolución 5858 de 2019, se establece el factor de relatividad para cada uno de los grupos etarios, los cuales se presentan a continuación en la tabla numero 63 y la gráfica 189, como se observa en presentan diferencias entre los años, la principal razón en la cantidad de actividades reportadas por las EPS presentes en la isla,



(a) Prueba de Rangos Zona Alejada Año 2019



(b) EPS Selección Zona Alejada Año 2019



(c) Total EPS Zona Alejada Año 2019

Descripción Grupo Etario	Sexo	Factor 2018	Factor 2019
Menor de un año	Ambos	0,8826	0,9055
Entre 1 y 5 Años	Ambos	0,6504	0,7803
Entre 6 y 11 Años	Ambos	0,3126	0,3839
Entre 12 y 17 Años	Femenino	0,2759	0,3575
Entre 12 y 17 Años	Masculino	0,2943	0,3207
Entre 18 y 28 Años	Femenino	0,3942	0,5861
Entre 18 y 28 Años	Masculino	0,4228	0,5259
Entre 29 y 44 Años	Femenino	0,5228	0,7632
Entre 28 y 44 Años	Masculino	0,5596	0,6847
Entre 45 y 49 Años	Ambos	0,6786	0,9744
Entre 50 y 54 Años	Ambos	0,8635	1,4205
Entre 55 y 59 Años	Ambos	1,21	1,6071
Entre 60 y 64 Años	Ambos	1,3518	1,9559
Entre 65 y 69 Años	Ambos	1,9043	2,6552
Entre 70 y 74 Años	Ambos	2,5844	3,1494
Entre 75 y 79 Años	Ambos	2,5308	4,8918
Mayores de 80 Años	Ambos	2,51	3,8992

Tabla 63: Factores de Relatividad Zona Alejada Años 2018 y 2019

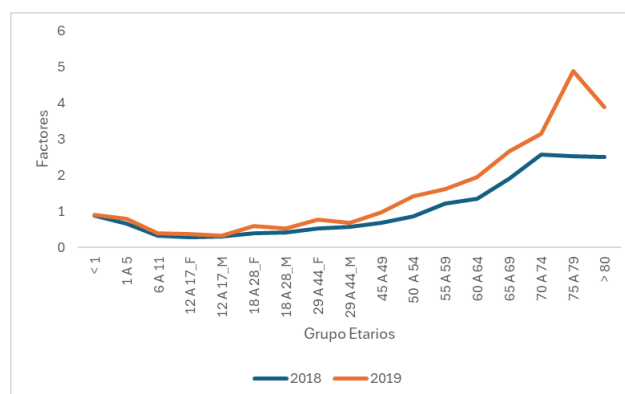


Figura 189: Factores de Relatividad Zona Alejada Años 2018 y 2019

Al comparar las frecuencias entre los años 2018 y 2019 se presenta un incremento considerable en todos los grupos etarios en el año 2019, lo anterior se presenta en las figuras números 190a y 190b.

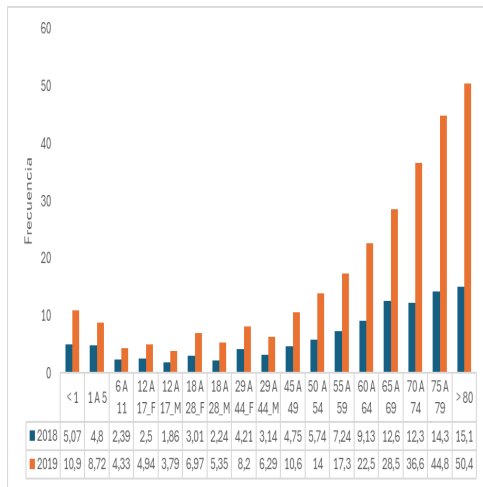
Por otro lado la severidad o costo promedio, presenta una disminución en todos los grupos etarios en el año 2019 comparado con el año 2018, lo anterior se presenta en las figuras 191a y 191b

#### 4.4.2. Zona Bogotá

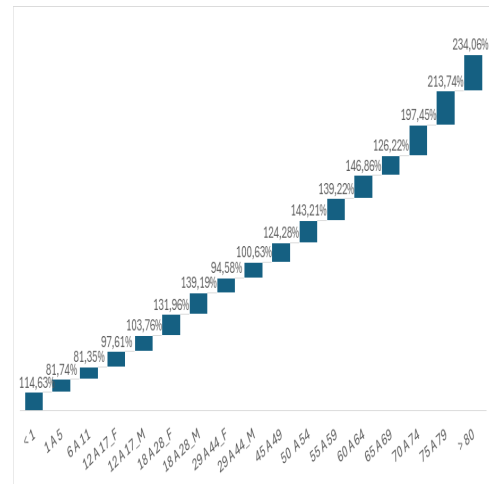
A continuación se presenta la prima pura correspondiente al año 2018 para cada uno de los grupos etarios para la Zona Bogotá, teniendo en cuenta que la misma se expresa como:

$$PrimaPura = \frac{Pérdidas}{Número \text{ de expuestos}} = Frecuencia * Severidad$$

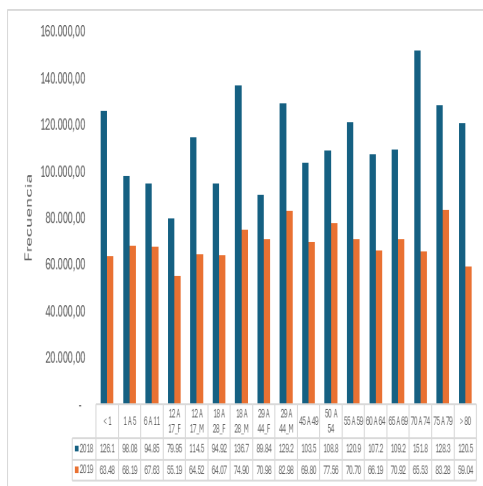
Es importante tener en cuenta que la frecuencia en el año 2018 sigue una distribución binomial negativa y la severidad una distribución lognormal.



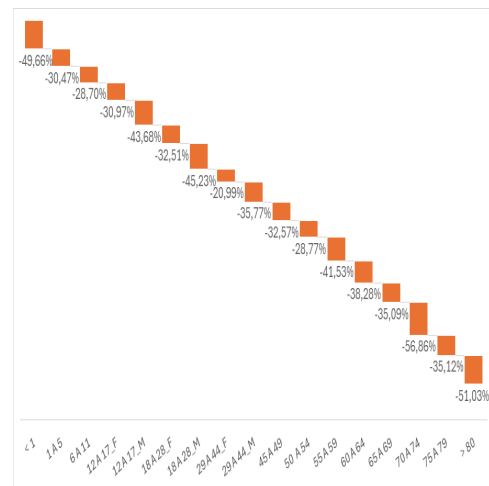
(a) Frecuencia Zona Alejada Años 2018 - 2019



(b) Variación Frecuencia Zona Alejada 2018 - 2019



(a) Severidad Zona Alejada Años 2018 - 2019



(b) Variación Severidad Zona Alejada 2018 - 2019

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	14,98050919	89.706,31	1.343.846,13
Entre 1 y 5 Años	Ambos	10,95133973	45.285,58	495.937,79
Entre 6 y 11 Años	Ambos	6,173606276	41.842,28	258.317,76
Entre 12 y 17 Años	Femenino	7,810542814	42.083,91	328.698,22
Entre 12 y 17 Años	Masculino	5,87125591	53.389,84	313.465,40
Entre 18 y 28 Años	Femenino	9,15711801	39.245,94	359.379,72
Entre 18 y 28 Años	Masculino	6,883488705	49.711,81	342.190,69
Entre 29 y 44 Años	Femenino	11,32886365	44.249,63	501.298,04
Entre 29 y 44 Años	Masculino	8,516009611	56.200,37	478.602,90
Entre 45 y 49 Años	Ambos	13,2736456	51.304,78	681.001,41
Entre 50 y 54 Años	Ambos	17,01012008	50.604,34	860.785,86
Entre 55 y 59 Años	Ambos	21,70344295	52.077,71	1.130.265,55
Entre 60 y 64 Años	Ambos	29,24216148	51.012,78	1.491.723,89
Entre 65 y 69 Años	Ambos	38,62471011	50.477,82	1.949.691,05
Entre 70 y 74 Años	Ambos	48,63085013	50.988,69	2.479.623,28
Entre 75 y 79 Años	Ambos	58,23922266	52.334,61	3.047.927,06
Mayores de 80 Años	Ambos	67,92465273	55.221,95	3.750.932,04

Tabla 64: Prima Pura Zona Bogotá Año 2018

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; Para lo anterior compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona alejada. Adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.

Resumen de contrastes de hipótesis				
	Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1	La mediana de diferencias entre EPS_SELECCION y TOTAL_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,795	Conserve la hipótesis nula.

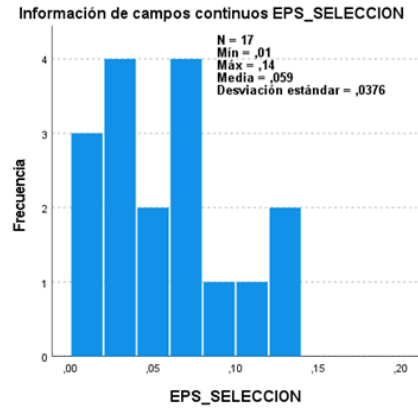
a. El nivel de significación es de ,050.  
b. Se muestra la significancia asintótica.

Figura 192: Prueba Wilcoxon Zona Bogotá Año 2018

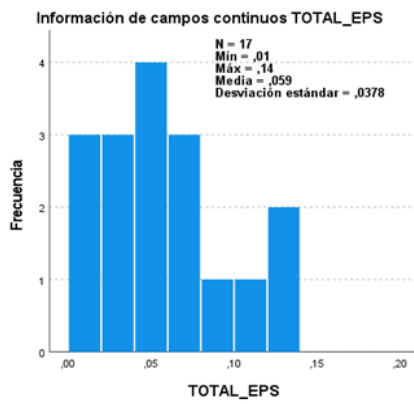
A continuación se presenta la prima pura correspondiente al año 2019 para cada uno de los grupos etarios, es importante tener en cuenta que la frecuencia en el año 2019 sigue una distribución binomial negativa y la severidad una distribución lognormal.



(a) Prueba de Rangos Zona Bogotá Año 2018



(b) EPS Selección Zona Bogotá Año 2018



(c) Total EPS Zona Bogotá Año 2018

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	14,49997949	87.637,88	1.270.747,48
Entre 1 y 5 Años	Ambos	11,5093166	42.339,74	487.301,52
Entre 6 y 11 Años	Ambos	6,435540118	43.147,91	277.680,07
Entre 12 y 17 Años	Femenino	7,978678788	41.900,42	334.309,96
Entre 12 y 17 Años	Masculino	6,01844647	52.177,29	314.026,24
Entre 18 y 28 Años	Femenino	9,6163685	40.600,49	390.429,29
Entre 18 y 28 Años	Masculino	7,253782311	50.525,71	366.502,49
Entre 29 y 44 Años	Femenino	11,95850302	44.870,94	536.589,24
Entre 29 y 44 Años	Masculino	9,020492268	55.955,32	504.744,57
Entre 45 y 49 Años	Ambos	14,35064748	50.376,74	722.938,81
Entre 50 y 54 Años	Ambos	18,72013222	49.494,69	926.547,09
Entre 55 y 59 Años	Ambos	24,02732382	48.302,98	1.160.591,32
Entre 60 y 64 Años	Ambos	32,62099855	48.477,99	1.581.400,58
Entre 65 y 69 Años	Ambos	43,31487472	47.439,35	2.054.829,29
Entre 70 y 74 Años	Ambos	54,87393328	47.183,80	2.589.160,95
Entre 75 y 79 Años	Ambos	66,16333981	48.405,67	3.202.680,86
Mayores de 80 Años	Ambos	75,83997976	49.284,32	3.737.722,15

Tabla 65: Prima Pura Zona Bogotá Año 2019

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones. Para lo anterior compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona alejada, adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.

Resumen de contrastes de hipótesis			
Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1 La mediana de diferencias entre EPS_SELECCION y TODAS_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,332	Conserve la hipótesis nula.

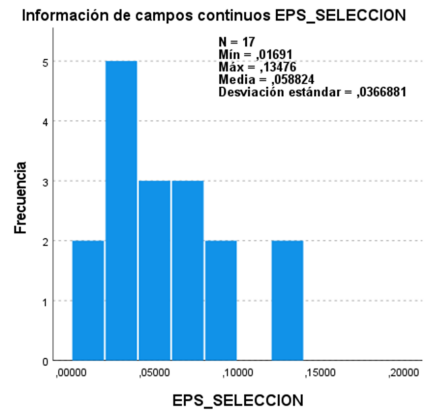
a. El nivel de significación es de ,050.  
b. Se muestra la significancia asintótica.

Figura 194: Prueba Wilcoxon Zona Bogotá Año 2019

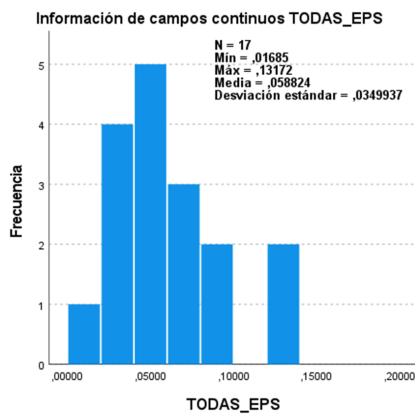
Teniendo como referencia la prima pura establecida para el año 2018 a través de la Resolución 5268 de 2017 y para el año 2019 a través de la Resolución 5858 de 2019, se establece el factor de relatividad para cada uno de los grupos etarios, los cuales se presentan a continuación en la tabla número 78 y la gráfica 196, no presenta cambios significativos en la composición de los factores por grupo de edad.



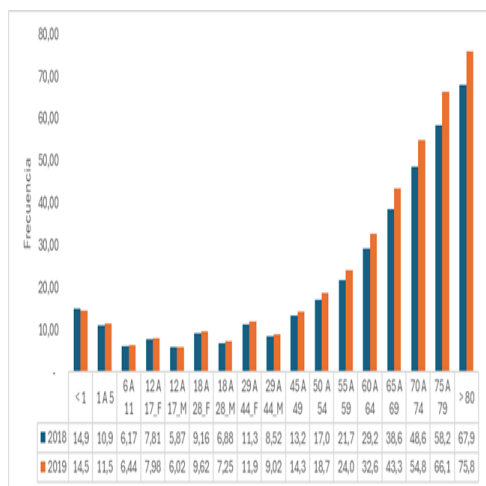
(a) Prueba de Rangos Zona Bogotá Año 2019



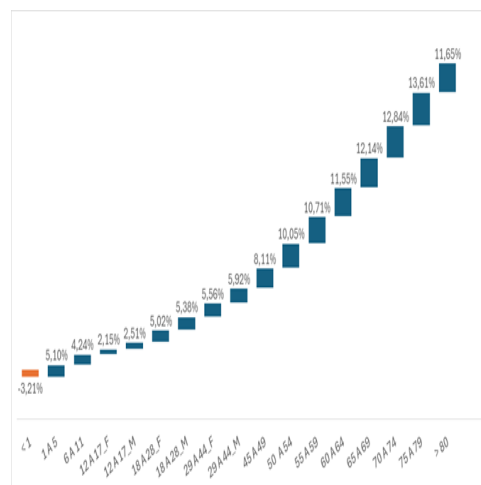
(b) EPS Selección Zona Bogotá Año 2019



(c) Total EPS Zona Bogotá Año 2019



(a) Frecuencia Zona Bogotá Años 2018 - 2019



(b) Variación Frecuencia Zona Bogotá 2018 - 2019

Descripción Grupo Etario	Sexo	Factor 2018	Factor 2019
Menor de un año	Ambos	1,8561	1,6666
Entre 1 y 5 Años	Ambos	0,6850	0,6391
Entre 6 y 11 Años	Ambos	0,3568	0,3642
Entre 12 y 17 Años	Femenino	0,4540	0,4385
Entre 12 y 17 Años	Masculino	0,4330	0,4119
Entre 18 y 28 Años	Femenino	0,4964	0,5121
Entre 18 y 28 Años	Masculino	0,4726	0,4807
Entre 29 y 44 Años	Femenino	0,6924	0,7038
Entre 28 y 44 Años	Masculino	0,6610	0,6620
Entre 45 y 49 Años	Ambos	0,9406	0,9482
Entre 50 y 54 Años	Ambos	1,1889	1,2152
Entre 55 y 59 Años	Ambos	1,5611	1,5222
Entre 60 y 64 Años	Ambos	2,0603	2,0741
Entre 65 y 69 Años	Ambos	2,6929	2,6950
Entre 70 y 74 Años	Ambos	3,4248	3,3958
Entre 75 y 79 Años	Ambos	4,2097	4,2004
Mayores de 80 Años	Ambos	5,1807	4,9022

Tabla 66: Factores de Relatividad Zona Bogotá Años 2018 y 2019

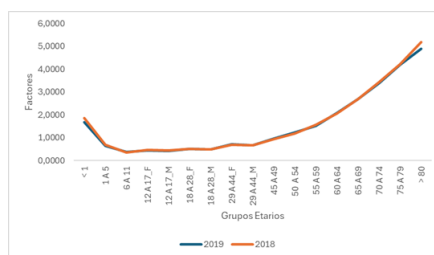
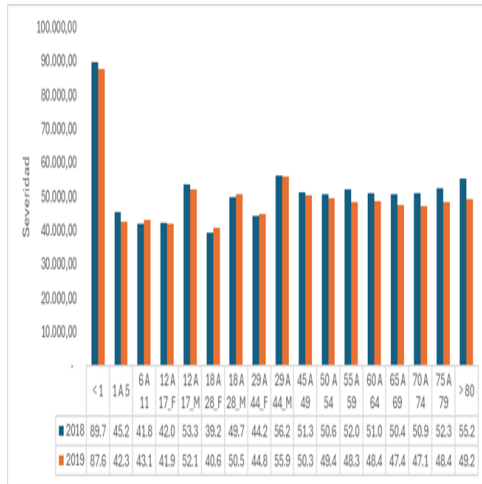


Figura 196: Factores de Relatividad Zona Bogotá Años 2018 y 2019

Al comparar las frecuencias entre los años 2018 y 2019 se presenta un incremento en todos los grupos etarios en el año 2019 con excepción del grupo etario menores de un año que presenta una disminución del 2.71%, lo anterior se presenta en las figuras números 197a y 197b.

Por otro lado, la severidad o costo promedio, en 13 de los 17 grupos etario, se presenta una disminución en el año 2019 comparado con el año 2018; lo anterior se presenta en las figuras 198a y 198b



(a) Frecuencia Zona Bogotá Años 2018 - 2019



(b) Variación Frecuencia Zona Bogota 2018 - 2019

#### 4.4.3. Zona Normal Mayor

A continuación se presenta la prima pura correspondiente al año 2018 para cada uno de los grupos etarios de la Zona Normal Mayor, teniendo en cuenta que la misma se expresa como:

$$PrimaPura = \frac{Pérdidas}{\text{Número de expuestos}} = Frecuencia * Severidad$$

Es importante tener en cuenta que la frecuencia en el año 2018 sigue una distribución binomial negativa y la severidad una distribución lognormal.

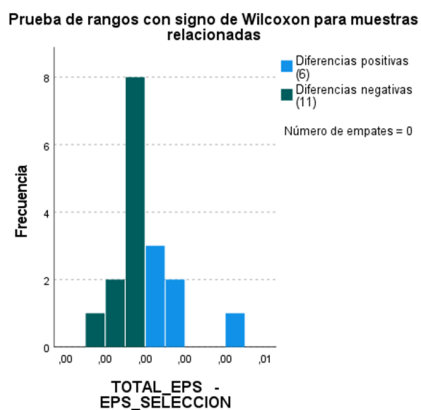
Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	13,87297451	72.972,60	1.012.347,07
Entre 1 y 5 Años	Ambos	11,2483084	41.796,26	470.137,23
Entre 6 y 11 Años	Ambos	6,020078274	45.499,66	273.911,50
Entre 12 y 17 Años	Femenino	7,354666837	46.521,37	342.149,21
Entre 12 y 17 Años	Masculino	5,503066322	57.270,53	315.163,51
Entre 18 y 28 Años	Femenino	8,778525138	41.889,09	367.724,46
Entre 18 y 28 Años	Masculino	6,568456072	51.463,41	338.035,16
Entre 29 y 44 Años	Femenino	11,50333857	44.514,25	512.062,50
Entre 29 y 44 Años	Masculino	8,607274329	54.752,92	471.273,37
Entre 45 y 49 Años	Ambos	13,6700743	50.858,00	695.232,59
Entre 50 y 54 Años	Ambos	18,04391906	50.338,47	908.303,28
Entre 55 y 59 Años	Ambos	23,24630761	49.506,91	1.150.852,95
Entre 60 y 64 Años	Ambos	30,82137123	48.540,13	1.496.073,51
Entre 65 y 69 Años	Ambos	40,13784769	48.263,67	1.937.200,02
Entre 70 y 74 Años	Ambos	48,94810156	48.590,34	2.378.404,75
Entre 75 y 79 Años	Ambos	57,14443519	49.869,50	2.849.764,28
Mayores de 80 Años	Ambos	64,59195982	52.181,76	3.370.522,05

Tabla 67: Prima Pura Zona Normal Mayor Año 2018

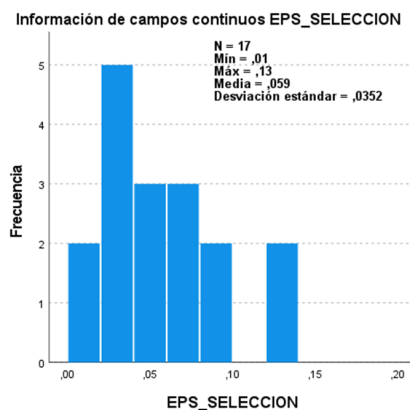
De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D’Abrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; para lo anterior compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

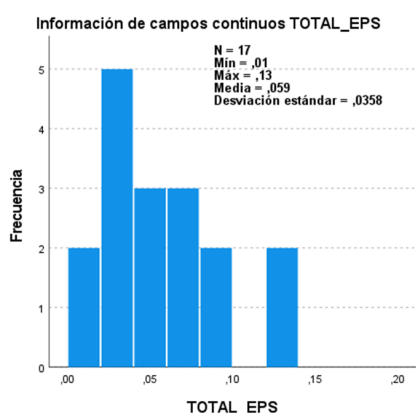
La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados,



(a) Prueba de Rangos Zona Normal Mayor Año 2018



(b) EPS Selección Zona Normal Mayor Año 2018



(c) Total EPS Zona Normal Mayor Año 2018

conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona normal mayor. Adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.

Resumen de contrastes de hipótesis				
	Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1	La mediana de diferencias entre EPS_SELECCION y TOTAL_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,381	Conserve la hipótesis nula.

a. El nivel de significación es de ,050.  
b. Se muestra la significancia asintótica.

Figura 199: Prueba Wilcoxon Zona Normal Mayor Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Normal Mayor para cada uno de los grupos etarios, la distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	14,00901086	67.796,31	949.759,29
Entre 1 y 5 Años	Ambos	12,37599654	37.613,42	465.503,59
Entre 6 y 11 Años	Ambos	6,687559423	41.224,32	275.690,06
Entre 12 y 17 Años	Femenino	8,302629657	40.069,08	332.678,74
Entre 12 y 17 Años	Masculino	6,286446258	49.324,16	310.073,67
Entre 18 y 28 Años	Femenino	9,667558982	37.249,31	360.109,88
Entre 18 y 28 Años	Masculino	7,31992061	45.787,52	335.161,01
Entre 29 y 44 Años	Femenino	12,66442424	40.554,95	513.605,13
Entre 29 y 44 Años	Masculino	9,589036926	49.934,06	478.819,51
Entre 45 y 49 Años	Ambos	15,36471116	45.158,54	693.847,99
Entre 50 y 54 Años	Ambos	20,35967053	44.653,58	909.132,20
Entre 55 y 59 Años	Ambos	26,25981272	44.676,64	1.173.200,28
Entre 60 y 64 Años	Ambos	34,92867248	44.661,02	1.559.950,28
Entre 65 y 69 Años	Ambos	45,84037457	44.163,57	2.024.474,67
Entre 70 y 74 Años	Ambos	56,18603756	44.020,07	2.473.313,15
Entre 75 y 79 Años	Ambos	66,31939447	45.297,48	3.004.101,34
Mayores de 80 Años	Ambos	75,40292146	47.442,38	3.577.293,79

Tabla 68: Prima Pura Zona Normal Mayor Año 2019

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; para lo anterior compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

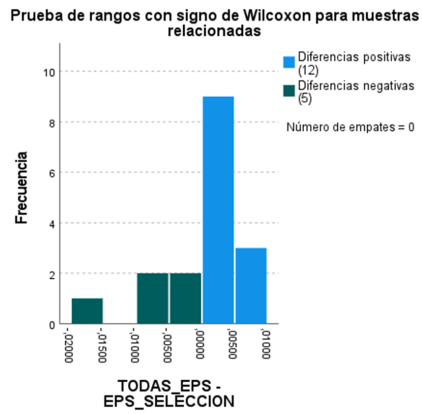
A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona alejada. Adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.

Resumen de contrastes de hipótesis			
Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1 La mediana de diferencias entre EPS_SELECCION y TODAS_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,381	Conserve la hipótesis nula.

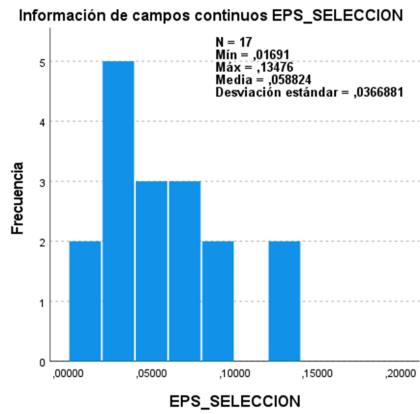
a. El nivel de significación es de ,050.  
b. Se muestra la significancia asintótica.

Figura 201: Prueba Wilcoxon Zona Normal Mayor Año 2019

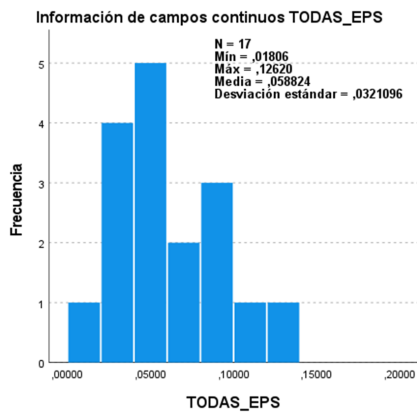
Teniendo como referencia la prima pura establecida para el año 2018 a través de la Resolución 5268 de 2017 y para el año 2019 a través de la Resolución 5858 de 2019, se establece el factor de relatividad para cada uno de los grupos etarios, los cuales se presentan a continuación en la tabla numero 69 y la gráfica 203, como se observa no presenta cambios significativos en la composición de los factores por grupo de edad.



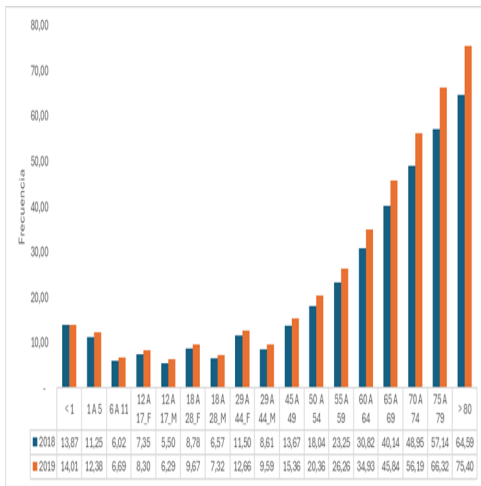
(a) Prueba de Rangos Zona Normal Mayor Año 2019



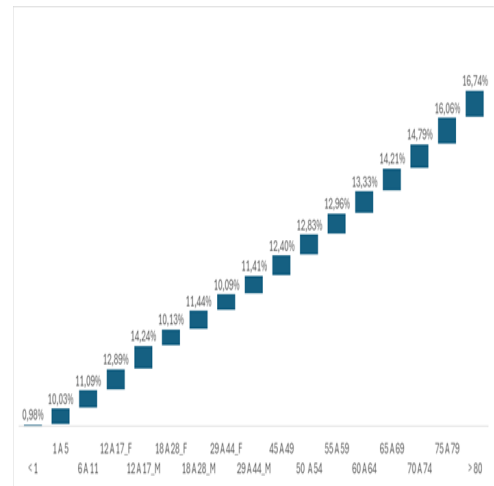
(b) EPS Selección Zona Normal Mayor Año 2019



(c) Total EPS Zona Normal Mayor Año 2019



(a) Frecuencia Zona Normal Mayor Años 2018 - 2019



(b) Variación Frecuencia Zona Normal Mayor 2018 - 2019

Descripción Grupo Etario	Sexo	Factor 2018	Factor 2019
Menor de un año	Ambos	1,3982	1,2456
Entre 1 y 5 Años	Ambos	0,6493	0,6105
Entre 6 y 11 Años	Ambos	0,3783	0,3616
Entre 12 y 17 Años	Femenino	0,4726	0,4363
Entre 12 y 17 Años	Masculino	0,4353	0,4067
Entre 18 y 28 Años	Femenino	0,5079	0,4723
Entre 18 y 28 Años	Masculino	0,4669	0,4396
Entre 29 y 44 Años	Femenino	0,7073	0,6736
Entre 28 y 44 Años	Masculino	0,6509	0,6280
Entre 45 y 49 Años	Ambos	0,9602	0,9100
Entre 50 y 54 Años	Ambos	1,2545	1,1924
Entre 55 y 59 Años	Ambos	1,5895	1,5387
Entre 60 y 64 Años	Ambos	2,0664	2,0459
Entre 65 y 69 Años	Ambos	2,6756	2,6552
Entre 70 y 74 Años	Ambos	3,2850	3,2438
Entre 75 y 79 Años	Ambos	3,9360	3,9400
Mayores de 80 Años	Ambos	4,6553	4,6918

Tabla 69: Factores de Relatividad Zona Normal Mayor Años 2018 y 2019

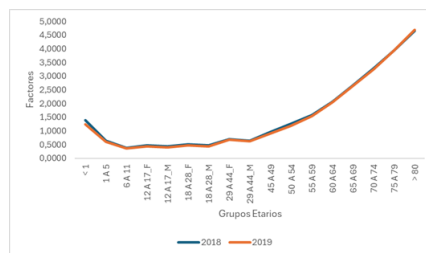
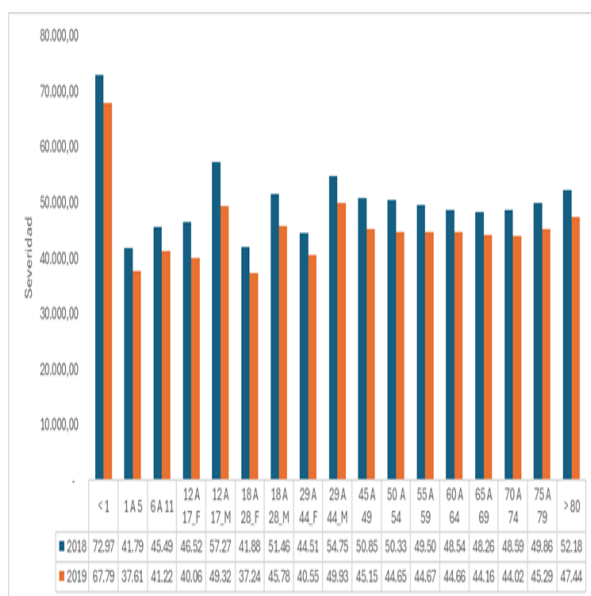


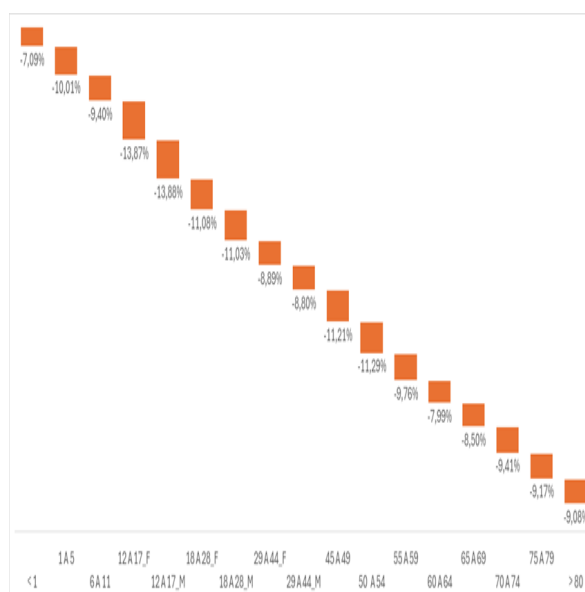
Figura 203: Factores de Relatividad Zona Normal Mayor Años 2018 y 2019

Al comparar las frecuencias entre los años 2018 y 2019 se presenta un incremento en todos los grupos etarios en el año 2019, lo anterior se presenta en las figuras números 204a y 204b.

Por otro lado, la severidad o costo promedio presenta una disminución en todos los grupos etarios en el año 2019 comparado con el año 2018; lo anterior se presenta en las figuras 205a y 205b



(a) Frecuencia Zona Normal Mayor Años 2018 - 2019



(b) Variación Frecuencia Zona Normal Mayor 2018 - 2019

#### 4.4.4. Zona Normal Medio

A continuación se presenta la prima pura correspondiente al año 2018 para cada uno de los grupos etarios para la Zona Normal Menor, teniendo en cuenta que la misma se expresa como:

$$PrimaPura = \frac{Pérdidas}{Número\ de\ expuestos} = Frecuencia * Severidad$$

Es importante tener en cuenta que la distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	11,83180792	76.508,31	905.231,61
Entre 1 y 5 Años	Ambos	9,568527471	38.007,58	363.676,53
Entre 6 y 11 Años	Ambos	5,241925231	40.134,37	210.381,39
Entre 12 y 17 Años	Femenino	6,465724929	44.279,15	286.296,81
Entre 12 y 17 Años	Masculino	4,586877588	54.338,52	249.244,15
Entre 18 y 28 Años	Femenino	7,084247071	43.773,96	310.105,58
Entre 18 y 28 Años	Masculino	5,025666027	53.706,77	269.912,30
Entre 29 y 44 Años	Femenino	9,440184133	45.776,19	432.135,67
Entre 29 y 44 Años	Masculino	6,697001418	56.211,41	376.447,92
Entre 45 y 49 Años	Ambos	10,93517186	50.870,99	556.283,05
Entre 50 y 54 Años	Ambos	14,43408616	51.897,31	749.090,22
Entre 55 y 59 Años	Ambos	18,71288113	51.589,75	965.392,95
Entre 60 y 64 Años	Ambos	24,98939216	52.128,97	1.302.671,27
Entre 65 y 69 Años	Ambos	32,39837649	51.503,89	1.668.642,34
Entre 70 y 74 Años	Ambos	39,88375874	52.302,84	2.086.033,92
Entre 75 y 79 Años	Ambos	46,96166731	55.205,92	2.592.561,94
Mayores de 80 Años	Ambos	51,23699178	57.526,16	2.947.467,23

Tabla 70: Prima Pura Zona Normal Medio Año 2018

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona normal mayor, Adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.

<b>Resumen de contrastes de hipótesis</b>				
	Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1	La mediana de diferencias entre EPS_SELECCION y TOTAL_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,381	Conserve la hipótesis nula.

a. El nivel de significación es de ,050.

b. Se muestra la significancia asintótica.

Figura 206: Prueba Wilcoxon Zona Normal Medio Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Normal Menor para cada uno de los grupos etarios, la distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	12,31013727	60.815,54	748.647,63
Entre 1 y 5 Años	Ambos	10,62408868	32.055,68	340.562,34
Entre 6 y 11 Años	Ambos	5,881374812	34.582,14	203.390,53
Entre 12 y 17 Años	Femenino	7,342190408	36.811,29	270.275,47
Entre 12 y 17 Años	Masculino	5,253050303	45.301,21	237.969,52
Entre 18 y 28 Años	Femenino	7,726614495	37.593,70	290.472,02
Entre 18 y 28 Años	Masculino	5,528090714	46.283,28	255.858,17
Entre 29 y 44 Años	Femenino	10,41379827	40.551,76	422.297,85
Entre 28 y 44 Años	Masculino	7,45066569	49.999,78	372.531,65
Entre 45 y 49 Años	Ambos	12,19641853	45.142,36	550.575,11
Entre 50 y 54 Años	Ambos	16,01066767	45.531,09	728.983,15
Entre 55 y 59 Años	Ambos	20,66848629	46.608,16	963.320,14
Entre 60 y 64 Años	Ambos	27,68663933	47.584,63	1.317.458,47
Entre 65 y 69 Años	Ambos	36,19554964	47.345,48	1.713.695,50
Entre 70 y 74 Años	Ambos	44,49502632	47.711,94	2.122.943,88
Entre 75 y 79 Años	Ambos	52,99645782	48.911,65	2.592.144,31
Mayores de 80 Años	Ambos	59,12786761	50.262,35	2.971.905,73

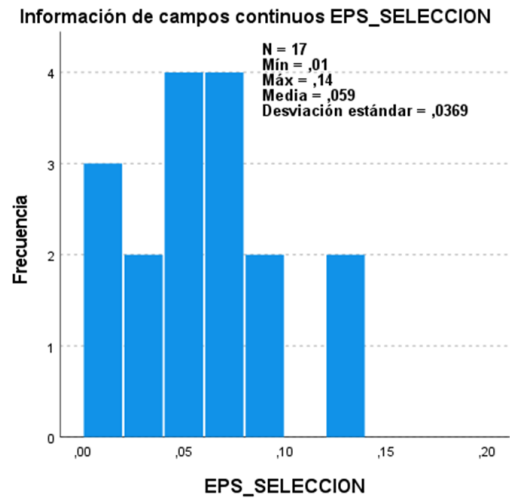
Tabla 71: Prima Pura Zona Normal Menor Año 2019

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; para lo anterior compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

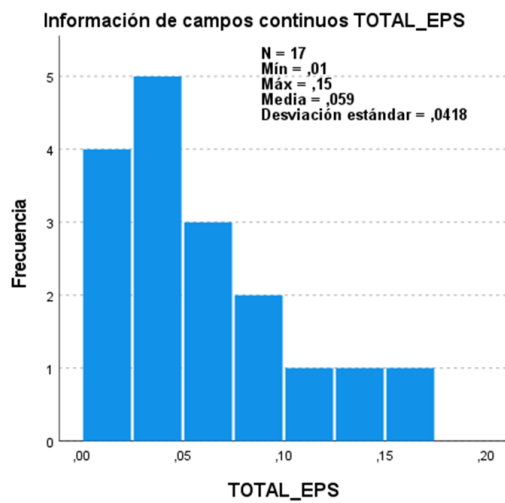
La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección cumplen con similitud con el total de afiliados en la zona alejada. Adicional a la prueba mencionada se presenta visualmente el comportamiento descrito.

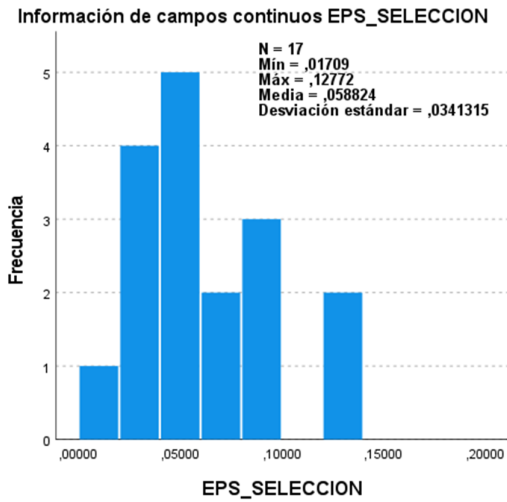
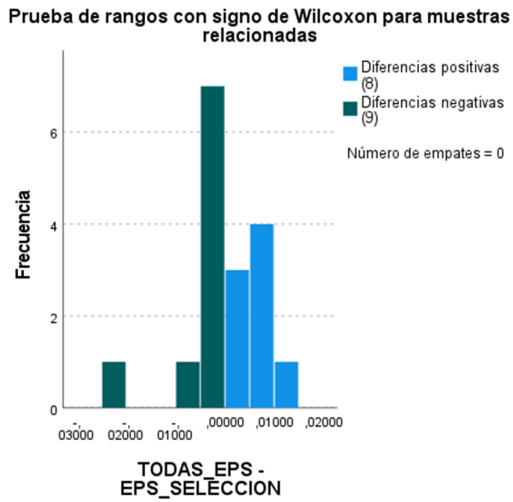


(a) Prueba de Rangos Zona Normal Medio Año 2018

(b) EPS Selección Zona Normal Medio Año 2018

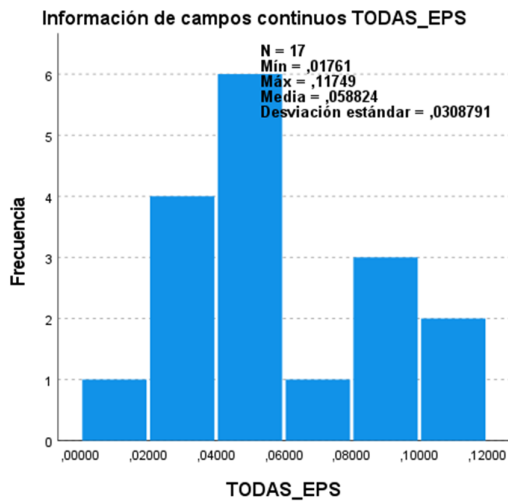


(c) Total EPS Zona Normal Medio Año 2018



(a) Prueba de Rangos Zona Normal Menor Año 2019

(b) EPS Selección Zona Normal Menor Año 2019



(c) Total EPS Zona Normal Menor Año 2019

**Resumen de contrastes de hipótesis**

Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1 La mediana de diferencias entre EPS_SELECCION y TODAS_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,758	Conserve la hipótesis nula.

a. El nivel de significación es de ,050.

b. Se muestra la significancia asintótica.

Figura 208: Prueba Wilcoxon Zona Normal Menor Año 2019

Teniendo como referencia la prima pura establecida para el año 2018 a través de la Resolución 5268 de 2017 y para el año 2019 a través de la Resolución 5858 de 2019, se establece el factor de relatividad para

cada uno de los grupos etarios, los cuales se presentan a continuación en la tabla numero 72 y la gráfica ??, como se observa no presenta cambios significativos en la composición de los factores por grupo de edad.

Descripción Grupo Etario	Sexo	Factor 2018	Factor 2019
Menor de un año	Ambos	1,2503	0,9819
Entre 1 y 5 Años	Ambos	0,5023	0,4467
Entre 6 y 11 Años	Ambos	0,2906	0,2668
Entre 12 y 17 Años	Femenino	0,3954	0,3545
Entre 12 y 17 Años	Masculino	0,3443	0,3121
Entre 18 y 28 Años	Femenino	0,4283	0,3810
Entre 18 y 28 Años	Masculino	0,3728	0,3356
Entre 29 y 44 Años	Femenino	0,5969	0,5539
Entre 28 y 44 Años	Masculino	0,5199	0,4886
Entre 45 y 49 Años	Ambos	0,7683	0,7221
Entre 50 y 54 Años	Ambos	1,0346	0,9561
Entre 55 y 59 Años	Ambos	1,3334	1,2634
Entre 60 y 64 Años	Ambos	1,7992	1,7279
Entre 65 y 69 Años	Ambos	2,3047	2,2476
Entre 70 y 74 Años	Ambos	2,8812	2,7843
Entre 75 y 79 Años	Ambos	3,5808	3,3997
Mayores de 80 Años	Ambos	4,0710	3,8978

Tabla 72: Factores de Relatividad Zona Normal Menor Años 2018 y 2019

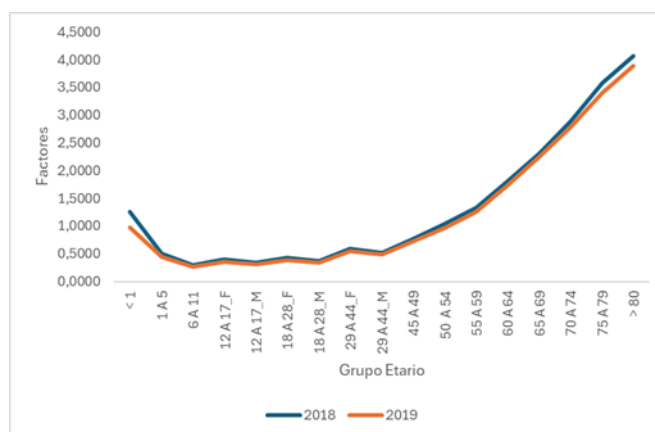


Figura 210: Factores de Relatividad Zona Normal Menor Años 2018 y 2019

Al comparar las frecuencias entre los años 2018 y 2019 se presenta un incremento en todos los grupos etarios en el año 2019, lo anterior se presenta en las figuras números 211a y 211b.

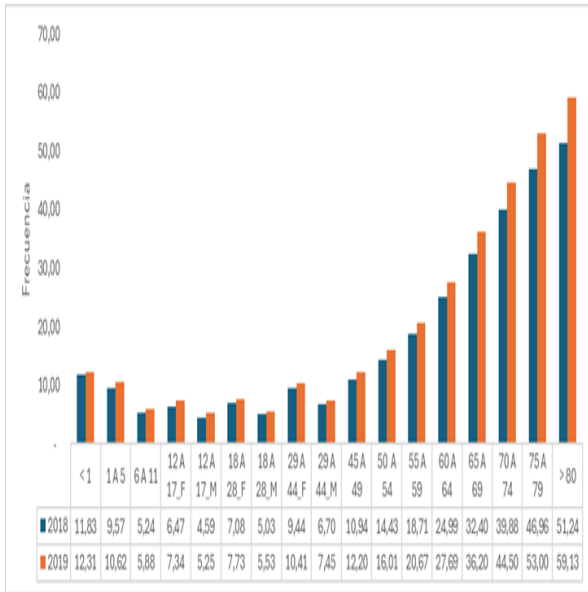
Por otro lado, la severidad o costo promedio presenta una disminución en todos los grupos etarios en el año 2019 comparado con el año 2018; lo anterior se presenta en las figuras 212a y 212b

#### 4.4.5. Zona Rural

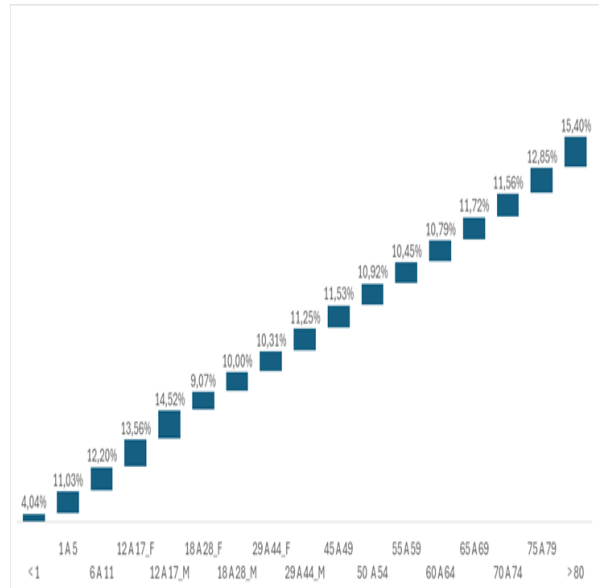
A continuación se presenta la prima pura correspondiente al año 2018 para cada uno de los grupos etarios para la Zona Rural, teniendo en cuenta que la misma se expresa como:

$$PrimaPura = \frac{Pérdidas}{Número\ de\ expuestos} = Frecuencia * Severidad$$

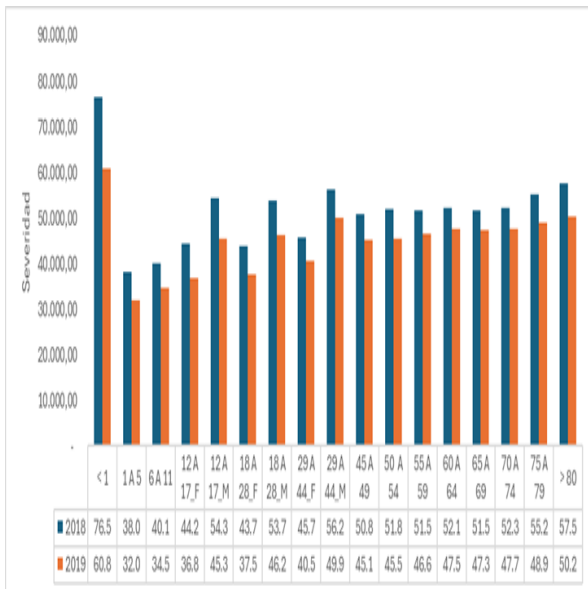
La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.



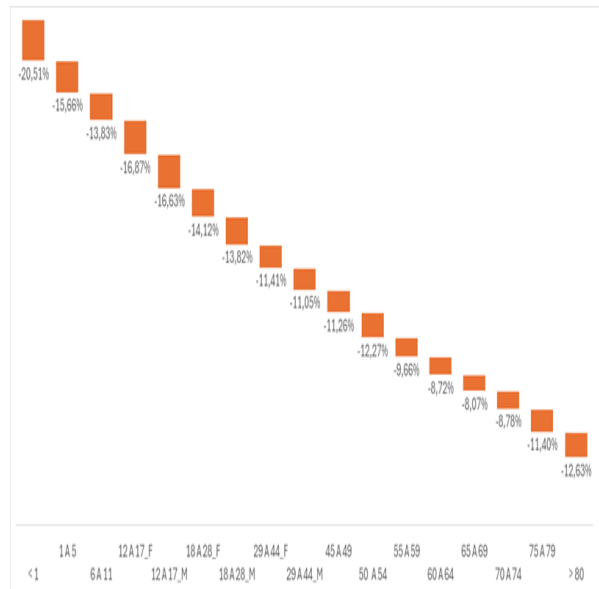
(a) Frecuencia Zona Normal Mayor Años 2018 - 2019



(b) Variación Frecuencia Zona Normal Mayor 2018 - 2019



(a) Frecuencia Zona Normal Menor Años 2018 - 2019



(b) Variación Frecuencia Zona Normal Menor 2018 - 2019

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	8,159177166	88.445,79	721.644,90
Entre 1 y 5 Años	Ambos	6,100542513	47.370,91	288.988,25
Entre 6 y 11 Años	Ambos	3,214646502	49.242,99	158.298,81
Entre 12 y 17 Años	Femenino	4,228367405	54.344,65	229.789,14
Entre 12 y 17 Años	Masculino	3,006411451	66.011,45	198.457,59
Entre 18 y 28 Años	Femenino	4,054291277	55.586,04	225.361,99
Entre 18 y 28 Años	Masculino	2,192985431	67.546,56	148.128,61
Entre 29 y 44 Años	Femenino	5,726152896	56.003,77	320.686,18
Entre 29 y 44 Años	Masculino	3,179743368	68.063,27	216.423,72
Entre 45 y 49 Años	Ambos	6,204242406	63.797,87	395.817,43
Entre 50 y 54 Años	Ambos	8,204422811	63.169,11	518.266,07
Entre 55 y 59 Años	Ambos	11,21768517	62.241,901	698.210,20
Entre 60 y 64 Años	Ambos	15,57701843	65.926,76	1.026.942,40
Entre 65 y 69 Años	Ambos	21,10235762	65.330,42	1.378.625,88
Entre 70 y 74 Años	Ambos	26,40251732	63.063,96	1.665.047,25
Entre 75 y 79 Años	Ambos	30,93592832	65.305,70	2.020.292,52
Mayores de 80 Años	Ambos	33,03680823	68.751,13	2.271.317,92

Tabla 73: Prima Pura Zona Rural Año 2018

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona normal mayor. Adicional a la prueba mencionada, se presenta visualmente el comportamiento descrito.

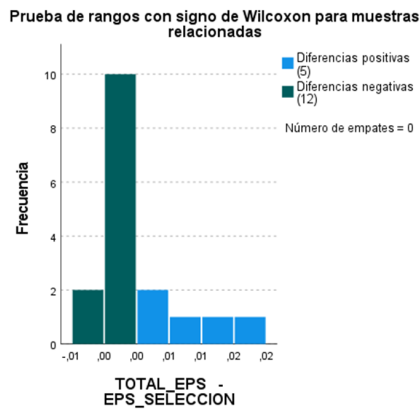
Resumen de contrastes de hipótesis				
	Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1	La mediana de diferencias entre EPS_SELECCION y TOTAL_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,381	Conserve la hipótesis nula.

a. El nivel de significación es de ,050.

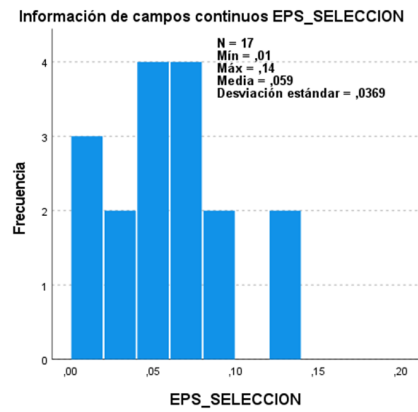
b. Se muestra la significancia asintótica.

Figura 213: Prueba Wilcoxon Zona Normal Rural Año 2018

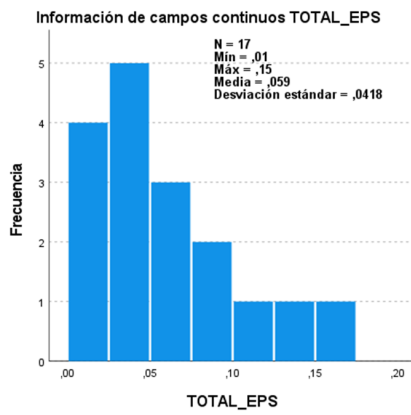
A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Rural para cada uno de los grupos etarios, la distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.



(a) Prueba de Rangos Zona Rural Año 2018



(b) EPS Selección Zona Rural Año 2018



(c) Total EPS Zona Rural Año 2018

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	9,425790751	56.636,53	533.844,11
Entre 1 y 5 Años	Ambos	7,249369892	32.391,97	234.821,34
Entre 6 y 11 Años	Ambos	4,036756237	36.457,83	147.171,36
Entre 12 y 17 Años	Femenino	5,384514165	39.078,47	210.418,58
Entre 12 y 17 Años	Masculino	3,541398907	48.411,15	171.443,18
Entre 18 y 28 Años	Femenino	4,197538736	40.004,82	167.921,78
Entre 18 y 28 Años	Masculino	2,760724299	49.582,25	136.882,92
Entre 29 y 44 Años	Femenino	6,254184629	43.571,50	272.504,19
Entre 29 y 44 Años	Masculino	4,113381808	54.096,31	222.518,77
Entre 45 y 49 Años	Ambos	7,3256377	49.620,29	363.500,25
Entre 50 y 54 Años	Ambos	9,736954563	50.859,64	495.217,98
Entre 55 y 59 Años	Ambos	13,16232009	51.730,23	680.889,81
Entre 60 y 64 Años	Ambos	18,43973378	53.263,33	982.161,61
Entre 65 y 69 Años	Ambos	25,01095698	52.545,66	1.314.217,35
Entre 70 y 74 Años	Ambos	32,22278406	52.560,52	1.693.646,16
Entre 75 y 79 Años	Ambos	38,58495814	51.587,60	1.990.505,46
Mayores de 80 Años	Ambos	43,57565831	51.107,55	2.227.044,93

Tabla 74: Prima Pura Zona Rural Año 2019

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

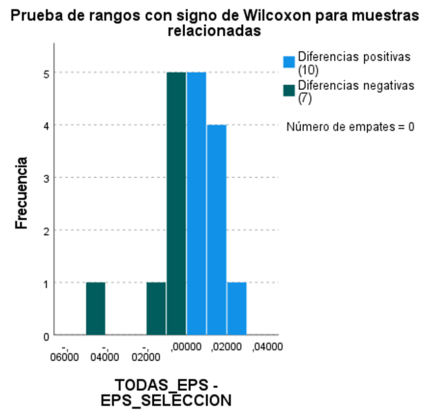
A continuación, se presentan los resultados, concluyendo que las EPS de la selección cumplen con similitud con el total de afiliados en la zona alejada. Adicional a la prueba mencionada, se presenta visualmente el comportamiento descrito.

Resumen de contrastes de hipótesis			
Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1 La mediana de diferencias entre EPS_SELECCION y TODAS_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	.619	Conserve la hipótesis nula.

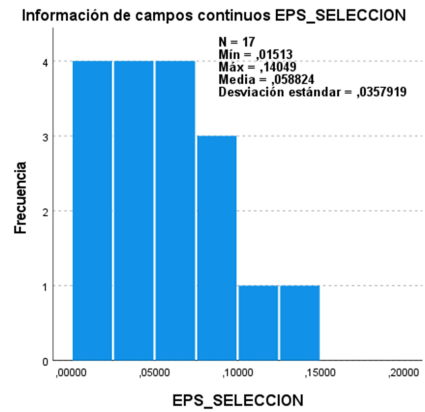
a. El nivel de significación es de .050.  
b. Se muestra la significancia asintótica.

Figura 215: Prueba Wilcoxon Zona Rural Año 2019

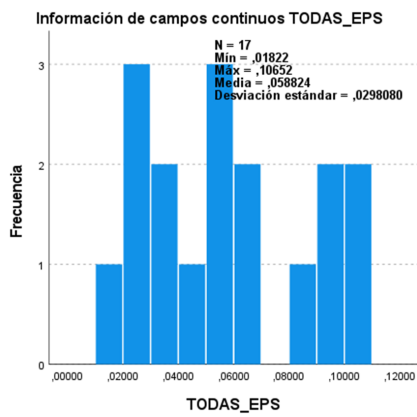
Teniendo como referencia la prima pura establecida para el año 2018 a través de la Resolución 5268 de 2017 y para el año 2019 a través de la Resolución 5858 de 2019, se establece el factor de relatividad para cada uno de los grupos etarios, los cuales se presentan a continuación en la tabla número 75 y la gráfica 217, como se observa, no presenta cambios significativos en la composición de los factores por grupo de edad; sin embargo, para esta zona, los factores del año 2018 son mayores a los presentados en el año 2019.



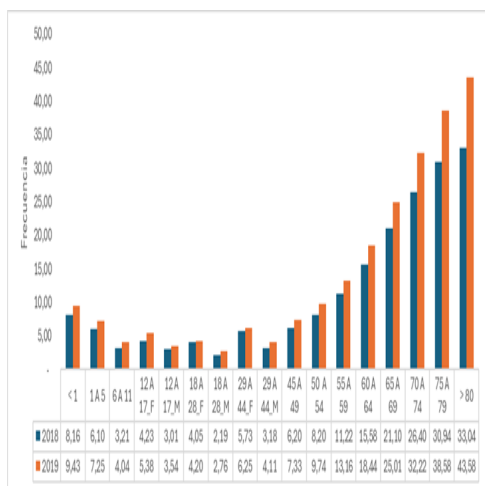
(a) Prueba de Rangos Zona Rural Año 2019



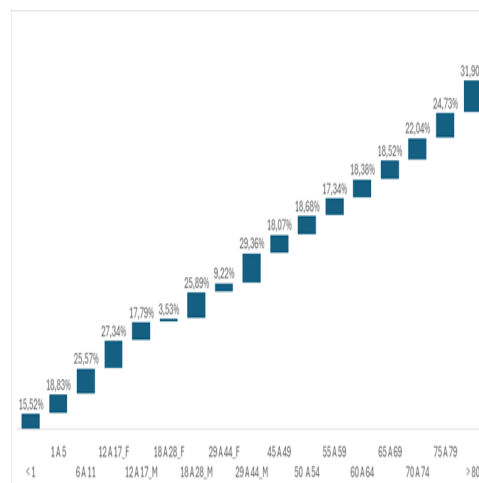
(b) EPS Selección Zona Rural Año 2019



(c) Total EPS Zona Rural Año 2019



(a) Frecuencia Zona Rural Años 2018 - 2019



(b) Variación Frecuencia Zona Rural 2018 - 2019

Descripción Grupo Etario	Sexo	Factor 2018	Factor 2019
Menor de un año	Ambos	0,9967	0,7002
Entre 1 y 5 Años	Ambos	0,3991	0,3080
Entre 6 y 11 Años	Ambos	0,2186	0,1930
Entre 12 y 17 Años	Femenino	0,3174	0,2760
Entre 12 y 17 Años	Masculino	0,2741	0,2249
Entre 18 y 28 Años	Femenino	0,3113	0,2202
Entre 18 y 28 Años	Masculino	0,2046	0,1795
Entre 29 y 44 Años	Femenino	0,4429	0,3574
Entre 28 y 44 Años	Masculino	0,2989	0,2918
Entre 45 y 49 Años	Ambos	0,5467	0,4767
Entre 50 y 54 Años	Ambos	0,7158	0,6495
Entre 55 y 59 Años	Ambos	0,9644	0,8930
Entre 60 y 64 Años	Ambos	1,4184	1,2881
Entre 65 y 69 Años	Ambos	1,9041	1,7236
Entre 70 y 74 Años	Ambos	2,2997	2,2213
Entre 75 y 79 Años	Ambos	2,7904	2,6106
Mayores de 80 Años	Ambos	3,1371	2,9209

Tabla 75: Factores de Relatividad Zona Rural Años 2018 y 2019

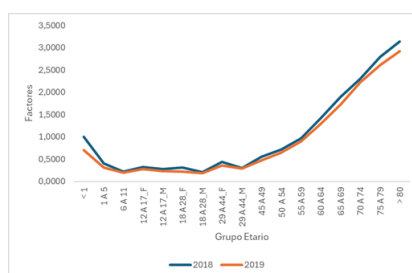
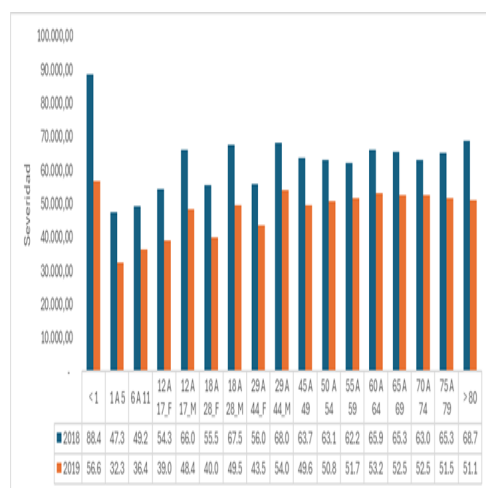


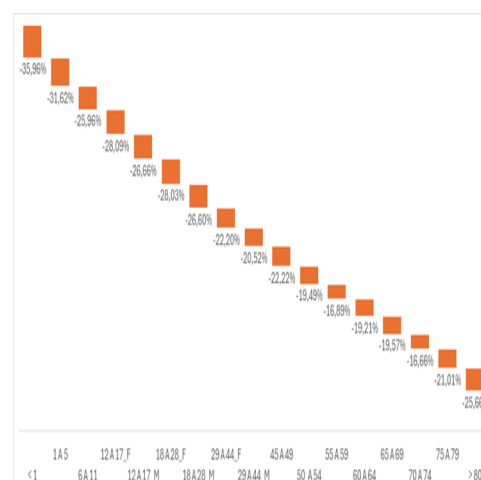
Figura 217: Factores de Relatividad Zona Rural Años 2018 y 2019

Al comparar las frecuencias entre los años 2018 y 2019 se presenta un incremento en todos los grupos etarios en el año 2019, lo anterior se presenta en las figuras números 218a y 218b.

Por otro lado, la severidad o costo promedio presenta una disminución en todos los grupos etarios en el año 2019 comparado con el año 2018; lo anterior se presenta en las figuras 219a y 219b



(a) Frecuencia Zona Rural Años 2018 - 2019



(b) Variación Frecuencia Zona Rural 2018 - 2019

#### 4.4.6. Zona Rural Disperso

A continuación se presenta la prima pura correspondiente al año 2018 para cada uno de los grupos etarios para la Zona Rural Disperso, teniendo en cuenta que la misma se expresa como:

$$PrimaPura = \frac{\text{Pérdidas}}{\text{Número de expuestos}} = Frecuencia * Severidad$$

La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

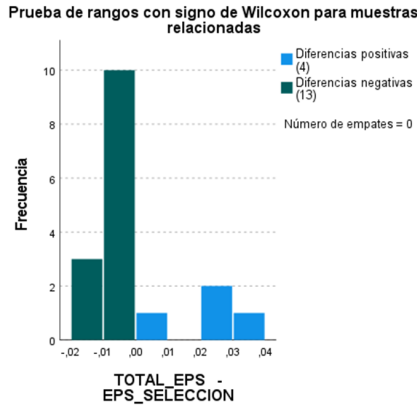
Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	9,031448206	74.573,29	673.504,83
Entre 1 y 5 Años	Ambos	6,579798037	34.258,27	225.412,50
Entre 6 y 11 Años	Ambos	3,227237979	34.724,65	112.064,70
Entre 12 y 17 Años	Femenino	4,360059292	40.383,58	176.074,80
Entre 12 y 17 Años	Masculino	2,618878036	56.350,66	147.575,51
Entre 18 y 28 Años	Femenino	2,439023039	39.916,08	97.356,24
Entre 18 y 28 Años	Masculino	1,4650039	55.677,95	81.568,42
Entre 29 y 44 Años	Femenino	3,624144663	43.423,81	157.374,15
Entre 29 y 44 Años	Masculino	2,176849492	60.731,27	132.202,83
Entre 45 y 49 Años	Ambos	4,290076254	65.933,33	282.859,01
Entre 50 y 54 Años	Ambos	5,555369419	53.270,81	295.939,05
Entre 55 y 59 Años	Ambos	8,047413855	59.142,31	475.942,68
Entre 60 y 64 Años	Ambos	10,21447097	66.459,85	678.852,22
Entre 65 y 69 Años	Ambos	14,56487279	68.513,65	997.892,62
Entre 70 y 74 Años	Ambos	19,81406451	83.278,10	1.650.077,68
Entre 75 y 79 Años	Ambos	22,96858025	56.922,02	1.307.418,04
Mayores de 80 Años	Ambos	26,73644283	78.898,25	2.109.458,57

Tabla 76: Prima Pura Zona Rural Disperso Año 2018

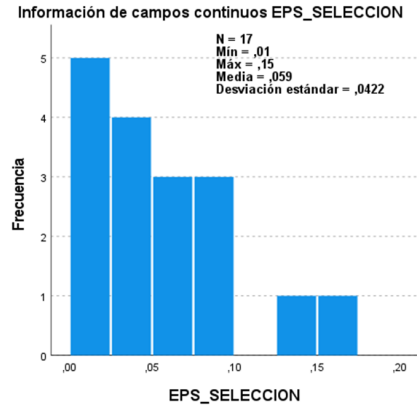
De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; para lo anterior, compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

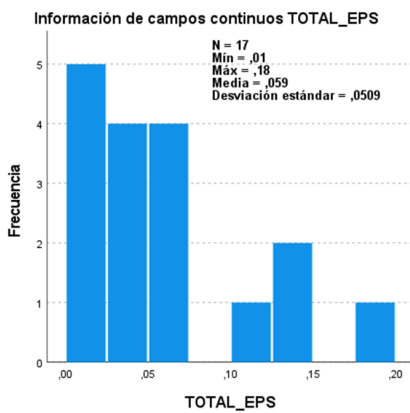
La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados,



(a) Prueba de Rangos Zona Rural Disperso Año 2018



(b) EPS Selección Zona Rural Disperso Año 2018



(c) Total EPS Zona Rural Disperso Año 2018

conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona rural disperso. Adicional a la prueba mencionada, se presenta visualmente el comportamiento descrito.

Resumen de contrastes de hipótesis				
	Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1	La mediana de diferencias entre EPS_SELECCION y TOTAL_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	.381	Conserve la hipótesis nula.

a. El nivel de significación es de .050.  
b. Se muestra la significancia asintótica.

Figura 220: Prueba Wilcoxon Zona Normal Rural Disperso Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Rural Disperso para cada uno de los grupos etarios, La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución Gamma.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	8,803944339	45.051,21	396.628,33
Entre 1 y 5 Años	Ambos	7,200371707	24.154,02	173.917,96
Entre 6 y 11 Años	Ambos	3,683416372	33.026,30	121.649,61
Entre 12 y 17 Años	Femenino	4,91416068	46.675,46	229.370,72
Entre 12 y 17 Años	Masculino	3,320438418	55.478,54	184.213,07
Entre 18 y 28 Años	Femenino	2,826357144	34.924,03	98.707,77
Entre 18 y 28 Años	Masculino	1,337997329	41.510,76	55.541,29
Entre 29 y 44 Años	Femenino	4,187976941	39.595,73	165.825,99
Entre 29 y 44 Años	Masculino	2,083303275	47.063,55	98.047,65
Entre 45 y 49 Años	Ambos	4,89420852	45.676,50	223.550,33
Entre 50 y 54 Años	Ambos	6,48966563	45.981,59	298.405,13
Entre 55 y 59 Años	Ambos	8,850698136	52.995,65	469.048,46
Entre 60 y 64 Años	Ambos	12,45735484	51.121,05	636.833,11
Entre 65 y 69 Años	Ambos	18,40413912	51.594,73	949.556,50
Entre 70 y 74 Años	Ambos	22,68068543	47.254,03	1.071.753,75
Entre 75 y 79 Años	Ambos	27,77426685	52.895,08	1.469.121,96
Mayores de 80 Años	Ambos	34,05061419	52.848,45	1.799.522,09

Tabla 77: Prima Pura Zona Rural Disperso Año 2019

De acuerdo a lo presentado en [de Salud de Protección Social, 2021], el valor estimado de la prima pura es aplicable para toda la población dado que las poblaciones son similares. Para lo anterior, se aplicó la prueba de los rangos con signo de Wilcoxon [Lehmann and D'Abbrera, 2006] a los datos de la zona por los grupos etarios.

La hipótesis nula corresponde a que no hay diferencias entre las poblaciones; para lo anterior, compara uno a uno los datos de las proporciones de la población afiliada de cada grupo etario respecto al total de la zona.

La prueba estadística de Wilcoxon antes mencionada tiene en cuenta la diferencia entre cada par de datos para el presente caso corresponde a los afiliados de la EPS de selección y el Total de Afiliados, conformando así los rangos positivos y negativos, y la suma de estos para calcular el estadístico de contraste.

A continuación, se presentan los resultados, concluyendo que las EPS de la selección del cumplen con similitud con el total de afiliados en la zona rural dispersa. Adicional a la prueba mencionada, se presenta visualmente el comportamiento descrito.

### Resumen de contrastes de hipótesis

Hipótesis nula	Prueba	Sig. <sup>a,b</sup>	Decisión
1 La mediana de diferencias entre EPS_SELECCION y TODAS_EPS es igual a 0.	Prueba de rangos con signo de Wilcoxon para muestras relacionadas	,723	Conserve la hipótesis nula.

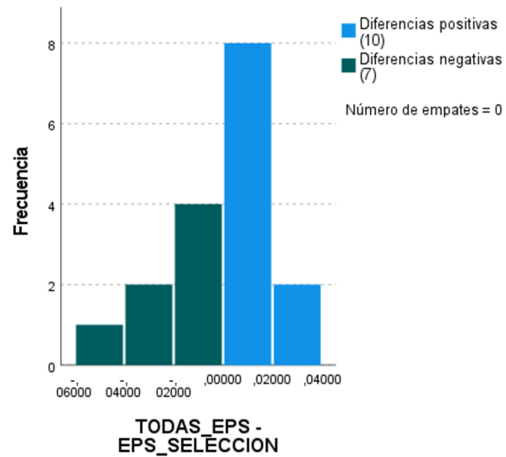
a. El nivel de significación es de ,050.

b. Se muestra la significancia asintótica.

Figura 222: Prueba Wilcoxon Zona Rural Disperso Año 2019

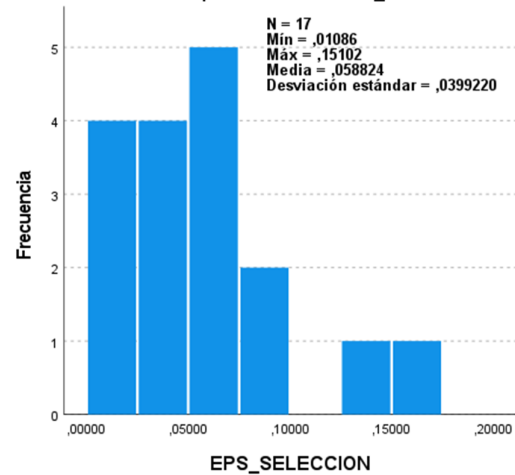
Teniendo como referencia la prima pura establecida para el año 2018 a través de la Resolución 5268 de 2017 y para el año 2019 a través de la Resolución 5858 de 2019, se establece el factor de relatividad para cada uno de los grupos etarios, los cuales se presentan a continuación en la tabla número ?? y la gráfica 224. Como se observa, no presenta cambios significativos en la composición de los factores por grupo de edad en casi todos los grupos, con excepción del grupo 70 a 74 años, que en el año 2018 presenta un factor de relatividad superior que en el año 2019.

Prueba de rangos con signo de Wilcoxon para muestras relacionadas



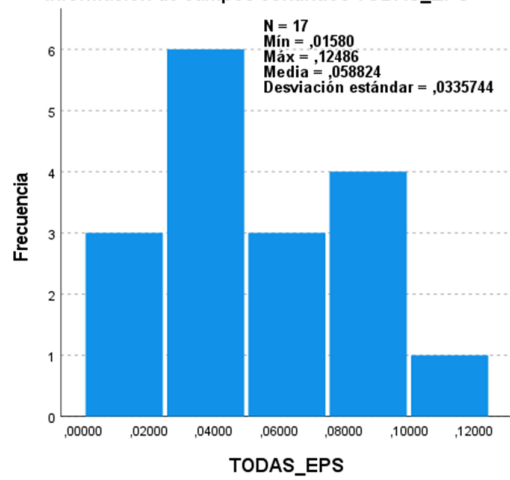
(a) Prueba de Rangos Zona Rural Disperso Año 2019

Información de campos continuos EPS\_SELECCION

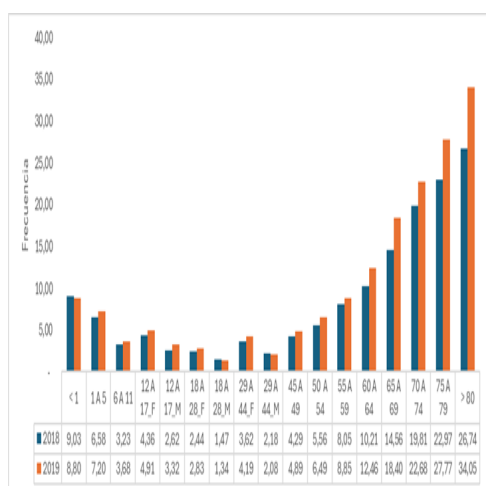


(b) EPS Selección Zona Rural Disperso Año 2019

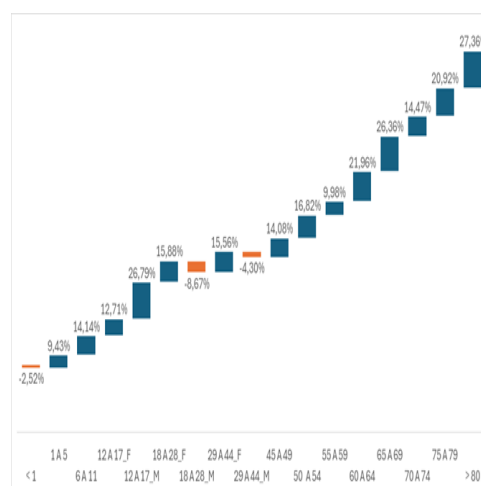
Información de campos continuos TODAS\_EPS



(c) Total EPS Zona Rural Disperso Año 2019



(a) Frecuencia Zona Rural Disperso Años 2018 - 2019



(b) Variación Frecuencia Zona Rural Disperso 2018 - 2019

Descripción Grupo Etario	Sexo	Factor 2018	Factor 2019
Menor de un año	Ambos	0,9302	0,5202
Entre 1 y 5 Años	Ambos	0,3113	0,2281
Entre 6 y 11 Años	Ambos	0,1548	0,1595
Entre 12 y 17 Años	Femenino	0,2432	0,3008
Entre 12 y 17 Años	Masculino	0,2038	0,2416
Entre 18 y 28 Años	Femenino	0,1345	0,1295
Entre 18 y 28 Años	Masculino	0,1127	0,0728
Entre 29 y 44 Años	Femenino	0,2174	0,2175
Entre 28 y 44 Años	Masculino	0,1826	0,1286
Entre 45 y 49 Años	Ambos	0,3907	0,2932
Entre 50 y 54 Años	Ambos	0,4087	0,3914
Entre 55 y 59 Años	Ambos	0,6574	0,6152
Entre 60 y 64 Años	Ambos	0,9376	0,8352
Entre 65 y 69 Años	Ambos	1,3783	1,2454
Entre 70 y 74 Años	Ambos	2,2791	1,4056
Entre 75 y 79 Años	Ambos	1,8058	1,9268
Mayores de 80 Años	Ambos	2,9135	2,3601

Tabla 78: Factores de Relatividad Zona Rural Disperso Años 2018 y 2019

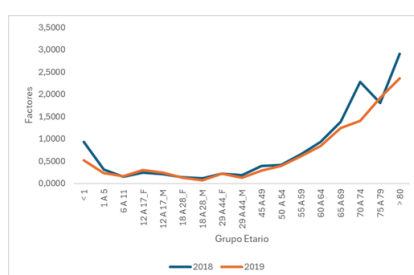
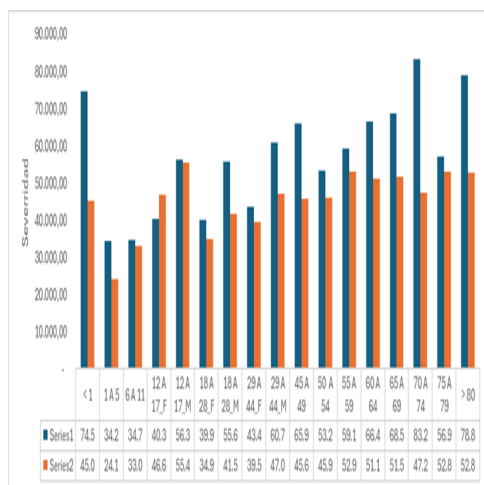


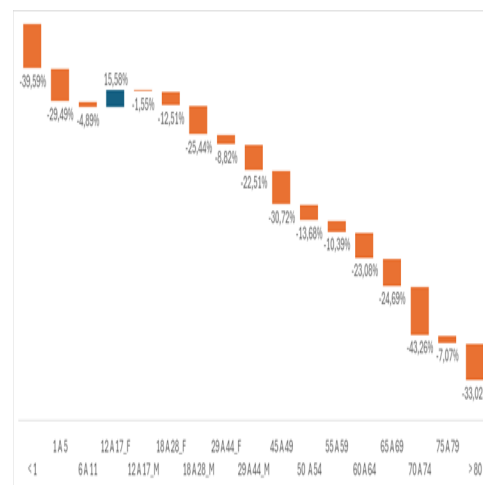
Figura 224: Factores de Relatividad Zona Rural Dispersa Años 2018 y 2019

Al comparar las frecuencias entre los años 2018 y 2019 se presenta un incremento en todos los grupos etarios en el año 2019, lo anterior se presenta en las figuras números 225a y 225b.

Por otro lado, la severidad o costo promedio presenta una disminución en todos los grupos etarios en el año 2019 comparado con el año 2018; lo anterior se presenta en las figuras 226a y 226b



(a) Frecuencia Zona Rural Disperso Mayor Años 2018 - 2019



(b) Variación Frecuencia Zona Rural Disperso 2018 - 2019

### 4.5. Prima Pura Zona Variable Explicativa

A continuación se presentan de manera comparativa los resultados de la prima pura en cada una de las zonas, y el MAPE calculado.

#### 4.5.1. Zona Alejada

La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	5,218319273	161.925,36	844.978,24
Entre 1 y 5 Años	Ambos	4,135669119	86.306,42	356.934,81
Entre 6 y 11 Años	Ambos	2,162434122	87.463,92	189.134,97
Entre 12 y 17 Años	Femenino	2,697633616	90.198,12	243.321,48
Entre 12 y 17 Años	Masculino	1,895039054	118.375,12	224.325,47
Entre 18 y 28 Años	Femenino	2,679943445	90.369,81	242.185,97
Entre 18 y 28 Años	Masculino	1,882612028	118.605,81	223.288,72
Entre 29 y 44 Años	Femenino	3,624044485	94.516,94	342.533,60
Entre 29 y 44 Años	Masculino	2,545826013	124.181,4	316.144,23
Entre 45 y 49 Años	Ambos	4,298830911	112.467,79	483.480,02
Entre 50 y 54 Años	Ambos	5,537916513	109.257,59	605.059,43
Entre 55 y 59 Años	Ambos	7,256457659	113.715,67	825.172,93
Entre 60 y 64 Años	Ambos	9,586794093	114.036,65	1.093.245,92
Entre 65 y 69 Años	Ambos	12,8048872	114.717,78	1.468.948,28
Entre 70 y 74 Años	Ambos	15,47330266	126.489,03	1.957.203,11
Entre 75 y 79 Años	Ambos	18,15955589	117.766,63	2.138.589,76
Mayores de 80 Años	Ambos	20,48435861	126.720,13	2.595.780,57

Tabla 79: Prima Pura Zona Alejada Año 2018 Modelo 2

Al analizar el MAPE para el año 2018 se observa que el mismo presenta un incremento en la zona alejada en la Frecuencia, pasando del 9.78 % al 15.96 %. Una situación similar se observa en la severidad, que pasa del 4.15 % al 10.18 %.

Lo anterior se evidencia en la prima pura del año 2018, lo cual se puede observar en la gráfica 227 que compara la prima pura de los dos modelos.

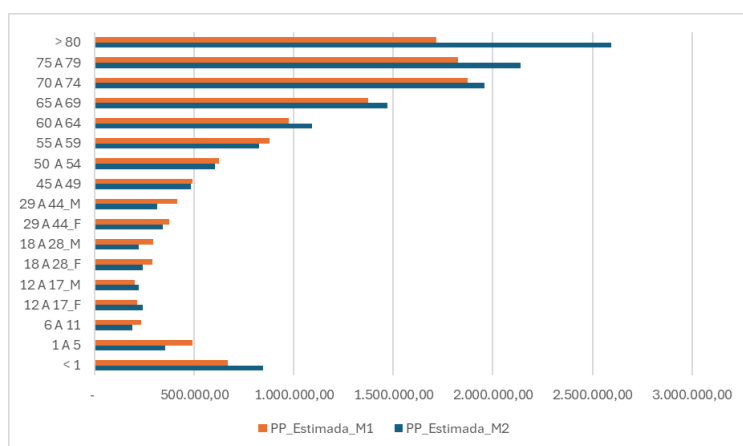


Figura 227: Prima Pura Comparada entre los modelos Zona Alejada Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Alejada para cada uno de los grupos etarios. La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	11,11251681	87.365,82	970.854,18
Entre 1 y 5 Años	Ambos	9,070718912	51.859,19	470.400,14
Entre 6 y 11 Años	Ambos	4,874492808	57.331,66	279.462,79
Entre 12 y 17 Años	Femenino	6,181537458	58.979,16	364.581,87
Entre 12 y 17 Años	Masculino	4,369350141	72.171,03	315.340,52
Entre 18 y 28 Años	Femenino	6,055837948	57.427,25	347.770,11
Entre 18 y 28 Años	Masculino	4,280500859	70.237,59	300.652,04
Entre 29 y 44 Años	Femenino	8,009112271	63.500,48	508.582,51
Entre 28 y 44 Años	Masculino	5,661150818	77.809,18	440.489,53
Entre 45 y 49 Años	Ambos	9,93904031	69.411,80	689.886,63
Entre 50 y 54 Años	Ambos	13,06748465	70.901,55	926.504,88
Entre 55 y 59 Años	Ambos	16,92311389	71.918,08	1.217.077,82
Entre 60 y 64 Años	Ambos	22,88508189	71.323,79	1.632.250,80
Entre 65 y 69 Años	Ambos	30,52588797	71.637,65	2.186.802,76
Entre 70 y 74 Años	Ambos	38,18645901	69.628,81	2.658.877,78
Entre 75 y 79 Años	Ambos	45,42897022	75.425,08	3.426.483,87
Mayores de 80 Años	Ambos	53,01752083	71.125,25	3.770.884,24

Tabla 80: Prima Pura Zona Alejada Año 2019 Modelo 2

Al analizar el MAPE para el año 2019 se observa que el mismo presenta un incremento en la zona alejada en la Frecuencia, pasando del 7.57 % al 14.74 %; situación similar se observa en la severidad que pasa del 2.87 % al 11.03 %.

Lo anterior se evidencia en la prima pura del año 2018, lo cual se puede observar en la gráfica 228 que compara la prima pura de los dos modelos.

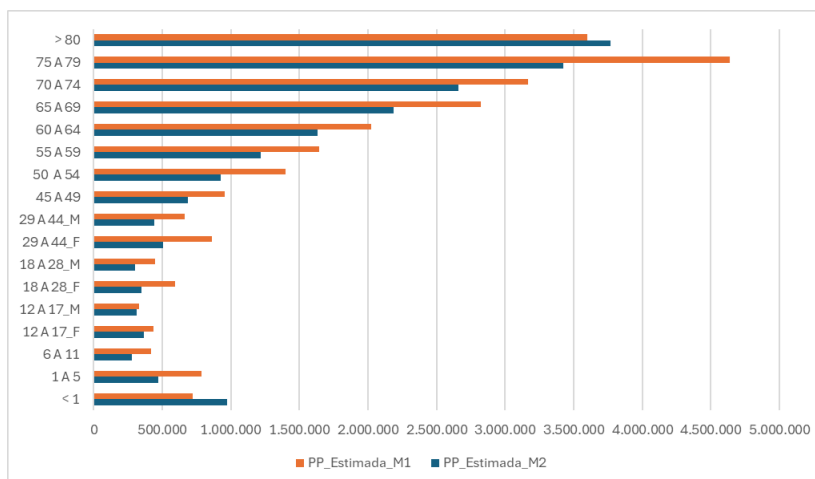


Figura 228: Prima Pura Comparada entre los modelos Zona Alejada Año 2019

Por lo prestado se puede inferir que el modelo de frecuencia y severidad realizado por cada una de las zonas presenta un mejor ajuste que el modelo que considera las zonas como variable explicativa del gasto.

#### 4.5.2. Zona Bogotá

La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución Lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	14,98050919	72.522,75	1.343.846,13
Entre 1 y 5 Años	Ambos	10,95133973	40.326,78	495.937,79
Entre 6 y 11 Años	Ambos	6,173606276	40.841,32	258.317,76
Entre 12 y 17 Años	Femenino	7,810542814	42.213,12	328.698,22
Entre 12 y 17 Años	Masculino	5,87125591	54.406,87	313.465,40
Entre 18 y 28 Años	Femenino	9,15711801	42.288,12	359.379,72
Entre 18 y 28 Años	Masculino	6,883488705	54.505,84	342.190,69
Entre 29 y 44 Años	Femenino	11,32886365	44.096,89	501.298,04
Entre 29 y 44 Años	Masculino	8,516009611	56.893,93	478.602,90
Entre 45 y 49 Años	Ambos	13,2736456	52.044,05	681.001,41
Entre 50 y 54 Años	Ambos	17,01012008	50.488,95	860.785,86
Entre 55 y 59 Años	Ambos	21,70344295	52.171,60	1.130.265,55
Entre 60 y 64 Años	Ambos	29,24216148	52.326,79	1.491.723,89
Entre 65 y 69 Años	Ambos	38,62471011	52.467,52	1.949.691,05
Entre 70 y 74 Años	Ambos	48,63085013	57.307,05	2.479.623,28
Entre 75 y 79 Años	Ambos	58,23922266	53.495,39	3.047.927,06
Mayores de 80 Años	Ambos	67,92465273	57.508,89	3.750.932,04

Tabla 81: Prima Pura Zona Bogotá Año 2018 Modelo 2

Al analizar el MAPE para el año 2018 se observa que el mismo presenta un incremento en la zona Bogotá en la Frecuencia, pasando del 7.55 % al 8.47 %. Situación similar se observa en la severidad que pasa del 1.17 % al 4.82 %.

Lo anterior se evidencia en la prima pura del año 2018, lo cual se puede observar en la gráfica 229 que compara la prima pura de los dos modelos.

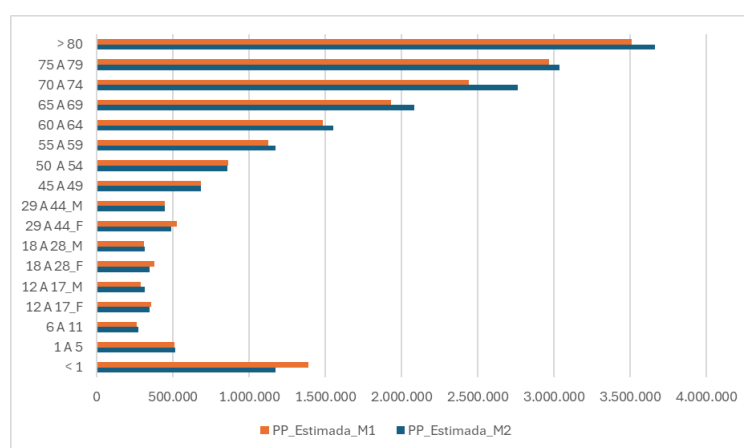


Figura 229: Prima Pura Comparada entre los modelos Zona Bogotá Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Bogotá para cada uno de los grupos etarios. La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	115,84613001	62.658,69	992.897,78
Entre 1 y 5 Años	Ambos	12,98046362	37.691,23	489.249,67
Entre 6 y 11 Años	Ambos	6,93616122	41.676,74	289.076,56
Entre 12 y 17 Años	Femenino	8,792897755	42.853,52	376.806,66
Entre 12 y 17 Años	Masculino	6,215160761	52.131,71	324.006,97
Entre 18 y 28 Años	Femenino	8,614097102	41.758,29	359.709,96
Entre 18 y 28 Años	Masculino	6,088777533	50.775,19	309.158,82
Entre 29 y 44 Años	Femenino	11,39252262	46.039,68	524.508,11
Entre 28 y 44 Años	Masculino	8,052676322	56.081,56	451.606,64
Entre 45 y 49 Años	Ambos	14,09727995	50.271,70	708.694,20
Entre 50 y 54 Años	Ambos	18,58916963	51.231,44	952.349,95
Entre 55 y 59 Años	Ambos	24,21917564	51.762,32	1.253.640,77
Entre 60 y 64 Años	Ambos	32,73805731	51.358,43	1.681.375,19
Entre 65 y 69 Años	Ambos	43,65416966	51.587,12	2.251.992,76
Entre 70 y 74 Años	Ambos	54,65521258	50.157,45	2.741.365,95
Entre 75 y 79 Años	Ambos	65,82287861	53.813,82	3.542.180,46
Mayores de 80 Años	Ambos	76,32357574	51.015,73	3.893.703,31

Tabla 82: Prima Pura Zona Bogotá Año 2019 Modelo 2

Al analizar el MAPE para el año 2019 se observa que el mismo presenta un incremento en la zona Bogotá en la Frecuencia, pasando del 7.51 % al 8.57 %. Una situación similar se observa en la severidad que pasa del 1.21 % al 6.63 %.

Lo anterior se evidencia en la prima pura del año 2019, lo cual se puede observar en la gráfica 230 que compara la prima pura de los dos modelos.

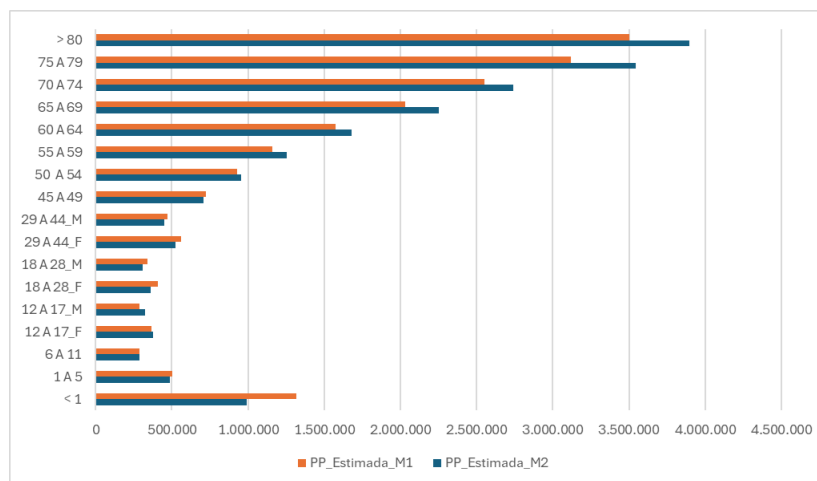


Figura 230: Prima Pura Comparada entre los modelos Zona Bogotá Año 2019

Por lo prestado se puede inferir que el modelo de frecuencia y severidad realizado por cada una de las zonas presenta un mejor ajuste que el modelo que considera las zonas como variable explicativa del gasto.

#### 4.5.3. Zona Normal Mayor

La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	16,25408127	70.198,77	1.141.016,46
Entre 1 y 5 Años	Ambos	12,83949321	39.185,12	503.117,09
Entre 6 y 11 Años	Ambos	6,708539398	39.694,52	266.292,26
Entre 12 y 17 Años	Femenino	8,327764674	41.013,13	341.547,73
Entre 12 y 17 Años	Masculino	5,850104773	52.823,94	309.025,61
Entre 18 y 28 Años	Femenino	8,273154003	41.085,80	339.909,17
Entre 18 y 28 Años	Masculino	5,81174176	52.919,77	307.556,04
Entre 29 y 44 Años	Femenino	11,1876533	42.838,29	479.259,89
Entre 29 y 44 Años	Masculino	7,859125054	55.231,96	434.074,91
Entre 45 y 49 Años	Ambos	13,13669148	50.712,65	666.196,37
Entre 50 y 54 Años	Ambos	17,00939037	49.179,94	836.520,80
Entre 55 y 59 Años	Ambos	22,4538964	50.766,64	1.139.908,97
Entre 60 y 64 Años	Ambos	29,71401705	50.836,06	1.510.543,68
Entre 65 y 69 Años	Ambos	39,74530143	51.064,58	2.029.577,02
Entre 70 y 74 Años	Ambos	48,20273955	55.770,76	2.688.303,63
Entre 75 y 79 Años	Ambos	56,81100495	52.015,56	2.955.056,27
Mayores de 80 Años	Ambos	63,49861606	56.065,46	3.560.079,14

Tabla 83: Prima Pura Zona Normal Mayor Año 2018 Modelo 2

Al analizar el MAPE para el año 2018 se observa que el mismo presenta un incremento en la zona Normal Mayor en la Frecuencia, pasando del 7.59 % al 10.65 %. Una situación similar se observa en la severidad que pasa del 0.64 % al 5.32 %.

Lo anterior se evidencia en la prima pura del año 2018, lo cual se puede observar en la gráfica 231 que compara la prima pura de los dos modelos.

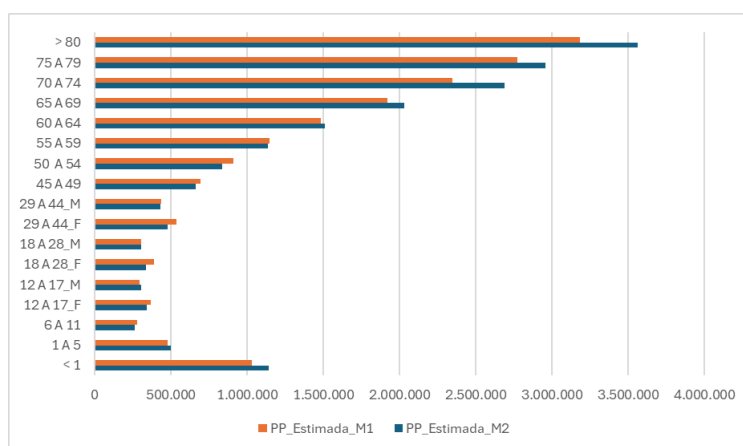


Figura 231: Prima Pura Comparada entre los modelos Zona Normal Mayor Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Normal Mayor para cada uno de los grupos etarios. La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	16,56311422	56.924,28	942.843,42
Entre 1 y 5 Años	Ambos	13,55478281	34.411,52	466.440,72
Entre 6 y 11 Años	Ambos	7,240511836	38.024,57	275.317,33
Entre 12 y 17 Años	Femenino	9,181430652	39.084,00	358.847,01
Entre 12 y 17 Años	Masculino	6,489790864	47.465,67	308.042,28
Entre 18 y 28 Años	Femenino	8,994729312	38.093,62	342.641,77
Entre 18 y 28 Años	Masculino	6,357823124	46.241,09	293.992,65
Entre 29 y 44 Años	Femenino	11,89592547	41.963,90	499.199,45
Entre 28 y 44 Años	Masculino	8,40850096	51.029,80	429.084,16
Entre 45 y 49 Años	Ambos	14,64199963	45.924,17	672.421,67
Entre 50 y 54 Años	Ambos	19,32023582	46.776,57	903.734,38
Entre 55 y 59 Años	Ambos	25,20227087	47.224,70	1.190.169,73
Entre 60 y 64 Años	Ambos	34,15094262	46.792,66	1.598.013,31
Entre 65 y 69 Años	Ambos	45,45774038	47.047,98	2.138.694,70
Entre 70 y 74 Años	Ambos	56,88739189	45.767,33	2.603.583,84
Entre 75 y 79 Años	Ambos	68,5971815	49.038,70	3.363.916,30
Mayores de 80 Años	Ambos	79,32202074	46.589,78	3.695.595,80

Tabla 84: Prima Pura Zona Normal Mayor Año 2019 Modelo 2

Al analizar el MAPE para el año 2019 se observa que el mismo presenta un incremento en la zona Normal Mayor en la Frecuencia, pasando del 6.77 % al 10.42 %. Una situación similar se observa en la severidad que pasa del 1.21 % al 5.66 %.

Lo anterior se evidencia en la prima pura del año 2019, lo cual se puede observar en la gráfica 232 que compara la prima pura de los dos modelos.

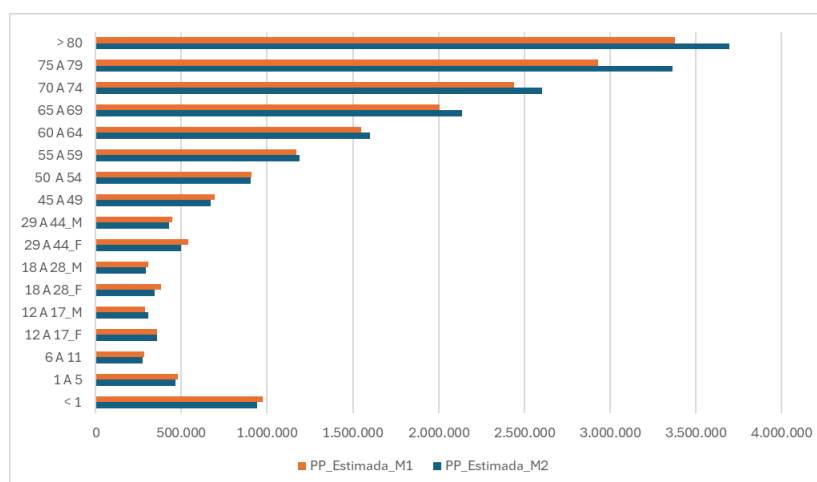


Figura 232: Prima Pura Comparada entre los modelos Zona Normal Mayor Año 2019

Por lo prestado se puede inferir que el modelo de frecuencia y severidad realizado por cada una de las zonas presenta un mejor ajuste que el modelo que considera las zonas como variable explicativa del gasto.

#### 4.5.4. Zona Normal Medio

La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	13,27628147	71.790,71	953.113,64
Entre 1 y 5 Años	Ambos	10,49768531	39.998,39	419.890,55
Entre 6 y 11 Años	Ambos	5,493918881	40.472,68	222.353,60
Entre 12 y 17 Años	Femenino	6,821691604	41.811,60	285.225,85
Entre 12 y 17 Años	Masculino	4,792115553	53.877,10	258.185,28
Entre 18 y 28 Años	Femenino	6,776957253	41.885,82	283.858,41
Entre 18 y 28 Años	Masculino	4,760690476	53.975,01	256.958,33
Entre 29 y 44 Años	Femenino	9,164370465	43.675,74	400.260,65
Entre 29 y 44 Años	Masculino	6,437805281	56.337,68	362.691,04
Entre 45 y 49 Años	Ambos	10,65965387	52.076,15	555.113,69
Entre 50 y 54 Años	Ambos	13,77371841	50.576,28	696.623,50
Entre 55 y 59 Años	Ambos	18,17239588	52.236,96	949.270,77
Entre 60 y 64 Años	Ambos	24,02505954	52.350,64	1.257.727,12
Entre 65 y 69 Años	Ambos	32,19116362	52.519,28	1.690.656,84
Entre 70 y 74 Años	Ambos	39,07377566	57.338,78	2.240.442,58
Entre 75 y 79 Años	Ambos	45,99495751	53.522,96	2.461.786,26
Mayores de 80 Años	Ambos	51,07604299	58.007,24	2.962.780,41

Tabla 85: Prima Pura Zona Normal Medio Año 2018 Modelo 2

Al analizar el MAPE para el año 2018 se observa que el mismo presenta un incremento en la zona Normal Medio en la Frecuencia, pasando del 8.34 % al 10.53 %. Una situación similar se observa en la severidad que pasa del 1.02 % al 3.27 %.

Lo anterior se evidencia en la prima pura del año 2018, lo cual se puede observar en la gráfica 233 que compara la prima pura de los dos modelos.

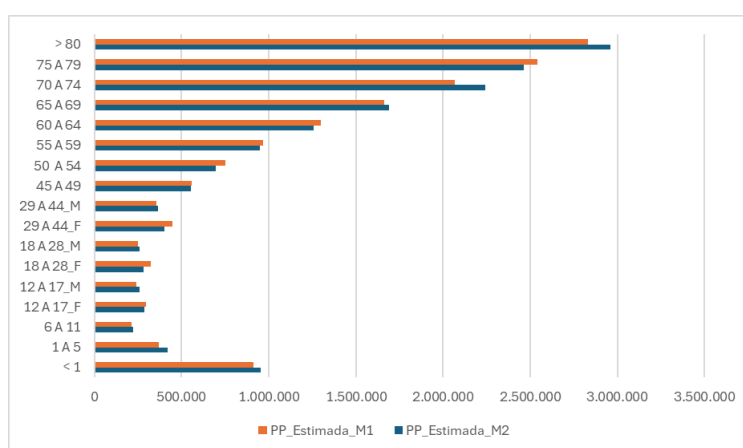


Figura 233: Prima Pura Comparada entre los modelos Zona Normal Medio Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Normal Medio para cada uno de los grupos etarios. La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	13,42495898	56.303,09	755.866,66
Entre 1 y 5 Años	Ambos	10,99630525	34.036,73	374.278,29
Entre 6 y 11 Años	Ambos	5,877664286	37.593,02	220.959,14
Entre 12 y 17 Años	Femenino	7,452750459	38.640,91	287.981,07
Entre 12 y 17 Años	Masculino	5,267892736	46.917,73	247.157,59
Entre 18 y 28 Años	Femenino	7,301201256	37.662,80	274.983,68
Entre 18 y 28 Años	Masculino	5,160771889	45.708,57	235.891,52
Entre 29 y 44 Años	Femenino	9,656160069	41.484,99	400.585,67
Entre 28 y 44 Años	Masculino	6,825348006	50.436,78	344.248,60
Entre 45 y 49 Años	Ambos	11,77485486	45.635,00	537.345,51
Entre 50 y 54 Años	Ambos	15,50195469	46.542,45	721.498,91
Entre 55 y 59 Años	Ambos	20,21820144	46.994,81	950.150,59
Entre 60 y 64 Años	Ambos	27,38245965	46.582,39	1.275.540,47
Entre 65 y 69 Años	Ambos	36,48412876	46.810,34	1.707.834,62
Entre 70 y 74 Años	Ambos	45,69697787	45.514,55	2.079.877,44
Entre 75 y 79 Años	Ambos	55,07741107	48.782,53	2.686.815,57
Mayores de 80 Años	Ambos	63,33422172	46.509,40	2.945.636,35

Tabla 86: Prima Pura Zona Normal Medio Año 2019 Modelo 2

Al analizar el MAPE para el año 2019 se observa que el mismo presenta un incremento en la zona Normal Medio en la Frecuencia, pasando del 8.33 % al 10.015 %. Una situación similar se observa en la severidad que pasa del 1.06 % al 3.96 %.

Lo anterior se evidencia en la prima pura del año 2019, lo cual se puede observar en la gráfica 234 que compara la prima pura de los dos modelos.

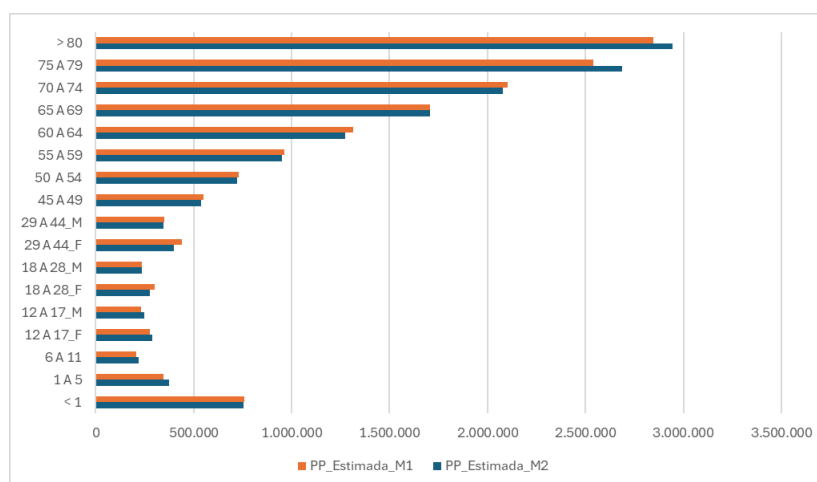


Figura 234: Prima Pura Comparada entre los modelos Zona Normal Mayor Año 2019

Por lo prestado se puede inferir que el modelo de frecuencia y severidad realizado por cada una de las zonas presenta un mejor ajuste que el modelo que considera las zonas como variable explicativa del gasto.

#### 4.5.5. Zona Rural

La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	8,266867917	87.979,42	727.314,28
Entre 1 y 5 Años	Ambos	6,546770286	48.445,29	317.160,17
Entre 6 y 11 Años	Ambos	3,422870137	49.064,69	167.942,06
Entre 12 y 17 Años	Femenino	4,253053645	50.685,32	215.567,39
Entre 12 y 17 Años	Masculino	2,987693625	65.611,67	196.027,57
Entre 18 y 28 Años	Femenino	4,225163554	50.776,92	214.540,79
Entre 18 y 28 Años	Masculino	2,968101338	65.733,07	195.102,42
Entre 29 y 44 Años	Femenino	5,713620825	52.986,87	302.746,91
Entre 29 y 44 Años	Masculino	4,013715777	68.663,62	275.596,26
Entre 45 y 49 Años	Ambos	6,518802315	64.275,24	418.997,61
Entre 50 y 54 Años	Ambos	8,416464761	62.427,69	525.420,48
Entre 55 y 59 Años	Ambos	11,10456849	64.523,31	716.503,53
Entre 60 y 64 Años	Ambos	14,70204124	64.608,18	949.872,11
Entre 65 y 69 Años	Ambos	19,729672	64.749,03	1.277.477,22
Entre 70 y 74 Años	Ambos	24,01113058	70.677,70	1.697.051,59
Entre 75 y 79 Años	Ambos	28,31912532	65.799,57	1.863.386,34
Mayores de 80 Años	Ambos	31,67812149	71.033,95	2.250.222,20

Tabla 87: Prima Pura Zona Rural Año 2018 Modelo 2

Al analizar el MAPE para el año 2018 se observa que el mismo presenta un incremento en la zona Rural en la Frecuencia, pasando del 6.84 % al 13.75 %. Una situación similar se observa en la severidad que pasa del 0.85 % al 3.42 %.

Lo anterior se evidencia en la prima pura del año 2018, lo cual se puede observar en la gráfica 235 que compara la prima pura de los dos modelos.

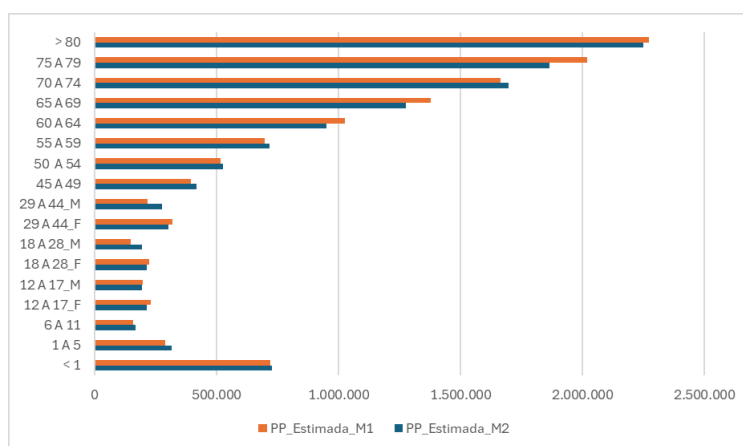


Figura 235: Prima Pura Comparada entre los modelos Zona Rural Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Rural para cada uno de los grupos etarios. La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	8,952971248	59.688,15	534.386,30
Entre 1 y 5 Años	Ambos	7,335190211	35.981,22	263.929,06
Entre 6 y 11 Años	Ambos	3,916009814	39.788,46	155.812,00
Entre 12 y 17 Años	Femenino	4,970586523	40.878,89	203.192,08
Entre 12 y 17 Años	Masculino	3,51340311	49.686,47	174.568,59
Entre 18 y 28 Años	Femenino	4,869511298	39.838,69	193.994,94
Entre 18 y 28 Años	Masculino	3,441959225	48.399,21	166.588,12
Entre 29 y 44 Años	Femenino	6,440143053	43.904,31	282.750,03
Entre 28 y 44 Años	Masculino	4,552142595	53.433,78	243.238,20
Entre 45 y 49 Años	Ambos	7,689171063	48.890,53	375.927,66
Entre 50 y 54 Años	Ambos	10,11624185	49.889,11	504.690,26
Entre 55 y 59 Años	Ambos	13,18551681	50.399,30	664.540,86
Entre 60 y 64 Años	Ambos	17,89342622	49.904,21	892.957,35
Entre 65 y 69 Años	Ambos	23,89469762	50.088,27	1.196.844,03
Entre 70 y 74 Años	Ambos	30,04705105	48.586,34	1.459.876,21
Entre 75 y 79 Años	Ambos	36,29518525	52.034,76	1.888.611,10
Mayores de 80 Años	Ambos	41,94769494	49.455,76	2.074.555,17

Tabla 88: Prima Pura Zona Rural Año 2019 Modelo 2

Al analizar el MAPE para el año 2019 se observa que el mismo presenta un incremento en la zona Rural en la Frecuencia, pasando del 10.02 % al 12.86 %. Una situación similar se observa en la severidad que pasa del 0.94 % al 4.49 %.

Lo anterior se evidencia en la prima pura del año 2019, lo cual se puede observar en la gráfica 236 que compara la prima pura de los dos modelos.

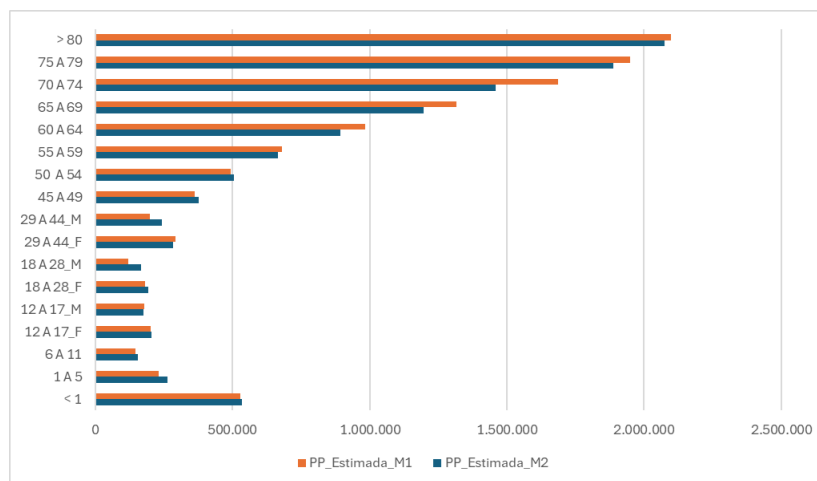


Figura 236: Prima Pura Comparada entre los modelos Zona Normal Mayor Año 2019

Por lo prestado se puede inferir que el modelo de frecuencia y severidad realizado por cada una de las zonas presenta un mejor ajuste que el modelo que considera las zonas como variable explicativa del gasto.

#### 4.5.6. Zona Rural Disperso

La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	6,537162746	78.202,68	511.223,65
Entre 1 y 5 Años	Ambos	5,156058593	43.454,10	224.051,89
Entre 6 y 11 Años	Ambos	2,695394879	44.008,32	118.619,81
Entre 12 y 17 Años	Femenino	3,346676529	45.474,49	152.188,40
Entre 12 y 17 Años	Masculino	2,350980017	58.714,34	138.036,24
Entre 18 y 28 Años	Femenino	3,324730153	45.555,85	151.460,89
Entre 18 y 28 Años	Masculino	2,335563083	58.821,89	137.382,23
Entre 29 y 44 Años	Femenino	4,495979196	47.518,29	213.641,23
Entre 29 y 44 Años	Masculino	3,158344452	61.417,49	193.977,58
Entre 45 y 49 Años	Ambos	5,100941301	57.750,60	294.582,42
Entre 50 y 54 Años	Ambos	6,546241127	56.341,32	368.823,86
Entre 55 y 59 Años	Ambos	8,640218412	58.202,92	502.885,94
Entre 60 y 64 Años	Ambos	11,37914067	58.499,45	665.673,45
Entre 65 y 69 Años	Ambos	15,42902756	58.202,98	898.015,41
Entre 70 y 74 Años	Ambos	18,79656321	63.434,06	1.192.342,35
Entre 75 y 79 Años	Ambos	22,04430915	59.342,40	1.308.162,22
Mayores de 80 Años	Ambos	24,74186519	63.871,83	1.580.308,24

Tabla 89: Prima Pura Zona Rural Disperso Año 2018 Modelo 2

Al analizar el MAPE para el año 2018 se observa que el mismo presenta un incremento en la zona Rural Disperso en la Frecuencia, pasando del 11.24 % al 23.91 %. Una situación similar se observa en la severidad que pasa del 4.11 % al 13.95 %.

Lo anterior se evidencia en la prima pura del año 2018, lo cual se puede observar en la gráfica 237 que compara la prima pura de los dos modelos.

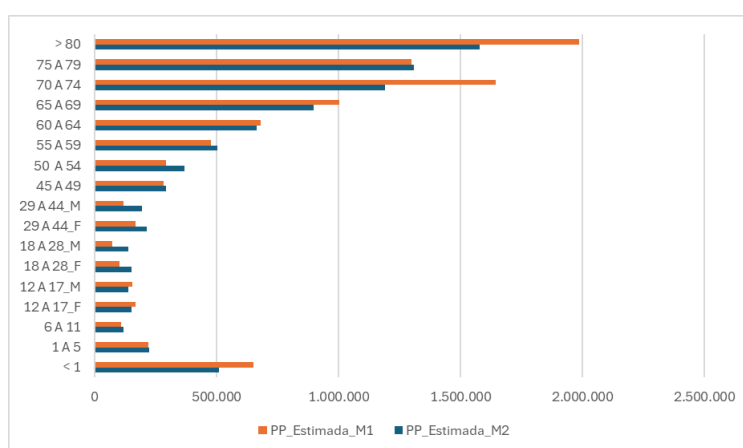


Figura 237: Prima Pura Comparada entre los modelos Zona Rural Disperso Año 2018

A continuación se presenta la prima pura correspondiente al año 2019 para la Zona Rural Disperso para cada uno de los grupos etarios. La distribución que sigue la frecuencia es una binomial negativa y la severidad sigue una distribución lognormal.

Descripción Grupo Etario	Sexo	Fq Estimada	Sev Estimada	Prima Pura
Menor de un año	Ambos	6,910443763	54.932,79	379.609,99
Entre 1 y 5 Años	Ambos	5,651697542	33.273,71	188.052,97
Entre 6 y 11 Años	Ambos	3,015202178	36.780,66	110.901,12
Entre 12 y 17 Años	Femenino	3,82746873	37.780,85	144.605,04
Entre 12 y 17 Años	Masculino	2,705403171	45.854,48	124.054,86
Entre 18 y 28 Años	Femenino	3,749638426	36.826,53	138.086,15
Entre 18 y 28 Años	Masculino	2,650389696	44.675,21	118.406,72
Entre 29 y 44 Años	Femenino	4,959061881	40.555,49	201.117,17
Entre 28 y 44 Años	Masculino	3,505257047	49.286,19	172.760,76
Entre 45 y 49 Años	Ambos	5,873037247	45.320,94	266.171,60
Entre 50 y 54 Años	Ambos	7,722618979	46.254,00	357.202,05
Entre 55 y 59 Años	Ambos	10,04508485	46.776,56	469.874,50
Entre 60 y 64 Años	Ambos	13,5461931	46.481,36	629.645,52
Entre 65 y 69 Años	Ambos	18,21846472	46.473,89	846.682,86
Entre 70 y 74 Años	Ambos	23,14966828	44.831,27	1.037.828,94
Entre 75 y 79 Años	Ambos	27,45370986	48.488,05	1.331.176,77
Mayores de 80 Años	Ambos	32,01449001	45.877,27	1.468.737,37

Tabla 90: Prima Pura Zona Rural Disperso Año 2019 Modelo 2

Al analizar el MAPE para el año 2019 se observa que el mismo presenta un incremento en la zona Rural Disperso en la Frecuencia, pasando del 7.57 % al 25.76 %. Una situación similar se observa en la severidad que pasa del 5.88 % al 12.68 %.

Lo anterior se evidencia en la prima pura del año 2019, lo cual se puede observar en la gráfica 238 que compara la prima pura de los dos modelos.

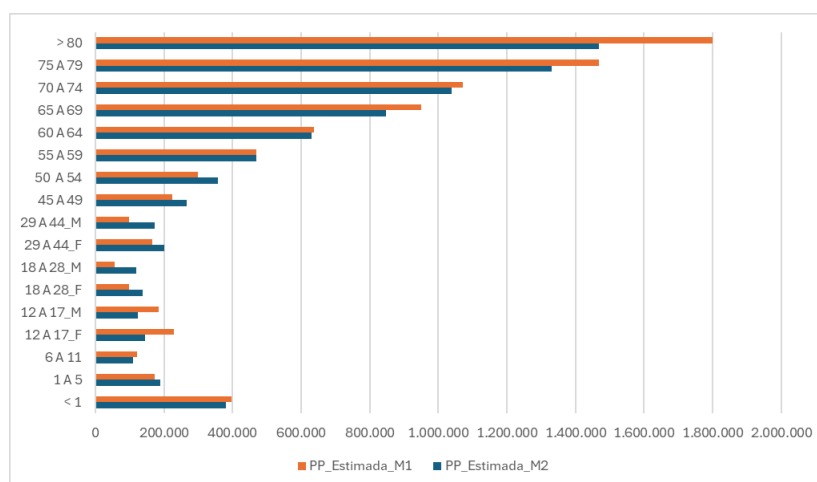


Figura 238: Prima Pura Comparada entre los modelos Zona Rural Disperso Año 2019

Por lo prestado se puede inferir que el modelo de frecuencia y severidad realizado por cada una de las zonas presenta un mejor ajuste que el modelo que considera las zonas como variable explicativa del gasto.

## 5. Discusión y Conclusiones

La Unidad de Pago por Capitación es la prima que el sistema general de seguridad social en salud reconoce a cada una de las EPS con el fin de garantizar la prestación de las tecnologías y servicios de salud definidos, el presente trabajo tiene como finalidad presentar una manera alternativa de estimación de la misma, toda vez que es necesario determinar el uso eficiente de los recursos que son escasos y cada vez mas existe una presión financiera sobre los sistemas de salud derivados de envejecimiento de la población, esta alternativa presenta nuevos ajustadores de riesgo que permitirían una distribución de riesgo que asegure la suficiencia de la prima para dar respuesta a las necesidades de la población.

El análisis y los modelos presentados se realizaron con la información proveniente de las EPS del régimen contributivo a la cual se aplican los procesos de calidad descritos en el Anexo 1 del presente trabajo, esta información constituyo la base para la estimación de la Unidad de Pago por Capitación por el Ministerio de Salud y Protección Social, ahora bien los resultados presentados con la aplicación de los Modelos GLM, no son comparables con la metodología actual utilizada por el órgano rector del sector salud, esto debido a que el presente modelo propone la modificación de las actuales zonas de reconocimiento de la UPC y de los factores de riesgo edad y sexo, lo cuales corresponden a la propuesta de curso de vida desarrollada a través de la Resolución 3280 de 2018.

Lo expresado de manera precedente, se expresa en los resultados de la prima pura de cada una de las zonas donde los los grupos con reconocimiento de un valor mayor son los menores de un año dado que en este periodo del curso de vida demandan mayores atenciones en salud por lo que la prima que el Sistema General de Seguridad Social en Salud reconoce es mayor, situación similar ocurre con los mayores 45 años y se incrementa dadas las diferentes enfermedades que tienen una relación directa con edad como lo son las enfermedades crónicas y por tanto demandan con mayor frecuencia la atención médica.

En lo presentado hay dos zonas que requerían un estudio mayor estas son la Zona Alejada y la Zona Rural Dispersa, que presentan diferencias al analizar la prima pura de cada uno de los grupos etarios, estas zonas se caracterizan por tener dificultades en acceso a los servicios de salud y de talento humano en salud, adicional que son zonas donde hay mayor afiliación al régimen subsidiado, y el presente trabajo analiza la información del régimen contributivo.

Este trabajo considera la aplicación de Modelos Lineales Generalizados, en cada una de las zonas propuestas, la frecuencia se estima mediante el modelo que sigue una distribución binomial negativa teniendo en cuenta que en todas las zonas la distribución poisson presenta sobredispersión, el modelo Lognormal fue el que mejor ajustó en casi todas las zonas y en los dos años de estudio, con excepción de la Zona Rural Disperso en el año 2019 que la mejor distribución es la Gamma.

Los resultados son representativos de toda la población del régimen contributivo, teniendo en cuenta en la prima pura los ajustadores de riesgo de edad y sexo. Es importante tener en cuenta que a lo largo de los años el régimen subsidiado no cuenta con información de calidad que sea representativa para la estimación de la prima pura.

Ahora bien, la aplicación de Modelos Lineales Generalizados como campo de conocimiento de la estadística aplicada presenta múltiples matices a tener en cuenta, sin embargo, al trabajar con información de la realidad de la prestación de las tecnologías y servicios de salud en Colombia y la variabilidad de práctica médica hace que la aplicación de los modelos sea distante de los resultados teóricos presentados.

A nivel internacional, cada sistema de salud tiene características propias que tienen como objetivo final mantener la salud de la población. En Colombia, el enfoque ha sido de aseguramiento público y colectivo, siendo uno de los países con mayores garantías en cuanto a tecnologías y servicios de salud financiadas. Por las características propias, no hay en la literatura estadística y actuarial estudios que se asemejen a lo abordado en el presente trabajo. Para la construcción, se tuvo en cuenta diferentes artículos, libros y demás revisión bibliográfica adaptadas al contexto colombiano.

Finalmente, la presión financiera y por tanto la sostenibilidad de los sistemas de salud en los últimos años se ha visto afectada por diferentes determinantes de la salud y conllevan a la pérdida de bienestar de la población, demandando así mayores servicios y tecnologías, es por ello que se buscan herramientas estadísticas y actuariales que disminuyan dicha presión y realicen un mejor ajuste de riesgo entre las diferentes aseguradoras.

## Referencias

- [et al., 2009] , A. S.-R., , B. M.-E., and G Sosa-Rubí, S. (2009). Análisis estadístico para datos de conteo: aplicaciones para el uso de los servicios de salud. *Salud Pública de México*.
- [Andrews, 2021] Andrews, M. (2021). *Doing Data Science in R: An Introduction for Social Scientists*. SAGE.
- [Arrow, 1981] Arrow, K. J. (1981). Incertidumbre y análisis de bienestar de las prestaciones médicas.
- [Boucher et al., 2007] Boucher, J.-P., Denuit, M., and Guillén, M. (2007). Risk classification for claim counts: a comparative analysis of various zeroinflated mixed poisson and hurdle models. *North American Actuarial Journal*, 11(4):110–131.
- [Caballero Otálora, 2020] Caballero Otálora, A. M. (2020). Modelo pago por capitación ajustada por la diabetes mellitus tipo 2 en el sistema de salud colombiano. Master’s thesis, Universidad Nacional de la Plata - Argentina.
- [de Salud de Protección Social, 2021] de Salud de Protección Social, M. (2021). *Estudio de suficiencia y de los mecanismos de ajuste de riesgo para el cálculo de la unidad de pago por capitación para garantizar el plan de beneficios en salud para el año 2020*. Ministerio de Salud y Protección Social.
- [de salud y Protección Social, ] de salud y Protección Social, M. Propuesta nuevas zonas upc. Documento de trabajo.
- [Duncan, 2018] Duncan, I. G. (2018). *Healthcare risk adjustment and predictive modeling*. Actex Publications.
- [Durbán, 2014] Durbán, M. (2014). Modelos lineales generalizados. Technical report, Universidad Carlos III.
- [Frees, 2014] Frees, E. W. (2014). *Frequency and Severity Models*, volume 1 of *International Series on Actuarial Science*, page 138–164. Cambridge University Press.
- [Giedion et al., 2014] Giedion, U., Bitrán, R., Tristao, I., En, C., Fuente Proporcionada Por, L., et al. (2014). Planes de beneficios en salud de américa latina. *Health Benefit Plans in Latin America (Spanish: Planes de beneficios en salud de América Latina)*.
- [Goldburd and Guller, 2019] Goldburd, Mark, K. A. T. D. and Guller, D. (2019). Generalized linear models for insurance rating. Master’s thesis, CASUALTY ACTUARIAL SOCIETY.
- [Grupo Curso de Vida, 2015] Grupo Curso de Vida, D. d. P. y. P. (2015). AbcÉ enfoque curso de vida. Technical report, Ministerio de Salud y Protección Social Colombia.
- [Gutiérrez Meléndez, 2022] Gutiérrez Meléndez, G. (2022). Comparativa de los modelos glm y gbm para la tarificación de una cartera de autos.
- [Lancheros, 2011] Lancheros, D. (2011). Tarificación: elemento central de la actividad aseguradora. aplicación a los seguros generales. *Bogotá, Colombia. Fasecolda*.
- [Lehmann and D’Abrera, 2006] Lehmann, E. L. E. L. and D’Abrera, H. J. M. (2006). *Nonparametrics : statistical methods based on ranks / Erich L. Lehmann ; with the special assistance of H.J.M. D’Abrera*. Springer, New York, rev. 1st ed. edition.
- [Li, 2022] Li, Y. (2022). Comparación de tarificación de seguro de auto por glm y redes neuronales.
- [Mecer, Raul, 2016] Mecer, Raul (2016). Enfoque curso de vida. fundamentos teóricos, metodológicos y aplicación en programas de salud infantil.
- [Organización Panamericana de la Salud, ] Organización Panamericana de la Salud, Organización Mundial de la Salud year=2014, p. Estrategia para la cobertura universal de salud: Resolución ce154.r17.
- [Ruiz Gómez, 2013] Ruiz Gómez, F. (2013). Mercados y sistema de salud en colombia.
- [Salinas-Rodríguez et al., 2009] Salinas-Rodríguez, A., Manrique-Espinoza, B., and Sosa-Rubí, S. G. (2009). Análisis estadístico para datos de conteo: aplicaciones para el uso de los servicios de salud. *salud pública de méxico*, 51:397–406.

[Werner and Modlin, 2010] Werner, G. and Modlin, C. (2010). Basic ratemaking. In *Casualty Actuarial Society*, volume 4, pages 1–320.

## A. Anexos

### A.1. Anexo 1. Tratamiento de datos

Las bases utilizadas en el presente trabajo de grado corresponden al reporte realizado por las EPS al Ministerio de Salud y Protección Social, con el fin de estimar el incremento de la Unidad de Pago por Capitación - UPC. Para lo anterior se ha desarrollado un proceso estructurado, que, a partir de unos parámetros de solicitud y presentación de la información, realiza un procesamiento tendiente a la verificación de la calidad y cobertura, conformación de la base de datos y retroalimentación. [de Salud de Protección Social, 2021]

El cual se presenta en el siguiente diagrama

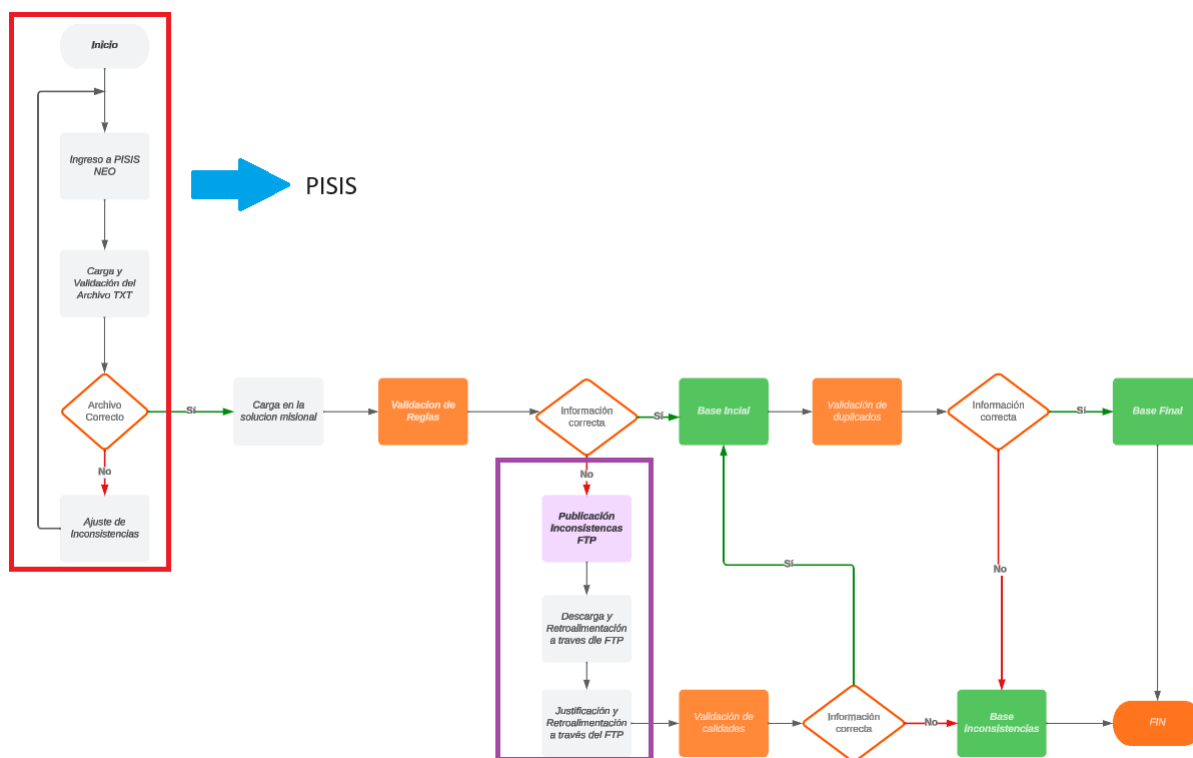


Figura 239: Flujo información UPC  
[de Salud de Protección Social, 2021]

Esta información ´pasa por 182 reglas de calidad agrupadas así:

- **Estructura del archivo:** Evalúa la completitud de los campos
- **Consistencia interna:** Evalúa si la información corresponde a tecnologías y servicios de salud financiados con la UPC Según las tablas de referencia de procedimientos, medicamentos e insumos.
- **Validación Cruzada:** Dentro de los procesos de calidad se valida que la forma de reconocimiento diferente a capita sea mayor a cero, así mismo se evalúa correspondencia entre Tabla de enfermedades establecidas en código internacional de enfermedades revisión 10 CIE10 y el sexo
- **Personas:** Evalúa el estado de afiliación de la persona al momento de la prestación del servicio, ase como correspondencia entre el reporte de insumos y correspondiente procedimiento.
- **Atenciones:** Evalúa aquellos procedimientos que clínicamente no pueden ser practicados más de una vez al día, al mes, o en la vida.
- **Dosis Máxima:** verifica la cantidad dispensada de medicamento en relación con la dosis máxima teórica calculada para el mismo.

- **Valores atípicos:** identifica los valores distantes del resto de los datos, con base en criterios paramétricos y no paramétricos.
  
- **Cruce CUPS con CIE-10:** Verifica la correspondencia del Procedimiento con el Diagnostico.

Una vez han pasado los procesos de calidad la selección de las EPS que entran a formar parte de la base para el calculo de la UPC se realiza teniendo en cuenta la relación entre lo reportado financieramente y lo soportado en bases de datos de prestación de servicios, cuyo criterio sea igual o superior al 90%.

Con esta base se realiza la siguiente depuración de datos:

1. Se elimina aquella información donde las variables bajo estudio se encuentran en cero (Reclamaciones, Valor de las Reclamaciones, Póliza, Días Compensados)
  
2. Se elimina aquella información donde se cuenta con reclamaciones pero las demás variables se encuentran en cero (Valor de las Reclamaciones, Póliza, Días Compensados), lo anterior pues no es consistente con el sistema general de seguridad social en salud

La base por tanto se compone de los registros con los que se cuenta:

1. Información en todas las variables bajo estudio los cuales afectan la póliza
  
2. Afiliados que se encuentren en protección laboral y por tanto cuentan con una "póliza extendida", es decir que se presentan reclamaciones y valor de reclamaciones pero no se cuenta con información de reconocimiento de a póliza (UPC y PyP)
  
3. Afiliados por los cuales se recibe recursos por la póliza, pero no la afectan es decir que no presentan demanda de servicios de salud durante el periodo de análisis.

La información se agrupa por zonas y grupos etarios y sexo que son las variables de análisis en los modelos seleccionados de frecuencia y severidad.

## A.2. Anexo 2. Aplicación de modelos alternos por zona

### A.3. Modelos de Frecuencias

#### A.3.1. Zona Alejada

GLM Poisson

A continuación se presentan los resultados del modelo poisson para los años 2018 y 2019, teniendo en cuenta las variables de Grupo de Edad, y Sexo que desde la literatura de seguros son las que pueden llegar afectar la frecuencia de uso de las personas.

```
call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2018_A)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.805124	0.012761	141.461	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO10	0.590541	0.014546	40.598	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO11	0.900220	0.014903	60.407	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO12	0.893033	0.015940	56.025	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO13	1.015113	0.017148	59.197	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO14	1.002672	0.016620	60.328	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.051221	0.015197	-3.371	0.00075	***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.746453	0.015655	-47.683	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.837197	0.016458	-50.869	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.589888	0.013996	-42.148	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.258410	0.013460	-19.199	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.034099	0.014911	-2.287	0.02221	*
GRUPO_ETAREO_RUTA_AGREGADO8	0.167604	0.014367	11.666	< 0.0000000000000002	***
GRUPO_ETAREO_RUTA_AGREGADO9	0.377497	0.014168	26.645	< 0.0000000000000002	***
SEXOM	-0.444784	0.004245	-104.774	< 0.0000000000000002	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 88905.1 on 27 degrees of freedom  
 Residual deviance: 7277.2 on 13 degrees of freedom  
 AIC: 7606.8

Number of Fisher Scoring iterations: 4

Figura 240: Modelo Poisson Zona Alejada Año 2018

```

Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2019_A)

Coefficients:
                Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.560004   0.008307  308.170 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.733580   0.009324   78.673 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  0.960867   0.009637   99.709 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.188423   0.009934  119.634 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.392491   0.010481  132.859 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.405717   0.010122  138.879 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.208389   0.010312  -20.208 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.921520   0.010699  -86.129 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.891389   0.011047  -80.689 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.495195   0.009126  -54.261 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.339618   0.008825  -38.482 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7   0.004632   0.009792   0.473      0.636
GRUPO_ETAREO_RUTA_AGREGADO8   0.286806   0.009332   30.733 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9   0.480542   0.009179   52.352 <0.0000000000000002 ***
SEXOM                -0.419882   0.002751 -152.623 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 295389  on 27  degrees of freedom
Residual deviance: 23361  on 13  degrees of freedom
AIC: 23715

Number of Fisher Scoring iterations: 4

```

Figura 241: Modelo Poisson Zona Alejada 2019

Para el 2018 y el 2019, las variables son significativas con una confianza del 95%.

### A.3.2. Zona Bogotá

- GLM Poisson El siguiente es el modelo para los años 2018 y 2019 teniendo en cuenta las variables de Grupo de Edad, y Sexo que desde la literatura de seguros son las que pueden llegar afectar la frecuencia de uso de las personas.

```
Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2018_BOG)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.8652618	0.0006739	4251.6	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	0.6760635	0.0007510	900.2	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11	0.9407550	0.0007539	1247.9	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12	1.1566634	0.0007671	1507.9	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13	1.3212605	0.0007875	1677.9	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14	1.4301829	0.0007521	1901.5	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.3121863	0.0008471	-368.5	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.8812144	0.0008425	-1046.0	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.7667424	0.0008398	-913.0	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.5671482	0.0007224	-785.1	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.3612533	0.0007068	-511.1	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.0837988	0.0007844	-106.8	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8	0.1601047	0.0007638	209.6	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9	0.3917371	0.0007535	519.9	<0.0000000000000002 ***
SEXOM	-0.3983495	0.0002018	-1973.7	<0.0000000000000002 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 58153878 on 27 degrees of freedom  
 Residual deviance: 1975373 on 13 degrees of freedom  
 AIC: 1975874

Number of Fisher Scoring iterations: 4

Figura 242: Modelo Poisson Zona Bogotá Año 2018

```

Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2019_BOG)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.8330418  0.0006965  4067.54 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.8135081  0.0007603  1070.05 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.0826389  0.0007621  1420.53 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.3066271  0.0007714  1693.83 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.4765776  0.0007899  1869.30 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.5697438  0.0007605  2064.04 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2   -0.2302882  0.0008610  -267.47 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3   -0.8095490  0.0008552  -946.63 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4   -0.7136268  0.0008605  -829.36 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5   -0.4879680  0.0007423  -657.40 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6   -0.2761991  0.0007259  -380.49 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7    0.0247132  0.0007959    31.05 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8    0.2859058  0.0007757    368.60 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9    0.5235630  0.0007638    685.43 <0.0000000000000002 ***
SEXOM                    -0.3896763  0.0001914  -2036.35 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 70582379 on 27 degrees of freedom
Residual deviance: 2228468 on 13 degrees of freedom
AIC: 2228971

Number of Fisher Scoring iterations: 4

```

Figura 243: Modelo Poisson Zona Bogotá Año 2019

Las variables en los años de estudio son significativas con una confianza del 95 %.

### A.3.3. Zona Normal Mayor

- GLM Poisson

El siguiente es el modelo teniendo en cuenta las variables de Grupo de Edad, y Sexo que desde la literatura de seguros son las que pueden llegar afectar la frecuencia de uso de las personas, para los años 2018 y 2019.

```

Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2018_NM)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.7804217  0.0005219  5327.51 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.8124838  0.0005738  1416.05 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.0663309  0.0005744  1856.55 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.2523298  0.0005832  2147.51 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.3877929  0.0005976  2322.30 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.4750278  0.0005726  2576.03 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.2022789  0.0006525  -310.00 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.8231211  0.0006576 -1251.76 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.7429814  0.0006576 -1129.78 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.5284622  0.0005614  -941.35 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.2664946  0.0005464  -487.76 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7   0.0300342  0.0006009   49.98 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8   0.3026212  0.0005834   518.68 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9   0.5447407  0.0005758   946.04 <0.0000000000000002 ***
SEXOM                -0.3952889  0.0001489 -2655.58 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 112394854 on 27 degrees of freedom
Residual deviance: 3635568 on 13 degrees of freedom
AIC: 3636085

Number of Fisher Scoring iterations: 4

```

Figura 244: Modelo Poisson Zona Normal Mayor Año 2018

```

Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2019_NM)

Coefficients:
                Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.7924882   0.0005207  5362.6 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.9194255   0.0005646  1628.6 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.1798124   0.0005643  2090.9 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.3718498   0.0005710  2402.6 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.5209812   0.0005842  2603.5 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.6117905   0.0005620  2867.9 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2   -0.1214413   0.0006388  -190.1 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3   -0.7347300   0.0006432 -1142.4 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4   -0.6354062   0.0006433  -987.8 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5   -0.4472808   0.0005567  -803.4 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6   -0.1849839   0.0005420  -341.3 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7    0.1283601   0.0005911   217.1 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8    0.4052246   0.0005751   704.6 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9    0.6485596   0.0005668  1144.2 <0.0000000000000002 ***
SEXOM                   -0.3782620   0.0001378 -2744.6 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 136649292  on 27  degrees of freedom
Residual deviance:  4192342  on 13  degrees of freedom
AIC: 4192864

Number of Fisher scoring iterations: 4

```

Figura 245: Modelo Poisson Zona Normal Mayor Año 2019

Las variables para los dos años son significativas con una confianza del 95 %.

#### A.3.4. Zona Normal Medio

- GLM Poisson El siguiente es el modelo teniendo en cuenta las variables de Grupo de Edad, y Sexo que desde la literatura de seguros son las que pueden llegar afectar la frecuencia de uso de las personas para los años 2018 y 2019.

```

Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2018_NME)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    2.6466179  0.0008941  2960.08 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.7782251  0.0010032   775.72 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.0233007  0.0010077  1015.52 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.2159406  0.0010265  1184.59 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.3601950  0.0010543  1290.17 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.4200057  0.0010123  1402.82 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.2070175  0.0011127  -186.05 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.8066344  0.0011057  -729.54 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.7339021  0.0011012  -666.45 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.6006088  0.0009677  -620.65 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.3239767  0.0009405  -344.49 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.0226210  0.0010400   -21.75 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8   0.2515766  0.0010150   247.86 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9   0.5015831  0.0010055   498.86 <0.0000000000000002 ***
SEXOM           -0.4614855  0.0002716 -1699.27 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 33009106 on 27 degrees of freedom
Residual deviance: 1418136 on 13 degrees of freedom
AIC: 1418620

Number of Fisher Scoring iterations: 4

```

Figura 246: Modelo Poisson Zona Normal Medio Año 2018

```

Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2019_NME)

Coefficients:
                Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.6883738  0.0008781  3061.6 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.8342084  0.0009724   857.8 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.0860468  0.0009744  1114.6 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.2804557  0.0009893  1294.3 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.4360246  0.0010158  1413.7 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.5137038  0.0009750  1552.5 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2   -0.1472400  0.0010758  -136.9 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3   -0.7375832  0.0010741  -686.7 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4   -0.6493999  0.0010674  -608.4 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5   -0.5582273  0.0009461  -590.1 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6   -0.2696144  0.0009188  -293.5 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7    0.0408623  0.0010089    40.5 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8    0.3092330  0.0009856   313.8 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9    0.5546498  0.0009751   568.8 <0.0000000000000002 ***
SEXOM                    -0.4496281  0.0002517 -1786.4 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 39274046 on 27 degrees of freedom
Residual deviance: 1663235 on 13 degrees of freedom
AIC: 1663723

Number of Fisher scoring iterations: 4

```

Figura 247: Modelo Poisson Zona Normal Medio Año 2019

Las variables en los dos años son significativas con una confianza del 95 %.

### A.3.5. Zona Rural

- GLM Poisson

El siguiente es el modelo teniendo en cuenta las variables de Grupo de Edad, y Sexo que desde la literatura de seguros son las que pueden llegar afectar la frecuencia de uso de las personas para los años 2018 y 2019.

Call:

```
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2018_RU)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.3171485	0.0018327	1264.33	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	0.7013316	0.0020967	334.49	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11	0.9891798	0.0021136	468.01	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12	1.1958791	0.0021640	552.63	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13	1.3435473	0.0022407	599.61	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14	1.3917991	0.0021280	654.03	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.2893935	0.0023107	-125.24	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.9252013	0.0023061	-401.21	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.7798785	0.0022718	-343.28	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.8905539	0.0020059	-443.96	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.5496733	0.0019327	-284.40	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.1870022	0.0021494	-87.00	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8	0.0893830	0.0021116	42.33	<0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9	0.3897368	0.0020916	186.34	<0.0000000000000002 ***
SEXOM	-0.5732989	0.0005883	-974.56	<0.0000000000000002 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 8217121 on 27 degrees of freedom  
 Residual deviance: 444148 on 13 degrees of freedom  
 AIC: 444589

Number of Fisher Scoring iterations: 4

Figura 248: Modelo Poisson Zona Rural Año 2018

```

Call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2019_RU)

Coefficients:
                Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.4511568   0.0016989  1442.76 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.7402611   0.0019268   384.19 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  1.0285900   0.0019405   530.06 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  1.2584937   0.0019811   635.24 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.4092427   0.0020516   686.89 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.4721545   0.0019489   755.37 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.2635761   0.0021264  -123.96 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.8394606   0.0021274  -394.60 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.7156723   0.0020964  -341.39 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.9314510   0.0018638  -499.76 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.5354375   0.0017878  -299.50 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.1604719   0.0019841   -80.88 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8   0.1226503   0.0019459   63.03 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9   0.4184326   0.0019260   217.25 <0.0000000000000002 ***
SEXOM              -0.5681017   0.0005319 -1068.09 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 10549952  on 27  degrees of freedom
Residual deviance:  562473  on 13  degrees of freedom
AIC: 562919

Number of Fisher Scoring iterations: 4

```

Figura 249: Modelo Poisson Zona Rural Año 2019

Las variables en los dos años son significativas con una confianza del 95 %.

### A.3.6. Zona Rural Disperso

- GLM Poisson

El siguiente es el modelo teniendo en cuenta las variables de Grupo de Edad, y Sexo que desde la literatura de seguros son las que pueden llegar afectar la frecuencia de uso de las personas para los años 2018 y 2019.

Call:

```
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2018_RUD)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.431448	0.005228	465.116	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	0.252495	0.006643	38.009	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11	0.569491	0.006825	83.439	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12	0.843537	0.007072	119.271	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13	0.990913	0.007534	131.523	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14	1.046487	0.006983	149.866	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.312926	0.006800	-46.016	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-1.029836	0.006974	-147.668	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.896365	0.006856	-130.745	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-1.422426	0.005858	-242.800	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-1.038578	0.005588	-185.873	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.610298	0.006448	-94.647	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.345027	0.006453	-53.471	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9	0.018306	0.006402	2.859	0.00425 **
SEXOM	-0.702416	0.002070	-339.325	< 0.0000000000000002 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 667370 on 27 degrees of freedom  
 Residual deviance: 45773 on 13 degrees of freedom  
 AIC: 46143

Number of Fisher Scoring iterations: 4

Figura 250: Modelo Poisson Zona Rural Disperso Año 2018

```

call:
glm(formula = RECLAMACIONES ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO +
     offset(log(EXPUESTOS)), family = poisson(), data = BASE_2019_RUD)

Coefficients:
                Estimate Std. Error z value      Pr(>|z|)
(Intercept)      2.452351   0.004869   503.71 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10  0.434065   0.006016    72.15 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO11  0.788260   0.006126   128.68 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO12  0.961153   0.006453   148.94 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO13  1.163689   0.006941   167.65 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO14  1.314871   0.006237   210.83 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO2   -0.214165   0.006176   -34.68 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO3   -0.882169   0.006378  -138.31 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO4   -0.734188   0.006267  -117.15 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO5   -1.378316   0.005465  -252.21 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO6   -0.984841   0.005207  -189.14 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO7   -0.481276   0.005944   -80.97 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO8   -0.204470   0.005904   -34.63 <0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO9    0.101675   0.005910    17.20 <0.0000000000000002 ***
SEXOM                   -0.705566   0.001867  -377.97 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 914010  on 27  degrees of freedom
Residual deviance:  59962  on 13  degrees of freedom
AIC: 60338

Number of Fisher scoring iterations: 4

```

Figura 251: Modelo Poisson Zona Rural Disperso Año 2019

Las variables son significativas para los dos años con una confianza del 95 %,

## A.4. Modelos de Severidad

### A.4.1. Zona Bogotá

- GLM Gamma

A continuación se presentan los resultados de la aplicación del modelo Gamma, para modelar la severidad de la zona Bogotá para el año 2018.

```

Call:
glm(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO,
     family = Gamma(link = "log"), data = BASE_2018_BOG)

Coefficients:
              Estimate Std. Error t value      Pr(>|t|)
(Intercept)    11.27790    0.04407  255.882 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10 -0.55233    0.06022   -9.172    0.0000004856 ***
GRUPO_ETAREO_RUTA_AGREGADO11 -0.56170    0.06022   -9.328    0.0000004009 ***
GRUPO_ETAREO_RUTA_AGREGADO12 -0.55030    0.06022   -9.138    0.0000005064 ***
GRUPO_ETAREO_RUTA_AGREGADO13 -0.52075    0.06022   -8.648    0.0000009438 ***
GRUPO_ETAREO_RUTA_AGREGADO14 -0.45164    0.06022   -7.500    0.0000044993 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.68142    0.06022  -11.316    0.0000000421 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.76298    0.06022  -12.670    0.0000000109 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.63135    0.06022  -10.485    0.0000001036 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.70164    0.06022  -11.652    0.0000000297 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.58088    0.06022   -9.646    0.0000002728 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.55437    0.06022   -9.206    0.0000004657 ***
GRUPO_ETAREO_RUTA_AGREGADO8  -0.56553    0.06022   -9.392    0.0000003709 ***
GRUPO_ETAREO_RUTA_AGREGADO9  -0.53459    0.06022   -8.878    0.0000007029 ***
SEXOM                0.24022    0.02276   10.554    0.0000000959 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.003626135)

Null deviance: 1.353619  on 27  degrees of freedom
Residual deviance: 0.047226  on 13  degrees of freedom
AIC: 540.33

Number of Fisher scoring iterations: 4

```

Figura 252: Modelo Gamma para Severidad Zona Bogotá Año 2018

En el modelo Gamma las variables son significativas al 95%.

A continuación se presenta los residuales del modelo gamma para la zona Bogotá.

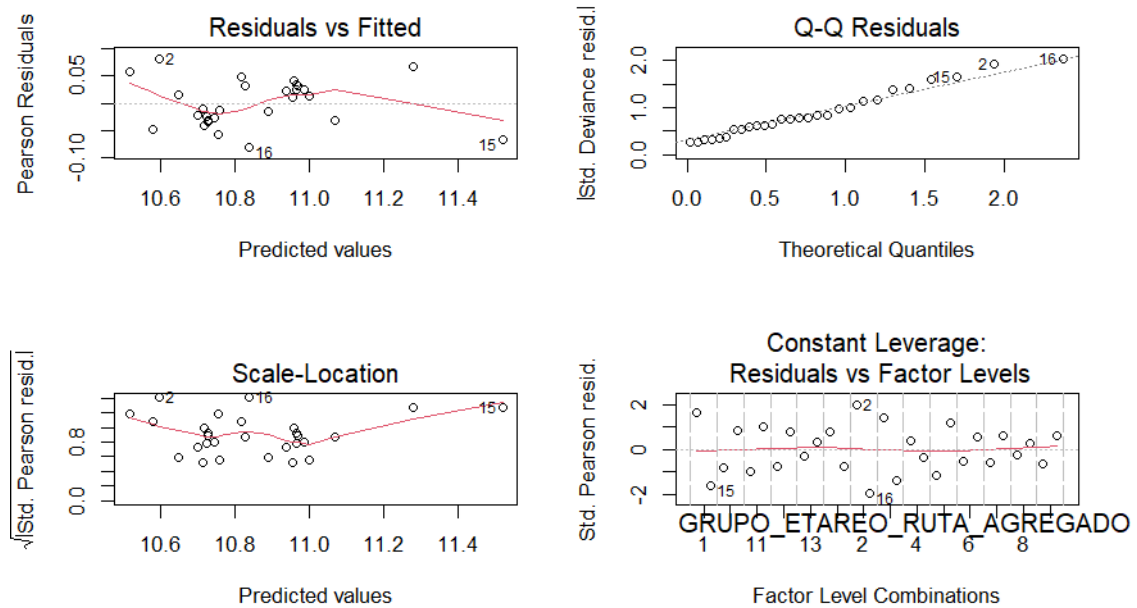


Figura 253: Residuales Modelo Gamma para Severidad Zona Bogotá Año 2018

Como se observa en la figura 253, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 91 y figura 254, presentando puntos influyentes en las observaciones 2,4,5 y16, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
2	2.2164273	0.5357143	0.300492564
4	0.3448445	0.5357143	0.009875402
5	-1.1745388	0.5357143	0.101576432
16	-2.3264835	0.5357143	0.300492636

Tabla 91: Puntos Influyentes Zona Bogotá Año 2018

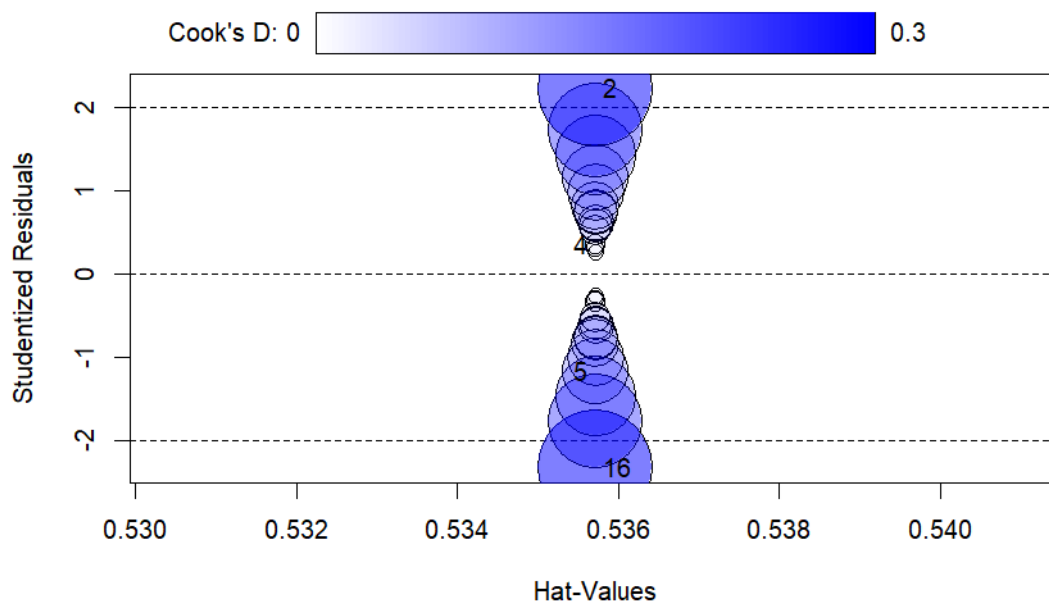


Figura 254: Puntos Influyentes Zona Bogotá Año 2018

Una vez realizados los análisis del modelo para el año 2018, se presentan los resultados de la estimación de la severidad para la zona Bogota.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	89.127,36	89.907,44
Entre 1 y 5 Años	Ambos	45.003,01	45.485,40
Entre 6 y 11 Años	Ambos	41.624,43	41.950,68
Entre 12 y 17 Años	Femenino	42.665,36	42.047,20
Entre 12 y 17 Años	Masculino	52.678,03	53.464,04
Entre 18 y 28 Años	Femenino	37.345,39	39.193,36
Entre 18 y 28 Años	Masculino	52.185,05	49.835,31
Entre 29 y 44 Años	Femenino	43.256,10	44.224,03
Entre 28 y 44 Años	Masculino	57.462,68	56.231,93
Entre 45 y 49 Años	Ambos	51.403,98	51.307,37
Entre 50 y 54 Años	Ambos	50.632,97	50.602,11
Entre 55 y 59 Años	Ambos	52.128,39	52.080,20
Entre 60 y 64 Años	Ambos	51.046,17	51.023,45
Entre 65 y 69 Años	Ambos	50.506,69	50.497,08
Entre 70 y 74 Años	Ambos	50.982,54	50.996,21
Entre 75 y 79 Años	Ambos	52.304,90	52.331,77
Mayores de 80 Años	Ambos	55.554,41	55.276,07

Tabla 92: Severidad Observada y Estimada Modelo Gamma Zona Bogota Año 2018

A continuación se presenta gráficamente la severidad observada y la severidad estimada del modelo Gamma.

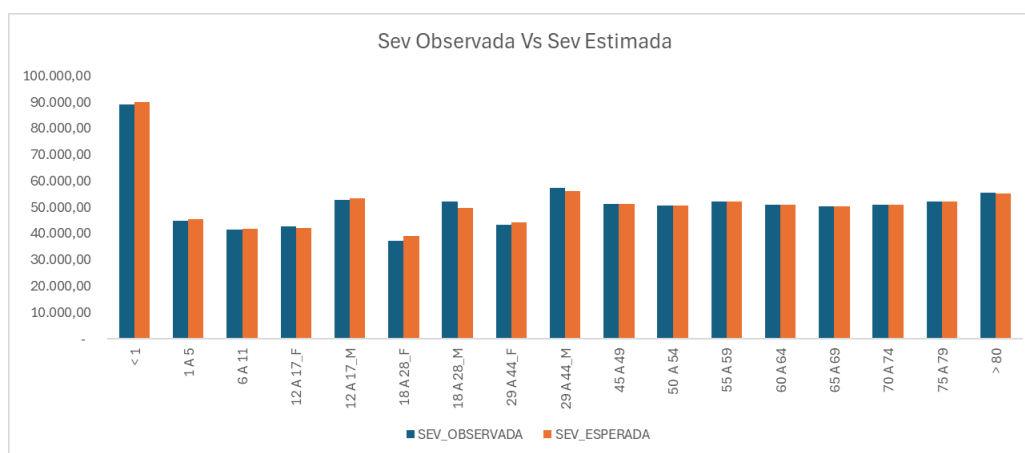


Figura 255: Severidad Observada Vs Severidad Estimada Zona Bogotá Año 2018

El modelo presenta un error porcentual absoluto medio (MAPE) del 1.2% lo cual se presenta un ajuste aceptable, y un pseudo R2 de 0.96.

A continuación se presentan los resultados de la aplicación del modelo Gamma, para modelar la severidad de la zona Bogotá para el año 2019.

```
call:
glm(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO,
     family = Gamma(link = "log"), data = BASE_2019_BOG)

Coefficients:
                Estimate Std. Error t value      Pr(>|t|)
(Intercept)      11.26331    0.03998 281.727 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10 -0.57922    0.05462 -10.604  0.00000009073 ***
GRUPO_ETAREO_RUTA_AGREGADO11 -0.59952    0.05462 -10.976  0.00000006047 ***
GRUPO_ETAREO_RUTA_AGREGADO12 -0.60406    0.05462 -11.059  0.00000005530 ***
GRUPO_ETAREO_RUTA_AGREGADO13 -0.57533    0.05462 -10.533  0.00000009820 ***
GRUPO_ETAREO_RUTA_AGREGADO14 -0.54322    0.05462  -9.945  0.00000019182 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.72536    0.05462 -13.279  0.00000000614 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.70752    0.05462 -12.953  0.00000000831 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.62110    0.05462 -11.371  0.00000003976 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.65142    0.05462 -11.926  0.00000002251 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.55189    0.05462 -10.104  0.00000015959 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.54771    0.05462 -10.027  0.00000017435 ***
GRUPO_ETAREO_RUTA_AGREGADO8  -0.56309    0.05462 -10.309  0.00000012629 ***
GRUPO_ETAREO_RUTA_AGREGADO9  -0.58525    0.05462 -10.715  0.00000008032 ***
SEXOM                    0.22093    0.02065  10.701  0.00000008152 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.002983605)

Null deviance: 1.24850 on 27 degrees of freedom
Residual deviance: 0.03885 on 13 degrees of freedom
AIC: 532.69

Number of Fisher Scoring iterations: 4
```

Figura 256: Modelo Gamma para Severidad Zona Bogotá Año 2019

En el modelo Gamma las variables para el año 2019 son significativas al 95%, por lo anterior se continua con el análisis, presentando el comportamiento de los residuales.

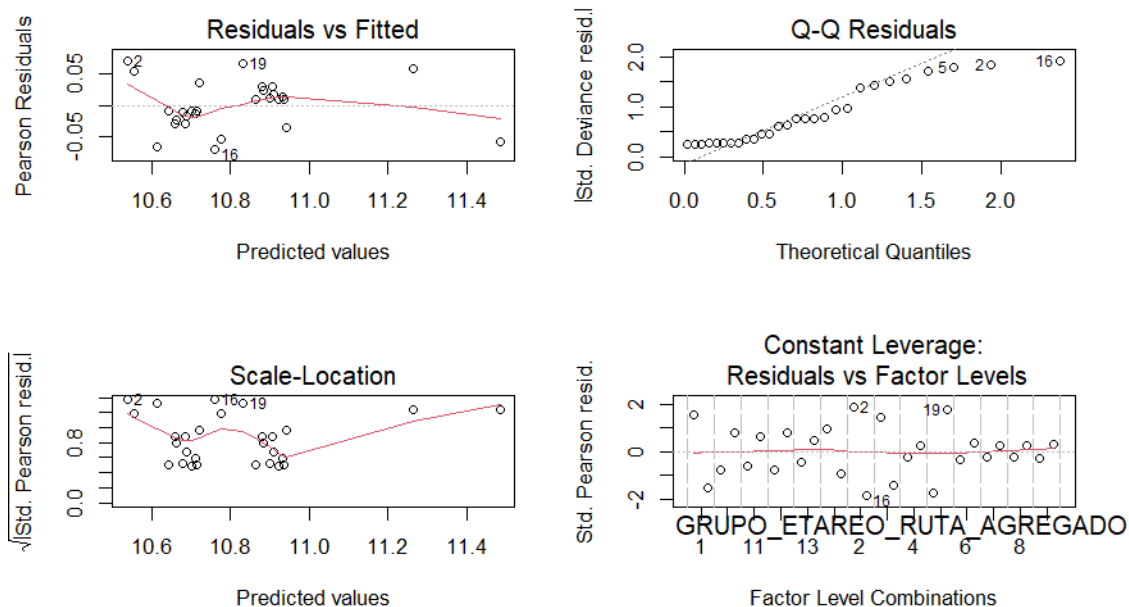


Figura 257: Residuales Modelo Gamma para Severidad Zona Bogotá Año 2019

Como se observa en la figura 257, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 93 y figura 258, presentando puntos influyentes en las observaciones 2,4,5 y16, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
2	2.0631992	0.5357143	0.269467162
4	-0.2518245	0.5357143	0.005249454
5	-1.9677180	0.5357143	0.237852884
16	-2.1447443	0.5357143	0.269467182

Tabla 93: Puntos Influyentes LogNormal Zona Bogotá Año 2019

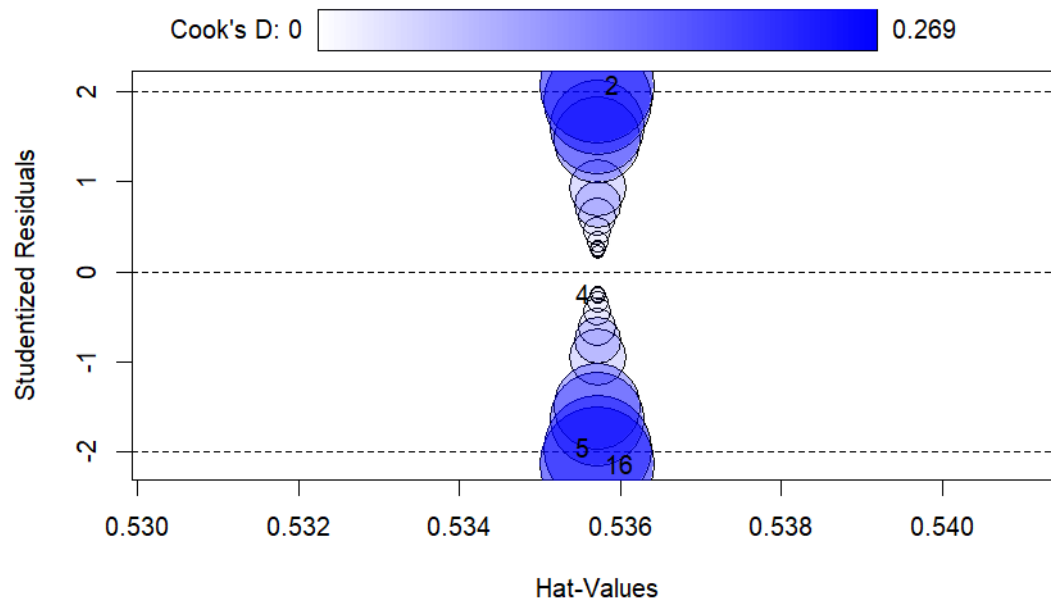


Figura 258: Puntos Influyentes LogNormal Zona Bogotá Año 2019

Una vez realizados los análisis del modelo para el año 2019, se presentan los resultados de la estimación de la severidad para la zona Bogota.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	87.103,19	87.770,17
Entre 1 y 5 Años	Ambos	42.094,76	42.482,66
Entre 6 y 11 Años	Ambos	42.944,88	43.239,70
Entre 12 y 17 Años	Femenino	41.457,88	41.864,93
Entre 12 y 17 Años	Masculino	52.722,94	52.215,26
Entre 18 y 28 Años	Femenino	37.956,76	40.614,88
Entre 18 y 28 Años	Masculino	53.971,45	50.656,16
Entre 29 y 44 Años	Femenino	44.290,69	44.865,06
Entre 28 y 44 Años	Masculino	56.673,49	55.957,12
Entre 45 y 49 Años	Ambos	50.406,70	50.374,23
Entre 50 y 54 Años	Ambos	49.512,33	49.492,49
Entre 55 y 59 Años	Ambos	48.313,30	48.300,77
Entre 60 y 64 Años	Ambos	48.489,93	48.483,46
Entre 65 y 69 Años	Ambos	47.431,01	47.440,39
Entre 70 y 74 Años	Ambos	47.167,59	47.188,69
Entre 75 y 79 Años	Ambos	48.365,48	48.403,43
Mayores de 80 Años	Ambos	49.621,38	49.332,59

Tabla 94: Severidad Observada y Estimada Modelo Gamma Zona Bogota Año 2019

A continuación se presenta gráficamente la severidad observada y la severidad estimada del modelo Gamma.

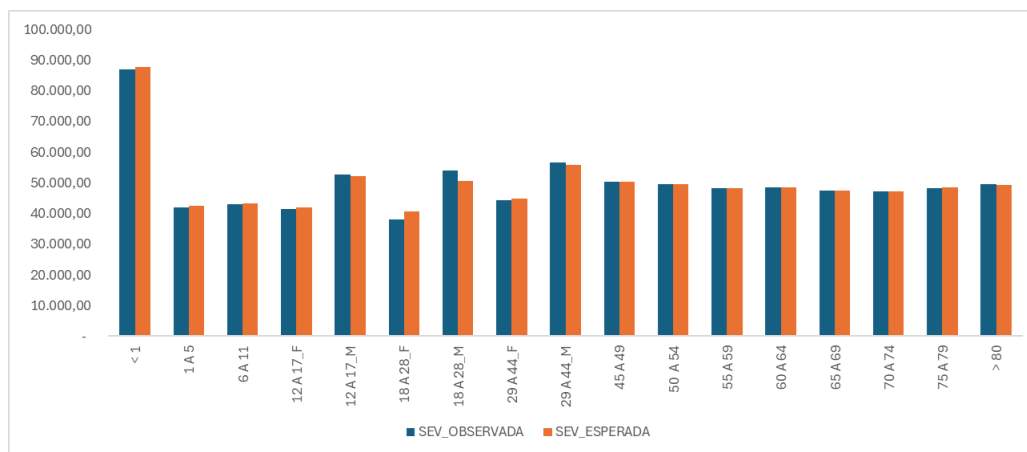


Figura 259: Severidad Observada Vs Severidad Estimada Zona Bogotá Año 2019

El modelo presenta un error porcentual absoluto medio (MAPE) del 1.22% lo cual se presenta un ajuste aceptable, y un pseudo R2 de 0.97

### A.4.2. Zona Normal Mayor

- GLM Gamma

A continuación se presentan los resultados de la aplicación del modelo Gamma, para modelar la severidad de la Zona Normal Mayor para el Año 2018.

```
call:
glm(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO,
     family = Gamma(link = "log"), data = BASE_2018_NM)

Coefficients:
                Estimate Std. Error t value      Pr(>|t|)
(Intercept)      11.09144    0.03278  338.398 < 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10 -0.39967    0.04478   -8.925   0.0000006619 ***
GRUPO_ETAREO_RUTA_AGREGADO11 -0.40601    0.04478   -9.066   0.0000005540 ***
GRUPO_ETAREO_RUTA_AGREGADO12 -0.39790    0.04478   -8.886   0.0000006959 ***
GRUPO_ETAREO_RUTA_AGREGADO13 -0.36819    0.04478   -8.222   0.0000016552 ***
GRUPO_ETAREO_RUTA_AGREGADO14 -0.31120    0.04478   -6.949   0.0000100757 ***
GRUPO_ETAREO_RUTA_AGREGADO2  -0.55933    0.04478  -12.490   0.0000000129 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.47701    0.04478  -10.652   0.0000000860 ***
GRUPO_ETAREO_RUTA_AGREGADO4  -0.34369    0.04478   -7.675   0.0000035114 ***
GRUPO_ETAREO_RUTA_AGREGADO5  -0.44937    0.04478  -10.035   0.0000001728 ***
GRUPO_ETAREO_RUTA_AGREGADO6  -0.38809    0.04478   -8.666   0.0000009213 ***
GRUPO_ETAREO_RUTA_AGREGADO7  -0.36206    0.04478   -8.085   0.0000019913 ***
GRUPO_ETAREO_RUTA_AGREGADO8  -0.36980    0.04478   -8.258   0.0000015770 ***
GRUPO_ETAREO_RUTA_AGREGADO9  -0.38368    0.04478   -8.568   0.0000010473 ***
SEXOM                  0.20762    0.01693   12.267   0.0000000160 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.002005336)

Null deviance: 0.743806 on 27 degrees of freedom
Residual deviance: 0.026098 on 13 degrees of freedom
AIC: 522.12

Number of Fisher Scoring iterations: 4
```

Figura 260: Modelo Gamma para Severidad Zona Normal Mayor Año 2018

En el modelo Gamma las variables son significativas al 95% para el año 2018, por lo cual se decide realizar análisis, teniendo en cuenta esta distribución, para la Zona Normal Mayor.

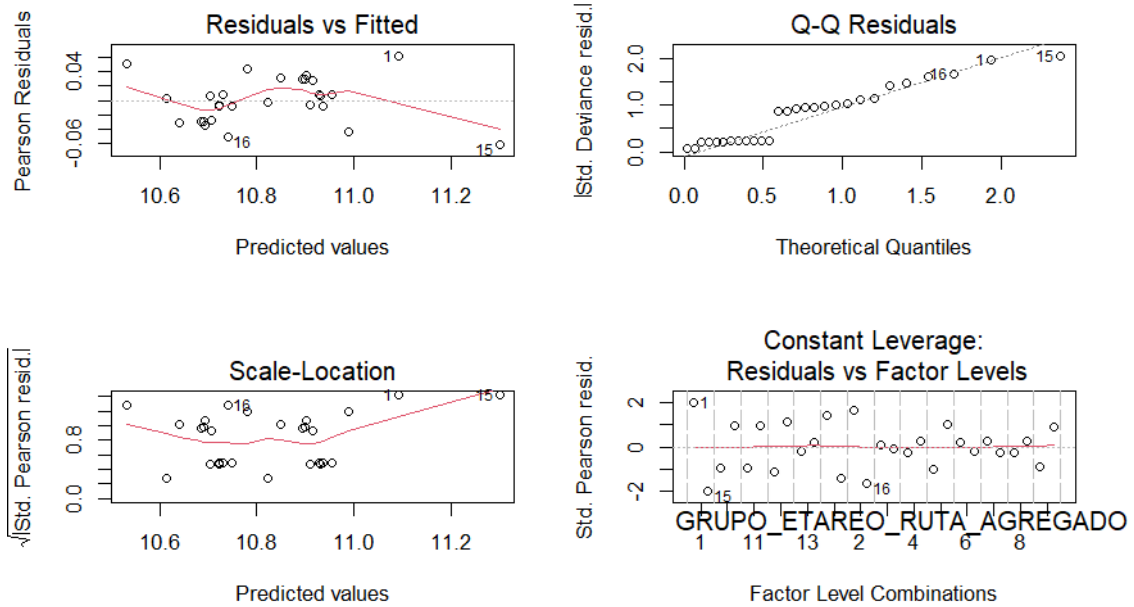


Figura 261: Residuales Modelo Gamma para Severidad Zona Normal Mayor Año 2018

Como se observa en la figura 261, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 95 y figura 262, presentando puntos influyentes en las observaciones 1,4,5 y 15, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
1	2.280115	0.5357143	0.310200225
4	-0.229783	0.5357143	0.004375578
5	-1.029285	0.5357143	0.080343981
15	-2.366992	0.5357143	0.310200243

Tabla 95: Frecuencias Observadas y Estimadas Zona Normal Mayor Año 2018

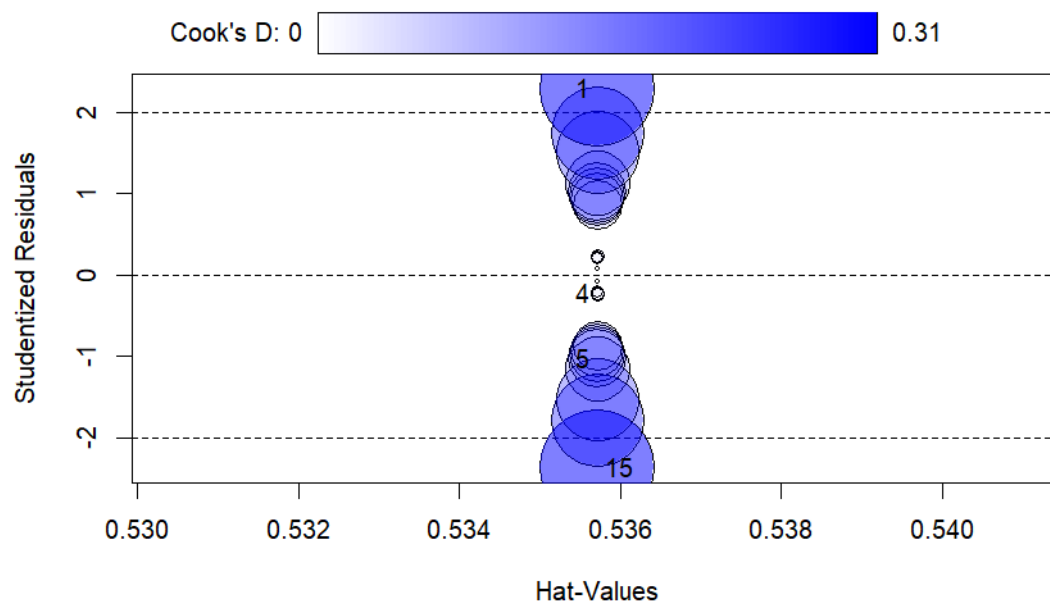


Figura 262: Puntos Influyentes Zona Normal Mayor Año 2018

A continuación se presenta la severidad observada y la severidad estimada en la Zona Normal Mayor utilizando la distribución gamma para el año 2018.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	72.694,02	73.130,73
Entre 1 y 5 Años	Ambos	41.636,72	41.877,21
Entre 6 y 11 Años	Ambos	45.493,11	45.506,75
Entre 12 y 17 Años	Femenino	46.186,65	46.525,23
Entre 12 y 17 Años	Masculino	57.677,67	57.260,96
Entre 18 y 28 Años	Femenino	40.554,01	41.859,36
Entre 18 y 28 Años	Masculino	53.124,99	51.518,43
Entre 29 y 44 Años	Femenino	44.794,79	44.504,43
Entre 28 y 44 Años	Masculino	54.416,50	54.773,86
Entre 45 y 49 Años	Ambos	50.829,78	50.863,33
Entre 50 y 54 Años	Ambos	50.359,71	50.336,61
Entre 55 y 59 Años	Ambos	49.560,49	49.512,25
Entre 60 y 64 Años	Ambos	48.548,90	48.547,55
Entre 65 y 69 Años	Ambos	48.279,52	48.270,13
Entre 70 y 74 Años	Ambos	48.594,26	48.602,86
Entre 75 y 79 Años	Ambos	49.852,47	49.867,13
Mayores de 80 Años	Ambos	52.589,09	52.258,23

Tabla 96: Severidad Observada y Estimada Modelo Gamma Zona Normal Mayor Año 2018

A continuación se presenta gráficamente la severidad observada y la severidad estimada del modelo Gamma.

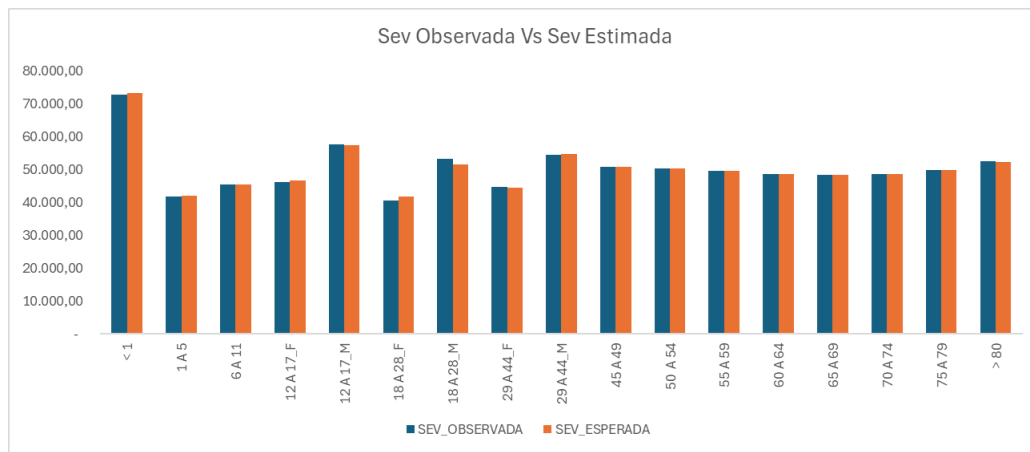


Figura 263: Severidad Observada Vs Severidad Estimada Zona Normal Mayor Año 2018

El modelo presenta un error porcentual absoluto medio (MAPE) del 0.65 % lo cual se presenta un ajuste aceptable y un pseudo R2 de 0.96

De acuerdo a la gráfica de severidad empírica el modelo Gamma no se ajusta a los datos de la zona normal mayor para el año 2019, por lo cual se decide no realizar este modelo.

#### A.4.3. Zona Normal Medio

- GLM Gamma

A continuación se presentan los resultados de la aplicación del modelo Gamma, para modelar la severidad de la Zona Normal Medio para el año 2018.

```
Call:
glm(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO + SEXO,
     family = Gamma(link = "log"), data = BASE_2018_NME)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11.13786	0.04309	258.464	< 0.0000000000000002 ***
GRUPO_ETAREO_RUTA_AGREGADO10	-0.38118	0.05888	-6.474	0.0000208558 ***
GRUPO_ETAREO_RUTA_AGREGADO11	-0.39289	0.05888	-6.673	0.0000153216 ***
GRUPO_ETAREO_RUTA_AGREGADO12	-0.37572	0.05888	-6.382	0.0000241222 ***
GRUPO_ETAREO_RUTA_AGREGADO13	-0.31879	0.05888	-5.415	0.000118 ***
GRUPO_ETAREO_RUTA_AGREGADO14	-0.27064	0.05888	-4.597	0.000501 ***
GRUPO_ETAREO_RUTA_AGREGADO2	-0.69864	0.05888	-11.866	0.0000000239 ***
GRUPO_ETAREO_RUTA_AGREGADO3	-0.64669	0.05888	-10.984	0.0000000599 ***
GRUPO_ETAREO_RUTA_AGREGADO4	-0.43917	0.05888	-7.459	0.0000047696 ***
GRUPO_ETAREO_RUTA_AGREGADO5	-0.45151	0.05888	-7.669	0.0000035415 ***
GRUPO_ETAREO_RUTA_AGREGADO6	-0.40609	0.05888	-6.897	0.0000108940 ***
GRUPO_ETAREO_RUTA_AGREGADO7	-0.41268	0.05888	-7.009	0.0000092109 ***
GRUPO_ETAREO_RUTA_AGREGADO8	-0.39119	0.05888	-6.644	0.0000160173 ***
GRUPO_ETAREO_RUTA_AGREGADO9	-0.39413	0.05888	-6.694	0.0000148341 ***
SEXOM	0.20543	0.02225	9.232	0.0000004513 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.003466338)

Null deviance: 1.035134 on 27 degrees of freedom  
 Residual deviance: 0.045127 on 13 degrees of freedom  
 AIC: 538.52

Number of Fisher Scoring iterations: 4

Figura 264: Modelo Gamma para Severidad Zona Normal Medio Año 2018

En el modelo Gamma las variables son significativas al 95 %, por lo cual se analizan los residuales con el fin de realizar la proyección de severidad, para la zona normal medio año 2018.

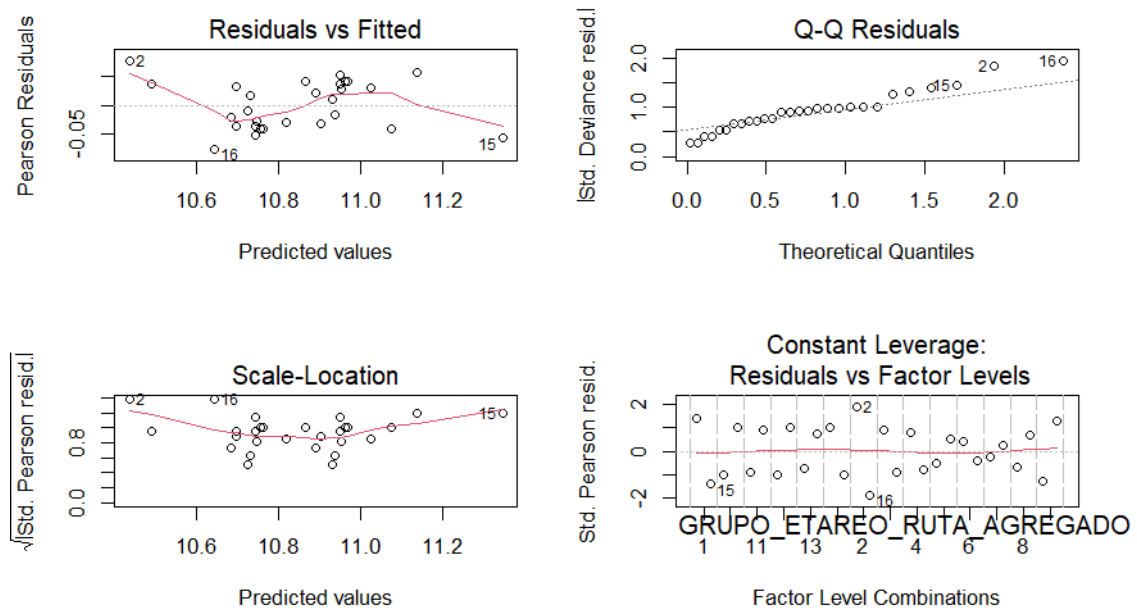


Figura 265: Residuales Modelo Gamma para Severidad Zona Normal Medio Año 2018

Como se observa en la figura 265, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla 97 y figura 266, presentando puntos influyentes en las observaciones 2,4,5 y 16, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
2	2.0935101	0.5357143	0.27594585
4	0.7539883	0.5357143	0.04574766
5	-0.5176654	0.5357143	0.02172574
16	-2.1853425	0.5357143	0.27594588

Tabla 97: Puntos Influyentes Distribución Gamma Zona Normal Medio Año 2018

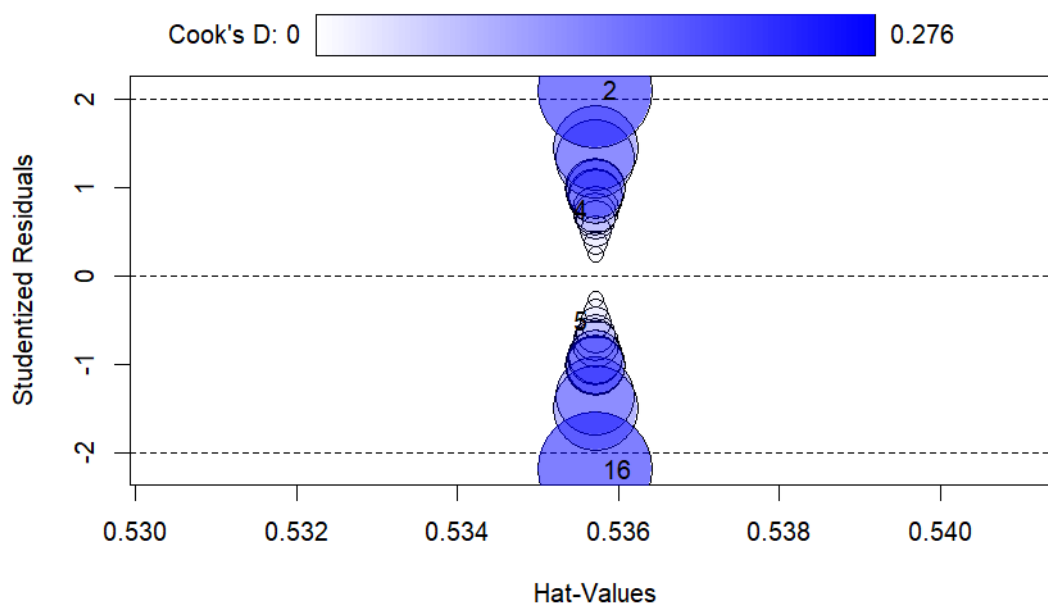


Figura 266: Puntos Influyentes Distribución Gamma Zona Normal Medio Año 2018

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la severidad esperada para el año 2018, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	76.152,84	76.642,11
Entre 1 y 5 Años	Ambos	37.796,98	38.157,58
Entre 6 y 11 Años	Ambos	40.004,89	40.185,79
Entre 12 y 17 Años	Femenino	45.667,95	44.297,50
Entre 12 y 17 Años	Masculino	52.716,60	54.399,59
Entre 18 y 28 Años	Femenino	42.821,71	43.754,56
Entre 18 y 28 Años	Masculino	54.878,41	53.732,83
Entre 29 y 44 Años	Femenino	46.523,72	45.787,41
Entre 28 y 44 Años	Masculino	55.325,05	56.229,28
Entre 45 y 49 Años	Ambos	50.945,37	50.869,38
Entre 50 y 54 Años	Ambos	52.075,69	51.901,89
Entre 55 y 59 Años	Ambos	51.914,89	51.633,89
Entre 60 y 64 Años	Ambos	52.307,24	52.150,12
Entre 65 y 69 Años	Ambos	51.656,23	51.520,67
Entre 70 y 74 Años	Ambos	52.439,16	52.324,29
Entre 75 y 79 Años	Ambos	55.251,59	55.213,36
Mayores de 80 Años	Ambos	57.675,06	57.597,42

Tabla 98: Severidad Observada y Estimada Modelo Gamma Zona Normal Medio Año 2018

A continuación se presenta gráficamente la severidad observada y la severidad estimada del modelo Gamma para el año 2018.

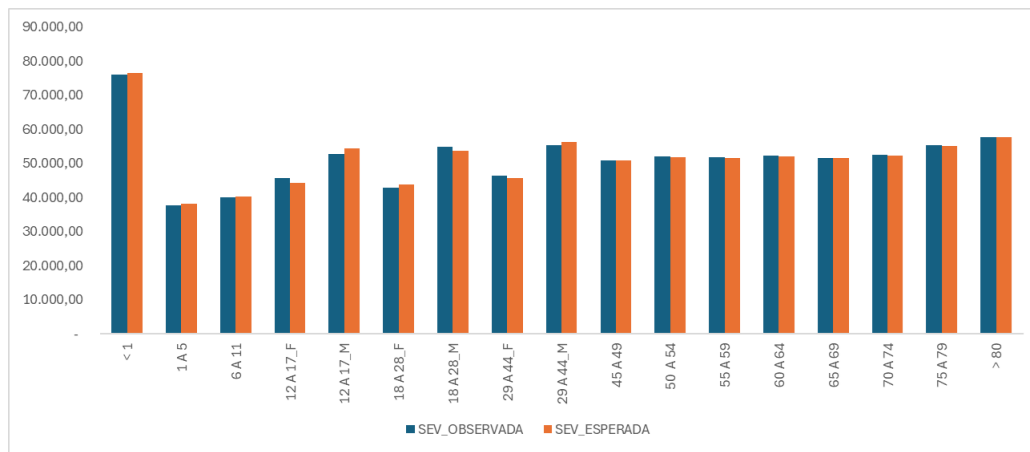


Figura 267: Severidad Observada Vs Severidad Estimada Zona Normal Medio Año 2018

El modelo presenta un error porcentual absoluto medio (MAPE) del 1.04 % lo cual se presenta un ajuste aceptable, y un pseudo R2 de 0.95

Para el año 2019 y de acuerdo lo presentado en las distribuciones empíricas de severidad, la distribución gamma no ajusta de manera aceptable por lo cual para este año no se realiza el modelo utilizando la distribución Gamma.

#### A.4.4. Zona Rural Disperso

A continuación se presentan los resultados de la aplicación del modelo LogNormal para el año 2019, para modelar la severidad de la Zona Rural Disperso.

```

GAMLSS-RS iteration 1: Global Deviance = 562.761
GAMLSS-RS iteration 2: Global Deviance = 562.761
*****
Family: c("LOGNO", "Log Normal")

Call: gamlss(formula = COSTO_MEDIO ~ GRUPO_ETAREO_RUTA_AGREGADO +
  SEXO, family = LOGNO(mu.link = "log"), data = BASE_2019_RUD)

Fitting method: RS()

-----
Mu link function: log
Mu Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.3627246  0.0087736 269.300 < 2e-16 ***
GRUPO_ETAREO_RUTA_AGREGADO10 0.0108853  0.0119167   0.913 0.378994
GRUPO_ETAREO_RUTA_AGREGADO11 0.0119499  0.0119104   1.003 0.335513
GRUPO_ETAREO_RUTA_AGREGADO12 0.0045121  0.0119544   0.377 0.712437
GRUPO_ETAREO_RUTA_AGREGADO13 0.0147635  0.0118939   1.241 0.238224
GRUPO_ETAREO_RUTA_AGREGADO14 0.0128241  0.0119055   1.077 0.302589
GRUPO_ETAREO_RUTA_AGREGADO2  -0.0596332  0.0123550  -4.827 0.000414 ***
GRUPO_ETAREO_RUTA_AGREGADO3  -0.0298331  0.0121641  -2.453 0.030451 *
GRUPO_ETAREO_RUTA_AGREGADO4   0.0076063  0.0119364   0.637 0.535942
GRUPO_ETAREO_RUTA_AGREGADO5  -0.0153325  0.0120743  -1.270 0.228210
GRUPO_ETAREO_RUTA_AGREGADO6  -0.0030051  0.0119994  -0.250 0.806489
GRUPO_ETAREO_RUTA_AGREGADO7   0.0001439  0.0119806   0.012 0.990617
GRUPO_ETAREO_RUTA_AGREGADO8   0.0007640  0.0119768   0.064 0.950186
GRUPO_ETAREO_RUTA_AGREGADO9   0.0141643  0.0118974   1.191 0.256856
SEXOM                0.0159088  0.0045375   3.506 0.004333 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
Sigma link function: log
Sigma Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.0537     0.1336  -15.37 2.94e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

-----
No. of observations in the fit: 28
Degrees of Freedom for the fit: 16
  Residual Deg. of Freedom: 12
                        at cycle: 2

Global Deviance:   562.761
      AIC:         594.761
      SBC:         616.0763
*****

```

Figura 268: Modelo LogNormal para Severidad Zona Rural Disperso Año 2019

El modelo LogNormal presenta 4 de las variables significativas al 95 %, sin embargo no es posible descartar alguno de los grupos etarios por lo cual se presenta a continuación los residuales

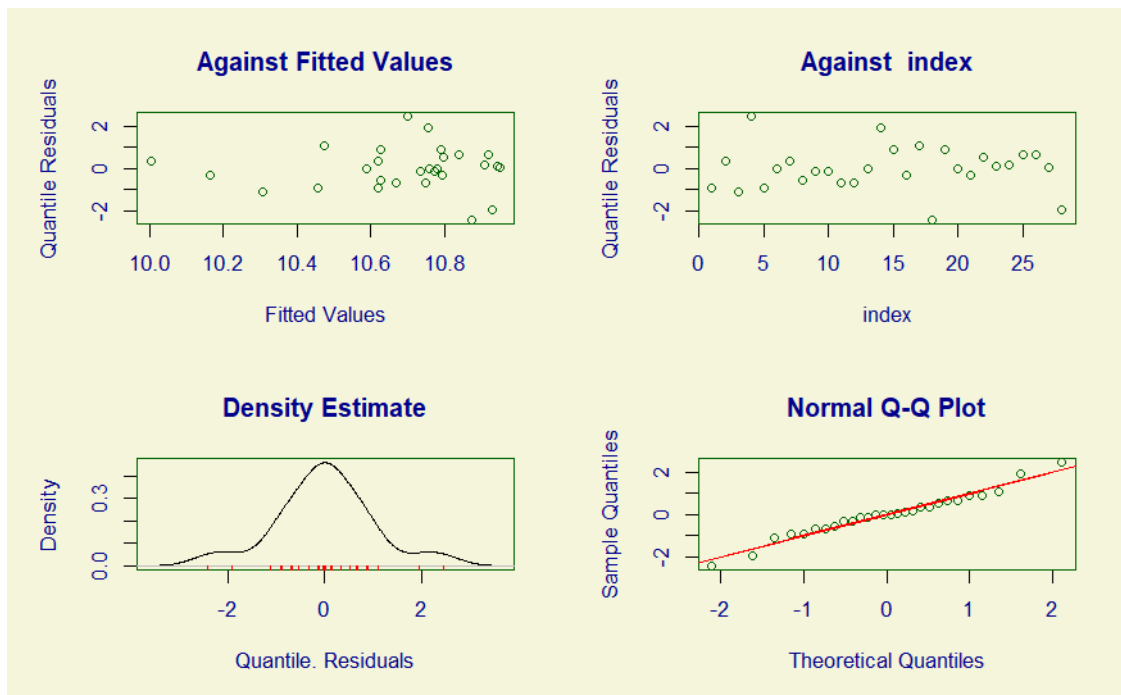


Figura 269: Residuales Modelo LogNormal para Severidad Zona Rural Disperso Año 2019

Como se observa en la figura numero 269, en el gráfico de residuales frente a las predicciones no se presenta patrón en los datos lo cual se corrobora en el gráfico de nivel de dispersión dado que la pendiente presentada es mínima.

Al observar el qq plot presenta un buen ajuste, la mayoría de los puntos se encuentran cercanos a la diagonal, finalmente se observan puntos influyentes, con el fin de detectar dichos puntos se presenta la siguiente tabla numero 99 y figura numero 270, presentando puntos influyentes en las observaciones 4,18,23 y 27, sin embargo dado el objetivo del presente trabajo no es dable establecer un modelo sin estas observaciones.

Observación	StudRes	Hat	CookD
4	3.17462861	0.5290154	4.443596e-01
18	-3.17462861	0.5437651	4.715153e-01
23	0.10812341	0.5442379	1.007252e-03
27	0.02479374	0.5442815	5.302242e-05

Tabla 99: Puntos Influyentes Distribución LogNormal Zona Rural Disperso 2019

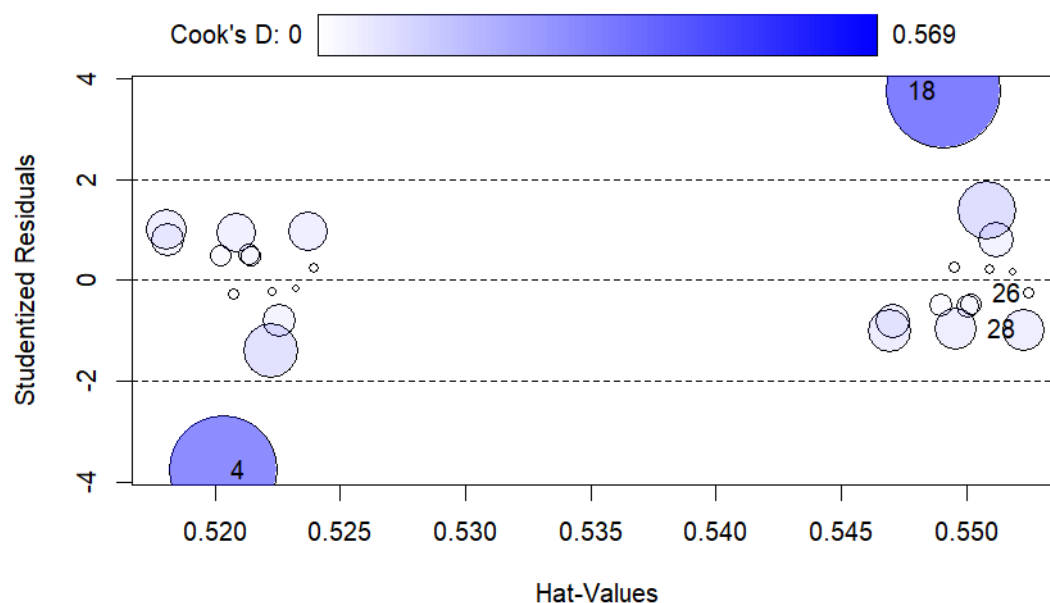


Figura 270: Puntos Influyentes Distribución LogNormal Zona Rural Disperso 2019

Ahora bien, teniendo en cuenta en análisis de los residuales presentados se realiza la proyección de la severidad esperada para el año 2019 en la zona rural disperso, cuyos resultados se presenta a continuación.

Descripción Grupo Etario	Sexo	Severidad Observada	Severidad Estimada
Menor de un año	Ambos	45.563,92	44.789,49
Entre 1 y 5 Años	Ambos	24.031,55	24.103,02
Entre 6 y 11 Años	Ambos	33.547,12	32.729,48
Entre 12 y 17 Años	Femenino	60.817,60	44.397,22
Entre 12 y 17 Años	Masculino	38.669,14	52.708,53
Entre 18 y 28 Años	Femenino	31.076,70	34.830,88
Entre 18 y 28 Años	Masculino	46.083,71	41.190,74
Entre 29 y 44 Años	Femenino	39.601,28	39.655,36
Entre 29 y 44 Años	Masculino	47.056,95	46.993,79
Entre 45 y 49 Años	Ambos	45.080,61	45.597,61
Entre 50 y 54 Años	Ambos	46.860,59	45.887,40
Entre 55 y 59 Años	Ambos	53.205,41	52.995,89
Entre 60 y 64 Años	Ambos	51.381,93	51.115,49
Entre 65 y 69 Años	Ambos	52.512,48	51.442,60
Entre 70 y 74 Años	Ambos	47.754,28	47.100,47
Entre 75 y 79 Años	Ambos	52.926,75	52.896,44
Mayores de 80 Años	Ambos	52.376,29	51.153,97

Tabla 100: Severidad Observada y Estimada Modelo LogNormal Zona Rural Disperso Año 2019

A continuación se presenta gráficamente la comparación entre la severidad observada y la severidad estimada con el modelo LogNormal para la zona Rural Disperso.

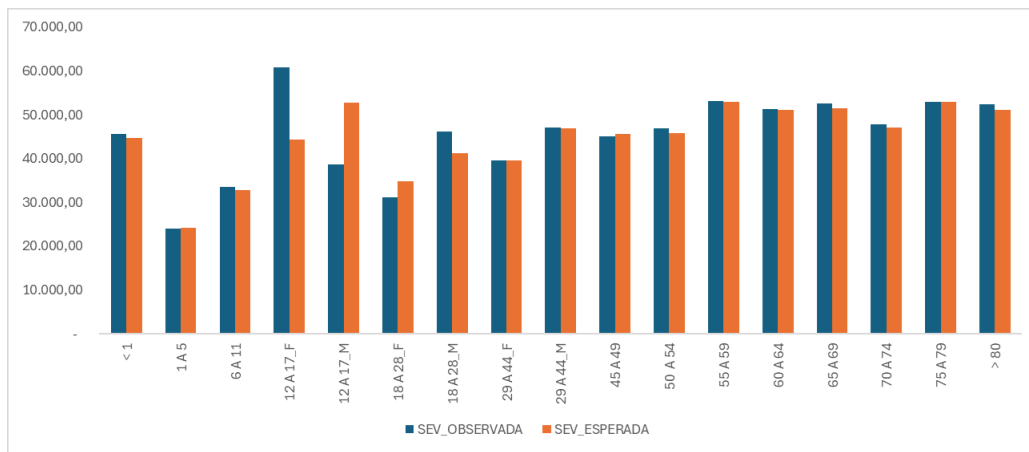


Figura 271: Severidad Observada vs Severidad Modelo LogNormal Estimada Zona Rural Disperso Año 2019

El MAPE presentado es de 5.92% y un Pseudo R2 de 0.75.

Ahora bien la elección del modelo para el año 2019 se hará teniendo en cuenta el menor MAPE en este caso lo presenta la distribución Gamma.